Hybrid Stochastic Local Unit Roots*

Offer Lieberman[†] and Peter C. B. Phillips[‡]

Revised, April 7, 2019

Abstract

Two approaches have dominated formulations designed to capture small departures from unit root autoregressions. The first involves deterministic departures that include local-to-unity (LUR) and mildly (or moderately) integrated (MI) specifications where departures shrink to zero as the sample size $n \to \infty$. The second approach allows for stochastic departures from unity, leading to stochastic unit root (STUR) specifications. This paper introduces a hybrid local stochastic unit root (LSTUR) specification that has both LUR and STUR components and allows for endogeneity in the time varying coefficient that introduces structural elements to the autoregression. This hybrid model generates trajectories that, upon normalization, have non-linear diffusion limit processes that link closely to models that have been studied in mathematical finance, particularly with respect to option pricing. It is shown that some LSTUR parameterizations have a mean and variance which are the same as a random walk process but with a kurtosis exceeding 3, a feature which is consistent with much financial data. We develop limit theory and asymptotic expansions for the process and document how inference in LUR and STUR autoregressions is affected asymptotically by ignoring one or the other component in the more general hybrid generating mechanism. In particular, we show how confidence belts constructed from the LUR model are affected by the presence of a STUR component in the generating mechanism. The import of these findings for empirical research are explored in an application to the spreads on US investment grade corporate debt.

Key words and phrases: Autoregression; Local unit root; Nonlinear diffusion; Stochastic unit root; Time-varying coefficient.

JEL Classification: C22

1 Introduction

For over four decades various devices have been employed to study and to model the progressive deterioration of Gaussian asymptotics in the simple first order autoregression (AR(1)) as the autoregressive coefficient (β) approaches unity from below. Edgeworth and saddlepoint approximations (Phillips, 1977, 1978) showed clearly with analytic formulae the extent of the error in the stationary asymptotics as $\beta \to 1$ and numerical computations (Evans and Savin, 1981) revealed that the unit root (UR)

*We thank the referees, Associate Editor, and Co-Editor for helpful comments on earlier versions of this paper.

[†]Bar-Ilan University. Support from Israel Science Foundation grant No. 1182-17 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. Research support from the Kelly Foundation, University of Auckland is gratefully acknowledged. Email: peter.phillips@yale.edu

limit distribution typically provides better approximations than stationary limit theory in the immediate neighborhood of unity. The use of local-to-unit root (LUR) autoregressions provided a direct approach to modeling processes with a root near unity. In independent work using different methods and assumptions, Chan and Wei (1987) and Phillips (1987) explored LUR models of the form

$$Y_t = \beta_n Y_{t-1} + \varepsilon_t,\ \beta_n = e^{c/n} \sim 1 + \frac{c}{n};\ t = 1, ..., n, \tag{1}$$

where c is constant and β_n is nearly nonstationary in the sense that c/n is necessarily small as the sample size $n \to \infty$.

Under quite general conditions on ε_t and the initial condition Y_0, the asymptotic distribution of the least squares estimator of β_n takes the form of a ratio of quadratic functionals of a linear diffusion process that depends on the localizing coefficient c in (1) and nonparametric quantities that depend on the one-sided and two-sided long run variances of ε_t. These results provided a natural path to the analysis of power functions (Phillips, 1987) and power envelopes for UR tests (Elliott *et. al.*, 1995; Elliott and Stock, 1996), as well as the construction of confidence intervals (Stock, 1991) and prediction intervals (Campbell and Yogo, 2006; Phillips, 2014) in models where persistence in the regressors is relevant in practical work.

The array mechanism of (1) has also proved useful in developing methods of uniform inference. Giraitis and Phillips (2006) established uniform asymptotic theory for the OLS estimator of β_n in models like (1) but where β_n is more distant from unity so that $(1 - \beta_n) n \to \infty$. These models allow values of stationary β_n that include neighborhoods of unity beyond the immediate $O\left(n^{-1}\right)$ vicinity of unity, such as when $\beta_n = 1 - L_n/n$, where $L_n \to \infty$ is slowly varying at infinity. These cases were explored in detail by Phillips and Magdalinos (2007a, 2007b) by using moderate deviations from unity of the form

$$\beta_n = 1 + \frac{c}{k_n}, \quad \text{with } c \text{ constant and } \frac{1}{k_n} + \frac{k_n}{n} \to 0. \tag{2}$$

Models with such roots are considered mildly integrated (MI) as β_n lies outside the LUR region as $n \to \infty$. Phillips and Magdalinos (2007a) developed central limit theory for the near-stationary case ($c < 0$) and, somewhat surprisingly, for the near-explosive case ($c > 0$), finding a Cauchy limit theory in the latter case that matched the known Cauchy limit that applies in the pure explosive case under Gaussian errors (White, 1958; Anderson, 1959). In a significant advance, Mikusheva (2007, 2012) demonstrated that careful approaches to confidence interval (CI) construction with appropriate centering were capable of producing uniform inferences about the true β_n in a wide interval that includes stationary, MI, LUR, and UR specifications.

A different approach was considered by Lieberman and Phillips (2014, 2017, 2018), who considered localized stochastic departures from unity via the stochastic unit root (STUR) model

$$\begin{aligned} Y_1 &= \varepsilon_1, \\ Y_t &= \mu + \exp\left(\frac{a' u_t}{\sqrt{n}}\right) Y_{t-1} + \varepsilon_t,\ t = 2, ..., n, \end{aligned} \tag{3}$$

where μ can be zero or otherwise and in which departures from unity are driven by a possibly endogenous $K \times 1$ vector of explanatory variables u_t. In their formulation, Lieberman and Phillips (2018) allowed $\{u_t, \varepsilon_t\}$ to follow a general linear process satisfying mild summability and moment conditions. STUR-type models have been successfully applied in the following contexts: Dual stocks (Lieberman, 2012), in which u_t can be an excess return on the stock in an 'away' market and Y_t can be the log of the price of the stock in a 'local' market, so that whether the autoregressive coefficient is less than-, equal to- or greater than unity, depends on the sign of u_t, for any given t; Exchange traded funds (Lieberman and Phillips, 2014); Option pricing (Lieberman and Phillips, 2017), where it was shown that in most moneyness-maturity categories the model substantially reduces the average percentage error of the Black-Scholes and Heston's (1993) stochastic volatility pricing schemes, and in some cases the reduction is up to 73%; Bond spreads (Lieberman and Phillips, 2018). This line of stochastic departure from a UR follows in the tradition of earlier contributions by Leybourne, McCabe and Mills (1996), Leybourne, McCabe and Tremayne (1996), Granger and Swanson (1997), McCabe and Smith, (1998), and Yoon (2006).

The present paper investigates a hybrid model that combines both LUR and STUR specifications in a localized stochastic unit root (LSTUR) model of the following form

$$\begin{aligned} Y_1 &= \varepsilon_1, \\ Y_t &= \beta_{nt} Y_{t-1} + \varepsilon_t, \ t = 2, ..., n, \end{aligned} \tag{4}$$

where

$$\beta_{nt} = \exp\left(\frac{c}{n} + \frac{a' u_t}{\sqrt{n}}\right), \tag{5}$$

$c \in \mathbb{R}$ and $a \in \mathbb{R}^K$. In this model the autoregressive coefficient is a stochastic time varying parameter that fluctuates in the vicinity of unity according to the properties of u_t, the value of the localizing constant c, and the size of the sample n. The time series $w_t = (u_t', \varepsilon_t)'$ is assumed to be generated according to a linear process framework that allows for both contemporaneous and serial cross dependence, thereby allowing the random coefficient β_{nt} to be endogenous. It is clear from the proof of Lemma 1 below that the initial condition is allowed to satisfy $Y_1 = o_p\left(n^{-1/2}\right)$ without affecting the asymptotic theory of the paper.

Denoting by $\mathcal{F}_t$ the σ-field generated by the w_t's up to time t, when w_t is an MDS our model satisfies

$$Var\left(Y_t|\mathcal{F}_{t-1}\right) = \sigma_\varepsilon^2 + \left(Var\left(\beta_{nt}|\mathcal{F}_{t-1}\right)\right) Y_{t-1}^2,$$

where σ_ε^2 is the variance of ε_t. In other words, the model has an ARCH(1) conditional variance dynamics and this class of models, is no doubt very relevant to finance.

The paper establishes limit theory for the normalized form of the output process Y_t in (4) and for nonlinear least squares (NLLS) estimation of the components, a and c, of β_{nt}. It turns out that the limiting output process of (4) is a nonlinear diffusion process that satisfies a nonlinear stochastic differential equation corresponding to a structural model of option pricing that has been considered in the continuous time mathematical finance literature (Föllmer and Schweizer, 1993). So the model may be considered a discrete time version of such a system. Working directly with this nonlinear continuous time system, Tao *et. al.* (2017) developed an estimation procedure for the structural parameters of the stochastic differential equation using a realized variance approach and established asymptotic properties of these estimates under infill asymptotics. The model considered in the present paper therefore links to the continuous time finance literature and to ongoing work on continuous time econometrics.

A primary goal of the current paper is to examine the properties of this hybrid model and, in doing so, study the implied empirical features of the model in comparison with the discrete time random walk (RW), LUR and STUR models. In particular, we show that certain LSTUR parameterization are consistent with a mean and variance which are equal to those of a RW process but with a kurtosis coefficient which is greater than 3 - a feature which is arguably consistent with much financial data. In particular, indices of financial returns typically display sample kurtosis that declines towards the Gaussian value 3 as the sampling interval increases. In work in progress we have used the LSTUR model to explain the phenomenon, with limit theory showing that LSTUR specifications provide two sources of excess kurtosis, both of which decline with the sampling interval. Applications to several financial indices demonstrate the usefulness of this approach.

The analysis helps to document how inference in LUR and STUR autoregressions is affected by the presence of the other component in the time varying autoregressive coefficient β_{nt} in the generating mechanism. In particular, we show how asymptotic confidence belts constructed using the LUR model (Stock, 1991) are affected by the omission of a random coefficient STUR component. There is ample evidence that bond spreads returns and stock market returns are negatively correlated (e.g., Kwan 1996) and so, for our empirical application we estimated an LSTUR model in which an AR(1) model for the log of the bonds spread of an index of investment grade assets has a time varying coefficient which is a function of the return on the S&P 500 index.

The plan for the rest of the paper is as follows. Notation, assumptions and limit theory for $n^{-1/2}Y_t$ are given in Section 2. Asymptotic theory for parameter estimation follows in Section 3 and extensions to the model are given in Section 4. Some further results including asymptotic expansions are given in Section 5. Robustness of the misspecified STUR-based NLLS and IV estimators of a and the covariance parameters are established in Section 6. A simulation study to the effects of an omitted STUR component on the confidence belts given by Stock (1991) for c and β in the LUR model is provided in Section 7. The empirical application supporting the analytical findings and simulations follows in Section 8. Section 9 concludes. All proofs are placed in the appendix except for proofs for the results of Section 4 which are provided in an online supplement to the paper.

2 Preliminary Limit theory for the LSTUR Model

We start with the following assumption that will be used in the sequel detailing the generating mechanism for w_t.

Assumption 1. *The vector w_t is a linear process satisfying*

$$w_t = D(L)\eta_t = \sum_{j=0}^{\infty} D_j \eta_{t-j}, \ \sum_{j=1}^{\infty} j \|D_j\| < \infty, \ D(1) \text{ has full rank } K+1, \tag{6}$$

η_t is iid, zero mean with $\mathbb{E}(\eta_t \eta_t') = \Sigma_\eta > 0$ and $\max_{i \leq K+1} \mathbb{E} |\eta_{i0}|^p < \infty$, for some $p > 4$.

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

$$n^{-1/2} \sum_{t=1}^{\lfloor n\cdot \rfloor} w_t \Rightarrow B(\cdot) \equiv \text{BM}\left(\Sigma^{\ell r}\right), \ \Sigma^{\ell r} = \begin{pmatrix} \Sigma_u^{\ell r} & \Sigma_{u\varepsilon}^{\ell r} \\ \Sigma_{u\varepsilon}^{\ell r\prime} & \left(\sigma_\varepsilon^{\ell r}\right)^2 \end{pmatrix}, \tag{7}$$

where $\lfloor \cdot \rfloor$ is the floor function and $B = (B_u, B_\varepsilon)'$ is a vector Brownian motion. The matrix $\Sigma^{\ell r} = D(1)\Sigma_\eta D(1)' > 0$ 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 (6) as

$$\begin{aligned} w_t &= \begin{pmatrix} u_t \\ \varepsilon_t \end{pmatrix} = \begin{pmatrix} D_{11}(L) & D_{12}(L) \\ D_{21}(L) & D_{22}(L) \end{pmatrix} \begin{pmatrix} \eta_{1t} \\ \eta_{2t} \end{pmatrix} = \begin{pmatrix} D_1(L) \\ D_2(L) \end{pmatrix} \begin{pmatrix} \eta_{1t} \\ \eta_{2t} \end{pmatrix} \\ &= \begin{pmatrix} \sum_{j=0}^{\infty} D_{1,j}\eta_{t-j} \\ \sum_{j=0}^{\infty} D_{2,j}\eta_{t-j} \end{pmatrix} \end{aligned} \tag{8}$$

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

We 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 $\sigma_{\varepsilon}^2 = \mathbb{E}\left(\varepsilon_t^2\right) > 0$. 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 $\Lambda_{u\varepsilon} = \sum_{h=1}^{\infty} \mathbb{E}(u_0 \varepsilon_h)$, $\lambda_{\varepsilon\varepsilon} = \sum_{h=1}^{\infty} \mathbb{E}(\varepsilon_0 \varepsilon_h)$, $\Delta_{u\varepsilon} = \sum_{h=0}^{\infty} \mathbb{E}(u_0 \varepsilon_h)$, $\Delta_{\varepsilon\varepsilon} = \sum_{h=0}^{\infty} \mathbb{E}(\varepsilon_0 \varepsilon_h)$.

We use H and L to denote the zero-one duplication and elimination matrices for which

$$vec(A) = H vech(A) \text{ and } vech(A) = L vec(A), \tag{9}$$

where A is a symmetric matrix of order $K+1$. Under Assumption 1, centred partial sums of $\eta_t \eta_t'$ satisfy the invariance principle

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor nr \rfloor} vech\left(\eta_t \eta_t' - \Sigma_\eta\right) \Rightarrow \xi(r), \tag{10}$$

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

$$\Sigma_{\eta \otimes \eta} = \mathbb{E}\left(L\left((\eta_t \otimes \eta_t) - \mathbb{E}(\eta_t \otimes \eta_t)\right)\left((\eta_t' \otimes \eta_t') - \mathbb{E}\left(\eta_t' \otimes \eta_t'\right)\right) L'\right).$$

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

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

Finally, the matrix of third order moments of η_t is denoted

$$M_3 = \mathbb{E}\left((\eta_t \otimes \eta_t) \eta_t'\right). \tag{11}$$

The limit process of the scaled time series Y_t is given in the following Lemma.

Lemma 1 *For the model (4), under Assumption 1,*

$$\frac{Y_{t=\lfloor nr \rfloor}}{\sqrt{n}} \Rightarrow G_{a,c}(r) := e^{rc + a' B_u(r)} \left(\int_0^r e^{-pc - a' B_u(p)} dB_\varepsilon(p) - a' \Delta_{u\varepsilon} \int_0^r e^{-pc - a' B_u(p)} dp \right). \tag{12}$$

Lemma 1 extends the limit theory for the special case where there is no LUR component ($c = 0$) and the case where there is no STUR component ($a = 0$). The latter case leads to the familiar limit

$$\frac{Y_t}{\sqrt{n}} \Rightarrow \sigma_{\varepsilon}^{\ell r} \int_0^r e^{(r-s)c} dW(s) =: \sigma_{\varepsilon}^{lr} J_c(r) = G_{0,c}(r) =: G_c(r), \text{ say}$$

where $W(r)$ is standard BM and $J_c(r)$ is a linear diffusion (Phillips, 1987). When the model includes a drift, the asymptotic properties are naturally different and are given in the following result.

Lemma 2 *For the extended model (4) in which $Y_1 = \mu + \varepsilon_1$, $Y_t = \mu + \beta_{nt} Y_{t-1} + \varepsilon_t$, $t = 2, ..., n$, under Assumption 1,*

$$\frac{Y_{t=\lfloor nr \rfloor}}{n} \Rightarrow \mu e^{rc + a' B_u(r)} \int_0^r e^{-pc - a' B_u(p)} dp.$$

The proof is similar to the subcase in which $c = 0$ (see, Lieberman and Phillips 2018) and is omitted for brevity. Note that in the drift case Y_t is normalized by n rather than by $\sqrt{n}$ and is asymptotically dominated by a stochastic drift.

3 Parameter Estimation

Let $\hat{a}_n$ and $\hat{c}_n$ denote the NLLS estimates of a and c, defined as solutions to the NLLS extremum estimation problem $\arg\min_{(a_1,c_1)\in\Theta} \left\{ \sum_{t=1}^n \left(Y_t - \beta_{nt}(a_1, c_1) Y_{t-1} \right)^2 \right\}$, where Θ is a compact subset of the parameter space $\mathbb{R}^K \times \mathbb{R}$ that contains the true value (a, c) in its interior.

This section presents the limit theory for these estimates in various cases. The method of proof of Theorem 5 below relies on providing an asymptotic approximation to the NLLS objective function which is asymptotically quadratic and convex in the parameters, so that the NLLS estimates exist and are unique for large enough n. We use the following sample covariance limit theory.

Lemma 3 *For the model (4), under Assumption 1*

$$\frac{1}{n} \sum_{t=2}^n \varepsilon_t Y_{t-1} \Rightarrow \int_0^1 G_{a,c}(r)\, dB_\varepsilon(r) + \Lambda'_{u\varepsilon} a \int_0^1 G_{a,c}(r)\, dr + \lambda_{\varepsilon\varepsilon}. \tag{13}$$

The limit in (13) reduces to the standard result $\int_0^1 G_{0,c}(r)\, dB_\varepsilon(r) + \lambda_{\varepsilon\varepsilon}$ when $a = 0$.

We start with the case where a is known, which enables us to relate results to earlier literature on the LUR model in a convenient way. This simplification is relaxed below.

Theorem 4 *For the model (4), under Assumption 1 and when a is known,*

$$\hat{c}_n - c \Rightarrow \left(\int_0^1 G_{a,c}^2(r)\, dr \right)^{-1} \left(\int_0^1 G_{a,c}(r)\, dB_\varepsilon(r) + \Delta'_{u\varepsilon} a \int_0^1 G_{a,c}(r)\, dr + \lambda_{\varepsilon\varepsilon} \right). \tag{14}$$

When $a = 0$ the result in (14) reduces to the standard limit theory for the least squares estimate $\hat{c}_n$ of the localizing coefficient c in a LUR model, viz.,

$$\hat{c}_n - c \Rightarrow \left(\int_0^1 G_c^2(r)\, dr \right)^{-1} \left(\int_0^1 G_c(r)\, dB_\varepsilon(r) + \lambda_{\varepsilon\varepsilon} \right). \tag{15}$$

The presence of the stochastic UR component alters the usual limit theory (15) by (i) modifying the limiting output process to $G_{a,c}(r)$ in which the effects of the random autoregressive coefficient figure, and (ii) introducing the additional bias term, $\Delta'_{u\varepsilon} a \int_0^1 G_{a,c}(r)\, dr$ to the limit distribution.

Next consider the case in which a is unknown.

