Taylor & Francis
Taylor & Francis Group

Econometric Reviews

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/lecr20

Latent local-to-unity models

Xiaohu Wang & Jun Yu

To cite this article: Xiaohu Wang & Jun Yu (2023) Latent local-to-unity models, Econometric Reviews, 42:7, 586-611, DOI: 10.1080/07474938.2023.2215034

To link to this article: https://doi.org/10.1080/07474938.2023.2215034

Published online: 29 Jun 2023.

ECONOMETRIC REVIEWS
2023, VOL. 42, NO. 7, 586–611
https://doi.org/10.1080/07474938.2023.2215034

Latent local-to-unity models

Xiaohu Wang[a] and Jun Yu[b,#]

[a]School of Economics, Fudan University, and Shanghai Institute of International Finance and Economics, Shanghai, China; [b]Singapore Management University, Singapore

ABSTRACT
The article studies a class of state-space models where the state equation is a local-to-unity process. The parameter of interest is the persistence parameter of the latent process. The large sample theory for the least squares (LS) estimator and an instrumental variable (IV) estimator of the persistent parameter in the autoregressive (AR) representation of the model is developed under two sets of conditions. In the first set of conditions, the measurement error is independent and identically distributed, and the error term in the state equation is stationary and fractionally integrated with memory parameter $d \in (-0.5, 0.5)$. For both estimators, the convergence rate and the asymptotic distribution crucially depend on d. The LS estimator has a severe downward bias, which is aggravated even more by the measurement error when $d \leq 0$. The IV estimator eliminates the effects of the measurement error and reduces the bias. In the second set of conditions, the measurement error is independent but not necessarily identically distributed, and the error term in the state equation is strongly mixing. In this case, the IV estimator still leads to a smaller bias than the LS estimator. Special cases of our models and results in relation to those in the literature are discussed.

ARTICLE HISTORY
Received 24 March 2022
Accepted 06 May 2023

KEYWORDS
Fractional Brownian motion; fractionally integrated errors; fractional O-U process; local-to-unity Instrumental variable; state-space.

JEL CLASSIFICATION:
C12, C22, G01.

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 tremendous attention in theoretical and empirical studies.[1]. The success of the local-to-unity model is 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 and financial time series; and (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 errors. This assumption can be too strong in practice. For example, when a time series is obtained from a survey, errors of many types are possible, such as 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. For another example, a time series is sometimes obtained from estimation. A well-known example that

CONTACT Jun Yu yujun@smu.edu.sg School of Economics and Lee Kong Chian School of Business, Singapore Management University, 90 Stamford Road, Singapore 178903.
Supplemental data for this article can be accessed online at https://doi.org/10.1080/07474938.2023.2215034.
#Present address: Faculty of Business Administration, University of Macau, China.
[1]An incomplete list of contributions includes Phillips (1988), Stock (1991), Cavanagh et al. (1995), Elliott et al. (1996), Wright (2000), Elliott and Stock (2001), Valkanov (2003), Gospodinov (2004), Torous et al. (2004), Rossi (2005), Campbell and Yogo (2006), Jansson and Moreira (2006), Buchmann and Chan (2007), Mikusheva (2007), Wang et al. (2023a); Wang et al. (2023b), Jiang et al. (2021), Dou and Müller (2021)

motivates this article is the daily time series of realized volatilities (RV), which are estimates of the daily integrated volatilities. 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 family of stochastic volatility (SV) models are among the interesting models in this example.

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

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

where μ is a constant, $\{\xi_t\}$ 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 the measurement error $\{w_t\}$ and also for $\{v_t\}$.

The main model considered in the article assumes that $\{w_t\}$ is independent and identically distributed (i.i.d.), and $\{v_t\}$ forms a fractionally integrated series with an order $d \in (-0.5, 0.5)$, that is, an I(d) process. The I(d) process with $d \in (0, 0.5)$ has positive serial correlations and an infinite long-run variance and has been widely applied to model long-memory time series in economics and finance; see, for example, Granger and Joyeux (1980), Baillie (1996), and Andersen et al. (2003). The interaction between cross-sectional/temporal aggregation and long-range dependence has been investigated by Andersen and Bollerslev (1997), Chambers (1998), and Lieberman and Phillips (2008). In contrast, when $d \in (-0.5, 0)$, $\{v_t\}$ becomes an antipersistent process that has negative serial dependence and zero long-run variance. As argued in Shi and Yu (2022) and Li et al. (2022), the local-to-unity process $\{\xi_t\}$ with antipersistent errors can also have positive autocorrelations decaying very slowly and, therefore, is capable of describing time series with long-range dependence. The models of this type share the same spirit as rough volatility models, which are becoming increasingly popular in modeling spot volatility and realized variance in the literature; see, for example, Gatheral et al. (2018) and Wang et al. (2023a). Moreover, when $c = 0$, the latent process ξ_t becomes an nonstationary $I(d^*)$ process with $d^* = d + 1 \in (0.5, 1.5)$. Diebold and Rudebusch (1989) have found evidence of $d^* = 0.68$ for the U.S. quarterly real GNP per capita.[2]

Our interest is in the estimation and inference of θ_T, which captures the persistence level of the latent process. Model (1) can be rewritten as a first-order autoregression (AR) with an intercept:

$$y_t = \alpha + \theta_T y_{t-1} + \varepsilon_t, \tag{2}$$

where $\alpha = (1 - \theta_T)\,\mu$ and

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

A simple way to estimate θ_T is via the least squares (LS) method based on the observations of $\{y_t\}$. Denote the LS estimator by $\hat{\theta}_T$. First, we will show that $\hat{\theta}_T$ is consistent, but both the convergence rate and the asymptotic distribution depend on the value of d. Second, we will show that $\hat{\theta}_T$ has a severe downward bias when $d \in (-0.5, 0]$. When $w_t \neq 0$, the variance of w_t appears in the limiting distribution of $\hat{\theta}_T$ and deteriorates the bias problem further. Third, when $d \in (0, 0.5)$, we will show that the limiting distribution is not affected by the existence of the measurement error $\{w_t\}$.

To avoid the influence of the measurement error $\{w_t\}$ and reduce the bias in estimating θ_T, we propose the use of the instrumental variable (IV) estimator, denoted by $\widehat{\theta}_T^{IV}$, with y_{t-2} as the IV. It is shown that the limiting distribution of $\widehat{\theta}_T^{IV}$ is not affected by $\{w_t\}$. In addition, when $d \in (-0.5, 0]$, $\widehat{\theta}_T^{IV}$ not only gets rid of the bias caused by the measurement error but also diminishes the bias induced by the serial dependence in $\{v_t\}$. In other words, even when the latent process ξ_t can be observed directly, $\widehat{\theta}_T^{IV}$ still

[2]Beran (1992) provides a nice review of results on estimation and statistical inference of the process with long-range dependence.

outperforms $\widehat{\theta}_T$ in terms of bias. Furthermore, when $d \in (0, 0.5)$, $\widehat{\theta}_T^{IV}$ leads to the same asymptotic theory as $\widehat{\theta}_T$.

For completeness of the theory and also for the comparison purpose, we study another model, where $\{v_t\}$ is a strongly mixing process and $\{w_t\}$ and $\{v_t\}$ have heteroskedasticity. Similar results are found in this model to those in the main model. The measurement error aggravates the bias of the LS estimator $\widehat{\theta}_T$ but plays no role in the asymptotics of the IV estimator $\widehat{\theta}_T^{IV}$. However, $\widehat{\theta}_T^{IV}$ may have more or less bias than $\widehat{\theta}_T$, depending on the sign of the probability limit of $\sum_{t=2}^T v_t v_{t-1}$. This result reveals the fact that the particular dependence structure of the I(d) process is the reason why $\widehat{\theta}_T^{IV}$ outperforms $\widehat{\theta}_T$ in the first model.

Our study is closely related to Hansen and Lunde (2014), where the latent process $\{\xi_t\}$ in (1) is assumed to be ARMA(p, q). They prove that the IV estimator of the persistence parameter with y_{t-j}, $j \geq \max\{p, q\}$, as an instrument, is consistent and can purge the influence of the errors $\{w_t\}$ in the observation equation. This article extends the model of Hansen and Lunde (2014) with a weakly dependent structure to that with a strongly dependent structure. Our results show that even with the I(d) error, the IV estimator with y_{t-2} as an instrument can eliminate the effects of the measurement error and, more importantly, can further reduce the bias.

The rest of the article is organized as follows. Section 2 introduces the model where $\{v_t\}$ is an $I(d)$ process and $\{w_t\}$ is an i.i.d. sequence. The large sample theory of the LS estimator and the IV estimator of θ_T is provided. Section 3 studies the model in which $\{v_t\}$ is strongly mixing and $\{w_t\}$ is an independent but not necessarily identically distributed sequence. Both $\{v_t\}$ and $\{w_t\}$ allow for heteroskedasticity. Simulation studies are presented in Section 4 to examine the finite sample performance of the derived asymptotic distributions and to compare the performance of the LS estimator and the IV estimator of θ_T. Section 5 concludes. Appendix collects the proof of the theoretical results. Throughout the article, we use $\xrightarrow{p}$, $\Rightarrow$, $\xrightarrow{d}$, $\overset{d}{=}$, $\overset{iid}{\sim}$ to denote convergence in probability, weak convergence, convergence in distribution, equivalence in distribution, and independent and identically distributed, respectively. The weak convergence is defined in the space of càdlàg functions equipped with the Skorokhod topology.

2. Latent model with $I(d)$ errors

In this section, we first introduce our main model of the article and connect it with some popular models in the volatility literature. Then, we focus on estimating the persistence parameter of the latent process and provide the large sample theory of the LS estimator and an IV estimator.

2.1. The model and motivations

Consider the latent local-to-unity model defined in (1) with the following assumptions for the error series $\{w_t\}$ and $\{v_t\}$.

Assumption 1. $w_t \overset{iid}{\sim} (0, \sigma_w^2)$.

Assumption 2. *v_t is an $I(d)$ process with $d \in (-0.5, 0.5)$, i.e., $(1-L)^d v_t = e_t \overset{iid}{\sim} (0, \sigma_e^2)$. There exists $k = \max\left\{4, \frac{2}{d+0.5} - 4\right\}$ such that $E|e_t|^k < \infty$.*

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

Remark 2.1. Assumption 1 assumes that the latent process $\{\xi_t\}$ is observed with i.i.d. errors. Assumption 2 takes that the error sequence of the latent process, $\{v_t\}$, is a fractionally integrated process with order $d \in (-0.5, 0.5)$. It means that

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

where $\Gamma(\cdot)$ denotes the Gamma function. The process $\{v_t\}$ is stationary and ergodic for all values of $d \in (-0.5, 0.5)$. The autocovariance function is given by

$$Cov\left(v_t, v_{t-k}\right) = \begin{cases} 0 \quad \text{for all} \quad k \neq 0, \text{if } d = 0 \\ \frac{\Gamma(1-2d)\Gamma(d+k)}{\Gamma(d)\Gamma(2-d)\Gamma(1-d+k)} \sim |k|^{2d-1} \quad \text{for large } |k|, \text{ if } d \neq 0 \end{cases}.$$

Note that when $d = 0$, $\{v_t\}$ becomes an i.i.d. sequence. Whereas when $d \neq 0$, $\{v_t\}$ has serial dependence at all lags. If $d > 0$, the autocovariance of v_t at any lag is positive. Additionally, the autocovariances decay more slowly than $|k|^{-1}$ and are not summable. Hence, $I(d)$ processes with $d \in (0, 0.5)$ are often adopted to model long-memory time series. In contrast, if $d < 0$, $\{v_t\}$ has negative autocovariance for any $k \neq 0$, which decays faster than $|k|^{-1}$ as $k \to \infty$. Moreover, the sum of autocovariances equals zero, and $\{v_t\}$ is called an antipersistent process. In this case, however, the autocovariance of the latent process ξ_t can still be positive and decay slowly at small and moderate lags due to the local-to-unity AR root (Li et al., 2022).

