The Form of the Optimal Nonlinear Instrument for Multiperiod Conditional Moment Restrictions

by

Stanislav Anatolyev*

New Economic School, Moscow

Phone: +7 095 129-1700

Fax: +7 095 129-3722

*This is a part of my PhD dissertation (Anatolyev, 2000). I am grateful to Kenneth West for introducing me to the subject and providing advice and encouragement. I also thank Bruce Hansen, Yuichi Kitamura, seminar participants at Indiana University, Pennsylvania State University, Rutgers University, University of British Columbia, University of Virginia and University of Wisconsin, the co-editor Richard Smith and two referees for helpful comments and suggestions. Research for this paper was supported in part by a fellowship from the Economics Education and Research Consortium, with funds provided by the Government of Sweden through the Eurasia Foundation. The opinions expressed in this paper are those of the author and do not necessarily reflect the views of the Government of Sweden, the Eurasia Foundation, or any other member of the Economics Education and Research Consortium.

Running head: FORM OF OPTIMAL INSTRUMENT

Author's address:

Stanislav Anatolyev

New Economic School

Nakhimovsky Prospekt, 47

Moscow, 117418 Russia

E-mail: sanatoly@nes.ru

Abstract

This note presents the form of the optimal instrument in a system of multi-period conditional moment restrictions in the presence of conditional heteroskedasticity. Using Hansen's (1985) and Hansen, Heaton and Ogaki's (1988) work on efficiency bounds for GMM estimators, we show that this form is an autoregressive recurrence parametrized by a few auxiliary processes that are defined through a system of non-linear stochastic restrictions, with a stability condition among them. In general, the system does not allow inversion and obtaining an explicit solution for the auxiliary parameters.

1 Introduction

A variety of intertemporal macroeconomic and financial models give rise to multiperiod conditional moment restrictions, where the moment function is serially correlated. Such restrictions typically result in models with overlapping prediction horizons (e.g. Hansen and Hodrick, 1980), with temporal aggregation (e.g. Hall, 1988), or with complex decision rules (e.g. Eichenbaum, Hansen and Singleton, 1988). An applied researcher typically estimates such models by converting them into an overidentified system of unconditional restrictions after having chosen a set of instruments, with successive application of the Generalized Method of Moments (Hansen, 1982). However, the set of possible instruments is infinite, which raises a question of their optimal choice. When the moment function has a martingale difference property, it is well known that the optimal instrument is an explicit function of certain conditional expectations as in Chamberlain (1987). In the context of multiperiod restrictions the conditions for an instrument to be optimal are significantly more complicated. For the special case of conditional homoskedasticity, Hansen (1985) derived the optimal instrument which takes the form of a certain recurrence relation. This optimal instrument was put to the test in Hansen and Singleton (1996), West and Wilcox (1996), and a few other papers.

For models where both serial correlation and conditional heteroskedasticity are in effect, Hansen (1985) and Hansen, Heaton and Ogaki (1988), using Gordin's (1969) martingale difference approximation and Hayashi and Sims' (1983) forward filtering idea, presented a characterization of the efficiency bound for GMM estimators that correspond to a given system of conditional moment restrictions. In this note, we obtain a more algorithmic description of Hansen, Heaton and Ogaki's (1988) result and present the form taken by the optimal instrument. The process followed by the optimal instrument turns out to be a recurrence relation that extends Hansen's formula (Hansen, 1985, lemma 5.7). The recurrence is parameterized by a few auxiliary processes whose law of motion is not explicit, but instead solves a system of nonlinear functional equations. In rare circumstances, it is possible to solve this system analytically, as in Heaton and Ogaki's (1991) example, but this is not typical.

The note is organized as follows. Section 2 presents a formal setup of the problem and a useful result. Section 3 derives the process for the optimal instrument. Section 4 concludes. The proofs are in the Appendix.

