5-2021

Latent local-to-unity models

Jun Yu
Singapore Management University, yujun@smu.edu.sg

SMU ECONOMICS & STATISTICS

Latent Local-to-Unity Models

Jun Yu

May 2021

Paper No. 04-2021

Latent Local-to-Unity Models*

Jun Yu
Singapore Management University

May 5, 2021

Abstract

This paper proposes a class of state-space models where the state equation is a local-to-unity process. The large sample theory is obtained for the least squares (LS) estimator of the autoregressive (AR) parameter in the AR representation of the model under two sets of conditions. In the first set of conditions, the error term in the observation equation is independent and identically distributed (iid), and the error term in the state equation is stationary and fractionally integrated with memory parameter $H \in (0, 1)$. It is shown that both the rate of convergence and the asymptotic distribution of the LS estimator depend on H. In the second set of conditions, the error term in the observation equation is independent but not necessarily identically distributed, and the error term in the state equation is strong mixing. When both error terms are iid, we also develop the asymptotic theory for an instrumental variable estimator. Special cases of our models are discussed.

JEL classification: C12, C22, G01
Keywords: State-space; Local-to-unity; O-U process; Fractional O-U process; Fractional Brownian motion; Fractional integration; Instrumental variable.

1 Introduction

Since the local-to-unity literature was initiated by Phillips (1987a) and Chan and Wei (1987), the local-to-unity model has received so much attention both in theoretical studies and in empirical studies.[1] The success of the local-to-unity model is

*I would like to thank Peter Phillips, Liyu Dou, Yiu Lim Lui, Jia Li, Shuping Shi, and Xiaohu Wang for helpful discussion. Jun Yu, School of Economics and Lee Kong Chian School of Business, Singapore Management University, 90 Stamford Road, Singapore 178903. Email: yujun@smu.edu.sg.

[1] An incomplete list of contributions include Stock (1991), Elliott and Stock (1994), Cavanagh et al. (1995), Wright (2000a), Elliott and Stock (2001), Gospodinov (2004), Valkanov (2003), Torous

not surprising because (1) the local-to-unity model is more general than the exact unit root model; (2) it well describes the dynamics of many macroeconomic time series and financial time series; (3) the resulting asymptotic distribution better approximates the finite sample distribution than the asymptotic distribution under the assumption of weak dependence.

However, the local-to-unity models used in practical applications assume the variable of interest is observed without any error. This assumption can be too strong in practice. For example, when a time series is obtained from a survey, many forms of errors are possible, including recall errors and sampling errors. These so-called measurement errors can occur with a systematic pattern that generates the difference between the respondents' answers to a question and the actual values. See Kasprzyk (2005) for possible sources of measurement errors and Bound et al. (2001) for certain econometric consequences of measurement errors. For another example, a time series is sometimes obtained from estimation. A well-known example that motivates the present paper is the daily time series of realized volatilities (RV), which are estimates of the daily integrated volatilities (IV). Andersen et al. (2003) and Corsi (2009) introduce alternative models for RV. For the third example, a latent time series may be related to an observed time series by definition or for structural reasons. The class of the DSGE models and the class of stochastic volatility (SV) models are among the interesting models in this example.

In this paper, we consider the following latent local-to-unity model:

$$\begin{cases} y_t = \xi_t + w_t \\ \xi_t = \theta_T \xi_{t-1} + v_t, \theta_T = 1 + \frac{c}{T}, \xi_0 \sim O_p(1) \end{cases}, t = 0, ..., T, \tag{1}$$

where $\{\xi_t\}$ is a latent process that is local-to-unity with $c \in (-\infty, \infty)$ being the local coefficient. When $\theta_T = \theta$, which is independent of T, and when $\{w_t\}$ and $\{v_t\}$ are serially independent Gaussian processes, the model is the popular linear Gaussian state-space model. We deviate from the literature on linear Gaussian state-space modeling by assuming θ_T is a function of T, and also, by allowing for more general stochastic behavior for w_t and v_t.

We are not the first one that is concerned about the latency of a persistent process. Motivated by the fact that RV is merely an estimate of IV, Hansen and Lunde (2014) also consider model (1). However, our model is different from their model in two aspects. First, Hansen and Lunde (2014) assume $\{w_t\}$ and $\{v_t\}$ are

et al. (2004), Rossi (2005), Campbell and Yogo (2006), Jansson and Moreira (2006), Mikusheva (2007), Wang et al. (2019), Jiang et al. (2021), Dou and Müller (2021).

serially independent. Our assumptions about $\{w_t\}$ and $\{v_t\}$ are more general, as it will become clear later. Second, Hansen and Lunde (2014) assume $\theta_T = \theta$ and consider two cases in particular, $\theta < 1$ and $\theta = 1$. We focus our attention to the case where $\theta_T = 1 + \frac{c}{T}$ that includes the unit root as a special case. Like Hansen and Lunde (2014), the parameter of interest in our paper is θ_T.

Model (1) may be rewritten as

$$y_t = \xi_t + w_t = \theta_T \xi_{t-1} + v_t + w_t = \theta_T y_{t-1} + v_t + w_t - \theta_T w_{t-1} := \theta_T y_{t-1} + \varepsilon_t, \quad (2)$$

where

$$\varepsilon_t = v_t + w_t - \theta_T w_{t-1}. \quad (3)$$

When $\{w_t\}$ and $\{v_t\}$ are serially independent, $\{\varepsilon_t\}$ is a local-to-unity moving averaging (MA) process plus an iid process. While $\{\varepsilon_t\}$ involves a local-to-unity MA component, a property shared by the model of Dou and Müller (2021), our model is very different from theirs because there is v_t in it. We will address the implication for this important difference later.

The paper is organized as follows. In Section 2, we assume $\{v_t\}$ is a stationary fractionally integrated process and $\{w_t\}$ is an iid sequence. In Section 3, we assume $\{v_t\}$ is strong mixing and $\{w_t\}$ is an independent but not necessarily identically distributed sequence. In the same section, we also investigate the large sample theory of a more efficient estimator when $\{v_t\}$ and $\{w_t\}$ are both iid. Section 4 concludes. Appendix collects the proof of the theoretical results. Throughout the paper, we use $\overset{p}{\to}$, $\Rightarrow$, $\overset{d}{\to}$, $\overset{d}{=}$, $\overset{iid}{\sim}$ to denote convergence in probability, weak convergence, convergence in distribution, and equivalence in distribution, independent and identically distributed, respectively.

2 Latent Model with Fractionally Integrated Errors

In the RV literature, a well-established stylized fact is a slowly decaying ACF for daily RV. However, how to model the slowly decaying ACF has recently been debated in the literature. An earlier attempt is to use a fractional process, namely $I(H - 0.5)$ with $H \in (0.5, 1)$, to model the slowly decaying ACF.[2] Important con-

[2] In the literature on fractional integration, parameter d has been traditionally used to represent the memory parameter. In this paper, we use H to represent the memory parameter. The two parameters are related to each other by $d = H - 0.5$.

tributions include Andersen et al. (2001), Andersen et al. (2001), Andersen et al. (2003). Andersen and Bollerslev (1997) provide an interesting explanation of a slowly decaying ACF in volatilities (i.e. ACF at lag k is of order k^{2-2H} with $H \in (0.5, 1)$ for large k so that the ACF is not absolutely summable) from the interactions of a large number of heterogeneous information processes.[3] Andersen et al. (2003) introduce the ARFIMA$(1, H-0.5, 0)$ model with $H \in (0.5, 1)$ for log RV and provide evidence that the model outperforms many alternative models in predicting RV and log RV, including GARCH-type models and other high-frequency models.

Gatheral et al. (2018) establish a new stylized fact, namely, the roughness of sample paths of RV. Consequently, a more recent attempt is to use continuous-time models based on the fractional Brownian motion (fBm), denoted by $B^H(t)$ with H being the Hurst parameter, to explain roughness in RV. To generate roughness, H in $B^H(t)$ must be in $(0, 0.5)$. Since the ACF of ARFIMA$(0, H-0.5, 0)$ and the first difference of $B^H(t)$ have the same asymptotic behavior, the use of $H < 0.5$ in the fBm is at odds with $H > 0.5$ used in the ARFIMA model.

