Further Results on Pseudo-Maximum Likelihood Estimation and Testing in the Constant Elasticity of Variance Continuous Time Model*

Emma M. Iglesias† Garry D. A. Phillips‡

University of A Coruña University of Exeter

This version: July 2019

Abstract

Constant elasticity volatility processes have been shown to be useful, for example, to encompass a number of existing models that have closed-form likelihood functions. In this paper we extend the existing literature in two directions: *First* we find explicit closed form solutions of the pseudo MLEs by discretizing the diffusion function and we provide their asymptotic theory in the context of the constant elasticity of variance model characterized by a general constant elasticity of variance parameter $\rho \geq 0$. *Second* we obtain bias expansions for those pseudo MLEs also in terms of $\rho \geq 0$. We provide a general framework since only the cases with $\rho = 0$ and $\rho = 0.5$ have been considered in the literature so far. When the time series is not positive almost surely, we need to impose the restriction that ρ is a non-negative integer.

Key words: Least Squares; Quasi-Maximum Likelihood; Continuous Record; Estimation; Testing; Bias Correction; Diffusion Processes. **JEL Classification:** C12; C22. **MSC Classification:** 62F12; 62M10; 62P20.

*We wish to thank the Editor and two referees for very helpful comments. The first author is very grateful for the financial support from the Spanish Ministry of Science and Innovation, projects ECO2015-63845-P and PGC2018-101327-B-100.

†Address: Department of Economics. Universidade da Coruña, Campus de A Coruña. 15071, A Coruña. Spain. E-mail: emma.iglesias@udc.es. Webpage: www.gcd.udc.es/emma

‡Address: Economics Dept., Business School, University of Exeter, Exeter EX4 4PU, United Kingdom. E-mail: G.D.A.Phillips@exeter.ac.uk

1 Introduction

Continuous time models have been proved to be very successful in economic theory (see e.g. Yu (2014)). Continuous time diffusion processes have been commonly and successfully used in economics and finance to model stochastic dynamics of financial securities following the paper by Black and Scholes (1973) which established the foundation of option pricing theory. An important family of diffusion processes is the general linear drift process proposed by Chan, Karolyi, Longstaff and Sanders (1992, CKLS from now onwards) where the diffusion function can accommodate a wide range of patterns of volatility. In this model, the diffusion function follows a form where a constant elasticity of variance (CEV) parameter ρ plays a crucial role (Bu et al (2011) have shown its usefulness, for example, to encompass a number of existing models that have closed-form likelihood functions). Important members of this family of models are the Vasicek (1977) model with $\rho = 0$; the CIR model (Cox et al (1985)) with $\rho = 0.5$; the Brennan and Schwartz (1980) model with $\rho = 1$, and the CIR- VR model (Cox et al. (1980) and Ahn and Gao (1999)) with $\rho = 1.5$.

It is well known in the literature via simulation studies (see e.g. Ball and Torous (1996) and Yu and Phillips (2001)) that the estimation of the drift parameters in the CKLS model yields biased estimators especially for the mean reversion parameter both in finite discrete samples and in large in-fill samples. Currently there are five papers in relation to this issue: (1) Tang and Chen (2009) derived analytical expressions for approximating the bias and variance of pseudo maximum likelihood estimators (PMLEs) using Nowman´s (1997) method that can be used to improve the estimation of the CKLS model. But their expressions are only valid for $\rho = 0$ and 0.5 and the performance of their bias formula is unsatisfactory in the near unit root situations (this corresponds to the slow mean reversion case which is empirically realistic for financial time series). (2) Yu (2012) adds an extra term to the bias approximations of Tang and Chen (2009) and this helps to improve the performance of the bias expressions when the speed of mean reversion is slow. But only the case where $\rho = 0$ is analyzed and only the bias of the estimator of the mean reversion parameter is given. (3) Iglesias (2014) shows that the expressions provided by Tang and Chen (2009) and Yu (2012) are not only useful for bias correction purposes in estimating continuous time models but also for testing using a t-statistic in the near unit root situation. Again, only the model where $\rho = 0$ is analyzed. And finally, (4) Bao et al (2015) give a bias approximation for the mean reversion estimator in continuous-time Lévy processes while more recently, (5) Bao et al (2017) focus on the case of $\rho = 0$ while deriving the exact distribution of the MLE. However, in practice, models with ρ different from 0 and 0.5 are needed; and

Ball and Torous (1996, 1999), Ahn and Gao (1999) and Czellar, Karloyi and Rochetti (2007) in particular show that this is the case.

In this paper, we plan to extend the results of the previous five papers in several directions: (1) we find explicit closed form solutions for the pseudo maximum likelihood estimators (PMLEs) in a general CKLS model for $\rho \geq 0$ and we also provide the asymptotic theory. (2) We derive analytical bias expressions that can be used when estimating a general CKLS (1992) model and we show the usefulness of the PMLEs versus alternative estimation methods such as the jackknife of Phillips and Yu (2005) and finally (3), we specialize our results for $\rho = 0$ and 0.5 to compare them with those of Tang and Chen (2009).

The plan of the paper is as follows. Section 2 presents the model that is the object of our study, the closed form solutions of the PMLEs and the asymptotic theory. Section 3 refers to the bias expansions for the PMLEs. Section 4 provides simulation results to show the usefulness of the bias corrected PMLEs. Finally, Section 5 concludes. A supporting information file contains the proofs of our main results.

2 Model, estimators and asymptotic theory

Following CKLS (1992), we analyze the stochastic process of CEV with mean-reverting drift

$$dx_t = \kappa\left(\mu - x_t\right)dt + \sigma x_t^{\rho} dB_t, \tag{1}$$

where B_t is a standard Brownian motion, σ is a volatility coefficient, κ is the mean reversion parameter and μ represents the long term mean. Parameter $\rho \geq 0$ shows the degree to which the standard deviation σx_t^{ρ} depends on x_t (i.e. the elasticity of volatility with respect to x_t). Examples already well known in the literature are $\rho = 0$ (Vasicek (1977)), $\rho = 0.5$ (Cox, Ingersoll and Ross (1985)), $\rho = 1$ (Brennan and Schwartz (1980)), and $\rho = 1.5$ (CIR VR model, Cox et al. (1980) and Ahn and Gao (1999), known as the *inverse square-root model*). We assume that $\kappa > 0$. See for example Hurn, Jeismand and Lindsay (2007) for a review of different estimation methods, and note that for cases different from $\rho = 0$ and 0.5, the behaviour of Nowman´s estimator has not yet been considered in the literature.

Let $x_0, x_\delta, ..., x_{n\delta}$ be discrete observations from the process (1) while δ is the sampling interval, n is the sample size and we define $n\delta = T$. In order to estimate (1), we use the discrete form. Nowman (1997), using Bergstrom´s (1984) approximation, discretized the diffusion function computing the following approximate discrete time series model

$$x_t = e^{-\delta\kappa} x_{t-1} + \mu\left(1 - e^{-\delta\kappa}\right) + \varepsilon_t, \tag{2}$$

where the ε_t are assumed normal, uncorrelated, with conditional mean $E(\varepsilon_t|x_{t-1}) = 0$ and conditional variance

$$Var\left(\varepsilon_t|x_{t-1}\right) = 0.5\sigma^2\kappa^{-1}\left(1-e^{-2\delta\kappa}\right)x_{t-1}^{2\rho}. \tag{3}$$

From (3), it is clear that if the time series can take negative values, we need ρ to be a general non-negative integer. If we denote $\theta = (\kappa, \mu, \sigma^2)'$, the PMLEs are given by maximizing the conditional pseudo log-likelihood $LogL\left(\theta\right) = -\frac{1}{2}\sum_{t=1}^{n}\left(\ln Var\left(\varepsilon_t|x_{t-1}\right) + \frac{\varepsilon_t^2}{\sqrt{Var(\varepsilon_t|x_{t-1})}}\right)$, where

PROPOSITION 1. *Assuming ε_t to be Gaussian in (2) and either (a) x_t is a non-negative series and $\rho > 0$ or (b) x_t is unrestricted and ρ is a general non-negative integer, the PMLEs $\widehat{\kappa}$, $\widehat{\mu}$ and $\widehat{\sigma}^2$ are given as $\widehat{\kappa} = -\delta^{-1}\log\left(\widehat{\beta}_1\right)$, $\widehat{\mu} = \widehat{\beta}_2$, $\widehat{\sigma}^2 = \frac{2\widehat{\kappa}\widehat{\beta}_3}{\left(1-\widehat{\beta}_1^2\right)}$, where we condition on the starting value and*

$$\widehat{\beta}_1 = \frac{\sum_{t=1}^{n} x_t x_{t-1}^{(1-2\rho)} \sum_{t=1}^{n} x_{t-1}^{-2\rho} - \sum_{t=1}^{n} x_t x_{t-1}^{-2\rho} \sum_{t=1}^{n} x_{t-1}^{(1-2\rho)}}{\sum_{t=1}^{n} x_{t-1}^{(2-2\rho)} \sum_{t=1}^{n} x_{t-1}^{-2\rho} - \left(\sum_{t=1}^{n}\left(x_{t-1}^{(1-2\rho)}\right)\right)^2}, \tag{4}$$

$$\widehat{\beta}_2 = \frac{\sum_{t=1}^{n}\left(x_t x_{t-1}^{-2\rho} - \widehat{\beta}_1 x_{t-1}^{(1-2\rho)}\right)}{\left(1-\widehat{\beta}_1\right)\sum_{t=1}^{n} x_{t-1}^{-2\rho}},\ \widehat{\beta}_3 = n^{-1}\sum_{t=1}^{n}\left(x_t - \widehat{\beta}_1 x_{t-1} - \widehat{\beta}_2\left(1-\widehat{\beta}_1\right)\right)^2 x_{t-1}^{-2\rho}. \tag{5}$$

If in Proposition 1 we set $\rho = 0$, and $\rho = 0.5$, we obtain the special cases given in Tang and Chen (2009, pages 66-67, equations (2.5) and (2.13)). Note that for $\rho = 0.5$ we are in the case of the time series not taking negative values. Moreover, moments of $\widehat{\kappa}$ may not exist at all since $\widehat{\beta}_1$ may be negative, and this is a characteristic of Nowman´s estimator. Also, from Proposition 1 we have the following two Remarks

Remark 1 *In Proposition 1, we need ρ to be known although obviously it would be more general if we could find an estimator for ρ. We have tried that, but the closed-form expressions we obtained for Nowman´s method become intractable.*

Remark 2 *It is important to note that from Proposition 1 and (3), if we want to estimate the CKLS model with our PMLEs and to use them with a series that may take positive and negative values, we need ρ to be a non-negative integer, since we need $x_t^{-2\rho}$ to exist.*

We generalize now the asymptotic theory of Nowman´s estimator in Tang and Chen (2009) in the following

THEOREM 1. *For a stationary CKLS process*[1], *as* $n \longrightarrow \infty$ *while* δ *(the sampling interval) is fixed, and assuming either (a)* x_t *is a series taking only non-negative values and* $\rho > 0$ *or (b)* x_t *is unrestricted and* ρ *is a general non-negative integer, let* $\widehat{\theta} = (\hat{\kappa}, \widehat{\mu},\ \widehat{\sigma}^2)'$, *and* $\widetilde{\theta} = (\kappa, \mu,\ \sigma^2 - B(\theta, \delta))'$, *where* $B(\theta, \delta)$ *is the inconsistency term related to* σ^2 *and let* $E(\varepsilon_t | x_{t-1}) = 0$. *Then* $\sqrt{n}\left(\widehat{\theta} - \widetilde{\theta}\right) \xrightarrow{d} N(0, \Omega)$ *where* $\Omega = \Lambda^{-1}$ *with* $\Lambda = E\left(-\frac{1}{n}\frac{\partial^2 LogL(\theta)}{\partial\theta\partial\theta'}\right) = \begin{pmatrix} \overline{A} & \overline{B} & \overline{C} \\ \overline{B} & \overline{D} & 0 \\ \overline{C} & 0 & \overline{E} \end{pmatrix}$, *and* $\overline{A} = \frac{\left(e^{-2\kappa\delta} + 2\kappa\delta e^{-2\kappa\delta} - 1\right)^2}{2\kappa^2\left(e^{-2\kappa\delta} - 1\right)^2} + \frac{2\delta^2 e^{-2\delta\kappa} E\left(\left(x_{t-1}^{(1-\rho)} - \mu x_{t-1}^{-\rho}\right)^2\right)}{\sigma^2\kappa^{-1}\left(1 - e^{-2\delta\kappa}\right)}$;
$\overline{B} = -\frac{2\kappa\delta e^{-\delta\kappa} E\left(x_{t-1}^{(1-2\rho)} - \mu x_{t-1}^{-2\rho}\right)}{\sigma^2\left(1 + e^{-\delta\kappa}\right)}$; $\overline{C} = \frac{\delta e^{-2\delta\kappa}}{\sigma^2\left(1 - e^{-2\delta\kappa}\right)} - \frac{1}{2\kappa\sigma^2}$; $\overline{D} = \frac{2\kappa\left(e^{\delta\kappa} - 1\right)}{\sigma^2\left(e^{\delta\kappa} + 1\right)} E\left(\frac{1}{x_{t-1}^{2\rho}}\right)$; $\overline{E} = \frac{1}{2\sigma^4}$.