2 Formulation of the problem

We consider a multiple equation model

$$\mathbf{f}(\boldsymbol{\beta}, \mathbf{x}_t) = \mathbf{e}_t, \tag{1}$$

where $\mathbf{e}_t$ is an $s \times 1$ vector of errors, $\mathbf{x}_t$ is a vector of observable variables, $\boldsymbol{\beta}$ is a $k \times 1$ vector of parameters to be estimated, and $\mathbf{f}(\boldsymbol{\beta}, \mathbf{x}_t)$ is a known up to $\boldsymbol{\beta}$ function which is possibly nonlinear in $\boldsymbol{\beta}$. In addition, we are given a vector $\mathbf{z}_t$ of observable *basic instruments* (as opposed to just *instruments* that may be generated from the basic ones). Some of $\mathbf{x}_t$'s may be among $\mathbf{z}_t$. Let us denote by $\Im_t$ the information embedded in $\mathbf{z}_t$ and all its history, i.e. $\Im_t \equiv \sigma(\mathbf{z}_t, \mathbf{z}_{t-1}, \ldots)$, and use notation $E_t[\cdot] \equiv E[\cdot|\Im_t]$. The conditional moment restriction[1]

$$E_t\left[\mathbf{e}_t\right] = \mathbf{0} \tag{2}$$

implies that all measurable functions of the basic instrument and their lags are valid instruments. This condition does not preclude correlatedness of the error with the leads of the basic instrument, since the latter may be endogenous (Hayashi and Sims, 1983). As a result, one generally cannot hope that the Wold innovation in $\mathbf{e}_t$ has a martingale difference property. This precludes backward filtering of (1) with subsequent taking care of heteroskedasticity. Define the $k \times s$ matrix

$$\mathrm{D}_t \equiv E_t\left[\frac{\partial \mathbf{f}(\boldsymbol{\beta}, \mathbf{x}_t)'}{\partial \boldsymbol{\beta}}\right]. \tag{3}$$

We use the Euclidean norm $|\mathbf{a}| \equiv \sqrt{\sum_i \mathbf{a}_i^2}$ for vectors and the induced spectral norm $|\mathrm{A}| = \sqrt{\varrho\left(\mathrm{A}'\mathrm{A}\right)}$, where $\varrho\left(\cdot\right)$ is the spectral radius, for matrices.[2]

We consider the class of instrumental variables (IV) estimators, indexed by $\Im_t$-measurable instruments. Let $\mathcal{Z}_t$ be the space of all $\Im_t$-measurable $k \times s$ matrices Z_t with finite $E\left[|\mathrm{Z}_t|^4\right]$. We are interested in the best choice of the instrument from $\mathcal{Z}_t$ in the sense of asymptotic theory:

Definition 1 *The instrument $\Xi_t \in \mathcal{Z}_t$ is called **optimal relative to** $\mathcal{Z}_t$ if the asymptotic covariance matrix of the corresponding IV estimator $\hat{\boldsymbol{\beta}}_\Xi$ does not exceed that of the IV estimator $\hat{\boldsymbol{\beta}}_\mathrm{Z}$ corresponding to any other instrument $\mathrm{Z}_t \in \mathcal{Z}_t$.*

The instrument CZ_t has the same asymptotic properties as Z_t for any nonsingular nonrandom $k \times k$ matrix C. Thus, if Ξ_t is optimal, then $\mathrm{C}\Xi_t$ is too for any such C, so the object of investigation represents an equivalence class. We make the following set of assumptions on $\mathbf{x}_t$, $\mathbf{z}_t$, $\Im_t$, $\boldsymbol{\beta}$, $\mathbf{f}(\cdot,\cdot)$, $\mathbf{e}_t$ and D_t.

Assumption 1 (Data-generating mechanism and regularity conditions)

(1) $\mathbf{x}_t$ *and* $\mathbf{z}_t$ *are jointly strictly stationary ergodic processes and* $E\left[|\mathbf{z}_t|^4\right]$ *is finite;*

(2) $\boldsymbol{\beta} \in int\,(\mathbb{B})$ *is a vector of parameters to be estimated, where* $\mathbb{B} \subseteq \mathbb{R}^k$ *is compact;*