Gatheral et al. (2018) find that the fBm with $H = 0.14$ has exceptionally good performance in forecasting RV or log RV out-of-sample. It outperforms the AR(5), AR(10), and HAR models in predicting the daily RV and log RV. Wang et al. (2019) propose a two-stage estimation method for the fractional Ornstein–Uhlenbeck (fO-U) process and develop large sample properties of the estimators. When applying the method to the daily log RV, daily log realized kernel (RK), and daily log bipower variation (BV), they find strong evidence of $H < 0.5$, although $H > 0.5$ is also allowed in their model. The same paper also reports evidence of the near unity root behavior, which leads to slowly decaying ACF at small and moderate lags. When examining the forecasting performance of the fO-U model out-of-sample, it is found that the fO-U model outperforms the random walk, AR(1), HAR, ARFIMA, fBm in predicting the daily RV, daily log RV, daily log RK, daily log BV.

Instead of modeling RV or log RV directly, Fukasawa et al. (2021) and Bolko et al. (2020) assume the log spot variance follows a continuous time model driven by the fBm. In both studies, high-frequency data are used to construct daily RV, and the estimation errors in the daily RV are taken seriously when the estimation equations

[3] Partly motivated by the presence of heterogeneous traders, Corsi (2009) proposes to use the heterogeneous autoregressive (HAR) model to capture the slowly decaying ACF. The HAR model has become a popular model in practice to forecast RV. An interesting observation of Table 2 in Corsi (2009) is that the sum of the three autoregressive parameter estimates is very close to one for USD/CHF and S&P500.

are set up. In particular, Fukasawa et al. (2021) obtain the approximate spectral density of the daily RV and then get the quasi-maximum likelihood estimate of H. Bolko et al. (2020) obtain the expressions for moments of the daily RV and then use the generalized method of moments to estimate H. When applying the proposed method to real data, both studies report strong evidence of $H \in (0, 0.5)$. In addition, Bolko et al. (2020) obtain strong evidence of the near unit root behavior in log spot variance. Liu et al. (2020) assume the log spot variance follows the fO-U process. Unlike Fukasawa et al. (2021) and Bolko et al. (2020), where high-frequency data are used to obtain the daily RV, Liu et al. (2020) only use daily returns, and treat log spot variance latent. A simulated maximum likelihood method is introduced to estimate the model. Strong evidence of $H \in (0, 0.5)$ and the near unit root behavior is found in several empirical studies in Liu et al. (2020).

Motivated by this debate in the literature, together with the possibility that IV is highly persistent, and the fact that daily RV is an approximation to daily IV, we consider model (1) with the following three assumptions.

Assumption 1 *$w_t \overset{iid}{\sim} (0, \sigma_w^2)$. There exists $k = \max\left\{4, \frac{2}{H} - 4\right\}$ where $H \in (0,1)$ such that $E\left|w_t\right|^k < \infty$.*

Assumption 2 *$(1-L)^{H-0.5} v_t = e_t \overset{iid}{\sim} (0, \sigma_e^2)$ with $\sigma_e^2 > 0$. There exists $k = \max\left\{4, \frac{2}{H} - 4\right\}$ where $H \in (0,1)$ such that $E\left|e_t\right|^k < \infty$.*

Assumption 3 *w_t and v_s are independent for any t and s.*

Remark 2.1 *Assumption 1 allows $\{\xi_t\}$ to be observed with iid errors.*

Remark 2.2 *Assumption 2 allows ξ_t to have an error term that is fractionally integrated, $I(H-0.5)$. When $H > 0.5$, v_t has a slowly decaying ACF such that the ACF is not summable. When $H = 0.5$, $v_t = e_t$ becomes an iid sequence. When $H < 0.5$, v_t is antipersistent and has a fast-decaying ACF although the ACF of ξ_t can decay slowly at small and moderate lags due to the local-to-unity. Assumption 2 implies that*

$$v_t = \sum_{k=0}^{\infty} a_k e_{t-k}, \text{ with } a_k = \frac{\Gamma(k+H-0.5)}{\Gamma(k+1)\Gamma(H-0.5)} \sim |k|^{H-1.5} \text{ for large } |k|.$$

An $I(H-0.5)$ process is always stationary when $H \in (0,1)$. From Sowell (1990), we know that

$$Var\left(\sum_{t=1}^{T} v_t\right) = \frac{\sigma_e^2 \Gamma(2-2H)}{2H\Gamma(H+0.5)\Gamma(1.5-H)} \left[\frac{\Gamma(H+0.5+T)}{\Gamma(0.5-H+T)} - \frac{\Gamma(H+0.5)}{\Gamma(0.5-H)}\right]$$

$$\sim \quad T^{2H}\left[\frac{\sigma_e^2\Gamma(2-2H)}{2H\Gamma(H+0.5)\Gamma(1.5-H)}\right] := T^{2H}\overline{\sigma}_v^2, \tag{4}$$

where $\overline{\sigma}_v^2 = \frac{\sigma_e^2\Gamma(2-2H)}{2H\Gamma(H+0.5)\Gamma(1.5-H)}$ *is often referred to as the long-run variance of* v_t *in the literature. The instantaneous variance of* v_t *(denoted by* σ_v^2*) is*

$$\sigma_v^2 := Var\left(v_t\right) = \frac{\sigma_e^2\Gamma(2-2H)}{\left(\Gamma(1.5-H)\right)^2}. \tag{5}$$

Since $Var\left(\sum_{t=1}^T v_t\right) \sim T^{2H}\overline{\sigma}_v^2$ *and* $E\left|e_t\right|^k < \infty$ *for* $k = \max\left\{4, \frac{2}{H}-4\right\}$*, according to Davydov (1970, Theorem 2), we have the following functional central limit theorem (FCLT):*

$$\frac{1}{T^H\overline{\sigma}_v}\sum_{t=1}^{[Tr]} v_t \Rightarrow B^H(r), \text{ as } T \to \infty,$$

where $[Tr]$ *denotes the integer part of* Tr *and* $B^H(r)$ *is an fBm that is a Gaussian process with mean zero and covariance function*

$$Cov\left(B^H(t), B^H(s)\right) = \frac{1}{2}\left(|t|^{2H} + |s|^{2H} - |t-s|^{2H}\right), \quad \forall t, s. \tag{6}$$

An alternative definition of fBm is given by Mandelbrot and van Ness (1968) as

$$B^H(t) = \frac{1}{\Gamma(H+0.5)}\left\{\int_{-\infty}^0\left[(t-s)^{H-0.5} - (-s)^{H-0.5}\right]dW(s) + \int_0^t (t-s)^{H-0.5}\,dW(s)\right\},$$

where $W(t)$ *is a standard Brownian motion. Clearly, if* $H = 0.5$*,* $B^H(r)$ *becomes a standard Brownian motion,* $W(r)$*.*

Remark 2.3 *Since* v_t *is a stationary* $I(H-0.5)$ *process, and* $\{w_t - \theta_T w_{t-1}\}$ *is an m-dependent process by Assumption 1,* ε_t *defined in (3) is also a stationary* $I(H-0.5)$ *process due to Assumption 3. Note that the limit of the instantaneous variance of* ε_t *(denoted by* σ_ε^2*) is*

$$\sigma_\varepsilon^2 := \lim_{T\to\infty} Var\left(\varepsilon_t\right) = \lim_{T\to\infty}\left[E\left(v_t^2\right) + E\left(w_t^2\right) + \theta_T E\left(w_t^2\right)\right] = \frac{\sigma_e^2\Gamma(2-2H)}{\left(\Gamma(1.5-H)\right)^2} + 2\sigma_w^2, \tag{7}$$

and that

$$\begin{aligned} Var\left(\sum_{t=1}^T \varepsilon_t\right) &= E\left(\sum_{t=1}^T v_t\right)^2 + E\left(\sum_{t=1}^T (w_t - \theta_T w_{t-1})\right)^2 \\ &= T^{2H}\overline{\sigma}_v^2 + \left(T\left(1-\theta_T\right)^2 - 2\theta_T\right)\sigma_w^2 \end{aligned}$$

$$
\begin{aligned}
&= T^{2H}\overline{\sigma}_v^2 + \left(\frac{c^2}{T} - 2\theta_T\right)\sigma_w^2 \\
&\sim T^{2H}\overline{\sigma}_v^2. \qquad (8)
\end{aligned}
$$

