IV AND GMM ESTIMATION AND TESTING OF MULTIVARIATE STOCHASTIC UNIT ROOT MODELS

By

Offer Lieberman and Peter C. B. Phillips

June 2016

COWLES FOUNDATION DISCUSSION PAPER NO. 2061

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

IV and GMM Estimation and Testing of Multivariate Stochastic Unit Root Models*

Offer Lieberman† and Peter C. B. Phillips‡

June 10, 2016

Abstract

Lieberman and Phillips (2015; 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

*We thank Tim Ginker for his research assistance.

†Bar-Ilan University. Support from Israel Science Foundation grant No. 1082-14 and from the Sapir Center in Tel Aviv University are gratefully acknowledged. Correspondence to: Department of Economics and Research Institute for Econometrics (RIE), Bar-Ilan University, Ramat Gan 52900, Israel. E-mail: offer.lieberman@gmail.com

‡Yale University, University of Auckland, Southampton University, and Singapore Management University. Support is acknowledged from the NSF under Grant No. SES 1258258.

the asymptotic distributions of the 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 a standard UR coefficient test 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.
Key words and phrases: Autoregression; Diffusion; Similarity; Stochastic unit root; Time-varying coefficients.

JEL Classification: C22

1 Introduction

The model under consideration is the stochastic unit root (STUR) system

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

where the STUR coefficient

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

depends nonlinearly on an $K \times 1$ 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 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 $a = c$ (Phillips, 1987; Chan and

Wei, 1987). In the STUR model (1)-(2), the coefficient $\beta_t(a;n)$ is similarly localized to unity 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 application of STUR models in finance are given in Lieberman and Phillips (2016, 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 (2016) 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 (rhs) of (1) contains a drift, LP (2016) 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 (2016) 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. 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 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 is developed in Section 6. Simulation experiments evaluating the adequacy of the limit theory is reported in Section 7. Section 8 concludes and proofs are given in the Appendix.

2 Preliminaries and Results on the NLLS

For the generating mechanism of the process $w_t = (u_t', \varepsilon_t)'$ 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*

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

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

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

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

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 $K \times 1$ and η_{2t} is scalar. Specifically,

$$\begin{aligned} u_t &= G_{1,0}\eta_t + G_{1,1}\eta_{t-1} + \cdots \\ \varepsilon_t &= G_{2,0}\eta_t + G_{2,1}\eta_{t-1} + \cdots, \end{aligned}$$

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

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

In the special case where w_t is an mds, $\Sigma = \Sigma^{\ell r}$. For that case, Lieberman and Phillips (2014, 2016) 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 (2016) to the present case of stationary driver variables and equation errors satisfying Assumption 1.

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

$$n^{-1/2} Y_{t=\lfloor nr \rfloor} \Rightarrow e^{a' B_u(r)} \left(\int_0^r e^{-a' B_u(p)} dB_\varepsilon(p) - a' \Delta_{u\varepsilon} \int_0^r e^{-a' B_u(p)} dp \right) := G_a(r). \tag{6}$$

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

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

limit process G_a whenever $a \neq 0$ and $\Delta_{u\varepsilon} \neq 0$. If $a = 0$, the limit process is standard Brownian motion B_ε as expected.

Denote by $\hat{a}_n$ the NLLS of a which minimizes the criterion $Q_n(a) = \sum_{t=2}^{n} \{Y_t - \beta_t(a;n) Y_{t-1}\}^2$. The following theorem was established in LP (2016) 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)

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

(2)

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

(3)

$$\sqrt{n}(\hat{a}_n - a) \Rightarrow \frac{1}{\int_0^1 G_a^2(r)\,dr} \Sigma_u^{-1} \left(\int_0^1 G_a(r)\,dB_{u\varepsilon}(r) + \{\mathbb{E}(\varepsilon_t u_t u_t')\}\, a \int_0^1 G_a(r)\,dr \right), \quad \text{if } \Sigma_{u\varepsilon} = 0. \tag{7}$$

(4)

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

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 $\Sigma_\eta = \mathbb{E}(\eta_t \eta_t')$, and in view of (6), on the long run covariances indirectly, through $G_a(r)$.*

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

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

because all odd moments of a symmetric distribution around zero are equal to zero. In this case, eq'n (7) reduces to

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

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 the model (1)-(4) in which it is assumed that the drift parameter is equal to zero. When the drift parameter is non-zero, Lieberman and Phillips (2016) 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 in the important case in which the model's localized STUR coefficient equals to zero. 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 eq'ns (1)-(2).

3 Instrumental Variable Estimation of the STUR Model

Let Z_t be an $q \times 1$ 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

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

Assumption 2: *The vector w_t^* satisfies*

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

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,

$$\begin{aligned} \mathbb{E}\left(Z_t \varepsilon_t\right) &= \sum_{j=0}^{\infty} G_{3j} \Sigma_{\eta^*} G_{2j}' = 0, & (8) \\ \mathbb{E}\left(Z_t u_t'\right) &= \Sigma_{Zu} = \sum_{j=0}^{\infty} G_{3j} \Sigma_{\eta^*} G_{1j}' \text{ has full rank } K. & (9) \end{aligned}$$

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:

$$\sum_{t=2}^{n} \left(Y_t - \beta_t\left(\hat{a}_n^{IV}; n\right) Y_{t-1}\right) Z_t = 0. \tag{10}$$

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

$$n^{-1/2} \sum_{t=1}^{\lfloor nr \rfloor} Z_t \varepsilon_t \Rightarrow B_{Z\varepsilon}\left(r\right), \tag{11}$$

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

$$\gamma_{Z\varepsilon}^{\ell r} = \sum_{h=-\infty}^{\infty} \mathbb{E}\left(Z_t Z_{t+h}' \varepsilon_t \varepsilon_{t+h}\right) = \sum_{h=-\infty}^{\infty} \mathbb{E}\left(Z_t Z_{t+h}'\right) \mathbb{E}(\varepsilon_t \varepsilon_{t+h}) \tag{12}$$

if Z_t is uncorrelated with ε_s for all s, t.

The asymptotic distribution of the IV estimator is given in the following theorem.

Theorem 3 *For the model (1)-(4), under Assumptions 2-3, for* $q = K$,

$$\sqrt{n}\left(\hat{a}_n^{IV} - a\right) \Rightarrow \frac{\Sigma_{Zu}^{-1} B_{Z\varepsilon}\left(1\right)}{\int_0^1 G_a\left(r\right) dr}. \tag{13}$$

It is emphasized that the matrix Σ_{Zu} appearing on the rhs of (13) is the contemporaneous covariance between Z and u. Several remarks are in place.

Remark 3 *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 4 *The role of the usual IV orthogonality condition (8) in Assumption 3 is evident in eq'n (45) 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 (8) renders $\hat{a}_n^{IV}$ *inconsistent as expected.*

Remark 5 *The limit distribution in (13) is not defined if the relevance condition (9) 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 6 *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,

$$\sqrt{n}\hat{a}_n^{IV} \Rightarrow \frac{N\left(0, \gamma_{Z\varepsilon}^{lr}\right)}{\sigma_{Zu} N\left(0, \left(\sigma_\varepsilon^{lr}\right)^2 / 3\right)} \tag{14}$$

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

$$\sqrt{n}\hat{a}_n^{IV} \Rightarrow \frac{N\left(0, \sigma_\varepsilon^2 \sigma_Z^2\right)}{\sigma_{Zu} N\left(0, \sigma_\varepsilon^2 / 3\right)}, \tag{15}$$

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

The limit distributions (14) and (15) 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 manifest 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 (14) as

$$\sqrt{n}\hat{a}_n^{IV} \Rightarrow \frac{N\left(0,\sigma_\varepsilon^2\sigma_Z^2\right)}{\sigma_{Zu}N\left(0,\sigma_\varepsilon^2/3\right)} = \frac{\xi_1}{\sigma_{Zu}\xi_2},.$$

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 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 $\sqrt{k_n}\left(\hat{B}-B\right)\Rightarrow N\left(0,V_B\right)$, then we will have

$$\begin{aligned}\sqrt{n}\hat{a}_n^{IV}-\hat{B} &= \sqrt{n}\hat{a}_n^{IV}-B+\left(\hat{B}-B\right)=\sqrt{n}\hat{a}_n^{IV}-B+O_p\left(\frac{1}{\sqrt{k_n}}\right)\\ &\Rightarrow \frac{\xi_{1.2}}{\sigma_{Zu}\xi_2}=\frac{1}{\sigma_{Zu}}\left(\frac{\sigma_{11.2}^\xi}{\sigma_{22}^\xi}\right)^{1/2}\mathbb{C}.\end{aligned}$$

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

Now, under Assumption 2,

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

If Z_t is independent of ε_t, as assumed in the second part of Remark 6, then $\sigma_{12}^{\xi} = 0$ and so, $B = 0$. Otherwise,

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

This term will be zero if η_t has a symmetric distribution. In these cases then, (14) simplifies to

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

and (15) reduces to

$$\sqrt{n}\hat{a}_n^{IV} \Rightarrow \frac{\sqrt{3}\sigma_Z}{\sigma_{Zu}}\mathbb{C}, \tag{16}$$

respectively. For inference then, the scaled Cauchy distribution should 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 then, t- and Wald tests might be constructed by simulating the standard deviation of the rhs of (13) and extacting 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 (13).