Also for $\rho = 0$, $B(\theta, \delta) = 0$.

There are many alternative estimators such as the Aït-Sahalia (1999, 2008) approximate likelihood method, simulation based methods (see for example Beskos et al (2009)) and many more that we may use to estimate the parameters in (1). However, the main advantage of Nowman´s (1997) method lies in its analytical tractability which justifies its use as an estimator that is studied and used in many papers such as in Tang and Chen (2009). Moreover, we show in Section 4 that Nowman´s method can be useful in practice and why it makes sense to analyze its asymptotic and finite sample properties. Tang and Chen (2009) propose the use of the bootstrap method and Bianchi and Cleur (1996) use indirect inference. But both the indirect inference and bootstrap methods are computationally expensive versus Nowman´s method.

3 Expansions for the bias parameter estimators in the CKLS model

We proceed now to extend the results of Tang and Chen (2009, Theorems 3.1.1 and 3.1.3, pages 68-69) by deriving analytical bias expressions that can be used in practice when estimating a general CKLS (1992) model. Following Tang and Chen (2009, proof of Theorem 3.1.3) we first note that from (2), then

$$E(x_j | x_i) = e^{-\delta_{ij}\kappa} x_i + \mu\left(1 - e^{-\delta_{ij}\kappa}\right) \text{ with } \delta_{ij} = \delta\,|j - i|\,. \tag{6}$$

Let $t_{1i} = x_i x_{i-1}^{(1-2\rho)} - \mu_1$, $t_{2i} = x_{i-1}^{-2\rho} - \mu_2$, $t_{3i} = x_i x_{i-1}^{-2\rho} - \mu_3$, $t_{4i} = x_{i-1}^{(1-2\rho)} - \mu_4$ and $t_{5i} = x_{i-1}^{(2-2\rho)} - \mu_5$, where from (6), $\mu_1 = e^{-\delta\kappa}\mu_5 + \mu\left(1 - e^{-\delta\kappa}\right)\mu_4$, $\mu_3 = e^{-\delta\kappa}\mu_4 + \mu\left(1 - e^{-\delta\kappa}\right)\mu_2$, $\mu_2 = E\left(x_{t-1}^{-2\rho}\right)$, $\mu_4 = E\left(x_{t-1}^{(1-2\rho)}\right)$

[1]See Conley, Hansen, Luttmer and Scheinkman (1997) for details of primitive conditions under which the CEV process is stationary and ergodic. Broze et al (1995) also provided conditions for second-order stationarity and ergodicity.

and $\mu_5 = E\left(x_{t-1}^{(2-2\rho)}\right)$. Also define $t_a = n^{-1}\sum_{i=1}^{n} t_{ai}$ which is $O_p(n^{\frac{-1}{2}})$ and let $\widetilde{t}_a = n^{-1}\sum_{i=1}^{n} t_{a(i-1)} = t_a + n^{-1}\left(t_{a0} - t_{an}\right) = t_a + O_p\left(n^{-1}\right)$, for $a = 1, .., 5$. In addition let $\mu_u = \mu_1\mu_2 - \mu_3\mu_4$, $\mu_d = \mu_5\mu_2 - \mu_4^2$.

In what follows, Theorem 2 shows the consistency and bias approximations when estimating κ, μ and σ^2 in model (2). Recalling that $\widehat{\kappa} = -\delta^{-1}\log\left(\hat{\beta}_1\right)$, our approach to analysing the bias of $\widehat{\kappa}$ is to first find a suitable expansion for $\hat{\beta}_1$,which subsumes the expansion used by Tang and Chen (2009), and then find an appropriate expansion for the transform. We first find that $\hat{\kappa} - \kappa = -\frac{1}{\delta\beta_1}\left(\widehat{\beta}_1 - \beta_1\right) + \frac{1}{2\delta\beta_1^2}\left(\widehat{\beta}_1 - \beta_1\right)^2 + O\left(n^{-2}\right)$ from which we obtain the following

THEOREM 2. *For a stationary CKLS process, as $n \longrightarrow \infty$ while δ is fixed and assuming either (a) x_t is a series taking only non-negative values and $\rho > 0$ or (b) x_t is unrestricted and ρ is a general non-negative integer, and $E(\varepsilon_t|x_{t-1}) = 0$, the bias of the estimator $\hat{\kappa}$ is given by*

$$\begin{aligned}
E(\hat{\kappa}-\kappa) = {} & \frac{\mu_2^2 var(t_1)}{2\delta e^{-2\kappa\delta}\mu_d^2} + (\mu_1\mu_5 + \frac{(\mu_1\mu_d - \mu_5\mu_u)^2}{2e^{-\kappa\delta}\mu_d^2} - \frac{\mu_u}{\mu_d}\mu_5^2)\frac{var(t_2)}{\delta e^{-\kappa\delta}\mu_d^2} + (2\mu_3\mu_4 + \frac{\mu_3^2}{2e^{-\kappa\delta}} - \mu_u - \frac{4\mu_u\mu_4^2}{\mu_d})\frac{var(t_4)}{\delta e^{-\kappa\delta}\mu_d^2} \\
& + \frac{\mu_4^2 var(t_3)}{2\delta e^{-2\kappa\delta}\mu_d^2} + (\frac{\mu_u^2}{2\mu_d^2 e^{-\kappa\delta}} - \frac{\mu_u}{\mu_d})\frac{\mu_2^2 var(t_5)}{\delta e^{-\kappa\delta}\mu_d^2} - [1 - \frac{\mu_2\mu_5}{\mu_d} - \frac{\mu_2}{e^{-\kappa\delta}}(\frac{\mu_1}{\mu_d} - \frac{\mu_5\mu_u}{\mu_d^2})]\frac{Cov(t_1t_2)}{\mu_d\delta e^{-\kappa\delta}} \\
& - \frac{\mu_2\mu_4 Cov(t_1t_3)}{\delta e^{-2\kappa\delta}\mu_d^2} - (2\mu_2\mu_4 + \frac{\mu_1\mu_3}{e^{-\kappa\delta}})\frac{Cov(t_1t_4)}{\mu_d^2\delta e^{-\kappa\delta}} + (\frac{\mu_2^2}{\mu_d} - \frac{\mu_2^2\mu_u}{\mu_d^2 e^{-\kappa\delta}})\frac{Cov(t_1t_5)}{\delta e^{-\kappa\delta}\mu_d} \\
& - [\frac{\mu_5}{\mu_d} + \frac{\mu_1}{e^{-\kappa\delta}}(\frac{1}{\mu_d} - \frac{\mu_u}{\mu_d^2})]\frac{\mu_4 Cov(t_2t_3)}{\delta e^{-\kappa\delta}\mu_d} + [\frac{4\mu_5\mu_4\mu_u}{\mu_d^2} - \frac{2\mu_1\mu_4 + \mu_3\mu_5}{\mu_d} - \frac{\mu_3}{e^{-\kappa\delta}}(\frac{\mu_1}{\mu_d} - \frac{\mu_5\mu_u}{\mu_d^2})]\frac{Cov(t_2t_4)}{\delta e^{-\kappa\delta}\mu_d} \\
& + [\frac{\mu_1\mu_2}{\mu_d} + \frac{\mu_u}{\mu_d} - \frac{2\mu_5\mu_2\mu_u}{\mu_d^2} + \frac{\mu_2}{e^{-\kappa\delta}}(\frac{\mu_u^2\mu_5}{\mu_d^3} - \frac{\mu_1\mu_u}{\mu_d^2})]\frac{Cov(t_2t_5)}{\delta e^{-\kappa\delta}\mu_d} + (1 + \frac{2\mu_4^2}{\mu_d} + \frac{\mu_3\mu_4}{\mu_d e^{-\kappa\delta}})\frac{Cov(t_3t_4)}{\delta e^{-\kappa\delta}\mu_d} \\
& + (\frac{\mu_u}{\mu_d^2 e^{-\kappa\delta}} - \frac{1}{\mu_d})\frac{Cov(t_3t_5)\mu_4\mu_2}{\delta e^{-\kappa\delta}\mu_d} + (-\frac{\mu_2\mu_3}{\mu_d} + 2\frac{(\mu_4 + \mu_1)\mu_u\mu_2}{\mu_d^2} + \frac{\mu_2\mu_3\mu_u}{\mu_d^2 e^{-\kappa\delta}})\frac{Cov(t_4t_5)}{\delta e^{-\kappa\delta}\mu_d} + o(n^{-1}).
\end{aligned}$$

for $e^{-\kappa\delta} \neq 1$ and where all variances and covariances are of order n^{-1}. Also $E\left(\widehat{\mu} - \mu\right) = O\left(n^{-2}\right)$ for $\rho = 0$, and assuming either (a) x_t is a series taking only positive values and $\rho > 0$ or (b) ρ is a general positive integer

$$E\left(\widehat{\mu} - \mu\right) = E\left(\frac{\sum_{t=1}^{n} x_{t-1}^{-2p}\varepsilon_t}{(1 - e^{-\kappa\delta})\sum_{t=1}^{n} x_{t-1}^{-2p}}\right) + E\left(\frac{\overline{F}\sum_{t=1}^{n} x_{t-1}^{-2p}\varepsilon_t}{(1 - e^{-\kappa\delta})\sum_{t=1}^{n} x_{t-1}^{-2p}}\right) + o\left(n^{-1}\right)$$

$$\text{with } \overline{F} = \frac{\mu_1 t_2 + \mu_2 t_1 - \mu_3 t_4 - \mu_4 t_3}{\mu_d} - \frac{\mu_u}{\mu_d}\left(\mu_5 t_2 + \mu_2 t_5\right) + o_p(n^{-1/2}).$$

The approximate discrete time series model used here given in (2) has the same form as the $AR(1)$ model which has been much studied in the time series literature. In fact, rewriting the equation in (2) in the form

$x_t = \beta x_{t-1} + \mu(1-\beta) + \varepsilon_t$, the well-known paper by Kendall (1954) derived the $O(T^{-1})$ bias for the slope coefficient to be $-(\frac{1-3\beta}{T})$ which is one of the best known approximations in the time series literature. The derivation assumed that the model was stationary and that the disturbances were NID $(0, \sigma^2)$. In the case we study here the disturbances are again assumed to be normally and independently distributed but with a variance that depends on $x_{t-1,}$ $i.e$, (3) for a given value of $\rho \geqslant 0$. It is assumed that the sampling interval δ is positive and fixed so that the above model, termed the discretized model, can be estimated directly and analysed as in the time series case; for example, asymptotic expansions for estimation errors can be employed along the lines of the original Kendall (1954) paper as is done here although the derivations of bias approximations are far more complicated.

4 Simulation results: Evidence of the usefulness of bias corrected PMLE

When estimating the parameters of the model, CKLS (1992) relied on the Generalized Method of Moments (GMM, see Hansen (1982)). Kladívko (2008) shows the importance of choosing a suitable variance-covariance matrix estimator of moment functions when applying GMM to the CKLS model. We also computed the GMM procedure of Kladívko (2008) in our simulations but we do not show the results because the empirical size distortions in the test procedure were very large (the test was extremely liberal). Another possibility could be to improve the GMM procedure, but it would be more computationally involved than PML and it is not the objective of this paper. Besides, when using our bias corrected PMLEs, we avoid specifying a variance-covariance matrix estimator of moment functions as in the GMM procedure.

Tang and Chen (2009, Tables 1, 2 and 3) already showed the usefulness of the bias corrected PMLEs from the estimation point of view in relation to bias and root mean squared error criteria when $\rho = 0$ and $\rho = 0.5$, so in the simulation section of this paper we focus on the testing side. Iglesias (2014) showed that for the case of $\rho = 0$, the bias expressions of Tang and Chen (2009) and Yu (2012) are useful also for testing purposes. In what follows, we show that our closed form solutions in Proposition 1 of the PMLEs and their bias corrected expressions of our Theorem 2, when estimating a CKLS model for a more general ρ, are very useful from the testing point of view when compared to alternative methods such as the Jackknife of Phillips and Yu (2005).

We consider the setting of two models as the data generating process in all simulations in Tang and Chen (2009): (1) CIR Model 2, where $\kappa = 0.223$, $\mu = 0.09$ and $\sigma^2 = 0.008$, $T = 10$ and $\delta = 1/12$ and

(2) CIR Model 3, where $\kappa = 0.148$, $\mu = 0.09$ and $\sigma^2 = 0.005$, $T = 10$ and $\delta = 1/12$. In CIR Model 3 the autoregressive coefficient of the discrete time model is 0.99 and the two models, as in Tang and Chen (2009), are designed to check the performance of the parameter estimation in the near unit root case. We draw 10000 simulations, and we construct a standard two-sided t-test for the null hypothesis $H_0 : \kappa = \kappa_0$ at 5% significance level[2] for different values of κ_0. We show the simulation results for $\rho = 0.5$ and $\rho = 1$[3].