Moreover, Assumption 1 and Assumption 2 imply that there exists $k = \max\left\{4, \frac{2}{H} - 4\right\}$ *such that* $E\left|\varepsilon_t\right|^k < \infty$*. Consequently, the FCLT of Davydov (1970) is applicable to the partial sum process of* $\{\varepsilon_t\}$*. The FCLT for the partial sum process of* $\{\varepsilon_t\}$ *is the source of limits of several sample moments that we state in the following lemma.*

Lemma 2.1 *Let* $\{y_t\}_{t=0}^T$ *be the time series generated by (1). Let Assumptions 1-3 hold. Then, as* $T \to \infty$,

1. $\frac{1}{T^H\overline{\sigma}_v} y_{[Tr]} \Rightarrow J_c^H(r)$;
2. $\frac{1}{T^{1+2H}\overline{\sigma}_v^2}\sum_{t=1}^T y_{t-1}^2 \xrightarrow{d} \int_0^1 J_c^H(r)^2 dr$;
3. $\frac{1}{T}\sum_{t=1}^T y_{t-1}\varepsilon_t \xrightarrow{d} -\frac{1}{2}\left(\frac{\sigma_e^2\Gamma(2-2H)}{(\Gamma(1.5-H))^2} + 2\sigma_w^2\right)$, *if* $H < 0.5$;
4. $\frac{1}{T}\sum_{t=1}^T y_{t-1}\varepsilon_t \xrightarrow{d} \frac{1}{2}\left(\overline{\sigma}_v^2 J_c(1)^2 - 2c\overline{\sigma}_v^2\int_0^1 J_c(r)dr - \left(\frac{\sigma_e^2\Gamma(2-2H)}{(\Gamma(1.5-H))^2} + 2\sigma_w^2\right)\right)$, *if* $H = 0.5$;
5. $\frac{1}{T^{2H}}\sum_{t=1}^T y_{t-1}\varepsilon_t \xrightarrow{d} \frac{\overline{\sigma}_v^2}{2}\left(J_c^H(1)^2 - 2c\int_0^1 J_c^H(r)dr\right)$, *if* $H > 0.5$.

   where $J_c(t)$ *is an O-U process defined by*

   $$dJ_c(t) = cJ_c(t)dt + dW(t), J_c(0) = 0,$$

   and $J_c^H(t)$ *is an fO-U process defined by*

   $$dJ_c^H(t) = cJ_c^H(t)dt + dB^H(t), J_c^H(0) = 0. \qquad (9)$$

We consider the following LS estimator of θ_T, denoted by $\widehat{\theta}_T$,

$$\widehat{\theta}_T = \frac{\sum_{t=1}^T y_t y_{t-1}}{\sum_{t=1}^T y_{t-1}^2} = \theta_T + \frac{\sum_{t=1}^T \varepsilon_t y_{t-1}}{\sum_{t=1}^T y_{t-1}^2}.$$

While the LS is not efficient, it is a simpler alternative to the maximum likelihood estimator (MLE) that can be obtained by the Kalman filter. However, MLE requires the full parametric assumption about the model. Moreover, it does not have an analytical expression.[4]

[4] In the case where $v_t \sim MA(q)$ with $q < \infty$, another more efficient estimator is the instrumental variable estimator that has the analytical expression (see Hall (1989) and Hansen and Lunde (2014)). However, in our model, $v_t \sim I(H - 0.5)$ that can only be expressed as an $MA(\infty)$ unless $H = 0.5$. As a result, it is not clear how to find valid instruments.

Theorem 2.1 *Let $\{y_t\}_{t=0}^T$ be the time series generated by (1). Let Assumptions 1-3 hold. Then, as $T \to \infty$,*

$$T^{2H}\left(\widehat{\theta}_T - \theta_T\right) \xrightarrow{d} -\frac{\frac{H\Gamma(H+0.5)}{\Gamma(1.5-H)} + \frac{2H\Gamma(H+0.5)\Gamma(1.5-H)}{\Gamma(2-2H)}\frac{\sigma_w^2}{\sigma_e^2}}{\int_0^1 J_c^H(r)^2 dr}, \text{ if } H < 0.5; \quad (10)$$

$$T\left(\widehat{\theta}_T - \theta_T\right) \xrightarrow{d} \frac{J_c(1)^2 - 2c\int_0^1 J_c(r)dr - \left(\frac{2H\Gamma(H+0.5)}{\Gamma(1.5-H)} + \frac{4H\Gamma(H+0.5)\Gamma(1.5-H)}{\Gamma(2-2H)}\frac{\sigma_w^2}{\sigma_e^2}\right)}{2\int_0^1 J_c(r)^2 dr}, \text{ if } H = 0.5; \quad (11)$$

$$T\left(\widehat{\theta}_T - \theta_T\right) \xrightarrow{d} \frac{J_c^H(1)^2 - 2c\int_0^1 J_c^H(r)dr}{2\int_0^1 J_c^H(r)^2 dr}, \text{ if } H > 0.5. \quad (12)$$

Remark 2.4 *When $\sigma_w^2 = 0$, there is no measurement error in ξ_t. If, in addition, $c = 0$ (and hence, $\theta = 1$), then the model is a unit root model without measurement error. This model is studied in Sowell (1990) where it is shown that, as $T \to \infty$,*

$$T\left(\widehat{\theta} - 1\right) \xrightarrow{d} \frac{\int_0^1 W(r)dW(r)}{\int_0^1 W(r)^2 dr}, \quad \text{if } H = 0.5, \quad (13)$$

$$T\left(\widehat{\theta} - 1\right) \xrightarrow{d} \frac{\frac{1}{2}B^H(1)^2}{\int_0^1 B^H(r)^2 dr}, \quad \text{if } H > 0.5, \quad (14)$$

$$T^{2H}\left(\widehat{\theta} - 1\right) \xrightarrow{d} -\frac{H\frac{\Gamma(0.5+H)}{\Gamma(1.5-H)}}{\int_0^1 B^H(r)^2 dr}, \quad \text{if } H < 0.5. \quad (15)$$

As $J_0(t) = W(t)$ and $J_0^H(t) = B^H(r)$, our results include those in Sowell (1990) as special cases.

Remark 2.5 *When $\sigma_w^2 = 0$, our model is closely related to the following model considered in Wang et al. (2021):*

$$\xi_t = \theta_T \xi_{t-1} + \varepsilon_t, \theta_T = \exp(c/T) \approx 1 + \frac{c}{T}, t = 1, ..., T, \quad (16)$$

where ε_t is a fractional Gaussian noise (FGN) with memory parameter $H \in (0,1)$ whose covariance function is

$$Cov\left(\eta_{t\Delta}^H, \eta_{s\Delta}^H\right) = \Delta^{2H}\frac{1}{2}\left[(k+1)^{2H} + (k-1)^{2H} - 2k^{2H}\right] \sim H(2H-1)k^{2H-2} \text{ for large } k = |t-s|.$$

For large k, the covariance function of an FGN is the same as that of $ARFIMA(0, H-0.5, 0)$ model. Wang et al. (2021) show that when ξ_t is observed, as $T \to \infty$, we have

$$T\left(\widehat{\theta}_T - \theta_T\right) \xrightarrow{d} \frac{\left(J_c(1)^2 - 2c\int_0^1 J_c(r)^2 dr - 1\right)/2}{\int_0^1 J_c(r)^2 dr}, \text{ if } H = 0.5;$$

$$T\left(\widehat{\theta}_T-\theta_T\right)\xrightarrow{d}\frac{\left(J_c^H(1)^2-2c\int_0^1 J_c^H(r)^2dr\right)/2}{\int_0^1 J_c^H(r)^2dr},\ \textit{if } H>0.5;$$

$$T^{2H}\left(\widehat{\theta}_T-\theta_T\right)\xrightarrow{d}\frac{-1/2}{\int_0^1 J_c^H(r)^2dr},\ \textit{if } H<0.5.$$

The results in Wang et al. (2021) are not the invariance principle as the error term in model (16) is assumed to be normally distributed. However, since the FGN has the same ACF as the $I(H-0.5)$ process for large lags, our asymptotic theory is very similar to that in Wang et al. (2021). The only difference is in the numerator of the limiting distribution. This difference arises because the instantaneous variance of the $I(H-0.5)$ process is different from that of FGN.

Let us conclude this section by relating our model with several interesting models proposed in the literature.