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

$$n\left(\hat{\beta}-1\right) \Rightarrow \begin{cases} \frac{\int_0^1 B_\varepsilon dB_\varepsilon + \lambda_{\varepsilon\varepsilon}}{\int_0^1 B_\varepsilon^2} & \text{under } H_0 : a = 0 \\ \frac{\int_0^1 G_a dB_\varepsilon + \lambda_{\varepsilon\varepsilon} + a' \int_0^1 dB_u G_a^2 + 2a'\Delta_{u\varepsilon}\int_0^1 G_a(r)dr}{\int_0^1 G_a^2} & \text{under } H_1 : a \neq 0, \end{cases} . \tag{17}$$

In particular, under the null the usual UR limit theory (Phillips, 1987) applies and under the alternative we have

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

which leads to the following limit behavior

$$\begin{aligned} n\left(\hat{\beta}-1\right) &= \frac{n^{-1}\sum_{t=1}^n Y_{t-1}\Delta Y_t}{n^{-2}\sum_{t=1}^n Y_{t-1}^2} \sim \frac{n^{-1}\left(\left(\sum_{t=1}^n Y_{t-1}\varepsilon_t\right) + \left(a'\sum_{t=1}^n \frac{u_t}{\sqrt{n}} Y_{t-1}^2\right)\right)}{n^{-2}\sum_{t=1}^n Y_{t-1}^2} \\ &\Rightarrow \frac{\left[\int_0^1 G_a dB_\varepsilon + \lambda_{\varepsilon\varepsilon}\right] + \left[a'\int_0^1 dB_u G_a^2 + 2a'\Delta_{u\varepsilon}\int_0^1 G_a(r)\,dr\right]}{\int_0^1 G_a^2}, \end{aligned} \tag{18}$$

where we use results from Ibragimov and Phillips (2008) for the sample covariance $n^{-3/2}\sum_{t=1}^n u_t Y_{t-1}^2$. The limit (18) 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 least squares-based estimators of σ_ε^2, Σ_u and $\Sigma_{u\varepsilon}$ are given by

$$\hat{\sigma}_{\varepsilon,n}^2 = \frac{1}{n}\sum_{t=2}^n \left(Y_t - e^{\hat{a}_n' u_t/\sqrt{n}} Y_{t-1}\right)^2, \; vech\left(\hat{\Sigma}_{u,n}\right) = \frac{1}{n}\sum_{t=1}^n vech\left(u_t u_t'\right), \tag{19}$$

and

$$\hat{\Sigma}_{u\varepsilon,n} = \frac{1}{n}\sum_{t=2}^{n}\left(Y_t - e^{\hat{a}_n' u_t/\sqrt{n}}Y_{t-1}\right)u_t,$$

respectively. Their limit theory was given in Theorem 3 of Lieberman and Phillips (2016) in the case where w_t is a strictly stationary and ergodic mds process and is presented here again for convenience.

Theorem 4 *For the model (1)–(4), if w_t is a strictly stationary and ergodic mds process,*

(1)

$$\hat{\sigma}_{\varepsilon,n}^2 - \sigma_{\varepsilon}^2 \Rightarrow -\frac{\left(\int_0^1 G_a(r)\,dr\right)^2}{\int_0^1 G_a^2(r)\,dr}\Sigma_{u\varepsilon}'\Sigma_u^{-1}\Sigma_{u\varepsilon}.$$

(2) If u_t has finite fourth moments, centred partial sums of $u_t u_t'$ satisfy the invariance principle

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{\lfloor nr\rfloor} vech\left(u_t u_t' - \Sigma_u\right) \Rightarrow \xi(r),$$

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

$$\Sigma_{u\otimes u} = \mathbb{E}\left(L_e\left(u_t \otimes u_t - \mathbb{E}\left(u_t \otimes u_t\right)\right)\left(u_t' \otimes u_t' - \mathbb{E}\left(u_t' \otimes u_t'\right)\right)L_e'\right)$$

and L_e is the elimination matrix satisfying $vech(uu') = L_e\,(u \otimes u)$.

(3)

$$\hat{\Sigma}_{u\varepsilon,n} - \Sigma_{u\varepsilon} \Rightarrow -\frac{\left(\int_0^1 G_a(r)\,dr\right)^2}{\int_0^1 G_a^2(r)\,dr}\Sigma_{u\varepsilon}.$$

The implication of Theorem 4 is that $\hat{\sigma}_{\varepsilon,n}^2$ and $\hat{\Sigma}_{u\varepsilon,n}$ are both inconsistent. Let

$$e_t^{IV} = Y_t - e^{\hat{a}_n^{IV\prime} u_t/\sqrt{n}}Y_{t-1},\ t = 2, ..., n. \tag{20}$$

We show in the following theorem that for $j = 0, 1, 2, ...$, the IV-based estimators

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

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

and

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

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

Theorem 5 *Under Assumptions 2-3, for $q = K$,*

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

We remark that $\hat{\Sigma}_{u,n}$, defined in (19), does not depend on a and is $\sqrt{n}$-consistent.

5 A Test Statistic with Nuisance Parameters Estimated

The limit distribution in (13) depends on the unknown parameter Σ_{Zu}. It is obvious from eq'n (45) of the Appendix that under Assumption 1,

$$\sqrt{n}\left(\int_0^1 G_a(r)\,dr\right)\Sigma_{Zu}\left(\hat{a}^{IV}_n - a\right) = B_{Z\varepsilon}(1) + O_p\left(n^{-1/2}\right). \qquad (21)$$

The lhs of (21) is equal to

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

Therefore, to first order, we may replace the rhs of (13) and (15) by

$$\frac{\left(\hat{\Sigma}_{Zu,n}\right)^{-1}B_{Z\varepsilon}(1)}{\int_0^1 G_a(r)\,dr} \tag{22}$$

and

$$\frac{B_{Z\varepsilon}(1)}{\hat{\sigma}_{Zu,n}\int_0^1 B_\varepsilon(r)\,dr}, \tag{23}$$

respectively. The long run covariances associated with the distributions in (22) and (23) can be consistently estimated using the results of Theorem 5 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 overidentification test. Let

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

$$\hat{a}_n^G = \hat{a}_n^G \left(\hat{W}\right) = \arg\min_a J_n \left(a, \hat{W}\right), \tag{24}$$

where

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

$\hat{W}$ 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 (9) but is $(q \times K)$, with the possibility that $q \geq K$. The limit theory for $\hat{a}_n^G$ is as follows.

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

$$\sqrt{n}\left(\hat{a}_n^G - a\right) \Rightarrow \left(\Sigma_{Zu}' W \Sigma_{Zu}\right)^{-1} \Sigma_{Zu}' W \frac{B_{Z\varepsilon}(1)}{\left(\int_0^1 G_a(r)\, dr\right)}. \tag{25}$$

Remark 7 *If the model is just identified, the result of the Theorem collapses to (13).*

Remark 8 *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

$$\hat{\delta}_n^G = \left(S_{ZX}' \hat{W} S_{ZX}\right)^{-1} S_{ZX}' \hat{W} s_{Xy}, \tag{26}$$

where

$$S_{ZX} = \frac{1}{n}\sum_{t=1}^{n} Z_t x_t' \text{ and } s_{Xy} = \frac{1}{n}\sum_{t=1}^{n} x_t Y_t.$$

The correspondence between (25) and the usual linear formulation (26) 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 (12). The limit theory for this statistic has the usual χ^2_{q-K} form, as given in (27) below.

Theorem 7 *For the model (1)-(4), under Assumptions 2-3 and for $q \geq K$,*

$$J_n\left(\hat{a}_n^G\left(\left(\hat{\gamma}_{Z\varepsilon,n}^{\ell r}\right)^{-1}\right), \left(\hat{\gamma}_{Z\varepsilon,n}^{\ell r}\right)^{-1}\right) = n\hat{g}_n'\left(\hat{\gamma}_{Z\varepsilon,n}^{\ell r}\right)^{-1}\hat{g}_n \Rightarrow \chi^2_{q-K}. \tag{27}$$

Remark 9 *The χ^2_{q-K} limit distribution for the overidentifying test again corresponds to that in linear model discussed in Remark 8.*

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 (23), the efficient GMM estimator $\hat{a}_n^G$, and the overidentification test. We consider the following scenarios.