Theorem 5 *For the model (4) under Assumption 1 with* $\Sigma_{u\varepsilon} \neq 0$

$$\left(\hat{a}_n - a\right) \Rightarrow \left(\int_0^1 G_{a,c}^2\left(r\right) dr\right)^{-1} \left(\int_0^1 G_{a,c}\left(r\right) dr\right) \Sigma_u^{-1} \Sigma_{u\varepsilon}, \tag{16}$$

and

$$\begin{aligned}\left(\hat{c}_n - c\right) \Rightarrow & \frac{\left(\int_0^1 G_{a,c}\left(r\right) dB_\varepsilon\left(r\right) + \Lambda_{u\varepsilon}' a \int_0^1 G_{a,c}\left(r\right) dr + \lambda_{\varepsilon\varepsilon}\right)}{\int_0^1 G_{a,c}^2\left(r\right) dr} \\ & - \Sigma_{u\varepsilon}' \Sigma_u^{-1} \frac{\int_0^1 G_{a,c}^2\left(r\right) dB_u\left(r\right) + 2\left(\Lambda_{uu}' a \int_0^1 G_{a,c}^2\left(r\right) dr + \Lambda_{u\varepsilon} \int_0^1 G_{a,c}\left(r\right) dr\right)}{\left(\int_0^1 G_{a,c}^2\left(r\right) dr\right)^2} \int_0^1 G_{a,c}\left(r\right) dr \\ & - a' \Sigma_{u\varepsilon} \frac{\int_0^1 G_{a,c}\left(r\right) dr}{\int_0^1 G_{a,c}^2\left(r\right) dr} - \frac{1}{2} \Sigma_{u\varepsilon}' \Sigma_u^{-1} \Sigma_{u\varepsilon} \left(\frac{\int_0^1 G_{a,c}\left(r\right) dr}{\int_0^1 G_{a,c}^2\left(r\right) dr}\right)^2.\end{aligned} \tag{17}$$

When $\Sigma_{u\varepsilon} = 0$,

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

and

$$\left(\hat{c}_n - c\right) \Rightarrow \left(\int_0^1 G_{a,c}^2\left(r\right) dr\right)^{-1} \left(\int_0^1 G_{a,c}\left(r\right) dB_\varepsilon\left(r\right) + \Lambda_{u\varepsilon}' a \int_0^1 G_{a,c}\left(r\right) dr + \lambda_{\varepsilon\varepsilon}\right).$$

The distribution of $\hat{a}_n$ depends on the localizing coefficient c through $G_{a,c}\left(r\right)$. The convergence rates in the theorem imply that the estimator is consistent when $\Sigma_{u\varepsilon} = 0$. When $\Sigma_{u\varepsilon} \neq 0$, the parameter a may be estimated consistently using instrumental variables (Lieberman and Phillips, 2018), see Section 6 below, or by infill asymptotics via a two-stage process involving realized variance when high frequency data is available (Tao et al., 2017)). Unlike $\hat{a}_n$, $\hat{c}_n$ is inconsistent irrespective of whether $\Sigma_{u\varepsilon} = 0$ and this accords with known results for simpler models without STUR effects (Phillips, 1987).

The next result concerns the OLS estimator of the autoregressive coefficient β_{nt}, defined in the usual way as $\hat{\beta}_n = \sum Y_t Y_{t-1} / \sum Y_{t-1}^2$ as if β_{nt} were a fixed parameter. The asymptotic distribution of $\hat{\beta}_n$ and that of the usual t-ratio for testing the hypothesis of a unit root are used later in the paper to construct confidence intervals for the autoregressive parameter.

Theorem 6 *Under Assumption 1, the OLS estimator* $\hat{\beta}_n$ *of* β_{nt}*, treated as a fixed coefficient in the* *model (4), satisfies*

$$n\left(\hat{\beta}_n - 1\right) \Rightarrow c + \frac{a'\Sigma_u a}{2}$$
$$+\frac{a'\left(\int_0^1 G_{a,c}^2(r)\,dB_u(r) + 2\left(\Lambda_{uu}' a\int_0^1 G_{a,c}^2(r)\,dr + \frac{3}{2}\Lambda_{u\varepsilon}\int_0^1 G_{a,c}(r)\,dr\right)\right) + \int_0^1 G_{a,c}(r)\,dB_\varepsilon(r) + \lambda_{\varepsilon\varepsilon}}{\int_0^1 G_{a,c}^2(r)\,dr}. \tag{18}$$

When w_t is a martingale difference, the one-sided long run covariances are zero and the limit result reduces to

$$n\left(\hat{\beta}_n - 1\right) \Rightarrow c + \frac{a'\Sigma_u a}{2} + \frac{a'\int_0^1 G_{a,c}^2(r)\,dB_u(r) + \int_0^1 G_{a,c}(r)\,dB_\varepsilon(r)}{\int_0^1 G_{a,c}(r)^2\,dr}. \tag{19}$$

4 Extensions

We provide in this subsection some possible extensions to the model. All proofs for the results of this subsection are provided in an online supplement to the paper. First, our model requires u_t to have a zero mean. Suppose that instead the autoregressive coefficient is

$$\beta_{nt}(c, a, \mu_u) = \exp\left(\frac{c}{n} + \frac{a'u_t^*}{\sqrt{n}}\right),\ u_t^* = u_t - \mu_u, \tag{20}$$

where $\mu_u = E(u_t)$ is unknown. Under Assumption 1, u_t is a linear process with coefficients that satisfy the condition of one-summability. Therefore, its sample mean, $\bar{\mu}_{u,n} = n^{-1}\sum u_t$, satisfies $\sqrt{n}\left(\bar{\mu}_{u,n} - \mu_u\right) \Rightarrow N\left(0, \Sigma_u^{\ell r}\right) \equiv Z_{\mu_u}$, where $\Sigma_u^{\ell r}$ is given in (7). Let $\hat{a}_n^p$ and $\hat{c}_n^p$ be the NLLS estimates of a and c, defined as solutions to the NLLS objective function $\arg\min_{a_1,c_1}\left\{\sum_{t=1}^n \left(Y_t - \beta_{nt}\left(a_1, c_1, \bar{\mu}_{u,n}\right)Y_{t-1}\right)^2\right\}$, in which $\bar{\mu}_u$ is plugged in for μ_u. We have the following results.

Theorem 7 *For the model (4) under Assumption 1 and with β_{nt} given by (20), the asymptotic distribution of $\hat{a}_n^p$ is equal to that of $\hat{a}_n$, whereas when $\Sigma_{u\varepsilon} \neq 0$*

$$\left(\hat{c}_n^p - c\right) \Rightarrow \frac{\int_0^1 G_{a,c}(r)\,dB_\varepsilon(r) + \Lambda_{u\varepsilon}' a\int_0^1 G_{a,c}(r)\,dr + \lambda_{\varepsilon\varepsilon}}{\int_0^1 G_{a,c}^2(r)\,dr}$$
$$-\Sigma_{u\varepsilon}'\Sigma_u^{-1}\frac{\int_0^1 G_{a,c}^2(r)\,dB_u(r) + 2\left(\Lambda_{uu}' a\int_0^1 G_{a,c}^2(r)\,dr + \Lambda_{u\varepsilon}\int_0^1 G_{a,c}(r)\,dr\right)}{\left(\int_0^1 G_{a,c}^2(r)\,dr\right)^2}\int_0^1 G_{a,c}(r)\,dr$$
$$-a'\Sigma_{u\varepsilon}\frac{\int_0^1 G_{a,c}(r)\,dr}{\int_0^1 G_{a,c}^2(r)\,dr} - \frac{1}{2}\Sigma_{u\varepsilon}'\Sigma_u^{-1}\Sigma_{u\varepsilon}\left(\frac{\int_0^1 G_{a,c}(r)\,dr}{\int_0^1 G_{a,c}^2(r)\,dr}\right)^2 + \left(a + \Sigma_u^{-1}\Sigma_{u\varepsilon}\frac{\int_0^1 G_{a,c}(r)\,dr}{\int_0^1 G_{a,c}^2(r)\,dr}\right)' Z_{\mu_u}.$$

and when $\Sigma_{u\varepsilon} = 0$,

$$\left(\hat{c}_n^p - c\right) \Rightarrow \left(\int_0^1 G_{a,c}^2(r)\,dr\right)^{-1}\left(\int_0^1 G_{a,c}(r)\,dB_\varepsilon(r) + \Lambda_{u\varepsilon}' a\int_0^1 G_{a,c}(r)\,dr + \lambda_{\varepsilon\varepsilon}\right) + a'Z_{\mu_u}.$$

Comparing Theorems 5 and 7, the asymptotic distribution of $\hat{c}_n^p$ has the additional terms

$$\left(a + \Sigma_u^{-1}\Sigma_{u\varepsilon}\frac{\int_0^1 G_{a,c}(r)\,dr}{\int_0^1 G_{a,c}^2(r)\,dr}\right)' Z_{\mu_u}$$

and $a'Z_{\mu_u}$, in the $\Sigma_{u\varepsilon} \neq 0$ and $\Sigma_{u\varepsilon} = 0$ cases respectively, which do not appear when $\mu_u = 0$. On the other end, the asymptotic distribution of the NLLS of a is unaffected by the plug-in of $\bar{\mu}_{u,n}$ in place of μ_u.

For unit root testing we make the following remark.

Remark 8 *The OLS estimator $\hat{\beta}_n^p$ of β_{nt}, treated as a fixed coefficient, when β_{nt} is given by (20) and μ_u is unknown, has the same asymptotic distribution as given by (18), because $\hat{\beta}_n^p$ depends directly on the observations on Y_t and not on unknown parameter estimates.*

Next, we address the question of how the distribution of the OLS estimator of β_{nt} would be affected by an inclusion of a constant, or a constant and trend in the dgp. To this end, let

$$\tilde{Y} = \mu_Y + Y_t \tag{21}$$

where Y_t is given by (4) and (5). The model is equivalent to

$$\tilde{Y}_t = \mu_Y(1 - \beta_{nt}) + \beta_{nt}\tilde{Y}_{t-1} + \varepsilon_t.$$

Treating the coefficients of this model as if they were fixed, the OLS estimator of the autoregressive coefficient β_{nt} is defined as $\hat{\beta}_n^\mu = \sum \tilde{Y}_t\tilde{Y}_{t-1}^\mu / \sum \left(\tilde{Y}_{t-1}^\mu\right)^2$, where $\tilde{Y}_t^\mu = \tilde{Y}_t - (n-1)^{-1}\sum_{t=1}^{n-1}\tilde{Y}_t$.

Theorem 9 *For the model given by (4), (5) and (21), under Assumption 1,*

$$\begin{aligned} n\left(\hat{\beta}_n^\mu - 1\right) &\Rightarrow c + \frac{a'\Sigma_u a}{2} \\ &+ \frac{a'\left(\int_0^1 G_{a,c}^2(r)\,dB_u(r) + 2\left(\Lambda_{uu}'a\int_0^1 G_{a,c}^2(r)\,dr + \frac{3}{2}\Lambda_{u\varepsilon}\int_0^1 G_{a,c}(r)\,dr\right)\right) + \int_0^1 G_{a,c}(r)\,dB_\varepsilon(r) + \lambda_{\varepsilon\varepsilon}}{\int_0^1 G_{a,c}^2(r)\,dr - \left(\int_0^1 G_{a,c}(r)\,dr\right)^2} \\ &- \frac{\left(\int_0^1 G_{a,c}(r)\,dr\right)a'\left(\int_0^1 G_{a,c}(r)\,dB_u(r) + \Lambda_{uu}'a\int_0^1 G_{a,c}(r)\,dr + \Lambda_{u\varepsilon}\right) + B_\varepsilon(1)\int_0^1 G_{a,c}(r)\,dr}{\int_0^1 G_{a,c}^2(r)\,dr - \left(\int_0^1 G_{a,c}(r)\,dr\right)^2}. \end{aligned}$$

The last line of the asymptotic distribution of $n\left(\hat{\beta}_n^\mu - 1\right)$ contain additional terms which do not exist in the $\mu_Y = 0$ case, given in Theorem 6 and the denominator changes to $\int_0^1 G_{a,c}^2(r)\,dr -$

$\left(\int_0^1 G_{a,c}(r)\,dr\right)^2$. In the MDS case, the formula reduced to

$$n\left(\hat{\beta}_n^{\mu}-1\right)\Rightarrow c+\frac{a'\Sigma_u a}{2}$$
$$+\frac{a'\left(\int_0^1 G_{a,c}^2(r)\,dB_u(r)-\int_0^1 G_{a,c}(r)\,dr\int_0^1 G_{a,c}(r)\,dB_u(r)\right)+\int_0^1 G_{a,c}(r)\,dB_\varepsilon(r)-B_\varepsilon(1)\int_0^1 G_{a,c}(r)\,dr}{\int_0^1 G_{a,c}^2(r)\,dr-\left(\int_0^1 G_{a,c}(r)\,dr\right)^2}. \tag{22}$$

Finally, the model

$$\tilde{Y}=\mu_Y+\delta_Y t+Y_t, \tag{23}$$

where Y_t is given by (4) and (5), is equivalent to

$$\tilde{Y}_t=\left[\mu_Y(1-\beta_{nt})+\beta_{nt}\delta_Y\right]+\delta_Y(1-\beta_{nt})\,t+\beta_{nt}\tilde{Y}_{t-1}+\varepsilon_t.$$

Treating the coefficients of this model as if they were fixed, the OLS estimator of the autoregressive coefficient β_{nt} is defined as $\hat{\beta}_n^{\tau}=\sum\tilde{Y}_t\tilde{Y}_{t-1}^{\tau}/\sum\left(\tilde{Y}_{t-1}^{\mu}\right)^2$, where $\tilde{Y}_{t-1}^{\tau}$ is the residual from the regression of $\tilde{Y}_{t-1}$ on a constant and a linear time trend. In the online supplement to the paper, we show that

$$\frac{\tilde{Y}_{t-1}^{\tau}}{\sqrt{n}}\Rightarrow G_{a,c}(r)-\int_0^1 G_{a,c}(r)\,dr-6\left(2\int_0^1 rG_{a,c}(r)\,dr-\int_0^1 G_{a,c}(r)\,dr\right)\left(r-\frac{1}{2}\right)\equiv G_{a,c}^{\tau}(r),$$

say.

Theorem 10 *For the model given by (4), (5) and (23), under Assumption 1,*

$$\begin{aligned}
n\left(\hat{\beta}_n^{\tau}-1\right)\Rightarrow&\left(\int_0^1\left(G_{a,c}^{\tau}(r)\right)^2dr\right)^{-1}\\
&\times\left(a'\left(\int_0^1 G_{a,c}^2(r)\,dB_u(r)+2\left(\Lambda_{uu}'a\int_0^1 G_{a,c}^2(r)\,dr+\Lambda_{u\varepsilon}\int_0^1 G_{a,c}(r)\,dr\right)\right)\right.\\
&-6\left(2\int_0^1 rG_{a,c}(r)\,dr-\int_0^1 G_{a,c}(r)\,dr\right)a'\left(\int_0^1 rG_{a,c}(r)\,dB_u(r)+\Lambda_{uu}'a\int_0^1 rG_{a,c}(r)\,dr+\frac{1}{2}\Lambda_{u\varepsilon}\right)\\
&+\left(6\int_0^1 rG_{a,c}(r)\,dr-4\int_0^1 G_{a,c}(r)\,dr\right)a'\left(\int_0^1 G_{a,c}(r)\,dB_u(r)+\Lambda_{uu}'a\int_0^1 G_{a,c}(r)\,dr+\Lambda_{u\varepsilon}\right)\\
&+\left(c+\frac{1}{2}a'\Sigma_u a\right)\int_0^1 G_{a,c}(r)\,G_{a,c}^{\tau}(r)\,dr+\int_0^1 G_{a,c}(r)\,dB_\varepsilon(r)+\Lambda_{u\varepsilon}'a\int_0^1 G_{a,c}(r)\,dr+\lambda_{\varepsilon\varepsilon}\\
&\left.-B_\varepsilon(1)\int_0^1 G_{a,c}(r)\,dr-6\left(2\int_0^1 rG_{a,c}(r)\,dr-\int_0^1 G_{a,c}(r)\,dr\right)\left(\int_0^1 r\,dB_\varepsilon(r)-\frac{1}{2}B_\varepsilon(1)\right)\right).
\end{aligned}$$

In the MDS case, the formula simplifies to

$$
\begin{aligned}
&n\left(\hat{\beta}_n^{\tau}-1\right) \Rightarrow \left(\int_0^1 \left(G_{a,c}^{\tau}(r)\right)^2 dr\right)^{-1} \left(a'\left(\int_0^1 G_{a,c}^2(r)\, dB_u(r)\right)\right. \\
&-6\left(2\int_0^1 rG_{a,c}(r)\, dr - \int_0^1 G_{a,c}(r)\, dr\right) a'\left(\int_0^1 rG_{a,c}(r)\, dB_u(r)\right) \\
&+\left(6\int_0^1 rG_{a,c}(r)\, dr - 4\int_0^1 G_{a,c}(r)\, dr\right) a'\left(\int_0^1 G_{a,c}(r)\, dB_u(r)\right) \\
&+\left(c+\frac{1}{2}a'\Sigma_u a\right)\int_0^1 G_{a,c}(r)\, G_{a,c}^{\tau}(r)\, dr + \int_0^1 G_{a,c}(r)\, dB_{\varepsilon}(r) - B_{\varepsilon}(1)\int_0^1 G_{a,c}(r)\, dr \\
&\left.-6\left(2\int_0^1 rG_{a,c}(r)\, dr - \int_0^1 G_{a,c}(r)\, dr\right)\left(\int_0^1 r dB_{\varepsilon}(r) - \frac{1}{2}B_{\varepsilon}(1)\right)\right).
\end{aligned}
$$

If in addition, $a = c = 0$, then $G_{a,c}(r) = B_{\varepsilon}(r)$, $G_{a,c}^{\tau}(r) = B_{\varepsilon}^{\tau}(r)$ and

$$
\begin{aligned}
n\left(\hat{\beta}_n^{\tau}-1\right) &\Rightarrow \frac{\int_0^1 B_{\varepsilon}(r)\, dB_{\varepsilon}(r) - B_{\varepsilon}(1)\int_0^1 B_{\varepsilon}(r)\, dr}{\int_0^1 \left(B_{\varepsilon}^{\tau}(r)\right)^2 dr} \\
&\frac{-6\left(2\int_0^1 rB_{\varepsilon}(r)\, dr - \int_0^1 B_{\varepsilon}(r)\, dr\right)\left(\int_0^1 r dB_{\varepsilon}(r) - \frac{1}{2}B_{\varepsilon}(1)\right)}{\int_0^1 \left(B_{\varepsilon}^{\tau}(r)\right)^2 dr}.
\end{aligned}
$$

It is straightforward to verify that this result agrees with the well known result for this case (see, for instance, Hayashi (2000, equation (9.3.19)).

5 Empirical Implications and Further Results

This section explores the relationships among the RW, LUR and LSTUR models in more detail in the univariate case ($K = 1$) with $\Sigma_{u\varepsilon} = 0$ and for iid (u_t', ε_t). This special case highlights the distinguishing features of these models and some key elements in their relationships that are important for empirical work. The output limit process (12) in this case has the simpler form

$$G_{a,c}(r) = e^{rc+aB_u(r)}\int_0^r e^{-pc-a'B_u(p)} dB_{\varepsilon}(p), \tag{24}$$

which satisfies the generating differential equation

$$dG_{a,c}(r) = aG_{a,c}(r)\, dB_u(r) + dB_{\varepsilon}(r) + \left(c+\frac{b}{2}\right) G_{a,c}(r)\, dr, \tag{25}$$

where we recall that $b = (a\sigma_u)^2$. The covariance kernel and moments of the output process $G_{a,c}(r)$ are given in the following result.

Lemma 11 *For the model (4), under the assumptions that $K = 1$, $\Sigma_{u\varepsilon} = 0$, and u_t and ε_t are iid,*

$$\mathbb{E}\left(G_{a,c}(r)\right) = 0,$$

$$Cov\left(G_{a,c}\left(r\right),G_{a,c}\left(s\right)\right)=\sigma_{\varepsilon}^{2}e^{\left(c+\frac{b}{2}\right)\left(r\vee s-r\wedge s\right)}\frac{e^{2\left(c+b\right)r\wedge s}-1}{2\left(c+b\right)}=:\gamma_{G_{a,c}}\left(r,s\right), \tag{26}$$

and

$$\mathbb{E}\left(G_{a,c}^{4}\left(r\right)\right)=\frac{3\sigma_{\varepsilon}^{4}e^{4\left(c+2b\right)r}}{c+b}\left(\frac{1-e^{-2\left(c+3b\right)r}}{2\left(c+3b\right)}-\frac{1-e^{-4\left(c+2b\right)r}}{4\left(c+2b\right)}\right). \tag{27}$$

An immediate consequence of Lemma 11 is that

$$Var\left(G_{a,c}\left(r\right)\right)=\mathbb{E}\left(G_{a,c}^{2}\left(r\right)\right)=\sigma_{\varepsilon}^{2}\frac{e^{2\left(c+b\right)r}-1}{2\left(c+b\right)}. \tag{28}$$

The function $\left(e^{zr}-1\right)/z$ is monotonically increasing and equals r at $z=0$. It follows from (28) that an LSTUR process with $c=-b$ has a limit process with variance $\sigma_{\varepsilon}^{2}r$, which is the variance of a Brownian motion. However, the process $G_{a,c}\left(r\right)$ is non-Gaussian in this case and has covariance kernel $\gamma_{G_{a,c=-b}}\left(r,s\right)=\sigma_{\varepsilon}^{2}e^{-\frac{b}{2}\left(r\vee s-r\wedge s\right)}r\wedge s\neq r\wedge s$. Thus, the particular case where $c+b=0$ provides an interesting example of a non-Gaussian LSTUR limit process whose first two moments match those of Brownian motion. For $c+b<0$ the variance of the LSTUR limit is less than that of Brownian motion and for $c+b>0$ the variance is larger and increasing with the value of $c+b$. In particular, given c, the variance of the process increases with b (equivalently, with either $|a|$ or σ_{u}). Alternatively, given b, the variance of the process increases with c. A small b expansion of (28) yields

$$Var\left(G_{a,c}\left(r\right)\right)=\sigma_{\varepsilon}^{2}\left(\frac{e^{2cr}-1}{2c}+\frac{\left(1+e^{2cr}\left(2cr-1\right)\right)}{2c^{2}}b+O\left(b^{2}\right)\right),$$

showing that the lead term of the variance is the variance of the linear diffusion LUR process, as expected, coupled with a second linear term in b.

