Econometric Theory, **34**, 2018, 1065–1100.
doi:10.1017/S0266466617000330

IV AND GMM INFERENCE IN ENDOGENOUS STOCHASTIC UNIT ROOT MODELS

OFFER LIEBERMAN
Bar-Ilan University

PETER C.B. PHILLIPS
Yale University, University of Auckland, Southampton University, and Singapore Management University

Lieberman and Phillips (2017; LP) introduced a multivariate stochastic unit root (STUR) model, which allows for random, time varying local departures from a unit root (UR) model, where nonlinear least squares (NLLS) may be used for estimation and inference on the STUR coefficient. In a structural version of this model where the driver variables of the STUR coefficient are endogenous, the NLLS estimate of the STUR parameter is inconsistent, as are the corresponding estimates of the associated covariance parameters. This paper develops a nonlinear instrumental variable (NLIV) as well as GMM estimators of the STUR parameter which conveniently addresses endogeneity. We derive the asymptotic distributions of the NLIV and GMM estimators and establish consistency under similar orthogonality and relevance conditions to those used in the linear model. An overidentification test and its asymptotic distribution are also developed. The results enable inference about structural STUR models and a mechanism for testing the local STUR model against a simple UR null, which complements usual UR tests. Simulations reveal that the asymptotic distributions of the NLIV and GMM estimators of the STUR parameter as well as the test for overidentifying restrictions perform well in small samples and that the distribution of the NLIV estimator is heavily leptokurtic with a limit theory which has Cauchy-like tails. Comparisons of STUR coefficient and standard UR coefficient tests show that the one-sided UR test performs poorly against the one-sided STUR coefficient test both as the sample size and departures from the null rise. The results are applied to study the relationships between stock returns and bond spread changes.

This paper is a revised version of an earlier paper entitled "IV and GMM Estimation and Testing of Multivariate Stochastic Unit Root Models." We thank the CoEditor, Pentti Saikkonen and two referees for helpful comments and Tim Ginker for research assistance. Support from the NSF under Grant No. SES 1258258 is acknowledged by Peter C.B. Phillips. Support from the Israel Science Foundation grant No. 1082-14 and from The Pinhas Sapir Center for Development at Tel Aviv University is gratefully acknowledged by Offer Lieberman. Address correspondence to Offer Lieberman, Department of Economics and Research Institute for Econometrics (RIE), Bar-Ilan University, Ramat Gan 52900, Israel; e-mail: offer.lieberman@gmail.com.

© Cambridge University Press 2017

https://doi.org/10.1017/S0266466617000330 Published online by Cambridge University Press

1. INTRODUCTION

Persistence is widely acknowledged to be one of the primary characteristics in economic and financial time series. This feature is often well captured empirically by using models with unit autoregressive roots or roots in the general vicinity of unity. The use of a local unit root (LUR) as a modeling and inferential tool has grown significantly since the early developmental research on these models (Chan and Wei, 1987; Phillips, 1987a) and the methodology now provides useful mechanisms for uniform autoregressive inference in both time series (Giraitis and Phillips, 2006; Mikusheva, 2007, 2012) and panels (Chao and Phillips, 2017). The LUR mechanism for characterizing local departures from unity uses a standard Pitman drift specification. This approach is convenient mathematically for studying local power properties. But it does not accommodate random departures from unity, formulations that allow for time variation in the coefficients, or structural dependence in the system between the coefficients and the equation errors, all of which provide a richer environment for practical empirical work.

An alternative approach uses stochastic departures from unity, allows for endogeneity, and leads to what we call an endogenous stochastic unit root (STUR) model. The present paper is concerned with such models and consistent methods of estimation of the localizing coefficients. The resulting asymptotic theory, as will be shown, offers a methodology for testing unit root persistence directly against random departures from unity even in the presence of endogeneity. The model under consideration is the structural stochastic unit root (STUR) system given by

$$\begin{aligned} &Y_1 = \varepsilon_1, \\ &Y_t = \beta_t(a;n)\, Y_{t-1} + \varepsilon_t, t = 2, \ldots, n, \end{aligned} \tag{1}$$

where the STUR coefficient

$$\beta_t(a;n) = \exp\left(\frac{a' u_t}{\sqrt{n}}\right), \tag{2}$$

depends nonlinearly on a $K \times 1$ vector of observed stationary variables u_t that are assumed to drive the localizing coefficient $\beta_t(a;n)$. In the important case where the vector $a = 0$, the model reduces to a simple unit root (UR) time series model. When the vector $a \neq 0$ but has certain components that are zero then a certain subvector of u_t comprises the driver variables of $\beta_t(a;n)$. These submodels are of considerable interest in cases where the UR model itself appears too restrictive and localized departures from unity are considered more relevant, especially when there are potential endogenous driver variables that are thought to influence the degree of persistence.

Examples of empirical models with roots in the vicinity of unity abound in the literature and this phenomenon has motivated the use of alternative models such as the local UR (LUR) model, where the coefficient $\beta_t(a;n) = e^{\frac{c}{n}}$ is fixed for given n and some unknown scalar $a = c$ (Chan and Wei, 1987; Phillips, 1987a).

https://doi.org/10.1017/S0266466617000330 Published online by Cambridge University Press

In the STUR model (1)–(2), the coefficient $\beta_t(a;n)$ is similarly localized in an array format but is dependent on a group of stationary covariates u_t with a localizing decay rate of $n^{-1/2}$ that is compatible with the (assumed) stationarity of u_t and enables an asymptotic development. Some examples of empirical applications of STUR models in finance include dual stocks pricing (Lieberman, 2012), Exchange Traded Funds pricing (Lieberman and Phillips, 2014) and call option pricing (Lieberman and Phillips, 2017, hereafter, LP). STUR models have the advantage that, under certain conditions, the coefficients may be identified and consistently estimated, thereby enabling investigators to test for the presence of relevant driver variables that influence departures of the coefficient $\beta_t(a;n)$ from unity. As will be shown in the present paper, we may also allow for structural model formulations in which the driver variables u_t that appear in (2) are endogeneous.

Under the assumption that (u_t, ε_t) is a martingale difference sequence (mds), LP (2017) showed that in the limit as the sample size $n \to \infty$, the standardized output $n^{-1/2}Y_t$ of (1)–(2) takes the form of a nonlinear diffusion, extending the well-known linear diffusion result for the LUR model. The asymptotic distribution of the nonlinear least squares (NLLS) estimator $\hat{a}_n$ of the localizing coefficient a in (2) then depends on this nonlinear diffusion. The LP results show that $\hat{a}_n$ is inconsistent in the structural model case where u_t and ε_t are correlated. Thus, in a structural version of (1)–(2), endogeneity bias is present in NLLS estimation in the limit, just as in linear models. However, when the right hand side of (1) contains a drift, LP (2017) showed that $\hat{a}_n$ is $\sqrt{n}$-consistent whether or not u_t and ε_t are correlated, a result due to the stronger regression signal that is present in the lagged variable regressor in (1) in this case.

The main goals of the present paper are as follows. First, we extend the central result of LP (2017) and derive the limit process of the standardized output $n^{-1/2}Y_t$ when u_t and ε_t are general linear processes. As expected, this extension induces additional terms in the limit which do not appear in the mds case. Second, we derive the asymptotic distribution of $\hat{a}_n$ in the model (1)–(2) for general weakly dependent u_t and ε_t. The asymptotic theory provides extensive implementation of recent weak convergence results of sample covariances between nonlinear functionals of integrated processes and short memory time series (Ibragimov and Phillips, 2008; Liang, Phillips, Wang, and Wang, 2016). We are particularly interested in the structural model case where u_t and ε_t are correlated because it seems important to allow for such correlation in practical work. Since the NLLS estimator $\hat{a}_n$ is inconsistent when u_t and ε_t are correlated, it is important to develop an alternative procedure that enables identification and consistent estimation.

As in the case of linear structural models, the primary alternative procedure involves instrumental variables. The present paper develops a nonlinear instrumental variable (NLIV) as well as (the more general) GMM estimators of a, derives their asymptotic distribution, and shows consistency under similar conditions to those used in the linear model. Furthermore, we derive the asymptotic

https://doi.org/10.1017/S0266466617000330 Published online by Cambridge University Press

distribution of the Sargan-Hansen test for overidentifying restrictions in this model. The limit theory facilitates statistical testing of the STUR model (1)–(2) against the simple unit root model where $a = 0$. Such tests are valuable in empirical applications where the relevance of potential driver variables warrants investigation.

The plan for the remainder of the paper is as follows. In Section 2 we set up model assumptions, characterize the asymptotic limit process form of $n^{-1/2}Y_t$, and derive the limit distribution of $\hat{a}_n$. The theory for the NLIV estimator is presented in Section 3. Asymptotic theory for estimation of the covariance parameters follows in Section 4 and a test statistic which accounts for the estimation of nuisance parameters is suggested in Section 5. Limit theory for GMM estimation and a test for overidentifying restrictions are developed in Section 6. Simulation experiments evaluating the adequacy of the limit theory are reported in Section 7 and a real data application is given in Section 8. Section 9 concludes and proofs are given in the Appendix.

2. PRELIMINARIES AND RESULTS ON THE NLLS

For the generating mechanism of the process $w_t = \left(u_t', \varepsilon_t\right)'$ we adopt a linear process framework similar to Ibragimov and Phillips (IP, 2008), making the following assumption.

Assumption 1. The vector w_t is a linear process satisfying

$$w_t = G(L)\,\eta_t = \sum_{j=0}^{\infty} G_j \eta_{t-j},\ \sum_{j=1}^{\infty} j \left\| G_j \right\| < \infty,\ G(1) \text{ has full rank } K+1, \tag{3}$$

η_t is iid, zero mean with $\mathbb{E}\left(\eta_t \eta_t'\right) = \Sigma_\eta > 0$ and $\max \mathbb{E}\left|\eta_{i0}\right|^p < \infty$, for some $p > 4$.

Under Assumption 1, w_t is zero mean, strictly stationary and ergodic, with partial sums satisfying the invariance principle

$$n^{-1/2} \sum_{t=1}^{\lfloor n\cdot \rfloor} w_t \Rightarrow B(\cdot) \equiv \mathrm{BM}\left(\Sigma^{\ell r}\right),\ \Sigma^{\ell r} = \begin{pmatrix} \Sigma_u^{\ell r} & \Sigma_{u\varepsilon}^{\ell r} \\ \Sigma_{u\varepsilon}^{\ell r\prime} & \left(\sigma_\varepsilon^{\ell r}\right)^2 \end{pmatrix}, \tag{4}$$

where $\lfloor \cdot \rfloor$ is the floor function and $B = (B_u, B_\varepsilon)'$ is a vector Brownian motion.[1] The matrix $\Sigma^{\ell r} = G(1)\,\Sigma_\eta G(1)' > 0$ is the long run covariance matrix of w_t, with $K \times K$ submatrix $\Sigma_u^{\ell r} > 0$, scalar $\left(\sigma_\varepsilon^{\ell r}\right)^2 > 0$ and $K \times 1$ vector $\Sigma_{u\varepsilon}^{\ell r}$. In component form, we write (3) as

$$\begin{aligned} w_t &= \begin{pmatrix} u_t \\ \varepsilon_t \end{pmatrix} = \begin{pmatrix} G_{11}(L)\ G_{12}(L) \\ G_{21}(L)\ G_{22}(L) \end{pmatrix} \begin{pmatrix} \eta_{1t} \\ \eta_{2t} \end{pmatrix} = \begin{pmatrix} G_1(L) \\ G_2(L) \end{pmatrix} \begin{pmatrix} \eta_{1t} \\ \eta_{2t} \end{pmatrix} \\ &= \begin{pmatrix} \sum_{j=0}^{\infty} G_{1,j} \eta_{t-j} \\ \sum_{j=0}^{\infty} G_{2,j} \eta_{t-j} \end{pmatrix}, \end{aligned} \tag{5}$$

where η_{1t} is $K \times 1$, η_{2t} is scalar, $G_{1,j}$ is $K \times (K+1)$ and $G_{2,j}$ is $1 \times (K+1)$.

https://doi.org/10.1017/S0266466617000330 Published online by Cambridge University Press

Denote the contemporaneous covariance matrix of w_t by $\Sigma > 0$, with corresponding components $\Sigma_u = \mathbb{E}\left(u_t u_t'\right) > 0$, $\Sigma_{u\varepsilon} = \mathbb{E}\left(u_t \varepsilon_t\right)$ and $\sigma_\varepsilon^2 = \mathbb{E}\left(\varepsilon_t^2\right) > 0$. The one-sided long run covariance matrices are similarly denoted by $\Lambda = \sum_{h=1}^{\infty} \mathbb{E}\left(w_0 w_h'\right)$ and $\Delta = \sum_{h=0}^{\infty} \mathbb{E}\left(w_0 w_h'\right) = \Lambda + \Sigma$, with corresponding component submatrices $\Lambda_{u\varepsilon} = \sum_{h=1}^{\infty} \mathbb{E}\left(u_0 \varepsilon_h\right)$, $\Lambda_{\varepsilon\varepsilon} = \sum_{h=1}^{\infty} \mathbb{E}\left(\varepsilon_0 \varepsilon_h\right)$ and $\Delta_{u\varepsilon} = \sum_{h=0}^{\infty} \mathbb{E}\left(u_0 \varepsilon_h\right)$, $\Delta_{\varepsilon\varepsilon} = \sum_{h=0}^{\infty} \mathbb{E}\left(\varepsilon_0 \varepsilon_h\right)$.

In the special case where w_t is an mds, $\Sigma = \Sigma^{\ell r}$. For that case, Lieberman and Phillips (2014, 2017) showed that the standardized output process $n^{-1/2} Y_{t=\lfloor n\cdot \rfloor}$ of (1) converges weakly to a nonlinear diffusion process. The following Lemma extends the result of LP (2017) to the present case of stationary driver variables and equation errors satisfying Assumption 1.

LEMMA 1. *For the model (1)–(2), under Assumption 1,*

$$n^{-1/2} Y_{\lfloor nr \rfloor} \Rightarrow e^{a' B_u(r)} \left(\int_0^r e^{-a' B_u(p)} dB_\varepsilon(p) - a' \Delta_{u\varepsilon} \int_0^r e^{-a' B_u(p)} dp \right) := G_a(r). \quad \textbf{(6)}$$

Importantly, the quantity $a' \Delta_{u\varepsilon}$ in (6) involves the one-sided long run covariance matrix $\Delta_{u\varepsilon}$ between u and ε. This quantity measures the importance of the random drift effect, $\int_0^r e^{a'(B_u(r) - B_u(p))} dp$, that is induced in the limit process G_a whenever $a \neq 0$ and $\Delta_{u\varepsilon} \neq 0$. If $a = 0$, the limit process is the Brownian motion B_ε, as expected.

Denote by $\hat{a}_n$ the NLLS of a which minimizes the criterion $Q_n(a) = \sum_{t=2}^{n} \{Y_t - \beta_t(a; n) Y_{t-1}\}^2$. By H and L we denote the

$$(K+1)^2 \times (K+1)(K+2)/2 \text{ and } ((K+1)(K+2)/2) \times (K+1)^2,$$

duplication and elimination matrices, respectively, of zeros and ones, which for a $(K+1) \times (K+1)$ matrix A, satisfy

$$vec(A) = Hvech(A) \text{ and } vech(A) = Lvec(A). \quad \textbf{(7)}$$

If η_t has finite fourth moments, centred partial sums of $\eta_t \eta_t'$ satisfy the invariance principle

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor nr \rfloor} vech\left(\eta_t \eta_t' - \Sigma_\eta\right) \Rightarrow \upsilon(r), \quad \textbf{(8)}$$

where $\upsilon(r)$ is vector Brownian motion with covariance matrix

$$\begin{aligned} \Sigma_{\eta\otimes\eta} &= E\left(vech\left(\eta_t \eta_t' - \Sigma_\eta\right) \left(vech\left(\eta_t \eta_t' - \Sigma_\eta\right)\right)' \right) \\ &= E\left(L\left((\eta_t \otimes \eta_t) - E(\eta_t \otimes \eta_t)\right) \left(\left(\eta_t' \otimes \eta_t'\right) - E\left(\eta_t' \otimes \eta_t'\right)\right) L' \right). \end{aligned}$$

Furthermore, for any $l \neq 0$, by $\zeta(r)$ we denote the vector Brownian motion with covariance matrix

$$E\left(vec\left(\eta_t \eta_{t-l}'\right) \left(vec\left(\eta_t \eta_{t-l}'\right)\right)'\right) = \Sigma_\eta \otimes \Sigma_\eta.$$

https://doi.org/10.1017/S0266466617000330 Published online by Cambridge University Press

Finally, we denote the matrix of third order moments of η by

$$M_3 = E\left(\left(\eta_t \otimes \eta_t\right)\eta_t'\right). \tag{9}$$

The following theorem was established in LP (2017) for the case where η_t is a strictly stationary mds and is extended below to the case where η_t is a zero mean, strictly stationary and ergodic process, satisfying Assumption 1.