Case 1:

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

Case 2:

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

Case 3:

$$u_t \overset{iid}{\sim} N\left(0,\sigma_u^2\right), \sigma_u^2 = 0.1, \varepsilon_t \overset{iid}{\sim} N\left(0,\sigma_u^2\right),$$

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

Case 4:

$$u_t \overset{iid}{\sim} N\left(0, \sigma_u^2\right), \sigma_u^2 = 0.1, \varepsilon_t \overset{iid}{\sim} N\left(0, \sigma_u^2\right) + 2\eta_t,$$

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

For each case, we simulated 5000 replications with $n = 100$, 1000, 10000 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	$= 0$	$\neq 0$	$\neq 0$

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

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

whereas in Case 2, $\Sigma_{Z\varepsilon} \neq 0$. Similarly, Cases 3-4 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 was simulated according to (1) and (2) and the IV estimator was solved in each replication as a solution to the nonlinear moment condition (10). 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 lhs of (13) against its limit distribution (coded as rhs) and against the estimated normal distribution of the lhs were formed. A selection of the plots is presented. The typical situation is given in Figures 1-2. As expected, for cases 1,3,5, in which Assumption 3 holds, the lhs and rhs 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 lhs and rhs become very different as n grows in cases

2,4,6, 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 1,3,5, in which Assumption 3 holds, the asymptotic distribution of the test statistic has a peaked distribution compared with the normal distribution. This is evident in the PP plots and Figure 4, which provides kernel density estimates of the lhs, rhs and estimated normal density of the lhs. Finally, for the case $a = 0$, the comparisons drawn in Figures 5-7 between the distribution of $\sqrt{n}\hat{a}_n^{IV}$ and the scaled Cauchy variate, given in (16), show a perfect fit.

In the second part of the analysis we investigated the rejection rates (RR) of the hypothesis test $H_0 : a = 0$ using (23). The results are of interest for applications with small to moderate n and/or a. Each of the experiments were based on 2000 replications. We considered the process

$$\eta_{1t} \overset{iid}{\sim} N\left(0, \sigma^2_{\eta_1}\right), \eta_{2t} \overset{iid}{\sim} N\left(0, \sigma^2_{\eta_2}\right), \eta_{3t} \overset{iid}{\sim} N\left(0, \sigma^2_{\eta_3}\right) \tag{28}$$

$$\begin{aligned} \sigma^2_{\eta_1} &= 0.673, \sigma^2_{\eta_2} = 0.129, \sigma^2_{\eta_3} = 0.5 \\ u_t &= \eta_{1,t} + 0.432\eta_{1,t-1} - 0.21\eta_{2,t-1} \\ \varepsilon_t &= \eta_{2,t} - 0.251\eta_{1,t-1} + 0.12\eta_{2,t-1} \\ Z_t &= \eta_{3,t} + 0.3\eta_{1t} + 0.4\eta_{3,t-1}. \end{aligned} \tag{29}$$

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 $n = 2000$ and varied the true a over the values 0, 0.2, 0.5, 1, 2, 5. The results are given in Table 2 and Figure 8. For the second part of the analysis, we fixed a at 0.2 and 1 and varied n over the values 100, 500, 1000, 1500, 2000, 5000. The results for this part of the analysis are given in Tables 3-4. Clearly, as a increases, both the one-sided and two-sided RRs increase from about 0.5 to about 0.99 and the sample mean of $\hat{a}_n^{IV}$ is accurate. In Table 3-4 we also observe an increase in the RRs and a decrease in the standard deviations of the estimator, as n increases, as expected.

For the same setting we simulated RRs 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 generated under the null hypothesis and under all the true parameters of the process as given by (28)-(29). Accordingly, the simulation reports performance of an 'ideal' DF test because the parameters $\lambda_{\varepsilon\varepsilon}$ and σ^2_ε that are needed for the simulation of the rhs of (17) 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 applied in a single direction. More specifically, Table 2 reveals that one-sided DF test power increases very slowly (and much slower than the coefficient based test) as a increases. Tables 3-4 show that this relative performance does not change as the sample size increases and that the power function of the DF test is essentially flat and close to 0.5 over the full range $n \in \{100, 500, 1000, 1500, 2000, 5000\}$ when $a = 0.2$. In contrast, this paper's coefficient test has power that increases from 0.75 to 0.95 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 this test. We expect the one-sided DF test against a STUR model 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 aid the significance of the left sided UR test. This transpires in the simulations in the reduction in power to approximately 50%. Figures[3] 9-10 emphasize this point in showing that the kernel density estimates of the difference between $n\left(\hat{\beta}-1\right)$ and the limit null distribution in the $a = 0.2$ and 1 cases are centered at zero. In summary, our one-sided test has good power which increases with both a and n - all better than the usual DF test.

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 5000 samples of $n = 100$, 500, according to the following law:

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

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

[3] In the construction of Figures 9-10 we have used the same generating mechanism error inputs with $a = 0$ (for the null) and $a = 0.2, 1$ (for the alternatives), so that the simulation kernel density estimates are essentially equivalent to those of $\{n\left(\hat{\beta}-1\right)$ under the alternative$\}$ $-$ $\{n\left(\hat{\beta}-1\right)$ under the null$\}$.

$$Z_{1t} = \left(g_{32,0}\eta_{2t} + g_{32,1}\eta_{2,t-1}\right) + \left(g_{33,0}\eta_{3t} + g_{33,1}\eta_{3,t-1}\right) + \left(g_{34,0}\eta_{4t} + g_{34,1}\eta_{4,t-1}\right)$$
$$Z_{2t} = \left(g_{42,0}\eta_{2t} + g_{42,1}\eta_{2,t-1}\right) + \left(g_{43,0}\eta_{3t} + g_{43,1}\eta_{3,t-1}\right) + \left(g_{44,0}\eta_{4t} + g_{44,1}\eta_{4,t-1}\right).$$

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 $K = 1$. Figures 11-12 show that the pp-plots for the distribution of $\hat{a}_n^G$ against the rhs of (25) is very accurate for as little as $n = 100$ observations, in both the $a = 0$ and $a \neq 0$ cases. Figures 13-16 reveal 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 Conclusion

In the context of the multivariate driftless STUR model with endogenous driver variables that appear in the STUR AR coefficient, nonlinear least squares estimation of the localizing STUR coefficient is known to be inconsistent. This paper explores the structural STUR model and extends the existing limit theory for both the output process and for the NLLS estimator of the localizing STUR coefficient in the general 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 a structural version of the STUR model under orthogonality and relevance conditions that mirror those used in other implementations of IV. The limit distribution of the NLIV 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}, as in the linear case.

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 well be endogenous and correlated with the errors. The

IV and GMM procedures given here enable inference about structural STUR models and a mechanism for testing the local STUR model against a simple UR null. The STUR tests appear to have promising power performance characteristics against a standard UR coefficient test both as the sample size rises and departures from the null increase.

References

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

Ibragimov, R. and P. C. B. Phillips (2008). "Regression Asymptotics Using Martingale Convergence Methods," *Econometric Theory*, 24, 888–947.

Liang, H., P. C. B. Phillips, H. Wang, and Q. Wang (2016). "Weak Convergence to Stochastic Integrals for Econometric Applications," *Econometric Theory* (forthcoming).

Lieberman, O. and 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. and P. C. B. Phillips (2016). "A Multivariate Stochastic Unit Root Model with an Application to Derivative Pricing," *Journal of Econometrics* (forthcoming).

Phillips, P. C. B. (1992). "Time Series Regression with a Unit Root," *Econometrica*, 55, 277-301.

Phillips, P. C. B. and V. Solo (1992). "Asymptotics for Linear Processes," *Annals of Statistics,* 20, 971–1001.

Appendix

Proof of Lemma 1. In view of the functional law (4), in an appropriately expanded probability space we may write, for $t = \lfloor nr \rfloor$ and any $r > 0$,

$$n^{-1/2} \sum_{j=1}^{t} \eta_j = B\left(t/n\right) + o_p\left(1\right), \tag{30}$$

so that

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

$$\begin{aligned}
&= n^{-1/2} e^{a' B_u(t/n)} \\
&\times \sum_{s=1}^{t-1} e^{-\{a' B_u((s-1)/n) + o_p(1)\}} \left(1 - \frac{a' u_s}{\sqrt{n}} + O_p\left(n^{-1}\right)\right) \varepsilon_s + o_p\left(1\right) \\
&= e^{a' B_u(t/n)} \sum_{s=1}^{t-1} e^{-a' B_u((s-1)/n)} \frac{\varepsilon_s}{\sqrt{n}} - e^{a' B_u(t/n)} \sum_{s=1}^{t-1} e^{-a' B_u((s-1)/n)} \left(\frac{a' u_s \varepsilon_s}{n}\right) + o_p\left(1\right).
\end{aligned} \tag{31}$$

Setting $t = \lfloor nr \rfloor$ and noting that $\mathbb{E}\left(e^{-a' B_u(p)}\right)^2 < \infty$, the first term on the rhs of (31) has the following limit

$$\begin{aligned}
& e^{a' B_u\left(\frac{t}{n}\right)} \sum_{s=1}^{t-1} e^{-a' B_u\left(\frac{s-1}{n}\right)} dB_\varepsilon\left(\frac{s}{n}\right) \\
& \to_p e^{a' B_u(r)} \left\{\int_0^r e^{-a' B_u(p)} dB_\varepsilon\left(p\right) - \Lambda'_{u\varepsilon} a \int_0^r e^{-a' B_u(p)} dp\right\} =: G_a^*\left(r\right).
\end{aligned} \tag{32}$$

The limit (32) 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, theorem 3.1; Liang et al, 2016, theorems 2.3 and 3.1). Again, as in (30), we assume that the probability space has been expanded to permit the representation of (32) as a limit in probability.