The process $\{\xi_t\}$ with no observation errors has been extensively studied in the literature. For example, Diebold and Rudebusch (1989) apply the process $\{\xi_t\}$ with $\theta_T = 1$ and $d = -0.32$ to model the U.S. quarterly real GNP per capita. Sowell (1990) develops the large sample theory of the LS estimator of θ_T when $\theta_T = 1$ and no intercept is involved in the regression. Buchmann and Chan (2007) extend Sowell's results to the local-to-unity case.

In the following, we motivate our study by linking our model to some existing models in the volatility literature. It is known that RV serves as a consistent estimate of the integrated variance as the sampling frequency of the return observations shrinks to zero. A well-established stylized fact is the daily log RV series has a slowly decaying autocorrelation function (ACF). A standard procedure to model the daily log RV series is to use a fractional process, namely $I(d)$ with $d \in (0, 0.5)$. Significant contributions include Andersen et al. (2001a, 2001b) and 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^{2d-1} with $d \in (0, 0.5)$ for large k so that the ACF is not absolutely summable) from the interactions of a large number of heterogeneous information processes.[3] The interaction between aggregation and long-range dependence is also explored in Chambers (1998) and Lieberman and Phillips (2008). Andersen et al. (2003) introduce the ARFIMA$(1, d, 0)$ model for log RV and provide evidence of $d \in (0, 0.5)$ (i.e., long-memory errors) and weak short-run dynamic based on a semiparametric estimation method. They also provide evidence that the ARFIMA$(1, d, 0)$ model with long-memory errors and weak short-run dynamics outperforms many alternative models in predicting RV and log RV, including GARCH-type models and other high-frequency models.

However, a recent attempt to model daily log RV series is to use the local-to-unity model with antipersistent errors. Shi and Yu (2022) find that many daily log RV series are better fitted by the local-to-unity model with $I(d)$ errors where $d \in (-0.5, 0)$. Li et al. (2022) argue that there is a weak identification issue between the pure $I(d)$ process with $d > 0$ and the local-to-unity process with $I(d)$ errors where $d < 0$. A debate about how to model slow-decaying ACFs of log RV begins with Gatheral et al. (2018), where the new stylized fact termed "roughness" is established for the sample path of RV. The partial sum of an $I(d)$ process converges weakly to the fractional Brownian motion (fBm), denoted by $B^H(t)$ with $H = d + 0.5$ being the Hurst parameter. Comte and Renault (1996) model the spot volatility of asset price by using the fractional Ornstein-Uhlenbeck (fO-U) process driven by $B^H(t)$ with $H > 0.5$ (i.e., $d > 0$).

[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.

Gatheral et al. (2018) point out that Comte and Renault's model tends to generate smoother sample paths than the actual daily RV series. Gatheral et al. (2018) advocate the usage of fBm with $H < 0.5$ (i.e., $d < 0$). This model generates sample paths that are rougher than that from the standard Brownian motion. In the empirical studies, Gatheral et al. use fBm with $H = 0.14$ to model log RV and forecast RV and log RV out-of-sample. It is shown that fBm with $H = 0.14$ performs better than the AR(5), AR(10), and HAR models. The fBm with $H = 0.14$ corresponds to a unit root process with antipersistent errors in discrete time.

Wang et al. (2023a) apply the fO-U process to model log RV. Under the infill sampling scheme, the fO-U model is a local-to-unity model with fractionally integrated errors. They introduce a two-stage method to estimate parameters and develop large sample properties of the estimators. When applying the approach to the daily log RV, daily log realized kernel (RK), and daily log bipower variation (BV), they find strong evidence of $H \in (0.05, 0.25) < 0.5$ (once again antipersistent errors), although $H > 0.5$ is also allowed in their model. This finding supports that the log RV follows a local-to-unity model with antipersistent errors. In terms of out-of-sample forecasting performance, their empirical studies show that the fO-U model outperforms the random walk, AR(1), HAR, ARFIMA, and fBm in predicting the daily RV, log RV, log RK, and log BV.

Instead of modeling RV or log RV directly, Bolko et al. (2023) assume the log spot variance follows an fO-U and take the estimation errors in the daily RV seriously. In particular, under the infill sampling scheme, the expressions for moments of the daily RV are obtained, facilitating the implementation of the generalized method of moments (GMM). When applying the GMM method to real data, Bolko et al. (2023) report strong evidence of $H \in (0.01, 0.2) < 0.5$. Therefore, Bolko et al. (2023) present strong evidence of local-to-unity and antipersistent errors in log spot variance.

Motivated by the ongoing debate about how to model the daily log RV series, together with the possibility that the integrated variance of an asset is highly persistent and the fact that daily RV is an approximation to daily integrated variance, we consider the latent model defined in (1) under Assumptions 1-3. Some concrete examples are given below to relate our model with several popular models proposed in the volatility literature.

Example 2.1. Breidt et al. (1998) propose the following long-memory SV model:

$$r_t = \sigma e^{\xi_t/2}\varepsilon_t, \ \ \varepsilon_t \overset{iid}{\sim} N(0,1),$$
$$\xi_t = \theta\xi_{t-1} + v_t, \ \ (1-L)^d v_t = e_t \overset{iid}{\sim} N\left(0,\sigma_e^2\right).$$

where r_t is the return of a financial asset. It is easy to get that

$$y_t := \log\left(r_t^2\right) = \mu + \xi_t + w_t,$$

where $\mu = \log\sigma^2 + 1.27$ and $w_t = \log\left(\varepsilon_t^2\right) - 1.27 \overset{iid}{\sim} \log\left(\chi^2_{(1)}\right) - 1.27$ with $\chi^2_{(1)}$ standing for the chi-squared distribution with one degree of freedom. By allowing $\theta = \theta_T = 1 + c/T$, the model becomes a special case of ours.

Example 2.2. Liu et al. (2021) propose the following fractional SV model:

$$r_{t\Delta} = \sigma e^{\xi_{t\Delta}/2}\varepsilon_{t\Delta}, \ \varepsilon_{t\Delta} \overset{iid}{\sim} N(0,1),$$
$$\xi_{t\Delta} = (1+\gamma\Delta)\,\xi_{(t-1)\Delta} + \sigma_v\eta^H_{t\Delta}, \quad H \in (0,1)$$

where $\{r_{t\Delta}\}$ is a return series with Δ denoting the sampling frequency and $\eta^H_{t\Delta} := B^H(t\Delta) - B^H((t-1)\,\Delta)$ is the first-order difference of the fBm $B^H(t\Delta)$. $\left\{\eta^H_{t\Delta}\right\}$ is often called a fractional Gaussian noise (FGN). Again, it is easy to get that $y_{t\Delta} := \log\left(r^2_{t\Delta}\right) = \mu + \xi_{t\Delta} + w_t$, where $\mu = \log\sigma^2 + 1.27$ and $w_t \overset{iid}{\sim} \log\left(\chi^2_{(1)}\right) - 1.27$. If we normalize the time span of the data to be one, then it has $\Delta = 1/T$ and $1 + \gamma\Delta = 1 + \gamma/T$ with T being the sample size. Hence, the latent process $\xi_{t\Delta}$ has a local-to-unity root as $T \to \infty$. Because the FGN $\left\{\eta_{t\Delta}^{H}\right\}$ has an autocovariance function that converges to zero at the same rate as those of the $I(d)$ process with $d = H - 0.5$ (see, e.g., Samorodnitsky and Taqqu, 1994), the results developed for our model can be easily extended to the fractional SV model of Liu et al. (2021).

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

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

Note that RV is an observed proxy of the integrated variance. Hence, if X_t is the log integrated variance series, we have the model

$$\begin{aligned} \log RV_t &= X_t + w_t = \mu + \xi_t + w_t, \\ \xi_t &= \theta \xi_{t-1} + v_t, \end{aligned}$$

where w_t denotes the observation error and $\xi_t := X_t - \mu$. By further allowing $\theta = \theta_T = 1 + c/T$, our model in (1) is obtained.

Example 2.4. Wang et al. (2023a) apply the exact discretization of an fO-U process to model log RV:

$$X_{t\Delta} = e^{-\kappa\Delta} X_{(t-1)\Delta} + \left(1 - e^{-\kappa\Delta}\right)\mu + \varepsilon_{t\Delta},\ \text{with}\ \ \varepsilon_{t\Delta} = \sigma \int_{(t-1)\Delta}^{t\Delta} e^{-\kappa(t\Delta - s)} dB^H(s),$$

where κ, μ, and σ are three constants, and Δ denotes the sampling frequency. Again, if $X_{t\Delta}$ is the log integrated variance series, we have the model

$$\begin{aligned} \log RV_{t\Delta} &= X_{t\Delta} + w_t = \mu + \xi_{t\Delta} + w_t, \\ \xi_{t\Delta} &= e^{-\kappa\Delta}\xi_{t-1} + \varepsilon_{t\Delta}, \end{aligned}$$

where w_t denotes the observation error and $\xi_t := X_t - \mu$. Cheridito et al. (2003) show that the autocovariance of $\{\varepsilon_{t\Delta}\}$ decays to zero at the same rate as that of the $I(d)$ process with $d = H - 0.5$, as the lag number increases. Hence, the results developed for our model can be easily extended to the new model.

Example 2.5. Comte and Renault (1996) specify the following fO-U model with $H > 0.5$ for the log spot variance:

$$dX(t) = -\kappa X(t)dt + \sigma dB^H(t), \tag{4}$$

while Gatheral et al. (2018) propose $H < 0.5$. The discretization of the fO-U process with Δ being the sampling frequency is

$$X_{t\Delta} = e^{-\kappa\Delta} X_{(t-1)\Delta} + \varepsilon_{t\Delta},\ \text{with}\ \ \varepsilon_{t\Delta} = \sigma \int_{(t-1)\Delta}^{t\Delta} e^{-\kappa(t\Delta - s)} dB^H(s),$$

which is the model discussed in Example 2.4 with $\mu = 0$. It is also well-known in the literature that the following discrete-time model

$$\xi_{t\Delta} = e^{-\kappa\Delta}\xi_{(t-1)\Delta} + (1-L)^{-d} e_{t\Delta}, e_{t\Delta} \overset{iid}{\sim} \left(0, \frac{1-e^{-2\kappa\Delta}}{2\kappa}\sigma^2\right), t = 1..., T, \tag{5}$$

with $d = H - 0.5$, weakly converges to model (4), that is, $\frac{\delta_H \Gamma(H+0.5)}{T^H}\xi_{\lfloor Tr \rfloor} \Rightarrow X(r)$ as $\Delta \to 0$, where δ_H is a function of H and $T = 1/\Delta$ is the sample size (Tanaka, 2013). Clearly, model (5) is a local-to-unity model with the error term satisfying Assumption 2. Since the spot variance is only observed with errors, our model in (1) is useful for studying the dynamics of the spot variance.