THEOREM 2. *For the model (1)–(2), under Assumption 1, the asymptotic behavior of $\hat{a}_n$ is given by*:

(1) $\left(\hat{a}_n - a\right) \Rightarrow \frac{\int_0^1 G_a(r)dr}{\int_0^1 G_a^2(r)dr}\Sigma_u^{-1}\Sigma_{u\varepsilon}$, *if* $\Sigma_{u\varepsilon} \neq 0$.

(2) $\hat{a}_n \Rightarrow \frac{\int_0^1 B_\varepsilon(r)dr}{\int_0^1 B_\varepsilon^2(r)dr}\Sigma_u^{-1}\Sigma_{u\varepsilon}$, *if* $\Sigma_{u\varepsilon} \neq 0$ *and* $a = 0$.

(3)

$$\begin{aligned}
&\sqrt{n}\left(\hat{a}_n - a\right)\\
&\quad\Rightarrow \frac{1}{\int_0^1 G_a^2(r)\,dr}\Sigma_u^{-1}\left(\sum_{j=0}^{\infty}\left(G_{2,j}\otimes G_{1,j}\right)H\int_0^1 G_a(r)\,d\upsilon(r)\right.\\
&\qquad+\sum_{j=1}^{\infty}\left(G_{2,j}\otimes G_{1,j}\right)M_3\left(\left(\sum_{i=0}^{j-1}G_{1,i}\right)'a\int_0^1 G_a(r)\,dr+\left(\sum_{i=0}^{j-1}G_{2,i}\right)'\right)\\
&\qquad+\sum_{j\neq k}\left(G_{2,k}\otimes G_{1,j}\right)\int_0^1 G_a(r)\,d\zeta(r)\\
&\qquad\left.+\left\{\mathbb{E}\left(\varepsilon_t u_t u_t'\right)\right\}a\int_0^1 G_a(r)\,dr\right),\quad \text{if } \Sigma_{u\varepsilon}=0.
\end{aligned} \tag{10}$$

(4)

$$\begin{aligned}
\sqrt{n}\hat{a}_n \Rightarrow{}& \frac{1}{\int_0^1 B_\varepsilon^2(r)\,dr}\Sigma_u^{-1}\left(\sum_{j=0}^{\infty}\left(G_{2,j}\otimes G_{1,j}\right)H\int_0^1 B_\varepsilon(r)\,d\upsilon(r)\right.\\
&+\sum_{j=1}^{\infty}\left(G_{2,j}\otimes G_{1,j}\right)M_3\left(\sum_{i=0}^{j-1}G_{2,i}\right)'\\
&\left.+\sum_{j\neq k}\left(G_{2,k}\otimes G_{1,j}\right)\int_0^1 B_\varepsilon(r)\,d\zeta(r)\right),
\end{aligned}$$

if $\Sigma_{u\varepsilon} = 0$ *and* $a = 0$.

Remark 1. The results in Theorem 2 depend directly on the contemporaneous covariance matrices $\Sigma_u = \sum_{j=0}^{\infty} G_{1j}\Sigma_\eta G_{1j}'$ and $\Sigma_{u\varepsilon} = \sum_{j=0}^{\infty} G_{1j}\Sigma_\eta G_{2j}'$, where $\Sigma_\eta = \mathbb{E}\left(\eta_t\eta_t'\right)$, and in view of (6), on the long run covariances indirectly, through $G_a(r)$.

https://doi.org/10.1017/S0266466617000330 Published online by Cambridge University Press

Remark 2. If η_t has a symmetric distribution around zero and $K = 1$, then

$$\mathbb{E}\left(\varepsilon_t u_t u_t'\right) = \mathbb{E}\left(\varepsilon_t u_t^2\right) = \sum_{j=0}^{\infty} G_{2,j} G_{1,j}^2 \mathbb{E}\left(\eta_{t-j}^3\right) = 0, \text{ and } M_3 = 0$$

because all odd moments of a symmetric distribution around zero are equal to zero. In this case, equation (10) reduces to

$$\sqrt{n}\left(\hat{a}_n - a\right) \Rightarrow \frac{1}{\int_0^1 G_a^2(r)\,dr} \Sigma_u^{-1} \left(\sum_{j=0}^{\infty} \left(G_{2,j} \otimes G_{1,j}\right) H \int_0^1 G_a(r)\,d\upsilon(r) + \sum_{j \neq k} \left(G_{2,k} \otimes G_{1,j}\right) \int_0^1 G_a(r)\,d\zeta(r) \right),$$

if $\Sigma_{u\varepsilon} = 0$.

Remark 3. If $\left(u_t', \varepsilon_t\right)'$ is a vector MDS, $G_{2,j} = 0$ and $G_{1,j} = 0\ \forall j \neq 0$ and equation (10) simplifies to

$$\sqrt{n}\left(\hat{a}_n - a\right) \Rightarrow \frac{1}{\int_0^1 G_a^2(r)\,dr} \Sigma_u^{-1} \left(\int_0^1 G_a(r)\,dB_{u\varepsilon}(r) + \left\{\mathbb{E}\left(\varepsilon_t u_t u_t'\right)\right\} a \int_0^1 G_a(r)\,dr \right),$$

if $\Sigma_{u\varepsilon} = 0$.

It follows from Theorem 2 that $\hat{a}_n$ is inconsistent when $\Sigma_{u\varepsilon} \neq 0$. It is emphasized that this result pertains to model (1)–(4) in which it is assumed that the drift parameter equals zero. When the drift parameter is non-zero, LP (2017) showed that the least squares estimator of a is consistent even when $\Sigma_{u\varepsilon} \neq 0$.

The present paper develops consistent IV and GMM estimators for a and derives their limit distributions. The conditions imposed on the instruments are similar to those that are used in linear model IV. The results are used to test the null hypothesis of a unit root against the STUR alternative which is given by equations (1)–(2).

3. IV ESTIMATION OF THE STUR MODEL

Let Z_t be a $q \times 1$ vector of instruments for u_t, $q \geq K$, and η_t^* be a $(K + q + 1) \times 1$- random vector. We extend the setup of (5) by letting

$$w_t^* = \begin{pmatrix} u_t \\ \varepsilon_t \\ Z_t \end{pmatrix} = G^*(L)\,\eta_t^* = \sum_{j=0}^{\infty} G_j^* \eta_{t-j}^* = \begin{pmatrix} G_1(L) \\ G_2(L) \\ G_3(L) \end{pmatrix} \begin{pmatrix} \eta_{1t} \\ \eta_{2t} \\ \eta_{3t} \end{pmatrix}$$
$$= \begin{pmatrix} G_{11}(L)\ G_{12}(L)\ G_{13}(L) \\ G_{21}(L)\ G_{22}(L)\ G_{23}(L) \\ G_{31}(L)\ G_{32}(L)\ G_{33}(L) \end{pmatrix} \begin{pmatrix} \eta_{1t} \\ \eta_{2t} \\ \eta_{3t} \end{pmatrix} = \begin{pmatrix} \sum_{j=0}^{\infty} G_{1,j} \eta_{t-j} \\ \sum_{j=0}^{\infty} G_{2,j} \eta_{t-j} \\ \sum_{j=0}^{\infty} G_{3,j} \eta_{t-j} \end{pmatrix}.$$

https://doi.org/10.1017/S0266466617000330 Published online by Cambridge University Press

Assumption 2. The vector w_t^* satisfies

$$\sum_{j=1}^{\infty} j\left\|G_j^*\right\| < \infty,\ G^*(1) \text{ is full rank,}$$

η_t^* is iid, zero mean with $\mathbb{E}\left(\eta_t^*\eta_t^{*\prime}\right) = \Sigma_{\eta^*} > 0$ and $\max \mathbb{E}\left|\eta_{i0}^*\right|^p < \infty$, for some $p > 4$.

This framework is sufficiently rich to include many known models, including the stationary and ergodic ARMA model.

Assumption 3. For all t,

$$\mathbb{E}\left(Z_t\varepsilon_t\right) = \sum_{j=0}^{\infty} G_{3,j}\,\Sigma_{\eta^*}\,G_{2,j}' = 0, \tag{11}$$

$$\mathbb{E}\left(Z_t u_t'\right) = \Sigma_{Zu} = \sum_{j=0}^{\infty} G_{3,j}\,\Sigma_{\eta^*}\,G_{1,j}' \text{ has full rank } K. \tag{12}$$

In the remainder of this section we shall consider the $q = K$ case, in which the IV estimator, $\hat{a}_n^{IV}$, solves the K-moment conditions:

$$\sum_{t=2}^{n}\left(Y_t - \beta_t\left(\hat{a}_n^{IV}; n\right)Y_{t-1}\right)Z_t = 0. \tag{13}$$

The more general $q \geq K$ case will be discussed in Section 6. Under Assumption 2,

$$n^{-1/2}\sum_{t=1}^{\lfloor nr\rfloor} Z_t\varepsilon_t \Rightarrow B_{Z\varepsilon}(r), \tag{14}$$

where $B_{Z\varepsilon}(r)$ is a Brownian motion with covariance matrix

$$\gamma_{Z\varepsilon}^{\ell r} = \sum_{h=-\infty}^{\infty}\mathbb{E}\left(Z_t Z_{t+h}'\varepsilon_t\varepsilon_{t+h}\right) = \sum_{h=-\infty}^{\infty}\mathbb{E}\left(Z_t Z_{t+h}'\right)\mathbb{E}\left(\varepsilon_t\varepsilon_{t+h}\right), \tag{15}$$

if Z_t is independent of ε_s for all s, t.

The asymptotic distribution of the IV estimator is as follows.

THEOREM 3. *For the model (1)–(4), under Assumptions 2 and 3, for $q = K$,*

$$\sqrt{n}\left(\hat{a}_n^{IV} - a\right) \Rightarrow \frac{\Sigma_{Zu}^{-1}B_{Z\varepsilon}(1)}{\int_0^1 G_a(r)\,dr}. \tag{16}$$

It is emphasized that the matrix Σ_{Zu} appearing on the right side of (16) is the contemporaneous covariance between Z and u. Several remarks are in place.

https://doi.org/10.1017/S0266466617000330 Published online by Cambridge University Press

Remark 4. Unlike the least squares estimator, Theorem 3 implies that $\hat{a}_n^{IV}$ is consistent for a, whether or not $\Sigma_{u\varepsilon} = 0$.

Remark 5. The role of the usual IV orthogonality condition (11) in Assumption 3 is evident in equation (A.22) of the Appendix, where it is clear that if $\mathbb{E}(Z_t\varepsilon_t) \neq 0$, then

$$\Sigma_{Zu}\left(\int_0^1 G_a(r)\,dr\right)\left(\hat{a}_n^{IV} - a\right) \Rightarrow \mathbb{E}(Z_t\varepsilon_t).$$

Hence, a violation of condition (11) renders $\hat{a}_n^{IV}$ inconsistent as expected.

Remark 6. The limit distribution in (16) is not defined if the relevance condition (12) of Assumption 3, that Σ_{Zu} has full rank, is violated. In particular, if some instruments are irrelevant and Σ_{Zu} has deficient rank, then the IV estimator will be inconsistent, although in such cases some contrasts (linear combinations) of a may be consistently estimable.

Remark 7. In the $K = 1$ case, under the null hypothesis $H_0 : a = 0$,

$$\int_0^1 G_a(r)\,dr = \int_0^1 B_\varepsilon(r)\,dr =_d N\left(0, \left(\sigma_\varepsilon^{lr}\right)^2/3\right),$$

so that,

$$\sqrt{n}\hat{a}_n^{IV} \Rightarrow \frac{N\left(0, \gamma_{Z\varepsilon}^{lr}\right)}{\sigma_{Zu}N\left(0, \left(\sigma_\varepsilon^{lr}\right)^2/3\right)}, \tag{17}$$

where $\sigma_{Zu} = Cov(Z, u)$. If, in addition, Z_t and ε_t are independent mds processes, then

$$\sqrt{n}\hat{a}_n^{IV} \Rightarrow \frac{N\left(0, \sigma_\varepsilon^2\sigma_Z^2\right)}{\sigma_{Zu}N\left(0, \sigma_\varepsilon^2/3\right)}, \tag{18}$$

where $\sigma_Z^2 = Var(Z_t)$, $\forall t$.

The limit distributions (17) and (18) are scaled ratios of normal variates which have heavy Cauchy tails because the denominator has positive density at the origin and is not perfectly correlated with the numerator.[2] This feature of the limit distribution is manifested in finite samples and affects the simulation results of Section 7, where large outliers occurred in the computation of simulated means and variances.

To explore this issue further, we rewrite the result (17) as

$$\sqrt{n}\hat{a}_n^{IV} \Rightarrow \frac{\xi_1}{\sigma_{Zu}\xi_2},$$

https://doi.org/10.1017/S0266466617000330 Published online by Cambridge University Press

say. The vector $(\xi_1, \xi_2)'$ is $N\left(0, \Sigma^{\xi}\right)$, Σ^{ξ} is positive definite and with components $\left\{\sigma_{11}^{\xi}, \sigma_{12}^{\xi}, \sigma_{22}^{\xi}\right\}$. Then

$$\sqrt{n}\hat{a}_n^{IV} \Rightarrow \frac{\xi_{1.2}}{\sigma_{Zu}\xi_2} + \frac{\sigma_{12}^{\xi}}{\sigma_{Zu}\sigma_{22}^{\xi}} =: \frac{\xi_{1.2}}{\sigma_{Zu}\xi_2} + B,$$

say, giving the bias $B = \sigma_{12}^{\xi}/\sigma_{Zu}\sigma_{22}^{\xi}$ in the limit distribution. When the covariance parameters in B are estimated, which will be at a $\sqrt{n}$ rate if the variables form an mds or, more typically, at a lower than $\sqrt{n}$ rate if they are weakly dependent and long run variances/covariances need to be estimated, we will effectively end up with centred asymptotics of the following form. Thus, if the rate of convergence is $\sqrt{k_n}$ for $\frac{k_n}{n} \to 0$ and we have $\sqrt{k_n}\left(\hat{B} - B\right) \Rightarrow N(0, V_B)$, then we will have

$$\sqrt{n}\hat{a}_n^{IV} - \hat{B} = \sqrt{n}\hat{a}_n^{IV} - B + \left(\hat{B} - B\right) = \sqrt{n}\hat{a}_n^{IV} - B + O_p\left(\frac{1}{\sqrt{k_n}}\right)$$

$$\Rightarrow \frac{\xi_{1.2}}{\sigma_{Zu}\xi_2} = \frac{1}{\sigma_{Zu}}\left(\frac{\sigma_{11.2}^{\xi}}{\sigma_{22}^{\xi}}\right)^{1/2}\mathbb{C}.$$

Now, under Assumption 2,

$$\sigma_{12}^{\xi} = Cov(\xi_1, \xi_2) = E\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n} Z_t\varepsilon_t\right)\left(\frac{1}{n^{3/2}}\sum_{s=1}^{n}\sum_{j=1}^{s}\varepsilon_j\right).$$

If Z_t is independent of ε_t, as assumed in the second part of Remark 7, then $\sigma_{12}^{\xi} = 0$ and so, $B = 0$. Otherwise,

$$\begin{aligned}
\sigma_{12}^{\xi} &= \frac{1}{n^2}\sum_{t=1}^{n}\sum_{s=1}^{n} E\left(Z_t\varepsilon_t\sum_{j=1}^{s}\varepsilon_j\right) \\
&= \frac{1}{n^2}\sum_{t=1}^{n}\sum_{s=1}^{n} E\left(\sum_{k=0}^{\infty} G_{3,k}\eta_{t-k}\sum_{l=0}^{\infty} G_{2,l}\eta_{t-l}\sum_{j=1}^{s}\sum_{m=0}^{\infty} G_{2,m}\eta_{j-m}\right) \\
&= \frac{1}{n^2}\sum_{t=1}^{n}\sum_{s=1}^{n} E\left(\sum_{k=0}^{\infty}\sum_{l=0}^{\infty}\sum_{j=1}^{s}\sum_{m=0}^{\infty} G_{3,k}G_{2,l}G_{2,m}\eta_{t-k}\eta_{t-l}\eta_{j-m}\right) \\
&= \frac{1}{n^2}\sum_{t=1}^{n}\sum_{s=1}^{n}\sum_{k=0}^{\infty}\sum_{j=1}^{s}\sum_{m=0}^{\infty} G_{3,k}G_{2,k}G_{2,m}E\left(\eta_{t-k}^3\right)1\{t-k = j-m\}.
\end{aligned}$$

This term will be zero if η_t has a symmetric distribution. In these cases, (17) simplifies to

$$\sqrt{n}\hat{a}_n^{IV} \Rightarrow \frac{1}{\sigma_{Zu}}\left(\frac{\sigma_{11.2}^{\xi}}{\sigma_{22}^{\xi}}\right)^{1/2}\mathbb{C} = \frac{1}{\sigma_{Zu}}\left(\frac{\sigma_{11}^{\xi}}{\sigma_{22}^{\xi}}\right)^{1/2}\mathbb{C} = \frac{1}{\sigma_{Zu}}\left(\frac{\gamma_{Z\varepsilon}^{lr}}{\left(\sigma_{\varepsilon}^{lr}\right)^2/3}\right)^{1/2}\mathbb{C},$$

https://doi.org/10.1017/S0266466617000330 Published online by Cambridge University Press

and (18) reduces to

$$\sqrt{n}\hat{a}_n^{IV} \Rightarrow \frac{\sqrt{3}\sigma_Z}{\sigma_{Zu}}\mathbb{C}, \tag{19}$$

respectively. For inference, this scaled Cauchy distribution can be used.