The second term on the rhs of (31) is

$$
\begin{aligned}
& -e^{a'B_u(t/n)}\sum_{s=1}^{t-1}e^{-a'B_u\left(\frac{s-1}{n}\right)}\left(\frac{a'u_s\varepsilon_s}{n}\right) = -a'e^{a'B_u(t/n)}\sum_{s=1}^{t-1}e^{-a'B_u\left(\frac{s-1}{n}\right)} \\
& \times\left(\frac{u_s\varepsilon_s-\Sigma_{u\varepsilon}}{n}+\frac{\Sigma_{u\varepsilon}}{n}\right) \\
& = -a'\Sigma_{u\varepsilon}e^{a'B_u(t/n)}\frac{1}{n}\sum_{s=1}^{t-1}e^{-a'B_u((s-1)/n)}+O_p\left(n^{-1/2}\right) \\
& \to_p \; -a'\Sigma_{u\varepsilon}^{s}e^{a'B_u(r)}\int_0^r e^{-a'B_u(p)}dp.
\end{aligned} \tag{33}
$$

Hence,

$$
\begin{aligned}
n^{-1/2}Y_{\lfloor nr\rfloor} \to_p & G_a^*(r)-a'\Sigma_{u\varepsilon}e^{a'B_u(r)}\int_0^r e^{-a'B_u(p)}dp \\
& = e^{a'B_u(r)}\left(\int_0^r e^{-a'B_u(p)}dB_\varepsilon(p)-\Lambda_{u\varepsilon}'a\int_0^r e^{-a'B_u(p)}dp\right. \\
& \left.-a'\Sigma_{u\varepsilon}\int_0^r e^{-a'B_u(p)}dp\right) \\
& = e^{a'B_u(r)}\left(\int_0^r e^{-a'B_u(p)}dB_\varepsilon(p)-a'\Delta_{u\varepsilon}\int_0^r e^{-a'B_u(p)}dp\right)
\end{aligned} \tag{34}
$$

giving (6), as required. ■

The following lemma will be used in the sequel.

Lemma 8 *Under Assumption 1,*

$$
\sum_{t=2}^{n}u_t\varepsilon_tY_{t-1}=n^{3/2}\Sigma_{u\varepsilon}\int_0^1 G_a(r)\,dr+n\int G_a(r)\,dB_{u\varepsilon}(r)+o_p(n).
$$

Proof of Lemma 8: We have

$$\frac{1}{n^{3/2}}\sum_{t=2}^{n}u_t\varepsilon_tY_{t-1}=\frac{1}{n^{3/2}}\sum_{t=2}^{n}\left(\sum_{j=0}^{\infty}G_{1,j}\eta_{t-j}\right)\left(\sum_{k=0}^{\infty}G_{2,k}\eta_{t-k}\right)'Y_{t-1}.$$

The leading term is

$$\frac{1}{n^{3/2}}\sum_{t=2}^{n}\left(\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}G_{1,j}\mathbb{E}\left(\eta_{t-j}\eta_{t-k}'\right)G_{2,k}'\right)Y_{t-1}=\frac{1}{n^{3/2}}\sum_{t=2}^{n}\left(\sum_{j=0}^{\infty}G_{1,j}\Sigma_\eta G_{2,j}'\right)Y_{t-1}$$
$$\Rightarrow\Sigma_{u\varepsilon}\int_0^1G_a\left(r\right)dr.\qquad(35)$$

The second order term satisfies

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

In the mds case, this term is equal to $\int_0^1G_a\left(r\right)dB_{u\varepsilon}\left(r\right)$. For $j=k=0$,

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

For $j=0$ and $k=1$,

$$\begin{aligned}&\frac{1}{n}\sum_{t=2}^{n}\left(G_{1,0}\eta_t\eta_{t-1}'G_{2,1}'-\mathbb{E}\left(G_{1,0}\eta_t\eta_{t-1}'G_{2,1}'\right)\right)Y_{t-1}\\=\ &\frac{1}{n}\sum_{t=2}^{n}G_{1,0}\eta_t\eta_{t-1}'G_{2,1}'\left\{\left(1+\frac{a'u_{t-1}}{\sqrt{n}}+o_p\left(n^{-1/2}\right)\right)Y_{t-2}+\varepsilon_{t-1}\right\}\end{aligned}$$

$$\begin{aligned}=\ &G_{1,0}\int G_a\left(r\right)dB_{\eta\eta_{-1}'}\left(r\right)G_{2,1}'+\frac{1}{n^{3/2}}\sum_{t=2}^{n}\mathbb{E}\left\{G_{1,0}\eta_t\eta_{t-1}'G_{2,1}'a'u_{t-1}\right\}Y_{t-2}\\&+\frac{1}{n}\sum_{t=2}^{n}\mathbb{E}\left\{G_{1,0}\eta_t\eta_{t-1}'G_{2,1}'\varepsilon_{t-1}\right\}+o_p\left(1\right)\Rightarrow G_{1,0}\int G_a\left(r\right)dB_{\eta\eta_{-1}'}\left(r\right)G_{2,1}'.\end{aligned}$$

Similarly, for $j = k = 1$,

$$\begin{aligned}
& \frac{1}{n}\sum_{t=2}^{n}\left(G_{1,1}\eta_{t-1}\eta_{t-1}' G_{2,1}' - \mathbb{E}\left(G_{1,1}\eta_{t-1}\eta_{t-1}' G_{2,1}'\right)\right) Y_{t-1} \\
= \quad & \frac{1}{n}\sum_{t=2}^{n}\left(G_{1,1}\eta_{t-1}\eta_{t-1}' G_{2,1}' - G_{1,1}\Sigma_{\eta} G_{2,1}'\right)\left\{\left(1 + \frac{a' u_{t-1}}{\sqrt{n}} + o_p\left(n^{-1/2}\right)\right) Y_{t-2} + \varepsilon_{t-1}\right\} \\
\Rightarrow \quad & G_{1,1}\int G_a(r)\, dB_{\eta_{-1}\eta_{-1}'}(r)\, G_{2,1}'.
\end{aligned}$$

Higher order lags can be treated in the same fashion. Therefore,

$$\begin{aligned}
& \frac{1}{n}\sum_{t=2}^{n}\left(\sum_{j=0}^{\infty}\sum_{k=0}^{\infty} G_{1,j}\eta_{t-j}\eta_{t-j}' G_{2,k}' - \mathbb{E}\left(\sum_{j=0}^{\infty}\sum_{k=0}^{\infty} G_{1,j}\eta_{t-j}\eta_{t-k}' G_{2,k}'\right)\right) Y_{t-1} \\
& = \frac{1}{n}\sum_{t=2}^{n}\left(\sum_{j=0}^{\infty}\sum_{k=0}^{\infty} G_{1,j}\eta_{t-j}\eta_{t-j}' G_{2,k}' - \Sigma_{u\varepsilon}\right) Y_{t-1} \\
& \Rightarrow \sum_{j=0}^{\infty} G_{1,j}\int_0^1 G_a(r)\, dB_{\eta_{-j}\eta_{-k}}(r)\left(\sum_{k=0}^{\infty} G_{2,j}\right)' \\
& = \int_0^1 G_a(r)\, dB_{u\varepsilon}(r)\,. \qquad (36)
\end{aligned}$$

Eq'ns (35) and (36) give the stated result. ■

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

$$Q_n(a) = \sum_{t=2}^{n}\left\{Y_t - \beta_t(a)\, Y_{t-1}\right\}^2. \qquad (37)$$

Minimizing (37) with respect to a yields

$$\begin{aligned}\dot{Q}_n(\hat{a}_n) &= -2\sum_{t=2}^{n}\{Y_t - \beta_t(\hat{a}_n)Y_{t-1}\}\dot{\beta}_t(\hat{a}_n)Y_{t-1} = 0 \\ &\Longrightarrow \sum_{t=2}^{n}\{Y_t - \beta_t(\hat{a}_n)Y_{t-1}\}u_t\beta_t(\hat{a}_n)Y_{t-1} = 0 \\ &\Longrightarrow \sum_{t=2}^{n}Y_t u_t\beta_t(\hat{a}_n)Y_{t-1} = \sum_{t=2}^{n}u_t\beta_t^2(\hat{a}_n)Y_{t-1}^2. \end{aligned} \tag{38}$$