2.2. Large sample theory

Throughout this article, our parameter of interest is θ_T that controls the persistence level of the latent process ξ_t. It is straightforward to rewrite Model (1) as an AR model:

$$y_t = \alpha + \theta_T y_{t-1} + \varepsilon_t,$$

where $\alpha = (1 - \theta_T)\,\mu$ and

$$\varepsilon_t = v_t + w_t - \theta_T w_{t-1}.$$

There are two simple ways to estimate θ_T, the LS method and the IV method with y_{t-2} as an instrument. The resulting estimators are

$$\widehat{\theta}_T = \frac{\sum_{t=1}^{T}\left(y_{t-1} - \overline{y}_{-1}\right) y_t}{\sum_{t=1}^{T}\left(y_{t-1} - \overline{y}_{-1}\right)^2}, \tag{6}$$

and

$$\widehat{\theta}_T^{IV} = \frac{\sum_{t=2}^{T}\left(y_{t-2} - \overline{y}_{-2}\right) y_t}{\sum_{t=2}^{T}\left(y_{t-2} - \overline{y}_{-2}\right) y_{t-1}}, \tag{7}$$

respectively, where $\overline{y}_{-1} = T^{-1}\sum_{t=1}^{T} y_{t-1}$ and $\overline{y}_{-2} = (T-1)^{-1}\sum_{t=2}^{T} y_{t-2}$. These two estimators have closed-form expressions and are hence easy to apply.[4]

When both $\{w_t\}$ and $\{v_t\}$ are i.i.d. series, the error term ε_t becomes the sum of a white noise process and a first-order moving average (MA(1)) process. In this case, we have

$$Corr(\varepsilon_t, \varepsilon_{t-1}) = \frac{-\theta_T \sigma_w^2}{\sigma_v^2 + \left(1 + \theta_T^2\right)\sigma_w^2} \to -\frac{1}{\sigma_v^2/\sigma_w^2 + 2}, \quad \text{as } T \to \infty,$$

and

$$Cov(\varepsilon_t, \varepsilon_{t-k}) = Corr(\varepsilon_t, \varepsilon_{t-k}) = 0 \text{ for all } k = 2, 3....$$

where $Corr(\cdot,\cdot)$ denotes the correlation function. Hence, ε_t is an MA(1) process with a negative root, which is due to the measurement error $\{w_t\}$. Note that when $\sigma_v^2/\sigma_w^2 \to 0$, $Corr(\varepsilon_t, \varepsilon_{t-1}) \to -1/2$. This means that as the ratio between the variance of the signal v_t and that of the noise w_t shrinks to zero, the root of the MA(1) process ε_t goes to minus unity. How to test for a unit root when the error term follows an MA(1) process with a negative root has received much attention in the unit root literature; see, for example, Schwert (1989) and Ng and Perron (2001). These studies show that the negative root in the MA(1) process makes the LS estimator of the AR(1) coefficient severely downward biased, leading to severe size distortions for the conventional Dickey-Fuller and augmented Dickey-Fuller tests. Hall (1989) proposes to use the IV estimator with y_{t-2} as an instrument to purge the influence of the negative MA root. Hansen and Lunde (2014) develop the asymptotic theory of the IV estimator with $y_{t-j}, j \geq \max\left(p, q\right)$, as an instrument when y_t is an ARMA$\left(p, q\right)$ process. Dou and Müller (2021) studied a generalized local-to-unity model that allows y_t to be an ARMA $(p, p-1)$ process with p AR roots close to unity and $p-1$ MA roots close to negative unity.

Our study contributes to the literature in two aspects. First, by allowing $\{v_t\}$ to be an $I(d)$ process, we show that the influence of $\{w_t\}$ on the LS estimator depends on the value of d. When $d \leq 0$, the existence of $\{w_t\}$ deteriorates the bias and increases the difficulty of making inferences by involving additional nuisance parameters in the limiting distribution. Whereas, when $d > 0$, $\{w_t\}$ plays no role in the asymptotics of the LS estimator. Second, we show that when $\{v_t\}$ is an $I(d)$ process, the IV estimator with y_{t-2} as an instrument can remove the impact of $\{w_t\}$ successfully. More importantly, even in the case without measurement error $\{w_t\}$, $\widehat{\theta}_T^{IV}$ continues to outperform the LS estimator $\widehat{\theta}_T$ in terms of bias if $d \leq 0$. When $d > 0$, $\widehat{\theta}_T^{IV}$ and $\widehat{\theta}_T$ share the same asymptotic distribution.

[4]The maximum likelihood estimator, although could be more efficient, does not lead to an analytical expression.

Before presenting the large sample theory of the estimators, we first introduce some notations that will be heavily used in the rest of this subsection. The $I(d)$ process $\{v_t\}$ has a variance of

$$\sigma_v^2 := Var(v_t) = \frac{\sigma_e^2 \Gamma(1-2d)}{\{\Gamma(1-d)\}^2}. \tag{8}$$

Sowell (1990) has shown that

$$\begin{aligned} Var\left(\sum_{t=1}^{T} v_t\right) &= \frac{\sigma_e^2 \Gamma(1-2d)}{(2d+1)\,\Gamma(1+d)\Gamma(1-d)}\left[\frac{\Gamma(1+d+T)}{\Gamma(1-d+T)} - \frac{\Gamma(1+d)}{\Gamma(1-d)}\right] \\ &\sim T^{(2d+1)}\left[\frac{\sigma_e^2 \Gamma(1-2d)}{(2d+1)\,\Gamma(1+d)\Gamma(1-d)}\right] := T^{(2d+1)}\overline{\sigma}_v^2, \end{aligned} \tag{9}$$

where

$$\overline{\sigma}_v^2 = \frac{\sigma_e^2 \Gamma(1-2d)}{(2d+1)\,\Gamma(1+d)\Gamma(1-d)}$$

is often referred to as the long-run variance of v_t in the literature. According to Davydov (1970, Theorem 2) and Taqqu (1975, Theorem 2.1), under Assumption 2, the following functional central limit theorem (FCLT) holds:

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

where $H = d + 0.5$, $\lfloor Tr \rfloor$ denotes the integer part of Tr for any $r \in [0, 1]$ and $B^H(r)$ is an fBm that is a Gaussian process with mean zero and covariance function

$$Cov\left(B^H(t), B^H(s)\right) = \frac{1}{2}\left(|t|^{2H} + |s|^{2H} - |t-s|^{2H}\right), \ \ \forall t, s. \tag{10}$$

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, $W(r)$. We further define the O-U process $J_c(t)$ by the stochastic diffusion function of

$$dJ_c(t) = cJ_c(t)dt + dW(t), \ J_c(0) = 0, \tag{11}$$

and the fO-U process $J_c^H(t)$ by

$$dJ_c^H(t) = cJ_c^H(t)dt + dB^H(t), \ J_c^H(0) = 0, \tag{12}$$

where c is the local parameter in $\theta_T = 1 + c/T$.

Theorem 2.1 gives the large sample theory of the LS estimator $\widehat{\theta}_T$.

Theorem 2.1. *Let $\{y_t\}_{t=0}^T$ be the time series generated by (1), $\widehat{\theta}_T$ be the LS estimator defined in (6). Let Assumptions 1-3 hold. Then, as $T \to \infty$, it has*

$$T^{(2d+1)}\left(\widehat{\theta}_T - \theta_T\right) \xrightarrow{d} \frac{-\sigma_v^2/2 - \sigma_w^2}{\overline{\sigma}_v^2\left(\int_0^1 J_c^H(r)^2 dr - \left(\int_0^1 J_c^H(r)dr\right)^2\right)}, \text{ if } d < 0; \tag{13}$$

$$T\left(\widehat{\theta}_T - \theta_T\right) \xrightarrow{d} \frac{\int_0^1 \bar{J}_c(r)dW(r) - \sigma_w^2/\sigma_e^2}{\int_0^1 \bar{J}_c(r)^2 dr}, \text{ if } d = 0; \tag{14}$$

$$T\left(\widehat{\theta}_T-\theta_T\right)\xrightarrow{d}\frac{\frac{1}{2}J_c^H(1)^2-c\int_0^1 J_c^H(r)^2dr-B^H(1)\int_0^1 J_c^H(r)dr}{\int_0^1 J_c^H(r)^2dr-\left(\int_0^1 J_c^H(r)dr\right)^2},\ \textit{if } d>0, \tag{15}$$

where $H=d+0.5$, $\bar{J}_c(r)=J_c(r)-\int_0^1 J_c(s)ds$ is the de-meaned O-U process, the parameters σ_v^2 and $\overline{\sigma}_v^2$ are defined in (8), (9), and σ_w^2 and σ_e^2 are defined in Assumptions 1 and 2, respectively.

Remark 2.2. Theorem 2.1 reveals several facts about the asymptotics of $\widehat{\theta}_T$. First, $\widehat{\theta}_T$ is a consistent estimator of θ_T. Second, both the convergence rate and the limiting distribution of $\widehat{\theta}_T$ depend on the value of d. The centered estimator $\widehat{\theta}_T-\theta_T$ converges to zero at the same rate when $d\geq 0$, which is higher than that of $d<0$. Third, the measurement error influences the limiting distribution when $d\leq 0$ because σ_w^2 appears in the limiting distribution. In particular, the measurement error leads to a more severe downward bias in $\widehat{\theta}_T$. Fourth, when $d>0$, σ_w^2 plays no role in the asymptotics of $\widehat{\theta}_T$.

Remark 2.3. To understand the phenomenon that the variance of the measurement error, σ_w^2, presents in the limiting distribution of the LS estimator only when $d\leq 0$, we pay attention to the asymptotic properties of the term $\sum_{t=1}^T y_{t-1}\varepsilon_t$, which appears in the numerator of the centered LS estimator. From Eq. (2), it is easy to get

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

The large sample theory of $\sum_{t=1}^T y_{t-1}\varepsilon_t$ is jointly determined by y_T^2, $\left(\theta_T^2-1\right)\sum_{t=1}^T y_{t-1}^2$, and $\sum_{t=1}^T\varepsilon_t^2$. Note that $y_t=\mu+\xi_t+w_t$, that is, imposing a relatively "high frequency" noise w_t on the persistent signal ξ_t gives the observation series y_t. Therefore, the asymptotic properties of y_t should be the same as those of ξ_t. As shown in Lemma A.1, $y_T^2=O_p\left(T^{2d+1}\right)$ and $\left(\theta_T^2-1\right)\sum_{t=1}^T y_{t-1}^2=O_p\left(T^{2d+1}\right)$, whose limiting distributions, after they are normalized by corresponding convergence rates, are free from the impact of w_t. In contrast, the limit of $T^{-1}\sum_{t=1}^T\varepsilon_t^2$ is the variance of $\varepsilon_t=v_t+w_t-\theta_T w_{t-1}$, that is,

$$\frac{1}{T}\sum_{t=1}^T\varepsilon_t^2=\frac{1}{T}\sum_{t=1}^T\left(v_t^2+w_t^2+\theta_T^2w_{t-1}^2\right)+o_p(1)\xrightarrow{p}\sigma_v^2+2\sigma_w^2,$$

which involves the variance of the measurement error w_t. It becomes clear now that when $d<0$, $\sum_{t=1}^T\varepsilon_t^2$ dominates the other two terms in the decomposition of $\sum_{t=1}^T y_{t-1}\varepsilon_t$ and hence, determines the limiting distribution of $\sum_{t=1}^T y_{t-1}\varepsilon_t$. Consequently, we see the role of σ_w^2 in the limiting distribution of the LS estimator. Whereas, when $d>0$, the dominant terms in the decomposition are $y_T^2=O_p\left(T^{2d+1}\right)$ and $\left(\theta_T^2-1\right)\sum_{t=1}^T y_{t-1}^2=O_p\left(T^{2d+1}\right)$. As a result, the limiting distribution of $\widehat{\theta}_T-\theta_T$ in the case of $d>0$ is not affected by the measurement error w_t.

Remark 2.4. When $\sigma_w^2=0$, $\mu=0$, and $c=0$, the latent process ξ_t in Model (1) becomes observable and is described by an integrated first-order autoregression with $I(d)$ errors:

$$\xi_t=\xi_{t-1}+v_t.$$

Sowell (1990) develops the large sample theory of the LS estimator of the AR root, that is $\widetilde{\theta}_T=\sum_{t=1}^T\xi_{t-1}\xi_t/\sum_{t=1}^T\xi_{t-1}^2$, as

$$T^{(2d+1)}\left(\widetilde{\theta}_T-1\right)\xrightarrow{d}-\frac{\sigma_v^2/2}{\overline{\sigma}_v^2\int_0^1 B^H(r)^2dr},\quad \text{if } d<0,$$

$$T\left(\widetilde{\theta}_T-1\right)\xrightarrow{d}\frac{\int_0^1 W(r)dW(r)}{\int_0^1 W(r)^2dr},\ \text{if } d=0,$$

$$T\left(\widetilde{\theta}_T-1\right)\xrightarrow{d}\frac{\frac{1}{2}B^H(1)^2}{\int_0^1 B^H(r)^2dr},\ \text{if } d>0.$$

Buchmann and Chan (2007) extend Sowell's results to the case where $c\neq 0$, that is, the AR root of ξ_t becomes $\theta_T=1+c/T$. Our results in Theorem 2.1 extend the study to the case where ξ_t are observed with measurement errors ($\sigma_w^2\neq 0$), and an AR regression with an unknown intercept is used.