We note that, unlike the ADF t-test in the linear case, the estimated standard deviation of $\hat{a}_n^{IV}$ does not have a closed form. In principle, t-ratio and Wald tests might be constructed by simulating the standard deviation of the right side of (16) and extracting the corresponding limit theory of the ratio. Such a construction substantially complicates implementation relative to the coefficient test and it is unclear whether this approach brings any benefit over the simpler coefficient test implied by (16).

To complete this section we compare the STUR approach to a direct DF test of the UR null implied by $H_0 : a = 0$. Simple calculations based on the earlier asymptotic theory show that the usual UR coefficient test of $\beta_t(a; n) = \beta = 1, \forall t$, has the following limit theory

$$n\left(\hat{\beta} - 1\right) \Rightarrow \frac{\int_0^1 B_\varepsilon dB_\varepsilon + \Lambda_{\varepsilon\varepsilon}}{\int_0^1 B_\varepsilon^2}, \text{ under } H_0 : a = 0. \tag{20}$$

Under the alternative we have

$$\Delta Y_t = \left(e^{a'u_t/\sqrt{n}} - 1\right)Y_{t-1} + \varepsilon_t = \frac{a'u_t}{\sqrt{n}}Y_{t-1} + \varepsilon_t + o_p(1),$$

and using the results[3]

$$\frac{1}{n}\sum_{t=2}^{n}\varepsilon_t Y_{t-1} \Rightarrow \int_0^1 G_a(r)\,dB_\varepsilon(r) + \Lambda'_{u\varepsilon}a\int_0^1 G_a(r)\,dr + \Lambda_{\varepsilon\varepsilon}, \tag{21}$$

$$\frac{1}{n^{3/2}}\sum_{t=2}^{\infty}u_t Y_{t-1}^2 \Rightarrow \int_0^1 G_a^2(r)\,dB_u(r) + 2\left(\Lambda'_{uu}a\int_0^1 G_a^2(r)\,dr + \Lambda_{u\varepsilon}\int_0^1 G_a(r)\,dr\right), \tag{22}$$

and

$$n^{-2}\sum_{t=1}^{n}Y_{t-1}^2 \Rightarrow \int_0^1 G_a^2(r)\,dr,$$

we obtain the following limit behavior

$$\begin{aligned} n\left(\hat{\beta} - 1\right) &= \frac{n^{-1}\sum_{t=1}^{n}Y_{t-1}\Delta Y_t}{n^{-2}\sum_{t=1}^{n}Y_{t-1}^2} \\ &= \frac{n^{-1}\left(\sum_{t=1}^{n}Y_{t-1}\varepsilon_t\right) + n^{-3/2}a'\sum_{t=1}^{n}u_t Y_{t-1}^2 + o_p(1)}{n^{-2}\sum_{t=1}^{n}Y_{t-1}^2} \end{aligned}$$

https://doi.org/10.1017/S0266466617000330 Published online by Cambridge University Press

$$\Rightarrow \left(\int_0^1 G_a^2(r)\,dr\right)^{-1}\left(\int_0^1 G_a(r)\,dB_\varepsilon(r) + \Lambda'_{u\varepsilon}a\int_0^1 G_a(r)\,dr + \Lambda_{\varepsilon\varepsilon} + a'\left(\int_0^1 G_a^2(r)\,dB_u(r) + 2\left(\Lambda'_{uu}a\int_0^1 G_a^2(r)\,dr + \Lambda_{u\varepsilon}\int_0^1 G_a(r)\,dr\right)\right)\right). \quad (23)$$

The limit (23) shows that standard UR tests based on the estimate $\hat{\beta}$ have local power which depends on the magnitude of a and the process $G_a(r)$. Finite sample performance is investigated numerically in Section 7.

4. ESTIMATION OF THE COVARIANCE PARAMETERS

The limit theory of the least squares-based estimators of σ_ε^2, Σ_u and $\Sigma_{u\varepsilon}$, $\hat{\sigma}_{\varepsilon,n}^2$, $\hat{\Sigma}_{u,n}$ and $\hat{\Sigma}_{u\varepsilon,n}$, respectively, was given in Theorem 3 of LP (2017), in the case where w_t is a strictly stationary and ergodic mds process. It follows from their results that $\hat{\sigma}_{\varepsilon,n}^2$ and $\hat{\Sigma}_{u\varepsilon,n}$ are inconsistent. Let

$$e_t^{IV} = Y_t - e^{\hat{a}_n^{IV\prime}u_t/\sqrt{n}}Y_{t-1},\ t = 2,\ldots,n. \quad (24)$$

We show in the next result that for $j = 0, 1, 2, \ldots$, the IV-based estimators

$$\hat{\gamma}_{\varepsilon,n}^{IV}(j) = \frac{1}{n}\sum_{t=j+2}^{n} e_t^{IV}e_{t-j}^{IV},\ \hat{\gamma}_{u,\varepsilon,n}^{IV}(j) = \frac{1}{n}\sum_{t=j+2}^{n} u_t e_{t-j}^{IV},$$

$$\hat{\gamma}_{Z\varepsilon,n}^{IV}(j) = \frac{1}{n}\sum_{t=j+2}^{n} Z_t Z'_{t-j}e_t^{IV}e_{t-j}^{IV},$$

of $\gamma_\varepsilon(j) = Cov(\varepsilon_t, \varepsilon_{t-j})$, $\gamma_{u,\varepsilon}(j) = Cov(u_t, \varepsilon_{t-j})$ and $\gamma_{Z\varepsilon}(j) = Cov(Z_t\varepsilon_t, Z'_{t-j}\varepsilon_{t-j})$ are all consistent. In particular, $\hat{\gamma}_{\varepsilon,n}^{IV}(0) = \left(\hat{\sigma}_{\varepsilon,n}^{IV}\right)^2$ is consistent for σ_ε^2, and $\hat{\Sigma}_{Zu,n} = \frac{1}{n}\sum_{t=2}^{n} Z_t u'_t$ is consistent for Σ_{Zu} by ergodicity.

THEOREM 4. *Under Assumptions 2 and 3, for $q = K$,*

1. $\hat{\gamma}_{\varepsilon,n}^{IV}(j) - \gamma_\varepsilon(j) = O_p\left(n^{-1/2}\right)$.
2. $\hat{\gamma}_{u,\varepsilon,n}^{IV}(j) - \gamma_{u,\varepsilon}(j) = O_p\left(n^{-1/2}\right)$.
3. $\hat{\gamma}_{Z\varepsilon,n}^{IV}(j) - \gamma_{Z\varepsilon}(j) = O_p\left(n^{-1/2}\right)$.

The variance estimate $\hat{\Sigma}_{u,n}$, defined through $vech\left(\hat{\Sigma}_{u,n}\right) = \frac{1}{n}\sum_{t=1}^{n} vech\left(u_t u'_t\right)$, does not depend on a and is $\sqrt{n}$-consistent.

https://doi.org/10.1017/S0266466617000330 Published online by Cambridge University Press

5. A TEST STATISTIC WITH NUISANCE PARAMETERS ESTIMATED

The limit distribution in (16) depends on the unknown parameter Σ_{Zu}. In fact, under Assumption 1 and from equation (A.22) of the Appendix, we have

$$\sqrt{n}\left(\int_0^1 G_a(r)\,dr\right)\Sigma_{Zu}\left(\hat{a}_n^{IV}-a\right)=B_{Z\varepsilon}(1)+O_p\left(n^{-1/2}\right). \tag{25}$$

The left side of (25) is equal to

$$\begin{aligned}&\sqrt{n}\left(\int_0^1 G_a(r)\,dr\right)\hat{\Sigma}_{Zu,n}\left(\hat{a}_n^{IV}-a\right)+\sqrt{n}\left(\int_0^1 G_a(r)\,dr\right)\left(\Sigma_{Zu}-\hat{\Sigma}_{Zu,n}\right)\left(\hat{a}_n^{IV}-a\right)\\&\quad=\sqrt{n}\left(\int_0^1 G_a(r)\,dr\right)\hat{\Sigma}_{Zu,n}\left(\hat{a}_n^{IV}-a\right)+O_p\left(n^{-1/2}\right)\\&\quad=B_{Z\varepsilon}(1)+O_p\left(n^{-1/2}\right).\end{aligned}$$

Therefore, to first order we may replace the right sides of (16) and (18) by

$$\frac{\left(\hat{\Sigma}_{Zu,n}\right)^{-1}B_{Z\varepsilon}(1)}{\int_0^1 G_a(r)\,dr} \tag{26}$$

and

$$\frac{B_{Z\varepsilon}(1)}{\hat{\sigma}_{Zu,n}\int_0^1 B_\varepsilon(r)\,dr}, \tag{27}$$

respectively. The long run covariances associated with the distributions in (26) and (27) can be consistently estimated using the results of Theorem 4 and a standard tapering argument. For instance, the covariance matrix $\gamma_{Z\varepsilon}^{lr}$ of $B_{Z\varepsilon}(1)$ may be consistently estimated by a Bartlett-Newey-West HAC estimator using the autocovariance estimates $\hat{\gamma}_{Z\varepsilon,n}^{IV}(j)$.

6. GMM ESTIMATION AND A TEST FOR OVERIDENTIFYING RESTRICTIONS

The approach may be extended to allow for $q > K$ instruments in Z_t. In such cases, we may estimate a by GMM and a Sargan-Hansen-type test may be used to test for overidentifying restrictions. This section develops this analysis and provides limit theory for the GMM estimator and the overidentification test. Let

$$g_n(a)=\frac{1}{n}\sum_{t=1}^{n}\left(Y_t-\beta_t(a)\,Y_{t-1}\right)Z_t,$$

$$\hat{a}_n^G=\hat{a}_n^G\left(\hat{W}\right)=\arg\min_a J_n\left(a,\hat{W}\right), \tag{28}$$

https://doi.org/10.1017/S0266466617000330 Published online by Cambridge University Press

where

$$J_n\left(a,\hat{W}\right)=ng_n'(a)\,\hat{W}g_n(a),$$

$\hat{W}$ is a $q\times q$, symmetric positive definite matrix, possibly dependent on the sample, such that $\hat{W}\to_p W$, and W is a weighting matrix. In this case Σ_{Zu} is given by (12) but is $(q\times K)$, with the possibility that $q\geq K$. The limit theory for $\hat{a}_n^G$ is as follows.

THEOREM 5. *For the model (1)–(4), under Assumptions 2 and 3 and for* $q\geq K$,

$$\sqrt{n}\left(\hat{a}_n^G-a\right)\Rightarrow\left(\Sigma_{Zu}'W\Sigma_{Zu}\right)^{-1}\Sigma_{Zu}'W\frac{B_{Z\varepsilon}(1)}{\int_0^1 G_a(r)\,dr}. \tag{29}$$

Remark 8. If the model is just identified, the result of the Theorem collapses to (16).

Remark 9. Consider the linear model

$$Y_t=x_t'\delta+\varepsilon_t,$$

where x_t is $K\times 1$. Let Z_t be a $q\times 1$ vector of instruments for x_t, with $q\geq K$, and $\hat{W}$ is defined above. It is well known (e.g., Hayashi, 2000) that the GMM estimator $\hat{\delta}_n^G$ of δ in this model is

$$\hat{\delta}_n^G=\left(S_{ZX}'\hat{W}S_{ZX}\right)^{-1}S_{ZX}'\hat{W}s_{Xy}, \tag{30}$$

where

$$S_{ZX}=\frac{1}{n}\sum_{t=1}^n Z_tx_t' \text{ and } s_{Xy}=\frac{1}{n}\sum_{t=1}^n x_tY_t.$$

The correspondence between (29) and the usual linear formulation (30) is clear.

A Sargan-Hansen-type test for overidentifying restrictions in this context can be based on the statistic

$$J_n\left(\hat{a}_n^G\left(\left(\hat{\gamma}_{Z\varepsilon,n}^{\ell r}\right)^{-1}\right),\left(\hat{\gamma}_{Z\varepsilon,n}^{\ell r}\right)^{-1}\right)=n\hat{g}_n'\left(\hat{\gamma}_{Z\varepsilon,n}^{\ell r}\right)^{-1}\hat{g}_n,$$

where

$$\hat{g}_n=g_n\left(\hat{a}_n^G\left(\left(\hat{\gamma}_{Z\varepsilon,n}^{\ell r}\right)^{-1}\right)\right)=\frac{1}{n}\sum_{t=1}^n Z_t\left(Y_t-\beta_t\left(\hat{a}_n^G\left(\left(\hat{\gamma}_{Z\varepsilon,n}^{\ell r}\right)^{-1}\right)\right)Y_{t-1}\right),$$

and $\hat{\gamma}_{Z\varepsilon,n}^{\ell r}$ is a consistent estimator of $\gamma_{Z\varepsilon,n}^{\ell r}$, defined in (15). The limit theory for this statistic has the usual χ^2_{q-K} form, as given in (31) below.

https://doi.org/10.1017/S0266466617000330 Published online by Cambridge University Press

THEOREM 6. *For the model (1)–(4), under Assumptions 2 and 3 and for* $q \geq K$,

$$J_n\left(\hat{a}_n^G\left(\left(\hat{\gamma}_{Z\varepsilon,n}^{\ell r}\right)^{-1}\right),\left(\hat{\gamma}_{Z\varepsilon,n}^{\ell r}\right)^{-1}\right) = n\hat{g}_n'\left(\hat{\gamma}_{Z\varepsilon,n}^{\ell r}\right)^{-1}\hat{g}_n \Rightarrow \chi^2_{q-K}. \tag{31}$$

Remark 10. The χ^2_{q-K} limit distribution for the overidentifying test again corresponds to that in the linear model discussed in Remark 9.

7. SIMULATIONS

This section reports an investigation of the finite sample performance of the limit theory for the coefficient estimator $\hat{a}_n^{IV}$, the coefficient test (27), the efficient GMM estimator $\hat{a}_n^G$, and the overidentification test. We consider the following scenarios.