Even though the special case $c+b=0$ matches the first two moments of the LSTUR limit process with a Brownian motion, the kurtosis of the processes differ. In particular, using Lemma 11, we have

$$\lim_{b+c\rightarrow0}\mathbb{E}\left(G_{a,c}^{4}\left(r\right)\right)=\frac{3\sigma_{\varepsilon}^{4}\left(e^{-4cr}+4cr-1\right)}{8c^{2}}=3\sigma_{\varepsilon}^{4}\left(r^{2}+O\left(c\right)\right),\text{ and }\lim_{b+c\rightarrow0}\mathbb{E}\left(G_{a,c}^{2}\left(r\right)\right)=\sigma_{\varepsilon}^{2}r, \tag{29}$$

so that in this case the kurtosis of the process, $\left\{3\sigma_{\varepsilon}^{4}\left(r^{2}+O\left(c\right)\right)\right\}/\left(\sigma_{\varepsilon}^{2}r\right)^{2}=3+O\left(c\right)$, matches that of Brownian motion when $c\rightarrow0$ because the variances are the same when $c+b=0$. However, kurtosis exceeds 3 in the case $c+b=0$ and $c<0$ and kurtosis increases as c becomes more negative when $c+b=0$. The case $c+b=0$ and $c>0$ is excluded because $b=\left(a\sigma_{u}\right)^{2}\geq0$.

We introduce a new instantaneous kurtosis measure for the process increments $dG_{a,c}(r)$ at r, which extends the idea of instantaneous drift and instantaneous volatility. The kurtosis measure is defined as

$$\kappa_{b,c}\left(r\right)=\frac{\mathbb{E}\left(\mathbb{E}\left[\left(dG_{a,c}\left(r\right)\right)^{4}|\mathcal{F}_{r}\right]\right)}{\left\{\mathbb{E}\left(\mathbb{E}\left[\left(dG_{a,c}\left(r\right)\right)^{2}|\mathcal{F}_{r}\right]\right)\right\}^{2}},$$

which has the following explicit form for the diffusion process (25)

$$\kappa_{b,c}(r) = 3 + \frac{3b^2\left[\mathbb{E}\left(G_{a,c}^4(r)\right) - \left(\mathbb{E}\left(G_{a,c}^2(r)\right)\right)^2\right]}{b^2\left(\mathbb{E}\left(G_{a,c}^2(r)\right)\right)^2 + \sigma_\varepsilon^4 + 2b\sigma_\varepsilon^2\mathbb{E}\left(G_{a,c}^2(r)\right)}, \tag{30}$$

as shown in Lemma 18 of the Appendix. The second term on the right side of (30) shows the excess kurtosis in the process increments arising from the non-Gaussianity of $G_{a,c}(r)$. As $b \to 0$ we have $\kappa_{b,c}(r) \to 3$, as expected since in that case $G_{a,c}(r) \to G_c(r) = \int_0^r e^{-(r-p)c} dB_\varepsilon(p) = \sigma_\varepsilon J_c(r)$, which is a linear Gaussian diffusion. But when $c \to 0$, $G_{a,c}(r) \to G_a(r) = e^{aB_u(r)} \int_0^r e^{-a'B_u(p)} dB_\varepsilon(p)$ which is still non-Gaussian and $\kappa_{b,0}(r) > 3$. A large b expansion of (30) shows that $\kappa_{b,c}(r) \sim \frac{9}{6}e^{4br}$, with kurtosis increasing exponentially with $b = a^2\sigma_u^2$, measuring the impact of non-Gaussianity in the process $G_{a,c}(r)$ as either a^2 or σ_u^2 rise, which originates in the nonlinear dependence of $G_{a,c}(r)$ on $aB_u(r)$.

These results are summarized in the following remark.

Remark 12 *For the model (4) with $K = 1$, $\Sigma_{u\varepsilon} = 0$, and iid (u_t, ε_t), the instantaneous kurtosis measure of the increment process $dG_{a,c}(r)$ is*

$$\kappa_{b,c}(r) = 3 + \frac{3b^2 Var\left(G_{a,c}^2(r)\right)}{b^2\left(\mathbb{E}\left(G_{a,c}^2(r)\right)\right)^2 + \sigma_\varepsilon^4 + 2b\sigma_\varepsilon^2\mathbb{E}\left(G_{a,c}^2(r)\right)},$$

and the kurtosis of the process $G_{a,c}(r)$ itself satisfies

$$\lim_{c+b\to 0} \frac{\mathbb{E}\left(G_{a,c}^4(r)\right)}{\left(\mathbb{E}\left(G_{a,c}^2(r)\right)\right)^2} = \frac{3\left(e^{-4cr} + 4cr - 1\right)}{8(cr)^2},$$

which rises as $c \to -\infty$ and has minimum of 3 at $c = 0$.

Financial data are well known to resemble trajectories generated by a RW but with the important exception that the kurtosis coefficient of asset returns exceeds 3, typically by a large margin. This stylized feature of financial times series matches the corresponding characteristic of the LSTUR limit process $G_{a,c}(r)$, which has random wandering behavior similar to a Gaussian RW but with kurtosis of its increments in excess of Gaussian increments. These features give the LSTUR process a desirable property for empirical work.

In spite of their common features, the limit processes corresponding to RW, LUR, and LSTUR time series are very different, including the special parameter configuration $c + b = 0$ in LSTUR. In particular, when $K = 1$, $\Sigma_{u\varepsilon} = 0$, and (u_t, ε_t) are iid, the limit process $G_{a,c}(r)$ satisfies the stochastic differential equation (25). Non-Gaussianity in the process $G_{a,c}(r)$ is then governed by the magnitude of the coefficient $b = a^2\sigma_u^2$. The following result sheds light on the composition of the process $G_{a,c}(r)$ when the parameter b is small.

Lemma 13 *For the model (4) when $K = 1$, $\Sigma_{u\varepsilon} = 0$, and u_t and ε_t are iid,*

$$G_{a,c}(r) = G_c(r) + V_{c,a}(r) + O_p(b), \tag{31}$$

where $G_c(r) = \int_0^r e^{(r-p)c} dB_\varepsilon(p)$ is a Gaussian process, $V_{c,a}(r) = a\int_0^r e^{(r-p)c}\left(B_u(r) - B_u(p)\right) dB_\varepsilon(p)$

is a mixed Gaussian process, and $G_c(r)$ and $V_{c,a}(r)$ are uncorrelated. To first order in b

$$Var\left(G_{a,c}(r)\right) = \sigma_\varepsilon^2 \left(\frac{e^{2cr}-1}{2c} + \frac{b\left(e^{2cr}(2cr-1)+1\right)}{(2c)^2} \right) + O(b). \tag{32}$$

According to (31) and (32) the STUR component effect is small when $b = a^2\sigma_u^2$ is small, in which case the limit process $G_{a,c}(r)$ is approximately mixed Gaussian, with variance that exceeds the variance of the LUR process component, viz.,

$$\sigma_\varepsilon^2 \left(\frac{e^{2cr}-1}{2c} + \frac{b\left(e^{2cr}(2cr-1)+1\right)}{(2c)^2} \right) \geq \sigma_\varepsilon^2 \left(\frac{e^{2cr}-1}{2c} \right).$$

In the special configuration $c + b = 0$ when b is small, c is also small and then the LSTUR process is approximately Brownian motion with variance $\sigma_\varepsilon^2 r$.

6 Robustness to Misspecification

This section explores the robustness of STUR-based NLLS and IV parameter estimation to misspecification that arises from an LSTUR generating mechanism. Let $\left(\tilde{a}_n, \tilde{\sigma}_{\varepsilon,n}^2\right)$ be the STUR-based NLLS estimates of $\left(a, \sigma_\varepsilon^2\right)$, so that

$$\tilde{a}_n = \arg\min_a \sum_t \left(Y_t - e^{a' u_t/\sqrt{n}} Y_{t-1}\right)^2, \; \tilde{\sigma}_{\varepsilon,n}^2 = \frac{1}{n}\sum_t \left(Y_t - e^{\tilde{a}_n' u_t/\sqrt{n}} Y_{t-1}\right)^2.$$

When $\Sigma_{u\varepsilon} = 0$, (u_t, ε_t) is iid, and the generating mechanism is LSTUR, $\tilde{a}_n$ and $\tilde{\sigma}_{\varepsilon,n}^2$ are still consistent for a and σ_ε^2, as shown below.

Lemma 14 *For the model (4) when $\Sigma_{u\varepsilon} = 0$ and (u_t, ε_t) is iid,*

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

(ii) $\tilde{\sigma}_{\varepsilon,n}^2 \to_p \sigma_\varepsilon^2,$

(iii) If $\tilde{Y}_t^p = \left(1 + \frac{\tilde{a}_n' u_t}{\sqrt{n}} + \frac{(\tilde{a}_n' u_t)^2}{2n}\right) Y_{t-1}$ *and* $\hat{Y}_t^p = \left(1 + \frac{\tilde{a}_n' u_t}{\sqrt{n}} + \frac{1}{n}\left(c + \frac{(\tilde{a}_n' u_t)^2}{2}\right)\right) Y_{t-1}$ *are in-sample predictors based on STUR and LSTUR specifications, then*

$$\frac{1}{\sqrt{n}} \sum_t \left(Y_t - \tilde{Y}_t^p\right) \Rightarrow B_\varepsilon(1) + c\int_0^1 G_{a,c}(r)\,dr, \quad \frac{1}{\sqrt{n}} \sum_t \left(Y_t - \hat{Y}_t^p\right) \Rightarrow B_\varepsilon(1),$$

and $\sum_t \left(\tilde{Y}_t^p - \hat{Y}_t^p\right)^2 \Rightarrow c^2 \int_0^1 G_{a,c}^2(r)\,dr.$

Parts (i) and (ii) of Lemma 14 are obtained in the same way as Theorems 2 and 3 of Lieberman and Phillips (2017). The only difference in the limit distribution in (i) compared to the case where STUR is the correct specification the limit process is now $G_{a,c}(r)$ rather than $G_a(r)$. An implication of this result is that the n^{-1}-normalized sum of squared errors of (the misspecified) STUR and LSTUR will be identical asymptotically and therefore, for large enough n, AIC and BIC should always favor STUR over LSTUR, even when LSTUR is the true DGP. This finding corresponds with the known result that information criteria such as BIC are typically blind to local departures of the LUR variety (Phillips and Ploberger, 2003; Leeb and Pötscher, 2005).

In part (iii) of the Lemma, $\tilde{Y}_t^p$ and $\hat{Y}_t^p$ are the STUR- and LSTUR-based predictors of Y_t. The latter is infeasible as c is unknown but may be replaced by an inconsistent estimate or by imposing a special restriction such as $c = -b$, which is discussed in Section 5. In this case, the $n^{-1/2}$-normalized error sums differ by the term $c\int_0^1 G_{a,c}(r)\,dr$ and the sum of squared discrepancies between the two predictors converges to $c^2\int_0^1 G_{a,c}(r)^2\,dr$ so that the value of the localizing coefficient c affects these differentials directly as well as through the correct limit process $G_{a,c}(r)$ corresponding to LSTUR rather than $G_a(r)$

In the case $\Sigma_{u\varepsilon} \neq 0$, even the correctly specified LSTUR-based NLLS estimator is inconsistent. Fortunately, for the LSTUR model the misspecified STUR-based IV estimators (Lieberman and Phillips, 2018) of a and the covariance parameters are still consistent. Let $\tilde{a}_n^{IV}$, $\tilde{\gamma}_{\varepsilon,n}^{IV}(j)$ and $\tilde{\gamma}_{u,\varepsilon,n}^{IV}(j)$ be the STUR-based IV estimators of a, $\gamma_\varepsilon(j) = Cov(\varepsilon_t, \varepsilon_{t-j})$ and $\gamma_{u,\varepsilon}(j) = Cov(u_t, \varepsilon_{t-j})$ for $(j = 0, 1, 2, ...)$. That is, $\tilde{a}_n^{IV}$ solves the K-moment conditions

$$\sum_{t=2}^{n}\left(Y_t - \beta_{nt}\left(\tilde{a}_n^{IV}\right)Y_{t-1}\right)Z_t = 0, \tag{33}$$

where Z_t is a vector of instruments which satisfy Assumption 3 of Lieberman and Phillips (2018), viz., for all t, $\mathbb{E}(Z_t\varepsilon_t) = 0$ and $\mathbb{E}(Z_t u_t') = \Sigma_{Zu}$, a matrix of full rank K,

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

and

$$e_t^{IV} = Y_t - e^{\tilde{a}_n^{IV\prime}u_t/\sqrt{n}}Y_{t-1},\ t = 2, ..., n. \tag{34}$$

An example of an instrument in an empirical example will be given in Section 9. In the misspecified model case, the STUR-IV estimators are still consistent. In particular, for the model (4) and under Assumptions 2-3 of Lieberman and Phillips (2018), we have $\tilde{a}_n^{IV} - a = O_p\left(n^{-1/2}\right)$, $\tilde{\gamma}_{\varepsilon,n}^{IV}(j) - \gamma_\varepsilon(j) = O_p\left(n^{-1/2}\right)$ and $\tilde{\gamma}_{u,\varepsilon,n}^{IV}(j) - \gamma_{u,\varepsilon}(j) = O_p\left(n^{-1/2}\right)$ for all fixed and finite j. The proof follows the arguments given in Theorems 3 and 4 of Lieberman and Phillips (2018) and is omitted.

The results established hitherto enable us to conduct hypothesis tests using plug-in estimators in place of nuisance parameters. In addition, even though $\hat{c}_n$ is inconsistent, as $\hat{a}_n$ is consistent when u_t is not endogenous, the LSTUR model can be consistently estimated in this case under the important restriction $c + b = 0$. As under endogeneity a can still can still be consistently estimated using the IV method, the LSTUR model under the restriction $c + b = 0$ can be consistently estimated in this case as well. In contrast, the LUR model cannot be consistently estimated, unless certain conditions are satisfied, such as when the data support joint large span and infill asymptotics (Tao *et. al.*, 2017). These results are employed in the empirical section below.

7 The Effects of Misspecification on CI Construction

Stock (1991) constructed confidence belts for the localizing coefficient c in the LUR model from which confidence intervals (CIs) valid within a vicinity of unity for the autoregressive coefficient β could be deduced from unit root tests. Application of this methodology to the Nelson Plosser (1982) data produced very wide confidence bands. Hansen (1999) showed how the accuracy of these simulation-based CIs deteriorated as the stationary region was approached. He suggested a grid bootstrap procedure for the construction of the CIs which helped to improve coverage accuracy of the bands[1]. By Theorem 2 of Hansen (1991) the grid bootstrap is consistent for an LUR process. However, there is no theory for LSTUR processes in general and so this method is not suitable for our purposes. Phillips (2014) provided an asymptotic analysis that explained the deterioration of the CIs as the generating process moves deeper into the LUR region and ultimately the stationary region, reinforcing the work of Hansen (1999) and Mikusheva (2007) on the role of correctly centred statistics in the development of uniformly valid confidence bands. In this literature the primary interest lies in developing CIs for the autoregressive parameter and the LUR specification is simply a vehicle for delivering such a CI that has validity over stationary and nonstationary regions.

The research referred to above was all conducted using LUR formulations of departures from unity. The present section addresses the issue of how confidence band accuracy is affected by an underspecification of an LSTUR process as an LUR process. To this end, we consider the limit distribution given in (19). The t-ratio for the UR hypothesis is given by

$$\begin{aligned} t_{\hat{\beta}} &= \frac{n\left(\hat{\beta}_n - 1\right)}{\left(\hat{\sigma}_\varepsilon^2 / n^{-2} \sum_t Y_{t-1}^2\right)^{1/2}} \\ &\Rightarrow \frac{1}{\sigma_\varepsilon}\left(\int_0^1 G_{a,c}(r)^2\, dr\right)^{1/2}\left(c + \frac{a\prime\Sigma_u a}{2} + \frac{a' \int_0^1 G_{a,c}^2(r)\, dB_u(r) + \int_0^1 G_{a,c}(r)\, dB_\varepsilon(r)}{\int_0^1 G_{a,c}(r)^2\, dr}\right), \end{aligned} \tag{35}$$

where $\hat{\sigma}_\varepsilon^2$ is a consistent estimator of σ_ε^2, such as the IV estimator $\tilde{\gamma}_{\varepsilon,n}^{IV}(0)$, discussed in Section 7. When $a = 0$, the result (35) reduces to equation (2) of Phillips (2014), or equation (5) of Stock (1991)[2]. To get an idea of how the confidence belts of Stock (1991) would be affected by the omission of a stochastic component, we simulated the right side of (35) with parameter settings $\sigma_\varepsilon^2 = 1$, $\Sigma_{u\varepsilon} = 0$, $a = (0, 1, 2, 3, 4)$, $\sigma_u^2 = (0.1, 1)$ and $c = (1, 0, -1, -5, -10, ..., -35)$. As $b = (a\sigma_u)^2$, the setting includes parameter combinations under which $-35 < c + b < 17$. Table 1 was constructed with 5000 replications and 400 integral points and includes the 5th, 10th, 50th, 90th and 95th percentiles of the simulated asymptotic distribution, as well as the width of the 80%- and 90%-CIs in each case. For brevity, in Table 1 we only display a subset of the cases. The full Table is supplied in the online supplement.

The most striking feature of the results is that the CIs become wider as the value of $c + b$ increases. In other words, *given* a c-value, the effect of misspecification becomes more pronounced as the value of a and/or σ_u^2 increases. This is expected, as a very negative value of $(c + b)$, for instance, is consistent with a dominant LUR - relative to STUR - component. Each value of c gives a point on each confidence line in the Stock (1991) confidence belts, from which the permissible values of the test statistic can be implied, given a confidence level, and vice versa. Therefore, wider CIs for the test statistic for larger a and/or σ_u^2 values translate to wider CIs for c and for β, implying that Stock's (1991) conclusion that the CIs for β are typically wide applies with greater force in the presence of a STUR component in the process. In effect, the CIs grow wider as the STUR signal becomes more dominant. For instance, suppose that the observed value $t_{\hat{\beta}} = -2.0$. Reading from Table 1, the value $c = 0$ is not in the 90%

[1] In his simulation experiment comparing between various methods, Hansen (1991, p. 600) commented that the only nongrid method that provides good coverage is Stock's confidence interval, which provides very accurate coverage rates for the autoregressive parameter near unity (including the explosive case).

[2] To be precise, Stock (1991, equation (5)) used a demeaned ADF t-statistic in constructing the confidence belts.

CI if $a = 0, 1, 2$ and $\sigma_u^2 = 0.1$, but it is inside the 90% CI if $a = 3, 4$ and $\sigma_u^2 = 0.1$. Put differently, when $b = [0, 0.4]$, $c = 0$ is not in the 90% CI, given a $t_{\hat{\beta}}$-value of -2.0, but for larger b-values, the value of $c = 0$ is within the 90% CI.

The above discussion pertains to a given c-value. In practice, as shown in the next section, a fitted LSTUR model may lead to a substantially narrower CI for c, compared with the CI for c that would be obtained from an LUR model. The results shown in this simulation are simply illustrative of the implications of having a generating mechanism that involves random as well as deterministic departures from unity. Comprehensive tabulation is a multidimensional task, involving a constellation of conceivable parameter values, and the limit theory is non-pivotal so that practical work would require consistent estimates of many unknown parameters and an approach that led to uniformly valid (over LUR and STUR departures from unity as well as stationary departures) confidence intervals. Such a program is beyond the scope of the present paper.

8 An Empirical Application

Lieberman and Phillips (2018) estimated a STUR model in which the dependent variable is the log spread between an index of U.S. dollar denominated investment grade rated corporate debt publicly issued in the U.S. domestic market and the spot Treasury curve. The variable $u_t - \bar{\mu}_{u,n}$ 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. The sample correlation between u_t and ΔY was -0.52, supporting Kwan's (1996) report of a negative correlation between stock returns and bond spread changes. In this case the NLLS estimator is inconsistent. The IV estimator, which is consistent, was estimated with 1454 daily observations over the period January 5, 2010, through to December 30, 2015, giving a value $\hat{a}_n^{IV} = -0.245$. In addition, the misspecified STUR-based IV estimators of the covariance parameters are consistent as discussed in Section 6.