The third line in (38) 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

$$\begin{aligned}&\frac{1}{\sqrt{n}}\sum_{t=2}^{n}u_t u_t'(\hat{a}_n - a)Y_{t-1}^2 + \frac{1}{2n}\sum_{t=2}^{n}u_t\left((2\hat{a}_n'u_t)^2 - \left((\hat{a}_n+a)'u_t\right)^2\right)Y_{t-1}^2 \\ &= \sum_{t=2}^{n}u_t\varepsilon_tY_{t-1} + \frac{1}{\sqrt{n}}\sum_{t=2}^{n}\varepsilon_t u_t u_t'\hat{a}_nY_{t-1}.\end{aligned} \tag{39}$$

Now,

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

The second term above is $O_p(n(\hat{a}_n - a))$. In the $\Sigma_{u\varepsilon} \neq 0$, we only need to maintain terms in (39) 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 $\hat{a}_n - a = O_p\left(n^{-1/2}\right)$, 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

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

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 (39) is $\sum_{t=2}^{n} u_t \varepsilon_t Y_{t-1}$, whose behavior is given by Lemma 8. The last term in (39) 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 (39) as

$$\begin{aligned} & n^{3/2} \Sigma_u \left(\hat{a}_n - a\right) \int_0^1 G_a^2(r)\, dr \\ = \; & n^{3/2} \Sigma_{u\varepsilon} \int_0^1 G_a(r)\, dr + n \left(\int G_a(r)\, dB_{u\varepsilon}(r) + \mathbb{E}\left(\varepsilon_t u_t u_t' \hat{a}_n\right) \int_0^1 G_a(r)\, dr \right), \end{aligned}$$

giving the results stated in the theorem. ■

Proof of Theorem 3. Expanding (10), we obtain

$$\begin{aligned}
&\sum_{t=2}^{n}\left(Y_t-\beta_t\left(\hat{a}_n^{IV};n\right)Y_{t-1}\right)Z_t\\
&=\sum_{t=2}^{n}\left(\beta_t\left(a;n\right)Y_{t-1}+\varepsilon_t-\beta_t\left(\hat{a}_n^{IV};n\right)Y_{t-1}\right)Z_t\\
&=\sum_{t=2}^{n}\left\{\beta_t\left(a;n\right)-\beta_t\left(\hat{a}_n^{IV};n\right)\right\}Y_{t-1}Z_t+\sum_{t=2}^{n}Z_t\varepsilon_t\\
&=\sum_{t=2}^{n}\left\{\frac{\left(a-\hat{a}_n^{IV}\right)'u_t}{\sqrt{n}}+\frac{\left(a'u_t\right)^2-\left(\hat{a}_n^{IV\prime}u_t\right)^2}{2n}+o_p\left(\frac{1}{n}\right)\right\}Y_{t-1}Z_t+\sum_{t=2}^{n}Z_t\varepsilon_t\\
&=0. \qquad (40)
\end{aligned}$$

Under Assumption 2,

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

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

$$\Sigma_{Z\otimes u}^{\ell r}=\sum_{h=-\infty}^{\infty}\Gamma_{Z\otimes u}(h),\ \text{ where } \Gamma_{Z\otimes u}(h)=\mathbb{E}\left(Z_tZ_{t+h}'\otimes u_tu_{t+h}'\right)-\mathbb{E}\left(Z_t\otimes u_t\right)\mathbb{E}\left(Z_t'\otimes u_t'\right).$$

By an application of Lemma 8, we obtain

$$\sum_{t=1}^{n}vec\left(Z_tu_t'\right)Y_{t-1}\sim n^{3/2}vec\left(\Sigma_{Zu}\right)\int_0^1G_a\left(r\right)dr+n\int_0^1G_a\left(r\right)dB_{Zu}\left(r\right)+o_p\left(n\right). \qquad (41)$$

Hence,

$$\frac{\left(\hat{a}_n^{IV}-a\right)'}{\sqrt{n}}\sum_{t=2}^{n}u_tY_{t-1}Z_t=\frac{1}{\sqrt{n}}\left(\sum_{t=2}^{n}Y_{t-1}Z_tu_t'\right)\left(\hat{a}_n^{IV}-a\right)=O_p\left(n\left(\hat{a}_n^{IV}-a\right)\right). \qquad (42)$$

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

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

which appears in (40), is

$$\frac{\sqrt{n}}{2}\mathbb{E}\left\{\left(a-\hat{a}_n^{IV}\right)'u_t\left(a+\hat{a}_n^{IV}\right)'u_tZ_t\right\}\int_0^1 G_a(r)\,dr=O_p\left(\sqrt{n}\left(a-\hat{a}_n^{IV}\right)\right). \tag{43}$$

Finally, by (11),

$$\sum_{t=2}^{n}Z_t\varepsilon_t=n\mathbb{E}\left(Z_t\varepsilon_t\right)+\sqrt{n}B_{Z\varepsilon}(1)+o_p(n), \tag{44}$$

where, temporarily, we have not imposed Assumption 3 requiring $\mathbb{E}\left(Z_t\varepsilon_t\right)=0$, in order to examine its role in Remark 4. Collecting the dominant terms in (41)-(44), we need a solution to the equation

$$\left(n\Sigma_{Zu}\int_0^1 G_a(r)\,dr+O_p\left(\sqrt{n}\right)\right)\left(\hat{a}_n^{IV}-a\right)=\sqrt{n}B_{Z\varepsilon}(1)+n\mathbb{E}\left(Z\varepsilon\right). \tag{45}$$

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

Proof of Theorem 5. Using (20), we have

$$\begin{aligned}
e_t^{IV}&=Y_t-\beta_t\left(\hat{a}_n^{IV}\right)Y_{t-1}\\
&=Y_t-\beta_t(a)\,\beta_t\left(\hat{a}_n^{IV}-a\right)Y_{t-1}\\
&=Y_t-\beta_t(a)\left(1+\frac{\left(\hat{a}_n^{IV}-a\right)'u_t}{\sqrt{n}}+o_p\left(\sqrt{n}\right)\right)Y_{t-1}\\
&=\varepsilon_t-\beta_t(a)\left(\frac{\left(\hat{a}_n^{IV}-a\right)'u_t}{\sqrt{n}}+o_p\left(\sqrt{n}\right)\right)Y_{t-1}.
\end{aligned} \tag{46}$$

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 the expansion (46) 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,

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

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

Proof of Theorem 6. The solution to (24) 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,

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

so, we need to solve

$$\left(\sum_{t=1}^{n}\beta_t\left(\hat{a}_n^G\right)Y_{t-1}u_tZ_t'\right)\hat{W}\left(\sum_{t=1}^{n}\left(Y_t-\beta_t\left(\hat{a}_n^G\right)Y_{t-1}\right)Z_t\right) \tag{47}$$
$$=\left(\sum_{t=1}^{n}\beta_t\left(\hat{a}_n^G\right)Y_{t-1}u_tZ_t'\right)\hat{W}\left(\sum_{t=1}^{n}\left(\left(\beta_t\left(a\right)-\beta_t\left(\hat{a}_n^G\right)\right)Y_{t-1}+\varepsilon_t\right)Z_t\right)$$
$$=0.$$

We shall need the following results.

$$\begin{aligned}\sum_{t=1}^{n}\beta_t\left(\hat{a}_n^G\right)Y_{t-1}u_tZ_t' &= \sum_{t=1}^{n}\left(1+\frac{\hat{a}_n^{G\prime}u_t}{\sqrt{n}}+o_p\left(n^{-1/2}\right)\right)u_tZ_t'Y_{t-1}\\ &= n^{3/2}\Sigma_{Zu}'\left(\int_0^1 G_a\left(r\right)dr+o_p\left(1\right)\right)+O_p\left(n\right)\\ &\quad+nE\left(\hat{a}_n^{G\prime}u_tu_tZ_t'\right)\left(\int_0^1 G_a\left(r\right)dr+o_p\left(1\right)\right).\end{aligned} \tag{48}$$

By (44),

$$\sum_{t=1}^{n}\varepsilon_tZ_t=\sqrt{n}B_{Z\varepsilon}\left(1\right)+o_p\left(\sqrt{n}\right), \tag{49}$$

Finally,

$$\begin{aligned}&\sum_{t=1}^{n}\left(\beta_t\left(a\right)Y_{t-1}-\beta_t\left(\hat{a}_n^G\right)Y_{t-1}\right)Z_t\\ &=\sum_{t=1}^{n}\left(\frac{\left(a-\hat{a}_n^G\right)'u_t}{\sqrt{n}}+o_p\left(n^{-1/2}\right)\right)Y_{t-1}Z_t\\ &=n\Sigma_{Zu}\left(a-\hat{a}_n^G\right)\int_0^1 G_a\left(r\right)dr+o_p\left(n\left(a-\hat{a}_n^G\right)\right).\end{aligned} \tag{50}$$

The result of the theorem follows upon substitution of (48)-(50) into (47). ■

Proof of Theorem 7. We know from (44) 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

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

Therefore,

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

The second term above is equal to

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

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

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

The middle term in the last line is equal to

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

say. Thus,

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

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 $q-K$. ■

Table 2. Rejection rates (RR) and estimates

a	0	0.2	0.5	1	2	5
One sided RR	0.503	0.934	0.943	0.943	0.974	0.995
DF one sided RR	0.489	0.502	0.521	0.546	0.565	0.878
Two sided RR	0.492	0.675	0.828	0.900	0.953	0.986
$\overline{\hat{a}_n}$	-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: $n = 2000$, the number of replications is equal to 2000. The values were obtained for the model (1) and (2) with a 1% trimming from each tail.