Example 2.1 *In Equation (35), Comte and Renault (1996) specify the following model for the log spot variance:*

$$dX(t)=-\kappa X(t)dt+\sigma dB^H(t),H>0.5. \tag{17}$$

It is well-known that the following discrete-time model

$$\xi_{t\Delta}=e^{-\kappa\Delta}\xi_{(t-1)\Delta}+(1-L)^{0.5-H}e_{t\Delta},\ e_{t\Delta}\overset{iid}{\sim}\left(0,\frac{1-e^{-2\kappa\Delta}}{2\kappa}\sigma^2\right),\ t=1,...,T, \tag{18}$$

weakly converges to model (17), that is, $\frac{\delta_H\Gamma(H+1/2)}{T^H}\xi_{\lfloor Tr\rfloor}\Rightarrow X(r)$, as $\Delta\to 0$ (Tanaka, 2013). Clearly, model (18) is a local-to-unity model with the error term satisfying Assumption 2. Comte and Renault (1996) impose the restriction that $H>0.5$. This restriction together with the local-to-unity feature will make the sample path of $X(t)$ very smooth.

Example 2.2 *Breidt et al. (1998) propose the following long memory SV model:*

$$r_t = \sigma e^{\xi_t/2}\varepsilon_t,\ \varepsilon_t\overset{iid}{\sim}N(0,1), \tag{19}$$

$$\xi_t = \theta\xi_{t-1}+v_t,(1-L)^{H-0.5}v_t=e_t\overset{iid}{\sim}N\left(0,\sigma_e^2\right). \tag{20}$$

Let $y_t=\log\left(r_t^2\right)=\log\sigma^2+1.27+\xi_t+w_t$ where $w_t\overset{iid}{\sim}\log\left(\chi^2_{(1)}\right)-1.27$. The estimated value for $H-0.5$ reported in Breidt et al. (1998) is 0.444, implying that $H=0.944$. If $\theta=1+\frac{c}{T}$, then the model is a special case of our model.

Example 2.3 *Andersen et al. (2003) propose the following model for log RV,*

$$\xi_t = \mu + \theta \xi_{t-1} + v_t, (1-L)^{H-0.5} v_t = e_t \overset{iid}{\sim} N\left(0, \sigma_e^2\right).$$

If ξ_t is the log IV, $\theta = 1 + \frac{c}{T}$, and ξ_t is proxied by the observed log RV that is equal to $\xi_t + w_t$ with $\{w_t\}$ being an iid sequence, then the model is a special case of our model.

Example 2.4 *Wang et al. (2019) specify the following model for log RV:*

$$d\xi(t) = -\kappa \xi(t) dt + \sigma dB^H(t),$$

Different from the model in Comte and Renault (1996), a general $H \in (0,1)$ is allowed in Wang et al. (2019). Since the discrete-time model specified in (18) weakly converges to model (17) as $\Delta \to 0$ for any $H \in (0,1)$, the model of Wang et al. (2019) is closely related to a local-to-unity model with fractionally integrated errors. If $\xi(t)$ is the log IV and $\xi(t)$ is proxied by the log RV that perturbs $\xi(t)$ with an iid error, then the model is closely related to our model.

Example 2.5 *Liu et al. (2020) propose the following fractional SV model:*

$$r_{t\Delta} = \sigma e^{\xi_{t\Delta}/2} \varepsilon_{t\Delta}, \ \varepsilon_{t\Delta} \overset{iid}{\sim} N(0,1), \tag{21}$$

$$\xi_{t\Delta} = (1 + \gamma \Delta) \xi_{(t-1)\Delta} + \sigma_v \eta_{t\Delta}^H, \tag{22}$$

where $\eta_{t\Delta}^H := B^H(t\Delta) - B^H((t-1)\Delta)$ is a FGN. As shown earlier, the covariance function of a FGN is the same as that of ARFIMA$(0, H-0.5, 0)$ model for large lags. Let $y_{t\Delta} = \log(r_{t\Delta}^2) = \log \sigma^2 + 1.27 + \xi_{t\Delta} + w_{t\Delta}$ where $w_{t\Delta} \overset{iid}{\sim} \log\left(\chi_{(1)}^2\right) - 1.27$. Then the fractional SV model is closely related to our model if $\Delta \to 0$.

3 Latent Model with Strong Mixing Errors

While Assumptions 1-3 allow for fractionally integrated errors in the latent local-to-unity model, no heteroskedasticity is allowed in $\{w_t\}$ or $\{v_t\}$. It is possible that $\{w_t\}$ and/or $\{v_t\}$ involve heteroskedasticity in practice and hence, it is important to relax the requirement of homoskedasticity.

For example, heteroskedasticity may be in presence in $\{v_t\}$ when ξ_t is the spot variance. The well-known square root model of Heston (1993) and the GARCH diffusion model of Nelson (1990) are two widely used specifications for the spot variance that allow for heteroskedasticity in the error term of the discretized representation via the Euler scheme. In the discretized square root model, the variance of v_t is a linear function of ξ_t. In the discretized GARCH diffusion model, the variance of v_t is a square function of ξ_t. This is the reason why we would like to relax the identical assumption about v_t.

For another example, heteroskedasticity arises in $\{w_t\}$ when one uses daily RV to estimate daily IV. To compute the daily RV for a trading day t, let the intra day return based on a particular sampling frequency M be

$$r_{i,t} = p_{i/M,t} - p_{(i-1)/M,t}, \text{ where } i = 1, 2, \cdots, M, \tag{23}$$

where $p_{i/M,t}$ is the log price at time i/M on day t. The RV on day t is

$$RV_t\left(M\right) = \sum_{i=1}^{M} r_{i,t}^2. \tag{24}$$

As $M \to \infty$,

$$RV_t\left(M\right) = \sum_{i=1}^{M} r_{i,t}^2 \xrightarrow{p} \int_{t-1}^{t} \sigma_s^2 ds := IV_t, \tag{25}$$

where σ_s^2 is the spot variance. Moreover, according to Barndorff-Nielsen and Shephard (2002), as $M \to \infty$,

$$\sqrt{M}\left(RV_t(M) - IV_t\right) \xrightarrow{d} N(0, 2IQ_t), \tag{26}$$

where

$$IQ_t = \int_{t-1}^{t} \sigma_s^4 ds \tag{27}$$

is the integrated quarticity (IQ).

To improve the accuracy of the asymptotic approximation, Barndorff-Nielsen and Shephard (2005) suggest using the log RV to approximate the log IV and develop the asymptotic distribution for the log RV, that is, as $M \to \infty$,

$$\sqrt{M}\left(\log\left(RV_t(M)\right) - \log\left(IV_t\right)\right) \xrightarrow{d} N(0, 2IQ_t/IV_t^2), \tag{28}$$

The asymptotic theory given by (26) and (28) suggests the presence of heteroskedasticity when approximating IV (or log IV) by RV (or log RV).[5] This is the reason why we would like to relax the iid assumption about w_t.

Unfortunately, for the FCLT to be applicable when the assumption of homoskedasticity is relaxed, a form of strong mixing condition for $\{v_t\}$ is required as a trade-off.

[5] However, IQ_t/IV_t^2 is usually less time varying than IQ_t.