When $\rho = 0.5$, and if $2\kappa\mu \geq \sigma^2$, $\kappa > 0$, $\mu > 0$ and $\sigma^2 > 0$ holds, the CIR model is well-defined and it has a steady-state (marginal) distribution. The marginal density is gamma distributed (see Feller (1951)). Note also that a chi-squared random variable with d degrees of freedom (χ_d^2) is equal in distribution to the gamma distribution $\Upsilon(d/2, 1/2)$ (which is the unconditional distribution of the CIR process). Therefore we set $\mu = x_0$ and we simulate the initial condition from a Gamma distribution $\Upsilon(d/2, 1/2)$ with $d = 3$ in order not to violate the condition which ensures stationarity of the CIR model (see Feller (1951)).

In Figures 1-4, we show the results of the empirical power of the t-test using our explicit expressions in Proposition 1 (named POWER), the bias corrected PMLEs given in Theorem 2 (named POWERBC) and using the Jackknife of Phillips and Yu (2005) (named POWERJACKK). When using the Jackknife of Phillips and Yu (2005) and following their suggestion, we construct 4 consecutive non-overlapping blocks of observations. When using our bias corrected estimator from our Theorem 2, we have used a parametric bootstrap by drawing 1000 bootstrap resamples to approximate the variances/covariances. Our results from Figure 1 for CIR Model 2 show that the three methods are very conservative and have an empirical size of 0 under the null hypothesis. This suggests that the asymptotic theory of all these tests works poorly in finite samples in view of the null rejection rate being much lower than the nominal rate; and therefore alternative methods should be investigated in further research mainly to improve on the size results. The main advantage of our proposal comes when analysing the empirical power: the use of our bias corrected PMLEs improves versus using the Jackknife or the PMLEs without bias correcting. The same results hold in Figure 2 for CIR Model 3. Figures 3 and 4 provide the same simulation results as in Figures 1 and 2 but now when $\rho = 1$, and the power gains from using the bias corrected method increase in relation to Figures 1-2. Therefore out of the three methods, we recommend that bias corrected PMLEs be used in practice.

[2] All the simulation results have been obtained in MATLAB.

[3] Broze et al (1995) showed conditions for stationarity when $\rho \leq 1$.

Figures 1-4: Empirical size and power, $H_0 : \kappa = \kappa_0$, for different values of κ_0.

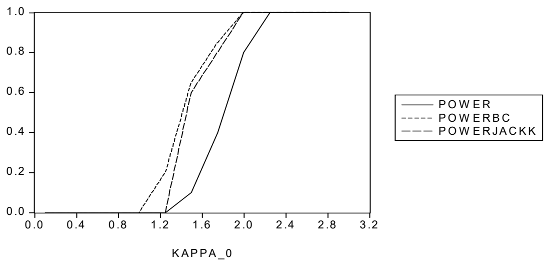

Figure 1: $\rho = 0.5$ CIR Model 2.

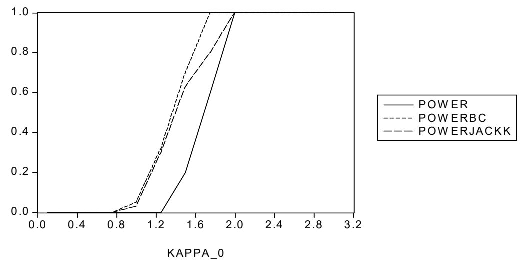

Figure 2: $\rho = 0.5$ CIR Model 3.

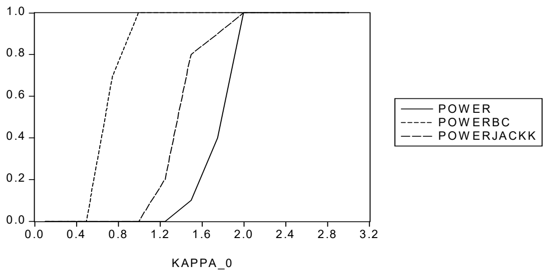

Figure 3: $\rho = 1$ Model 2.

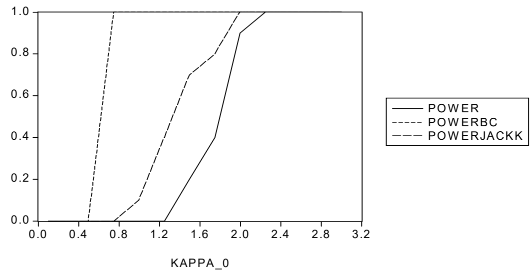

Figure 4: $\rho = 1$ Model 3.

5 Conclusions

We have extended the results in Tang and Chen (2009), Yu (2012) and Bao et al (2015) in two directions. *First* we find explicit closed form solutions of the PMLEs for the general CKLS (1992) model characterized by a general non-negative integer parameter ρ. Our assumption of having a non-negative integer parameter ρ is very simple, and it allows the nesting of popular models in the literature such as the Vasicek (1977) model with $\rho = 0$; the CIR model (Cox et al (1985)) with $\rho = 0.5$ due to the χ^2 nature of the time series in this case; the Brennan and Schwartz (1980) model with $\rho = 1$; but if we impose positivity of the time series, our theory works for any $\rho \geq 0$. We also provide the asymptotic theory for those PMLEs. *Second* we obtain bias expansions for the parameter estimators when used in a general CKLS (1992) model, while again only the cases with $\rho = 0$ and $\rho = 0.5$ were analyzed in the literature so far. We show *inter alia* that the bias of the long term mean parameter estimator is $O\left(n^{-1}\right)$ for any positive ρ value, contradicting the results of Tang and Chen (2009) where it was claimed to be $O\left(n^{-2}\right)$ for $\rho = 0.5$. Finally, we show in simulations the usefulness of our results. Wang, Phillips and Yu (2011) point out that one can often get a lower bias using a cruder approximation than Nowman´s, such as the Euler approximation, as the biases resulting from the discretisation and the estimation often partially cancel one another, so this may be a subject for further research.

6 Supporting information

Additional Supporting Information may be found online in the supporting information tab for this article.

7 Data availability statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

References

[1] Ahn, D. and B. Gao (1999), A parametric nonlinear model of term structure dynamics, *Review of Financial Studies* 12, 721–762.

[2] Aït-Sahalia, Y. (1999), Transition densities for interest rate and other nonlinear diffusions, *Journal of Finance* 54, 1361 1395.

[3] Aït-Sahalia, Y. (2008), Closed-form likelihood expansions for multivariate diffusions, *Annals of Statistics* 36, 906 937.

[4] Ball, C. A. and W. N. Torous, (1996), Unit roots and the estimation of interest rate dynamics, *Journal of Empirical Finance* 3, 215-238.

[5] Ball, C. A. and W. N. Torous (1999), The Stochastic Volatility of Short-Term Interest Rates: Some International Evidence, *The Journal of Finance* 54, 6, 2339-2359.

[6] Bao, Y., A. Ullah, Y. Wang and J. Yu (2015), Bias in the estimation of Mean Reversion in Continuous-Time Lévy Processes, *Economics Letters* 134, 16-19.

[7] Bao, Y., A. Ullah and Y. Wang (2017), Distribution of the Mean Reversion Estimator in the Ornstein-Uhlenbeck Process, *Econometric Reviews* 36, 6-9, 1039-1056.

[8] Bergstrom, A. R. (1984), Continuous time stochastic models and issues of aggregation over time. In: Griliches, Z., Intriligator, M.D. (Eds.), *Handbook of Econometrics*, vol. 2.

[9] Beskos, A., O. Papaspiliopoulos, and G. Roberts (2009), Monte Carlo maximum likelihood estimation for discretely observed diffusion processes, *Annals of Statistics* 37, 1, 223-245.

[10] Bianchi, C. and E. M. Cleur (1996), Indirect estimation of stochastic differential equation models: some computational experiments, *Computational Economics* 9, 3, 257-274.

[11] Black, F. and M. Scholes (1973), The pricing of options and corporate liabilities, *Journal of Political Economy* 81, 3, 637-654.

[12] Brennan, M. J. and E. S. Schwartz (1980), Analysing Convertible Bonds, *Journal of Financial and Quantitative Analysis* 15, 907-929.

[13] Broze, L., O. Scaillet and J. M. Zakoian (1995), Testing for Continuous Time Models of the Short Term Interest Rate, *Journal of Empirical Finance* 2, 199-223.

[14] Bu, R., L. Giet, K. Hadri and M. Lubrano (2011), Modeling Multivariate Interest Rates Using Time-Varying Copulas and Reducible Nonlinear Stochastic Differential Equations, *Journal of Financial Econometrics* 9, 1, 198-236.

[15] Chan, K. C., Karolyi, G. A., Longstaff, F. A. and Sanders A. B. (1992), An Empirical Comparison of Alternative Models of the Short-term Interest Rate, *The Journal of Finance* XLVII, 3, 1209-1227.

[16] Conley, T. G., L. P. Hansen, E. G. J. Luttmer and J. A. Scheinkman (1997), Short Term Interest Rates as Subordinated Diffusions, *Review of Financial Studies* 10, 3, 525-577.

[17] Cox, J. C., Ingersoll J. E. and Ross S. A. (1980), An Analysis of Variable Rate Loan Contracts, *Journal of Finance* 35, 389-403.

[18] Cox, J. C., Ingersoll J. E. and Ross S. A. (1985), A Theory of the Term Structure of Interest Rates, *Econometrica* 53, 385-407.

[19] Czellar, V., G. A. Karolyi and E. Ronchetti (2007), Indirect robust estimation of the short-term interest rate process, *Journal of Empirical Finance* 14, 4, 546-563.

[20] Feller, W. (1951), Two Singular Diffusion Problems, *Annals of Mathematics*, 2nd Series 54, 1, 173-182.

[21] Hansen, L. P. (1982), Large Sample Properties of Generalized Method of Moments Estimators, *Econometrica* 50, 1029–1054.

[22] Hurn A. S., J. I. Jeismand and K. A. Lindsay (2007), Seeing the Wood for the Trees: A Critical Evaluation of Methods to Estimate the Parameters of Stochastic Differential Equations, *Journal of Financial Econometrics* 5, 3, 390-455.

[23] Iglesias, E. M. (2014), Testing of the Mean Reversion Parameter in Continuous Time Models, *Economics Letters* 120, 146-148.

[24] Kendall, M.G. (1954), A note on the bias in the estimation of autocorrelation, *Biometrika* 41, 403-404.

[25] Kladívko, K. (2008), The General Method of Moments (GMM) using MATLAB: The practical guide based on the CKLS interest rate model, *Mezinárodní konference Technical Computing Prague 2008.*

[26] Nowman, K. (1997), Gaussian Estimation of Single-factor Continuous Time Models of the Term Structure of Interest Rates, *Journal of Finance* 52, 1695–1706.

[27] Phillips, P.C.B. and J. Yu (2005), Jackknifing bond option prices, *Review of Financial Studies* 18, 707-742.

[28] Tang, C. Y. and S. X. Chen (2009), Parameter Estimation and Bias Correction for Diffusion Processes, *Journal of Econometrics* 149, 65-81.

[29] Vasicek, O. (1977), An Equilibrium Characterization of the Term Structure, *Journal of Financial Economics* 5, 177-188.

[30] Wang, X., P. C. B. Phillips and J. Yu (2011), Bias in Estimating Multivariate and Univariate Diffusions, *Journal of Econometrics* 161, 2, 228-245.

[31] Yu, J. (2012), Bias in the Estimation of the Mean Reversion Parameter in Continuous Time Models, *Journal of Econometrics* 169, 114-122.

[32] Yu, J. (2014), Econometric Analysis of Continuous Time Models: A Survey of Peter Phillips´s Work and Some New Results, *Econometric Theory* 30, 737-774.

[33] Yu, J. and Phillips, P.C.B. (2001), A Gaussian approach for estimating continuous time models of short term interest rates, *The Econometrics Journal* 4, 211-225.

Supplement to: "Further Results on Pseudo-Maximum Likelihood Estimation and Testing in the Constant Elasticity of Variance Continuous Time Model".

In this file, we collect first Proposition 1 and the corresponding proof (Appendix A), Theorem 1 and Corollary 1 with the corresponding proofs (Appendix B) and Theorem 2 with the proof (Appendix C), where

$$dx_t = \kappa\left(\mu - x_t\right) dt + \sigma x_t^{\rho} dB_t. \tag{1}$$

Moreover

$$x_t = e^{-\delta\kappa} x_{t-1} + \mu\left(1 - e^{-\delta\kappa}\right) + \varepsilon_t, \tag{2}$$

and

$$Var\left(\varepsilon_t | x_{t-1}\right) = 0.5\sigma^2\kappa^{-1}\left(1 - e^{-2\delta\kappa}\right) x_{t-1}^{2\rho}. \tag{3}$$

After that, we proceed to provide four other types of results:

1. *first*, we show more details on the proof of Theorem 1, Corollary 1 and an extra Corollary 2 when $\rho = 0$ and $\delta \to 0$ (see Appendix D).
2. *second*, in order to make the analysis clearer, we show how the PMLEs can be interpreted as IV estimators. This help us to study the order of the bias expansions in a more transparent way (See Appendix E).
3. *third*, we provide the corresponding proofs for the bias expressions using the IV approach (see Appendix F).
4. *fourth,* we show simulation evidence of the order of the bias expressions (see Appendix G).