Table 3. Rejection rates (RR) and estimates

n	100	500	1000	1500	2000	5000
One sided RR	0.748	0.888	0.915	0.934	0.934	0.950
DF one sided RR	0.508	0.501	0.515	0.502	0.502	0.505
Two sided RR	0.494	0.587	0.627	0.656	0.675	0.770
$\overline{\hat{a}_n}$	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 2000. The values were obtained for the model (1) and (2) with a 1% trimming from each tail.

Table 4. Rejection rates (RR) and estimates

n	100	500	1000	1500	2000	5000
One sided RR	0.836	0.908	0.930	0.944	0.947	0.958
DF one sided RR	0.507	0.528	0.539	0.543	0.546	0.514
Two sided RR	0.636	0.808	0.865	0.876	0.895	0.926
$\overline{\hat{a}_n}$	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 2000. The values were obtained for the model (1) and (2) with a 1% trimming from each tail.

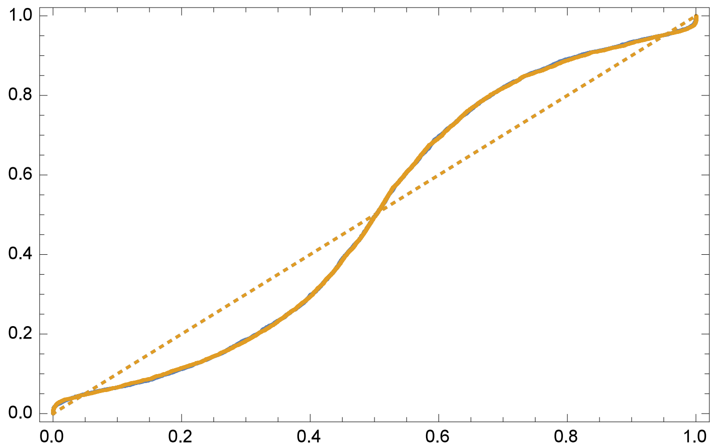

Figure 1: PP plot of $\sqrt{n}\left(\hat{a}_n^{IV} - a_0\right)$ (blue) and $B_{\varepsilon Z}(1) / \left(\sigma_{Zu} \int G_a(r)\, dr\right)$ (gold) against the estimated normal distribution of $\sqrt{n}\left(\hat{a}_n^{IV} - a_0\right)$, with $n = 100$, $\Sigma_{u\varepsilon} \neq 0$, $\Sigma_{Z\varepsilon}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

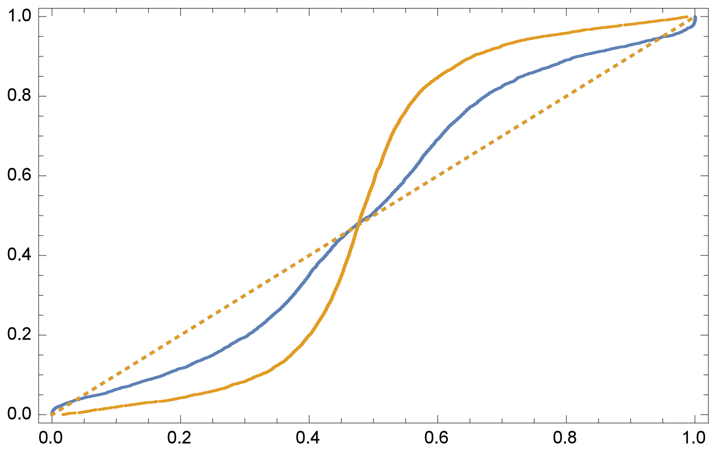

Figure 2: PP plot of $\sqrt{n}\left(\hat{a}_n^{IV} - a_0\right)$ (blue) and $B_{\varepsilon Z}(1) / \left(\sigma_{Zu} \int G_a(r)\, dr\right)$ (gold) against the estimated normal distribution of $\sqrt{n}\left(\hat{a}_n^{IV} - a_0\right)$, with $n = 1000$, $\Sigma_{u\varepsilon} \neq 0$, $\Sigma_{Z\varepsilon} \neq 0$, $\sigma_{Zu} \neq 0$, $a = 0.15$.

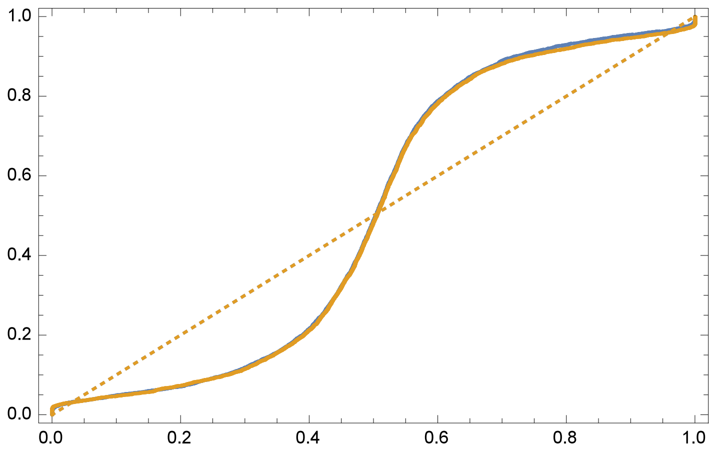

Figure 3: PP plot of $\sqrt{n}\left(\hat{a}_n^{IV} - a_0\right)$ (blue) and $B_{\varepsilon Z}(1) / \left(\hat{\sigma}_{Zu,n} \int G_a(r)\, dr\right)$ (gold) against the estimated normal distribution of $\sqrt{n}\left(\hat{a}_n^{IV} - a_0\right)$, with $n = 100$, $\Sigma_{u\varepsilon} \neq 0$, $\Sigma_{Z\varepsilon}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

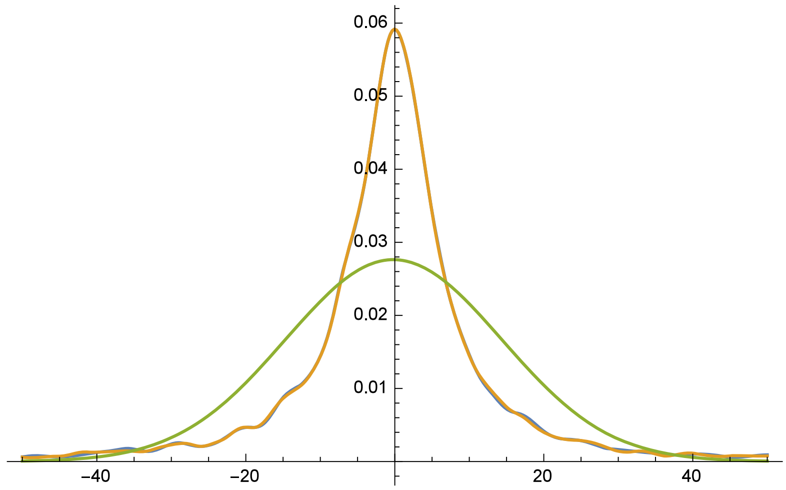

Figure 4: Kernel density estimates of $\sqrt{n}\left(\hat{a}_n^{IV} - a_0\right)$ (blue) and $B_{\varepsilon Z}(1) / \left(\sigma_{Zu} \int G_a(r)\, dr\right)$ (gold) against the estimated normal distribution of $\sqrt{n}\left(\hat{a}_n^{IV} - a_0\right)$, with $n = 10000$, $\Sigma_{u\varepsilon} \neq 0$, $\Sigma_{Z\varepsilon} = 0$, $\sigma_{Zu} \neq 0$, $a = 0.15$.