Case 1.

$$u_t \overset{iid}{\sim} N\left(0,\sigma_u^2\right), \sigma_u^2 = 0.1, \varepsilon_t \overset{iid}{\sim} bu_t + \eta_t, \eta_t \overset{iid}{\sim} U[-1,1], Z_t \overset{iid}{\sim} u_t - 3b\sigma_u^2\eta_t.$$

Case 2.

$$u_t \overset{iid}{\sim} N\left(0,\sigma_u^2\right), \sigma_u^2 = 0.1, \varepsilon_t \overset{iid}{\sim} bu_t + 2\eta_t, \eta_t \overset{iid}{\sim} U[-1,1], Z_t \overset{iid}{\sim} u_t - 3b\sigma_u^2\eta_t.$$

Case 3.

$$u_t \overset{iid}{\sim} N\left(0,\sigma_u^2\right), \sigma_u^2 = 0.1, \varepsilon_t \overset{iid}{\sim} N\left(0,\sigma_u^2\right),$$
$$u_t \text{ are independent of } \varepsilon_t, \eta_t \overset{iid}{\sim} U[-1,1], Z_t \overset{iid}{\sim} u_t - 3b\sigma_u^2\eta_t.$$

Case 4.

$$u_t \overset{iid}{\sim} N\left(0,\sigma_u^2\right), \sigma_u^2 = 0.1, \varepsilon_t \overset{iid}{\sim} N\left(0,\sigma_u^2\right) + 2\eta_t,$$
$$u_t \text{ are independent of } \varepsilon_t, \eta_t \overset{iid}{\sim} U[-1,1], Z_t \overset{iid}{\sim} u_t - 3b\sigma_u^2\eta_t.$$

For each case, we simulated 5,000 replications with $n = 100$, 1,000, 10,000 and $b = 0.2$. These scenarios are summarized in Table 1 below.

TABLE 1. Covariances in each case

Case	$\Sigma_{u\varepsilon}$	$\Sigma_{Z\varepsilon}$	Σ_{Zu}
1	$\neq 0$	$= 0$	$\neq 0$
2	$\neq 0$	$\neq 0$	$\neq 0$
3	$= 0$	$= 0$	$\neq 0$
4	$= 0$	$\neq 0$	$\neq 0$

https://doi.org/10.1017/S0266466617000330 Published online by Cambridge University Press

Cases 1 and 2 correspond to the situation in which $\Sigma_{u\varepsilon} \neq 0$, but Assumption 3 holds for Case 1, because

$$E(\varepsilon_t Z_t) = E\left((bu_t + \eta_t)\left(u_t - 3b\sigma_u^2 \eta_t\right)\right) = b\sigma_u^2 - 3b\sigma_u^2 Var(\eta_t) = 0,$$

whereas in Case 2, $\Sigma_{Z\varepsilon} \neq 0$. Similarly, Cases 3 and 4 correspond to the situation in which $\Sigma_{u\varepsilon} = 0$, but Assumption 3 holds for Case 3, whereas in Case 4, $\Sigma_{Z\varepsilon} \neq 0$. For each case we consider also two subcases: $a = 0.15$ and $a = 0$. The data were simulated according to (1) and (2) and the IV estimator was solved in each replication as a solution to the nonlinear moment condition (13). To assess the adequacy of the results of Section 4, we have also added Cases 5 and 6, which are Cases 1 and 3 with $\hat{\Sigma}_{Zu,n}$ replacing Σ_{Zu}. A 1%-trimming was enforced in the simulations because large outliers were encountered, some due to the limit distribution being a scaled ratio of normals and some due to the fact that large simulated u_t-values can result in large exponentials and consequently, numerically unstable results.

PP plots of the left side of (16) against its limit distribution (coded as right side) and against the estimated normal distribution of the left side were formed. A selection of the plots is presented and more details are provided in the working paper version (Lieberman and Phillips, 2016). The typical situation is given in Figures 1 and 2. As expected, for Cases 1, 3, and 5, in which Assumption 3 holds, the left side and right side in each case are very close for as little as $n = 100$. There is no noticeable difference between the $\Sigma_{u\varepsilon} = 0$ and $\Sigma_{u\varepsilon} \neq 0$ cases. On the other hand, as expected, the left and right sides become very different as n grows in Cases 2, 4, and 6, when $\Sigma_{Z\varepsilon} \neq 0$. Figure 3 reveals that replacing Σ_{Zu} by $\hat{\Sigma}_{Zu,n}$ does not cause a noticeable difference. In Cases 1, 3, and 5, in which Assumption 3 holds, the asymptotic distribution of the test statistic has a peaked

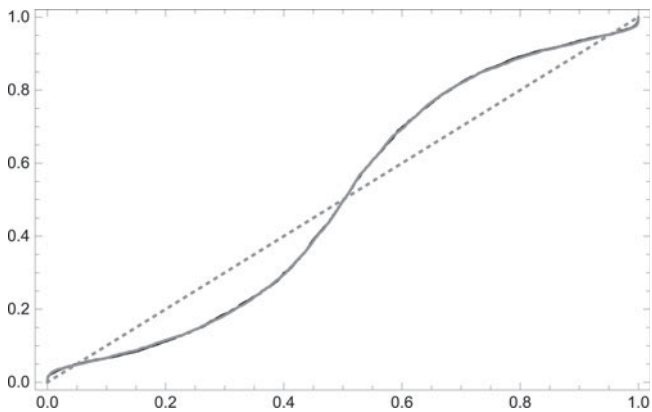


FIGURE 1. PP plots of the distributions of $\sqrt{n}\left(\hat{a}_n^{IV} - a_0\right)$ (black) and $B_{\varepsilon Z}(1) / \left(\sigma_{Zu} \int G_a(r)\, dr\right)$ (gray) against the estimated normal distribution of $\sqrt{n}\left(\hat{a}_n^{IV} - a_0\right)$, with $n = 100$, $\Sigma_{u\varepsilon} \neq 0$, $\Sigma_{Z\varepsilon} = 0$, $\sigma_{Zu} \neq 0$, $a = 0.15$.

https://doi.org/10.1017/S0266466617000330 Published online by Cambridge University Press

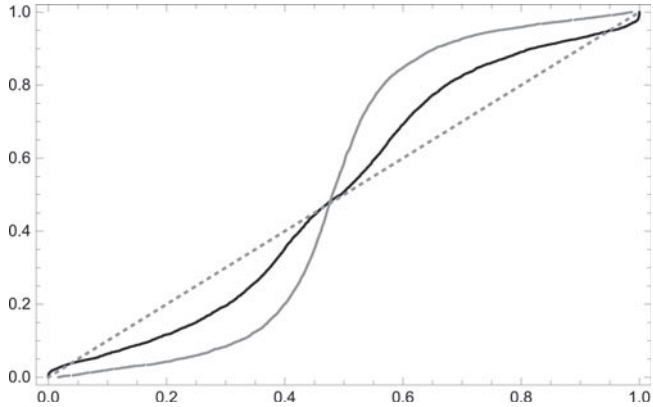


FIGURE 2. PP plots of the distributions of $\sqrt{n}\left(\hat{a}_n^{IV} - a_0\right)$ (black) and $B_{\varepsilon Z}(1) / \left(\sigma_{Zu} \int G_a(r)\, dr\right)$ (gray) against the estimated normal distribution of $\sqrt{n}\left(\hat{a}_n^{IV} - a_0\right)$, with $n = 1,000$, $\Sigma_{u\varepsilon} \neq 0$, $\Sigma_{Z\varepsilon} \neq 0$, $\sigma_{Zu} \neq 0$, $a = 0.15$.

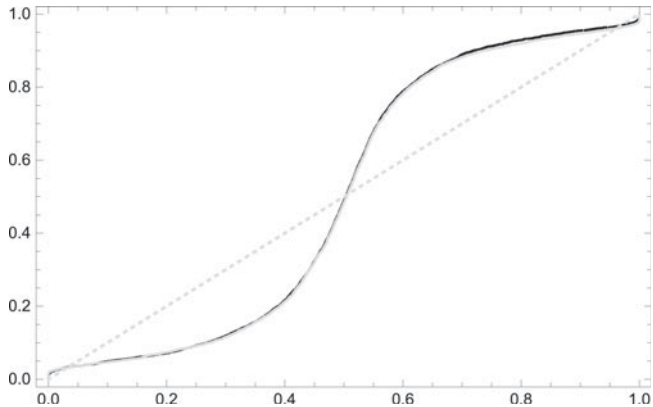


FIGURE 3. PP plots of the distributions of $\sqrt{n}\left(\hat{a}_n^{IV} - a_0\right)$ (black) and $B_{\varepsilon Z}(1) / \left(\hat{\sigma}_{Zu,n} \int G_a(r)\, dr\right)$ (gray) against the estimated normal distribution of $\sqrt{n}\left(\hat{a}_n^{IV} - a_0\right)$, with $n = 100$, $\Sigma_{u\varepsilon} \neq 0$, $\Sigma_{Z\varepsilon} = 0$, $\Sigma_{Zu} = \sigma_{Zu} \neq 0$, $a = 0.15$.

distribution compared with the normal distribution. This is evident in the PP plots and Figure 4, which provide kernel density estimates of the left and right sides and estimated normal density of the left side. Finally, for the case $a = 0$, the comparisons drawn in Figure 5 between the distribution of $\sqrt{n}\hat{a}_n^{IV}$ and the scaled Cauchy variate, given in (19), show a near perfect fit.

In the second part of the analysis we investigated the empirical p-values (PV) of the hypothesis test $H_0 : a = 0$ using (27). The results are of interest for applications with small to moderate n and/or a. Each of the experiments was based on 2,000 replications. We considered the process

https://doi.org/10.1017/S0266466617000330 Published online by Cambridge University Press

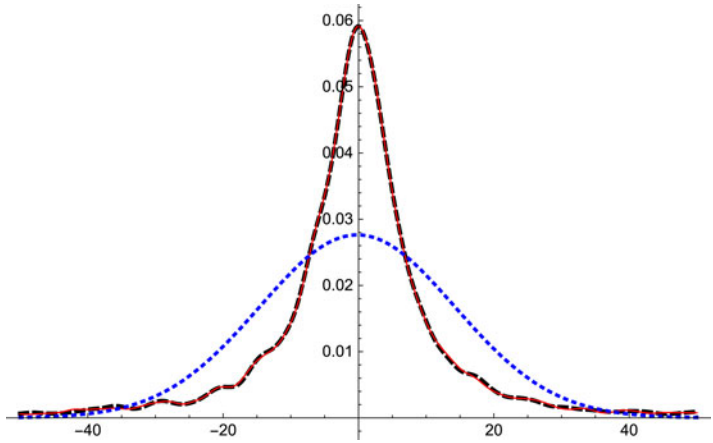


FIGURE 4. Kernel density estimates of $\sqrt{n}\left(\hat{a}_n^{IV} - a_0\right)$ (dashed black) and $B_{\varepsilon Z}(1) / \left(\sigma_{Zu} \int G_a(r)\, dr\right)$ (solid red) against the estimated normal distribution of $\sqrt{n}\left(\hat{a}_n^{IV} - a_0\right)$ (dotted blue), with $n = 10,000$, $\Sigma_{u\varepsilon} \neq 0$, $\Sigma_{Z\varepsilon} = 0$, $\sigma_{Zu} \neq 0$, $a = 0.15$.

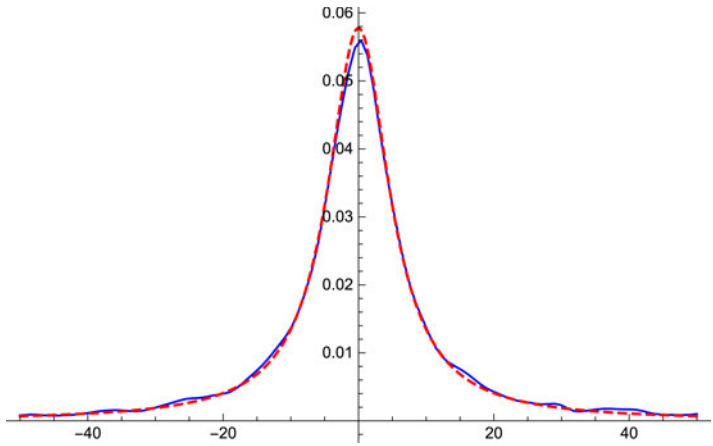


FIGURE 5. Kernel density estimate of $\sqrt{n}\left(\hat{a}_n^{IV} - a_0\right)$ (solid blue) against the scaled Cauchy variate (dashed red), $n = 100$, $\Sigma_{u\varepsilon} \neq 0$, $\Sigma_{Z\varepsilon} = 0$, $\sigma_{Zu} \neq 0$, $a = 0$.

$$\eta_{1t} \overset{iid}{\sim} N\left(0, \sigma_{\eta_1}^2\right),\ \eta_{2t} \overset{iid}{\sim} N\left(0, \sigma_{\eta_2}^2\right),\ \eta_{3t} \overset{iid}{\sim} N\left(0, \sigma_{\eta_3}^2\right), \tag{32}$$

$$\begin{aligned} &\sigma_{\eta_1}^2 = 0.673, \sigma_{\eta_2}^2 = 0.129, \sigma_{\eta_3}^2 = 0.5, \\ &u_t = \eta_{1,t} + 0.432\eta_{1,t-1} - 0.21\eta_{2,t-1}, \\ &\varepsilon_t = \eta_{2,t} - 0.251\eta_{1,t-1} + 0.12\eta_{2,t-1}, \\ &Z_t = \eta_{3,t} + 0.3\eta_{1t} + 0.4\eta_{3,t-1}. \end{aligned} \tag{33}$$

https://doi.org/10.1017/S0266466617000330 Published online by Cambridge University Press

In this setting, $Cov(u, \varepsilon) \neq 0$, $Cov(u, Z) \neq 0$ and $Cov(\varepsilon, Z) = 0$. For the first part of the analysis, we set $n = 2,000$ and varied the true a over the values 0, 0.2, 0.5, 1, 2, and 5. The results are given in Table 2 and Figure 6. For the second part of the analysis, we fixed a at 0.2 and 1 and varied n over the values 100, 500, 1,000, 1,500, 2,000, and 5,000. The results for this part of the analysis are given in Tables 3 and 4. In Table 2, clearly, as a increases, both the one-sided and two-sided PVs decrease from about 0.5 to about 0.01 and 0.02, respectively, and the sample mean of $\hat{a}_n^{IV}$ is accurate. In Tables 3 and 4 we also observe a decrease in the PVs and a decrease in the standard deviations of the estimator, as n increases, as expected.

For the same setting we simulated PVs of the DF statistic against the distribution of $\left(n^{-1}\sum_{t=1}^{n} Y_{t-1}\Delta Y_t\right) / \left(n^{-2}\sum_{t=1}^{n} Y_{t-1}^2\right)$ with Y_t being generated under

TABLE 2. Simulated p-values (PV) and estimates

a	0	0.2	0.5	1	2	5
One sided PV	0.502	0.311	0.179	0.096	0.042	0.010
DF one sided PV	0.500	0.500	0.496	0.501	0.449	0.134
Two sided PV	0.502	0.374	0.226	0.109	0.052	0.022
$\overline{\hat{a}_n}$	−0.047	0.231	0.482	0.999	2.0029	5.003
$\hat{\sigma}\left(\hat{a}_n\right)$	0.676	0.623	0.532	0.602	0.513	0.147

Note: $n = 2,000$, the number of replications is equal to 2,000. The values were obtained for the model (1)–(2) with a 1% trimming from each tail.

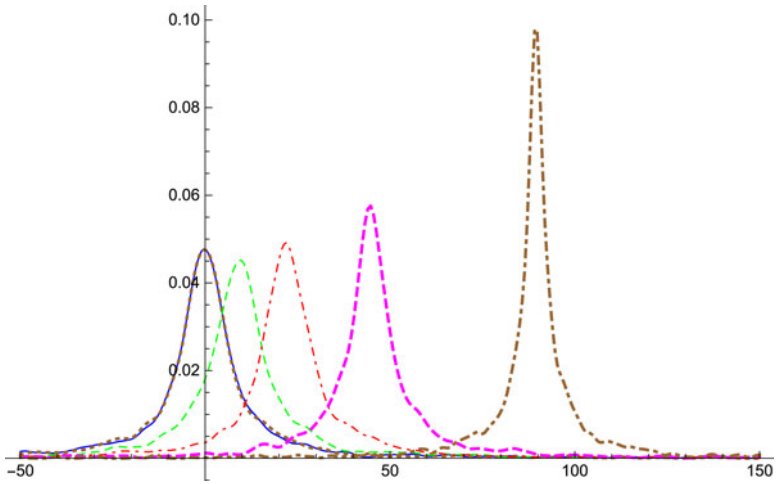


FIGURE 6. Trimmed (1% from each side) RHS (solid blue) and LHS kernel distributions, based on 2,000 replications with $n = 2,000$. $a = 0$ (dotted brown), $a = 0.2$ (dashed green), $a = 0.5$ (dot dashed red), $a = 1$ (bold dashed magenta), $a = 2$ (bold dot dashed brown).

https://doi.org/10.1017/S0266466617000330 Published online by Cambridge University Press

all the true parameters of the process as given by (32) and (33). Accordingly, the simulation reports performance of an 'ideal' DF test because the parameters $\Lambda_{\varepsilon\varepsilon}$ and σ_{ε}^2 that are needed for the simulation of the right side of (20) were taken to be known. Even with this prior advantage that the (one-sided) DF test has over this paper's coefficient test in which nuisance parameters are estimated, it is clear from Tables 2–4 that the DF test lacks power when it is applied, as is conventional, in a single direction. More specifically, Table 2 reveals that one-sided DF test PV does not change for small to moderate values of a and decreases very slowly (and much slower than the IV-STUR coefficient based test) as a increases from 1 to 5. Tables 3 and 4 show that for small to moderate values of a, as the sample size increases the PV of the DF test is essentially fixed and close to 0.5 over the full range $n \in \{100, 500, 1,000, 1,500, 2,000, 5,000\}$. In contrast, the IV-STUR coefficient test has PVs that decrease from 0.46 to 0.25 when $a = 0.2$ and from 0.3 to 0.06, when $a = 1$, over this range of sample sizes.