(3) $\mathbf{f}(\boldsymbol{\beta}, \mathbf{x}_t)$ *is a Borel measurable function for all* $\boldsymbol{\beta} \in \mathbb{B}$ *and is continuously differentiable in the first argument for all* $\boldsymbol{\beta} \in \mathbb{B}$ *for all* $\mathbf{x}_t$ *in its support;*

(4) $E\left[|\mathrm{D}_t|^2\right]$ *is finite and* $E\left[\mathrm{D}_t\mathrm{D}_t'\right]$ *is of rank* k*;*

(5) $E\left[\sup_{|\mathbf{b}-\boldsymbol{\beta}|<\delta}\left|\frac{\partial \mathbf{f}(\mathbf{b},\mathbf{x}_t)}{\partial \mathbf{b}'} - \frac{\partial \mathbf{f}(\boldsymbol{\beta},\mathbf{x}_t)}{\partial \boldsymbol{\beta}'}\right|^2\right] < \infty$ *for some* $\delta > 0$.

Later we will add assumptions about the structure of the error $\mathbf{e}_t$. Assumptions 1(1)-1(3) are self-explanatory. Assumption 1(4) guarantees local identification of $\boldsymbol{\beta}$. From assumption 1(5) it follows that $\mathrm{Z}_t\mathbf{f}(\mathbf{b}, \mathbf{x}_t)$ is first-moment continuous for all Z_t from $\mathcal{Z}_t$ at all $\mathbf{b} \in \mathbb{B}$, which is needed to apply Hansen's (1982) GMM theory.

We will need the following terminology and a result from the theory of random matrices and a so called *generalized autoregression* for a $k \times m$ matrix process Ψ_t:

$$\Psi_t = \Psi_{t-1}\mathrm{A}_t + \mathrm{B}_t, \tag{4}$$

where A_t is $m \times m$ and B_t – $k \times m$ matrix processes. Denote $\max(a, 0)$ by a^+.

Definition 2 *The **top (dominant) Lyapunov exponent** associated with the sequence* $\{\mathrm{A}_t\}$ *is defined as*

$$\lambda(\mathrm{A}) = \lim_{T\to\infty} \frac{1}{T} \log |\mathrm{A}_T \mathrm{A}_{T-1} \cdots \mathrm{A}_2 \mathrm{A}_1| \,. \tag{5}$$

Lemma 1 *Suppose that a* $k \times m$ *matrix process* Ψ_t *satisfies the recurrence (4), and:*

(1) A_t *and* B_t *are stationary and ergodic;*

(2) $E\left[(\log|\mathrm{A}_t|)^+\right]$ *and* $E\left[(\log|\mathrm{B}_t|)^+\right]$ *are finite;*

(3) the top Lyapunov exponent associated with $\{\mathrm{A}_t\}$ *satisfies* $\lambda(\mathrm{A}) < 0$.

Then there exists a unique stationary ergodic solution Ψ_t *of (4).*

3 Main result

Let us define $p+1$ fundamental quantities

$$\Omega_t \equiv E_t[\mathbf{e}_t\mathbf{e}_t'], \ \Gamma_{1,t} \equiv E_t[\mathbf{e}_{t-1}\mathbf{e}_t'], \ \cdots, \ \Gamma_{p,t} \equiv E_t[\mathbf{e}_{t-p}\mathbf{e}_t'] \tag{1}$$

to be used throughout. These strictly stationary and ergodic $\Im_t$ - measurable processes may be viewed as $s \times s$ infinite-dimensional parameters that index conditional heteroskedasticity, Ω_t being the conditional variance, and each $\Gamma_{j,t}$ – the conditional j^{th} order autocovariance of the errors. We assume that $\mathbf{e}_t$ satisfies

Assumption 2 (Error term) *The error* $\mathbf{e}_t$ *satisfies (2), has finite* $E[|\mathbf{e}_t|^4]$, *and*

(1) $\mathbf{e}_t$ *is conditionally* p^{th} *order serially correlated:* $E_t[\mathbf{e}_{t-j}\mathbf{e}_t'] = 0$ *for* $j > p$;

(2) $\operatorname{essinf} \lambda_{\min}(\Gamma_{p,t}\Gamma_{p,t}') > 0$, *where* $\lambda_{\min}(\mathrm{A})$ *denotes the smallest eigenvalue of* A.