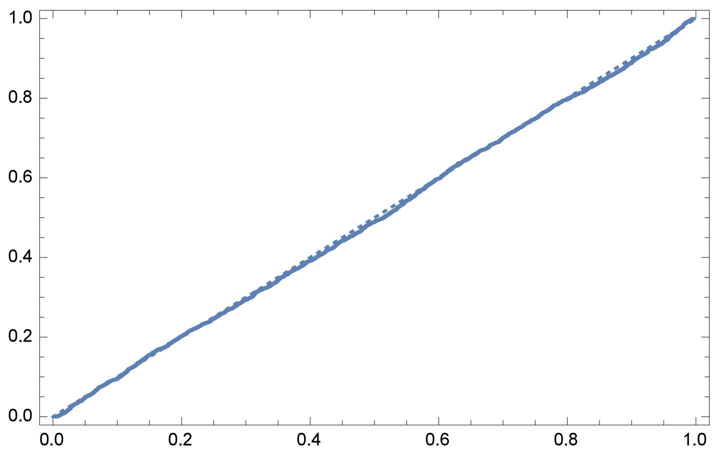

Figure 5: PP plot of $\sqrt{n}\left(\hat{a}_n^{IV} - a_0\right)$ against the scaled Cauchy variate, $n = 100$, $\Sigma_{u\varepsilon} \neq 0$, $\Sigma_{Z\varepsilon} = 0$, $\sigma_{Zu} \neq 0$, $a = 0$.

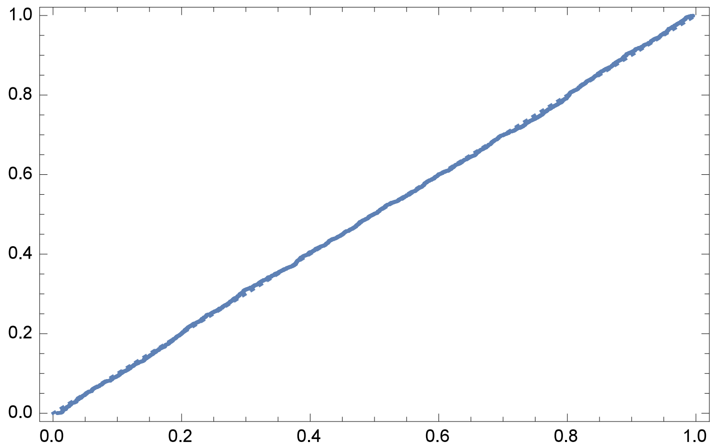

Figure 6: PP plot of $\sqrt{n}\left(\hat{a}_n^{IV} - a_0\right)$ against the scaled Cauchy variate, with the scaling factor estimated, $n = 100$, $\Sigma_{u\varepsilon} \neq 0$, $\Sigma_{Z\varepsilon} = 0$, $\sigma_{Zu} \neq 0$, $a = 0$.

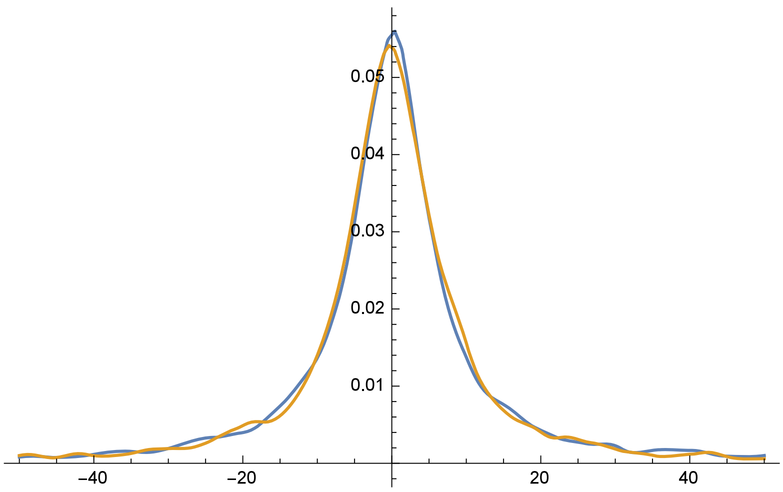

Figure 7: Kernel density estimates of $\sqrt{n}\left(\hat{a}_n^{IV} - a_0\right)$ (blue) against the scaled Cauchy variate (brown), $n = 100$, $\Sigma_{u\varepsilon} \neq 0$, $\Sigma_{Z\varepsilon} = 0$, $\sigma_{Zu} \neq 0$, $a = 0$.

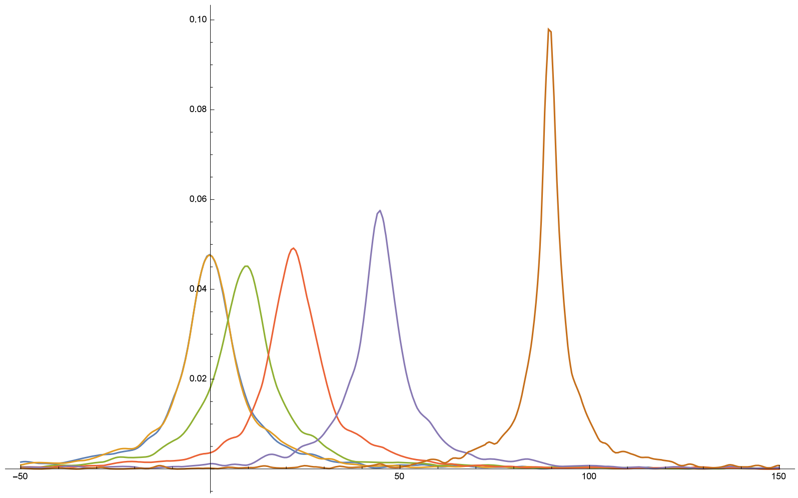

Figure 8: Trimmed (1% from each side) RHS (blue) and LHS kernel distributions, corresponds to 2000 replications with $n = 2000$. light brown ($a = 0$), green ($a = 0.2$), red ($a = 0.5$), magenta ($a = 1$), dark brown ($a = 2$).

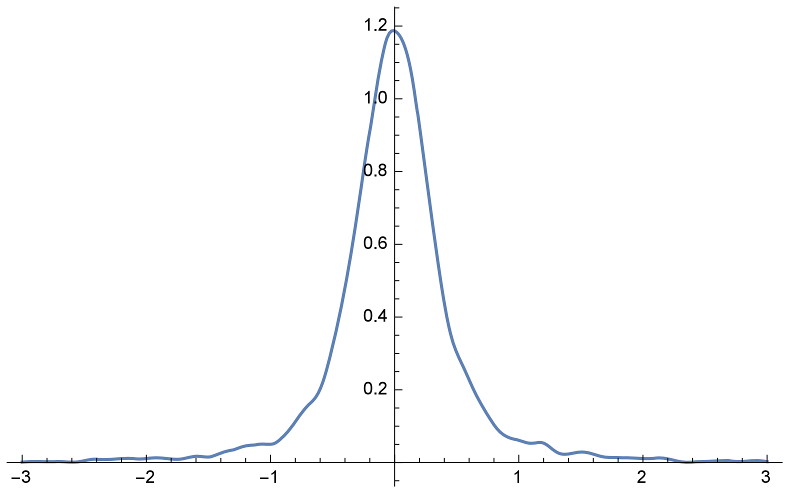

Figure 9: Kernel density estimate of $n\left(\hat{\beta}-1\right)-\left(\int_0^1 B_\varepsilon dB_\varepsilon+\lambda_{\varepsilon\varepsilon}\right)/\int_0^1 B_\varepsilon^2$, with $a = 0.2$, $n = 2000$.

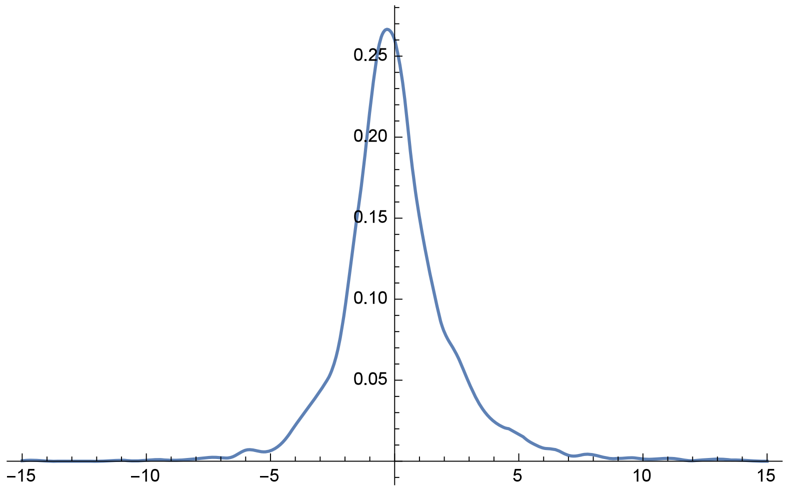

Figure 10: Kernel density estimate of $n\left(\hat{\beta}-1\right)-\left(\int_0^1 B_\varepsilon dB_\varepsilon+\lambda_{\varepsilon\varepsilon}\right)/\int_0^1 B_\varepsilon^2$, with $a=1$, $n=2000$.