Remark 2.5. When $\sigma_w^2=0$, our model is also closely related to the following model considered by Wang et al. (2023b):

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

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

$$\begin{aligned}Cov\left(\varepsilon_t,\varepsilon_s\right)&=\frac{1}{2}\left[(k+1)^{2H}+(k-1)^{2H}-2k^{2H}\right]\\&\sim H(2H-1)k^{2H-2}\ \text{for large } k=|t-s|.\end{aligned}$$

As k increases, the autocovariance of the FGN decays to zero at the same rate as that of an $I(d)$ process with $d=H-0.5$. Wang et al. (2023b) derive the limiting distributions of both $\widehat{\theta}_T$ (i.e., the LS estimator with fitted intercept) and $\widetilde{\theta}_T$ (i.e., the LS estimator without fitted intercept) under a general initial condition. However, the results in Wang et al. (2023b) are not the invariance principle as the error term in the model (16) is assumed to be normally distributed. By betting ε_t be an $I\left(d\right)$ process without specific distributional assumption, the results in the present paper establish an invariance principle.

To purge the influence of the measurement error and reduce the bias in estimating θ_T, we propose to use the IV estimator $\widehat{\theta}_T^{IV}$ defined in (7). Theorem 2.2 presents the large sample theory.

Theorem 2.2. *Let $\{y_t\}_{t=0}^T$ be the time series generated by (1), $\widehat{\theta}_T^{IV}$ be the IV estimator defined in (7). Let Assumptions 1-3 hold. Then, as $T\to\infty$, it has*

$$T^{(2d+1)}\left(\widehat{\theta}_T^{IV}-\theta_T\right)\xrightarrow{d}\frac{-\left(0.5+d\right)\Gamma\left(2+d\right)/\Gamma\left(2-d\right)}{\int_0^1 J_c^H(r)^2dr-\left(\int_0^1 J_c^H(r)dr\right)^2},\ \text{if } d<0;\tag{17}$$

$$T\left(\widehat{\theta}_T^{IV}-\theta_T\right)\xrightarrow{d}\frac{\int_0^1 \bar{J}_c(r)dW(r)}{\int_0^1 \bar{J}_c(r)^2dr},\ \text{if } d=0;\tag{18}$$

$$T\left(\widehat{\theta}_T^{IV}-\theta_T\right)\xrightarrow{d}\frac{\frac{1}{2}J_c^H\left(1\right)^2-c\int_0^1 J_c^H\left(r\right)^2dr-B^H(1)\int_0^1 J_c^H(r)dr}{\int_0^1 J_c^H(r)^2dr-\left(\int_0^1 J_c^H(r)dr\right)^2},\ \text{if } d>0,\tag{19}$$

where $d\in(-0.5,0.5)$ and $H=d+0.5$.

Remark 2.6. Theorem 2.2 shows that the measurement error $\{w_t\}$ (i.e., σ_w^2) does not affect the IV estimator $\widehat{\theta}_T^{IV}$ in asymptotics. Moreover, the limiting distribution of $\widehat{\theta}_T^{IV}$ in every case has only d and c as nuisance parameters. It further has $c=0$ in unit-root testing. In contrast, the limiting distribution of the LS estimator $\widehat{\theta}_T$ depends on extra parameters, $\sigma_w^2,\sigma_e^2,\sigma_v^2,\overline{\sigma}_v^2$. Hence, it is more convenient to make inferences about θ based on $\widehat{\theta}_T^{IV}$ than on $\widehat{\theta}_T$.

Remark 2.7. It is known that $\int_0^1 \bar{J}_c(r)dW(r)/\int_0^1 \bar{J}_c(r)^2dr$ is left-skewed for $c \geq 0$, suggesting that the IV estimator is downward biased when $d = 0$. In this case, it is straightforward to compare the limiting distribution in (18) with that in (14) and find that $\widehat{\theta}_T^{IV}$ has a less downward bias than $\widehat{\theta}_T$ because σ_w^2 is eliminated from the asymptotic distribution.

Remark 2.8. To facilitate the comparison between the limiting distribution of $\widehat{\theta}_T^{IV}$ and that of $\widehat{\theta}_T$ when $d < 0$, we give an equivalent representation of the limiting distribution given in (17):[5]

$$T^{(2d+1)}\left(\widehat{\theta}_T^{IV} - \theta_T\right) \xrightarrow{d} \frac{-\sigma_v^2/2 - Cov\left(v_t, v_{t-1}\right)}{\overline{\sigma}_v^2\left[\int_0^1 J_c^H(r)^2dr - \left(\int_0^1 J_c^H(r)dr\right)^2\right]}, \tag{20}$$

where

$$Cov\left(v_t, v_{t-1}\right) = \frac{\Gamma\left(1-2d\right)\Gamma\left(1+d\right)}{\Gamma\left(d\right)\Gamma\left(1-d\right)\Gamma\left(2-d\right)}\sigma_e^2 = \frac{d}{1-d}\sigma_v^2. \tag{21}$$

Comparing to the limiting distribution of $\widehat{\theta}_T$ given in (13), the term $-\sigma_w^2$ is replaced with the term $-Cov\left(v_t, v_{t-1}\right)$ in the limiting distribution of $\widehat{\theta}_T^{IV}$. Since $-\sigma_w^2$ is always negative and $-Cov\left(v_t, v_{t-1}\right)$ is always positive when $d < 0$, $\widehat{\theta}_T^{IV}$ is not only less biased than $\widehat{\theta}_T$ when there is measurement error but also less biased than $\widehat{\theta}_T$ when there is no measurement error. Combining (20) and (21), we can also get

$$T^{(2d+1)}\left(\widehat{\theta}_T^{IV} - \theta_T\right) \xrightarrow{d} \frac{-\frac{1}{2}\frac{1+d}{1-d}\sigma_v^2}{\overline{\sigma}_v^2\left(\int_0^1 J_c^H(r)^2dr - \left(\int_0^1 J_c^H(r)dr\right)^2\right)},$$

which suggests that the bias of $\widehat{\theta}_T^{IV}$ is negative.

Remark 2.9. For the case of $d > 0$, the formula in (19) shows that the IV estimator $\widehat{\theta}_T^{IV}$ has the same limiting distribution as $\widehat{\theta}_T$. Hence, $\widehat{\theta}_T^{IV}$ does not lead to any loss of asymptotic efficiency relative to $\widehat{\theta}_T$.

Remark 2.10. The advantages of $\widehat{\theta}_T^{IV}$ over $\widehat{\theta}_T$ in the case of $d = 0$ have been explored by Hall (1989) and Hansen and Lunde (2014). However, when $d \neq 0$, $\{v_t\}$ has serial dependence. In this scenario, the choice of the optimal instrumental variable y_{t-j} and the performance of the resulting IV estimator have not been studied in the literature. Theorem 2.2 partially fulfills this task by showing that y_{t-2} is a good instrument, especially when $d < 0$.

3. Latent model with strongly mixing errors

3.1. The model and motivations

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 present in $\{v_t\}$ when ξ_t measures 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.

Heteroskedasticity may also arise in the measurement error $\{w_t\}$. One example is using daily RV to estimate daily integrated variance. To compute the daily RV for a trading day t, let the intraday return based on a particular sampling frequency M be

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

[5]This formula is obtained in the proof of Theorem 2.2, which is included in the Appendix.

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

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

As $M \to \infty$, it has

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

where σ_s^2 is the spot variance and IV_t denotes the integrated variance on day t. Barndorff-Nielsen and Shephard (2002) have shown that, as $M \to \infty$,

$$\sqrt{M}\,(RV_t(M) - IV_t) \xrightarrow{d} MN(0, 2IQ_t), \tag{22}$$

where MN stands for mixed normality and

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

is the integrated quarticity. To improve the accuracy of the asymptotic approximation, Barndorff-Nielsen and Shephard (2005) suggest using the log RV_t to approximate the log IV_t and develop the following asymptotic theory when $M \to \infty$:

$$\sqrt{M}\left[\log\left(RV_t(M)\right) - \log\left(IV_t\right)\right] \xrightarrow{d} MN(0, 2IQ_t/IV_t^2), \tag{23}$$

where IQ_t/IV_t^2 can be much less time-varying than IQ_t. The asymptotic theory given by (22) and (23) suggests the presence of heteroskedasticity when approximating IV_t (or log IV_t) by RV_t (or log RV_t). This example gives a practical reason why one would like to relax the i.i.d. 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. Hence, in this section, we study the latent local-to-unity model defined in (1), but with the following assumptions on the error sequences $\{w_t\}$ and $\{v_t\}$.

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

Assumption 5. *$E(v_t) = 0$ for all t. $\{v_t\}$ is strong mixing with mixing coefficients α_m satisfying $\sum_{m=1}^{\infty} \alpha_m^{1-\frac{2}{k}} < \infty$. There exists $k > 2$ such that $\sup_t E|v_t|^k < \infty$. Assume $\sigma_v^2 = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} E(v_t^2) \in (0,\infty)$, $\overline{\sigma}_v^2 = \lim_{T\to\infty} \frac{1}{T} E\left(\sum_{t=1}^{T} v_t\right)^2 \in (0,\infty)$, and $\gamma_1 = \text{plim}_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} v_t v_{t-1}$ exists.*

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

Assumption 4 allows the latent process ξ_t to be observed with independent but not necessarily identically distributed errors. Assumption 5 allows ξ_t to have an error term that could be serially dependent and heteroscedastic. This 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) applies to the partial sum process of $\{v_t\}$, that is, as $T \to \infty$,

$$\frac{1}{T^{1/2}\overline{\sigma}_v} \sum_{t=1}^{\lfloor Tr \rfloor} v_t \Rightarrow W(r) \quad \text{for } r \in [0,1].$$

3.2. Asymptotics for the model with strongly mixing errors

This subsection develops the large sample theory for the LS estimator $\widehat{\theta}_T$ and the IV estimator $\widehat{\theta}_T^{IV}$, defined in (6) and (7), respectively, under Assumptions 4-6. Compared to $\widehat{\theta}_T$, it is found that the estimator $\widehat{\theta}_T^{IV}$ can purge the influence of the observation errors $\{w_t\}$, but not necessarily lead to a smaller bias. The last finding is distinct from that for the model with $I(d)$ errors, in which case $\widehat{\theta}_T^{IV}$ always leads to a smaller bias than $\widehat{\theta}_T$.

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

$$T\left(\widehat{\theta}_T - \theta_T\right) \xrightarrow{d} \frac{\int_0^1 \bar{J}_c(r)dW(r) + \frac{1}{2}\left(\overline{\sigma}_v^2 - \sigma_v^2 - 2\sigma_w^2\right)/\overline{\sigma}_v^2}{\int_0^1 \bar{J}_c(r)^2 dr}, \tag{24}$$

and

$$T\left(\widehat{\theta}_T^{IV} - \theta_T\right) \xrightarrow{d} \frac{\int_0^1 \bar{J}_c(r)dW(r) + \frac{1}{2}\left(\overline{\sigma}_v^2 - \sigma_v^2 - 2\gamma_1\right)/\overline{\sigma}_v^2}{\int_0^1 \bar{J}_c(r)^2 dr}, \tag{25}$$

where σ_w^2, γ_1, $\overline{\sigma}_v^2$, and σ_v^2 are defined in Assumptions 4-5, and $\bar{J}_c(r)$ is defined in Theorem 2.1.