The sum of squared errors (SSE) of the fitted LSTUR model under the restriction $c + b = 0$ is 0.102, for the fitted STUR model it is 0.102, and for the unrestricted AR(1) model it is 0.135. For the latter, the estimated coefficient was equal to 0.9997. Thus, LSTUR reduces the SSE of the random walk model by 24.4% and the STUR and LSTUR SSE's are equal to a third decimal place, corroborating the results of Section 7. The sample kurtosis of the returns is equal to 13.32 whereas that of the returns of the fitted values of the LSTUR model - 12.48, in line with the results of Section 6 and further enforcing the good fit of the model. The graph of the LSTUR-based $\hat{\beta}_{nt}$ against t is displayed in Figure 1, from which it is obvious that there is a substantial variation over the unit root. In particular, the mean, variance, skewness and kurtosis of $\hat{\beta}_{nt}$ are 0.99998, 0.08, 0.57 and 7.94, respectively and the range is $[0.97, 1.05]$, showing that large deviations from the unit root do occur over the sample.

Using the IV estimation results we calculated the t-statistic (35) with error variance estimated by $\tilde{\gamma}_{\varepsilon,n}^{IV}(0)$ obtaining a value of $t_{\hat{\beta}} = -0.659$. We recall that by Remark 8 demeaning of u_t does not affect the asymptotic distribution of the test statistic. The 5th, 10th, 50th, 90th and 95th percentiles of the asymptotic distribution were simulated[3] using the rhs of (35) with parameters replaced by their IV-consistent estimates. The intersection points of $t_{\hat{\beta}} = -0.659$ with the percentiles are shown in Figure 2 and Table 2. The 90% CI for c is given by the intersection of the horizontal line $t_{\hat{\beta}} = -0.659$ and the 5th and 95th percentiles lines, in the $\left(c, t_{\hat{\beta}}\right)$ plane. As shown in Figure 2, the CI lower and upper

[3] A MATHEMATICA program was written to evaluate the percentiles using 400 integration points, 5000 replications and a grid of 0.1 over the c-values.

limits are $c_L^a = -4.15$ and $c_U^a = 3.16$ and the median unbiased estimate of c in the LSTUR model is $\hat{c}_{med} = -0.42$. The procedure was repeated for the LUR model, where the asymptotic distribution is given by (35), with $a = 0$, $\sigma_u^2 = 0$, $\Sigma_{u\varepsilon} = 0$. The results are shown in Table 2. For this model we obtain the 90% CI limits $c_L = -4.25$ and $c_U = 3.24$.

Table 2 reveals that the 90% CI for c, which is LSTUR-based, is narrower and is in fact fully within the 90% LUR-based CI. Thus, at least in this case, LSTUR attenuates the estimated impact of c on the time varying autoregressive coefficient β_{nt}. The induced 90% CI for β_{nt} which is LUR-based is approximately $[1 - 4.25/n, 1 + 3.24/n]$, whereas the variation of u_t needs to be accounted for in the construction of an LSTUR-based 90% CI for β_{nt}. Conditional on u_t and on the values of the nuisance parameters, the LSTUR-based 90% CI for β_{nt} is $\left[e^{-4.15/n + au_t/\sqrt{n}}, e^{3.16/n + au_t/\sqrt{n}}\right]$, so that the width of the interval is approximately $7.31/n$, compared with a width of $7.49/n$ for the LUR-based CI. The means of the CI bounds, taken with respect to u_t and assuming that w_t is multivariate normal, are $\mathbb{E}e^{c_L^a + au_t/\sqrt{n}} = e^{\left(c_L^a + b/2\right)/n}$ and $\mathbb{E}e^{c_U^a + au_t/\sqrt{n}} = e^{\left(c_U^a + b/2\right)/n}$. Plugging in the IV estimates, $\hat{a}_n^{IV} = -0.245$ and $\hat{\sigma}_u^2 = n^{-1}\sum\left(u_t - \bar{\mu}_{u,n}\right)^2 = 0.983$[4] into these formulae, the estimated means of the bounds are $1 - 4.12/n$ and $1 + 3.19/n$, which are slightly smaller in absolute values than the respective LUR-based bounds. Furthermore, Given the model parameters, and assuming that w_t is multivariate normal,

$$\Pr\left(e^{\frac{c}{n} + \frac{au_t}{\sqrt{n}}} < L\right) = \alpha \ \text{ iff } L = e^{\frac{c}{n} + \frac{c_\alpha \sqrt{b}}{\sqrt{n}}},$$

where c_α is the α'th percentile of the standard normal distribution. Thus, given the model parameters and the distribution of w_t, the induced 90% CI for β_{nt} is

$$\left[e^{-\frac{4.15}{n} - \frac{0.4}{\sqrt{n}}}, e^{\frac{3.16}{n} + \frac{0.4}{\sqrt{n}}}\right].$$

So, the width of the CI is approximately $0.8/\sqrt{n} + 7.31/n$. Compared with the LUR-based induced CI for β_{nt}, the LSTUR-based induced CI has a term which is $O\left(n^{-1/2}\right)$, to account for the additional variability in β_{nt} which is due to u_t. On the other hand, the $O\left(n^{-1}\right)$ term in the CI which is due to c and b in LSTUR and due to c only in LUR, is smaller in absolute value in the LSTUR-based bounds than in LUR.

We remark that an 'exact' analytical CI which accounts for the variability in the estimates of a and the covariance parameters is analytically intractable, because these estimates influence both the percentiles of $t_{\hat{\beta}}$ (and, hence, the values c_L^a and c_U^a) as well as the summand $\hat{a}_n^{IV} u_t/\sqrt{n}$. Nevertheless, qualitatively, the message from the empirical application is that the reported CI for c can be wrong and, in reality, wider when an LSTUR process is misspecified as a LUR model. On the other hand, unconditionally, the induced CI for β_{nt} is wider when a STUR component is present as is expected from the additional random variability that is embodied in the LSTUR representation of the time variation in the autoregressive coefficient.

9 Discussion

It is widely acknowledged that with much economic and financial data the unit root hypothesis may only hold approximately or in some sense on average over a given sample. A more general modeling

[4] The estimator $\hat{\sigma}_u^2$ is consistent as it does not depend on a.

perspective that offers greater flexibility is that the generating mechanism may involve temporary departures from unity at any sample point that can move the process in stationary or explosive directions. Recognition of this type of functional coefficient flexibility and its relevance for applied work has led to the literature on LUR, functional LUR (Bykhovskaya and Phillips, 2017, 2018), and STUR models, which seek to capture certain non-random and random departures from an autoregressive unit root process. The hybrid model introduced in this paper incorporates two streams of this literature as special cases and the limit theory generalizes results already known for the LUR and STUR models. As expected, ignoring one or other of these component departures introduces inferential bias. Both simulations and empirics reveal how the construction of uniform confidence intervals for autoregressive coefficients using a LUR model formulation are affected by misspecification in which the random departures of the LSTUR mechanism are neglected. Of particular relevance in applications is the fact that an LSTUR process, may have the same mean and variance as a Gaussian random walk but with kurtosis that is well in excess of 3, a feature that is consonant with the heavy tails of much observed financial return data.

References

Anderson, T. W. (1959). "On asymptotic distributions of estimates of parameters of stochastic difference equations," *The Annals of Mathematical Statistics* 30(3), 676–687.

Bykhovskaya, A. and P. C. B. Phillips (2017). "Point Optimal Testing with Roots that are Functionally Local to Unity", *Journal of Econometrics* (forthcoming).

Bykhovskaya, A. and P. C. B. Phillips (2018). "Boundary Limit Theory for Functional Local to Unity Regression", *Journal of Time Series Analysis* 39, 523–562.

Campbell, J.Y. and M. Yogo (2006). "Efficient tests of stock return predictability," *Journal of Financial Economics* 81(1), 27–60.

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

Evans, G. B. A. and N. E. Savin (1981). "Testing for a unit root: 1" *Econometrica* 49(3), 753–779.

Elliott, G., T. J. Rothenberg and J. H. Stock (1996). "Efficient tests of an autoregressive unit root," *Econometrica,* Vol. 64, 813 - 836.

Föllmer, H. and M. Schweizer (1993). "A microeconomic approach to diffusion models for stock prices," *Mathematical Finance* 3(1), 1–23.

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

Granger, C. W. and N. R. Swanson (1997). "An introduction to stochastic unit-root processes," *Journal of Econometrics 80(1), 35–62.*

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.

Kwan, S. H. (1996). "Firm-specific information and the correlation between individual stocks and bonds," *Journal of Financial Economics,* 40, 63–80.

Leadbetter, M. R., Lindgren, G, and H. Rootzén (1982). *Extremes and Related Propoerties of Random Sequences and Processes.* Springer–Verlag.

Leeb H. and B. Pötscher (2005). "Model selection and inference: facts and fiction," *Econometric Theory*, 21, pp. 21-59.

Leybourne, S. J., McCabe, B. P., and T. C. Mills (1996). "Randomized unit root processes for modelling and forecasting financial time series: theory and applications," *Journal of Forecasting* 15(3), 253-270.

Leybourne, S. J., McCabe, B. P. and A. R. Tremayne (1996). "Can economic time series be differenced to stationarity?," *Journal of Business and Economic Statistics* 14(4), 435-446.

Lieberman, O. (2012). “A similarity-based approach to time-varying coefficient non-stationary autoregression," *Journal of Time Series Analysis* 33, 484–502.

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 (2017). “A multivariate stochastic unit root model with an application to derivative pricing,” *Journal of Econometrics,* 196, 99-110.

Lieberman, O. and P. C. B. Phillips (2018). “IV and GMM Inference in Endogenous Stochastic Unit Root Models". *Econometric Theory,* 34, 1065–1100.

McCabe, B. P. and R. J. Smith (1998). “The power of some tests for difference stationarity under local heteroscedastic integration," *Journal of the American Statistical Association* 93(442), 751-761.

Mikusheva, A. (2007). “Uniform inference in autoregressive models," *Econometrica* 75(5), 1411–1452.

Mikusheva, A. (2012). “One-dimensional inference in autoregressive models with the potential presence of a unit root," *Econometrica* 80(1), 173–212.

Phillips, P. C. B. (1987). “Towards a unified asymptotic theory for autoregression,” *Biometrika* 74, 535–547.

Phillips, P. C. B. (1977). “Approximations to Some Finite Sample Distributions Associated with a First Order Stochastic Difference Equation,” *Econometrica* 45:2, 463–485.

Phillips, P. C. B. (1978). “Edgeworth and saddlepoint approximations in a first order non-circular autoregression,” *Biometrika* 65:1, 91–98.

Phillips, P. C. B. (2014). “On confidence intervals for autoregressive roots and predictive regression," *Econometrica* 82(3), 1177-1195.

Phillips, P. C. B. and T. Magdalinos (2007a). “Limit theory for moderate deviations from a unit root," *Journal of Econometrics* 136, 115-130.

Phillips, P. C. B. and T. Magdalinos (2007b), "Limit Theory for Moderate Deviations from a Unit Root Under Weak Dependence," in G. D. A. Phillips and E. Tzavalis (Eds.) *The Refinement of Econometric Estimation and Test Procedures: Finite Sample and Asymptotic Analysis.* Cambridge: Cambridge University Press, pp.123-162.

Ploberger, W. and P. C. B. Phillips (2003). “Empirical limits for time series econometric models” *Econometrica,* 71, 627-673.

Stock, J. H. (1991). “Confidence intervals for the largest autoregressive root in U.S. macroeconomic time series," *Journal of Monetary Economics* 28(3), 435–459.

Tao, Y., P. C. B. Phillips and J. Yu (2017). “Random coefficient continuous systems: testing for extreme sample path behavior”, Working paper, Singapore Management University.

White, J. S. (1958). “The limiting distribution of the serial correlation coefficient in the explosive case," *The Annals of Mathematical Statistics* 29(4), 1188-1197.

Yoon, G. (2006). "A note on some properties of STUR processes," *Oxford Bulletin of Economics and Statistics* 68(2), 253-260.

Table 1. Percentiles and Confidence Intervals for $t_{\hat{\beta}}$, $\sigma_u^2 = 0.1$

c	a	5	10	50	90	95	80% CI	90% CI
1	0	-1.724	-1.361	0.3	2.466	3.144	3.827	4.868
1	1	-1.73	-1.355	0.273	2.588	3.423	3.943	5.152
1	2	-1.716	-1.367	0.151	2.935	4.199	4.302	5.915
1	3	-1.807	-1.4917	-0.1127	3.988	6.432	5.479	8.239
1	4	-1.974	-1.641	-0.467	4.921	9.538	6.562	11.512
0	0	-1.913	-1.62	-0.508	0.9445	1.311	2.564	3.224
0	1	-1.919	-1.607	-0.534	0.907	1.374	2.514	3.293
0	2	-1.976	-1.641	-0.562	1.068	1.753	2.71	3.729
0	3	-2.03	-1.712	-0.736	1.25	2.264	2.962	4.294
0	4	-2.054	-1.817	-0.94	1.882	3.437	3.698	5.492
-1	0	-2.145	-1.852	-0.877	0.125	0.432	1.977	2.576
-1	1	-2.136	-1.812	-0.887	0.101	0.472	1.914	2.608
-1	2	-2.195	-1.889	-0.955	0.171	0.592	2.059	2.786
-1	3	-2.195	-1.905	-1.021	0.21	0.862	2.115	3.057
-1	4	-2.255	-1.986	-1.187	0.32	1.277	2.306	3.532
-5	0	-2.746	-2.472	-1.642	-0.996	-0.81	1.476	1.936
-5	1	-2.755	-2.501	-1.656	-0.996	-0.802	1.505	1.953
-5	2	-2.757	-2.496	-1.664	-1.01	-0.819	1.486	1.938
-5	3	-2.788	-2.52	-1.7	-1.023	-0.804	1.497	1.983
-5	4	-2.875	-2.556	-1.755	-1.102	-0.838	1.454	2.037
-10	0	-3.283	-3.047	-2.257	-1.628	-1.468	1.419	1.815
-10	1	-3.274	-3.026	-2.269	-1.63	-1.463	1.396	1.811
-10	2	-3.297	-3.053	-2.285	-1.642	-1.475	1.411	1.821
-10	3	-3.308	-3.064	-2.272	-1.624	-1.462	1.44	1.846
-10	4	-3.35	-3.116	-2.299	-1.685	-1.496	1.431	1.854

Note: The entries in the table are the percentiles- and confidence interval width (last two columns) of the limit distribution of the statistic $t_{\hat{\beta}}$, based on 5000 replications and 400 integral points, with $\sigma_\varepsilon^2 = 1$, $\Sigma_{u\varepsilon} = 0$. The full table is given in the online supplement to the paper.

Table 2. c-values percentiles.

	Percentiles				
Model	5th	10th	50th	90th	95th
LUR	3.24	2.26	-0.41	-3.16	-4.25
LSTUR	3.16	2.23	–0.42	-3.21	-4.15

Note: The figures in the Table are percentile c-values, obtained from the intersections of the line $t_{\hat{\beta}} = -0.659$ with the confidence lines for the LUR and LSTUR models, using Stock's (1991) method.

10 Proofs

10.1 Proofs of Lemmas and Supplementary Results

Proof of Lemma 3. From (8), $\frac{1}{n}\sum_{t=2}^{n}\varepsilon_t Y_{t-1} = \frac{1}{n}\sum_{t=2}^{n}\sum_{j=0}^{\infty} D_{2,j}\eta_{t-j}Y_{t-1}$. As Y_{t-1} is uncorrelated with η_t,

$$\frac{1}{n}D_{2,0}\sum_{t=2}^{n}\eta_t Y_{t-1} \Rightarrow D_{2,0}\int_0^1 G_{a,c}(r)\, dB_\eta(r)\,. \tag{36}$$

Next decompose the contemporaneous sample covariance as

$$\begin{aligned}\frac{1}{n}D_{2,1}\sum_t \eta_{t-1}Y_{t-1} &= \frac{1}{n}D_{2,1}\sum_t \eta_{t-1}\left(\exp\left(\frac{c}{n}+\frac{a'u_{t-1}}{\sqrt{n}}\right)Y_{t-2}+\varepsilon_{t-1}\right)\\ &= \frac{1}{n}D_{2,1}\sum_t \eta_{t-1}\left\{\left(1+\frac{a'u_{t-1}}{\sqrt{n}}+o_p\left(n^{-1/2}\right)\right)Y_{t-2}+\varepsilon_{t-1}\right\},\end{aligned} \tag{37}$$

where $\frac{1}{n}\sum_t \eta_{t-1}Y_{t-2} \Rightarrow \int_0^1 G_{a,c}(r)\, dB_\eta(r)$ from (36) and

$$\begin{aligned}&\frac{1}{n^{3/2}}\sum_t \eta_{t-1}a'u_{t-1}Y_{t-2} = \frac{1}{n^{3/2}}\sum_{t=2}^{\infty}\eta_{t-1}\left(\sum_{j=0}^{\infty}D_{1,j}\eta_{t-1-j}\right)' aY_{t-2}\\ &= \frac{1}{n^{3/2}}\sum_{t=2}^{\infty}\eta_{t-1}\eta_{t-1}'D_{1,0}'aY_{t-2} + \frac{1}{n^{3/2}}\sum_{t=2}^{\infty}\eta_{t-1}\sum_{j=1}^{\infty}\eta_{t-1-j}'D_{1,j}'a\left(\exp\left(\frac{c}{n}+\frac{a'u_{t-2}}{\sqrt{n}}\right)Y_{t-3}+\varepsilon_{t-2}\right)\\ &= \Sigma_\eta D_{1,0}'a\int_0^1 G_{a,c}(r)\, dr + \frac{1}{n^{3/2}}\sum_{t=2}^{\infty}\eta_{t-1}\sum_{j=1}^{\infty}\eta_{t-1-j}'D_{1,j}'a\varepsilon_{t-2} + o_p(1)\,.\end{aligned} \tag{38}$$

Further,

$$\begin{aligned}&\frac{1}{n^{3/2}}\sum_{t=2}^{\infty}\eta_{t-1}\sum_{j=1}^{\infty}\eta_{t-1-j}'D_{1,j}'a\varepsilon_{t-2} = \frac{1}{n^{3/2}}\sum_{t=2}^{\infty}\eta_{t-1}\sum_{j=1}^{\infty}\eta_{t-1-j}'D_{1,j}'a\sum_{k=0}^{\infty}D_{2,k}\eta_{t-2-k}\\ &= \frac{1}{n^{3/2}}\sum_{t=2}^{\infty}\eta_{t-1}\sum_{j=1}^{\infty}\sum_{k=0}^{\infty}D_{2,k}\eta_{t-2-k}\eta_{t-1-j}'D_{1,j}'a = O_p\left(n^{-1}\right),\end{aligned}$$

so that

$$\frac{1}{n^{3/2}}\sum_t \eta_{t-1}a'u_{t-1}Y_{t-2} \Rightarrow \Sigma_\eta D'_{1,0}a\int_0^1 G_{a,c}(r)\,dr. \tag{39}$$

The last non-vanishing term in (37) involves

$$\frac{1}{n}\sum_{t=2}^{n}\eta_{t-1}\varepsilon_{t-1} = \frac{1}{n}\sum_{t=2}^{n}\eta_{t-1}\sum_{k=0}^{\infty}D_{2,k}\eta_{t-1-k} \to_p \Sigma_\eta D'_{2,0}, \tag{40}$$

and so

$$\frac{1}{n}D_{2,1}\sum_t \eta_{t-1}Y_{t-1} \Rightarrow D_{2,1}\left(\int_0^1 G_{a,c}(r)\,dB_\eta(r) + \Sigma_\eta D'_{1,0}a\int_0^1 G_{a,c}(r)\,dr + \Sigma_\eta D'_{2,0}\right). \tag{41}$$

Continuing,

$$\begin{aligned} Y_{t-1} &= \exp\left(\frac{c}{n}+\frac{a'u_{t-1}}{\sqrt{n}}\right)Y_{t-2}+\varepsilon_{t-1} \\ &= \exp\left(\frac{c}{n}+\frac{a'u_{t-1}}{\sqrt{n}}\right)\left(\exp\left(\frac{c}{n}+\frac{a'u_{t-2}}{\sqrt{n}}\right)Y_{t-3}+\varepsilon_{t-2}\right)+\varepsilon_{t-1} \\ &= \exp\left(\frac{2c}{n}+\frac{a'(u_{t-1}+u_{t-2})}{\sqrt{n}}\right)Y_{t-3}+\exp\left(\frac{c}{n}+\frac{a'u_{t-1}}{\sqrt{n}}\right)\varepsilon_{t-2}+\varepsilon_{t-1}. \end{aligned} \tag{42}$$