1 APPENDIX A

1.1 Explicit closed form solutions of the PMLEs

Note that equation (2) with $E(\varepsilon_t|x_{t-1}) = 0$ is a valid representation of any diffusion model with linear drift (see Aït-Sahalia (1996)) given in (1). Thus equation (2) can be used for consistent estimation of the drift parameters. On the other hand, the conditional variance given in equation (3) is not generally correct. But this will only affect the efficiency of the resulting PMLEs and under weak regularity conditions, the PMLEs of the drift parameters will still be consistent and asymptotically normally distributed. We will show this in our Theorem 1. If we denote $\theta = (\kappa, \mu, \sigma^2)'$, the PMLEs are given by maximizing the conditional pseudo log-likelihood $LogL(\theta) = -\frac{1}{2}\left(\sum_{t=1}^{n} \ln Var(\varepsilon_t|x_{t-1}) + \frac{\varepsilon_t^2}{\sqrt{Var(\varepsilon_t|x_{t-1})}}\right)$, where we obtain

Proposition 1. *Assuming ε_t to be Gaussian in (2) and either (a) x_t is a non-negative series and $\rho > 0$ or (b) x_t is unrestricted and ρ is a general non-negative integer, the PMLEs $\widehat{\kappa}$, $\widehat{\mu}$ and $\widehat{\sigma}^2$ are given as $\widehat{\kappa} = -\delta^{-1} \log\left(\widehat{\beta}_1\right)$, $\widehat{\mu} = \widehat{\beta}_2$, $\widehat{\sigma}^2 = \frac{2\widehat{\kappa}\widehat{\beta}_3}{\left(1-\widehat{\beta}_1^2\right)}$, where we condition on the starting value and*

$$\widehat{\beta}_1 = \frac{\sum_{t=1}^{n} x_t x_{t-1}^{(1-2\rho)} \sum_{t=1}^{n} x_{t-1}^{-2\rho} - \sum_{t=1}^{n} x_t x_{t-1}^{-2\rho} \sum_{t=1}^{n} x_{t-1}^{(1-2\rho)}}{\sum_{t=1}^{n} x_{t-1}^{(2-2\rho)} \sum_{t=1}^{n} x_{t-1}^{-2\rho} - \left(\sum_{t=1}^{n}\left(x_{t-1}^{(1-2\rho)}\right)\right)^2}, \tag{4}$$

$$\widehat{\beta}_2 = \frac{\sum_{t=1}^{n}\left(x_t x_{t-1}^{-2\rho} - \widehat{\beta}_1 x_{t-1}^{(1-2\rho)}\right)}{\left(1-\widehat{\beta}_1\right)\sum_{t=1}^{n} x_{t-1}^{-2\rho}}, \quad \widehat{\beta}_3 = n^{-1}\sum_{t=1}^{n}\left(x_t - \widehat{\beta}_1 x_{t-1} - \widehat{\beta}_2\left(1-\widehat{\beta}_1\right)\right)^2 x_{t-1}^{-2\rho}. \tag{5}$$

If in Proposition 1 we set $\rho = 0$, and $\rho = 0.5$, we obtain the special cases given in Tang and Chen (2009, pages 66-67, equations (2.5) and (2.13)). Note that for $\rho = 0.5$ we are in the case of the time series not taking negative values. Proposition 1 also allows one to obtain the PMLEs in other popular models such as for $\rho = 1$ (Brennan and Schwartz (1980)). Note that moments of $\widehat{\kappa}$ may not exist at all since $\widehat{\beta}_1$ may be negative, and this is a characteristic of Nowman´s estimator. Also, from Proposition 1 we have the following two Remarks

Remark 1 *In Proposition 1, we need ρ to be known although obviously it would be more general if we could find an estimator for ρ. We have tried that, but the closed-form expressions we obtained for Nowman´s method become non-tractable.*

Remark 2 *It is important to note that from Proposition 1 and (3), if we want to estimate the CKLS model with our PMLEs and to use them with a series that may take positive and negative values, we need ρ to be a non-negative integer, since we need $x_t^{-2\rho}$ to exist. For example, when $\rho = 1/4$, we have $x_t^{-2\rho} = x_t^{-1/2}$ which does not exist if x_t is negative.*

1.2 PROOF of Proposition 1

Under the conditions of Proposition 1, the conditional loglikelihood function from (2)-(3) is

$$LogL(\theta) = -\frac{1}{2}\sum_{t=1}^{n}\left(\ln\left(0.5\sigma^2\kappa^{-1}\left(1-e^{-2\delta\kappa}\right)x_{t-1}^{2\rho}\right) + \frac{\left(x_t - e^{-\delta\kappa}x_{t-1} - \mu\left(1-e^{-\delta\kappa}\right)\right)^2}{0.5\sigma^2\kappa^{-1}\left(1-e^{-2\delta\kappa}\right)x_{t-1}^{2\rho}}\right).$$

Setting

$$\frac{\partial LogL}{\partial\widehat{\kappa}} = 0;\quad \frac{\partial LogL}{\partial\widehat{\mu}} = 0 \text{ and } \frac{\partial LogL}{\partial\widehat{\sigma}^2} = 0$$

we obtain for $\widehat{\kappa}$

$$\begin{aligned} & e^{-\delta\widehat{\kappa}}\left(\sum_{t=1}^{n}x_{t-1}^{(2-2\rho)} - \left(\sum_{t=1}^{n}\left(x_{t-1}^{(1-2\rho)}\right)\right)^2\left(\sum_{t=1}^{n}x_{t-1}^{-2\rho}\right)^{-1}\right) \\ = & \sum_{t=1}^{n}x_t x_{t-1}^{(1-2\rho)} - \sum_{t=1}^{n}\left(x_t x_{t-1}^{-2\rho}\right)\left(\sum_{t=1}^{n}x_{t-1}^{-2\rho}\right)^{-1}\sum_{t=1}^{n}x_{t-1}^{(1-2\rho)} \end{aligned}$$

and taking logarithms

$$\begin{aligned} \widehat{\kappa} &= -\delta^{-1}\log\left(\frac{\sum_{t=1}^{n}x_t x_{t-1}^{(1-2\rho)} - \sum_{t=1}^{n}\left(x_t x_{t-1}^{-2\rho}\right)\left(\sum_{t=1}^{n}x_{t-1}^{-2\rho}\right)^{-1}\sum_{t=1}^{n}x_{t-1}^{(1-2\rho)}}{\sum_{t=1}^{n}x_{t-1}^{(2-2\rho)} - \left(\sum_{t=1}^{n}\left(x_{t-1}^{(1-2\rho)}\right)\right)^2\left(\sum_{t=1}^{n}x_{t-1}^{-2\rho}\right)^{-1}}\right) \\ &= -\delta^{-1}\log\left(\frac{\sum_{t=1}^{n}x_t x_{t-1}^{(1-2\rho)}\sum_{t=1}^{n}x_{t-1}^{-2\rho} - \sum_{t=1}^{n}x_t x_{t-1}^{-2\rho}\sum_{t=1}^{n}x_{t-1}^{(1-2\rho)}}{\sum_{t=1}^{n}x_{t-1}^{(2-2\rho)}\sum_{t=1}^{n}x_{t-1}^{-2\rho} - \left(\sum_{t=1}^{n}\left(x_{t-1}^{(1-2\rho)}\right)\right)^2}\right) = -\delta^{-1}\log\left(\widehat{\beta}_1\right). \end{aligned}$$

Also for $\widehat{\mu}$ we obtain

$$\widehat{\mu} = \frac{\sum_{t=1}^{n}\left(x_t - e^{-\delta\widehat{\kappa}}x_{t-1}\right)x_{t-1}^{-2\rho}}{\sum_{t=1}^{n}\left(1-e^{-\delta\widehat{\kappa}}\right)x_{t-1}^{-2\rho}} = \frac{\sum_{t=1}^{n}\left(x_t x_{t-1}^{-2\rho} - \widehat{\beta}_1 x_{t-1}^{(1-2\rho)}\right)}{\left(1-\widehat{\beta}_1\right)\sum_{t=1}^{n}x_{t-1}^{-2\rho}} = \widehat{\beta}_2.$$

Finally

$$\begin{aligned} \widehat{\sigma}^2 &= n^{-1}\sum_{t=1}^{n}\frac{2\widehat{\kappa}\left(x_t - e^{-\delta\widehat{\kappa}}x_{t-1} - \widehat{\mu}\left(1-e^{-\delta\widehat{\kappa}}\right)\right)^2}{\left(1-e^{-2\delta\widehat{\kappa}}\right)x_{t-1}^{2\rho}} \\ &= \frac{2\widehat{\kappa}}{\left(1-\widehat{\beta}_1^2\right)}n^{-1}\sum_{t=1}^{n}\left(x_t - \widehat{\beta}_1 x_{t-1} - \widehat{\beta}_2\left(1-\widehat{\beta}_1\right)\right)^2 x_{t-1}^{-2\rho}. \end{aligned}$$

■

2 APPENDIX B

2.1 Asymptotic theory

Continuous time models have been proved to be very successful in economic theory (see e.g. Merton (1990)). We generalize now the asymptotic theory of Nowman´s estimator given in Tang and Chen (2009) for $\rho = 0$ and 0.5. We show the following Theorem[1]

THEOREM 1. *For a stationary CKLS process[2], as $n \longrightarrow \infty$ while δ (the sampling interval) is fixed, and assuming either (a) x_t is a series taking only non-negative values and $\rho > 0$ or (b) x_t is unrestricted and ρ is a general non-negative integer, let $\widehat{\theta} = (\hat{\kappa}, \widehat{\mu}, \widehat{\sigma}^2)'$, and $\widetilde{\theta} = (\kappa, \mu, \sigma^2 - B(\theta, \delta))'$, where $B(\theta, \delta)$ is the inconsistency term related to σ^2 and let $E(\varepsilon_t | x_{t-1}) = 0$. Then $\sqrt{n}\left(\widehat{\theta} - \widetilde{\theta}\right) \xrightarrow{d} N(0, \Omega)$ where $\Omega = \Lambda^{-1}$ with $\Lambda = E\left(-\frac{1}{n}\frac{\partial^2 LogL(\theta)}{\partial\theta\partial\theta'}\right) = \begin{pmatrix} \overline{A} & \overline{B} & \overline{C} \\ \overline{B} & \overline{D} & 0 \\ \overline{C} & 0 & \overline{E} \end{pmatrix}$, and $\overline{A} = \frac{\left(e^{-2\kappa\delta} + 2\kappa\delta e^{-2\kappa\delta} - 1\right)^2}{2\kappa^2\left(e^{-2\kappa\delta} - 1\right)^2} + \frac{2\delta^2 e^{-2\delta\kappa} E\left(\left(x_{t-1}^{(1-\rho)} - \mu x_{t-1}^{-\rho}\right)^2\right)}{\sigma^2\kappa^{-1}\left(1 - e^{-2\delta\kappa}\right)}$;*

$\overline{B} = -\frac{2\kappa\delta e^{-\delta\kappa} E\left(x_{t-1}^{(1-2\rho)} - \mu x_{t-1}^{-2\rho}\right)}{\sigma^2\left(1 + e^{-\delta\kappa}\right)}$;

$$\overline{C} = \frac{\delta e^{-2\delta\kappa}}{\sigma^2\left(1 - e^{-2\delta\kappa}\right)} - \frac{1}{2\kappa\sigma^2};\ \overline{D} = \frac{2\kappa\left(e^{\delta\kappa} - 1\right)}{\sigma^2\left(e^{\delta\kappa} + 1\right)} E\left(\frac{1}{x_{t-1}^{2\rho}}\right);\ \overline{E} = \frac{1}{2\sigma^4}.$$ *Also for $\rho = 0$, $B(\theta, \delta) = 0$.*

Note that we leave Theorem 1 in terms of expectations since we provide the theory for a general ρ and therefore the expectations of x_t will be different depending on the ρ value we select. In the following Corollary 1, we apply Theorem 1 for the specific case of $\rho = 0$.