Assumption 4 *$E(w_t) = 0$ for all t. $\{w_t\}$ is independent over t. Let $\sigma_w^2 = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} E(w_t^2)$ exist. There exists $k > 2$ such that $\sup_t E\left|w_t\right|^k < \infty$.*

Assumption 5 *$E(v_t) = 0$ for all t. $\{v_t\}$ is strong mixing with mixing coefficient α_m satisfying $\sum_{m=0}^{\infty} \alpha_m^{1-\frac{2}{k}} < \infty$. There exists $k > 2$ such that $\sup_t E\left|v_t\right|^k < \infty$. Let both $\sigma_v^2 = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} E(w_t^2) > 0$ and $\overline{\sigma}_v^2 = \lim_{T\to\infty} \frac{1}{T} E\left(\sum_{t=1}^{T} v_t\right)^2 > 0$ exist.*

Assumption 6 *w_t and v_s are independent for any t and s.*

Remark 3.1 *Assumption 4 allows $\{\xi_t\}$ to be observed with independent but not necessarily identically distributed errors.*

Remark 3.2 *Assumption 5 allows ξ_t to have an error term that could be serially dependent and heteroskedastic. The assumption, also adopted in Phillips (1987a, 1987b), includes many stationary ARMA models as special cases. According to Phillips (1987a, 1987b), the FCLT of Herrndorf (1983) is applied to the partial sum process of $\{v_t\}$, that is, as $T \to \infty$,*

$$\frac{1}{T^{0.5}\overline{\sigma}_v} \sum_{t=1}^{[Tr]} v_t \Rightarrow W(r).$$

Remark 3.3 *Since $\{v_t\}$ is strong mixing with mixing coefficient α_m, and $\{w_t - \theta_T w_{t-1}\}$ is an m-dependent process by Assumption 4, ε_t defined in (3) is also strong mixing with mixing coefficient α_m due to Assumption 5. Note that*

$$\sigma_\varepsilon^2 := \lim_{T\to\infty} Var\left(\varepsilon_t\right) = \lim_{T\to\infty} \left[E\left(v_t^2\right) + E\left(w_t^2\right) + \theta_T E\left(w_t^2\right)\right] = \sigma_v^2 + 2\sigma_w^2, \tag{29}$$

and that

$$\begin{aligned} \overline{\sigma}_\varepsilon^2 := \lim_{T\to\infty} \frac{1}{T} Var\left(\sum_{t=1}^{T} \varepsilon_t\right) &= \lim_{T\to\infty} \frac{1}{T} E\left(\sum_{t=1}^{T} v_t\right)^2 + \lim_{T\to\infty} \frac{1}{T} E\left(\sum_{t=1}^{T} \left(w_t - \theta_T w_{t-1}\right)\right)^2 \\ &= \overline{\sigma}_v^2 + 2\sigma_w^2 - 2\sigma_w^2 \\ &= \overline{\sigma}_v^2. \end{aligned} \tag{30}$$

Consequently, the FCLT of Herrndorf (1983) is also applicable to the partial sum process of $\{\varepsilon_t\}$.

Remark 3.4 *In Dou and Müller (2021), a generalized local-to-unity model is proposed where the MA component of the model is local-to-unity, that is,* $\varepsilon_t = w_t - \left(1 + \frac{\gamma}{T}\right) w_{t-1}$. *Assuming* $\{w_t\}$ *is iid, we can show that*

$$\overline{\sigma}_{\varepsilon}^2 := \lim_{T\to\infty} \frac{1}{T} Var\left(\sum_{t=1}^{T} \varepsilon_t\right) = \lim_{T\to\infty} \frac{1}{T} Var\left(w_T - w_0 - \frac{\gamma}{T}\sum_{t=1}^{T} w_{t-1}\right) = 0.$$

This suggests that the FCLT cannot be applied to $\{\varepsilon_t\}$ *in their model although the FCLT is applicable to their model with* p *autoregressive roots and* $p-1$ *moving-average roots, all local-to-unity with* $p > 1$.

Lemma 3.1 *Let* $\{y_t\}_{t=0}^{T}$ *be the time series generated by (1). Let Assumptions 4-6 hold. Then, as* $T \to \infty$,

1. $\frac{1}{T^{0.5}\overline{\sigma}_v} y_{[Tr]} \Rightarrow J_c(r)$;
2. $\frac{1}{T^2\overline{\sigma}_v^2} \sum_{t=1}^{T} y_{t-1}^2 \xrightarrow{d} \int_0^1 J_c(r)^2 dr$;
3. $\frac{1}{T\overline{\sigma}_v^2} \sum_{t=1}^{T} y_{t-1}\varepsilon_t \xrightarrow{d} \int_0^1 J_c(r) dW(r) + \frac{1}{2}\left(1 - \frac{\sigma_v^2 + 2\sigma_w^2}{\overline{\sigma}_v^2}\right)$.

Theorem 3.1 *Let* $\{y_t\}_{t=0}^{T}$ *be the time series generated by (1). Let Assumptions 4-6 hold. Then, as* $T \to \infty$,

$$T\left(\widehat{\theta}_T - \theta_T\right) \xrightarrow{d} \frac{\int_0^1 J_c(r)dW(r) + \frac{1}{2}\left(1 - \frac{\sigma_v^2+2\sigma_w^2}{\overline{\sigma}_v^2}\right)}{\int_0^1 J_c(r)^2 dr}. \tag{31}$$

Remark 3.5 *When* $c = 0$, $\theta = 1$. *In this case, the model has a unit root and is a special case of the model considered in Phillips (1987b). When* $c = 0$, $J_0(r) = W(r)$ *and* $\int_0^1 J_0(r)dW(r) = \frac{1}{2}\left(W(1)^2 - 1\right)$ *and hence, the asymptotic theory in (31) can be rewritten as*

$$T\left(\widehat{\theta} - 1\right) \xrightarrow{d} \frac{\frac{1}{2}\left(W(1)^2 - \frac{\sigma_v^2+2\sigma_w^2}{\overline{\sigma}_v^2}\right)}{\int_0^1 W(r)^2 dr}.$$

This is the same as his Theorem 3.1(c).

Remark 3.6 *When* $\sigma_w^2 = 0$, *there is no measurement error. In this case, the model is a special case of the model considered in Phillips (1987a). That is why equation (31) is the same as that in his Theorem 1(a) with* $\sigma_v^2 + 2\sigma_w^2$ *and* $\overline{\sigma}_v^2$ *being the instantaneous variance and long-run variance.*

If both $\{w_t\}$ and $\{v_t\}$ are iid sequences (hence $\overline{\sigma}_v^2 = \sigma_v^2$), then ε_t is an m-dependent process. In this case more efficient estimation of θ_T can be achieved by using valid instrument variables. Following Hall (1989) and Hansen and Lunde (2014), we use y_{t-2} as the instrument variable and consider the following instrumental variable estimator of θ_T, denoted by $\widehat{\theta}_T^{IV}$,

$$\widehat{\theta}_T^{IV} = \frac{\sum_{t=2}^T y_t y_{t-2}}{\sum_{t=2}^T y_{t-1} y_{t-2}} = \theta_T + \frac{\sum_{t=2}^T y_{t-2}\varepsilon_t}{\sum_{t=2}^T y_{t-1} y_{t-2}}.$$

Lemma 3.2 *Let $\{y_t\}_{t=0}^T$ be the time series generated by (1). Let $\{w_t\}$ and $\{v_t\}$ are iid sequences. Then, as $T \to \infty$,*

1. $\frac{1}{T^2\sigma_v^2}\sum_{t=2}^T y_{t-1}y_{t-2} \xrightarrow{d} \int_0^1 J_c(r)^2 dr$;

2. $\frac{1}{T\sigma_v^2}\sum_{t=2}^T y_{t-2}\varepsilon_t \xrightarrow{d} \int_0^1 J_c(r) dW(r)$.

Theorem 3.2 *Let $\{y_t\}_{t=0}^T$ be the time series generated by (1). Let $\{w_t\}$ and $\{v_t\}$ are iid sequences. Then, as $T \to \infty$,*

$$T\left(\widehat{\theta}_T^{IV} - \theta_T\right) \xrightarrow{d} \frac{\int_0^1 J_c(r) dW(r)}{\int_0^1 J_c(r)^2 dr}. \tag{32}$$

Remark 3.7 *Our result in (32) extends the result of Hall (1989) to the local-to-unity case. Compared to the asymptotic theory in (31), the asymptotic distribution of the IV estimator only depends on c, not on the variances of the errors.*

Remark 3.8 *In this case, it is straightforward to show that the asymptotic theory for the OLS estimator given in (31) can be rewritten as*

$$T\left(\widehat{\theta}_T - \theta_T\right) \xrightarrow{d} \frac{\int_0^1 J_c(r) dW(r) + \frac{1}{2}\left(1 - \frac{\sigma_v^2 + 2\sigma_w^2}{\sigma_v^2}\right)}{\int_0^1 J_c(r)^2 dr}. \tag{33}$$

Since $\frac{\sigma_v^2+2\sigma_w^2}{\sigma_v^2} > 1$ as long as $\sigma_w^2 > 0$, compared to the asymptotic distribution in (32), the asymptotic distribution in (33) involves an additional term, $\frac{\frac{1}{2}\left(1-\frac{\sigma_v^2+2\sigma_w^2}{\sigma_v^2}\right)}{\int_0^1 J_c(r)^2 dr}$, which only has the negative support.

Let us conclude this section by relating our model with several interesting models proposed in the literature.