Remark 3.1. Theorem 3.1 shows that the measurement error $\{w_t\}$ affects the limiting distribution of the LS estimator $\widehat{\theta}_T$. In contrast, it does not affect the limiting distribution of the IV estimator $\widehat{\theta}_T^{IV}$.

Remark 3.2. Although $\widehat{\theta}_T^{IV}$ purges the influence of σ_w^2 in asymptotics, γ_1 appears in the limiting distribution. Whether or not $\widehat{\theta}_T^{IV}$ outperforming $\widehat{\theta}_T$ in terms of bias depends on the sign of $E\left\{\left[\int_0^1 \bar{J}_c(r)dW(r) + \frac{1}{2}\left(\overline{\sigma}_v^2 - \sigma_v^2\right)/\overline{\sigma}_v^2\right] / \int_0^1 \bar{J}_c(r)^2 dr\right\}$ and the sign and magnitude of γ_1 relative to σ_w^2.

Remark 3.3. Under Assumptions 4-6, it can be shown that

$$\frac{1}{T}\sum_{t=1}^T \varepsilon_t^2 = \frac{1}{T}\sum_{t=1}^T \left(v_t^2 + w_t^2 + \theta_T^2 w_{t-1}^2\right) + o_p(1) \xrightarrow{p} \sigma_v^2 + 2\sigma_w^2,$$

by McLeish (1975, Theorem 2.10 with condition (2.12)). Hence, $\sigma_v^2 + 2\sigma_w^2$ as a whole can be consistently estimated by

$$\frac{1}{T}\sum_{t=1}^T \widehat{\varepsilon}_t^2 = \frac{1}{T}\sum_{t=1}^T \left(y_t - \widehat{\alpha} - \widehat{\theta}_T y_{t-1}\right)^2.$$

Whereas, estimating σ_v^2 separately is not easily achieved. Thus, compared to $\widehat{\theta}_T$, it is more challenging to use $\widehat{\theta}_T^{IV}$ to make inferences.

Remark 3.4. For a special case where both $\{w_t\}$ and $\{v_t\}$ are i.i.d. sequences and $\theta_T = 1$ with $c = 0$, Hall (1989) and Hansen and Lunde (2014) prove that

$$T\left(\widehat{\theta}_T - \theta_T\right) \xrightarrow{d} \frac{\int_0^1 \bar{J}_c(r)dW(r) - \sigma_w^2/\overline{\sigma}_v^2}{\int_0^1 \bar{J}_c(r)^2 dr},$$

and

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

In this case, $\widehat{\theta}_T^{IV}$ has a smaller bias than $\widehat{\theta}_T$. Moreover, using $\widehat{\theta}_T^{IV}$ to make inferences is more convenient than using $\widehat{\theta}_T$, because its limiting distribution of $\widehat{\theta}_T^{IV}$ is free from nuisance parameters. Our results in Theorem 3.1 extend those in the literature to the local-to-unity case allowing for heteroskedasticity.

4. Finite sample performance of asymptotic distributions

To check the finite sample performance of the asymptotic distributions, we focus on the case where Assumptions 1-3 hold. That is, we check the quality of approximation of the asymptotic distributions derived in Theorem 2.1 and Theorem 2.2 in finite sample.

We simulate data $\{y_t\}_{t=0}^T$ from Model (1) under Assumptions 1-3 with $w_t \overset{iid}{\sim} N(0, 0.1)$, $e_t \overset{iid}{\sim} N(0, 1)$, $\mu = 0$, $c = -5$, and $T = 256$, 1024, and 4096, respectively. The finite sample distributions of the estimators $\widehat{\theta}_T$ and $\widehat{\theta}_T^{IV}$ are obtained from 10,000 replications. The asymptotic distributions are calculated based on the formulae given in Theorem 2.1 and Theorem 2.2, with the integrals replaced by the corresponding sums with 10,000 interior points.

Figure 1 plots three finite sample distributions of $T\left(\widehat{\theta}_T - \theta_T\right)$, its asymptotic distribution, three finite sample distributions of $T\left(\widehat{\theta}_T^{IV} - \theta_T\right)$ and its asymptotic distribution when $d = 0$. It can be seen that even when $T = 256$, the finite sample distributions are very close to their asymptotic distributions. All of them are asymmetric and skewed to the left. It is much more likely for $T\left(\widehat{\theta}_T - \theta_T\right)$ and $T\left(\widehat{\theta}_T^{IV} - \theta_T\right)$ to take a negative number than a positive number. It is clear that the IV estimator is less biased and more efficient than the LS estimator when $d = 0$.

Figure 2 plots three finite sample distributions plus the asymptotic distribution of $T\left(\widehat{\theta}_T - \theta_T\right)$, as well as those of $T\left(\widehat{\theta}_T^{IV} - \theta_T\right)$ when $d = -0.3$. The four densities centered around -2 are for $\widehat{\theta}_T$ while the four densities centered around -0.8 are for $\widehat{\theta}_T^{IV}$. Three features are apparent in the figure. First, the finite sample distributions of the LS estimator are close to the asymptotic distribution. In contrast, the finite sample distributions of the IV estimator are noticeably different from the asymptotic distribution, even when $T = 4096$, although the finite sample distribution gets closer to the asymptotic distribution as T increases. Second, the limiting distributions of both estimators have only negative support. Hence, both estimators have a downward bias. But, the IV estimator is much less biased than the LS estimator. Third, the distributions of the IV estimator are more concentrated than those of the LS estimator. Therefore, $\widehat{\theta}_T^{IV}$ is more efficient than $\widehat{\theta}_T$ when $d = -0.3$.

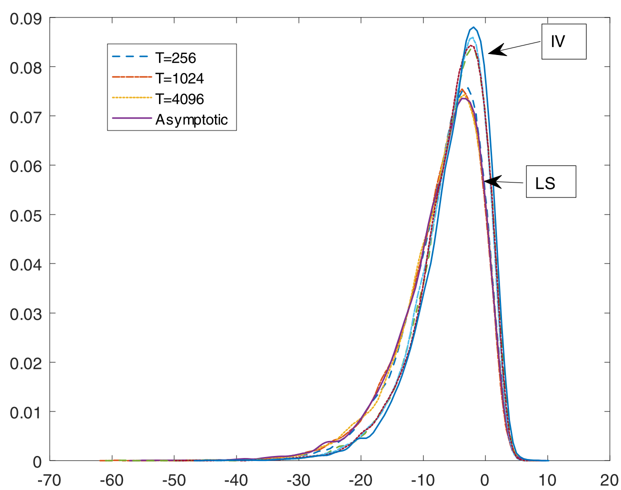

Figure 1. The graph plots three finite sample distributions and the asymptotic distribution of $T\left(\widehat{\theta}_T - \theta_T\right)$ and those of $T\left(\widehat{\theta}_T^{IV} - \theta_T\right)$, when $d = 0$ and $T = 256$, 1024, and 4096, respectively.

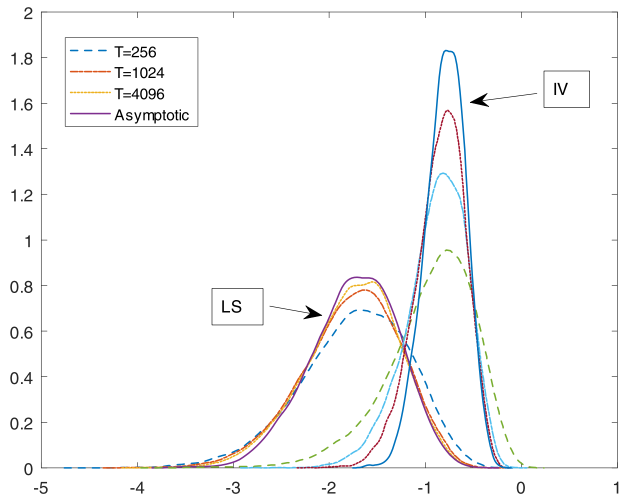

Figure 2. Three finite sample distributions and the asymptotic distribution of $T^{2d+1}\left(\widehat{\theta}_T - \theta_T\right)$ and those of $T^{2d+1}\left(\widehat{\theta}_T^{IV} - \theta_T\right)$ when $d = -0.3$ and $T = 256$, 1024, and 4096, respectively.

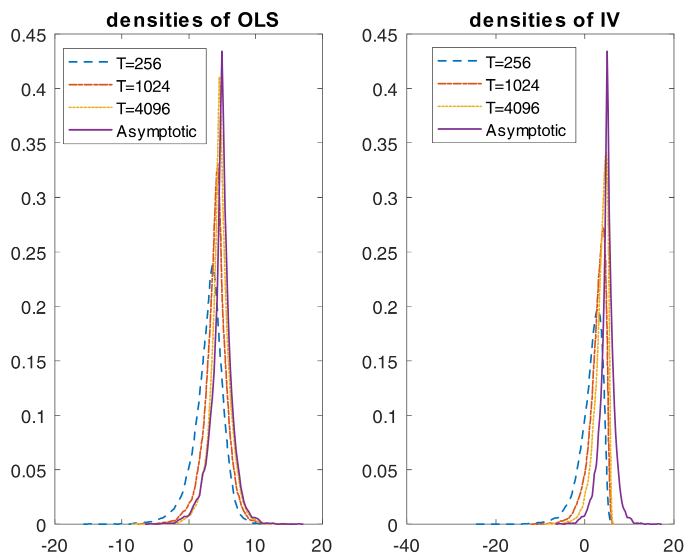

Figure 3. The finite sample and asymptotic distributions of $T\left(\widehat{\theta}_T - \theta_T\right)$ (the left panel) and those of $T\left(\widehat{\theta}_T^{IV} - \theta_T\right)$ (the right panel) when $d = 0.3$, $T = 256$, 1024, and 4096, respectively.

Figure 3 plots three finite sample distributions of $T\left(\widehat{\theta}_T - \theta_T\right)$ and its asymptotic distribution (the left panel) as well as those distributions of $T\left(\widehat{\theta}_T^{IV} - \theta_T\right)$ (the right panel) when $d = 0.3$. The two asymptotic distributions are the same, as has been proven in Theorem 2.1 and Theorem 2.2. The figure shows that, for each estimator, the finite sample distribution has some distance to the asymptotic distribution when $T = 256$. As T increases, the finite sample distribution gets closer to its limiting distribution. Also shown in the figure is that both $T\left(\widehat{\theta}_T - \theta_T\right)$ and $T\left(\widehat{\theta}_T^{IV} - \theta_T\right)$ are more likely to take a positive number than a negative one. Hence, both two estimators are upward biased when $d = 0.3$.

To further compare the two estimators' finite sample performance in the case of $d = 0.3$, Figure 4 plots the finite sample distributions of $T\left(\widehat{\theta}_T - \theta_T\right)$ and $T\left(\widehat{\theta}_T^{IV} - \theta_T\right)$ for $T = 256$ only, plus the associated

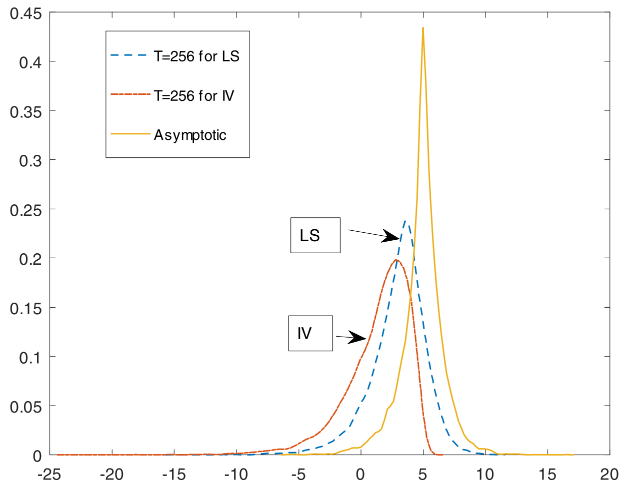

Figure 4. The finite sample distributions of $T\left(\widehat{\theta}_T - \theta_T\right)$ and $T\left(\widehat{\theta}_T^{IV} - \theta_T\right)$ when $d = 0.3$ and $T = 256$, and the associated asymptotic distribution.

asymptotic distribution. It shows that the LS estimator has a more significant upward bias than the IV estimator. This is because the finite sample distribution of the LS estimator is closer to the asymptotic distribution, which has a support that is almost always positive. Moreover, there might have a bias-variance trade-off between $\widehat{\theta}_T$ and $\widehat{\theta}_T^{IV}$ when d is positive, as the finite sample distribution of $\widehat{\theta}_T$ seems to be more concentrated than that of $\widehat{\theta}_T^{IV}$.