Assumption 2(1) says that $\mathbf{e}_t$ is a $(p+1)$-period ahead forecast error. Assumption 2(2) means that the p^{th} order of conditional serial correlation does not come too close to order $p-1$ for any realizations of $\Im_t$. The following theorem obtains a parametrization of the process followed by one of optimal instruments, given that there are such in $\mathcal{Z}_t$.

Theorem 1 *Let the model satisfy Assumptions 1 and 2. Consider the following recurrence relation:*

$$\Xi_t = \Xi_{t-1}\Phi_{1,t} + \Xi_{t-2}\Phi_{2,t} + \cdots + \Xi_{t-p}\Phi_{p,t} + \Delta_t \mathrm{P}_t, \tag{2}$$

where the stationary ergodic $\Im_t$ *- measurable processes,* $s \times s$ *matrix* $\Phi_{1,t}$, $\Phi_{2,t}$, $\cdots$, $\Phi_{p,t}$, $k \times s$ *matrix* Δ_t, *and symmetric almost surely positive definite* $s \times s$ *matrix* P_t, *satisfy the following system:*

$$\begin{pmatrix}\Gamma_{1,t}' & \cdots & \Gamma_{p,t}'\end{pmatrix} + \Omega_t \mathrm{J}\Phi_t' + E_t\left[\begin{pmatrix}\Gamma_{p,t+p} & \cdots & \Gamma_{1,t+1}\end{pmatrix}\Phi_{t+p}'\cdots\Phi_{t+1}'\right]\Phi_t' = \mathrm{O}_{s\times ps}, \tag{3}$$

$$\mathrm{P}_t = -\Gamma_{p,t}^{-1}\Phi_{p,t}, \tag{4}$$

$$\Delta_t = \mathrm{D}_t - \sum_{j=1}^{p} E_t\left[\Delta_{t+j}\mathrm{P}_{t+j}\mathrm{J}\Phi_{t+j+1}\cdots\Phi_{t+p}\begin{pmatrix}\Gamma_{p,t+p} & \cdots & \Gamma_{1,t+1}\end{pmatrix}'\right], \tag{5}$$

where $\underset{ps\times ps}{\Phi_t} = \begin{pmatrix} \Phi_{1,t} & \\ \Phi_{2,t} & \mathrm{I}_{(p-1)s} \\ \cdots & \\ \Phi_{p,t} & \mathrm{O}_{s\times(p-1)s} \end{pmatrix}$, $\mathrm{J} = \begin{pmatrix}\mathrm{I}_s & \mathrm{O}_{s\times(p-1)s}\end{pmatrix}$, *and for integers* m *and* n, I_m *is* $m \times m$ *identity matrix, and* $\mathrm{O}_{m\times n}$ *is* $m \times n$ *zero matrix. If a solution to (3)–(5) satisfies* $\lambda(\Phi) < 0$, $E\left[(\log|\Delta_t|)^+\right] < \infty$, *then there exists unique stationary solution* Ξ_t *of (2). If in addition* $E\left[|\Xi_t|^4\right] < \infty$, *then* Ξ_t *is optimal relative to* $\mathcal{Z}_t$.

Theorem 1 does not address the issue of existence of optimal instruments or solutions to the system and implicitly presumes existence of all needed conditional expectations. Making these conditions primitive is problematic due to a highly nonlinear environment and essentially requires knowledge of the solution to (3)–(5). The proof of Theorem 1 proceeds in two steps. We obtain the equation (3) from Hansen, Heaton and Ogaki's (1988) conditional forward moving-average representation for the error term. Then we "solve out" Hansen's (1985) optimality condition treating it as a stochastic difference equation of order $2p+1$ and obtain (2), (4) and (5). For details, see the Appendix.