Example 3.1 *The random walk model has been widely used in the literature as the benchmark for examining the forecasting performance. If one assumes the log IV evolves according to the following random walk model*

$$\xi_t = \xi_{t-1} + v_t, v_t \overset{iid}{\sim} N\left(0, \sigma_e^2\right), t = 1, ..., T, \tag{34}$$

and assumes the log IV is related to the log RV by

$$y_t = \xi_t + w_t, w_t \overset{iid}{\sim} N\left(0, \frac{2}{M}\frac{IQ_t}{IV_t^2}\right), \tag{35}$$

then this model is a special case of our model.

Example 3.2 *If one replaces equation (34) by the following local-to-unity AR(1) model*

$$\xi_t = \left(1 + \frac{c}{T}\right)\xi_{t-1} + v_t, v_t \overset{iid}{\sim} N\left(0, \sigma_e^2\right), t = 1, ..., T, \tag{36}$$

and keep equation (35) intact, one would have another special case of our model.

Example 3.3 *The HAR model proposed by Corsi (2008) is of the form*

$$\xi_t = \beta_1\xi_{t-1} + \beta_2\xi_{t-5} + \beta_3\xi_{t-22} + \widetilde{v}_t, \tag{37}$$

where ξ_t is the daily RV. We can rewrite Model (37) as

$$\begin{aligned}\xi_t &= \theta\xi_{t-1} - \beta_2\sum_{j=1}^{4}\left(\xi_{t-j} - \xi_{t-j-1}\right) - \beta_3\sum_{j=1}^{21}\left(\xi_{t-j} - \xi_{t-j-1}\right) + \widetilde{v}_t,\\ &= \theta\xi_{t-1} + v_t,\end{aligned}$$

where $\theta = \beta_1 + \beta_2 + \beta_3$ and

$$v_t = -\beta_2\sum_{j=1}^{4}\left(\xi_{t-j} - \xi_{t-j-1}\right) - \beta_3\sum_{j=1}^{21}\left(\xi_{t-j} - \xi_{t-j-1}\right) + \widetilde{v}_t.$$

Since in practice, $\beta_1, \beta_2, \beta_3 > 0$ and $\beta_1 + \beta_2 + \beta_3$ is close to one, we could assume $\theta = 1 + \frac{c}{T}$. In this case, as $\beta_2 + \beta_3 < 1$, $\{v_t\}$ is stationary and satisfies the strong mixing in Assumption (5). If we assumes ξ_t is daily IV that is related to daily RV by

$$y_t = \xi_t + w_t, w_t \overset{iid}{\sim} N\left(0, \frac{2IQ_t}{M}\right),$$

then the HAR model is a special case of our model.

Example 3.4 *A well-studied SV model is given by*

$$\begin{aligned} r_t &= \sigma e^{\xi_t/2}\varepsilon_t,\ \varepsilon_t \overset{iid}{\sim} N(0,1), \\ \xi_t &= \theta\xi_{t-1} + v_t,\ v_t \overset{iid}{\sim} N\left(0,\sigma_v^2\right). \end{aligned}$$

Let $y_t = \log(r_t^2) = \log\sigma^2 + 1.27 + \xi_t + w_t$ *where* $w_t \overset{iid}{\sim} \log\left(\chi^2_{(1)}\right) - 1.27$*. If* $\theta = 1 + \frac{c}{T}$*, then the model is a special case of our model.*

Example 3.5 *Hansen and Lunde (2014) consider the following model*

$$\begin{aligned} y_t &= \xi_t + w_t, \\ \theta(L)\xi_t &= \varphi(L)v_t, \end{aligned}$$

where $\theta(L)$ *and* $\varphi(L)$ *are lag polynomial of orders* p*,* q *with* $p, q < \infty$*. An important special case considered by Hansen and Lunde (2014) is* $p = 1$*,* $q = 0$*, that is,* $\theta(L) = 1 - \theta L$*,* $\varphi(L) = 1$*. In this case, their model with* $\theta = 1$ *is a special case of our model. Although Hansen and Lunde (2014) is interested in the case where* $\{\xi_t\}$ *is stationary and the case where* $\{\xi_t\}$ *is I(1), their empirical estimates of* θ *are always very close to unity, indicating strong support to our model.*

4 Conclusion

In this paper, we study the asymptotic properties of the LS estimator of the AR(1) parameter in latent local-to-unity models. Two different sets of conditions are considered. In the first class of models, the error term in the observation equation is iid, and the error term in the state equation is a stationary and fractionally integrated sequence. In this case, when if $H \leq 0.5$, the rate of convergence is T; when if $H > 0.5$, the rate of convergence is T^{2H} which is faster. The asymptotic distribution depends on H. The asymptotic distribution has an additional term when $H = 0.5$ than when $H > 0.5$ and two additional terms than when $H < 0.5$. The discontinuity in the asymptotic distribution is due to the order of the sum of squared errors in the AR(1) representation of the model. In the second class of models, the error term in the observation equation is independent and not necessarily identically distributed, and the error term in the state equation is strong mixing. In this case, the rate of convergence is T, and the asymptotic distribution is similar to what Phillips (1987a) obtains for the observed local-to-unity model.

While the models considered in our paper do not have an intercept, such restriction can be lifted by adding an intercept to the measurement equation, that is, by considering the following model,

$$\begin{cases} y_t = \alpha + \xi_t + w_t \\ \xi_t = \theta_T \xi_{t-1} + v_t, \theta_T = 1 + \frac{c}{T}, \xi_0 \sim O_p(1) \end{cases}, t = 0, ..., T, \tag{38}$$

The asymptotic distributions of the LS estimator and the instrumental variable estimator will be slightly different as demeaning is needed in the stochastic integrals. This extension can be done in the same way as how Phillips and Perron (1988) extend the results of Phillips (1987b).

Our model is similar to the class of models recently introduced in Dou and Müller (2021) in the sense that the local-to-unity feature exists both in the autoregressive component and in the moving average component. However, the local-to-unity feature comes from the state-space modeling strategy in our model, and hence, has a natural structural interpretation. It would be interesting to compare the empirical relevance of these two non-nested modeling strategies.

While the LS estimator has an analytical expression in our models, it is not efficient as it ignores the dependence of the error term in the AR(1) representation of the model. The instrumental variable estimator is obviously more efficient than the LS estimator when the two error terms are serially independent, as it is manifest in our asymptotic theory. If, in addition, the distributions of two errors are Gaussian, another efficient estimator can be constructed based on the Kalman filter that maximizes the likelihood function. In unreported simulation studies, we have found evidence that the instrumental variable estimator and the MLE share similar efficiency in finite samples for our model. How to obtain the asymptotic theory for the MLE and what is the relative asymptotic efficiency of the instrumental variable estimator and the MLE are important questions to be investigated.

5 Appendix

Proof of Lemma 2.1

In model (2), under Assumptions (1)-(3), we have $Var\left(\sum_{t=1}^{T} \varepsilon_t\right) \sim T^{2H}\overline{\sigma}_v^2$ as $T \to \infty$ by equation (8). By Assumption 1 and Assumption 2, there exists $k = \max\left\{4, \frac{2}{H} - 4\right\}$ where $H \in 0, 1)$ such that $E\left|\varepsilon_t\right|^k < \infty$. By the FCLT of Davydov

(1970), we have, as $T \to \infty$,

$$\frac{1}{T^H \overline{\sigma}_v} \sum_{t=1}^{[Tr]} \varepsilon_t \Rightarrow B^H(r).$$

From model (2), we have

$$y_t = \theta_T y_{t-1} + \varepsilon_t = \theta_T^t y_0 + \sum_{j=0}^{t-1} \theta_T^j \varepsilon_{t-j} = \theta_T^t y_0 + \sum_{s=1}^{t} \theta_T^{t-s} \varepsilon_s.$$

Let $S_j = \sum_{i=1}^{j} \varepsilon_i$ and, when $\frac{j-1}{T} \leq r < \frac{j}{T}$,

$$X_T(r) = \frac{1}{T^H \overline{\sigma}_v} S_{[Tr]} = \frac{1}{T^H \overline{\sigma}_v} S_{j-1}, \text{ for } j = 1, ..., T.$$

Then, since $y_0 \sim O_p(1)$ and $\theta_T^{[Tr]} \sim O(1)$, we have

$$\begin{aligned}
\frac{1}{T^H \overline{\sigma}_v} y_{[Tr]} &= \frac{1}{T^H \overline{\sigma}_v} \theta_T^{[Tr]} y_0 + \frac{1}{T^H \overline{\sigma}_v} \sum_{s=1}^{[Tr]} \theta_T^{t-s} \varepsilon_s \\
&= \frac{1}{T^H \overline{\sigma}_v} \sum_{s=1}^{[Tr]} \left(1 + \frac{c}{T}\right)^{t-s} \varepsilon_s + O_p\left(T^{-H}\right) \\
&= \sum_{s=1}^{[Tr]} e^{([Tr]-j)c/T} \int_{(j-1)/T}^{j/T} dX_T(s) + O_p\left(T^{-H}\right) + O_p(T^{-1}) \\
&= \int_0^1 e^{(r-s)c} dX_T(s) + O_p\left(T^{-H}\right) \\
&= J_c^H(r) + O_p\left(T^{-H}\right).
\end{aligned}$$

This proves (1) of the lemma. Part (2) of the lemma follows when we applying part (1) and the continuous mapping theorem.