5. Conclusion

In this article, the primary consideration is given to the latent local-to-unity model under the conditions that (i) the error in the observation equation is an i.i.d. sequence, and (ii) the error in the state equation is an I(d) series with $d \in (-0.5, 0.5)$. We develop the large sample theory for the LS estimator of the AR root of the latent process. Two properties are found. First, both the convergence rate and the limiting distribution crucially depend on the value of d. Second, the variance of the measurement error appears in the limiting distribution and deteriorates the downward bias of the LS estimator when $d \leq 0$.

To purge the influence of the measurement error, we propose the IV estimator with y_{t-2} as an instrument to estimate the AR root of the latent process. It is shown that the IV method not only eliminates the effect of the measurement error but also leads to a less biased estimator than LS when $d \leq 0$.

We also study another model specification, in which (i) the error term in the observation equation is independent and not necessarily identically distributed, and (ii) the error term in the state equation is strong mixing. In this case, the IV estimator can still remove the influence of the measurement error in asymptotics. However, it is not necessarily less biased than the LS estimator.

Our model is similar to the class of models recently introduced by Dou and Müller (2021) in that the local-to-unity feature exists in both the AR and the MA components. However, in our model, the local-to-unity feature is only a part of the MA component, which comes from the state-space modeling strategy. Hence, it has a natural structural interpretation. It would be interesting to compare the empirical relevance of these two non-nested modeling strategies.

We have not considered the boundary case where $d = 0.5$. The model with $d = 0.5$ has attracted some attention in the literature in different contexts; see, for example, Duffy and Kasparis (2021). Extending our results to cover the case of $d = 0.5$ is to be investigated.

We should also point out that making statistical inferences about θ_T based on our asymptotic theory is generally not feasible, as it requires c and d to be consistently estimated. Although d is consistently estimable as shown in Wang et al. (2023b), c is not consistently estimable in our setup. However, our asymptotic theory can be used to examine the local power of unit root tests. To develop a feasible inferential framework, one may use the grid bootstrap method of Hansen (1999). The formal asymptotic justification of the grid bootstrap method in the context of the latent local-to-unity model is beyond the scope of this article. We plan to report it in the future study.

A. Appendix

Lemma A.1. *Let $\{y_t\}_{t=0}^T$ be the time series generated by (1). Let Assumptions 1-3 hold, $H = d + 0.5$, and $r \in [0,1]$. Then, as $T \to \infty$, it has*

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

where σ_v^2 and $\overline{\sigma}_v^2$ are defined in (8), (9), σ_w^2 and σ_e^2 are defined in Assumptions 1 and 2, $J_c^H(r)$ is the fO-U process defined in (12), and $J_c(r)$ is $J_c^H(r)$ with $H = 0.5$.

Lemma A.2. *Let $\{y_t\}_{t=0}^T$ be the time series generated by (1) and Assumptions 4-6 hold. Then, as $T \to \infty$, it has*

1. $\frac{1}{T^{1/2}} y_{[Tr]} \Rightarrow \overline{\sigma}_v J_c(r)$;
2. $\frac{1}{T^{3/2}}\sum_{t=1}^{T} y_{t-1} \xrightarrow{d} \overline{\sigma}_v \int_0^1 J_c(r)dr$;
3. $\frac{1}{T^2}\sum_{t=1}^{T} y_{t-1}^2 \xrightarrow{d} \overline{\sigma}_v^2 \int_0^1 J_c(r)^2 dr$;
4. $\frac{1}{T}\sum_{t=1}^{T} y_{t-1}\varepsilon_t \xrightarrow{d} \overline{\sigma}_v^2 \int_0^1 J_c(r)dW(r) + \frac{1}{2}\left(\overline{\sigma}_v^2 - \sigma_v^2 - 2\sigma_w^2\right)$;
5. $\frac{1}{T}\sum_{t=2}^{T} y_{t-2}\varepsilon_t \xrightarrow{d} \overline{\sigma}_v^2 \int_0^1 J_c(r)dW(r) + \frac{1}{2}\left(\overline{\sigma}_v^2 - \sigma_v^2 - 2\gamma_1\right)$;
6. $\frac{1}{T^2}\sum_{t=2}^{T} y_{t-2}y_{t-1} \xrightarrow{d} \overline{\sigma}_v^2 \int_0^1 J_c(r)^2 dr$,

where σ_w^2, γ_1, $\overline{\sigma}_v^2$, and σ_v^2 are the constants defined in Assumptions 4-5, and $\bar{J}_c(r)$ is defined in Theorem 2.1.

Proof of Lemma A.1. Part (1): Under Assumption (2), the FCLT of Davydov (1970) and Taqqu (1975) shows that, as $T \to \infty$,

$$\frac{1}{T^H}\sum_{t=1}^{\lfloor Tr \rfloor} v_t \Rightarrow \overline{\sigma}_v B^H(r).$$

Note that $\varepsilon_t = v_t + w_t - \theta_T w_{t-1}$ and

$$\sum_{t=1}^{\lfloor Tr \rfloor} \varepsilon_t = \left(\sum_{t=1}^{\lfloor Tr \rfloor} v_t\right) + w_{\lfloor Tr \rfloor} - w_0 - \frac{c}{T}\sum_{t=1}^{\lfloor Tr \rfloor} w_t,$$

where $w_t \overset{iid}{\sim} \left(0, \sigma_w^2\right)$ as presumed in Assumption 1. Hence, as long as $H = d + 0.5 > 0$, it has

$$\begin{aligned}\sup_{r\in(0,1)} \frac{1}{T^H}\left|\sum_{t=1}^{\lfloor Tr\rfloor} \varepsilon_t - \sum_{t=1}^{\lfloor Tr\rfloor} v_t\right| &= \sup_{r\in(0,1)} \frac{1}{T^H}\left|w_{\lfloor Tr\rfloor} - w_0 - \frac{c}{T}\sum_{t=1}^{\lfloor Tr\rfloor} w_t\right| \\ &\leq \sup_{r\in(0,1)}\left|\frac{w_{\lfloor Tr\rfloor} - w_0}{T^H}\right| + \left|\frac{c}{T^{H+1}}\sum_{t=1}^{\lfloor Tr\rfloor} w_t\right| \\ &\xrightarrow{p} 0.\end{aligned}$$

Therefore, $\frac{1}{T^H}\sum_{t=1}^{\lfloor Tr\rfloor}\varepsilon_t$ and $\frac{1}{T^H}\sum_{t=1}^{\lfloor Tr\rfloor} v_t$ have the same limit. This gives the proof of Part (1).

Part (2): Buchmann and Chan (2007) have proved that, under Assumption (2), the latent process ξ_t given in Model (1) has the following large sample theory (see also Tanaka, 2013):

$$\frac{\xi_{\lfloor Tr\rfloor}}{T^H} \Rightarrow \overline{\sigma}_v J_c^H(r), \text{ as } T \to \infty,$$

where $J_c^H(r)$ is the fO-U process defined in (12). Note that $y_t = \mu + \xi_t + w_t$. Under Assumption (1), it is easy to get

$$\sup_{r\in(0,1)}\left|\frac{y_{\lfloor Tr\rfloor}}{T^H} - \frac{\xi_{\lfloor Tr\rfloor}}{T^H}\right| = \sup_{r\in(0,1)}\left|\frac{\mu + w_{\lfloor Tr\rfloor}}{T^H}\right| \xrightarrow{p} 0.$$

Therefore, it has $\frac{y_{\lfloor Tr\rfloor}}{T^H} \Rightarrow \overline{\sigma}_v J_c^H(r)$ too.

Parts (3)-(4) are obtained by directly applying Part (2) and the continuous mapping theorem.

To prove Parts (5)-(7), we turn to the AR representation of y_t that is given in Eq. (2):

$$y_t = (1 - \theta_T)\,\mu + \theta_T y_{t-1} + \varepsilon_t,$$

which leads to

$$y_t^2 = \frac{\mu^2 c^2}{T^2} + \theta_T^2 y_{t-1}^2 + \varepsilon_t^2 + 2\theta_T y_{t-1}\varepsilon_t + 2\frac{\mu c}{T}\left(\theta_T y_{t-1} + \varepsilon_t\right),$$

and

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

Under Assumptions 1-3, the law of large numbers yields the following results for any $H \in (0, 1)$:

$$\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t = \frac{1}{T}\sum_{t=1}^{T}\left(v_t + w_t - \theta_T w_{t-1}\right) \xrightarrow{p} 0,$$

and

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2 &= \frac{1}{T}\sum_{t=1}^{T}\left[v_t^2 + \left(w_t - \theta_T w_{t-1}\right)^2 + 2v_t\left(w_t - \theta_T w_{t-1}\right)\right] \\ &= \frac{1}{T}\sum_{t=1}^{T}\left(v_t^2 + w_t^2 + \theta_T^2 w_{t-1}^2\right) + o_p(1) \xrightarrow{p} \sigma_v^2 + 2\sigma_w^2.\end{aligned}$$

Hence, $\sum_{t=1}^{T}\varepsilon_t^2 = O_p(T)$ for any $H \in (0, 1)$. However, the results in Parts (3)-(4) show that $y_T^2 = O_p(T^{2H})$, $\sum_{t=1}^{T} y_{t-1}^2 = O_p(T^{1+2H})$, and $\sum_{t=1}^{T} y_{t-1} = O_p(T^{1+H})$. Those terms have orders crucially depending on H, and hence, can dominate or be dominated by the term $\sum_{t=1}^{T}\varepsilon_t^2$ for different values of H. As a result, the convergence rate and the limiting distribution of $\sum_{t=1}^{T} y_{t-1}\varepsilon_t$ also depend on the value of H.

When $H < 0.5$, it has

$$\frac{1}{T}\sum_{t=1}^{T} y_{t-1}\varepsilon_t = -\frac{1}{2\theta_T}\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2 + o_p(1) \xrightarrow{p} \frac{\sigma_v^2}{2} + \sigma_w^2.$$

When $H = 0.5$, it has

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

$$\xrightarrow{d} \frac{1}{2}\left\{\overline{\sigma}_v^2 J_c(1)^2 - 2c\overline{\sigma}_v^2 \int_0^1 J_c(r)^2 dr - \sigma_v^2 - 2\sigma_w^2\right\}$$

$$= \frac{\sigma_e^2}{2}\left\{J_c(1)^2 - 2c\int_0^1 J_c(r)^2 dr - 1\right\} - \sigma_w^2,$$

where the last equation comes from that fact that $\overline{\sigma}_v^2 = \sigma_v^2 = \sigma_e^2$ when $H = 0.5$.