The performance of the DF test in these simulations is obviously affected by the (conventional) one-sided implementation of the test. The one-sided DF test is naturally expected to have power limited by the fact that in the STUR model we get mildly explosive departures 50% of the time with a symmetric u_t-distribution. In such (subperiod) cases the departures will not contribute power to a left sided UR test against stationarity. This manifests itself in the simulations. In summary,

TABLE 3. Simulated p-values (PV) and estimates

n	100	500	1,000	1,500	2,000	5,000
One sided PV	0.458	0.393	0.363	0.329	0.303	0.246
DF one sided PV	0.498	0.500	0.498	0.500	0.500	0.499
Two sided PV	0.508	0.458	0.431	0.399	0.371	0.297
$\overline{\hat{a}_n}$	0.196	0.173	0.177	0.163	0.231	0.200
$\hat{\sigma}\left(\hat{a}_n\right)$	1.719	1.202	0.817	0.812	0.623	0.415

Note: $a = 0.2$, the number of replications is equal to 2,000. The values were obtained for the model (1)–(2) with a 1% trimming from each tail.

TABLE 4. Simulated p-values (PV) and estimates

n	100	500	1,000	1,500	2,000	5,000
One sided PV	0.302	0.158	0.128	0.099	0.084	0.055
DF one sided PV	0.501	0.501	0.498	0.502	0.492	0.508
Two sided PV	0.394	0.211	0.161	0.114	0.108	0.065
$\overline{\hat{a}_n}$	0.893	0.967	1.029	1.024	0.966	0.990
$\hat{\sigma}\left(\hat{a}_n\right)$	1.723	1.033	0.969	0.668	0.534	0.443

Note: $a = 1$, the number of replications is equal to 2,000. The values were obtained for the model (1)–(2) with a 1% trimming from each tail.

https://doi.org/10.1017/S0266466617000330 Published online by Cambridge University Press

the one-sided directional IV-STUR test has good power which increases with both a and n, and in all cases gives better discriminatory power than the usual directional DF test against stationarity.

In the third part of the analysis we analyzed the small sample performance of the distributions of the efficient two-step efficient GMM estimator and the J_n-test for overidentifying restrictions based on it. The first step weighting matrix was taken to be $\hat{W} = n\left(\sum_{t=1}^{n} Z_t Z_t'\right)^{-1}$ – see, for instance, Hayashi (2000, p. 213). To this end we generated 5,000 samples of $n = 100, 500$, according to the following law:

$$\eta_t \sim N(0, I_4),\ \varepsilon_t = \eta_{1t},$$

$$u_t = g_{21}\eta_{1t} + \left(g_{22,0}\eta_{2t} + g_{22,1}\eta_{2,t-1}\right) + \left(g_{23,0}\eta_{3t} + g_{23,1}\eta_{3,t-1}\right) + \left(g_{24,0}\eta_{4t} + g_{24,1}\eta_{4,t-1}\right),$$

$$Z_{1t} = \left(g_{32,0}\eta_{2t} + g_{32,1}\eta_{2,t-1}\right) + \left(g_{33,0}\eta_{3t} + g_{33,1}\eta_{3,t-1}\right) + \left(g_{34,0}\eta_{4t} + g_{34,1}\eta_{4,t-1}\right),$$

$$Z_{2t} = \left(g_{42,0}\eta_{2t} + g_{42,1}\eta_{2,t-1}\right) + \left(g_{43,0}\eta_{3t} + g_{43,1}\eta_{3,t-1}\right) + \left(g_{44,0}\eta_{4t} + g_{44,1}\eta_{4,t-1}\right).$$

The g's were generated from a uniform distribution $U[0.1, 0.35]$ once only and $a = 0$ and 0.15. Clearly, Assumptions 2–3 are satisfied with $q = 2$ and $K = 1$. Additional figures supplied in the working paper version of the paper (Lieberman and Phillips, 2016) show that the PP-plots for the distribution of $\hat{a}_n^G$ against the right side of (29) are very accurate for as little as $n = 100$ observations, in both the $a = 0$ and $a \neq 0$ cases and that the asymptotic distribution of J_n against the $\chi^2(1)$ limit distribution is reasonable for $n = 100$ and is excellent for $n = 500$.

8. EMPIRICAL APPLICATION

We estimated the STUR model (1)–(2) with Y_t being the log spread between an index of U.S. dollar denominated investment grade rated corporate debt publically issued in the U.S. domestic market, and the spot Treasury curve. To qualify for inclusion in the index, securities must have an investment grade rating (based on an average of Moody's, S&P, and Fitch) and an investment grade rated country of risk (based on an average of Moody's, S&P, and Fitch foreign currency long term sovereign debt ratings). Each security must have greater than 1 year of remaining maturity, a fixed coupon schedule, and a minimum amount outstanding of $250 million. It was calculated from the BofA (Bank of America) Merrill Lynch US Corporate Master Option-Adjusted Spread. The variable u_t was taken to be the demeaned $100\log(SP_{US,t}/SP_{US,t-1})$, where $SP_{US,t}$ is the opening rate of the SPDR S&P 500 ETF Trust (SPY). The instrument Z_t is calculated as the demeaned $100\log(SP_{US,t}/SP_{AX,USD,t})$, where $SP_{AX,USD,t}$ is the Australian currency adjusted iShares Core S&P 500 ETF (IVV.AX) opening rate. The data are measured at the daily frequency and cover the period from January 5, 2010 to December 30, 2015, giving a total of 1,454 observations. The stock data were obtained from the Yahoo Finance Website (https://au.finance.yahoo.com/) and the

https://doi.org/10.1017/S0266466617000330 Published online by Cambridge University Press

Spread data were retrieved from the Federal Reserve Bank of St. Louis (FRED). For further details regarding the construction methodology of this series, the reader is referred to https://fred.stlouisfed.org/series/BAMLC0A0CM.

The model chosen for the application follows up the findings of Kwan (1996) that returns on stocks and yield changes on bonds are negatively correlated and that lagged stock returns have explanatory power for bond yield changes, so that stocks lead bonds in reflecting firm-specific information. To this end, it is emphasized that in our data the information on u_t and Z_t is recorded prior to the information on y_t on any given day.

The sample correlations in the data are given by $\hat{\rho}_{u,\Delta y} = -0.52$, $\hat{\rho}_{u,z} = 0.26$ and $\hat{\rho}_{z,\Delta y} = -0.08$. Thus, for the data under consideration, there is indeed a negative correlation between stock returns and bond spread changes, as Kwan (1996) reported. This also means that the NLLS estimates of the STUR model are inconsistent. The figures for $\hat{\rho}_{u,z}$ and $\hat{\rho}_{z,\Delta y}$ justify Z_t as a good instrument for u_t.

The results for the model estimation are as follows. First, $\hat{a}_n^{IV} = -0.245$ whereas $\hat{a}_n = -0.297$. The mean squared prediction error (MSPE) of the random walk model is 32.8% larger than the one obtained by the NLIV estimated STUR model. The ADF test of a unit root in y_t yielded a p-value of 0.76. For the hypothesis $H_0 : a = 0$, the simulated p-value using (18) is 0.188 and using (19) with covariance estimates it is 0.207. To assess how significant this value is, we simulated 2,000 replications of two processes with n given by the empirical sample size: a random walk with covariances given by the sample covariances of the data and a STUR model with $a = \hat{a}_n^{IV}$ and covariances as in the random walk simulation. Against $H_1 : a < 0$, the simulated p-value in the random walk model was 0.502, whereas in the STUR simulation it was 0.303. The ADF test p-values in the two simulations were 0.500. Table 3 also reveals that for a simulated STUR process with $a = 0.2$ and $n = 1,500$ observations (albeit with different covariance configurations), one should expect a p-value of 0.329. Therefore, the p-values obtained in the application, together with the NLIV estimate $\hat{a}_n^{IV} = -0.245$ and the improvement in terms of MSPE compared with the random walk model are all strongly indicative of a STUR process in the spread data. On the other hand, for n and a of the magnitudes given here, the ADF test does not appear to identify departures from the random walk model in the direction of a STUR alternative.

9. CONCLUSION

This paper explores a structural version of the STUR autoregressive model and extends the existing limit theory for both the output process and the nonlinear least squares estimator of the localizing STUR coefficient in the weakly dependent time series case. Just as in linear and nonlinear models involving only stationary variables, instrumental variables are shown here to be useful in providing consistent estimates of the localizing coefficients of the driver variables in the structural version of the STUR model under orthogonality and relevance conditions that mirror

https://doi.org/10.1017/S0266466617000330 Published online by Cambridge University Press

those used in other implementations of IV. The limit distribution of the nonlinear IV estimators in the just identified case turns out to be Cauchy-like and involves a bias term. The limit distribution of the Sargan-Hansen test for overidentifying restrictions turns out to be χ^2_{q-K} just as in the linear case, facilitating inference.

It is of particular interest in empirical applications of STUR models to be able to test for the presence of driver variables in determining the STUR coefficient. The coefficient-based test for the relevance of driver variables that is proposed in the present paper has a convenient limit theory and simulations show that its performance in finite samples is satisfactory. The theory is potentially useful in cases where the data generating process can only be approximately described by a unit root process and which is more likely to fit data with a time dependent coefficient that is influenced by covariates that may be endogenous. The IV and GMM procedures given here enable inference about structural STUR models and provide a mechanism for testing the local STUR model against a simple UR null. The STUR test appears to have promising power performance characteristics against standard directional UR coefficient tests both as the sample size rises and departures from the null increase. The STUR test gains this advantage from its greater specificity concerning the alternative model, which enables focus on potential driver variables that influence departures from unit root behavior.

NOTES

1. Primitive conditions under which the functional law (4) holds are given, for example, in Phillips and Solo (1992) and Phillips (1987b).

2. If $\xi = \frac{\xi_1}{\xi_2}$ where $(\xi_1, \xi_2) \sim \mathcal{N}\left(0, \Sigma^\xi\right)$ and $\Sigma^\xi > 0$ have components $\left\{\sigma^\xi_{11}, \sigma^\xi_{12}, \sigma^\xi_{22}\right\}$, then $\xi = \frac{\xi_{1.2}}{\xi_2} + \frac{\sigma^\xi_{12}}{\sigma^\xi_{22}}$, where $\xi_{1.2} = \xi_1 - \frac{\sigma^\xi_{12}}{\sigma^\xi_{22}}\xi_2 \sim \mathcal{N}\left(0, \sigma^\xi_{11.2}\right)$, and $\sigma^\xi_{11.2} = \sigma^\xi_{11} - \frac{\sigma^\xi_{12}}{\sigma^\xi_{22}}^2 > 0$. Since $\xi_{1.2}$ is independent of ξ_2, the ratio $\frac{\xi_{1.2}}{\xi_2} \sim \left(\frac{\sigma^\xi_{11.2}}{\sigma^\xi_{22}}\right)^{1/2} \mathbb{C}$, where $\mathbb{C}$ is standard Cauchy, and so $\xi = \frac{\sigma^\xi_{12}}{\sigma^\xi_{22}} + \left(\frac{\sigma^\xi_{11.2}}{\sigma^\xi_{22}}\right)^{1/2} \mathbb{C}$ is non-central Cauchy and has Cauchy tails.

3. The proofs of (21) and (22) are similar to that of Lemma 8 and are available on request from the authors.

REFERENCES

Chan, N.H. & C.Z. Wei (1987) Asymptotic Inference for nearly nonstationary AR(1) processes. *Annals of Statistics* 15, 1050–1063.

Chao, J. & P.C.B. Phillips (2017) Uniform Inference in Panel Autoregression. Cowles Foundation Discussion paper, Yale University.

Giraitis, L. & P.C.B. Phillips (2006) Uniform limit theory for stationary autoregression. *Journal of Time Series Analysis* 27, 51–60.

Hayashi, F. (2000) *Econometrics*. Princeton University Press.

Ibragimov, R. & P.C.B. Phillips (2008) Regression asymptotics using martingale convergence methods. *Econometric Theory* 24, 888–947.

https://doi.org/10.1017/S0266466617000330 Published online by Cambridge University Press

Kwan, S.H. (1996) Firm-specific information and the correlation between individual stocks and bonds. *Journal of Financial Economics* 40, 63–80.
Liang, H., P.C.B. Phillips, H. Wang, & Q. Wang (2016) Weak convergence to stochastic integrals for econometric applications. *Econometric Theory* 32, 1349–1375.
Lieberman, O. (2012) A similarity-based approach to time-varying coefficient non-stationary autoregression. *Journal of Time Series Analysis* 33, 484–502.
Lieberman, O. & P.C.B. Phillips (2014) Norming rates and limit theory for some time-varying coefficient autoregressions. *Journal of Time Series Analysis* 35, 592–623.
Lieberman, O. & P.C.B. Phillips (2016) IV and GMM Inference in Estimation and Testing of Multivariate Stochastic Unit Root Models. Cowles Foundation Discussion paper no. 2061, Yale University.
Lieberman, O. & P.C.B. Phillips (2017) A multivariate stochastic unit root model with an application to derivative pricing. *Journal of Econometrics* 196, 99–110.
Mikusheva, A. (2007) Uniform inference in autoregressive models. *Econometrica* 75, 1411–1452.
Mikusheva, A. (2012) One-dimensional inference in autoregressive models with the potential presence of a unit root. *Econometrica* 80, 173–212.
Phillips, P.C.B. (1987a) Towards a unified asymptotic theory for autoregression. *Biometrika* 74, 535–547.
Phillips, P.C.B. (1987b) Time series regression with a unit root. *Econometrica* 55, 277–301.
Phillips, P.C.B. & V. Solo (1992) Asymptotics for linear processes. *Annals of Statistics* 20, 971–1001.