COROLLARY 1. *For a stationary Vasicek (1977) process, as $n \longrightarrow \infty$ while δ is fixed and for $\rho = 0$, let $\widehat{\theta} = (\hat{\kappa}, \widehat{\mu}, \widehat{\sigma}^2)'$, and $\theta = (\kappa, \mu, \sigma^2)'$, then $\sqrt{n}\left(\widehat{\theta} - \theta\right) \xrightarrow{d} N(0, \Omega_1)$ where*

$$\Omega_1 = \begin{pmatrix} \delta^{-2}\left(e^{2\kappa\delta} - 1\right) & 0 & -\frac{\sigma^2\left(1 - e^{2\kappa\delta} + 2\kappa\delta\right)}{\kappa\delta^2} \\ 0 & \frac{\sigma^2\left(1 + e^{\kappa\delta}\right)}{2\kappa\left(e^{\delta\kappa} - 1\right)} & 0 \\ -\frac{\sigma^2\left(1 - e^{2\kappa\delta} + 2\kappa\delta\right)}{\kappa\delta^2} & 0 & \frac{\sigma^4\left(2\kappa^2\delta^2\left(1 + e^{2\kappa\delta}\right) + 4\kappa\delta\left(1 - e^{2\kappa\delta}\right) + e^{4\kappa\delta} - 2e^{2\kappa\delta} + 1\right)}{\delta^2\kappa^2\left(e^{2\kappa\delta} - 1\right)} \end{pmatrix}.$$

[1] As noted in our Remark 2, we may replace the restriction that ρ has to be a non-negative integer with assumptions on the parameters to ensure that x_t has positive support. This can be done both in Theorems 1 and 2.

[2] See Conley, Hansen, Luttmer and Scheinkman (1997) for details of primitive conditions under which the CEV process is stationary and ergodic. Broze et al (1995) also provided conditions for second-order stationarity and ergodicity.

Remark 3 *Tang and Chen (2009, Theorem 3.1.2) analyzed also the case of* $\rho = 0$ *and they obtained a diagonal variance-covariance matrix with the same main diagonal components as in our* Ω_1 *matrix in Corollary 1, except that the third component is replaced by* $\sigma^4(\kappa\delta)^{-2}\left(e^{\kappa\delta} - e^{-\kappa\delta}\right)(1 - \frac{2\kappa\delta e^{-2\kappa\delta}}{\left(1-e^{-2\kappa\delta}\right)})$*. When we specialize our Corollary 1 for* $\delta \to 0$ *(high frequency case), we obtain the same result as in Theorem 3.2.2 of Tang and Chen (2009).*

2.2 PROOF of Theorem 1

Let $\beta_1 = e^{-\kappa\delta}$, $\beta_2 = \mu$ and $\beta_3 = \sigma^2(2\kappa)^{-1}\left(1 - e^{-2\delta\kappa}\right)$ and $\beta = (\beta_1, \beta_2, \beta_3)'$ be the 1-1 mapping from $\theta = (\kappa, \mu, \sigma^2)'$. Then for $\rho = 0$, $\widehat{\beta}_3$ is consistent but for any positive ρ, $\widehat{\beta}_3$ is inconsistent, and it can be shown that $E\left(\widehat{\beta}_3\right) = \beta_3 + B(\theta, \delta) + O\left(n^{-1}\right)$ where $B(\theta, \delta)$ is the inconsistency term related to $\widehat{\beta}_3$ (see for example Aït-Sahalia (1996, equation (2.4)) which shows that equation (2) with $E(\varepsilon_t | x_{t-1}) = 0$ is a valid representation of any diffusion model with linear drift. Thus equation (2) can be used for consistent estimation of the drift parameters but not for the diffusion parameter). For $\rho = 0$, $B(\theta, \delta) = 0$ as shown in Tang and Chen (2009). Let $\widetilde{\beta} = (\beta_1, \beta_2, \beta_3 + B(\theta, \delta))'$. Then,

$$LogL(\theta) = -\frac{1}{2}\sum_{t=1}^{n}\left(\ln\left(0,5\sigma^2\kappa^{-1}\left(1 - e^{-2\delta\kappa}\right)x_{t-1}^{2\rho}\right) + \frac{\left(x_t - e^{-\delta\kappa}x_{t-1} - \mu\left(1 - e^{-\delta\kappa}\right)\right)^2}{0,5\sigma^2\kappa^{-1}\left(1 - e^{-2\delta\kappa}\right)x_{t-1}^{2\rho}}\right),$$

$$\frac{\partial LogL(\theta)}{\partial\kappa} = \sum_{t=1}^{n}\left((2\kappa)^{-1} - \frac{\delta e^{-2\delta\kappa}}{(1 - e^{-2\delta\kappa})} - \frac{\left(x_t - e^{-\delta\kappa}x_{t-1} - \mu\left(1 - e^{-\delta\kappa}\right)\right)\delta e^{-\delta\kappa}(x_{t-1} - \mu)}{0.5\sigma^2 x_{t-1}^{2\rho}\kappa^{-1}\left(1 - e^{-2\delta\kappa}\right)}\right)$$

$$+\sum_{t=1}^{n}\left(\frac{2\delta e^{-2\delta\kappa}\left(x_t - e^{-\delta\kappa}x_{t-1} - \mu\left(1 - e^{-\delta\kappa}\right)\right)^2}{\sigma^2 x_{t-1}^{2\rho}\kappa^{-1}\left(1 - e^{-2\delta\kappa}\right)^2} - \frac{\left(x_t - e^{-\delta\kappa}x_{t-1} - \mu\left(1 - e^{-\delta\kappa}\right)\right)^2}{\sigma^2 x_{t-1}^{2\rho}\left(1 - e^{-2\delta\kappa}\right)}\right),$$

$$\frac{\partial LogL(\theta)}{\partial\mu} = \sum_{t=1}^{n}\frac{\left(\left(x_t - e^{-\delta\kappa}x_{t-1}\right) - \mu\left(1 - e^{-\delta\kappa}\right)\right)\left(1 - e^{-\delta\kappa}\right)}{0,5\sigma^2\kappa^{-1}\left(1 - e^{-2\delta\kappa}\right)x_{t-1}^{2\rho}},$$

$$\frac{\partial LogL(\theta)}{\partial\sigma^2} = -\frac{1}{2\sigma^2}\sum_{t=1}^{n}\left(1 - \frac{\left(x_t - e^{-\delta\kappa}x_{t-1} - \mu\left(1 - e^{-\delta\kappa}\right)\right)^2}{\sigma^2 0,5\kappa^{-1}\left(1 - e^{-2\delta\kappa}\right)x_{t-1}^{2\rho}}\right).$$

$LogL(\theta)$ can be regarded as $LogL(\beta)$ after re-parametrization. Following the proof of Tang and Chen (2009, Theorem 3.1.4), we apply first a Taylor series expansion to the pseudo-likelihood score equations for $\widehat{\beta} = (\widehat{\beta}_1, \widehat{\beta}_2, \beta_3)'$

$$0=\frac{\partial LogL\left(\widehat{\beta}\right)}{\partial\beta}\approx\frac{\partial LogL\left(\widetilde{\beta}\right)}{\partial\beta}+\left(\widehat{\beta}-\widetilde{\beta}\right)\frac{\partial^2 LogL\left(\widetilde{\beta}\right)}{\partial\beta\partial\beta'}$$

Later, we apply a central limit theorem for mixing sequences (Bosq (1998)), and by Slutsky´s Theorem and transforming back to $\widehat{\theta}$ as a function of the asymptotically normal vector, we obtain that

$$\begin{aligned}\sqrt{n}\left(\widehat{\theta}-\widetilde{\theta}\right) &= -\sqrt{n}\frac{\partial LogL\left(\widetilde{\theta}\right)}{\partial\theta}\left(\frac{\partial^2 LogL\left(\widetilde{\theta}\right)}{\partial\theta\partial\theta'}\right)^{-1}\\ &= \left(\frac{1}{\sqrt{n}}\frac{\partial LogL\left(\widetilde{\theta}\right)}{\partial\theta}\right)\left(-\frac{1}{n}\frac{\partial^2 LogL\left(\widetilde{\theta}\right)}{\partial\theta\partial\theta'}\right)^{-1}\quad\longrightarrow_d N\left(0,\Omega\right)\end{aligned}$$

where

$$\Omega=E\left(-\frac{1}{n}\frac{\partial^2 LogL\left(\theta\right)}{\partial\theta\partial\theta'}\right)^{-1}$$

with

$$\begin{aligned}\frac{\partial^2 LogL\left(\theta\right)}{\partial\kappa^2}&=\sum_{t=1}^{n}\left(-\frac{1}{2\kappa^2}+\frac{\left(1-e^{-2\delta\kappa}\right)2\delta^2e^{-2\delta\kappa}+2\delta^2e^{-4\delta\kappa}}{\left(1-e^{-2\delta\kappa}\right)^2}\right)\\
&-\sum_{t=1}^{n}\frac{\left(1-e^{-2\delta\kappa}\right)\left[e^{-\delta\kappa}\left(\delta e^{-\delta\kappa}x_{t-1}-\mu\delta e^{-\delta\kappa}\right)+\left(x_t-e^{-\delta\kappa}x_{t-1}-\mu\left(1-e^{-\delta\kappa}\right)\right)\delta e^{-\delta\kappa}\right]}{0.5\left(x_{t-1}-\mu\right)^{-1}\delta^{-1}\sigma^2x_{t-1}^{2\rho}\kappa^{-1}\left(1-e^{-2\delta\kappa}\right)^2}\\
&+\sum_{t=1}^{n}\frac{e^{-\delta\kappa}\left(x_t-e^{-\delta\kappa}x_{t-1}-\mu\left(1-e^{-\delta\kappa}\right)\right)\left[-\kappa^{-1}\left(1-e^{-2\delta\kappa}\right)+2\delta e^{-2\delta\kappa}\right]}{0.5\left(x_{t-1}-\mu\right)^{-1}\delta^{-1}\sigma^2x_{t-1}^{2\rho}\kappa^{-1}\left(1-e^{-2\delta\kappa}\right)^2}\\
&+\sum_{t=1}^{n}\frac{2e^{-2\delta\kappa}\left(x_t-e^{-\delta\kappa}x_{t-1}-\mu\left(1-e^{-\delta\kappa}\right)\right)\left(\delta e^{-\delta\kappa}x_{t-1}-\mu\delta e^{-\delta\kappa}\right)-2\delta e^{-2\delta\kappa}\left(x_t-e^{-\delta\kappa}x_{t-1}-\mu\left(1-e^{-\delta\kappa}\right)\right)^2}{\left(2\delta\right)^{-1}\sigma^2x_{t-1}^{2\rho}\kappa^{-1}\left(1-e^{-2\delta\kappa}\right)^2}\\
&-\sum_{t=1}^{n}\frac{e^{-2\delta\kappa}\left(x_t-e^{-\delta\kappa}x_{t-1}-\mu\left(1-e^{-\delta\kappa}\right)\right)^2\left[-\kappa^{-1}\left(1-e^{-2\delta\kappa}\right)+4\delta e^{-2\delta\kappa}\right]}{\left(2\delta\right)^{-1}\sigma^2x_{t-1}^{2\rho}\kappa^{-1}\left(1-e^{-2\delta\kappa}\right)^3}\\
&-\sum_{t=1}^{n}\frac{2\left(x_t-e^{-\delta\kappa}x_{t-1}-\mu\left(1-e^{-\delta\kappa}\right)\right)\left(\delta e^{-\delta\kappa}x_{t-1}-\mu\delta e^{-\delta\kappa}\right)}{\sigma^2x_{t-1}^{2\rho}\left(1-e^{-2\delta\kappa}\right)}+\sum_{t=1}^{n}\frac{2\delta e^{-2\delta\kappa}\left(x_t-e^{-\delta\kappa}x_{t-1}-\mu\left(1-e^{-\delta\kappa}\right)\right)^2}{\sigma^2x_{t-1}^{2\rho}\left(1-e^{-2\delta\kappa}\right)^2},\end{aligned}$$

$$\frac{\partial^2 LogL(\theta)}{\partial\mu^2} = \sum_{t=1}^{n} \frac{-\left(1-e^{-\delta\kappa}\right)^2}{0,5\sigma^2\kappa^{-1}\left(1-e^{-2\delta\kappa}\right)x_{t-1}^{2\rho}},$$

$$\frac{\partial^2 LogL(\theta)}{\partial\sigma^4} = \frac{n}{2\sigma^4} - \sum_{t=1}^{n} \frac{2\left(x_t - e^{-\delta\kappa}x_{t-1} - \mu\left(1-e^{-\delta\kappa}\right)\right)^2}{\sigma^6\kappa^{-1}\left(1-e^{-2\delta\kappa}\right)x_{t-1}^{2\rho}},$$

$$\frac{\partial^2 LogL(\theta)}{\partial\mu\partial\sigma^2} = \sum_{t=1}^{n} \frac{-\left(x_t - e^{-\delta\kappa}x_{t-1} - \mu\left(1-e^{-\delta\kappa}\right)\right)\left(1-e^{-\delta\kappa}\right)}{0,5\sigma^4\kappa^{-1}\left(1-e^{-2\delta\kappa}\right)x_{t-1}^{2\rho}},$$

$$\begin{aligned}\frac{\partial^2 LogL(\theta)}{\partial\kappa\partial\sigma^2} &= \frac{1}{\sigma^2}\sum_{t=1}^{n}\left(\frac{\left(x_t - e^{-\delta\kappa}x_{t-1} - \mu\left(1-e^{-\delta\kappa}\right)\right)\left(\delta e^{-\delta\kappa}x_{t-1} - \mu\delta e^{-\delta\kappa}\right)}{\sigma^2 0,5 x_{t-1}^{2\rho}\kappa^{-1}\left(1-e^{-2\delta\kappa}\right)}\right) \\ &\quad - \frac{1}{\sigma^2}\sum_{t=1}^{n}\left(\frac{\left(x_t - e^{-\delta\kappa}x_{t-1} - \mu\left(1-e^{-\delta\kappa}\right)\right)^2\left[\kappa^{-1}2\delta e^{-2\delta\kappa} - \kappa^{-2}\left(1-e^{-2\delta\kappa}\right)\right]}{\sigma^2 x_{t-1}^{2\rho}\kappa^{-2}\left(1-e^{-2\delta\kappa}\right)^2}\right).\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 LogL(\theta)}{\partial\mu\partial\kappa} &= \sum_{t=1}^{n} \frac{(x_t - e^{-\delta\kappa}x_{t-1} - \mu\left(1-e^{-\delta\kappa}\right))\delta e^{-\delta\kappa} + \left(1-e^{-\delta\kappa}\right)\left(\delta e^{-\delta\kappa}x_{t-1} - \delta\mu e^{-\delta\kappa}\right)}{0,5\sigma^2 x_{t-1}^{2\rho}\kappa^{-1}\left(1-e^{-2\delta\kappa}\right)} \\ &\quad + \sum_{t=1}^{n} \frac{\left(x_t - e^{-\delta\kappa}x_{t-1} - \mu\left(1-e^{-\delta\kappa}\right)\right)\left(1-e^{-\delta\kappa}\right)\left[\kappa^{-1}2\delta e^{-2\delta\kappa} - \kappa^{-2}\left(1-e^{-2\delta\kappa}\right)\right]}{0,5\sigma^2 x_{t-1}^{2\rho}\kappa^{-2}\left(1-e^{-2\delta\kappa}\right)^2}.\end{aligned}$$