When $H > 0.5$,

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

$$\xrightarrow{d} \frac{1}{2}\left\{\overline{\sigma}_v^2 J_c^H(1)^2 - 2c\overline{\sigma}_v^2 \int_0^1 J_c^H(r)^2 dr\right\}.$$

This completes the proof of Lemma A.1. □

Proof of Theorem 2.1. When $d < 0$, it has $H = d + 0.5 < 0.5$. From the results in Lemma A.1.1-5, it is straightforward to get

$$T^{2H}\left(\widehat{\theta}_T - \theta_T\right) = \frac{T^{-1}\sum_{t=1}^{T} y_{t-1}\varepsilon_t - T^{2H-1}\left(T^{-H}\sum_{t=1}^{T}\varepsilon_t\right)\left(T^{-1-H}\sum_{t=1}^{T} y_{t-1}\right)}{T^{-2H-1}\sum_{t=1}^{T} y_{t-1}^2 - \left(T^{-H-1}\sum_{t=1}^{T} y_{t-1}\right)^2}$$

$$\xrightarrow{d} \frac{-\sigma_v^2/2 - \sigma_w^2}{\overline{\sigma}_v^2\left\{\int_0^1 J_c^H(r)^2 dr - \left(\int_0^1 J_c^H(r) dr\right)^2\right\}}.$$

When $d = 0$, it has $H = 0.5$. Applying the results in Lemma A.1.1-4 and Lemma A.1.6, we have

$$T\left(\widehat{\theta}_T - \theta_T\right) = \frac{T^{-1}\sum_{t=1}^{T} y_{t-1}\varepsilon_t - \left(T^{-0.5}\sum_{t=1}^{T}\varepsilon_t\right)\left(T^{-3/2}\sum_{t=1}^{T} y_{t-1}\right)}{T^{-2}\sum_{t=1}^{T} y_{t-1}^2 - \left(T^{-3/2}\sum_{t=1}^{T} y_{t-1}\right)^2}$$

$$\xrightarrow{d} \frac{\frac{\sigma_e^2}{2}\left\{J_c(1)^2 - 2c\int_0^1 J_c(r)^2 dr - 1\right\} - \sigma_w^2 - \overline{\sigma}_v^2 W(1)\int_0^1 J_c(r) dr}{\overline{\sigma}_v^2\int_0^1 J_c(r)^2 dr - \overline{\sigma}_v^2\left(\int_0^1 J_c(r) dr\right)^2}$$

$$\stackrel{d}{=} \frac{\int_0^1 J_c(r) dW(r) - \sigma_w^2/\sigma_e^2 - W(1)\int_0^1 J_c(r) dr}{\int_0^1 J_c(r)^2 dr - \left(\int_0^1 J_c(r) dr\right)^2},$$

where the last equation is from the fact that $\overline{\sigma}_v^2 = \sigma_e^2$ when $H = 0.5$, and

$$\frac{1}{2}\left\{J_c(1)^2 - 2c\int_0^1 J_c(r)^2 dr - 1\right\} = \int_0^1 J_c(r) dW(r).$$

With the definition of $\overline{J}_c(r) := J_c(r) - \int_0^1 J_c(r) dr$, it is easy to get

$$\int_0^1 \overline{J}_c(r) dW(r) = \int_0^1 J_c(r) dW(r) - W(1)\int_0^1 J_c(r) dr,$$

and

$$\int_0^1 \bar{J}_c(r)^2 dr = \int_0^1 J_c(r)^2 dr - \left(\int_0^1 J_c(r) dr\right)^2 .$$

The limiting distribution reported in (14) in Theorem 2.1 is then obtained.

For the case of $d > 0$, which makes $H > 0.5$, it has

$$\begin{aligned} T\left(\widehat{\theta}_T - \theta_T\right) &= \frac{T^{-2H} \sum_{t=1}^{T} y_{t-1}\varepsilon_t - \left(T^{-H} \sum_{t=1}^{T} \varepsilon_t\right)\left(T^{-H-1} \sum_{t=1}^{T} y_{t-1}\right)}{T^{-2H-1} \sum_{t=1}^{T} y_{t-1}^2 - \left(T^{-H-1} \sum_{t=1}^{T} y_{t-1}\right)^2} \\ &\xrightarrow{d} \frac{\frac{1}{2}\left\{\overline{\sigma}_v^2 J_c^H(1)^2 - 2c\overline{\sigma}_v^2 \int_0^1 J_c^H(r)^2 dr\right\} - \overline{\sigma}_v^2 B^H(1) \int_0^1 J_c^H(r) dr}{\overline{\sigma}_v^2 \int_0^1 J_c^H(r)^2 dr - \overline{\sigma}_v^2 \left(\int_0^1 J_c^H(r) dr\right)^2} \\ &= \frac{\frac{1}{2}\left\{J_c^H(1)^2 - 2c \int_0^1 J_c^H(r)^2 dr\right\} - B^H(1) \int_0^1 J_c^H(r) dr}{\int_0^1 J_c^H(r)^2 dr - \left(\int_0^1 J_c^H(r) dr\right)^2}, \end{aligned}$$

where the limit comes from the results in Lemma A.1.1-4 and Lemma A.1.7. □

Proof of Theorem 2.2. The IV estimator $\widehat{\theta}_T^{IV}$ defined in (7) can be rewritten as

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

From Lemma A.1, it is straightforward to get that, as $T \to \infty$,

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

and

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

From the AR representation of y_t given in Eq. (2), we have

$$\sum_{t=2}^{T} y_{t-2}\varepsilon_t = \frac{1}{\theta_T}\sum_{t=2}^{T}\left[y_{t-1} - (1-\theta_T)\,\mu - \varepsilon_{t-1}\right]\varepsilon_t$$
$$= \frac{1}{\theta_T}\sum_{t=2}^{T} y_{t-1}\varepsilon_t - \frac{1}{\theta_T}\sum_{t=2}^{T}\varepsilon_{t-1}\varepsilon_t - o_p(1).$$

Under Assumptions 1-3, as $T \to \infty$, it has

$$\frac{1}{T}\sum_{t=2}^{T}\varepsilon_{t-1}\varepsilon_t = \frac{1}{T}\sum_{t=2}^{T}(v_{t-1} + w_{t-1} - \theta_T w_{t-2})(v_t + w_t - \theta_T w_{t-1})$$
$$= \frac{1}{T}\sum_{t=2}^{T} v_{t-1}v_t - \frac{\theta_T}{T}\sum_{t=2}^{T} w_{t-1}^2 + o_p(1)$$
$$\xrightarrow{d} Cov(v_{t-1}, v_t) - \sigma_w^2,$$

where

$$Cov(v_{t-1}, v_t) = \frac{\Gamma(1-2d)\,\Gamma(1+d)}{\Gamma(d)\,\Gamma(1-d)\,\Gamma(2-d)}\sigma_e^2.$$

For the case of $d < 0$, it has $H = d + 0.5 < 0.5$. Lemma A.1.5 shows that $\sum_{t=1}^{T} y_{t-1}\varepsilon_t = O_p(T)$. We then get, as $T \to \infty$,

$$\frac{1}{T}\sum_{t=2}^{T} y_{t-2}\varepsilon_t = \frac{1}{T\theta_T}\sum_{t=2}^{T} y_{t-1}\varepsilon_t - \frac{1}{T\theta_T}\sum_{t=2}^{T}\varepsilon_{t-1}\varepsilon_t - o_p(1)$$
$$\xrightarrow{p} -\sigma_v^2/2 - \sigma_w^2 - Cov(v_{t-1}, v_t) + \sigma_w^2$$
$$= -\sigma_v^2/2 - Cov(v_{t-1}, v_t).$$

Note that the variance of the measurement error, σ_w^2, disappears in the limit of $T^{-1}\sum_{t=2}^{T} y_{t-2}\varepsilon_t$. As it becomes clear later, this is why the limiting distribution of the IV estimator is free from σ_w^2 when $d < 0$. The covariance term $Cov(v_{t-1}, v_t)$ appears in the limit of $T^{-1}\sum_{t=2}^{T} y_{t-2}\varepsilon_t$ because of the inclusion of the cross product $T^{-1}\sum_{t=2}^{T}\varepsilon_{t-1}\varepsilon_t$. It is important to note that $Cov(v_{t-1}, v_t) < 0$ when $d < 0$. Hence, $\sigma_v^2/2$ and $Cov(v_{t-1}, v_t)$ could partially cancel out each other.

With the results above, it is ready to get that when $d < 0$, it has $H = d + 0.5 < 0.5$ and

$$T^{2H}\left(\widehat{\theta}_T^{IV} - \theta_T\right) = \frac{\frac{1}{T}\sum_{t=2}^{T} y_{t-2}\varepsilon_t - \frac{T^{2H}}{T-1}\left(\frac{1}{T^H}\sum_{t=2}^{T}\varepsilon_t\right)\left(\frac{1}{T^{1+H}}\sum_{t=2}^{T} y_{t-2}\right)}{\frac{1}{T^{1+2H}}\sum_{t=2}^{T} y_{t-2}y_{t-1} - \frac{T}{T-1}\left(\frac{1}{T^{1+H}}\sum_{t=2}^{T} y_{t-2}\right)\left(\frac{1}{T^{1+H}}\sum_{t=2}^{T} y_{t-1}\right)}$$
$$\xrightarrow{d} \frac{-\sigma_v^2/2 - Cov(v_{t-1}, v_t)}{\overline{\sigma}_v^2\left\{\int_0^1 J_c^H(r)^2 dr - \left(\int_0^1 J_c^H(r)dr\right)^2\right\}}.$$

Eqs. (8) and (9) give the expressions of σ_v^2 and $\overline{\sigma}_v^2$, respectively. Some simple calculations yield the following formula:

$$-\frac{\sigma_v^2}{2\overline{\sigma}_v^2}-\frac{Cov\left(v_{t-1},v_t\right)}{\overline{\sigma}_v^2}=-\frac{(1+2d)\,\Gamma\,(1+d)}{2\Gamma\,(1-d)}-\frac{(1+2d)\,[\Gamma\,(1+d)]^2}{\Gamma\,(d)\,\Gamma\,(2-d)}$$
$$=-\frac{(0.5+d)\,\Gamma\,(2+d)}{\Gamma\,(2-d)}.$$

This equation completes the proof of the formula (17) in Theorem 2.2.

When $d=0$, v_t becomes an i.i.d. sequence that has zero autocovariances and $\sigma_v^2=\sigma_e^2$. Then, it has $Cov\left(v_{t-1},v_t\right)=0$,

$$\frac{1}{T}\sum_{t=2}^{T}\varepsilon_{t-1}\varepsilon_t=\frac{1}{T}\sum_{t=2}^{T}v_{t-1}v_t-\frac{\theta_T}{T}\sum_{t=2}^{T}w_{t-1}^2+o_p\,(1)\xrightarrow{p}-\sigma_w^2,$$

and

$$\frac{1}{T}\sum_{t=2}^{T}y_{t-2}\varepsilon_t=\frac{1}{T\theta_T}\sum_{t=2}^{T}y_{t-1}\varepsilon_t-\frac{1}{T\theta_T}\sum_{t=2}^{T}\varepsilon_{t-1}\varepsilon_t-o_p\,(1)$$
$$\xrightarrow{p}\frac{\sigma_e^2}{2}\left\{J_c(1)^2-2c\int_0^1 J_c(r)^2dr-1\right\}-\sigma_w^2+\sigma_w^2$$
$$=\sigma_e^2\int_0^1 J_c(r)dW\,(r)\,,$$

where the term σ_w^2 is cancelled out. We can further obtain

$$T\left(\widehat{\theta}_T^{IV}-\theta_T\right)=\frac{\frac{1}{T}\sum\limits_{t=2}^{T}y_{t-2}\varepsilon_t-\frac{T}{T-1}\left(\frac{1}{T^{1/2}}\sum\limits_{t=2}^{T}\varepsilon_t\right)\left(\frac{1}{T^{3/2}}\sum\limits_{t=2}^{T}y_{t-2}\right)}{\frac{1}{T^2}\sum\limits_{t=2}^{T}y_{t-2}y_{t-1}-\frac{T}{T-1}\left(\frac{1}{T^{3/2}}\sum\limits_{t=2}^{T}y_{t-2}\right)\left(\frac{1}{T^{3/2}}\sum\limits_{t=2}^{T}y_{t-1}\right)}$$
$$\xrightarrow{d}\frac{\int_0^1 J_c(r)dW(r)-W(1)\int_0^1 J_c(r)dr}{\int_0^1 J_c(r)^2dr-\left(\int_0^1 J_c(r)dr\right)^2}=\frac{\int_0^1\overline{J}_c(r)dW(r)}{\int_0^1\overline{J}_c(r)^2dr},$$

which gives the result in the formula (18) of Theorem 2.2.