APPENDIX

Proof of Lemma 1. In what follows, the notation $o_p(1)$ and $O_p(1)$ stands for orders in probability which are uniform wrt t. In view of the functional law (4), in an appropriately expanded probability space we may write, for $t = \lfloor nr \rfloor$ and any $r > 0$,

$$n^{-1/2}\sum_{j=1}^{t}\eta_j = B(t/n) + o_p(1), \tag{A.1}$$

so that

$$\begin{aligned}
n^{-1/2}Y_t &= n^{-1/2}\sum_{s=1}^{t} e^{\frac{a'}{\sqrt{n}}\sum_{j=s+1}^{t}u_j}\varepsilon_s + O_p\left(n^{-1/2}\right)\\
&= n^{-1/2}e^{\frac{a'}{\sqrt{n}}\sum_{j=1}^{t}u_j}\sum_{s=1}^{t} e^{-\frac{a'}{\sqrt{n}}\sum_{j=1}^{s}u_j}\varepsilon_s + O_p\left(n^{-1/2}\right)\\
&= n^{-1/2}e^{\{a'B_u(t/n)+o_p(1)\}}\sum_{s=1}^{t} e^{\left\{-\frac{a'}{\sqrt{n}}\sum_{j=0}^{s-1}u_j-\frac{a'}{\sqrt{n}}u_s\right\}}\varepsilon_s + O_p\left(n^{-1/2}\right)\\
&= n^{-1/2}e^{a'B_u(t/n)}\sum_{s=1}^{t} e^{-\{a'B_u((s-1)/n)+o_p(1)\}}\left(1-\frac{a'u_s}{\sqrt{n}}+O_p\left(n^{-1}\right)\right)\varepsilon_s + o_p(1)\\
&= e^{a'B_u(t/n)}\sum_{s=1}^{t} e^{-a'B_u((s-1)/n)}\frac{\varepsilon_s}{\sqrt{n}} - e^{a'B_u(t/n)}\sum_{s=1}^{t} e^{-a'B_u((s-1)/n)}\left(\frac{a'u_s\varepsilon_s}{n}\right) + o_p(1).
\end{aligned} \tag{A.2}$$

Noting that $\mathbb{E}\left(e^{-a'B_u(p)}\right)^2 < \infty$, the first term on the right side of (A.2) has the following limit in an appropriately defined space

https://doi.org/10.1017/S0266466617000330 Published online by Cambridge University Press

$$e^{a'B_u\left(\frac{t}{n}\right)}\sum_{s=1}^{t-1}e^{-a'B_u\left(\frac{s-1}{n}\right)}\frac{\varepsilon_s}{\sqrt{n}}$$
$$\Rightarrow\ e^{a'B_u(r)}\left\{\int_0^r e^{-a'B_u(p)}dB_\varepsilon(p)-\Lambda'_{u\varepsilon}a\int_0^r e^{-a'B_u(p)}dp\right\}=:G_a^*(r). \tag{A.3}$$

The limit (A.3) makes use of a result on the weak convergence to stochastic integrals with random drift of sample covariances involving functions of partial sums (see Ibragimov and Phillips, 2008, Thm. 3.1; Liang et al., 2016, Thms. 2.3 and 3.1). Again, as in (A.1), we assume that the probability space has been expanded to permit the representation of (A.3) as a limit in probability.

The second term on the right side of (A.2) is

$$\begin{aligned}-e^{a'B_u(t/n)}\sum_{s=1}^{t}e^{-a'B_u\left(\frac{s-1}{n}\right)}\left(\frac{a'u_s\varepsilon_s}{n}\right)&=-a'e^{a'B_u(t/n)}\sum_{s=1}^{t}e^{-a'B_u\left(\frac{s-1}{n}\right)}\\&\quad\times\left(\frac{u_s\varepsilon_s-\Sigma_{u\varepsilon}}{n}+\frac{\Sigma_{u\varepsilon}}{n}\right)\\&=-a'\Sigma_{u\varepsilon}e^{a'B_u(t/n)}\frac{1}{n}\sum_{s=1}^{t}e^{-a'B_u((s-1)/n)}+O_p\left(n^{-1/2}\right)\\&\Rightarrow\ -a'\Sigma_{u\varepsilon}e^{a'B_u(r)}\int_0^r e^{-a'B_u(p)}dp.\end{aligned} \tag{A.4}$$

Hence,

$$\begin{aligned}n^{-1/2}Y_{\lfloor nr\rfloor}&\Rightarrow G_a^*(r)-a'\Sigma_{u\varepsilon}e^{a'B_u(r)}\int_0^r e^{-a'B_u(p)}dp\\&=e^{a'B_u(r)}\left(\int_0^r e^{-a'B_u(p)}dB_\varepsilon(p)-\Lambda'_{u\varepsilon}a\int_0^r e^{-a'B_u(p)}dp-a'\Sigma_{u\varepsilon}\int_0^r e^{-a'B_u(p)}dp\right)\\&=e^{a'B_u(r)}\left(\int_0^r e^{-a'B_u(p)}dB_\varepsilon(p)-a'\Delta_{u\varepsilon}\int_0^r e^{-a'B_u(p)}dp\right),\end{aligned} \tag{A.5}$$

giving (6), as required. ■

The following lemma will be used in the sequel. The terms H, $v(r)$ and M_3, which appear in it, were defined in (7), (8), and (9), respectively.

LEMMA A.1. *Under Assumption 1,*

$$\begin{aligned}\sum_{t=2}^{n}u_t\varepsilon_tY_{t-1}&=n^{3/2}\Sigma_{u\varepsilon}\int_0^1 G_a(r)\,dr\\&\quad+n\left(\sum_{j=0}^{\infty}\left(G_{2,j}\otimes G_{1,j}\right)H\int_0^1 G_a(r)\,d\xi(r)\right.\\&\quad+\sum_{j=1}^{\infty}\left(G_{2,j}\otimes G_{1,j}\right)M_3\left(\left(\sum_{i=0}^{j-1}G_{1,i}\right)'a\int_0^1 G_a(r)\,dr+\left(\sum_{i=0}^{j-1}G_{2,i}\right)'\right)\\&\quad\left.+\sum_{j\neq k}\left(G_{2,k}\otimes G_{1,j}\right)\int_0^1 G_a(r)\,d\zeta(r)\right)+o_p(n).\end{aligned}$$

https://doi.org/10.1017/S0266466617000330 Published online by Cambridge University Press

Proof of Lemma A.1. Throughout we use column vectorization. The term under consideration is

$$\frac{1}{n^{3/2}}\sum_{t=2}^{n}u_t\varepsilon_t Y_{t-1}=\frac{1}{n^{3/2}}\sum_{t=2}^{n}\left(\sum_{j=0}^{\infty}G_{1,j}\eta_{t-j}\right)\left(\sum_{j=0}^{\infty}G_{2,j}\eta_{t-j}\right)' Y_{t-1}$$
$$=\frac{1}{n^{3/2}}\sum_{t=2}^{n}\left(\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}G_{1,j}\eta_{t-j}\eta'_{t-k}G_{2,k}-\sum_{j=0}^{\infty}G_{1,j}\Sigma_\eta G'_{2,j}+\Sigma_{u\varepsilon}Y_{t-1}\right).$$

The leading term is

$$\frac{1}{n^{3/2}}\Sigma_{u\varepsilon}\sum_{t=2}^{n}Y_{t-1}\Rightarrow\Sigma_{u\varepsilon}\int_0^1 G_a(r)\,dr,$$

but we also need to consider

$$\frac{1}{n}\sum_{t=2}^{n}\left(\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}G_{1,j}\eta_{t-j}\eta'_{t-k}G'_{2,k}-\sum_{j=0}^{\infty}G_{1,j}\Sigma_\eta G'_{2,j}\right)Y_{t-1}.$$

First,

$$\frac{1}{n}\sum_{t=2}^{n}vec\left(G_{1,0}\left(\eta_t\eta'_t-\Sigma_\eta\right)G'_{2,0}\right)Y_{t-1}=\frac{1}{n}\left(G_{2,0}\otimes G_{1,0}\right)H\sum_{t=2}^{n}vech\left(\eta_t\eta'_t-\Sigma_\eta\right)Y_{t-1}$$
$$\Rightarrow\left(G_{2,0}\otimes G_{1,0}\right)H\int_0^1 G_a(r)\,d\upsilon(r). \qquad \textbf{(A.6)}$$

Further, since $E\left(G_{1,0}\eta_t\eta'_{t-1}G'_{2,1}\right)=0$, we have

$$\frac{1}{n}\sum_{t=2}^{n}\left(G_{1,0}\eta_t\eta'_{t-1}G'_{2,1}-E\left(G_{1,0}\eta_t\eta'_{t-1}G'_{2,1}\right)\right)Y_{t-1}$$
$$=\frac{1}{n}\sum_{t=2}^{n}G_{1,0}\eta_t\eta'_{t-1}G'_{2,1}\left\{\left(1+\frac{a'u_{t-1}}{\sqrt{n}}+o_p\left(n^{-1/2}\right)\right)Y_{t-2}+\varepsilon_{t-1}\right\}$$
$$=\frac{1}{n}\sum_{t=2}^{n}G_{1,0}\eta_t\eta'_{t-1}G'_{2,1}Y_{t-2}+\frac{1}{n^{3/2}}\sum_{t=2}^{n}G_{1,0}\eta_t\eta'_{t-1}G'_{2,1}a'u_{t-1}Y_{t-2}$$
$$+\frac{1}{n}\sum_{t=2}^{n}G_{1,0}\eta_t\eta'_{t-1}G'_{2,1}\varepsilon_{t-1}+o_p(1). \qquad \textbf{(A.7)}$$

As in the developments leading to (A.6), but because $\eta_t\eta'_{t-1}$ is not a symmetric matrix, the first term in (A.7) yields

$$\frac{1}{n}\sum_{t=2}^{n}G_{1,0}\eta_t\eta'_{t-1}G'_{2,1}Y_{t-2}=\left(G_{2,1}\otimes G_{1,0}\right)\frac{1}{n}\sum_{t=2}^{n}vec\left(\eta_t\eta'_{t-1}\right)Y_{t-2}$$
$$\Rightarrow\left(G_{2,1}\otimes G_{1,0}\right)\int_0^1 G_a(r)\,d\zeta(r), \qquad \textbf{(A.8)}$$

https://doi.org/10.1017/S0266466617000330 Published online by Cambridge University Press

where $\zeta(r)$ is defined following (8). The second term in (A.7) yields

$$\frac{1}{n^{3/2}}\sum_{t=2}^{n} G_{1,0}\eta_t\eta_{t-1}'G_{2,1}'a'u_{t-1}Y_{t-2} = \frac{1}{n^{3/2}}\sum_{t=2}^{n} E\left(G_{1,0}\eta_t\eta_{t-1}'G_{2,1}'a'u_{t-1}\right)Y_{t-2} + o_p(1)$$
$$= o_p(1).$$

The third term in (A.7) yields

$$\frac{1}{n}\sum_{t=2}^{n} G_{1,0}\eta_t\eta_{t-1}'G_{2,1}'\varepsilon_{t-1} = E\left(G_{1,0}\eta_t\eta_{t-1}'G_{2,1}'\varepsilon_{t-1}\right) + o_p(1) = o_p(1).$$

Therefore,

$$\frac{1}{n}\sum_{t=2}^{n}\left(G_{1,0}\eta_t\eta_{t-1}'G_{2,1}' - E\left(G_{1,0}\eta_t\eta_{t-1}'G_{2,1}'\right)\right)Y_{t-1} \Rightarrow \left(G_{2,1}\otimes G_{1,0}\right)\int_0^1 G_a(r)\,d\zeta(r).$$

Further,

$$\begin{aligned}
&\frac{1}{n}\sum_{t=2}^{n}\left(G_{1,1}\eta_{t-1}\eta_{t-1}'G_{2,1}' - E\left(G_{1,1}\eta_{t-1}\eta_{t-1}'G_{2,1}'\right)\right)Y_{t-1}\\
&\quad= \frac{1}{n}\sum_{t=2}^{n} G_{1,1}\left(\eta_{t-1}\eta_{t-1}' - \Sigma_\eta\right)G_{2,1}'\left\{\left(1+\frac{a'u_{t-1}}{\sqrt{n}} + o_p\left(n^{-1/2}\right)\right)Y_{t-2} + \varepsilon_{t-1}\right\}\\
&\quad= \frac{1}{n}\sum_{t=2}^{n} G_{1,1}\left(\eta_{t-1}\eta_{t-1}' - \Sigma_\eta\right)G_{2,1}'Y_{t-2} + \frac{1}{n^{3/2}}\sum_{t=2}^{n} G_{1,1}\left(\eta_{t-1}\eta_{t-1}' - \Sigma_\eta\right)G_{2,1}'a'u_{t-1}Y_{t-2}\\
&\qquad+ \frac{1}{n}\sum_{t=2}^{n} G_{1,1}\left(\eta_{t-1}\eta_{t-1}' - \Sigma_\eta\right)G_{2,1}'\varepsilon_{t-1} + o_p(1). \qquad \textbf{(A.9)}
\end{aligned}$$

As in (A.6), the first term in (A.9) is

$$\frac{1}{n}\sum_{t=2}^{n} G_{1,1}\left(\eta_{t-1}\eta_{t-1}' - \Sigma_\eta\right)G_{2,1}'Y_{t-2} \Rightarrow \left(G_{2,1}\otimes G_{1,1}\right)H\int_0^1 G_a(r)\,d\upsilon(r).$$

The second term in (A.9) is

$$\begin{aligned}
&\frac{1}{n^{3/2}}\sum_{t=2}^{n} G_{1,1}\left(\eta_{t-1}\eta_{t-1}' - \Sigma_\eta\right)G_{2,1}'a'u_{t-1}Y_{t-2}\\
&\quad\Rightarrow E\left(G_{1,1}\left(\eta_{t-1}\eta_{t-1}' - \Sigma_\eta\right)G_{2,1}'a'u_{t-1}\right)\int_0^1 G_a(r)\,dr\\
&\quad= E\left[G_{1,1}\left(\eta_{t-1}\eta_{t-1}' - \Sigma_\eta\right)G_{2,1}'\left(a'G_{1,0}\eta_{t-1}\right)\right]\int_0^1 G_a(r)\,dr\\
&\quad= E\left[G_{1,1}\eta_{t-1}\eta_{t-1}'G_{2,1}'\left(a'G_{1,0}\eta_{t-1}\right)\right]\int_0^1 G_a(r)\,dr\\
&\quad= E\left[G_{1,1}\eta_{t-1}\eta_{t-1}'G_{2,1}'\eta_{t-1}'G_{1,0}'a\right]\int_0^1 G_a(r)\,dr
\end{aligned}$$

https://doi.org/10.1017/S0266466617000330 Published online by Cambridge University Press

$$= \left(G_{2,1} \otimes G_{1,1}\right) E\left(\left(\eta_{t-1} \otimes \eta_{t-1}\right) \eta'_{t-1}\right) G'_{1,0} a \int_0^1 G_a(r)\, dr$$

$$= \left(G_{2,1} \otimes G_{1,1}\right) M_3 G'_{1,0} a \int_0^1 G_a(r)\, dr.$$

The third term in (A.9) is

$$\frac{1}{n}\sum_{t=2}^{n} G_{1,1}\left(\eta_{t-1}\eta'_{t-1} - \Sigma_\eta\right) G'_{2,1}\varepsilon_{t-1} \Rightarrow E\left[G_{1,1}\left(\eta_{t-1}\eta'_{t-1} - \Sigma_\eta\right) G'_{2,1}\varepsilon_{t-1}\right]$$

$$= E\left[G_{1,1}\eta_{t-1}\eta'_{t-1}G'_{2,1}G_{2,0}\eta_{t-1}\right]$$

$$= \left(G_{2,1} \otimes G_{1,1}\right) M_3 G'_{2,0}.$$

Hence,

$$\frac{1}{n}\sum_{t=2}^{n}\left(G_{1,1}\eta_{t-1}\eta'_{t-1}G'_{2,1} - E\left(G_{1,1}\eta_{t-1}\eta'_{t-1}G'_{2,1}\right)\right) Y_{t-1}$$

$$\Rightarrow \left(G_{2,1} \otimes G_{1,1}\right)\left(H\int_0^1 G_a(r)\, d\upsilon(r) + M_3\left(G'_{1,0} a \int_0^1 G_a(r)\, dr + G'_{2,0}\right)\right).$$

Further,

$$Y_{t-1} = \exp\left(\frac{a' u_{t-1}}{\sqrt{n}}\right) Y_{t-2} + \varepsilon_{t-1}$$

$$= \exp\left(\frac{a' u_{t-1}}{\sqrt{n}}\right)\left(\exp\left(\frac{a' u_{t-2}}{\sqrt{n}}\right) Y_{t-3} + \varepsilon_{t-2}\right) + \varepsilon_{t-1}$$

$$= \exp\left(\frac{a'\left(u_{t-1} + u_{t-2}\right)}{\sqrt{n}}\right) Y_{t-3} + \exp\left(\frac{a' u_{t-1}}{\sqrt{n}}\right)\varepsilon_{t-2} + \varepsilon_{t-1}.$$

So

$$\frac{1}{n}\sum_{t=2}^{n}\left(G_{1,1}\eta_{t-1}\eta'_{t-2}G'_{2,2} - E\left(G_{1,1}\eta_{t-1}\eta'_{t-2}G'_{2,2}\right)\right) Y_{t-1}$$

$$= \frac{1}{n}\sum_{t=2}^{n} G_{1,1}\eta_{t-1}\eta'_{t-2}G'_{2,2}Y_{t-1} = \frac{1}{n}\sum_{t=2}^{n} G_{1,1}\eta_{t-1}\eta'_{t-2}G'_{2,2}Y_{t-3}$$

$$+ \frac{1}{n^{3/2}}\sum_{t=2}^{n} G_{1,1}\eta_{t-1}\eta'_{t-2}G'_{2,2}a'\left(u_{t-1} + u_{t-2}\right) Y_{t-3}$$

$$+ \frac{1}{n}\sum_{t=2}^{n} G_{1,1}\eta_{t-1}\eta'_{t-2}G'_{2,2}\exp\left(\frac{a' u_{t-1}}{\sqrt{n}}\right)\varepsilon_{t-2}$$

$$+ \frac{1}{n}\sum_{t=2}^{n} G_{1,1}\eta_{t-1}\eta'_{t-2}G'_{2,2}\varepsilon_{t-1}. \qquad \textbf{(A.10)}$$

https://doi.org/10.1017/S0266466617000330 Published online by Cambridge University Press