We need to show that $\Lambda = E\left(-\frac{1}{n}\frac{\partial^2 LogL(\theta)}{\partial\theta\partial\theta'}\right)$ is positive definite. We commence by showing each of the components of Λ as a function of $\rho \geq 0$ and we obtain

$$\Lambda = E\left(-\frac{1}{n}\frac{\partial^2 LogL(\theta)}{\partial\theta\partial\theta'}\right) = \begin{pmatrix} \overline{A} & \overline{B} & \overline{C} \\ \overline{B} & \overline{D} & 0 \\ \overline{C} & 0 & \overline{E} \end{pmatrix},$$

with

$$\begin{aligned}\overline{A} &= \frac{\left(e^{-2\kappa\delta} + 2\kappa\delta e^{-2\kappa\delta} - 1\right)^2}{2\kappa^2\left(e^{-2\kappa\delta} - 1\right)^2} + \frac{2\delta^2 e^{-2\delta\kappa}E\left(\left(x_{t-1}^{(1-\rho)} - \mu x_{t-1}^{-\rho}\right)^2\right)}{\sigma^2\kappa^{-1}\left(1-e^{-2\delta\kappa}\right)};\ \overline{B} = -\frac{2\kappa\delta e^{-\delta\kappa}E\left(x_{t-1}^{(1-2\rho)} - \mu x_{t-1}^{-2\rho}\right)}{\sigma^2\left(1+e^{-\delta\kappa}\right)}, \\ \overline{C} &= \frac{\delta e^{-2\delta\kappa}}{\sigma^2\left(1-e^{-2\delta\kappa}\right)} - \frac{1}{2\kappa\sigma^2};\ \overline{D} = \frac{2\kappa\left(e^{\delta\kappa} - 1\right)}{\sigma^2\left(e^{\delta\kappa}+1\right)}E\left(\frac{1}{x_{t-1}^{2\rho}}\right);\ \overline{E} = \frac{1}{2\sigma^4}.\end{aligned}$$

Hence, Λ will be positive definite if for any non-zero column vector z with entries a, b and c, $z'\Lambda z > 0$.

In our case we may write

$$
\begin{aligned}
z\acute{}\Lambda z &= a^2\overline{A} + 2ab\overline{C} + 2ac\overline{C} + b^2\overline{D} + c^2\overline{E} \\
&= \frac{2}{\sigma^2\kappa^{-1}} E[\frac{\delta^2 e^{-2\delta\kappa} a^2 \left(x_{t-1}^{(1-\rho)} - \mu x_{t-1}^{-\rho}\right)^2}{\left(1-e^{-\delta\kappa}\right)\left(1+e^{-\delta\kappa}\right)} - \frac{2\delta e^{-\delta\kappa}\left(x_{t-1}^{(1-2\rho)} - \mu x_{t-1}^{-2\rho}\right) ab}{\left(1+e^{-\delta\kappa}\right)} + \frac{\left(e^{\delta\kappa}-1\right) x_{t-1}^{-2\rho} b^2}{\left(e^{\delta\kappa}+1\right)}] \\
&\quad + \frac{\left(e^{-2\kappa\delta} + 2\kappa\delta e^{-2\kappa\delta} - 1\right)^2 a^2}{2\kappa^2\left(e^{-2\kappa\delta}-1\right)^2} - \frac{\left(e^{-2\delta\kappa} + 2\kappa\delta e^{-2\delta\kappa} - 1\right) ac}{\sigma^2\left(e^{-2\kappa\delta}-1\right)\kappa} + \frac{c^2}{2\sigma^4} \\
&= \frac{2\left(e^{\delta\kappa}+1\right)}{\sigma^2\kappa^{-1}\left(e^{\delta\kappa}-1\right)} E\left[\left(\frac{\delta a\left(x_{t-1}^{(1-\rho)} - \mu x_{t-1}^{-\rho}\right)}{\left(e^{\delta\kappa}+1\right)} - \frac{\left(e^{\delta\kappa}-1\right) x_{t-1}^{-\rho} b}{\left(e^{\delta\kappa}+1\right)}\right)^2\right] \\
&\quad + \frac{1}{2}\left(\frac{\left(e^{-2\kappa\delta} + 2\kappa\delta e^{-2\kappa\delta} - 1\right) a}{\kappa\left(e^{-2\kappa\delta}-1\right)} - \frac{c}{\sigma^2}\right)^2 > 0,
\end{aligned}
$$

since

$$
\begin{aligned}
& E[\frac{\delta^2 a^2\left(x_{t-1}^{(1-\rho)} - \mu x_{t-1}^{-\rho}\right)^2}{\left(e^{\delta\kappa}-1\right)\left(e^{\delta\kappa}+1\right)} - \frac{2\delta\left(x_{t-1}^{(1-2\rho)} - \mu x_{t-1}^{-2\rho}\right) ab}{\left(e^{\delta\kappa}+1\right)} + \frac{\left(e^{\delta\kappa}-1\right) x_{t-1}^{-2\rho} b^2}{\left(e^{\delta\kappa}+1\right)}] \\
= \; & \frac{\left(e^{\delta\kappa}+1\right)}{\left(e^{\delta\kappa}-1\right)} E\left[\frac{\delta^2 a^2\left(x_{t-1}^{(1-\rho)} - \mu x_{t-1}^{-\rho}\right)^2}{\left(e^{\delta\kappa}+1\right)^2} - \frac{2\delta\left(x_{t-1}^{(1-2\rho)} - \mu x_{t-1}^{-2\rho}\right)\left(e^{\delta\kappa}-1\right) ab}{\left(e^{\delta\kappa}+1\right)^2} + \frac{\left(e^{\delta\kappa}-1\right)^2 x_{t-1}^{-2\rho} b^2}{\left(e^{\delta\kappa}+1\right)^2}\right] \\
= \; & \frac{\left(e^{\delta\kappa}+1\right)}{\left(e^{\delta\kappa}-1\right)} E\left[\left(\frac{\delta a\left(x_{t-1}^{(1-\rho)} - \mu x_{t-1}^{-\rho}\right)}{\left(e^{\delta\kappa}+1\right)} - \frac{\left(e^{\delta\kappa}-1\right) x_{t-1}^{-\rho} b}{\left(e^{\delta\kappa}+1\right)}\right)^2\right].
\end{aligned}
$$

Therefore, we are able to show that $z\acute{}\Lambda z > 0$ (as a sum of squares -therefore non-negative-, and that is zero only if $a = b = c = 0$, that is when z is the zero vector). Then we conclude that Λ is positive definite. Finally, $\Lambda^{-1} = \Omega$. ■

2.3 PROOF of Corollary 1 when $\rho = 0$

Now, for $\rho = 0$

$$
\begin{aligned}
E\left(-\frac{1}{n}\frac{\partial^2 LogL(\theta)}{\partial\kappa^2}\right) &= \frac{\left(e^{-2\kappa\delta} + 2\kappa\delta e^{-2\kappa\delta} - 1\right)^2}{2\kappa^2\left(e^{-2\kappa\delta}-1\right)^2} + \frac{2\delta^2 e^{-2\delta\kappa}}{\sigma^2\kappa^{-1}\left(1-e^{-2\delta\kappa}\right)} 0,5\sigma^2\kappa^{-1} \\
&= \frac{\left(e^{-2\kappa\delta} + 2\kappa\delta e^{-2\kappa\delta} - 1\right)^2 + 2\kappa^2\delta^2 e^{-2\delta\kappa}\left(1-e^{-2\delta\kappa}\right)}{2\kappa^2\left(1-e^{-2\kappa\delta}\right)^2}
\end{aligned}
$$

since

$$\begin{aligned}
x_t - e^{-\delta\kappa} x_{t-1} - \mu\left(1 - e^{-\delta\kappa}\right) &= \varepsilon_t, \\
x_t - \mu &= e^{-\delta\kappa}\left(x_{t-1} - \mu\right) + \varepsilon_t, \\
E\left(x_t - \mu\right) &= e^{-\delta\kappa} E\left(x_{t-1} - \mu\right) + E\left(\varepsilon_t\right), \\
\left(x_t - \mu\right)^2 &= \left(e^{-\delta\kappa}\left(x_{t-1} - \mu\right) + \varepsilon_t\right)^2, \\
\left(1 - e^{-2\delta\kappa}\right) E\left(x_t - \mu\right)^2 &= 0,5\sigma^2\kappa^{-1}\left(1 - e^{-2\delta\kappa}\right) \Longrightarrow E\left(x_t - \mu\right)^2 = 0,5\sigma^2\kappa^{-1},
\end{aligned}$$

from (2). Also

$$E\left(-\frac{1}{n}\frac{\partial^2 LogL\left(\theta\right)}{\partial\mu^2}\right) = \sigma^{-2} 2\kappa\left(e^{\delta\kappa} - 1\right)\left(e^{\delta\kappa} + 1\right)^{-1},$$

Moreover

$$E\left(-\frac{1}{n}\frac{\partial^2 LogL\left(\theta\right)}{\partial\sigma^4}\right) = \frac{1}{2\sigma^4}.$$

and

$$E\left(-\frac{1}{n}\frac{\partial^2 LogL}{\partial\mu\partial\kappa}\right) = -\frac{2\kappa\delta e^{-\delta\kappa}}{\sigma^2\left(1 + e^{-\delta\kappa}\right)} E\left(x_{t-1} - \mu\right) = 0$$

since $E\left(x_t - \mu\right) = e^{-\delta\kappa} E\left(x_{t-1} - \mu\right) + E\left(\varepsilon_t\right)$ from (2); $E\left(x_t\right) = e^{-\delta\kappa} x_0 + \mu\left(1 - e^{-\delta\kappa}\right)$ and $E\left(x_t - \mu\right) = e^{-\delta\kappa} x_0 - \mu e^{-\delta\kappa} = 0$ where we assume the initial condition $x_0 = \mu$. Also

$$E\left(-\frac{1}{n}\frac{\partial^2 LogL}{\partial\kappa\partial\sigma^2}\right) = \frac{\delta e^{-2\delta\kappa}}{\sigma^2\left(1 - e^{-2\delta\kappa}\right)} - \frac{1}{2\kappa\sigma^2}.$$

Finally

$$E\left(-\frac{1}{n}\frac{\partial^2 LogL}{\partial\mu\partial\sigma^2}\right) = E\left(\frac{\left(x_t - e^{-\delta\kappa} x_{t-1} - \mu\left(1 - e^{-\delta\kappa}\right)\right)\left(1 - e^{-\delta\kappa}\right)}{0,5\sigma^4\kappa^{-1}\left(1 - e^{-2\delta\kappa}\right)}\right) = 0$$