The key relation is (2). It is an extention of Hansen's (1985, lemma 5.7) formula for the process followed by the optimal instrument in a homoskedastic environment. Here, in contrast to Hansen (1985), Φ_t is time varying, and the product $\Delta_t \mathrm{P}_t$ is a generalization of the projection of the discounted sum of future D_t-variables onto the space of instruments. The condition $\lambda(\Phi) < 0$ rules out unstable solutions of the nonlinear equation (3). For example, if in the one-equation MA(1) case we have (using small plain font for scalar quantities) $e_t = \varepsilon_{t+1} - \theta\varepsilon_t$, where $\theta \in (-1, 1)$, $\{\varepsilon_t\}$ is a martingale difference sequence relative to $\Im_t$ with $\sigma_t^2 \equiv E_t[\varepsilon_t^2]$, then $\gamma_{1,t} = -\theta\sigma_t^2$, $\omega_t = E_t[\sigma_{t+1}^2] + \theta^2\sigma_t^2$, equation (3) can be rewritten as

$$(1 - \theta\phi_t)\sigma_t^2 + \phi_t E_t[(\phi_{t+1} - \theta^{-1})\sigma_{t+1}^2] = 0,$$

and one notices immediately that $\phi_t = \theta^{-1}$ for all t is an unstable solution. Unfortunately, the stable one is *not* (except under homoskedasticity or other special circumstances) simply $\phi_t = \theta$. Moreover, in general $|\phi_t|$ may exceed unity for a set of realizations of $\Im_t$ of nonzero measure. One case where the condition $\lambda(\Phi) < 0$ fails is: conditional homoskedasticity, unit roots in $\mathbf{e}_t$.

4 Conclusion

We have presented the form of the process followed by the optimal instrument in a system of multiperiod conditional moment restrictions in the presence of conditional heteroskedasticity. Using Hansen's (1985) and Hansen, Heaton and Ogaki's (1988) work on efficiency bounds for GMM estimators, we have shown that this form is a recurrence parametrized by a few auxiliary processes that are defined through a system of nonlinear stochastic restrictions, with a stability condition among them. In general, the system does not allow inversion and obtaining an explicit solution for the auxiliary parameters.

Further research will show how to implement this result. The construction of a feasible instrument should take the derived form as a point of departure and attempt to either estimate the auxiliary processes directly from the system by designing a contractive iterative scheme, or by approximating the system in such a way that an explicit though approximate solution can be obtained and estimated. Either way requires nonparametric estimation of conditional expectations that may depend on entire history of the state vector.

Notes

1. Chances are that (2) is an implication of the decision rule rather than the decision rule itself, because the decision maker may use information beyond that in $\Im_t$.

2. We use the spectral norm instead of the more usual $\sqrt{\mathrm{tr}(\mathrm{A}'\mathrm{A})}$ because it is desirable to keep $|\mathrm{A}|$ smaller.

References

Anatolyev, S. (2000) *Optimal Instrumental Variables Estimation in Stationary Time Series.* PhD dissertation, University of Wisconsin-Madison.

Bougerol, P. & N. Picard (1992) Strict Stationarity of Generalized Autoregressive Processes. *Annals of Probability* 20, 1714–1730.

Brandt, A. (1986) The Stochastic Equation $Y_{n+1} = A_n Y_n + B_n$ with Stationary Coefficients. *Advances in Applied Probability* 18, 211–220.

Chamberlain, G. (1987) Asymptotic Efficiency in Estimation with Conditional Moment Restrictions. *Journal of Econometrics* 34, 305–334.

Eichenbaum, M.S., L.P. Hansen & K.J. Singleton (1988) A Time Series Analysis of Representative Agent Models of Consumption and Leisure Choice Under Uncertainty. *Quarterly Journal of Economics* 103, 51–78.

Gordin, M.I. (1969) The Central Limit Theorem for Stationary Processes. *Soviet Mathematics Doklady* 10, 1174–1176.

Hall, R.E. (1988) Intertemporal Substitution in Consumption. *Journal of Political*

Economy 96, 339–357.

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

Hansen, L.P. (1985) A Method for Calculating Bounds on the Asymptotic Variance-Covariance Matrices of Generalized Method of Moments Estimators. *Journal of Econometrics* 30, 203–228.

Hansen, L.P., J.C. Heaton & M. Ogaki (1988) Efficiency Bounds Implied by Multiperiod Conditional Moment Restrictions. *Journal of the American Statistical Association* 83, 863–871.