We therefore have

$$\frac{1}{n}D_{2,2}\sum_t \eta_{t-2}Y_{t-1} = \frac{1}{n}D_{2,2}\sum_t \eta_{t-2}\left(\exp\left(\frac{2c}{n}+\frac{a'(u_{t-1}+u_{t-2})}{\sqrt{n}}\right)Y_{t-3}+\exp\left(\frac{c}{n}+\frac{a'u_{t-1}}{\sqrt{n}}\right)\varepsilon_{t-2}+\varepsilon_{t-1}\right) \tag{43}$$

$$= \frac{1}{n}D_{2,2}\sum_t \eta_{t-2}\left(1+\frac{a'(u_{t-1}+u_{t-2})}{\sqrt{n}}\right)Y_{t-3} \tag{44}$$

$$+\frac{1}{n}D_{2,2}\sum_t \eta_{t-2}\exp\left(\frac{c}{n}+\frac{a'u_{t-1}}{\sqrt{n}}\right)\varepsilon_{t-2} \tag{45}$$

$$+\frac{1}{n}D_{2,2}\sum_t \eta_{t-2}\varepsilon_{t-1}+o_p(1). \tag{46}$$

To deal with (44), we write

$$\frac{1}{n}\sum_t \eta_{t-2}\left(1+\frac{a'(u_{t-1}+u_{t-2})}{\sqrt{n}}\right)Y_{t-3} = \int_0^1 G_{a,c}(r)\,dB_\eta(r)+\frac{1}{n^{3/2}}\sum_t \eta_{t-2}a'(u_{t-1}+u_{t-2})Y_{t-3}+o_p(1), \tag{47}$$

and using (39) gives

$$\frac{1}{n^{3/2}}\sum_t \eta_{t-2}a'u_{t-2}Y_{t-3} \Rightarrow \Sigma_\eta D'_{1,0}a\int_0^1 G_{a,c}(r)\,dr, \tag{48}$$

so that

$$\frac{1}{n^{3/2}}\sum_t \eta_{t-2}a'u_{t-1}Y_{t-3}=\frac{1}{n^{3/2}}\sum_t \eta_{t-2}\left(\sum_{j=0}^{\infty}D_{1,j}\eta_{t-1-j}\right)' aY_{t-3}=\frac{1}{n^{3/2}}\sum_t \eta_{t-2}\eta_{t-1}'D_{1,0}'aY_{t-3}$$
$$+\frac{1}{n^{3/2}}\sum_t \eta_{t-2}\eta_{t-2}'D_{1,1}'aY_{t-3}+\frac{1}{n^{3/2}}\sum_t \eta_{t-2}\sum_{j=3}^{\infty}\eta_{t-j}'D_{1,j-1}'aY_{t-3}. \tag{49}$$

Two types of terms occur in (49): one with equal lags of η_t and the other with non-equal lags. Since $Y_{t-3}=\exp\left(\frac{c}{n}+\frac{a'u_{t-3}}{\sqrt{n}}\right)Y_{t-4}+\varepsilon_{t-3}$, the first term is

$$\begin{aligned}&\frac{1}{n^{3/2}}\sum_t \eta_{t-2}\eta_{t-1}'D_{1,0}'aY_{t-3}=\frac{1}{n^{3/2}}\sum_t \eta_{t-2}\eta_{t-1}'D_{1,0}'a\varepsilon_{t-3}+o_p(1)\\ =\;&\frac{1}{n^{3/2}}\sum_t \eta_{t-2}\eta_{t-1}'D_{1,0}'a\left(D_{2,0}\eta_{t-3}+\cdots\right)+o_p(1)=o_p(1).\end{aligned} \tag{50}$$

The second term in (49) gives

$$\frac{1}{n^{3/2}}\sum_t \eta_{t-2}\eta_{t-2}'D_{1,1}'aY_{t-3}\Rightarrow \Sigma_\eta D_{1,1}'a\int_0^1 G_{a,c}(r)\,dr, \tag{51}$$

and the third term in (49) is

$$\frac{1}{n^{3/2}}\sum_t \eta_{t-2}\sum_{j=3}^{\infty}\eta_{t-j}'D_{1,j-1}'aY_{t-3}=\frac{1}{n^{3/2}}\sum_t \eta_{t-2}\sum_{j=3}^{\infty}\eta_{t-j}'D_{1,j-1}'a\varepsilon_{t-3}+o_p(1)=o_p(1). \tag{52}$$

It follows from (49)-(52) that

$$\frac{1}{n^{3/2}}\sum_t \eta_{t-2}a'u_{t-1}Y_{t-3}\Rightarrow \Sigma_\eta D_{1,1}'a\int_0^1 G_{a,c}(r)\,dr. \tag{53}$$

Then, from (47), (48) and (53) we deduce that

$$\frac{1}{n}\sum_t \eta_{t-2}\left(1+\frac{a'\left(u_{t-1}+u_{t-2}\right)}{\sqrt{n}}\right)Y_{t-3}\Rightarrow \int_0^1 G_{a,c}(r)\,dB_\eta(r)+\Sigma_\eta\sum_{j=0}^{1}D_{1,j}'a\int_0^1 G_{a,c}(r)\,dr. \tag{54}$$

Using (40), we have

$$\frac{1}{n}\sum_t \eta_{t-2}\exp\left(\frac{c}{n}+\frac{a'u_{t-1}}{\sqrt{n}}\right)\varepsilon_{t-2}\Rightarrow \Sigma_\eta D_{2,0}', \tag{55}$$

and

$$\frac{1}{n}\sum_t \eta_{t-2}\varepsilon_{t-1}=\frac{1}{n}\sum_t \eta_{t-2}\sum_{j=0}^{\infty}\eta_{t-1-j}'D_{2,j}'\varepsilon_{t-1}\Rightarrow \Sigma_\eta D_{2,1}'. \tag{56}$$

Combining results from (42), (54), (55), and (56) gives

$$\frac{1}{n}D_{2,2}\sum_{t}\eta_{t-2}Y_{t-1}\Rightarrow D_{2,2}\left(\int_{0}^{1}G_{a,c}\left(r\right)dB_{\eta}\left(r\right)+\Sigma_{\eta}\sum_{j=0}^{1}D_{1,j}^{\prime}a\int_{0}^{1}G_{a,c}\left(r\right)dr+\Sigma_{\eta}\sum_{j=0}^{1}D_{2,j}^{\prime}\right) \tag{57}$$

For $n^{-1}D_{2,j}\sum_{t}\eta_{t-j}Y_{t-1}$, $j\geq 3$, the pattern is similar to the one eatablished above. Under Assumption 1, the D's satisfy the one-summability condition, so that, by the Dominated Convergence Theorem, for any real r and any $m>0$,

$$\lim_{m\to\infty}E\left(\exp\left(r\frac{1}{n}\sum_{t=2}^{n}\sum_{j=0}^{m}D_{2,j}\eta_{t-j}Y_{t-1}\right)\right)=E\exp\left(r\frac{1}{n}\sum_{t=2}^{n}\sum_{j=0}^{\infty}D_{2,j}\eta_{t-j}Y_{t-1}\right).$$

In view of (36), (41) and (57) then, we have

$$\frac{1}{n}\sum_{t=2}^{n}\varepsilon_{t}Y_{t-1}=\frac{1}{n}\sum_{t=2}^{n}\sum_{j=0}^{\infty}D_{2,j}\eta_{t-j}Y_{t-1}\Rightarrow\sum_{m=0}^{\infty}D_{2,m}\int_{0}^{1}G_{a,c}\left(r\right)dB_{\eta}\left(r\right)$$
$$+\sum_{m=1}^{\infty}D_{2,m}\left(\Sigma_{\eta}\sum_{j=0}^{m-1}D_{1,m}^{\prime}a\int_{0}^{1}G_{a,c}\left(r\right)dr+\Sigma_{\eta}\sum_{j=0}^{m-1}D_{2,j}^{\prime}\right).$$

Now, $\sum_{m=0}^{\infty}D_{2,m}=D_{2}\left(1\right)$ and B_{η} has variance matrix Σ_{η}, so that $D_{2}\left(1\right)B_{\eta}\left(r\right)=B_{\varepsilon}\left(r\right)$ has variance matrix $D_{2}\left(1\right)\Sigma_{\eta}D_{2}\left(1\right)^{\prime}$ and

$$\left(\sum_{m=0}^{\infty}D_{2,m}\right)\int_{0}^{1}G_{a,c}\left(r\right)dB_{\eta}\left(r\right)=\int_{0}^{1}G_{a,c}\left(r\right)dB_{\varepsilon}\left(r\right).$$

Further, recalling that $\Lambda_{u\varepsilon}=\sum_{h=1}^{\infty}\mathbb{E}\left(u_{0}\varepsilon_{h}\right)$, we obtain

$$\begin{aligned}
&\Lambda_{u\varepsilon}^{\prime}a=\mathbb{E}\left\{\left(D_{1,0}\eta_{t}+D_{1,1}\eta_{t-1}+\cdots\right)\left[\left(D_{2,0}\eta_{t+1}+D_{2,1}\eta_{t}+\cdots\right)+\left(D_{2,0}\eta_{t+2}+D_{2,1}\eta_{t+1}+\cdots\right)\right]\right\}^{\prime}a\\
&=\left\{D_{1,0}\Sigma_{\eta}\left(D_{2,1}+D_{2,2}+\cdots\right)^{\prime}+D_{1,1}\Sigma_{\eta}\left(D_{2,2}+D_{2,3}+\cdots\right)^{\prime}+\cdots\right\}^{\prime}a\\
&=\left\{D_{1,0}\Sigma_{\eta}D_{2,1}^{\prime}+\left(D_{1,0}+D_{1,1}\right)\Sigma_{\eta}D_{2,2}^{\prime}\cdots\right\}^{\prime}a\\
&=\left\{D_{2,1}\Sigma_{\eta}D_{1,0}^{\prime}+D_{2,2}\Sigma_{\eta}\left(D_{1,0}+D_{1,1}\right)^{\prime}\cdots\right\}a\\
&=\sum_{j=1}^{\infty}D_{2,j}\Sigma_{\eta}\left(\sum_{k=0}^{j-1}D_{1,k}^{\prime}\right)a. \qquad (58)
\end{aligned}$$

Similarly, $\lambda_{\varepsilon\varepsilon} = \mathbb{E}\sum_{h=1}^{\infty} \varepsilon_t \varepsilon_{t+h}$ is

$$\begin{aligned}
& \mathbb{E}\left\{ \left(D_{2,0}\eta_t + D_{2,1}\eta_{t-1} + \cdots\right)\left[\left(D_{2,0}\eta_{t+1} + D_{2,1}\eta_t + \cdots\right) + \left(D_{2,0}\eta_{t+2} + D_{2,1}\eta_{t+1} + \cdots\right) + \cdots\right]'\right\} \\
= \; & D_{2,0}\Sigma_\eta \left(D_{2,1} + D_{2,2} + \cdots\right)' + D_{2,1}\Sigma_\eta \left(D_{2,2} + D_{2,3} + \cdots\right)' + \cdots \\
= \; & D_{2,0}\Sigma_\eta D_{2,1}{}' + \left(D_{2,0} + D_{2,1}\right)\Sigma_\eta D_{2,2}' + \cdots \\
= \; & D_{2,1}\Sigma_\eta D_{2,0}' + D_{2,2}\Sigma_\eta \left(D_{2,0} + D_{2,1}\right)' + \cdots = \sum_{j=1}^{\infty} D_{2,j}\Sigma_\eta \left(\sum_{k=0}^{j-1} D_{2,k}'\right). \qquad (59)
\end{aligned}$$

This implies the result given in (13) and the Lemma is established. ■

Lemma 15 *For the model (4), under Assumption 1,*

$$\frac{1}{n^2}\sum_t u_t u_t' Y_{t-1}^2 \Rightarrow \Sigma_u \int_0^1 G_{a,c}^2(r)\,dr.$$

Proof of Lemma 15. The result follows immediately from

$$\frac{1}{n^2}\sum_t u_t u_t' Y_{t-1}^2 = \Sigma_u \frac{1}{n}\sum_t \left(\frac{Y_{t-1}}{\sqrt{n}}\right)^2 + \frac{1}{\sqrt{n}}\sum_t \left(\frac{u_t u_t' - \Sigma_u}{\sqrt{n}}\right)\left(\frac{Y_{t-1}}{\sqrt{n}}\right)^2 \Rightarrow \Sigma_u \int_0^1 G_{a,c}^2(r)\,dr.$$

■

Lemma 16 *For the model (4), under Assumption 1,*

$$\frac{1}{n^{3/2}}\sum_t u_t Y_{t-1}^2 \Rightarrow \int_0^1 G_{a,c}^2(r)\,dB_u(r) + 2\left(\Lambda_{uu}' a \int_0^1 G_{a,c}^2(r)\,dr + \Lambda_{u\varepsilon}\int_0^1 G_{a,c}(r)\,dr\right).$$

Proof of Lemma 16. We have

$$\frac{1}{n^{3/2}}\sum_t u_t Y_{t-1}^2 = \frac{1}{n^{3/2}}\sum_{t=2}^{n}\sum_{j=0}^{\infty} D_{1,j}\eta_{t-j} Y_{t-1}^2.$$

For $j = 0$,

$$\frac{1}{n^{3/2}}\sum_{t=2}^{n} D_{1,0}\eta_t Y_{t-1}^2 \Rightarrow D_{1,0}\int_0^1 G_{a,c}^2(r)\,dB_\eta(r).$$

For $j = 1$,

$$\begin{aligned}
& \frac{1}{n^{3/2}}\sum_{t=2}^{n} D_{1,1}\eta_{t-1} Y_{t-1}^2 = \frac{1}{n^{3/2}}\sum_{t=2}^{n} D_{1,1}\eta_{t-1}\left(\exp\left(\frac{c}{n} + \frac{a' u_{t-1}}{\sqrt{n}}\right) Y_{t-2} + \varepsilon_{t-1}\right)^2 \\
= \; & D_{1,1}\int_0^1 G_{a,c}^2(r)\,dB_\eta(r) + \frac{1}{n^{3/2}}\sum_{t=2}^{n} D_{1,1}\eta_{t-1}\frac{2a' u_{t-1}}{\sqrt{n}} Y_{t-2}^2 + \frac{2}{n^{3/2}}\sum_{t=2}^{n} D_{1,1}\eta_{t-1} Y_{t-2}\varepsilon_{t-1} + o_p(1) \\
\Rightarrow \; & D_{1,1}\int_0^1 G_{a,c}^2(r)\,dB_\eta(r) + 2D_{1,1}\Sigma_\eta D_{1,0}' a \int_0^1 G_{a,c}^2(r)\,dr + 2D_{1,1}\Sigma_\eta D_{2,0}\int_0^1 G_{a,c}(r)\,dr.
\end{aligned}$$

For $j=2$,

$$\begin{aligned}
&\frac{1}{n^{3/2}}\sum_{t=2}^{n}D_{1,2}\eta_{t-2}Y_{t-1}^{2}\\
&= \frac{1}{n^{3/2}}\sum_{t=2}^{n}D_{1,2}\eta_{t-2}\left(\exp\left(\frac{2c}{n}+\frac{a'\left(u_{t-1}+u_{t-2}\right)}{\sqrt{n}}\right)Y_{t-3}+\exp\left(\frac{c}{n}+\frac{a'u_{t-1}}{\sqrt{n}}\right)\varepsilon_{t-2}+\varepsilon_{t-1}\right)^{2}
\end{aligned}$$

$$\begin{aligned}
&= D_{1,2}\int_{0}^{1}G_{a,c}^{2}\left(r\right)dB_{\eta}\left(r\right)+\frac{1}{n^{3/2}}\sum_{t=2}^{n}D_{1,2}\eta_{t-2}\frac{2a'\left(u_{t-1}+u_{t-2}\right)}{\sqrt{n}}Y_{t-3}^{2}\\
&\quad+\frac{2}{n^{3/2}}\sum_{t=2}^{n}D_{1,2}\eta_{t-2}Y_{t-3}\left(\varepsilon_{t-2}+\varepsilon_{t-1}\right)+o_{p}\left(1\right)\\
&\Rightarrow D_{1,2}\int_{0}^{1}G_{a,c}^{2}\left(r\right)dB_{\eta}\left(r\right)+2D_{1,2}\Sigma_{\eta}\left(D_{1,0}+D_{1,1}\right)'a\int_{0}^{1}G_{a,c}^{2}\left(r\right)dr+2D_{1,2}\Sigma_{\eta}\left(D_{2,0}+D_{2,1}\right)'\int_{0}^{1}G_{a,c}\left(r\right)dr.
\end{aligned}$$

Continuing this scheme and using summability, we deduce that

$$\frac{1}{n^{3/2}}\sum_{t}u_{t}Y_{t-1}^{2}\Rightarrow\sum_{j=0}^{\infty}D_{1,j}\int_{0}^{1}G_{a,c}^{2}\left(r\right)dB_{\eta}\left(r\right)+2\sum_{j=1}^{\infty}D_{1,j}\Sigma_{\eta}\left(\sum_{k=0}^{j-1}D_{1,k}'a\int_{0}^{1}G_{a,c}^{2}\left(r\right)dr+\sum_{k=0}^{j-1}D_{2,k}'\int_{0}^{1}G_{a,c}\left(r\right)dr\right).$$

and the proof of the Lemma is completed by using (58) and (59). ■

Lemma 17 *For the model (4), under Assumption 1,*

$$\begin{aligned}
\sum_{t=2}^{n}u_{t}\varepsilon_{t}Y_{t-1} &= n^{3/2}\Sigma_{u\varepsilon}\int_{0}^{1}G_{a,c}\left(r\right)dr+n\left\{\sum_{j=0}^{\infty}\left(D_{2,j}\otimes D_{1,j}\right)H\int_{0}^{1}G_{a,c}\left(r\right)d\xi\left(r\right)\right.\\
&\quad+\sum_{j=1}^{\infty}\left(D_{2,j}\otimes D_{1,j}\right)M_{3}\left(\left(\sum_{i=0}^{j-1}D_{1,i}\right)'a\int_{0}^{1}G_{a,c}\left(r\right)dr+\left(\sum_{i=0}^{j-1}D_{2,i}\right)'\right)\\
&\quad\left.+\sum_{j\neq k}\left(D_{2,k}\otimes D_{1,j}\right)\int_{0}^{1}G_{a,c}\left(r\right)d\zeta\left(r\right)\right\}+o_{p}\left(n\right).
\end{aligned}$$

Proof of Lemma 17. The proof is similar to that of Lemma 8 of Lieberman and Phillips (2018) and is omitted.