Using (A.8), we obtain

$$\frac{1}{n}\sum_{t=2}^{n} G_{1,1}\eta_{t-1}\eta_{t-2}'G_{2,2}'Y_{t-3} \Rightarrow \left(G_{2,2}\otimes G_{1,1}\right)\int_0^1 G_a(r)\,d\zeta(r). \tag{A.11}$$

The second term in (A.10) is

$$\frac{1}{n^{3/2}}\sum_{t=2}^{n} G_{1,1}\eta_{t-1}\eta_{t-2}'G_{2,2}'a'\left(u_{t-1}+u_{t-2}\right)Y_{t-3} = o_p(1).$$

The third term in (A.10) is

$$\begin{aligned}\frac{1}{n}\sum_{t=2}^{n} G_{1,1}\eta_{t-1}\eta_{t-2}'G_{2,2}'\exp\left(\frac{a'u_{t-1}}{\sqrt{n}}\right)\varepsilon_{t-2} &= \frac{1}{n}\sum_{t=2}^{n} G_{1,1}\eta_{t-1}\eta_{t-2}'G_{2,2}'\varepsilon_{t-2}+o_p(1)\\ &= o_p(1)\end{aligned}$$

and the last term in (A.10) is $\frac{1}{n}\sum_{t=2}^{n} G_{1,1}\eta_{t-1}\eta_{t-2}'G_{2,2}'\varepsilon_{t-1} = o_p(1)$. Therefore,

$$\frac{1}{n}\sum_{t=2}^{n}\left(G_{1,1}\eta_{t-1}\eta_{t-2}'G_{2,2}'-E\left(G_{1,1}\eta_{t-1}\eta_{t-2}'G_{2,2}'\right)\right)Y_{t-1} \Rightarrow \left(G_{2,2}\otimes G_{1,1}\right)\int_0^1 G_a(r)\,d\zeta(r). \tag{A.12}$$

Further,

$$\begin{aligned}&\frac{1}{n}\sum_{t=2}^{n}\left(G_{1,2}\eta_{t-2}\eta_{t-2}'G_{2,2}'-E\left(G_{1,2}\eta_{t-2}\eta_{t-2}'G_{2,2}'\right)\right)Y_{t-1}\\ &\quad=\frac{1}{n}\sum_{t=2}^{n} G_{1,2}\left(\eta_{t-2}\eta_{t-2}'-\Sigma_\eta\right)G_{2,2}'Y_{t-1} = \frac{1}{n}\sum_{t=2}^{n} G_{1,2}\left(\eta_{t-2}\eta_{t-2}'-\Sigma_\eta\right)G_{2,2}'Y_{t-3}\\ &\qquad+\frac{1}{n^{3/2}}\sum_{t=2}^{n} G_{1,2}\left(\eta_{t-2}\eta_{t-2}'-\Sigma_\eta\right)G_{2,2}'a'\left(u_{t-1}+u_{t-2}\right)Y_{t-3}\\ &\qquad+\frac{1}{n}\sum_{t=2}^{n} G_{1,2}\left(\eta_{t-2}\eta_{t-2}'-\Sigma_\eta\right)G_{2,2}'\exp\left(\frac{a'u_{t-1}}{\sqrt{n}}\right)\varepsilon_{t-2}\\ &\qquad+\frac{1}{n}\sum_{t=2}^{n} G_{1,2}\left(\eta_{t-2}\eta_{t-2}'-\Sigma_\eta\right)G_{2,2}'\varepsilon_{t-1}.\end{aligned} \tag{A.13}$$

The first term in (A.13) is

$$\frac{1}{n}\sum_{t=2}^{n} G_{1,2}\left(\eta_{t-2}\eta_{t-2}'-\Sigma_\eta\right)G_{2,2}'Y_{t-3} \Rightarrow \left(G_{2,2}\otimes G_{1,2}\right)H\int_0^1 G_a(r)\,d\upsilon(r).$$

https://doi.org/10.1017/S0266466617000330 Published online by Cambridge University Press

The second term in (A.13) is

$$\frac{1}{n^{3/2}}\sum_{t=2}^{n} G_{1,2}\left(\eta_{t-2}\eta_{t-2}'-\Sigma_\eta\right)G_{2,2}'a'\left(u_{t-1}+u_{t-2}\right)Y_{t-3}$$
$$\Rightarrow E\left(G_{1,2}\eta_{t-2}\eta_{t-2}'G_{2,2}'\left(\left(G_{1,1}+G_{1,0}\right)\eta_{t-2}\right)'a\right)\int_0^1 G_a(r)\,dr$$
$$=\left(G_{2,2}\otimes G_{1,2}\right)M_3\left(G_{1,1}+G_{1,0}\right)'a\int_0^1 G_a(r)\,dr.$$

The third term in (A.13) is

$$\frac{1}{n}\sum_{t=2}^{n} G_{1,2}\left(\eta_{t-2}\eta_{t-2}'-\Sigma_\eta\right)G_{2,2}'\exp\left(\frac{a'u_{t-1}}{\sqrt{n}}\right)\varepsilon_{t-2}$$
$$\Rightarrow E\left(G_{1,2}\left(\eta_{t-2}\eta_{t-2}'-\Sigma_\eta\right)G_{2,2}'\varepsilon_{t-2}\right)=E\left(G_{1,2}\eta_{t-2}\eta_{t-2}'G_{2,2}'G_{2,0}\eta_{t-2}\right)$$
$$=\left(G_{2,2}\otimes G_{1,2}\right)M_3G_{2,0}'.$$

The fourth term in (A.13) is

$$\frac{1}{n}\sum_{t=2}^{n} G_{1,2}\left(\eta_{t-2}\eta_{t-2}'-\Sigma_\eta\right)G_{2,2}'\varepsilon_{t-1}\Rightarrow E\left(G_{1,2}\eta_{t-2}\eta_{t-2}'G_{2,2}'G_{2,1}\eta_{t-2}\right)$$
$$=\left(G_{2,2}\otimes G_{1,2}\right)M_3G_{2,1}'.$$

Therefore,

$$\frac{1}{n}\sum_{t=2}^{n}\left(G_{1,2}\eta_{t-2}\eta_{t-2}'G_{2,2}'-E\left(G_{1,1}\eta_{t-2}\eta_{t-2}'G_{2,2}'\right)\right)Y_{t-1}$$
$$\Rightarrow\left(G_{2,2}\otimes G_{1,2}\right)\left(H\int_0^1 G_a(r)\,dB_\xi(r)+M_3\left(\left(G_{1,1}+G_{1,0}\right)'a\int_0^1 G_a(r)\,dr+\left(G_{2,0}+G_{2,1}\right)'\right)\right).$$

In view of the above we deduce that for $j\neq k$,

$$\frac{1}{n}\sum_{t=2}^{n}\left(G_{1,j}\eta_{t-j}\eta_{t-k}'G_{2,k}'-E\left(G_{1,j}\eta_{t-j}\eta_{t-k}'G_{2,k}'\right)\right)Y_{t-1}\Rightarrow\left(G_{2,k}\otimes G_{1,j}\right)\int_0^1 G_a(r)\,d\zeta(r),$$

whereas for $j=k$,

$$\frac{1}{n}\sum_{t=2}^{n}\left(G_{1,j}\eta_{t-j}\eta_{t-j}'G_{2,j}'-E\left(G_{1,j}\eta_{t-j}\eta_{t-j}'G_{2,j}'\right)\right)Y_{t-1}$$
$$\Rightarrow\left(G_{2,j}\otimes G_{1,j}\right)\left(H\int_0^1 G_a(r)\,dv(r)\right.$$
$$\left.+M_3\times 1\{j\geq 1\}\left(\left(\sum_{i=0}^{j-1}G_{1,i}\right)'a\int_0^1 G_a(r)\,dr+\left(\sum_{i=0}^{j-1}G_{2,i}\right)'\right)\right).$$

The result of the lemma follows upon aggregation over the indices j and k. ■

https://doi.org/10.1017/S0266466617000330 Published online by Cambridge University Press

Proof of Theorem 2. We trace through the proof of Theorem 2 of LP and extend the derivations there to the linear process case. The objective function is

$$Q_n(a) = \sum_{t=2}^{n} \{Y_t - \beta_t(a) Y_{t-1}\}^2. \tag{A.14}$$

Minimizing (A.14) with respect to a yields

$$\begin{aligned}
\dot{Q}_n(\hat{a}_n) &= -2\sum_{t=2}^{n} \{Y_t - \beta_t(\hat{a}_n) Y_{t-1}\} \dot{\beta}_t(\hat{a}_n) Y_{t-1} = 0 \\
&\Longrightarrow \sum_{t=2}^{n} \{Y_t - \beta_t(\hat{a}_n) Y_{t-1}\} u_t \beta_t(\hat{a}_n) Y_{t-1} = 0 \\
&\Longrightarrow \sum_{t=2}^{n} Y_t u_t \beta_t(\hat{a}_n) Y_{t-1} = \sum_{t=2}^{n} u_t \beta_t^2(\hat{a}_n) Y_{t-1}^2.
\end{aligned} \tag{A.15}$$

The third line in (A.15) is equivalent to

$$\sum_{t=2}^{n} u_t \beta_t(\hat{a}_n) \{\beta_t(a) Y_{t-1} + \varepsilon_t\} Y_{t-1} = \sum_{t=2}^{n} u_t \beta_t^2(\hat{a}_n) Y_{t-1}^2$$

or

$$\sum_{t=2}^{n} u_t \beta_t^2(\hat{a}_n) Y_{t-1}^2 - \sum_{t=2}^{n} u_t \beta_t(\hat{a}_n + a) Y_{t-1}^2 = \sum_{t=2}^{n} u_t \varepsilon_t \beta_t(\hat{a}_n) Y_{t-1}.$$

As $\beta_t^2(\hat{a}_n) = \beta_t(2\hat{a}_n)$, to second order the last expression equals

$$\begin{aligned}
&\frac{1}{\sqrt{n}} \sum_{t=2}^{n} u_t u_t'(\hat{a}_n - a) Y_{t-1}^2 + \frac{1}{2n} \sum_{t=2}^{n} u_t \left((2\hat{a}_n' u_t)^2 - \left((\hat{a}_n + a)' u_t \right)^2 \right) Y_{t-1}^2 \\
&\quad = \sum_{t=2}^{n} u_t \varepsilon_t Y_{t-1} + \frac{1}{\sqrt{n}} \sum_{t=2}^{n} \varepsilon_t u_t u_t' \hat{a}_n Y_{t-1}.
\end{aligned} \tag{A.16}$$

Now,

$$\begin{aligned}
&\frac{1}{\sqrt{n}} \sum_{t=2}^{n} u_t u_t'(\hat{a}_n - a) Y_{t-1}^2 \\
&\quad = n^{3/2} \Sigma_u (\hat{a}_n - a) \int_0^1 G_a^2(r)\, dr + n \sum_{t=2}^{n} \left(\frac{u_t u_t' - \Sigma_u}{\sqrt{n}} \right) (\hat{a}_n - a) \left(\frac{Y_{t-1}}{\sqrt{n}} \right)^2.
\end{aligned}$$

The second term above is $O_p\left(n(\hat{a}_n - a)\right)$. In the $\Sigma_{u\varepsilon} \neq 0$ case, we only need to maintain terms in (A.16) which are $O_p\left(n^{3/2}\right)$, so the second term above can be neglected. In the $\Sigma_{u\varepsilon} = 0$ case, we only need to retain terms which are $O_p(n)$. Because in this case

https://doi.org/10.1017/S0266466617000330 Published online by Cambridge University Press

$\hat{a}_n - a = O_p\left(n^{-1/2}\right)$, this term is negligible also in this case. Thus, we only need to consider in both cases

$$\frac{1}{\sqrt{n}}\sum_{t=2}^{n} u_t u_t' \left(\hat{a}_n - a\right) Y_{t-1}^2 \sim n^{3/2}\Sigma_u \left(\hat{a}_n - a\right) \int_0^1 G_a^2 (r)\, dr.$$

The next term to consider is

$$\frac{1}{2n}\sum_{t=2}^{n} u_t \left(\left(2\hat{a}_n' u_t\right)^2 - \left(\left(\hat{a}_n + a\right)' u_t\right)^2\right) Y_{t-1}^2 = \frac{1}{2n}\sum_{t=2}^{n} u_t \left\{\left(\hat{a}_n - a\right)' u_t \left(3\hat{a}_n + a\right)' u_t\right\} Y_{t-1}^2.$$

This term is $O_p(n)$ in the $\Sigma_{u\varepsilon} \neq 0$ case and $O_p\left(n^{1/2}\right)$ in the $\Sigma_{u\varepsilon} = 0$ case. So, this term is negligible in both cases. The next term in (A.16) is $\sum_{t=2}^{n} u_t \varepsilon_t Y_{t-1}$, whose behavior is given by Lemma A.1. The last term in (A.16) is

$$\frac{1}{\sqrt{n}}\sum_{t=2}^{n} \varepsilon_t u_t u_t' \hat{a}_n Y_{t-1} = n\mathbb{E}\left(\varepsilon_t u_t u_t' \hat{a}_n\right) \int_0^1 G_a (r)\, dr + o_p (n).$$

Ignoring negligible terms, we thus seek a solution to (A.16) as

$$\begin{aligned}
&n^{3/2}\Sigma_u \left(\hat{a}_n - a\right) \int_0^1 G_a^2 (r)\, dr \\
&\quad = n^{3/2}\Sigma_{u\varepsilon} \int_0^1 G_a (r)\, dr + n\left(\sum_{j=0}^{\infty} \left(G_{2,j} \otimes G_{1,j}\right) H \int_0^1 G_a (r)\, d\xi (r)\right. \\
&\qquad + \sum_{j=1}^{\infty} \left(G_{2,j} \otimes G_{1,j}\right) M_3 \left(\left(\left(\sum_{i=0}^{j-1} G_{1,i}\right)' a \int_0^1 G_a (r)\, dr + \left(\sum_{i=0}^{j-1} G_{2,i}\right)'\right)\right. \\
&\qquad + \sum_{j\neq k} \left(G_{2,k} \otimes G_{1,j}\right) \int_0^1 G_a (r)\, d\zeta (r) \\
&\qquad \left.\left. + \mathbb{E}\left(\varepsilon_t u_t u_t' \hat{a}_n\right) \int_0^1 G_a (r)\, dr\right),
\end{aligned}$$

giving the results stated in the theorem. ■

Proof of Theorem 3. Expanding the moment conditions (13) yields the (asymptotically equivalent) equation

$$\begin{aligned}
&\sum_{t=2}^{n} \left(Y_t - \beta_t \left(\hat{a}_n^{IV}; n\right) Y_{t-1}\right) Z_t \\
&\quad = \sum_{t=2}^{n} \left(\beta_t (a; n)\, Y_{t-1} + \varepsilon_t - \beta_t \left(\hat{a}_n^{IV}; n\right) Y_{t-1}\right) Z_t \\
&\quad = \sum_{t=2}^{n} \left\{\beta_t (a; n) - \beta_t \left(\hat{a}_n^{IV}; n\right)\right\} Y_{t-1} Z_t + \sum_{t=2}^{n} Z_t \varepsilon_t
\end{aligned}$$

https://doi.org/10.1017/S0266466617000330 Published online by Cambridge University Press

$$= \sum_{t=2}^{n} \left\{ \frac{(a - \hat{a}_n^{IV})' u_t}{\sqrt{n}} + \frac{(a' u_t)^2 - (\hat{a}_n^{IV\prime} u_t)^2}{2n} + o_p\left(\frac{1}{n}\right) \right\} Y_{t-1} Z_t + \sum_{t=2}^{n} Z_t \varepsilon_t$$

$$= 0. \tag{A.17}$$

Under Assumption 2,

$$n^{-1/2} \sum_{t=1}^{\lfloor nr \rfloor} vec\left(Z_t u_t' - \Sigma_{Zu}\right) \Rightarrow B_{Zu}(r),$$

where $B_{Zu}(r)$ is vector Brownian motion with covariance matrix

$$\Sigma_{Z\otimes u}^{\ell r} = \sum_{h=-\infty}^{\infty} \Gamma_{Z\otimes u}(h)$$

and

$$\Gamma_{Z\otimes u}(h) = \mathbb{E}\left(Z_t Z_{t+h}' \otimes u_t u_{t+h}'\right) - \mathbb{E}\left(Z_t \otimes u_t\right) \mathbb{E}\left(Z_t' \otimes u_t'\right).$$