Hansen, L.P & R.J. Hodrick (1980) Forward Exchange Rates as Optimal Predictors of Future Spot Rates: An Econometric Analysis. *The Journal of Political Economy* 88, 829–853.

Hansen, L.P. & K.J. Singleton (1996) Efficient Estimation of Linear Asset Pricing Models with moving-Average Errors. *Journal of Business and Economic Statistics* 14, 53–68.

Hayashi, F. & C. Sims (1983) Nearly Efficient Estimation of Time Series Models with Predetermined, but Not Exogenous, Instruments. *Econometrica* 51, 783–798.

Heaton, J.C. & M. Ogaki (1991) Efficiency Bounds Calculations for a Time Series Model, with Conditional Heteroskedasticity. *Economics Letters* 35, 167–171.

West, K.D. & D.W. Wilcox (1996) A Comparison of Alternative Instrumental Variables Estimators of a Dynamic Linear Model. *Journal of Business and Economic Statistics* 14, 281–293.

A Appendix

Proof of Lemma 1: See Brandt (1986) for a proof in the unidimensional case $k = s = 1$, and Bougerol and Picard (1992) for a generalization. ■

Proof of Theorem 1: Assumptions 1–3 of Hansen *et al* (1988) are implied by assumptions 1(1), 1(3) and 1(4). Let us first assume that assumption 7 of Hansen *et al* (1988) is also satisfied. Then, by lemma 4.1 of Hansen *et al* (1988) there exist $\Im_t$-measurable stationary ergodic $s \times s$ matrix processes W_t, $V_{1,t}$, $\cdots$, $V_{p,t}$ such that $E[|W_t|^2] < \infty$, $E[|V_{1,t}|^2] < \infty$, ...,$E[|V_{p,t}|^2] < \infty$, $P\{|W_t| = 0\} = 0$, and

$$\mathbf{e}_t = W_t \boldsymbol{\eta}_t + \sum_{j=1}^{p} V_{j,t+1} \boldsymbol{\eta}_{t+j},$$

where $\boldsymbol{\eta}_t$ is the normalized forward innovation with $E_t[\boldsymbol{\eta}_t] = \mathbf{0}$ and $E_t[\boldsymbol{\eta}_t \boldsymbol{\eta}_t'] = \mathrm{I}_s$. If assumption 7 of Hansen *et al* (1988) is not satisfied, from their proof of lemma 4.1 it still follows that $\mathbf{e}_t = \boldsymbol{\eta}_t^* + \sum_{j=1}^{p} \boldsymbol{\eta}_{j,t+j}^{**}$, where $\boldsymbol{\eta}_t^*$ is an unnormalized forward innovation such that $E_t[\boldsymbol{\eta}_t^*] = \mathbf{0}$, but now $P\{|\boldsymbol{\eta}_t^*| = 0\} > 0$. However, this contradicts our assumption 2(2), and hence can be ruled out.

Matching the 1^{st} through p^{th} conditional moments of e_t gives the following system:

$$\begin{aligned}
\Omega_t &= W_t W_t' + E_t\left[V_{1,t+1}V_{1,t+1}' + \cdots + V_{p,t+p}V_{p,t+p}'\right], \\
\Gamma_{1,t}' &= W_t V_{1,t}' + E_t\left[V_{1,t+1}V_{2,t+1}' + \cdots + V_{p-1,t+p-1}V_{p,t+p-1}'\right], \\
&\cdots \\
\Gamma_{p,t}' &= W_t V_{p,t}'.
\end{aligned}$$

Let us define $\Phi_{j,t} \equiv -V_{j,t} W_t^{-1}$ for $j = 1, \cdots, p$ (W_t is nonsingular almost surely).

Then each $\Phi_{j,t}$ is a stationary ergodic $\Im_t$-measurable stochastic process. By post-multiplying $E_{t+p}\left[\left(\Gamma_{p,t+p}\ \cdots\ \Gamma_{1,t+1}\right)\right]$ by $\Phi'_{t+p}, \cdots, \Phi'_{t+1}$ and Φ'_t, and taking $E_t\left[\cdot\right]$ of what emerges, it is straightforward to see that it boils down to $-\left(\Gamma'_{1,t}\ \cdots\ \Gamma'_{p,t}\right) - \Omega_t \mathrm{J}\Phi'_t$, hence (3) follows.