To prove part (3), note that

$$\sum_{t=1}^{T} y_{t-1} \varepsilon_t = \frac{1}{2\theta_T} \left[y_T^2 - y_0^2 - \left(\theta_T^2 - 1\right) \sum_{t=1}^{T} y_{t-1}^2 - \sum_{t=1}^{T} \varepsilon_t^2 \right].$$

By the law of large numbers, for any $H \in (0, 1)$,

$$\frac{1}{T} \sum_{t=1}^{T} \varepsilon_t^2 \xrightarrow{p} \sigma_\varepsilon^2 = \frac{\sigma_e^2 \Gamma(2 - 2H)}{\left(\Gamma(1.5 - H)\right)^2} + 2\sigma_w^2.$$

Hence, $\sum_{t=1}^{T} \varepsilon_t^2 = O_p(T)$ for any $H \in (0, 1)$. However, the order of $\sum_{t=1}^{T} \varepsilon_t^2$ relative to those of y_T^2 and $\left(\theta_T^2 - 1\right) \sum_{t=1}^{T} y_{t-1}^2$ depends on the value of H.

When $H = 0.5$,

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T} y_{t-1}\varepsilon_t &= \frac{1}{2\theta_T}\left[\frac{y_T^2 - y_0^2}{T} - T\left(\theta_T^2 - 1\right)\sum_{t=1}^{T}\left(\frac{y_{t-1}}{\sqrt{T}}\right)^2\frac{1}{T} - \frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2\right] \\ &\xrightarrow{d} \frac{1}{2}\left\{\overline{\sigma}_v^2 J_c(1)^2 - 2c\overline{\sigma}_v^2\int_0^1 J_c(r)dr - \left(\frac{\sigma_e^2\Gamma(2-2H)}{\left(\Gamma(1.5-H)\right)^2} + 2\sigma_w^2\right)\right\}.\end{aligned}$$

When $H > 0.5$,

$$\begin{aligned}\frac{1}{T^{2H}}\sum_{t=1}^{T} y_{t-1}\varepsilon_t &= \frac{1}{2\theta_T}\left[\frac{y_T^2 - y_0^2}{T^{2H}} - T\left(\theta_T^2 - 1\right)\sum_{t=1}^{T}\left(\frac{y_{t-1}}{T^H}\right)^2\frac{1}{T} - \frac{1}{T^{2H}}\sum_{t=1}^{T}\varepsilon_t^2\right] \\ &\xrightarrow{d} \frac{1}{2}\left\{\overline{\sigma}_v^2 J_c^H(1)^2 - 2c\overline{\sigma}_v^2\int_0^1 J_c^H(r)dr\right\}.\end{aligned}$$

When $H < 0.5$,

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T} y_{t-1}\varepsilon_t &= \frac{1}{2\theta_T}\left[\frac{y_T^2 - y_0^2}{T} - T\left(\theta_T^2 - 1\right)\sum_{t=1}^{T}\left(\frac{y_{t-1}}{\sqrt{T}}\right)^2\frac{1}{T} - \frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2\right] \\ &\xrightarrow{d} -\frac{1}{2}\left(\frac{\sigma_e^2\Gamma(2-2H)}{\left(\Gamma(1.5-H)\right)^2} + 2\sigma_w^2\right).\end{aligned}$$

This completes the proof of Lemma 2.1.

Proof of Theorem 2.1

The results in Theorem 2.1 follow directly from the continuous mapping theorem and Lemma 2.1.

Proof of Lemma 3.1

In model (2), under Assumptions (4)-(6), ε_t is strong mixing with mixing coefficient α_m. By Lemma 1 of Phillips (1987a), as $T \to \infty$,

$$\frac{1}{T^{0.5}\overline{\sigma}_v} y_{[Tr]} \Rightarrow J_c(r),$$

$$\frac{1}{T^2\overline{\sigma}_v^2}\sum_{t=1}^{T} y_{t-1}^2 \xrightarrow{d} \int_0^1 J_c(r)^2 dr,$$

and

$$\frac{1}{T\overline{\sigma}_v^2}\sum_{t=1}^{T} y_{t-1}\varepsilon_t \xrightarrow{d} \int_0^1 J_c(r)dW(r) + \frac{1}{2}\left(1 - \frac{\sigma_v^2 + 2\sigma_w^2}{\overline{\sigma}_v^2}\right),$$

since $\sigma_\varepsilon^2 = \sigma_v^2 + 2\sigma_w^2$ and $\overline{\sigma}_\varepsilon^2 = \overline{\sigma}_v^2$. This completes the proof of Lemma 3.1.

Proof of Theorem 3.1

The results in Theorem 3.1 follow directly from the continuous mapping theorem and Lemma 3.1.

Proof of Lemma 3.2

In model (2), when $\{v_t\}$ is an iid sequence, $\overline{\sigma}_v^2 = \sigma_v^2$ and hence, $\overline{\sigma}_\varepsilon^2 = \sigma_v^2$. By Lemma 1 of Phillips (1987a), as $T \to \infty$,

$$\frac{1}{T^{0.5}\sigma_v} y_{[Tr]} \Rightarrow J_c(r).$$

By the above FCLT and the continuous mapping theorem, we have

$$\begin{aligned}
\frac{1}{T^2\sigma_v^2}\sum_{t=1}^{T} y_{t-1}y_{t-2} &= \frac{1}{T^2\sigma_v^2}\sum_{t=1}^{T}\left(\left(1+\frac{c}{T}\right)y_{t-2}+\varepsilon_{t-1}\right)y_{t-2} \\
&= \frac{1}{T^2\sigma_v^2}\sum_{t=1}^{T} y_{t-2}^2 + \frac{c}{T^3\sigma_v^2}\sum_{t=1}^{T} y_{t-2}^2 + \frac{1}{T^2\sigma_v^2}\sum_{t=1}^{T} y_{t-2}\varepsilon_{t-1} \\
&\xrightarrow{d} \int_0^1 J_c(r)^2 dr,
\end{aligned}$$

since $\frac{c}{T^3\sigma_v^2}\sum_{t=1}^{T} y_{t-2}^2 \xrightarrow{p} 0$ and $\frac{1}{T^2\sigma_v^2}\sum_{t=1}^{T} y_{t-2}\varepsilon_{t-1} \xrightarrow{p} 0$.

Moreover,

$$\begin{aligned}
\frac{1}{T\sigma_v^2}\sum_{t=1}^{T} y_{t-2}\varepsilon_t &= \frac{1}{T\sigma_v^2}\sum_{t=1}^{T}\left(\frac{1}{\theta_T}y_{t-1} - \frac{1}{\theta_T}y_{t-1} + y_{t-2}\right)\varepsilon_t \\
&= \frac{1}{\theta_T T\sigma_v^2}\sum_{t=1}^{T} y_{t-1}\varepsilon_t - \frac{1}{\theta_T T^2\sigma_v^2}\sum_{t=1}^{T}\varepsilon_{t-1}\varepsilon_t \\
&= \frac{1}{2\theta_T^2\sigma_v^2}\left\{\frac{y_T^2 - y_0^2}{T} - T\left(\theta_T^2 - 1\right)\sum_{t=1}^{T}\left(\frac{y_{t-1}}{\sqrt{T}}\right)^2\frac{1}{T} - \frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2\right\} - \frac{1}{\theta_T T\sigma_v^2}\sum_{t=1}^{T}\varepsilon_{t-1}\varepsilon_t \\
&\xrightarrow{d} \frac{1}{2}\left\{J_c(1)^2 - 2c\int_0^1 J_c(r)^2 dr - \frac{\sigma_v^2 + 2\sigma_w^2}{\sigma_v^2}\right\} + \frac{\sigma_w^2}{\sigma_v^2} \\
&\stackrel{d}{=} \frac{1}{2}\left\{J_c(1)^2 - 2c\int_0^1 J_c(r)^2 dr - 1\right\} \\
&\stackrel{d}{=} \frac{1}{2}\int_0^1 J_c(r)dW(r),
\end{aligned}$$

where the third last step is due to the FCLT, continuous mapping theorem, and law of large numbers, and the last step is from the well-known relationship between the OU process and the standard Brownian motion (see equation (8) in Phillips (1987a) for example).