Lemma 18 *For the model (25) where $K=1$, $\Sigma_{u\varepsilon}=0$, u_t and ε_t are iid, and with the filtration $\mathcal{F}_r=\sigma\left\{\left(B_u\left(s\right),B_\varepsilon\left(s\right)\right),0\leq s\leq r\right\}$ the instantaneous kurtosis measure is*

$$\kappa_{b,c}\left(r\right)=\frac{\mathbb{E}\left(\mathbb{E}\left[\left(dG_{a,c}\left(r\right)\right)^{4}|\mathcal{F}_{r}\right]\right)}{\left\{\mathbb{E}\left(\mathbb{E}\left[\left(dG_{a,c}\left(r\right)\right)^{2}|\mathcal{F}_{r}\right]\right)\right\}^{2}}=3+\frac{3b^{2}\left[\mathbb{E}\left(G_{a,c}^{4}\left(r\right)\right)-\left(\mathbb{E}\left(G_{a,c}^{2}\left(r\right)\right)\right)^{2}\right]}{b^{2}\left(\mathbb{E}\left(G_{a,c}^{2}\left(r\right)\right)\right)^{2}+\sigma_{\varepsilon}^{4}+2b\sigma_{\varepsilon}^{2}\mathbb{E}\left(G_{a,c}^{2}\left(r\right)\right)} \tag{60}$$

Proof of Lemma 18. The process increments $dG_{a,c}(r)$ at r satisfy (25)

$$dG_{a,c}(r) = aG_{a,c}(r)\,dB_u(r) + dB_\varepsilon(r) + \left(c + \frac{b}{2}\right) G_{a,c}(r)\,dr, \tag{61}$$

where $b = a^2\sigma_u^2$. In the arguments that follow we evaluate conditional expectations and expectations of the second and fourth powers of the stochastic differential $dG_{a,c}(r)$ in order to calculate the instantaneous kurtosis measure $\kappa_{b,c}(r)$ at r. The components involve powers of the stochastic differentials $dB_u(r)$ and $dB_\varepsilon(r)$ and we use the fact that $(dW(r))^2 = d[W]_r = dr$ *a.s.* and is a nonstochastic differential, where $[W]_r = \int_0^r (dW)^2 = r$ is the quadratic variation of the standard Brownian motion W. The resulting expressions therefore involve deterministic differentials dr and these are treated in the usual manner according to their respective orders of magnitude. When these deterministic differentials are scaled by stochastic processes then O_p rather than O notation is used for the respective quantities in the following derivations. We have

$$\mathbb{E}\left[(dG_{a,c}(r))^4 |\mathcal{F}_r\right] = \mathbb{E}\left[aG_{a,c}(r)\,dB_u(r) + dB_\varepsilon(r) + \left(c+\frac{b}{2}\right) G_{a,c}(r)\,dr|\mathcal{F}_r\right]^4$$

$$\begin{aligned}
&= \mathbb{E}\left[(aG_{a,c}(r)\,dB_u(r) + dB_\varepsilon(r))^4 |\mathcal{F}_r\right] + 4\left(c+\frac{b}{2}\right)\mathbb{E}\left[(aG_{a,c}(r)\,dB_u(r) + dB_\varepsilon(r))^3 G_{a,c}(r)\,|\mathcal{F}_r\right] dr \\
&+ 6\left(c+\frac{b}{2}\right)^2 \mathbb{E}\left[(aG_{a,c}(r)\,dB_u(r) + dB_\varepsilon(r))^2 G_{a,c}(r)^2 |\mathcal{F}_r\right](dr)^2 \\
&+ 4\left(c+\frac{b}{2}\right)^3 \mathbb{E}\left[(aG_{a,c}(r)\,dB_u(r) + dB_\varepsilon(r))\,G_{a,c}(r)^3 |\mathcal{F}_r\right](dr)^3 + \left(c+\frac{b}{2}\right)^4 \mathbb{E}\left[G_{a,c}(r)^4 |\mathcal{F}_r\right](dr)^4 \\
&= \left[3b^2 G_{a,c}(r)^4 + 6b\sigma_\varepsilon^2 G_{a,c}(r)^2 + 3\sigma_\varepsilon^4\right](dr)^2 + 6\left(c+\frac{b}{2}\right)^2\left[bG_{a,c}(r)^4 + G_{a,c}(r)^2\sigma_\varepsilon^2\right](dr)^3 + O_p\left((dr)^4\right)
\end{aligned} \tag{62}$$

since $\mathbb{E}\left[(aG_{a,c}(r)\,dB_u(r) + dB_\varepsilon(r))^3 G_{a,c}(r)\,|\mathcal{F}_r\right] dr = 0$,

$$\begin{aligned}
\mathbb{E}\left\{\mathbb{E}\left[(aG_{a,c}(r)\,dB_u(r) + dB_\varepsilon(r))^4 |\mathcal{F}_r\right]\right\} &= 3a^4\sigma_u^4\mathbb{E}\left(G_{a,c}^4(r)\right) + 6a^2\sigma_u^2\sigma_\varepsilon^2\mathbb{E}\left(G_{a,c}^2(r)\right) + 3\sigma_\varepsilon^4 \\
&= 3b^2\mathbb{E}\left(G_{a,c}^4(r)\right) + 6b\sigma_\varepsilon^2\mathbb{E}\left(G_{a,c}^2(r)\right) + 3\sigma_\varepsilon^4,
\end{aligned} \tag{63}$$

and

$$\mathbb{E}\left[(aG_{a,c}(r)\,dB_u(r) + dB_\varepsilon(r))^2 G_{a,c}(r)^2 |\mathcal{F}_r\right] = \left[a^2\sigma_u^2 G_{a,c}^4(r) + \sigma_\varepsilon^2 G_{a,c}(r)^2\right] dr.$$

Similarly

$$\mathbb{E}\left[\left(dG_{a,c}\left(r\right)\right)^2|\mathcal{F}_r\right] = \mathbb{E}\left[aG_{a,c}\left(r\right)dB_u\left(r\right) + dB_\varepsilon\left(r\right) + \left(c+\frac{b}{2}\right)G_{a,c}\left(r\right)dr|\mathcal{F}_r\right]^2$$

$$\begin{aligned}
&= \mathbb{E}\left[\left(aG_{a,c}\left(r\right)dB_u\left(r\right) + dB_\varepsilon\left(r\right)\right)^2|\mathcal{F}_r\right] + \left(c+\frac{b}{2}\right)^2\mathbb{E}\left[G_{a,c}\left(r\right)^2|\mathcal{F}_r\right]\left(dr\right)^2 \\
&= \mathbb{E}\left[\left(a^2\sigma_u^2G_{a,c}\left(r\right)^2 + \sigma_\varepsilon^2\right)|\mathcal{F}_r\right]dr + \left(c+\frac{b}{2}\right)^2\mathbb{E}\left[G_{a,c}\left(r\right)^2|\mathcal{F}_r\right]\left(dr\right)^2 \\
&= \left[bG_{a,c}\left(r\right)^2 + \sigma_\varepsilon^2\right]dr + \left(c+\frac{b}{2}\right)^2G_{a,c}\left(r\right)^2\left(dr\right)^2. \qquad (64)
\end{aligned}$$

Using (62) - (64) and treating terms of order $o\left(\left(dr\right)^2\right)$ as zero gives

$$\begin{aligned}
\kappa_{b,c}\left(r\right) &= \frac{\mathbb{E}\left(\mathbb{E}\left[\left(aG_{a,c}\left(r\right)dB_u\left(r\right) + dB_\varepsilon\left(r\right)\right)^4|\mathcal{F}_r\right]\right) + o\left(\left(dr\right)^2\right)}{\left\{\mathbb{E}\left(\mathbb{E}\left[\left(aG_{a,c}\left(r\right)dB_u\left(r\right) + dB_\varepsilon\left(r\right)\right)^2|\mathcal{F}_r\right]\right) + o\left(\left(dr\right)^2\right)\right\}^2} \\
&= \frac{3b^2\mathbb{E}\left(G_{a,c}^4\left(r\right)\right) + 6b\sigma_\varepsilon^2\mathbb{E}\left(G_{a,c}^2\left(r\right)\right) + 3\sigma_\varepsilon^4}{\left(b\mathbb{E}\left(G_{a,c}^2\left(r\right)\right) + \sigma_\varepsilon^2\right)^2}
\end{aligned}$$

$$\begin{aligned}
&= \frac{3b^2\mathbb{E}\left(G_{a,c}^4\left(r\right)\right) + 6b\sigma_\varepsilon^2\mathbb{E}\left(G_{a,c}^2\left(r\right)\right) + 3\sigma_\varepsilon^4}{b^2\left(\mathbb{E}\left(G_{a,c}^2\left(r\right)\right)\right)^2 + \sigma_\varepsilon^4 + 2b\sigma_\varepsilon^2\mathbb{E}\left(G_{a,c}^2\left(r\right)\right)} \\
&= 3 + \frac{3b^2\left[\mathbb{E}\left(G_{a,c}^4\left(r\right)\right) - \left(\mathbb{E}\left(G_{a,c}^2\left(r\right)\right)\right)^2\right]}{b^2\left(\mathbb{E}\left(G_{a,c}^2\left(r\right)\right)\right)^2 + \sigma_\varepsilon^4 + 2b\sigma_\varepsilon^2\mathbb{E}\left(G_{a,c}^2\left(r\right)\right)},
\end{aligned}$$

as stated. For large b, note that

$$\begin{aligned}
\mathbb{E}\left(G_{a,c}^4\left(r\right)\right) &= \frac{3\sigma_\varepsilon^4e^{4(c+2b)r}}{c+b}\left(\frac{1}{2\left(c+3b\right)} - \frac{1}{4\left(c+2b\right)}\right) = \frac{3\sigma_\varepsilon^4e^{4(c+2b)r}}{4\left(c+b\right)}\frac{2\left(c+2b\right)-\left(c+3b\right)}{\left(c+3b\right)\left(c+2b\right)} \\
&= \frac{3\sigma_\varepsilon^4e^{4(c+2b)r}}{4\left(c+b\right)}\frac{c+b}{\left(c+3b\right)\left(c+2b\right)} = \frac{3\sigma_\varepsilon^4e^{4(c+2b)r}}{4\left(c+3b\right)\left(c+2b\right)},
\end{aligned}$$

and $\mathbb{E}\left(G_{a,c}^2\left(r\right)\right) = \sigma_\varepsilon^2\frac{e^{2(c+b)r}-1}{2(c+b)} \sim \sigma_\varepsilon^2\frac{e^{2(c+b)r}}{2(c+b)}$. Hence, as $b\to\infty$

$$\begin{aligned}
\kappa_{b,c}\left(r\right) &= \frac{3b^2\mathbb{E}\left(G_{a,c}^4\left(r\right)\right) + 6b\sigma_\varepsilon^2\mathbb{E}\left(G_{a,c}^2\left(r\right)\right) + 3\sigma_\varepsilon^4}{\left(b\mathbb{E}\left(G_{a,c}^2\left(r\right)\right) + \sigma_\varepsilon^2\right)^2} \sim \frac{3b^2\mathbb{E}\left(G_{a,c}^4\left(r\right)\right)}{b^2\mathbb{E}\left(G_{a,c}^2\left(r\right)\right)^2} \sim \frac{3\frac{3\sigma_\varepsilon^4e^{4(c+2b)r}}{4(c+3b)(c+2b)}}{\left(\sigma_\varepsilon^2\frac{e^{2(c+b)r}}{2(c+b)}\right)^2} \\
&= 9\frac{e^{4br}\left(c+b\right)^2}{\left(c+3b\right)\left(c+2b\right)} \sim \frac{9}{6}e^{4br},
\end{aligned}$$

and kurtosis of the process increments $dG_{a,c}(r)$ grows exponentially with b irrespective of the fixed value of c.

10.2 Proofs of the Main Results

Proof of Lemma 1. By repeated substitution, we obtain

$$Y_t = \sum_{j=1}^{t} \exp\left(\frac{(t-j)\,c}{n} + \frac{a' \sum_{i=j+1}^{t} u_i}{\sqrt{n}} \right) \varepsilon_j, \ t \geq 2. \tag{65}$$

Therefore, setting $t = \lfloor nr \rfloor$ and using a strong approximation version of the invariance principle (7), we have

$$\begin{aligned}
\frac{Y_{t=\lfloor nr \rfloor}}{\sqrt{n}} &= e^{rc + a' B_u(r) + o_p(1)} \sum_{j=1}^{\lfloor nr \rfloor} \exp\left(-\frac{jc}{n} - \frac{a' \sum_{i=1}^{j} u_i}{\sqrt{n}} \right) \frac{\varepsilon_j}{\sqrt{n}} \\
&= e^{rc + a' B_u(r)} \sum_{j=1}^{\lfloor nr \rfloor} \exp\left(-\frac{(j-1)\,c}{n} - \frac{a' \sum_{i=1}^{j-1} u_i}{\sqrt{n}} \right) \exp\left(-\frac{a' u_j}{\sqrt{n}} \right) \frac{\varepsilon_j}{\sqrt{n}} + o_p(1) \\
&= e^{rc + a' B_u(r)} \sum_{j=1}^{\lfloor nr \rfloor} \exp\left(-\frac{(j-1)\,c}{n} - \frac{a' \sum_{i=1}^{j-1} u_i}{\sqrt{n}} \right) \left(1 - \frac{a' u_j}{\sqrt{n}} + \frac{(a' u_j)^2}{2n} e^{\tilde{x}_j} \right) \frac{\varepsilon_j}{\sqrt{n}} + o_p(1) \\
&= e^{rc + a' B_u(r)} \left\{ \sum_{j=1}^{\lfloor nr \rfloor} e^{-\frac{(j-1)c}{n} - \frac{a' \sum_{i=1}^{j-1} u_i}{\sqrt{n}}} \frac{\varepsilon_j}{\sqrt{n}} - \sum_{j=1}^{\lfloor nr \rfloor} e^{-\frac{(j-1)c}{n} - a' B_u\left(\frac{j-1}{n}\right)} \left(\frac{a' \Sigma_{u\varepsilon}}{n} + \frac{a' (u_j \varepsilon_j - \Sigma_{u\varepsilon})}{n} \right) \right\} + o_p(1).
\end{aligned} \tag{66}$$

The error order in (66) holds uniformly over $j = 1, ..., n$ because the components u_{ij} of u_j satisfy

$$\begin{aligned}
\mathbb{P}\{\max_{j \leq n} \left| \frac{u_{ij}}{\sqrt{n}} \right| > \epsilon\} &= \mathbb{P}\{\frac{1}{n} \sum_{t=1}^{n} u_{it}^2 \mathbf{1}\left\{ |u_{it}| > \sqrt{n}\epsilon \right\} > \epsilon^2\} \\
\leq \ \frac{\sum_{t=1}^{n} \mathbb{E}\left[u_{it}^2 \mathbf{1}\left\{ |u_{it}| > \sqrt{n}\epsilon \right\} \right]}{n\epsilon^2} &= \frac{1}{\epsilon^2} \mathbb{E}\left[u_{it}^2 \mathbf{1}\left\{ |u_{it}| > \sqrt{n}\epsilon \right\} \right] \to 0
\end{aligned} \tag{67}$$

so that $\max_{j \leq n} \frac{|u_{ij}|}{\sqrt{n}} = o_p(1)$. Then, $\max_{j \leq n} |\tilde{x}_j| \leq \max_{j \leq n} \left| \frac{a' u_j}{\sqrt{n}} \right| = o_p(1)$. Similar arguments may be used in the arguments that follow in order to establish uniform error bounds as may be needed, although details are not always provided.

Now let $e^{-\frac{(j-1)c}{n} - \frac{a' \sum_{i=1}^{j-1} u_j}{\sqrt{n}}} =: f\left(-\frac{(j-1)c}{n} - a' \frac{1}{\sqrt{n}} \sum_{i=1}^{j-1} u_j \right)$. Then,

$$\frac{\partial}{\partial X} f\left(-\frac{(j-1)\,c}{n} - a' X \right)_{X = \frac{1}{\sqrt{n}} \sum_{i=1}^{j-1} u_i} = -a f\left(-\frac{(j-1)\,c}{n} - a' \frac{1}{\sqrt{n}} \sum_{i=1}^{j-1} u_i \right)$$

and by Ibragimov and Phillips (2008; equation (4.9)) we obtain the following sample covariance limit

$$\sum_{j=1}^{\lfloor nr \rfloor} \exp\left(-\frac{(j-1)\,c}{n} - \frac{a'\sum_{i=1}^{j-1} u_i}{\sqrt{n}}\right)\frac{\varepsilon_j}{\sqrt{n}} \Rightarrow -a'\Lambda_{u\varepsilon}\int_0^r e^{-pc-a'B_u(p)}dp + \int_0^r e^{-pc-a'B_u(p)}dB_\varepsilon(p)\,. \quad (68)$$

Furthermore,

$$\sum_{j=1}^{\lfloor nr \rfloor} \exp\left(-\frac{(j-1)\,c}{n} - a'B_u\left(\frac{j-1}{n}\right)\right)\frac{a'\Sigma_{u\varepsilon}}{n} = a'\Sigma_{u\varepsilon}\int_0^r e^{-pc-a'B_u(p)}dp + o_p(1)\,. \quad (69)$$

The last term in braces in (66) is

$$\sum_{j=1}^{\lfloor nr \rfloor} e^{-\frac{(j-1)c}{n}-a'B_u\left(\frac{j-1}{n}\right)}\frac{a'\left(u_j\varepsilon_j - \Sigma_{u\varepsilon}\right)}{n} = \frac{1}{\sqrt{n}}\sum_{j=1}^{\lfloor nr \rfloor} e^{-\frac{(j-1)c}{n}-a'B_u\left(\frac{j-1}{n}\right)}a'\left(dB_{u\varepsilon}\left(\frac{j}{n}\right) + o_p\left(\max_j\left(dB_{u\varepsilon}\left(\frac{j}{n}\right)\right)\right)\right). \quad (70)$$

As the differentials $dB_{u\varepsilon}\left(\frac{j}{n}\right)$ are independent Gaussian variables, each with a variance which is $O\left(n^{-1}\right)$, $\max_{j\le n} dB_{u\varepsilon}\left(\frac{j}{n}\right) = O_p\left(\sqrt{\frac{\log n}{n}}\right)$, see for instance, equation (1.7.2) of Leadbetter *et. al.* (1982). It follows that (70) is $O_p\left(n^{-1/2}\right)$ and therefore, from (66), (68) and (69), that

$$\begin{aligned}\frac{Y_t}{\sqrt{n}} &\Rightarrow e^{rc+a'B_u(r)}\left\{-a'\Lambda_{u\varepsilon}\int_0^r e^{-pc-a'B_u(p)}dp + \int_0^r e^{-pc-a'B_u(p)}dB_\varepsilon(p) - a'\Sigma_{u\varepsilon}\int_0^r e^{-pc-aB_u(p)}dp\right\} \\ &= e^{rc+a'B_u(r)}\left\{\int_0^r e^{-pc-a'B_u(p)}dB_\varepsilon(p) - a'\Delta_{u\varepsilon}\int_0^r e^{-pc-aB_u(p)}dp\right\},\end{aligned}$$

which is the stated result. ■

Proof of Theorem 4. Existence and uniqueness of the estimate $\hat{c}_n$ follows from the convexity of the NLLS objective function and monotonicity of dependence of the function $\beta_{nt} = \beta_{nt}(c)$ as a function of c, given $a'u_t$. The method of proof then relies on asymptotic expansions of the first order conditions. The NLLS $\hat{c}_n$ of c, given known a, is defined as the solution to the equation

$$\sum_t \left(Y_t - \beta_{nt}(\hat{c}_n, a)\,Y_{t-1}\right)\dot{\beta}_{nt}(\hat{c}_n, a)\,Y_{t-1} = 0, \quad (71)$$

where $\dot{\beta}_{nt}(c,a) = \frac{\partial \beta_{nt}(c,a)}{\partial c} = \frac{1}{n}\beta_{nt}(c,a)\,.$ The solution to (71) is equivalent to the solution of

$$\sum_t Y_t\beta_{nt}(\hat{c}_n, a)\,Y_{t-1} = \sum_t \beta_{nt}^2(\hat{c}_n, a)\,Y_{t-1}^2.$$

or

$$\sum_t \left(\beta_{nt}(c,a)\,Y_{t-1} + \varepsilon_t\right)\beta_{nt}(\hat{c}_n, a)\,Y_{t-1} = \sum_t \beta_{nt}(2\hat{c}_n, 2a)\,Y_{t-1}^2.$$

Rearranging the last equation, we seek a solution to

$$\sum_t e^{2a'u_t/\sqrt{n}} \left[e^{2\hat{c}_n/n} - e^{(c+\hat{c}_n)/n} \right] Y_{t-1}^2 = \sum_t e^{a'u_t/\sqrt{n}} e^{\hat{c}_n/n} \varepsilon_t Y_{t-1}. \tag{72}$$

Expanding the left side of (72) we get

$$\begin{aligned}
\sum_t e^{2a'u_t/\sqrt{n}} \left[\sum_{j=1}^{\infty} \frac{(2\hat{c}_n)^j - (c+\hat{c}_n)^j}{n^j j!} \right] Y_{t-1}^2 &= \sum_t e^{2a'u_t/\sqrt{n}} \left[\sum_{j=1}^{\infty} \frac{(2\hat{c}_n)^j - (2\hat{c}_n + [c-\hat{c}_n])^j}{n^j j!} \right] Y_{t-1}^2 \\
= \sum_t e^{2a'u_t/\sqrt{n}} \left[\sum_{j=1}^{\infty} \frac{(2\hat{c}_n)^j - \sum_{k=0}^{j} \binom{j}{k} (2\hat{c}_n)^k [c-\hat{c}_n]^{j-k}}{n^j j!} \right] Y_{t-1}^2 & \\
= \sum_t e^{2a'u_t/\sqrt{n}} \left[\sum_{j=1}^{\infty} \frac{-\sum_{k=0}^{j-1} \binom{j}{k} (2\hat{c}_n)^k [c-\hat{c}_n]^{j-k}}{n^j j!} \right] Y_{t-1}^2 & \\
= \sum_t e^{2a'u_t/\sqrt{n}} \left[(\hat{c}_n - c) \sum_{j=1}^{\infty} \frac{\sum_{k=0}^{j-1} \binom{j}{k} (2\hat{c}_n)^k [c-\hat{c}_n]^{j-1-k}}{n^j j!} \right] Y_{t-1}^2. &
\end{aligned}$$