By an application of Lemma A.1,

$$\sum_{t=1}^{n} Y_{t-1} Z_t u_t' = n^{3/2} \Sigma_{Zu} \int_0^1 G_a(r)\, dr + O_p(n). \tag{A.18}$$

Hence,

$$\frac{\left(\hat{a}_n^{IV} - a\right)'}{\sqrt{n}} \sum_{t=2}^{n} u_t Y_{t-1} Z_t = \frac{1}{\sqrt{n}} \left(\sum_{t=2}^{n} Y_{t-1} Z_t u_t' \right) \left(\hat{a}_n^{IV} - a\right) = O_p\left(n\left(\hat{a}_n^{IV} - a\right)\right). \tag{A.19}$$

Next,

$$\left(a' u_t\right)^2 - \left(\hat{a}_n^{IV\prime} u_t\right)^2 = \left(a - \hat{a}_n^{IV}\right)' u_t \left(a + \hat{a}_n^{IV}\right)' u_t.$$

Therefore, the leading term in the factor

$$\sum_{t=2}^{n} \left\{ \frac{\left(a' u_t\right)^2 - \left(\hat{a}_n^{IV\prime} u_t\right)^2}{2n} \right\} Y_{t-1} Z_t$$

which appears in (A.17), is

$$\frac{\sqrt{n}}{2} \mathbb{E}\left\{ \left(a - \hat{a}_n^{IV}\right)' u_t \left(a + \hat{a}_n^{IV}\right)' u_t Z_t \right\} \int_0^1 G_a(r)\, dr = O_p\left(\sqrt{n}\left(a - \hat{a}_n^{IV}\right)\right). \tag{A.20}$$

Finally, by (14),

$$\sum_{t=2}^{n} Z_t \varepsilon_t = n \mathbb{E}\left(Z_t \varepsilon_t\right) + \sqrt{n} B_{Z\varepsilon}(1) + o_p(n), \tag{A.21}$$

https://doi.org/10.1017/S0266466617000330 Published online by Cambridge University Press

where, temporarily, we have not imposed Assumption 3 requiring $\mathbb{E}(Z_t\varepsilon_t)=0$, in order to examine its role in Remark 5. Collecting the dominant terms in (A.18)–(A.21), we need a solution to the equation

$$\left(n\Sigma_{Zu}\int_0^1 G_a(r)\,dr+O_p\left(\sqrt{n}\right)\right)\left(\hat{a}_n^{IV}-a\right)=\sqrt{n}B_{Z\varepsilon}(1)+n\mathbb{E}(Z\varepsilon). \tag{A.22}$$

The desired result follows immediately upon imposition of Assumption 3. ■

Proof of Theorem 4. Using (24), we have

$$\begin{aligned}
e_t^{IV} &= Y_t-\beta_t\left(\hat{a}_n^{IV}\right)Y_{t-1}\\
&= Y_t-\beta_t(a)\,\beta_t\left(\hat{a}_n^{IV}-a\right)Y_{t-1}\\
&= Y_t-\beta_t(a)\left(1+\frac{\left(\hat{a}_n^{IV}-a\right)'u_t}{\sqrt{n}}+o_p\left(\sqrt{n}\right)\right)Y_{t-1}\\
&= \varepsilon_t-\beta_t(a)\left(\frac{\left(\hat{a}_n^{IV}-a\right)'u_t}{\sqrt{n}}+o_p\left(\sqrt{n}\right)\right)Y_{t-1}.
\end{aligned} \tag{A.23}$$

As $\left(\hat{a}_n^{IV}-a\right)=O_p\left(n^{-1/2}\right)$,

$$\begin{aligned}
\beta_t(a)\frac{\left(\hat{a}_n^{IV}-a\right)'u_t}{\sqrt{n}}Y_{t-1} &= \beta_t(a)\left(\hat{a}_n^{IV}-a\right)'u_t\left(G_a(r)+o_p(1)\right)\\
&= O_p\left(n^{-1/2}\right).
\end{aligned}$$

The next order term in expansion (A.23) is

$$\begin{aligned}
\beta_t(a)\frac{\left(\left(\hat{a}_n^{IV}-a\right)'u_t\right)^2}{2n}Y_{t-1} &= \beta_t(a)\frac{\left(\left(\hat{a}_n^{IV}-a\right)'u_t\right)^2}{2\sqrt{n}}\left(G_a(r)+o_p(1)\right)\\
&= O_p\left(n^{-3/2}\right).
\end{aligned}$$

Therefore,

$$\hat{\gamma}_{\varepsilon,n}^{IV}(j)=\frac{1}{n}\sum_{t=j+1}^{n}\left(\varepsilon_t+O_p\left(n^{-1/2}\right)\right)\left(\varepsilon_{t-j}+O_p\left(n^{-1/2}\right)\right)=\gamma_\varepsilon(j)+O_p\left(n^{-1/2}\right),$$

and part (1) of the Theorem is established. The proofs for parts (2) and (3) are similar and are therefore omitted. ■

Proof of Theorem 5. The solution to (28) must satisfy

$$\left(\frac{\partial g_n'\left(\hat{a}_n^G\right)}{\partial a}\right)\hat{W}g_n\left(\hat{a}_n^G\right)=0.$$

Now,

$$\frac{\partial g_n'\left(\hat{a}_n^G\right)}{\partial a}=-\frac{1}{n^{3/2}}\sum_{t=1}^{n}\beta_t\left(\hat{a}_n^G\right)Y_{t-1}u_tZ_t',$$

https://doi.org/10.1017/S0266466617000330 Published online by Cambridge University Press

so, we need to solve

$$\left(\sum_{t=1}^{n}\beta_t\left(\hat{a}_n^G\right)Y_{t-1}u_tZ_t'\right)\hat{W}\left(\sum_{t=1}^{n}\left(Y_t-\beta_t\left(\hat{a}_n^G\right)Y_{t-1}\right)Z_t\right)$$
$$=\left(\sum_{t=1}^{n}\beta_t\left(\hat{a}_n^G\right)Y_{t-1}u_tZ_t'\right)\hat{W}\left(\sum_{t=1}^{n}\left(\left(\beta_t(a)-\beta_t\left(\hat{a}_n^G\right)\right)Y_{t-1}+\varepsilon_t\right)Z_t\right)$$
$$=0. \tag{A.24}$$

We shall need the following results. First,

$$\sum_{t=1}^{n}\beta_t\left(\hat{a}_n^G\right)Y_{t-1}u_tZ_t'=\sum_{t=1}^{n}\left(1+\frac{\hat{a}_n^{G\prime}u_t}{\sqrt{n}}+o_p\left(n^{-1/2}\right)\right)u_tZ_t'Y_{t-1}$$
$$=n^{3/2}\Sigma_{Zu}'\left(\int_0^1 G_a(r)\,dr+o_p(1)\right)$$
$$+nE\left(\hat{a}_n^{G\prime}u_tu_tZ_t'\right)\left(\int_0^1 G_a(r)\,dr+o_p(1)\right)+O_p(n).$$

By (A.21),

$$\sum_{t=1}^{n}\varepsilon_tZ_t=\sqrt{n}B_{Z\varepsilon}(1)+o_p\left(\sqrt{n}\right). \tag{A.25}$$

Finally,

$$\sum_{t=1}^{n}\left(\beta_t(a)Y_{t-1}-\beta_t\left(\hat{a}_n^G\right)Y_{t-1}\right)Z_t=\sum_{t=1}^{n}\left(\frac{\left(a-\hat{a}_n^G\right)'u_t}{\sqrt{n}}+o_p\left(n^{-1/2}\right)\right)Y_{t-1}Z_t$$
$$=n\Sigma_{Zu}\left(a-\hat{a}_n^G\right)\int_0^1 G_a(r)\,dr+o_p\left(n\left(a-\hat{a}_n^G\right)\right). \tag{A.26}$$

The result of the theorem follows upon substitution of (A.25) and (A.26) into (A.24). ■

Proof of Theorem 6. We know from (A.21) that

$$\sqrt{n}\bar{g}_n\equiv\frac{1}{\sqrt{n}}\sum_{t=1}^{n}Z_t\varepsilon_t\Rightarrow B_{Z\varepsilon}(1)\sim N\left(0,\gamma_{Z\varepsilon}^{\ell r}\right).$$

It follows that

$$\hat{g}_n=g_n\left(\hat{a}_n^G\left(\left(\hat{\gamma}_{Z\varepsilon,n}^{\ell r}\right)^{-1}\right)\right)=\frac{1}{n}\sum_{t=1}^{n}Z_t\left(\beta_tY_{t-1}+\varepsilon_t-\beta_t\left(\hat{a}_n^G\left(\left(\hat{\gamma}_{Z\varepsilon,n}^{\ell r}\right)^{-1}\right)\right)Y_{t-1}\right)$$
$$=\frac{1}{n}\sum_{t=1}^{n}Z_t\varepsilon_t+\frac{1}{n}\sum_{t=1}^{n}Z_t\left(\beta_t(a)Y_{t-1}-\beta_t\left(\hat{a}_n^G\left(\left(\hat{\gamma}_{Z\varepsilon,n}^{\ell r}\right)^{-1}\right)\right)Y_{t-1}\right).$$

https://doi.org/10.1017/S0266466617000330 Published online by Cambridge University Press

Therefore,

$$\sqrt{n}\hat{g}_n = B_{Z\varepsilon}(1) + \frac{1}{\sqrt{n}}\sum_{t=1}^{n} Z_t \left(\frac{\left(a - \hat{a}_n^G\left(\left(\hat{\gamma}_{Z\varepsilon,n}^{\ell r}\right)^{-1}\right)\right)' u_t}{\sqrt{n}} + o_p\left(n^{-1/2}\right) \right) Y_{t-1} + o_p(1).$$

The second term above is equal to

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\left(n\Sigma_{Zu}\left(a - \hat{a}_n^G\left(\left(\hat{\gamma}_{Z\varepsilon,n}^{\ell r}\right)^{-1}\right)\right)\int_0^1 G_a(r)\,dr + o_p\left(n\left(a - \hat{a}_n^G\left(\left(\hat{\gamma}_{Z\varepsilon,n}^{\ell r}\right)^{-1}\right)\right)\right)\right) \\
&= \sqrt{n}\Sigma_{Zu}\left(a - \hat{a}_n^G\left(\left(\hat{\gamma}_{Z\varepsilon,n}^{\ell r}\right)^{-1}\right)\right)\int_0^1 G_a(r)\,dr + o_p(1) \\
&= -\Sigma_{Zu}\left(\int_0^1 G_a(r)\,dr\right)^{-1}\left(\Sigma_{Zu}'\left(\hat{\gamma}_{Z\varepsilon,n}^{\ell r}\right)^{-1}\Sigma_{Zu}\right)^{-1}\Sigma_{Zu}'\left(\hat{\gamma}_{Z\varepsilon,n}^{\ell r}\right)^{-1} \\
&\quad \times B_{Z\varepsilon}(1)\left(\int_0^1 G_a(r)\,dr\right) + o_p(1) \\
&= \left[I_q - \Sigma_{Zu}\left(\Sigma_{Zu}'\left(\hat{\gamma}_{Z\varepsilon,n}^{\ell r}\right)^{-1}\Sigma_{Zu}\right)^{-1}\Sigma_{Zu}'\left(\hat{\gamma}_{Z\varepsilon,n}^{\ell r}\right)^{-1}\right] B_{Z\varepsilon}(1) + o_p(1). \\
&= \hat{B}_n B_{Z\varepsilon}(1) + o_p(1),
\end{aligned}$$

say. The test for over-identifying restrictions is given by

$$J_n\left(\hat{a}_n^G\left(\left(\hat{\gamma}_{Z\varepsilon,n}^{\ell r}\right)^{-1}\right), \hat{\gamma}_{Z\varepsilon,n}^{\ell r}\right) = n\hat{g}_n'\left(\hat{\gamma}_{Z\varepsilon,n}^{\ell r}\right)^{-1}\hat{g}_n = n\bar{g}'\hat{B}_n\left(\hat{\gamma}_{Z\varepsilon,n}^{\ell r}\right)^{-1}\hat{B}_n\bar{g} + o_p(1).$$

The middle term in the last line is equal to

$$\begin{aligned}
&\hat{B}_n\left(\hat{\gamma}_{Z\varepsilon,n}^{\ell r}\right)^{-1}\hat{B}_n \\
&\quad = \left[I_q - \left(\hat{\gamma}_{Z\varepsilon,n}^{\ell r}\right)^{-1}\Sigma_{Zu}\left(\Sigma_{Zu}'\left(\hat{\gamma}_{Z\varepsilon,n}^{\ell r}\right)^{-1}\Sigma_{Zu}\right)^{-1}\Sigma_{Zu}'\right]\left(\hat{\gamma}_{Z\varepsilon,n}^{\ell r}\right)^{-1} \\
&\qquad \times \left[I_q - \Sigma_{Zu}\left(\Sigma_{Zu}'\left(\hat{\gamma}_{Z\varepsilon,n}^{\ell r}\right)^{-1}\Sigma_{Zu}\right)^{-1}\Sigma_{Zu}'\left(\hat{\gamma}_{Z\varepsilon,n}^{\ell r}\right)^{-1}\right] \\
&\quad = \left(\hat{\gamma}_{Z\varepsilon,n}^{\ell r}\right)^{-1} - \left(\hat{\gamma}_{Z\varepsilon,n}^{\ell r}\right)^{-1}\Sigma_{Zu}\left(\Sigma_{Zu}'\left(\hat{\gamma}_{Z\varepsilon,n}^{\ell r}\right)^{-1}\Sigma_{Zu}\right)^{-1}\Sigma_{Zu}'\left(\hat{\gamma}_{Z\varepsilon,n}^{\ell r}\right)^{-1} \\
&\quad = \left(\hat{\gamma}_{Z\varepsilon,n}^{\ell r}\right)^{-1/2}\left(I_q - \left(\hat{\gamma}_{Z\varepsilon,n}^{\ell r}\right)^{-1/2\prime}\Sigma_{Zu}\left(\Sigma_{Zu}'\left(\hat{\gamma}_{Z\varepsilon,n}^{\ell r}\right)^{-1}\Sigma_{Zu}\right)^{-1}\Sigma_{Zu}'\left(\hat{\gamma}_{Z\varepsilon,n}^{\ell r}\right)^{-1/2}\right)\left(\hat{\gamma}_{Z\varepsilon,n}^{\ell r}\right)^{-1/2\prime} \\
&\quad = \left(\hat{\gamma}_{Z\varepsilon,n}^{\ell r}\right)^{-1/2}\hat{M}_n\left(\hat{\gamma}_{Z\varepsilon,n}^{\ell r}\right)^{-1/2\prime},
\end{aligned}$$

say. Thus,

$$J_n\left(\hat{a}_n^G\left(\left(\hat{\gamma}_{Z\varepsilon,n}^{\ell r}\right)^{-1}\right), \left(\hat{\gamma}_{Z\varepsilon,n}^{\ell r}\right)^{-1}\right) = n\bar{g}'\left(\hat{\gamma}_{Z\varepsilon,n}^{\ell r}\right)^{-1/2}\hat{M}_n\left(\hat{\gamma}_{Z\varepsilon,n}^{\ell r}\right)^{-1/2\prime}\bar{g} + o_p(1).$$

The result of the theorem follows from the facts that $\left(\hat{\gamma}_{Z\varepsilon,n}^{\ell r}\right)^{-1/2\prime}\bar{g} \Rightarrow N\left(0, I_q\right)$ and $\hat{M}_n$ is symmetric and idempotent with rank $q - K$. ■

https://doi.org/10.1017/S0266466617000330 Published online by Cambridge University Press