For the case where $d>0$, it has $H>0.5$ and

$$\frac{1}{T^{2H}}\sum_{t=2}^{T}y_{t-2}\varepsilon_t=\frac{1}{T^{2H}\theta_T}\sum_{t=2}^{T}y_{t-1}\varepsilon_t-\frac{1}{T^{2H}\theta_T}\sum_{t=2}^{T}\varepsilon_{t-1}\varepsilon_t-o_p\,(1)$$
$$=\frac{1}{T^{2H}\theta_T}\sum_{t=2}^{T}y_{t-1}\varepsilon_t+o_p\,(1)\,.$$

As a result, the IV estimator $\widehat{\theta}_T^{IV}$ has the same limiting distribution as the LS estimator $\widehat{\theta}_T$, as given in the formula (19) of Theorem 2.2. □

Proof of Lemma A.2. Under Assumptions (5), Phillips (1987a) prove that, as $T\to\infty$

$$\frac{1}{\sqrt{T}}\xi_{\lfloor Tr\rfloor}\Rightarrow\overline{\sigma}_vJ_c(r).$$

Note that $y_t=\mu+\xi_t+w_t$. Under Assumption (4), it is easy to get

$$\sup_{r\in(0,1)}\left|\frac{y_{\lfloor Tr\rfloor}}{\sqrt{T}}-\frac{\xi_{\lfloor Tr\rfloor}}{\sqrt{T}}\right|=\sup_{r\in(0,1)}\left|\frac{\mu+w_{\lfloor Tr\rfloor}}{\sqrt{T}}\right|\xrightarrow{p}0.$$

Therefore, it has $T^{-1/2}y_{\lfloor Tr\rfloor}\Rightarrow\overline{\sigma}_vJ_c(r)$, as $T\to\infty$. Lemma A.2.1 is proved.

The results in Lemma A.2.2 and A.2.3 are obtained by directly applying the result of Lemma A.2.1 together with the continuous mapping theorem.

To prove Lemma A.2.4, we first note that, as $T \to \infty$, it has

$$\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t = \frac{1}{T}\sum_{t=1}^{T}(v_t + w_t - \theta_T w_{t-1}) \xrightarrow{p} 0,$$

and

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2 &= \frac{1}{T}\sum_{t=1}^{T}\left[v_t^2 + (w_t - \theta_T w_{t-1})^2 + 2v_t\left(w_t - \theta_T w_{t-1}\right)\right]\\ &= \frac{1}{T}\sum_{t=1}^{T}\left(v_t^2 + w_t^2 + \theta_T^2 w_{t-1}^2\right) + o_p(1)\\ &\xrightarrow{p} \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}E\left(v_t^2\right) + \lim_{T\to\infty}\frac{2}{T}\sum_{t=1}^{T}E\left(w_t^2\right) = \sigma_v^2 + 2\sigma_w^2,\end{aligned}$$

where the two limit results come from the result of McLeish (1975, Theorem 2.10 with condition (2.12)); see also Page 297 of Phillips, 1987b). From the AR representation of y_t given in Eq. (2), it is easy to get that

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

which, in turn, leads to

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

To prove Lemma A.2.5, note that $\gamma_1 = \operatorname{plim}_{T\to\infty} T^{-1}\sum\limits_{t=2}^{T} v_{t-1}v_t$ exists. Then, as $T \to \infty$, it has

$$\begin{aligned}\frac{1}{T}\sum_{t=2}^{T}\varepsilon_{t-1}\varepsilon_t &= \frac{1}{T}\sum_{t=2}^{T}\left(v_{t-1} + w_{t-1} - \theta_T w_{t-2}\right)\left(v_t + w_t - \theta_T w_{t-1}\right)\\ &= \frac{1}{T}\sum_{t=2}^{T}v_{t-1}v_t - \frac{\theta_T}{T}\sum_{t=2}^{T}w_{t-1}^2 + o_p(1)\\ &\xrightarrow{p} \gamma_1 - \sigma_w^2.\end{aligned}$$

We then have

$$\begin{aligned}\frac{1}{T}\sum_{t=2}^{T}y_{t-2}\varepsilon_t &= \frac{1}{T}\frac{1}{\theta_T}\sum_{t=2}^{T}\left[y_{t-1} - (1-\theta_T)\mu - \varepsilon_{t-1}\right]\varepsilon_t\\ &= \frac{1}{\theta_T}\frac{1}{T}\sum_{t=2}^{T}y_{t-1}\varepsilon_t - \frac{1}{\theta_T}\frac{1}{T}\sum_{t=2}^{T}\varepsilon_{t-1}\varepsilon_t + o_p\left(\frac{1}{T}\right)\\ &\xrightarrow{d} \overline{\sigma}_v^2\int_0^1 J_c(r)dW(r) + \frac{1}{2}\left(\overline{\sigma}_v^2 - \sigma_v^2 - 2\gamma_1\right),\end{aligned}$$

where the first equation is from the AR representation of y_t given in Eq. (2).

To prove Lemma A.2.6, it has

$$\frac{1}{T^2}\sum_{t=2}^{T} y_{t-2}y_{t-1} = \frac{1}{T^2}\sum_{t=2}^{T} y_{t-2}\left[(1-\theta_T)\,\mu + \theta_T y_{t-2} + \varepsilon_{t-1}\right]$$
$$= \frac{\theta_T}{T^2}\sum_{t=2}^{T} y_{t-2}^2 + o_p(1) \xrightarrow{d} \overline{\sigma}_v^2 \int_0^1 J_c(r)^2 dr.$$

The proof is completed. □

Proof of Theorem 3.1. The results in Theorem 3.1 follow directly from the continuous mapping theorem and Lemma A.2. Details are omitted. □

Acknowledgment

We would like to thank two referees, an associate editor, Peter Phillips, Liyu Dou, Yiu Lim Lui, Jia Li, Shuping Shi, and Yongmiao Hong for helpful discussion. Xiaohu Wang, School of Economics, Fudan University, and Shanghai Institute of International Finance and Economics, Shanghai, China. Email: wang_xh@fudan.edu.cn. 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.

Funding

Wang acknowledges financial support from National Natural Science Foundation of China (NSFC) (No. G0301), and from Shanghai Pujiang Program (No. 22P JC022). Yu acknowledges that this research/project is supported by the Ministry of Education, Singapore, under its Academic Research Fund (AcRF) Tier 2 (Award Number MOE-T2EP402A20-0002). He also acknowledges the financial support from the Lee Foundation.

References

Andersen, T. G., 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. (2001a). 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. (2001b). 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.

Baillie, R. (1996). Long memory processes and fractional integration in econometrics. *Journal of Econometrics* 73(1):5–59.

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

Beran, J. (1992). Statistical methods for data with long-range dependence. *Statistical Science* 7(4):404–416.

Bolko, A., Christensen, K., Pakkanen, M. S., Veliyev, B. (2023). A GMM approach to estimate the roughness of stochastic volatility. *Journal of Econometrics* 235(2):745–778. forthcoming.

Bound, J., Brown, C., Mathiowetz, N. (2001). Measurement error in survey data. *In: Handbook of Econometrics*, Heckman, J. J., Leamer, E., eds., Vol. 5, pp. 3705–3843. Elsevier. Amsterdam, Holland.

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.

Buchmann, B., Chan, N. G. (2007). Asymptotic theory of least squares estimators for nearly unstable processes under strong dependence. *Annals of Statistics* 35(5):2001-2017.

Campbell, J. Y., Motohiro, Y. (2006). Efficient tests of stock return predictability. *Journal of Financial Economics* 81:27–60.

Cavanagh, C. L., Elliott, G., Stock, J. H. (1995). Inference in models with nearly integrated regressors. *Econometric Theory* 11:1131–1147.

Chambers, M. (1998). Long memory and aggregation in macroeconomic time series. *International Economic Review* 39:1053–1072.
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.
Davydov, Y. A. (1970). The invariance principle for stationary processes. *Theory of Probability and Its Applications* 15: 487–489.
Diebold, F. X., Rudebusch, G. D. (1989). Long memory and persistence in aggregate output. *Journal of Monetary Economics* 24(2):189–209.
Dou, L., Müller, U. (2021). Generalized local-TO-Unity models. *Econometrica* 89:1825–1854.
Duffy, J. A., Kasparis, I. (2021). Estimation and inference in the presence of fractional $d = 0.5$ AND weakly nonstationary processes. *Annals of Statistics* 49(2):1195–1217.
Elliott, G., Rothenberg, T. J., Stock, J. H. (1996). Efficient tests for an autoregressive unit root. *Econometrica* 64:813–836.
Elliott, G., Stock, J. H. (2001). Confidence intervals for autoregressive coefficients near one. *Journal of Econometrics* 103:155–181.
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.
Granger, C. W. J., Joyeux, R. (1980). An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis* 1:15–29.
Hall, A. (1989). Testing for a unit root in the presence of moving average errors. *Biometrika* 76:49–56.
Hansen, B. (1999). The grid bootstrap and the autoregressive model. *Review of Economics and Statistics* 81(4):594–607.
Hansen, P., 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., 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. *In Household Sample Surveys in Developing and Transition Countries, Chapter IX*, pp. 171–198.
Li, J., Phillips, P. C. B., Shi, S., Yu, J. (2022). *Weak identification of long memory with implications for inference*. Working Paper. Singapore Management University, Singapore.
Lieberman, O., Phillips, P. C. B. (2008). Refined inference on long memory in realized volatility. *Econometric Reviews* 27:254–267.
Liu, X., Shi, S., Yu, J. (2021). Fractional stochastic volatility model. Working Paper. Singapore Management University, Singapore.
Mandelbrot, B., Van Ness, J. W. (1968). Fractional brownian motions, fractional noises and applications. *SIAM Review* 10:422–437.
Mcleish, D. L. (1975). A maximal inequality and dependent strong laws. *Annals of Probability* 3(5):829–839.
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(1-2):7–38.
Ng, S., Perron, P. (2001). Lag length selection and the construction of unit root tests with good size and power. *Econometrica* 69(6):1519–1554.
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.
Rossi, B. (2005). Confidence intervals for half-life deviations from purchasing power parity. *Journal of Business & Economic Statistics* 23:432–442.
Samorodnitsky, G., Taqqu, M. S. (1994). *Stable non-Gaussian random processes: Stochastic models with infinite variance. Stochastic modeling*. New York: Chapman & Hall.

Schwert, G. W. (1989). Tests for unit roots: A monte carlo investigation. *Journal of Business & Economic Statistics* 7:147–159.

Shi, S., Yu, J. (2022). Volatility puzzle: long memory or anti-persistency. *Management Science*, forthcoming.

Sowell, F. (1990). 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.

Taqqu, M. S. (1975). Weak convergence to fractional Brownian motion and to the Rosenblatt process. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 31:287–302.

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. (2023a). Modeling and forecasting realized volatility with the fractional Ornstein-Uhlenbeck process. *Journal of Econometrics* 232:389–415.

Wang, X., Xiao, W., Yu, J. (2023b). Asymptotic properties of least squares estimator in local to unity processes with the fractional Gaussian noises. *Advances in Econometrics*, 45A, 73–95.

Wright, J. H. (2000). Confidence intervals for univariate impulse responses with a near unit root. *Journal of Business & Economics Statistics* 18:368–373.