At the true value of c, the objective function $Q_n^a(c) = n^{-1}(Y_t - \beta_{nt}(c,a) Y_{t-1})^2$ converges in probability to σ_ε^2 and therefore the only term in the square brackets which contributes asymptotically is the first order term, $(\hat{c}_n - c)/n$. The leading term on the left side of (72) is therefore

$$\sum_t \left(1 + \frac{2a'u_t}{\sqrt{n}} + \frac{2a'\Sigma_u a}{n} + \frac{2a'(u_t u_t' - \Sigma_u) a}{n} + o_p\left(\frac{1}{n}\right) \right) \left[\frac{\hat{c}_n - c}{n} + O_p\left(n^{-2}\right) \right] Y_{t-1}^2.$$

Upon scaling by $1/n$, we have the following asymptotic form:

$$\begin{aligned}
& (\hat{c}_n - c) \left\{ \frac{1}{n} \sum_t \left(\frac{Y_{t-1}}{\sqrt{n}} \right)^2 + \frac{2}{n} \sum_t \frac{a'u_t}{\sqrt{n}} \left(\frac{Y_{t-1}}{\sqrt{n}} \right)^2 + \frac{2}{n} \sum_t a' \left(\frac{u_t u_t' - \Sigma_u}{\sqrt{n}} \right) a \left(\frac{Y_{t-1}}{\sqrt{n}} \right)^2 \right. \\
& \left. + \frac{2a'\Sigma_u a}{n^2} \sum_t \left(\frac{Y_{t-1}}{\sqrt{n}} \right)^2 + o_p\left(\frac{1}{n}\right) \right\} \\
= \; & (\hat{c}_n - c) \left\{ \frac{1}{n} \sum_t \left(\frac{Y_{t-1}}{\sqrt{n}} \right)^2 + O_p\left(\frac{1}{n}\right) \right\} \sim_a (\hat{c}_n - c) \int_0^1 G_{a,c}^2(r)\, dr,
\end{aligned} \tag{73}$$

where the notation $\sim_a$ means that the difference between the lhs and the rhs of the symbol is $o_p(1)$. Scaling by $1/n$, the dominant term on the right side of (72) is

$$\frac{1}{n} \sum_t \varepsilon_t Y_{t-1} + \frac{1}{n} \sum_t a'u_t \varepsilon_t \frac{Y_{t-1}}{\sqrt{n}},$$

and

$$\frac{1}{n^{3/2}}\sum_t a'u_t\varepsilon_t Y_{t-1} = \frac{a'\Sigma_{u\varepsilon}}{n}\sum_t \frac{Y_{t-1}}{\sqrt{n}} + \frac{1}{\sqrt{n}}\sum_t a'\left(\frac{u_t\varepsilon_t - \Sigma_{u\varepsilon}}{\sqrt{n}}\right)\frac{Y_{t-1}}{\sqrt{n}} \Rightarrow a'\Sigma_{u\varepsilon}\int_0^1 G_{a,c}(r)\,dr.$$

Using Lemma 3, we obtain

$$\frac{1}{n}\sum_t \varepsilon_t Y_{t-1} + \frac{1}{n^{3/2}}\sum_t a'u_t\varepsilon_t Y_{t-1} \Rightarrow \left(\int_0^1 G_{a,c}(r)\,dB_\varepsilon(r) + a'\Lambda_{u\varepsilon}\int_0^1 G_{a,c}(r)\,dr + \lambda_{\varepsilon\varepsilon}\right) + a'\Sigma_{u\varepsilon}\int_0^1 G_{a,c}(r)\,dr$$

$$= \int_0^1 G_{a,c}(r)\,dB_\varepsilon(r) + a'\Delta_{u\varepsilon}\int_0^1 G_{a,c}(r)\,dr + \lambda_{\varepsilon\varepsilon}. \tag{74}$$

The result of the theorem then follows from (73) and (74). ■

Proof of Theorem 5. We maintain the notation (a, c) for the true value of the parameter and denote by (a_1, c_1) any other parameter in the parameter space. The NLLS estimate of (a, c) satisfies the extremum problem

$$(\hat{a}_n, \hat{c}_n) = \arg\min_{(a_1,c_1)\in\Theta}\left\{\sum_{t=1}^n \left(Y_t - \beta_{nt}(a_1,c_1)\,Y_{t-1}\right)^2\right\}. \tag{75}$$

The objective function $\sum_t (Y_t - \beta_{nt}(a_1,c_1)\,Y_{t-1})^2$ is quadratic and hence convex in β_{nt} and the exponential function $\beta_{nt} = \beta_{nt}(a_1,c_1) = e^{\frac{c_1}{n}+\frac{a_1'u_t}{\sqrt{n}}}$ is monotonic and local to unity, which simplifies the development of asymptotics for $(\hat{a}_n, \hat{c}_n)$. We proceed by providing an asymptotic approximation to (75). As shown below the objective function is asymptotically quadratic and convex in the parameters, so that the NLLS estimates exist and are unique for large enough n.

We start by noting that in view of the asymptotic expansion

$$\begin{aligned}\beta_{nt} &= e^{\frac{c_1}{n}+\frac{a_1'u_t}{\sqrt{n}}} = 1 + \left(\frac{c_1}{n} + \frac{a_1'u_t}{\sqrt{n}}\right) + \frac{1}{2}\left(\frac{c_1}{n} + \frac{a_1'u_t}{\sqrt{n}}\right)^2 + O_p\left(\frac{1}{n^{3/2}}\right)\\ &= 1 + \frac{c_1 + \frac{1}{2}(a_1'u_t)^2}{n} + \frac{a_1'u_t}{\sqrt{n}} + O_p\left(\frac{1}{n^{3/2}}\right) =: 1 + \delta_{nt}(a_1,c_1),\end{aligned}$$

the model $Y_t = \beta_{nt}(a,c)\,Y_{t-1} + \varepsilon_t$ has a linear in variables and non-linear in parameters asymptotic representation of the form

$$\begin{aligned}\Delta Y_t &= \delta_{nt}(a,c)\,Y_{t-1} + \varepsilon_t\\ &= \frac{c}{\sqrt{n}}\frac{Y_{t-1}}{\sqrt{n}} + \frac{a'u_tY_{t-1}}{\sqrt{n}} + \frac{1}{2}\frac{(a'u_t)^2}{\sqrt{n}}\frac{Y_{t-1}}{\sqrt{n}} + \varepsilon_t + o_p\left(\frac{1}{\sqrt{n}}\right),\end{aligned}$$

where we use use the fact that $n^{-1/2}Y_{t-1} = O_p(1)$. The approximation holds uniformly over Θ by compactness and uniformly over $t = 1, 2, ..., n$ because $\max_{t\le n}\frac{|u_{it}|}{\sqrt{n}} = o_p(1)$, just as in (67) above. Similar arguments may be used below in order to establish uniform error bounds as may be needed.

Standardizing and centering the objective criterion in (75) we obtain the following asymptotic approximation as $n \to \infty$,

$$\begin{aligned}
R_n(a_1, c_1) &= \frac{1}{n}\sum_t \left(\Delta Y_t - \delta_{nt}(a_1, c_1) Y_{t-1}\right)^2 - \frac{1}{n}\sum_t (\Delta Y_t)^2 \\
&= -\frac{2}{n}\sum_t \delta_{nt}(a_1, c_1) \Delta Y_t Y_{t-1} + \frac{1}{n}\sum_t \delta_{nt}(a_1, c_1)^2 Y_{t-1}^2 \\
&= -\frac{2}{n}\sum_t \delta_{nt}(a_1, c_1) \{\delta_{nt}(a, c) Y_{t-1} + \varepsilon_t\} Y_{t-1} + \frac{1}{n}\sum_t \delta_{nt}(a_1, c_1)^2 Y_{t-1}^2 \\
&= -\frac{2}{n}\sum_t \delta_{nt}(a_1, c_1) \varepsilon_t Y_{t-1} - \frac{2}{n}\sum_t \delta_{nt}(a_1, c_1) \delta_{nt}(a, c) Y_{t-1}^2 + \frac{1}{n}\sum_t \delta_{nt}(a_1, c_1)^2 Y_{t-1}^2.
\end{aligned} \tag{76}$$

Next, substitute

$$\delta_{nt}(a_1, c_1) = \frac{c_1}{n} + \frac{a_1' u_t}{\sqrt{n}} + \frac{1}{2}\frac{(a_1' u_t)^2}{n} + O_p\left(\frac{1}{n^{3/2}}\right), \tag{77}$$

and consider successive terms of (76) as follows:

$$\begin{aligned}
&\frac{2}{n}\sum_t \delta_{nt}(a_1, c_1) \varepsilon_t Y_{t-1} = \frac{2}{n}\sum_t \left[\frac{c_1}{n} + \frac{a_1' u_t}{\sqrt{n}} + \frac{1}{2}\frac{(a_1' u_t)^2}{n}\right] \varepsilon_t Y_{t-1} + O_p\left(\frac{1}{\sqrt{n}}\right) \\
= \quad & \frac{2a_1'}{n}\sum_t u_t \varepsilon_t \frac{Y_{t-1}}{\sqrt{n}} + O_p\left(\frac{1}{n^{1/2}}\right),
\end{aligned}$$

$$\begin{aligned}
&\frac{2}{n}\sum_t \delta_{nt}(a_1, c_1) \delta_{nt}(a, c) Y_{t-1}^2 \\
= \quad & \frac{2}{n}\sum_t \left[\frac{c_1}{n} + \frac{a_1' u_t}{\sqrt{n}} + \frac{1}{2}\frac{(a_1' u_t)^2}{n}\right] \left[\frac{c}{n} + \frac{a' u_t}{\sqrt{n}} + \frac{1}{2}\frac{(a' u_t)^2}{n}\right] Y_{t-1}^2 + O_p\left(\frac{1}{n^{1/2}}\right) \\
= \quad & \frac{2}{n}\sum_t a_1' u_t u_t' a \left(\frac{Y_{t-1}}{\sqrt{n}}\right)^2 + O_p\left(\frac{1}{n^{1/2}}\right),
\end{aligned}$$

and

$$\begin{aligned}
&\frac{1}{n}\sum_t \delta_{nt}(a_1, c_1)^2 Y_{t-1}^2 = \frac{1}{n}\sum_t \left[\frac{c_1}{n} + \frac{a_1' u_t}{\sqrt{n}} + \frac{1}{2}\frac{(a_1' u_t)^2}{n}\right]^2 Y_{t-1}^2 + O_p\left(\frac{1}{n^{1/2}}\right) \\
= \quad & \frac{1}{n}\sum_t a_1' u_t u_t' a_1 \left(\frac{Y_{t-1}}{\sqrt{n}}\right)^2 + O_p\left(\frac{1}{n^{1/2}}\right).
\end{aligned}$$

Then

$$R_n(a_1,c_1) = -\frac{2}{n}\sum_t \delta_{nt}(a_1,c_1)\varepsilon_t Y_{t-1} - \frac{2}{n}\sum_t \delta_{nt}(a_1,c_1)\delta_{nt}(a,c)Y_{t-1}^2 + \frac{1}{n}\sum_t \delta_{nt}(a_1,c_1)^2 Y_{t-1}^2$$

$$= -\frac{2a_1'}{n}\sum_t u_t\varepsilon_t \frac{Y_{t-1}}{\sqrt{n}} - \frac{2}{n}\sum_t a_1' u_t u_t' a\left(\frac{Y_{t-1}}{\sqrt{n}}\right)^2 + \frac{1}{n}\sum_t a_1' u_t u_t' a_1 \left(\frac{Y_{t-1}}{\sqrt{n}}\right)^2 + O_p\left(\frac{1}{n^{1/2}}\right)$$

which is quadratic in a_1 with a unique minimum up to $O_p\left(n^{-1/2}\right)$. It follows that

$$\begin{aligned}\hat{a}_n &= \arg\min_{a_1} R_n(a_1,c_1) = \left(\frac{1}{n}\sum_t u_t u_t'\left(\frac{Y_{t-1}}{\sqrt{n}}\right)^2\right)^{-1}\left\{\frac{1}{n}\sum_t u_t\varepsilon_t\frac{Y_{t-1}}{\sqrt{n}} + \frac{1}{n}\sum_t u_t u_t' a\left(\frac{Y_{t-1}}{\sqrt{n}}\right)^2\right\} + O_p\left(\frac{1}{n^{1/2}}\right)\\ &= a + \left(\frac{1}{n}\sum_t u_t u_t'\left(\frac{Y_{t-1}}{\sqrt{n}}\right)^2\right)^{-1}\left(\frac{1}{n}\sum_t u_t\varepsilon_t\frac{Y_{t-1}}{\sqrt{n}}\right) + O_p\left(\frac{1}{n^{1/2}}\right),\end{aligned}$$

which is independent of c up to this order and as $n \to \infty$. We have

$$\hat{a}_n \Rightarrow a + \left(\Sigma_u \int_0^1 G_{a,c}^2(r)\,dr\right)^{-1}\Sigma_{u\varepsilon}\int_0^1 G_{a,c}(r)\,dr = a + \Sigma_u^{-1}\Sigma_{u\varepsilon}\left(\int_0^1 G_{a,c}^2(r)\,dr\right)^{-1}\int_0^1 G_{a,c}(r)\,dr = a + \varrho_a, \tag{78}$$

where

$$\varrho_a = \Sigma_u^{-1}\Sigma_{u\varepsilon}\left(\int_0^1 G_{a,c}^2\right)^{-1}\int_0^1 G_{a,c}.$$

Note that the limit distribution of $\hat{a}_n$ given in (78) depends on c only through the limit process $G_{a,c}$. It can be shown that to $O_p\left(n^{-1/2}\right)$ the solution to $\hat{a}_n - a$ in (78) is identical to the solution of the exact objective function and so, (16) is established. Continuing, the objective function $\sum\left(Y_t - \beta_t(\hat{a}_n,c_1)Y_{t-1}\right)^2$ is convex in c_1. We solve for $\hat{c}_n$ as

$$\begin{aligned}\hat{c}_n &= \arg\min_{c_1}\sum\left(Y_t - \beta_{nt}(\hat{a}_n,c_1)Y_{t-1}\right)^2\\ &= \arg\min_{c_1}\sum\left(Y_t - e^{\frac{c_1}{n}}e^{\frac{\hat{a}_n' u_t}{\sqrt{n}}}Y_{t-1}\right)^2.\end{aligned}$$

The equation is

$$\begin{aligned}& -\frac{2}{n}\sum\left(Y_t - \beta_{nt}(\hat{a}_n,\hat{c}_n)Y_{t-1}\right)\beta_{nt}(\hat{a}_n,\hat{c}_n)Y_{t-1}\\ =& -\frac{2}{n}\sum\left(\left(\beta_{nt}(a,c) - \beta_{nt}(\hat{a}_n,\hat{c}_n)\right)Y_{t-1} + \varepsilon_t\right)\beta_{nt}(\hat{a}_n,\hat{c}_n)Y_{t-1}\\ =& \frac{2}{n}\sum\left(\left(e^{2\left(\frac{\hat{c}_n}{n} + \frac{\hat{a}_n' u_t}{\sqrt{n}}\right)} - e^{\frac{c+\hat{c}_n}{n} + \frac{(a+\hat{a}_n)' u_t}{\sqrt{n}}}\right)Y_{t-1}^2 - \varepsilon_t e^{\frac{\hat{c}_n}{n} + \frac{\hat{a}_n' u_t}{\sqrt{n}}}Y_{t-1}\right)\\ =& \; 0.\end{aligned}$$

The solution is

$$
\begin{aligned}
& \frac{1}{n}\sum\left(\frac{\hat{c}_n-c}{n}+\frac{(\hat{a}_n-a)'u_t}{\sqrt{n}}+2\left(\frac{\hat{a}_n'u_t}{\sqrt{n}}\right)^2-\frac{1}{2}\left(\frac{(a+\hat{a}_n)'u_t}{\sqrt{n}}\right)^2+O_p\left(n^{-3/2}\right)\right)Y_{t-1}^2 \\
& -\frac{1}{n}\sum\varepsilon_t e^{\frac{\hat{c}_n}{n}+\frac{\hat{a}_n'u_t}{\sqrt{n}}}Y_{t-1} \\
= \ & 0.
\end{aligned}
$$

That is,

$$
\begin{aligned}
& \frac{\hat{c}_n-c}{n^2}\sum Y_{t-1}^2+\frac{(\hat{a}_n-a)'}{n^{3/2}}\sum u_tY_{t-1}^2+\frac{1}{n^2}\sum\left(2\left(\hat{a}_n'u_t\right)^2-\frac{1}{2}\left((a+\hat{a}_n)'u_t\right)^2\right)Y_{t-1}^2 \\
& -\frac{1}{n}\sum\varepsilon_tY_{t-1}-\frac{\hat{a}_n'}{n^{3/2}}\sum u_t\varepsilon_tY_{t-1}+O_p\left(n^{-1/2}\right)=0.
\end{aligned}
$$

Using (78), and Lemmas 15, 16 and 13, the solution to $\hat{c}_n-c$ is

$$
\begin{aligned}
& \hat{c}_n-c\Rightarrow-\frac{\varrho_a'\left(\int_0^1 G_{a,c}^2(r)\,dB_u(r)+2\left(\Lambda_{uu}'a\int_0^1 G_{a,c}^2(r)\,dr+\Lambda_{u\varepsilon}\int_0^1 G_{a,c}(r)\,dr\right)\right)}{\int_0^1 G_{a,c}^2(r)\,dr} \\
& \quad -2(a+\varrho_a)'\Sigma_u(a+\varrho_a)+\frac{1}{2}(2a+\varrho_a)'\Sigma_u(2a+\varrho_a) \\
& \quad +\frac{\int_0^1 G_{a,c}(r)\,dB_\varepsilon(r)+\Lambda_{u\varepsilon}'a\int_0^1 G_{a,c}(r)\,dr+\lambda_{\varepsilon\varepsilon}}{\int_0^1 G_{a,c}^2(r)\,dr}+\frac{(a+\varrho_a)'\Sigma_{u\varepsilon}\int_0^1 G_{a,c}(r)\,dr}{\int_0^1 G_{a,c}^2(r)\,dr}. \qquad (79)
\end{aligned}
$$

As

$$
\begin{aligned}
& -2(a+\varrho_a)'\Sigma_u(a+\varrho_a)+\frac{1}{2}(2a+\varrho_a)'\Sigma_u(2a+\varrho_a) \\
= \ & -2a'\Sigma_u\varrho_a-\frac{3}{2}\varrho_a'\Sigma_u\varrho_a \\
= \ & -2a'\Sigma_u\Sigma_u^{-1}\Sigma_{u\varepsilon}\frac{\int_0^1 G_{a,c}(r)\,dr}{\int_0^1 G_{a,c}^2(r)\,dr}-\frac{3}{2}\Sigma_{u\varepsilon}'\Sigma_u^{-1}\Sigma_u\Sigma_u^{-1}\Sigma_{u\varepsilon}\left(\frac{\int_0^1 G_{a,c}(r)\,dr}{\int_0^1 G_{a,c}^2(r)\,dr}\right)^2 \\
= \ & -2a'\Sigma_{u\varepsilon}\frac{\int_0^1 G_{a,c}(r)\,dr}{\int_0^1 G_{a,c}^2(r)\,dr}-\frac{3}{2}\Sigma_{u\varepsilon}'\Sigma_u^{-1}\Sigma_{u\varepsilon}\left(\frac{\int_0^1 G_{a,c}(r)\,dr}{\int_0^1 G_{a,c}^2(r)\,dr}\right)^2,
\end{aligned}
$$

and

$$\begin{aligned}
& -2a'\Sigma_{u\varepsilon}\frac{\int_0^1 G_{a,c}(r)\,dr}{\int_0^1 G_{a,c}^2(r)\,dr}-\frac{3}{2}\Sigma_{u\varepsilon}'\Sigma_u^{-1}\Sigma_{u\varepsilon}\left(\frac{\int_0^1 G_{a,c}(r)\,dr}{\int_0^1 G_{a,c}^2(r)\,dr}\right)^2+\frac{(a+\varrho_a)'\,\Sigma_{u\varepsilon}\int_0^1 G_{a,c}(r)\,dr}{\int_0^1 G_{a,c}^2(r)\,dr} \\
= & -2a'\Sigma_{u\varepsilon}\frac{\int_0^1 G_{a,c}(r)\,dr}{\int_0^1 G_{a,c}^2(r)\,dr}-\frac{3}{2}\Sigma_{u\varepsilon}'\Sigma_u^{-1}\Sigma_{u\varepsilon}\left(\frac{\int_0^1 G_{a,c}(r)\,dr}{\int_0^1 G_{a,c}^2(r)\,dr}\right)^2+\frac{a'\Sigma_{u\varepsilon}\int_0^1 G_{a,c}(r)\,dr}{\int_0^1 G_{a,c}^2(r)\,dr} \\
& +\frac{\Sigma_{u\varepsilon}'\Sigma_u^{-1}\Sigma_{u\varepsilon}\left(\int_0^1 G_{a,c}(r)\,dr\right)^2}{\left(\int_0^1 G_{a,c}^2(r)\,dr\right)^2} \\
= & -a'\Sigma_{u\varepsilon}\frac{\int_0^1 G_{a,c}(r)\,dr}{\int_0^1 G_{a,c}^2(r)\,dr}-\frac{1}{2}\Sigma_{u\varepsilon}'\Sigma_u^{-1}\Sigma_{u\varepsilon}\left(\frac{\int_0^1 G_{a,c}(r)\,dr}{\int_0^1 G_{a,c}^2(r)\,dr}\right)^2
\end{aligned}$$

we end up with (17).