Proof of Theorem 3.2

The results in Theorem 3.2 follow directly from the continuous mapping theorem and Lemma 3.2.

References

Andersen, T. G., and Bollerslev, T., 1997. Heterogeneous Information Arrivals and Return Volatility Dynamics: Uncovering the Long-Run in High Frequency Returns. *Journal of Finance* 52(3), 975-1005.

Andersen, T. G., Bollerslev, T., Diebold, F. X., Ebens, H., 2001. The Distribution of Realized Stock Return Volatility. *Journal of Financial Economics* 61(1), 43-76.

Andersen, T. G., Bollerslev, T., Diebold, F. X., Labys, P., 2001. The Distribution of Realized Exchange Rate Volatility. *Journal of the American Statistical Association* 96(453), 42-55.

Andersen, T. G., Bollerslev, T., Diebold, F. X., Labys, P., 2003. Modeling and forecasting realized volatility. *Econometrica* 71(2), 579–625.

Barndorff-Nielsen, O.E., and Shephard, N., 2002. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal Royal Statistical Society. Series B* 64 (2), 253–280.

Barndorff-Nielsen, O.E., and Shephard, N., 2005. How accurate is the asymptotic approximation to the distribution of realized volatility? In *Identification and Inference for Econometric Models*, eds by D.W.K. Andrews, J. Powell, P. Ruud and J. Stock, Cambridge University Press.

Billingsley, P., 1968. *Convergence of Probability Measures.* Wiley, New York.

Bolko, A., Christensen, K., Pakkanen, M. S., Veliyev, B., 2020. Roughness in spot variance? A GMM approach for estimation of fractional log-normal stochastic volatility models using realized measures. Working Paper, Aarhus University.

Bound, J., Brown, C., Mathiowetz, N., 2001. Measurement Error in Survey Data. *Handbook of Econometrics*, Editor(s): J. J. Heckman, E. Leamer, Elsevier, 5, 3705-3843.

Breidt, F. J., Crato, N., de Lima, P., 1998. The detection and estimation of long memory in stochastic volatility. *Journal of Econometrics* 83, 325-348

Campbell, J. Y., and Yogo, M., 2006. Efficient Tests of Stock Return Predictability. *Journal of Financial Economics*, 81, 27-60.

Cavanagh, C. L., G. Elliott, Stock, J. H., 1995. Inference in Models with Nearly Integrated Regressors. *Econometric Theory* 11, 1131-1147.

Chan, N. H, , Wei, C. Z., 1987. Asymptotic Inference for Nearly Nonstationary AR(1) Processes. *Annals of Statistics* 15, 1050-1063.

Cheridito, P., Kawaguchi, H., Maejima, M., 2003. Fractional Ornstein–Uhlenbeck processes. *Electronical Journal Probability* 8(3), 1-14.

Comte, F., Renault, E., 1996. Long memory continuous-time models. *Journal of Econometrics* 73(1), 101–149.

Corsi, F., 2009. A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics* 7(2), 174–196.

Dou, L., and Müller, U., 2021, Generalized Local-to-Unity Models. *Econometrica* forthcoming.

Davydov, Y. A., 1970. The Invariance Principle for Stationary Processes, *Theory of Probability and Its Applications* 15, 487-489.

Elliott, G., Rothenberg, T. J., Stock, J. H., 1996. Efficient Tests for an Autoregressive Unit Root, *Econometrica* 64, 813-836.

Elliott, G., and Stock, J. H., 2001. Confidence Intervals for Autoregressive Coefficients Near One. *Journal of Econometrics* 103, 155-181.

Fukasawa, M., Takabatake, T., Westphal, R., 2021. Consistent estimation for fractional stochastic volatility model under high-frequency asymptotics. *Mathematical Finance* forthcoming.

Gatheral, J., Jaisson, T., Rosenbaum, M., 2018. Volatility is rough. *Quantitative Finance* 18 (6), 933-949.

Gospodinov, N., 2004. Asymptotic Confidence Intervals for Impulse Responses of Near Integrated Processes. *Econometrics Journal* 7, 505-527.

Hall, A., 1989. Testing for a unit root in the presence of moving average errors. *Biometrika* 76, 49-56.

Hansen, P., and Lunde, A., 2014. Estimating the Persistence and the Autocorrelation Function of a Time Series that is Measured with Error. *Econometric Theory*, 30(1), 60-93.

Herrndorf, N., 1983. Stationary Strongly Mixing Sequences Not Satisfying the Central Limit Theorem. *Annals of Probability* 11, 809-813.

Heston S., 1993. A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options. *Review of Financial Studies*, 6, 327–343.

Jansson, M., and Moreira, M. J. , 2006. Optimal Inference in Regression Models with Nearly Integrated Regressors. *Econometrica* 74, 681-714.

Jiang, L., Wang, X., Yu, J., 2021. In-fill Asymptotic Theory for Structural Break Point in Autoregression. *Econometric Reviews*, 40, 359-386.

Kasprzyk, D., 2005. Measurement error in household surveys: sources and measurement. *Household Sample Surveys in Developing and Transition Countries* Chapter 5, 171-198.

Liu, X., Shi, S., Yu, J., 2020. Persistent and Rough Volatility, Working Paper, SMU.

Mikusheva, A., 2007. Uniform Inference in Autoregressive Models, *Econometrica* 75, 1411-1452.

Nelson D.B., 1990. ARCH Models as Diffusion Approximations, *Journal of Econometrics*, 45, 7–38

Phillips, P. C. B., 1987a. Toward a unified asymptotic theory for autoregression. *Biometrika* 74, 533-547.

Phillips, P. C. B., 1987b. Time series regression with a unit root. *Econometrica* 55, 277-301.

Phillips, P. C. B., 1988. Regression Theory for Near-Integrated Time Series. *Econometrica* 56, 1021-1043.

Phillips, P. C. B., and Perron, P., 1988. Testing for a Unit Root in Time Series Regression. *Biometrika* 75, 335-346.

Rossi, B., 2005. Confidence Intervals for Half-Life Deviations from Purchasing Power Parity. *Journal of Business & Economic Statistics* 23, 432-442.

Sowell, F., 2000. The Fractional Unit Root Distribution, *Econometrica* 58, 495-505.

Stock, J. H., 1991. Confidence Intervals for the Largest Autoregressive Root in U.S. Macroeconomic Time Series. *Journal of Monetary Economics* 28, 435-459.

Tanaka, K., 2013. Distributions of the maximum likelihood and minimum contrast estimators associated with the fractional Ornstein–Uhlenbeck process. *Statistical Inference for Stochastic Processes* 16 (3), 173–192.

Torous, W., Valkanov, R., Yan, S., 2004. On predicting stock returns with nearly integrated explanatory variables. *The Journal of Business* 77, 937-966.

Valkanov, R., 2003. Long-Horizon Regressions: Theoretical Results and Applications, *Journal of Financial Economics* 68, 201-232.

Wang, X., Xiao, W., Yu, J., 2019. Estimation and inference of fractional continuous-time model with discrete-sampled data. Working Paper, SMU.

Wang, X., Xiao, W., Yu, J., 2021. Asymptotic Properties of Least Squares Estimator in Local to Unity Processes with Fractional Gaussian Noises. Working paper, SMU.

Wright, J. H., 2000. Confidence Intervals for Univariate Impulse Responses with a Near Unit Root. *Journal of Business & Economics Statistics* 18, 368-373.