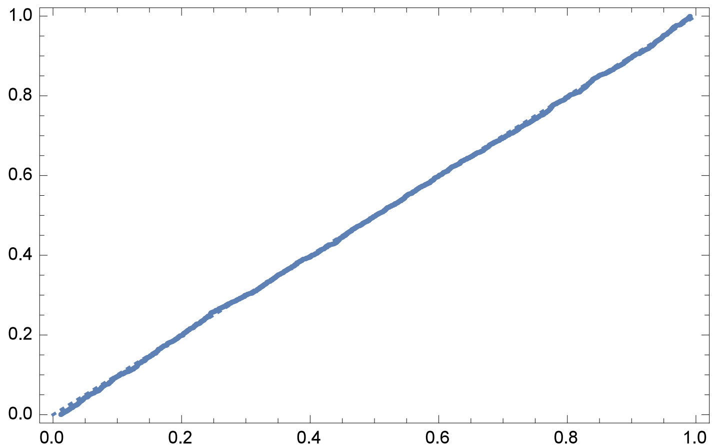

Figure 11: PP plot of $\sqrt{n}\left(\hat{a}_n^G-a_0\right)$ against $(\int\limits_0^1 G_a(r)dr)^{-1}(\Sigma_{Zu}'W\Sigma_{Zu})^{-1}\Sigma_{Zu}'B_{Z\varepsilon}(1)$, with $n=100$, $\Sigma_{u\varepsilon}\neq 0$, $\Sigma_{Z\varepsilon}=0$, $\sigma_{Zu}\neq 0$, $a=0.15$.

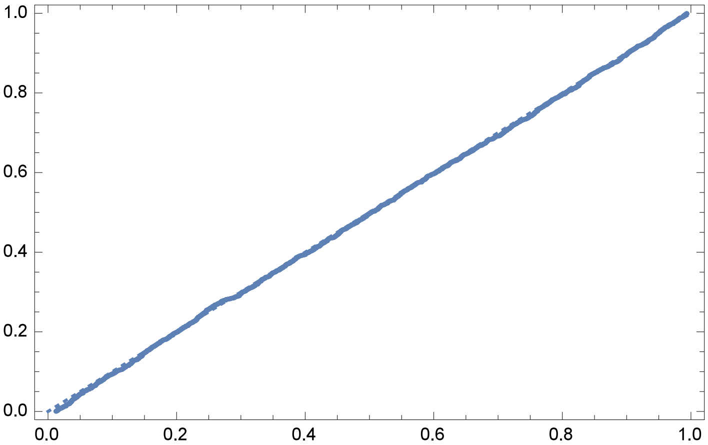

Figure 12: PP plot of $\sqrt{n}\left(\hat{a}_n^G - a_0\right)$ against $(\int_0^1 G_a(r)dr)^{-1}(\Sigma'_{Zu}W\Sigma_{Zu})^{-1}\Sigma'_{Zu}B_{Z\varepsilon}(1)$, with $n = 100$, $\Sigma_{u\varepsilon} \neq 0$, $\Sigma_{Z\varepsilon} = 0$, $\sigma_{Zu} \neq 0$, $a = 0$.

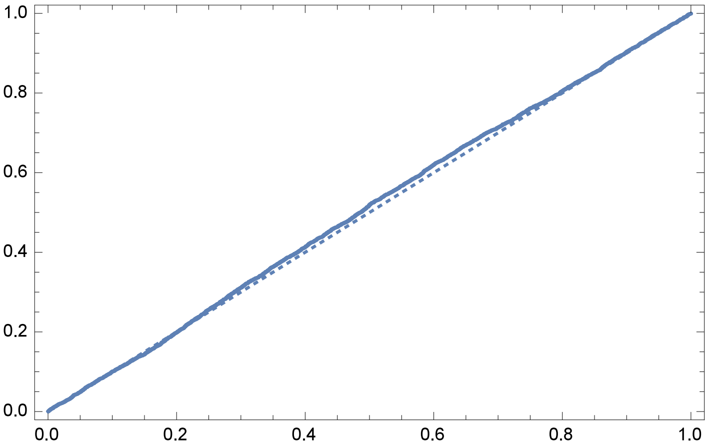

Figure 13: PP plot of $J_n(\hat{a}_n^G, \hat{W})$ against $\chi(1)$ distribution, with $n = 100$, $\Sigma_{u\varepsilon} \neq 0$, $\Sigma_{Z\varepsilon} = 0$, $\sigma_{Zu} \neq 0$, $a = 0.15$.

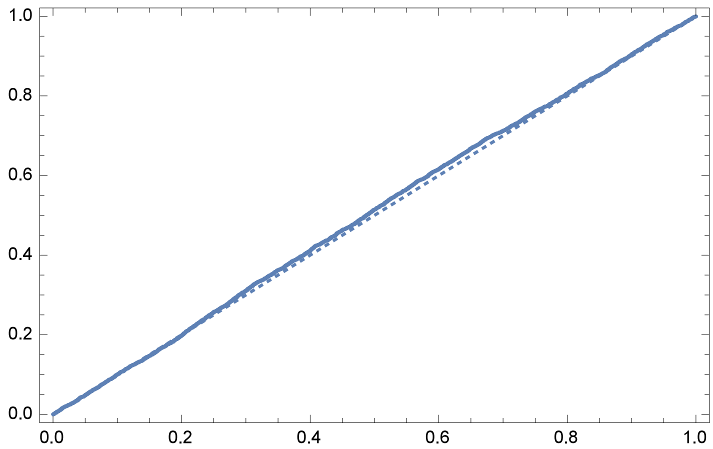

Figure 14: PP plot of $J_n(\hat{a}_n^G, \hat{W})$ against $\chi(1)$ distribution, with $n = 100$, $\Sigma_{u\varepsilon} \neq 0$, $\Sigma_{Z\varepsilon} = 0$, $\sigma_{Zu} \neq 0$, $a = 0$.

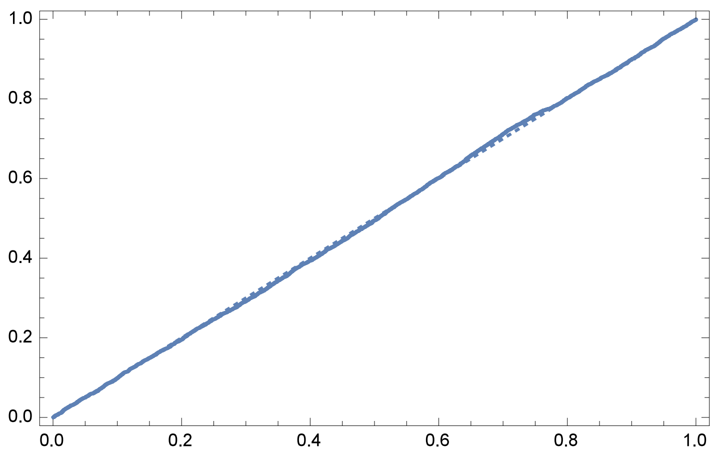

Figure 15: PP plot of $J_n(\hat{a}_n^G, \hat{W})$ against $\chi(1)$ distribution, with $n = 500$, $\Sigma_{u\varepsilon} \neq 0$, $\Sigma_{Z\varepsilon} = 0$, $\sigma_{Zu} \neq 0$, $a = 0.15$.

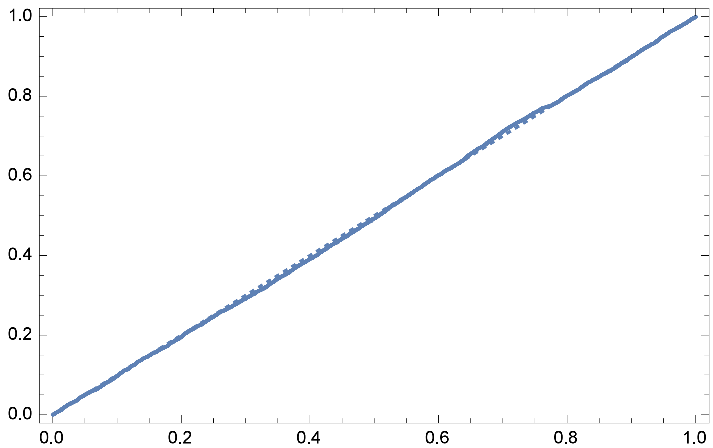

Figure 16: PP plot of $J_n(\hat{a}_n^G, \hat{W})$ against $\chi(1)$ distribution, with $n = 500$, $\Sigma_{u\varepsilon} \neq 0$, $\Sigma_{Z\varepsilon} = 0$, $\sigma_{Zu} \neq 0$, $a = 0$.