In the case $\Sigma_{u\varepsilon}=0$, it is clear from (78) that $\varrho_a=0$, so that $\hat{a}_n$ is consistent. The second and third terms in (79) are equal to zero and from (79) it follows that

$$\hat{c}_n-c\Rightarrow\frac{\int_0^1 G_{a,c}(r)\,dB_\varepsilon(r)+\Lambda_{u\varepsilon}'a\int_0^1 G_a(r)\,dr+\lambda_{\varepsilon\varepsilon}}{\int_0^1 G_{a,c}^2(r)\,dr}.$$

In this case we can solve

$$\hat{a}_n=\arg\min_{a_1}\sum\left(Y_t-\beta_t(\hat{c}_n,a_1)\,Y_{t-1}\right)^2,$$

leading to

$$\begin{aligned}
& -2\sum\left(Y_t-\beta_{nt}(\hat{c}_n,\hat{a}_n)\,Y_{t-1}\right)\beta_{nt}(\hat{c}_n,\hat{a}_n)\frac{u_t}{\sqrt{n}}Y_{t-1} \\
= & -\frac{2e^{\frac{\hat{c}_n}{n}}}{n}\sum\left(\left(\beta_{nt}(c,a)-\beta_{nt}(\hat{c}_n,\hat{a}_n)\right)Y_{t-1}+\varepsilon_t\right)e^{\frac{\hat{a}_n'u_t}{\sqrt{n}}}u_tY_{t-1} \\
= & -\frac{2e^{\frac{\hat{c}_n}{n}}}{n}\sum\left(\left(e^{\frac{c}{n}+\frac{(a+\hat{a}_n)'u_t}{\sqrt{n}}}-e^{\frac{\hat{c}_n}{n}+\frac{2\hat{a}_n'u_t}{\sqrt{n}}}\right)u_tY_{t-1}^2+e^{\frac{\hat{a}_n'u_t}{\sqrt{n}}}u_t\varepsilon_tY_{t-1}\right) \\
= & \ 0.
\end{aligned}$$

An expansion yields a solution

$$\begin{aligned}
& \frac{1}{n}\sum\left(1+\frac{c}{n}+\frac{(a+\hat{a}_n)'\,u_t}{\sqrt{n}}+\frac{\left((a+\hat{a}_n)'\,u_t\right)^2}{2n}+O_p\left(n^{-3/2}\right)\right)u_tY_{t-1}^2 \\
& -\frac{1}{n}\sum\left(1+\frac{\hat{c}_n}{n}+\frac{2\hat{a}_n'u_t}{\sqrt{n}}+\frac{2\left(\hat{a}_n'u_t\right)^2}{n}+O_p\left(n^{-3/2}\right)\right)u_tY_{t-1}^2 \\
& +\frac{1}{n}\sum\left(1+\frac{\hat{a}_n'u_t}{\sqrt{n}}+\frac{\left(\hat{a}_n'u_t\right)^2}{2n}+O_p\left(n^{-3/2}\right)\right)u_t\varepsilon_tY_{t-1} \\
= & \ 0.
\end{aligned}$$

Keeping the leading terms, we obtain

$$
\begin{aligned}
& \frac{(c-\hat{c}_n)}{\sqrt{n}} \sum \frac{u_t}{\sqrt{n}} \frac{Y_{t-1}^2}{n} + \frac{1}{n} \sum \frac{Y_{t-1}^2}{n} u_t u_t' \left(\sqrt{n}\,(a-\hat{a}_n)\right) \\
& + \frac{1}{\sqrt{n}} \sum \left[\frac{1}{2} \left((a+\hat{a}_n)' u_t\right)^2 - 2\left(\hat{a}_n' u_t\right)^2 \right] \frac{u_t}{\sqrt{n}} \frac{Y_{t-1}^2}{n} \\
& + \sum \frac{u_t \varepsilon_t}{\sqrt{n}} \frac{Y_{t-1}}{\sqrt{n}} \\
& + \frac{1}{n} \sum \varepsilon_t u_t u_t' \hat{a}_n \frac{Y_{t-1}}{\sqrt{n}} \\
= \; & O_p\left(n^{-1/2}\right).
\end{aligned}
$$

The first term above is negligible and by the consistency of $\hat{a}_n$,

$$
\begin{aligned}
& \frac{1}{\sqrt{n}} \sum \left[\frac{1}{2} \left((a+\hat{a}_n)' u_t\right)^2 - 2\left(\hat{a}_n' u_t\right)^2 \right] \frac{u_t}{\sqrt{n}} \frac{Y_{t-1}^2}{n} \\
= \; & \frac{1}{\sqrt{n}} \sum \left[\frac{1}{2} \left((2a+o_p(1))' u_t\right)^2 - 2\left((a+o_p(1))' u_t\right)^2 \right] \frac{u_t}{\sqrt{n}} \frac{Y_{t-1}^2}{n} \\
= \; & o_p\left(n^{-1/2}\right).
\end{aligned}
$$

Therefore, the solution is

$$
\left(\frac{1}{n} \sum \frac{Y_{t-1}^2}{n} u_t u_t'\right) \sqrt{n}\,(\hat{a}_n - a) = \sum \frac{u_t \varepsilon_t}{\sqrt{n}} \frac{Y_{t-1}}{\sqrt{n}} + \frac{1}{n} \sum \varepsilon_t u_t u_t' a \frac{Y_{t-1}}{\sqrt{n}} + O_p\left(n^{-1/2}\right),
$$

and the proof is completed upon an application of Lemmas 15 and 17. ■

Proof of Theorem 6: The ols estimator of β_{nt} in (4) satisfies

$$
\hat{\beta}_n = \frac{\sum_{t=2}^n Y_t Y_{t-1}}{\sum_{t=2}^n Y_{t-1}^2} = \frac{\sum_{t=2}^n \beta_{nt} Y_{t-1}^2}{\sum_{t=2}^n Y_{t-1}^2} + \frac{\sum_{t=2}^n \varepsilon_t Y_{t-1}}{\sum_{t=2}^n Y_{t-1}^2}.
$$

The first term above yields

$$
\begin{aligned}
\frac{\sum_{t=2}^n \beta_{nt} Y_{t-1}^2}{\sum_{t=2}^n Y_{t-1}^2} &= \frac{\sum_{t=2}^n \left(1 + \frac{a' u_t}{\sqrt{n}} + \frac{c}{n} + \frac{1}{2}\left(\frac{a' u_t}{\sqrt{n}}\right)^2 + o_p\left(n^{-1}\right)\right) Y_{t-1}^2}{\sum_{t=2}^n Y_{t-1}^2} \\
&= 1 + \frac{\sum_{t=2}^n \left(\frac{a' u_t}{\sqrt{n}} + \frac{c}{n} + \frac{1}{2}\left(\frac{a' u_t}{\sqrt{n}}\right)^2 + o_p\left(n^{-1}\right)\right) Y_{t-1}^2}{\sum_{t=2}^n Y_{t-1}^2}.
\end{aligned}
$$

Thus,

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

By Lemmas 15 and 16, the first term satisfies

$$\begin{aligned}&\frac{n^{-1}\sum_{t=2}^{n}\left(\frac{a'u_t}{\sqrt{n}} + \frac{c}{n} + \frac{1}{2}\left(\frac{a'u_t}{\sqrt{n}}\right)^2 + o_p\left(n^{-1}\right)\right)Y_{t-1}^2}{n^{-2}\sum_{t=2}^{n}Y_{t-1}^2}\\ \Rightarrow\ &\frac{a'\left(\int_0^1 G_{a,c}^2\left(r\right)dB_u\left(r\right) + 2\left(\Lambda_{uu}'a\int_0^1 G_{a,c}^2\left(r\right)dr + \Lambda_{u\varepsilon}\int_0^1 G_{a,c}\left(r\right)dr\right)\right)}{\int_0^1 G_{a,c}^2\left(r\right)dr}\\ &+\frac{\left(c+\frac{a'\Sigma_u a}{2}\right)\int_0^1 G_{a,c}^2\left(r\right)dr}{\int_0^1 G_{a,c}^2\left(r\right)dr},\end{aligned}$$

and by Lemma 3, the second term yields

$$\frac{n^{-1}\sum_{t=2}^{n}\varepsilon_t Y_{t-1}}{n^{-2}\sum_{t=2}^{n}Y_{t-1}^2} \Rightarrow \frac{\int_0^1 G_{a,c}\left(r\right)dB_\varepsilon\left(r\right) + \Lambda_{u\varepsilon}'a\int_0^1 G_{a,c}\left(r\right)dr + \lambda_{\varepsilon\varepsilon}.}{\int_0^1 G_{a,c}\left(r\right)^2 dr}.$$

Hence,

$$\begin{aligned}n\left(\hat{\beta}_n - 1\right) \Rightarrow\ & c + \frac{a'\Sigma_u a}{2} + \frac{a'\left(\int_0^1 G_{a,c}^2\left(r\right)dB_u\left(r\right) + 2\left(\Lambda_{uu}'a\int_0^1 G_{a,c}^2\left(r\right)dr + \Lambda_{u\varepsilon}\int_0^1 G_{a,c}\left(r\right)dr\right)\right)}{\int_0^1 G_{a,c}^2\left(r\right)dr}\\ &+\frac{\int_0^1 G_{a,c}\left(r\right)dB_\varepsilon\left(r\right) + \Lambda_{u\varepsilon}'a\int_0^1 G_{a,c}\left(r\right)dr + \lambda_{\varepsilon\varepsilon}}{\int_0^1 G_{a,c}\left(r\right)^2 dr},\end{aligned}$$

which simplifies to the stated result. ■

Proof of Lemma 11. As $B_u\left(p\right)$ is independent of $dB_\varepsilon\left(p\right)$, the expected value of $G_{a,c}\left(r\right)$ is zero.

The covariance of the process is given by

$$
\begin{aligned}
& Cov\left(G_{a,c}\left(r\right), G_{a,c}\left(s\right)\right) \\
= \ & \mathbb{E}\left(\int_0^r \int_0^s e^{(r+s-p-q)c+a(B_u(r)+B_u(s)-B_u(p)-B_u(q))} dB_\varepsilon\left(p\right) dB_\varepsilon\left(q\right)\right) \\
= \ & \sigma_\varepsilon^2 \int_0^{r\wedge s} e^{(r+s-2p)c} \mathbb{E}\left(e^{a(B_u(r\vee s)-B_u(r\wedge s)+2(B_u(r\wedge s)-B_u(p)))}\right) dp \\
= \ & \sigma_\varepsilon^2 \int_0^{r\wedge s} e^{(r\vee s+r\wedge s-2p)c} e^{\frac{a^2\sigma_u^2(r\vee s-r\wedge s)}{2}+2a^2\sigma_u^2(r\wedge s-p)} dp \\
= \ & \sigma_\varepsilon^2 \int_0^{r\wedge s} e^{(r\vee s-r\wedge s+2(r\wedge s-p))c} e^{\frac{b(r\vee s-r\wedge s)}{2}+2b(r\wedge s-p)} dp \\
= \ & \sigma_\varepsilon^2 e^{\left(c+\frac{b}{2}\right)(r\vee s-r\wedge s)} \frac{e^{2(c+b)r\wedge s}-1}{2\left(c+b\right)},
\end{aligned}
$$

as stated. The fourth order moment is given by

$$
\begin{aligned}
& \mathbb{E}\left(G_{a,c}^4\left(r\right)\right) = \int_0^r \cdots \int_0^r e^{\sum_{i=1}^4 (r-p_i)c} \mathbb{E}\left(e^{\sum_{i=1}^4 B_u(r)-B_u(p_i)}\right) \mathbb{E}\prod_{i=1}^4 dB_\varepsilon\left(p_i\right) \\
= \ & 6\sigma_\varepsilon^4 \int_0^r \int_0^q e^{2(2r-p-q)c} \mathbb{E}\left(e^{4a(B_u(r)-B_u(q))+2a(B_u(q)-B_u(p))}\right) dpdq \\
= \ & 6\sigma_\varepsilon^4 \int_0^r \int_0^q e^{2(2r-p-q)c+8a^2\sigma_u^2(r-q)+2a^2\sigma_u^2(q-p)} dpdq \\
= \ & \frac{3\sigma_\varepsilon^4}{c+b}\left[\frac{1-e^{4(c+2b)r}}{4\left(c+2b\right)} - e^{2(c+b)r}\frac{1-e^{2(c+3b)r}}{2\left(c+3b\right)}\right]
\end{aligned}
$$

which gives the stated result. ■

Proof of Lemma 13: Expansion of the limit process in this case yields

$$
\begin{aligned}
G_{a,c}\left(r\right) &= e^{rc+a'B_u(r)} \int_0^r e^{-pc-a'B_u(p)} dB_\varepsilon\left(p\right) = \int_0^r e^{(r-p)c}\left\{1+a\left(B_u\left(r\right)-B_u\left(p\right)\right)+O_p\left(a^2\right)\right\} dB_\varepsilon\left(p\right) \\
&= \int_0^r e^{(r-p)c} dB_\varepsilon\left(p\right) + a\int_0^r e^{(r-p)c}\left(B_u\left(r\right)-B_u\left(p\right)\right) dB_\varepsilon\left(p\right) + O_p\left(a^2\right) =: G_c\left(r\right)+V_{c,a}\left(r\right)+O_p\left(a^2\right).
\end{aligned} \tag{80}
$$

Here $G_c\left(r\right)$ is the limit process of the LUR process and has finite dimensional distribution $\mathcal{N}\left(0, \sigma_\varepsilon^2\left(e^{2cr}-1\right)/2c\right)$ (Phillips, 1987)). The process $V_{c,a}\left(r\right)$ has mean $\mathbb{E}\left(V_{c,a}\left(r\right)\right)=0$, variance

$$
Var\left(V_{c,a}\left(r\right)\right) = a^2\sigma_\varepsilon^2\sigma_u^2 \int_0^r e^{2(r-p)c}\left(r-p\right) dp = \frac{\sigma_\varepsilon^2 b}{4c^2}\left(1+e^{2cr}\left(2cr-1\right)\right), \tag{81}
$$

and fourth moment

$$\mathbb{E}\left(V_{c,a}^{4}(r)\right) = a^{4}\int_{0}^{r}\cdots\int_{0}^{r} e^{\sum_{i=1}^{4}(r-p_i)c}\mathbb{E}\prod_{i=1}^{4}\left(\left(B_u(r)-B_u(p_i)\right)\right)\mathbb{E}\prod_{i=1}^{4}dB_{\varepsilon}(p_i)$$

$$= 6a^{4}\sigma_u^{4}\sigma_{\varepsilon}^{4}\int_{0}^{r}\int_{0}^{q} e^{(4r-2(p+q))c}(r-p)(r-q)\,dpdq = 3a^{4}\sigma_u^{4}\sigma_{\varepsilon}^{4}\frac{\left(1+e^{2cr}(2cr-1)\right)^{2}}{16c^{4}}.$$

It follows that $\mathbb{E}\left(V_{c,a}^{4}(r)\right) = 3\left(Var\left(V_{c,a}(r)\right)\right)^{2}$ and so $V_{c,a}$ has kurtosis 3. Observe that, in view of the independence of B_u and B_{ε} in the present case, the process $V_{c,a}(r)$ is mixed normal ($\mathcal{MN}$) process with finite dimensional distribution

$$V_{c,a}(r) \sim_d \mathcal{MN}\left(0, a^{2}\int_{0}^{r} e^{2(r-p)c}\left(B_u(r)-B_u(p)\right)^{2}dp\right).$$

Finally,

$$\mathbb{E}\left(G_c(r)V_{c,a}(r)\right) = a\sigma_{\varepsilon}^{2}\int_{0}^{r} e^{2(r-p)c}\mathbb{E}\left(B_u(r)-B_u(q)\right)dp = 0,$$

and so

$$Var\left(G_{a,c}(r)\right) = Var\left(G_c(r)\right)+Var\left(V_{c,a}(r)\right)+O\left(b^{2}\right) = \sigma_{\varepsilon}^{2}\frac{e^{2cr}-1}{2c}+\frac{\sigma_{\varepsilon}^{2}b}{4c^{2}}\left(1+e^{2cr}(2cr-1)\right)+O\left(b^{2}\right), \tag{82}$$

giving (32). The moment expansion (82) is valid based on the stochastic expansion (80) because all moments of the component Gaussian processes $\left(B_u(r), B_{\varepsilon}(r)\right)$ are finite and bounded. ■

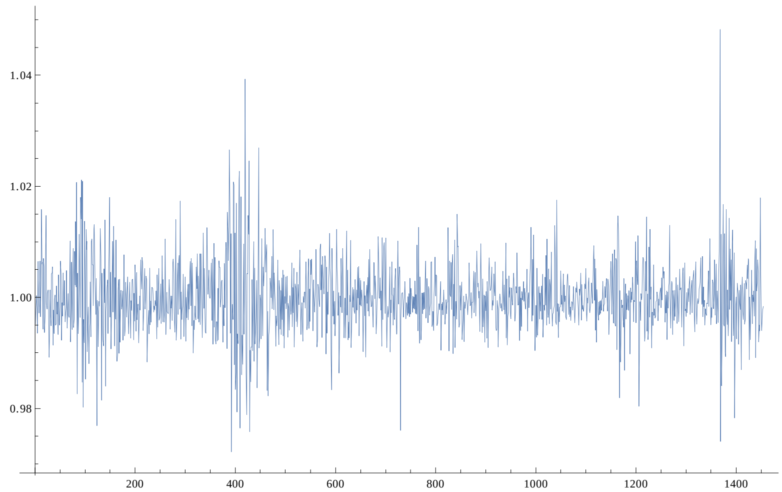

Figure 1: $\hat{\beta}_t$ of an LSTUR model under the restriction $c + b = 0$.

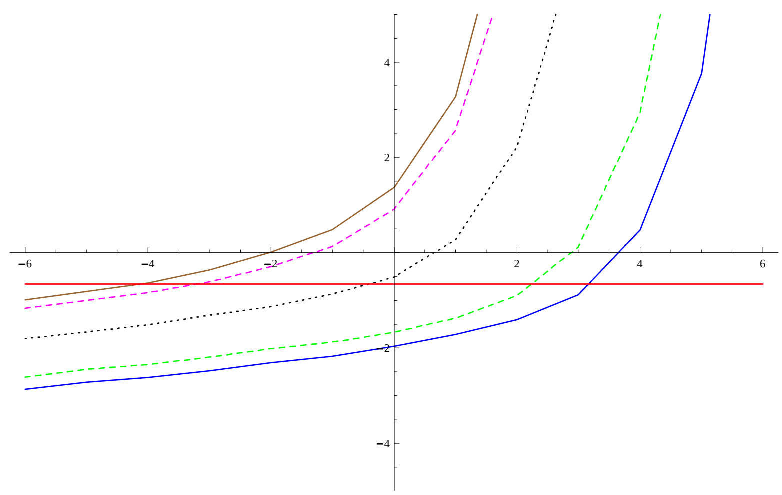

Figure 2: Asymptotic confidence belts for $t_{\hat{\beta}}$ - the LSTUR case, based on equation (31). Y-axis - $t_{\hat{\beta}}$ values, X-axis - c values, solid blue - 5th percentile belt, dashed green - 10th percentile belt, dotted black - median belt, dashed magenta, 90th - percentile belt, solid brown - 95th - percentile belt, horizontal red line - the sample's $t_{\hat{\beta}}$, $a = -0.245$, $\rho = -0.150$, $\sigma_u^2 = 0.983$, $\sigma_\varepsilon^2 = 7 \times 10^{-5}$, $\hat{t}_\beta = -0.659$. Calculated with a grid step of 1, 400 integral points and 5000 replications.