Finding now the inverse of

$$\begin{pmatrix}
\frac{\left(e^{-2\kappa\delta} + 2\kappa\delta e^{-2\kappa\delta} - 1\right)^2 + 2\kappa^2\delta^2 e^{-2\delta\kappa}\left(1 - e^{-2\delta\kappa}\right)}{2\kappa^2\left(1 - e^{-2\kappa\delta}\right)^2} & 0 & \frac{\delta e^{-2\delta\kappa}}{\sigma^2\left(1 - e^{-2\delta\kappa}\right)} - \frac{1}{2\kappa\sigma^2} \\
0 & \frac{2\kappa\left(e^{\delta\kappa} - 1\right)}{\sigma^2\left(e^{\delta\kappa} + 1\right)} & 0 \\
\frac{\delta e^{-2\delta\kappa}}{\sigma^2\left(1 - e^{-2\delta\kappa}\right)} - \frac{1}{2\kappa\sigma^2} & 0 & \frac{1}{2\sigma^4}
\end{pmatrix}$$

we obtain Ω_1

$$\Omega_1 = \begin{pmatrix}
\delta^{-2}\left(e^{2\kappa\delta} - 1\right) & 0 & -\frac{\sigma^2\left(1 - e^{2\kappa\delta} + 2\kappa\delta\right)}{\kappa\delta^2} \\
0 & \frac{\sigma^2\left(1 + e^{\kappa\delta}\right)}{2\kappa\left(e^{\delta\kappa} - 1\right)} & 0 \\
-\frac{\sigma^2\left(1 - e^{2\kappa\delta} + 2\kappa\delta\right)}{\kappa\delta^2} & 0 & \frac{\sigma^4\left(2\kappa^2\delta^2\left(1 + e^{2\kappa\delta}\right) + 4\kappa\delta\left(1 - e^{2\kappa\delta}\right) + e^{4\kappa\delta} - 2e^{2\kappa\delta} + 1\right)}{\delta^2\kappa^2\left(e^{2\kappa\delta} - 1\right)}
\end{pmatrix}.$$

■

3 APPENDIX C

3.1 Theorem 2

Following Tang and Chen (2009, proof of Theorem 3.1.3) we first note that from (2), the following holds

$$E\left(x_j|x_i\right) = e^{-\delta_{ij}\kappa}x_i + \mu\left(1 - e^{-\delta_{ij}\kappa}\right) \text{ with } \delta_{ij} = \delta\left|j-i\right|. \tag{6}$$

Let $t_{1i} = x_i x_{i-1}^{(1-2\rho)} - \mu_1$, $t_{2i} = x_{i-1}^{-2\rho} - \mu_2$, $t_{3i} = x_i x_{i-1}^{-2\rho} - \mu_3$, $t_{4i} = x_{i-1}^{(1-2\rho)} - \mu_4$ and $t_{5i} = x_{i-1}^{(2-2\rho)} - \mu_5$, where from (13)

$$\begin{aligned}\mu_1 &= E\left(x_i x_{i-1}^{(1-2\rho)}\right) = E\left(x_{t-1}^{(1-2\rho)}E\left(x_t|x_{t-1}\right)\right) = e^{-\delta\kappa}\mu_5 + \mu\left(1-e^{-\delta\kappa}\right)\mu_4,\\ \mu_3 &= E\left(x_t x_{t-1}^{-2\rho}\right) = E\left(x_{t-1}^{-2\rho}E\left(x_t|x_{t-1}\right)\right) = e^{-\delta\kappa}\mu_4 + \mu\left(1-e^{-\delta\kappa}\right)\mu_2,\end{aligned}$$

$\mu_2 = E\left(x_{t-1}^{-2\rho}\right)$, $\mu_4 = E\left(x_{t-1}^{(1-2\rho)}\right)$ and $\mu_5 = E\left(x_{t-1}^{(2-2\rho)}\right)$. Also define $t_a = n^{-1}\sum_{i=1}^{n} t_{ai}$ which is $O_p(n^{\frac{-1}{2}})$ and let $\widetilde{t}_a = n^{-1}\sum_{i=1}^{n} t_{a(i-1)} = t_a + n^{-1}\left(t_{a0} - t_{an}\right) = t_a + O_p\left(n^{-1}\right)$, for $a = 1, .., 5$. In addition let $\mu_u = \mu_1\mu_2 - \mu_3\mu_4$, $\mu_d = \mu_5\mu_2 - \mu_4^2$.

In what follows, Theorem 2 shows the consistency and bias approximations when estimating κ, μ and σ^2 in model (2)[3]. Recalling that $\widehat{\kappa} = -\delta^{-1}\log\left(\hat{\beta}_1\right)$, our approach to analysing the bias of $\widehat{\kappa}$ is to first find a suitable expansion for $\hat{\beta}_1$,which subsumes the expansion used by Tang and Chen (2009), and then find an appropriate expansion for the transform. We first find that $\hat{\kappa} - \kappa = -\frac{1}{\delta\beta_1}\left(\widehat{\beta}_1 - \beta_1\right) + \frac{1}{2\delta\beta_1^2}\left(\widehat{\beta}_1 - \beta_1\right)^2 + O\left(n^{-2}\right)$ from which we obtain the following

Theorem 2. *For a stationary CKLS process, as $n \longrightarrow \infty$ while δ is fixed and assuming either (a) x_t is a series taking only non-negative values and $\rho > 0$ or (b) x_t is unrestricted and ρ is a general non-negative integer, and $E(\varepsilon_t|x_{t-1}) = 0$, the bias of the estimator $\hat{\kappa}$ is given by*

$$\begin{aligned}E(\hat{\kappa}-\kappa) = &\frac{\mu_2^2 var(t_1)}{2\delta e^{-2\kappa\delta}\mu_d^2} + (\mu_1\mu_5 + \frac{\left(\mu_1\mu_d - \mu_5\mu_u\right)^2}{2e^{-\kappa\delta}\mu_d^2} - \frac{\mu_u}{\mu_d}\mu_5^2)\frac{var(t_2)}{\delta e^{-\kappa\delta}\mu_d^2} + (2\mu_3\mu_4 + \frac{\mu_3^2}{2e^{-\kappa\delta}} - \mu_u - \frac{4\mu_u\mu_4^2}{\mu_d})\frac{var(t_4)}{\delta e^{-\kappa\delta}\mu_d^2}\\ &+ \frac{\mu_4^2 var(t_3)}{2\delta e^{-2\kappa\delta}\mu_d^2} + (\frac{\mu_u^2}{2\mu_d^2 e^{-\kappa\delta}} - \frac{\mu_u}{\mu_d})\frac{\mu_2^2 var(t_5)}{\delta e^{-\kappa\delta}\mu_d^2} - [1 - \frac{\mu_2\mu_5}{\mu_d} - \frac{\mu_2}{e^{-\kappa\delta}}(\frac{\mu_1}{\mu_d} - \frac{\mu_5\mu_u}{\mu_d^2})]\frac{Cov(t_1 t_2)}{\mu_d\delta e^{-\kappa\delta}}\end{aligned}$$

[3] Tang and Chen (2009) are able to obtain the bias expressions in terms of gamma and hypergeometric functions since when $\rho = 0.5$, for any $j > i$, $cx_j|x_i \sim \chi_\nu^2(\lambda)$ where $\nu = 4\kappa\mu\sigma^{-2}$, $\lambda = cx_i e^{-(i-j)\kappa\delta}$ and $c = 4\kappa\sigma^{-2}\left(1 - e^{-(i-j)\kappa\delta}\right)$. In our case, we will have to leave the bias expression in more general terms (since we cannot rely on chi-square distributional assumptions, something that only holds for $\rho = 0.5$).

$$
\begin{aligned}
&-\frac{\mu_2\mu_4 Cov(t_1t_3)}{\delta e^{-2\kappa\delta}\mu_d^2}-(2\mu_2\mu_4+\frac{\mu_1\mu_3}{e^{-\kappa\delta}})\frac{Cov(t_1t_4)}{\mu_d^2\delta e^{-\kappa\delta}}+(\frac{\mu_2^2}{\mu_d}-\frac{\mu_2^2\mu_u}{\mu_d^2 e^{-\kappa\delta}})\frac{Cov(t_1t_5)}{\delta e^{-\kappa\delta}\mu_d}\\
&-[\frac{\mu_5}{\mu_d}+\frac{\mu_1}{e^{-\kappa\delta}}(\frac{1}{\mu_d}-\frac{\mu_u}{\mu_d^2})]\frac{\mu_4 Cov(t_2t_3)}{\delta e^{-\kappa\delta}\mu_d}+[\frac{4\mu_5\mu_4\mu_u}{\mu_d^2}-\frac{2\mu_1\mu_4+\mu_3\mu_5}{\mu_d}-\frac{\mu_3}{e^{-\kappa\delta}}(\frac{\mu_1}{\mu_d}-\frac{\mu_5\mu_u}{\mu_d^2})]\frac{Cov(t_2t_4)}{\delta e^{-\kappa\delta}\mu_d}\\
&+[\frac{\mu_1\mu_2}{\mu_d}+\frac{\mu_u}{\mu_d}-\frac{2\mu_5\mu_2\mu_u}{\mu_d^2}+\frac{\mu_2}{e^{-\kappa\delta}}(\frac{\mu_u^2\mu_5}{\mu_d^3}-\frac{\mu_1\mu_u}{\mu_d^2})]\frac{Cov(t_2t_5)}{\delta e^{-\kappa\delta}\mu_d}+(1+\frac{2\mu_4^2}{\mu_d}+\frac{\mu_3\mu_4}{\mu_d e^{-\kappa\delta}})\frac{Cov(t_3t_4)}{\delta e^{-\kappa\delta}\mu_d}\\
&\quad+(\frac{\mu_u}{\mu_d^2 e^{-\kappa\delta}}-\frac{1}{\mu_d})\frac{Cov(t_3t_5)\mu_4\mu_2}{\delta e^{-\kappa\delta}\mu_d}+(-\frac{\mu_2\mu_3}{\mu_d}+2\frac{(\mu_4+\mu_1)\mu_u\mu_2}{\mu_d^2}+\frac{\mu_2\mu_3\mu_u}{\mu_d^2 e^{-\kappa\delta}})\frac{Cov(t_4t_5)}{\delta e^{-\kappa\delta}\mu_d}+o(n^{-1}).
\end{aligned}
$$

for $e^{-\kappa\delta}\neq 1$ and where all variances and covariances are of order n^{-1}. Also $E(\widehat{\mu}-\mu)=O\left(n^{-2}\right)$ for $\rho=0$, and assuming either (a) x_t is a series taking only positive values and $\rho>0$ or (b) ρ is a general positive integer

$$
E(\widehat{\mu}-\mu)=E(\frac{\sum_{t=1}^{n}x_{t-1}^{-2p}\varepsilon_t}{(1-e^{-\kappa\delta})\sum_{t=1}^{n}x_{t-1}^{-2p}})+E(\frac{\overline{F}\sum_{t=1}^{n}x_{t-1}^{-2p}\varepsilon_t}{(1-e^{-\kappa\delta})\sum_{t=1}^{n}x_{t-1}^{-2p}})+o\left(n^{-1}\right)
$$

$$
\text{with } \overline{F}=\frac{\mu_1t_2+\mu_2t_1-\mu_3t_4-\mu_4t_3}{\mu_d}-\frac{\mu_u}{\mu_d}\left(\mu_5t_2+\mu_2t_5\right)+o_p(n^{-1/2}).
$$

Clearly the above is a somewhat cumbersome expression which cannot readily be simplified. However for a given value for ρ, all the individual terms could, in principle, be replaced by consistent estimates and an estimated bias will be obtained. For the case considered by Tang and Chen (2009), when $\rho=\frac{1}{2}$, we consider the following remark.

Remark 4 *In the Appendix we show that if in Theorem 2 we set $\rho=0.5$, we obtain the special case given in Tang and Chen (2009, Theorem 3.1.3) for the bias of $\widehat{\kappa}$. However, for $\rho=0.5$, Tang and Chen (2009, Theorems 3.1.3), obtain that the bias of $\widehat{\mu}$ is of order n^{-2}, while our results in Theorem 2 contradict the result of Tang and Chen (2009) in respect of the bias of $\widehat{\mu}$ since we show that the bias is of order n^{-1} when $\rho>0$.*

3.2 PROOF of Theorem 2

We proceed now to expand $\widehat{\kappa}$ and $\widehat{\mu}$. First we start analyzing $\widehat{\kappa}$, and its main component $\widehat{\beta}_1$.