Define $\mathrm{P}_t = -\Gamma_{p,t}^{-1}\Phi_{p,t}$. It follows that $\mathrm{P}_t = \left(W_t W'_t\right)^{-1}$, an almost surely symmetric positive definite matrix. Note also that $E\left[\log\left|\mathrm{P}_t\right|\right] \leq 2E\left[\log\left|W_t^{-1}\right|\right] \leq 2E\left[\log\left|\Gamma_{p,t}^{-1}\right|\right] + 2E\left[\log\left|V_{p,t}\right|\right]$, and $2E\left[\log\left|\Gamma_{p,t}^{-1}\right|\right] \leq 2\log\operatorname{esssup}\left|\Gamma_{p,t}^{-1}\right| = \log\operatorname{esssup}\left|\lambda_{\max}\left(\left(\Gamma_{p,t}\Gamma'_{p,t}\right)^{-1}\right)\right|$ $= -\log\operatorname{essinf}\left|\lambda_{\min}\left(\Gamma_{p,t}\Gamma'_{p,t}\right)\right| < \infty$, $E\left[\log\left|V_{p,t}\right|\right] \leq E[|V_{p,t}|^2] < \infty$. Consider Δ_t defined by (5). Since $E[(\log|\Delta_t \mathrm{P}_t|)^+] \leq E[(\log|\Delta_t|)^+] + E[(\log|\mathrm{P}_t|)^+] < \infty$, by lemma 1 applied to the companion form of (2), the matrix $(\Xi_t\ \Xi_{t-1}\ \cdots\ \Xi_{t-p+1})'$ is a uniquely defined stationary and ergodic process, and so is the matrix Ξ_t.

Consider the FOC (Hansen, 1985; Hansen *et al*, 1988) for optimality of Ξ_t:

$$E[\mathrm{Z}_t \mathrm{D}'_t] = E[\mathrm{Z}_t\Gamma'_{p,t}\Xi'_{t-p}] + \cdots + E[\mathrm{Z}_t\Gamma'_{1,t}\Xi'_{t-1}] + E[\mathrm{Z}_t\Omega_t\Xi'_t]$$
$$+E[\mathrm{Z}_t\Gamma_{1,t+1}\Xi'_{t+1}] + \cdots + E[\mathrm{Z}_t\Gamma_{p,t+p}\Xi'_{t+p}] \qquad \forall\ \mathrm{Z}_t \in \mathcal{Z}_t.$$

Since this equality must hold for all $\mathrm{Z}_t \in \mathcal{Z}_t$, it is equivalent to

$$\mathrm{D}_t = E_t\left[\left(\Xi_{t+p}\ \cdots\ \Xi_{t+1}\ \Xi_t\ \Xi_{t-1}\ \cdots\ \Xi_{t-p}\right)\left(\Gamma_{p,t+p}\ \cdots\ \Gamma_{1,t+1}\ \Omega_t\ \Gamma'_{1,t}\ \cdots\Gamma'_{p,t}\right)'\right]. \quad (6)$$

We now show that (6) is satisfied for the constructed Ξ_t, $\Phi_{1,t}, \cdots, \Phi_{p,t}$, P_t and Δ_t. Note that (2) implies $\left(\Xi_{t+p}\ \cdots\ \Xi_{t+1}\right) = \left(\Xi_{t+p-1}\ \cdots\ \Xi_t\right)\Phi_{t+p} + \Delta_{t+p}\mathrm{P}_{t+p}\mathrm{J}$. Make repeated substitutions of this equation and that for Ξ_t into (6) until there are no Ξ's dated later than $t-1$ on the right-hand side. It will turn out that the coefficient near $\left(\Xi_{t-1}\ \cdots\ \Xi_{t-p}\right)$ is zero matrix due to (3), and everything that is left is exactly (5). ■