3.2.1 BIAS EXPANSION OF $\widehat{\kappa}$

Consistency of $\widehat{\beta}_1$ We start with the expansion of $\widehat{\beta}_1$ given by

$$\widehat{\beta}_1 = \frac{\sum_{t=1}^n x_t x_{t-1}^{(1-2\rho)} \sum_{t=1}^n x_{t-1}^{-2\rho} - \sum_{t=1}^n x_t x_{t-1}^{-2\rho} \sum_{t=1}^n x_{t-1}^{(1-2\rho)}}{\sum_{t=1}^n x_{t-1}^{(2-2\rho)} \sum_{t=1}^n x_{t-1}^{-2\rho} - \left(\sum_{t=1}^n \left(x_{t-1}^{(1-2\rho)}\right)\right)^2}$$

$$= \frac{\mu_1\mu_2 - \mu_3\mu_4 + \mu_1\widetilde{t}_2 + \mu_2 t_1 + \widetilde{t}_2 t_1 - \mu_3\widetilde{t}_4 - \mu_4 t_3 - \widetilde{t}_4 t_3}{\mu_5\mu_2 - \mu_4^2 + \mu_5\widetilde{t}_2 + \mu_2\widetilde{t}_5 + \widetilde{t}_2\widetilde{t}_5 - 2\mu_4\widetilde{t}_4 - \widetilde{t}_4\widetilde{t}_4}$$

$$= \frac{\mu_u}{\mu_d} + \frac{\mu_1 t_2 + \mu_2 t_1 - \mu_3 t_4 - \mu_4 t_3}{\mu_d} - \frac{\mu_u}{\mu_d^2}\left(\mu_5 t_2 + \mu_2 t_5\right) + \frac{t_2 t_1 - t_4 t_3}{\mu_d} - \frac{\mu_u}{\mu_d^2}\left(t_2 t_5 - t_4^2\right)$$

$$+ \frac{\mu_u}{\mu_d^3}\left(\mu_5^2 t_2^2 + \mu_5\mu_2 t_2 t_5 - 2\mu_5\mu_4 t_2 t_4 + \mu_2\mu_5 t_2 t_5 + \mu_2^2 t_5^2 - 2\mu_1\mu_2 t_4 t_5 - 2\mu_4\mu_5 t_2 t_4 - 2\mu_4\mu_2 t_4 t_5 + 4\mu_4^2 t_4^2\right)$$

$$- \frac{1}{\mu_d^2}(\mu_1\mu_5 t_2^2 + \mu_1\mu_2 t_2 t_5 - 2\mu_1\mu_4 t_2 t_4 + \mu_2\mu_5 t_1 t_2 + \mu_2^2 t_1 t_5 - 2\mu_2\mu_4 t_1 t_4 - \mu_3\mu_5 t_4 t_2 - \mu_3\mu_2 t_4 t_5$$

$$+ 2\mu_3\mu_4 t_4^2 - \mu_4\mu_5 t_3 t_2 - \mu_4\mu_2 t_3 t_5 + 2\mu_4^2 t_3 t_4) + R_n, \quad (7)$$

where $\mu_u = \mu_1\mu_2 - \mu_3\mu_4$, $\mu_d = \mu_5\mu_2 - \mu_4^2$ and $\frac{\mu_u}{\mu_d} = \beta_1$, which is shown below. Also with δ fixed R_n is $o_p\left(n^{-1}\right)$ and we have used the notation given in Section 3 of the main paper.

Before proceeding to obtain the asymptotic bias of $\widehat{\beta}_1$, we check that (7) is the same expansion as the one used by Tang and Chen (2009, page 76, expression (A.1)) for the case of $\rho = 0.5$. For $\rho = 0.5$ we have

$$t_{1i} = x_i - \mu_1,\ t_{2i} = x_{i-1}^{-1} - \mu_2,\ t_{3i} = x_i x_{i-1}^{-1} - \mu_3,\ t_{4i} = 1 - \mu_4 = 0,\ \text{since } \mu_4 = 1$$

$$\text{and } t_{5i} = x_{i-1} - \mu_5,\ \text{with } \mu_1 = \mu_5,$$

and then specializing equation (7) for $\rho = 0.5$

$$\widehat{\beta}_1 = \frac{\mu_1\mu_2 - \mu_3}{\mu_1\mu_2 - 1} + \frac{\mu_1 t_2 + \mu_2 t_1 - t_3}{\mu_1\mu_2 - 1} - \frac{\mu_1\mu_2 - \mu_3}{\left(\mu_1\mu_2 - 1\right)^2}\left(\mu_1 t_2 + \mu_2 t_5\right) + \frac{t_2 t_1}{\mu_1\mu_2 - 1} - \frac{\mu_1\mu_2 - \mu_3}{\left(\mu_1\mu_2 - 1\right)^2} t_2 t_5$$

$$+ \frac{\mu_1\mu_2 - \mu_3}{\left(\mu_1\mu_2 - 1\right)^3}\left(\mu_1^2 t_2^2 + \mu_1\mu_2 t_2 t_5 + \mu_2\mu_5 t_2 t_5 + \mu_2^2 t_5^2\right)$$

$$- \frac{1}{\left(\mu_1\mu_2 - 1\right)^2}(\mu_1^2 t_2^2 + \mu_1\mu_2 t_2 t_5 + \mu_2\mu_1 t_1 t_2 + \mu_2^2 t_1 t_5 - \mu_1 t_3 t_2 - \mu_2 t_3 t_5) + R_n, \quad (8)$$

and (8) reduces to (A.1) in Tang and Chen (2009, page 76) when setting $t_5 = t_1$. We can use also (7) to check the consistency of $\widehat{\beta}_1$ by showing that $\beta_1 = \frac{\mu_u}{\mu_d}$ as follows

$$\frac{\mu_u}{\mu_d} = \frac{\mu_1\mu_2 - \mu_3\mu_4}{\mu_5\mu_2 - \mu_4^2} = \frac{\left(e^{-\delta\kappa}\mu_5 + \mu\left(1 - e^{-\delta\kappa}\right)\mu_4\right)\mu_2 - \left(e^{-\delta\kappa}\mu_4 + \mu\left(1 - e^{-\delta\kappa}\right)\mu_2\right)\mu_4}{\mu_5\mu_2 - \mu_4^2}$$

$$= \frac{e^{-\delta\kappa}\mu_5\mu_2 + \mu\left(1 - e^{-\delta\kappa}\right)\mu_4\mu_2 - e^{-\delta\kappa}\mu_4^2 - \mu\left(1 - e^{-\delta\kappa}\right)\mu_2\mu_4}{\mu_5\mu_2 - \mu_4^2} = \frac{e^{-\delta\kappa}\mu_5\mu_2 - e^{-\delta\kappa}\mu_4^2}{\mu_5\mu_2 - \mu_4^2} = e^{-\delta\kappa} = \beta_1.$$

This indicates that $\widehat{\beta}_1$ is a consistent estimator of β_1 for any non-negative even integer 2ρ (generalizing the result of Tang and Chen (2009, page 76 for $\rho = 0.5$).

Bias expansion of $\widehat{\beta}_1$ Now we proceed to analyze the bias $\widehat{\beta}_1$ of the case of non-negative even integer ρ by using a general expansion using (7). In the expansion of $\widehat{\kappa}$, we will show that we need to consider two terms: $E(\widehat{\beta}_1 - \beta_1)$ and $E(\widehat{\beta}_1 - \beta_1)^2$. We shall first consider $E(\widehat{\beta}_1 - \beta_1)$ to order $n^{-1}$and since $E(t_i) = 0$, $i = 1, 2, 3, 4, 5$ we find that when taking expectations we need only consider terms which involve a product of the t_i. We shall also use $\frac{\mu_u}{\mu_d} = \beta_1$ (due to the consistency of $\hat{\beta}_1$ noted previously). Rearranging (7), we find that

$$\begin{aligned}(\hat{\beta}_1 - \beta_1) \quad = \quad & \frac{\mu_1 t_2 + \mu_2 t_1 - \mu_3 t_4 - \mu_4 t_3}{\mu_d} - \frac{\mu_u}{\mu_d^2}(\mu_5 t_2 + \mu_2 t_5) + (\frac{\mu_u}{\mu_d^3}\mu_5^2 - \frac{1}{\mu_d^2}\mu_1\mu_5)t_2^2 \\ & + (\frac{\mu_u}{\mu_d^2} + \frac{\mu_u 4\mu_4^2}{\mu_d^3} - \frac{2\mu_3\mu_4}{\mu_d^2})t_4^2 + \frac{\mu_u}{\mu_d^3}\mu_2^2 t_5^2 + (\frac{1}{\mu_d} - \mu_2\mu_5\frac{1}{\mu_d^2})t_1 t_2 \\ & +2\mu_2\mu_4\frac{1}{\mu_d^2}t_1 t_4 - \frac{1}{\mu_d^2}\mu_2^2 t_1 t_5 + \frac{1}{\mu_d^2}\mu_4\mu_5 t_3 t_2 - (4\mu_5\mu_4\frac{\mu_u}{\mu_d^3} - 2\mu_1\mu_4\frac{1}{\mu_d^2} - \frac{1}{\mu_d^2}\mu_3\mu_5)t_2 t_4 \\ & +(2\frac{\mu_u}{\mu_d^3}\mu_5\mu_2 - \frac{1}{\mu_d^2}\mu_1\mu_2 - \frac{\mu_u}{\mu_d^2})t_2 t_5 - (\frac{1}{\mu_d} + \frac{2\mu_4^2}{\mu_d^2})t_3 t_4 + \frac{1}{\mu_d^2}\mu_4\mu_2 t_3 t_5 \\ & -(2\frac{\mu_u}{\mu_d^3}(\mu_4\mu_2 + \mu_1\mu_2) - \frac{1}{\mu_d^2}\mu_3\mu_2)t_4 t_5 + R_n.\end{aligned}$$

Finally taking expectations, noting that $E(t_i) = 0$, $i = 1, 2, 3, 4, 5$ yields

$$\begin{aligned}E(\hat{\beta}_1 - \beta_1) = & (\frac{\mu_u}{\mu_d^3}\mu_5^2 - \frac{1}{\mu_d^2}\mu_1\mu_5)var(t_2) + (\frac{\mu_u}{\mu_d^2} + \frac{\mu_u 4\mu_4^2}{\mu_d^3} - \frac{2\mu_3\mu_4}{\mu_d^2})var(t_4) + (\frac{\mu_u}{\mu_d^3}\mu_2^2)var(t_5) \\ & + (\frac{1}{\mu_d} - \mu_2\mu_5\frac{1}{\mu_d^2})Cov(t_1 t_2) + 2\mu_2\mu_4\frac{1}{\mu_d^2}Cov(t_1 t_4) - \frac{1}{\mu_d^2}\mu_2^2 Cov(t_1 t_5) + \frac{1}{\mu_d^2}\mu_4\mu_5 Cov(t_2 t_3) \\ & - (4\mu_5\mu_4\frac{\mu_u}{\mu_d^3} - 2\mu_1\mu_4\frac{1}{\mu_d^2} - \frac{1}{\mu_d^2}\mu_3\mu_5)Cov(t_2 t_4) + (2\frac{\mu_u}{\mu_d^3}\mu_5\mu_2 - \frac{1}{\mu_d^2}\mu_1\mu_2 - \frac{\mu_u}{\mu_d^2})Cov(t_2 t_5) \\ & + (-\frac{2\mu_4^2}{\mu_d^2} - \frac{1}{\mu_d})Cov(t_3 t_4) + \frac{1}{\mu_d^2}\mu_4\mu_2 Cov(t_3 t_5) + (\frac{1}{\mu_d^2}\mu_2\mu_3 - 2\frac{\mu_u}{\mu_d^3}(\mu_4\mu_2 + \mu_1\mu_2))Cov(t_4 t_5) + o(n^{-1}) \quad (9)\end{aligned}$$

We move now to the expansion of $E(\widehat{\beta}_1 - \beta_1)^2$. From (7) we can show that

$$\widehat{\beta}_1 - \beta_1 = \frac{\mu_1 t_2 + \mu_2 t_1 - \mu_3 t_4 - \mu_4 t_3}{\mu_d} - \frac{\mu_u}{\mu_d^2}(\mu_5 t_2 + \mu_2 t_5) + o_p(n^{-1/2}). \quad (10)$$

Then

$$
\begin{aligned}
(\widehat{\beta}_1 - \beta_1)^2 &= (\frac{\mu_1 t_2 + \mu_2 t_1 - \mu_3 t_4 - \mu_4 t_3}{\mu_d})^2 + (\frac{\mu_u}{\mu_d^2}(\mu_5 t_2 + \mu_2 t_5))^2 \\
&\quad -2(\frac{\mu_1 t_2 + \mu_2 t_1 - \mu_3 t_4 - \mu_4 t_3}{\mu_d})(\frac{\mu_u}{\mu_d^2}(\mu_5 t_2 + \mu_2 t_5)) + o_p(n^{-1})
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{\mu_d^2}(\mu_1 t_2 + \mu_2 t_1 - \mu_3 t_4 - \mu_4 t_3)^2 + \left(\frac{\mu_u}{\mu_d^2}\right)^2 (\mu_2 t_5 + \mu_5 t_2)^2 - \frac{2\mu_u}{\mu_d^3}(\mu_2 t_5 + \mu_5 t_2)(\mu_1 t_2 + \mu_2 t_1 - \mu_3 t_4 - \mu_4 t_3) = \\
&\quad \frac{1}{\mu_d^2}(\mu_1^2 t_2^2 + 2\mu_1\mu_2 t_1 t_2 - 2\mu_1\mu_3 t_2 t_4 - 2\mu_1\mu_4 t_2 t_3 + \mu_2^2 t_1^2 - 2\mu_2\mu_3 t_1 t_4 - 2\mu_2\mu_4 t_1 t_3 + \mu_3^2 t_4^2 \\
&\quad + 2\mu_3\mu_4 t_3 t_4 + \mu_4^2 t_3^2) + (\frac{\mu_u}{\mu_d^2})^2(\mu_5^2 t_2^2 + \mu_2^2 t_5^2 + 2\mu_5 t_2 \mu_2 t_5) - 2\frac{\mu_u}{\mu_d^3}(\mu_1\mu_5 t_2^2 + \mu_1\mu_2 t_2 t_5 + \mu_2\mu_5 t_1 t_2 \\
&\quad + \mu_2^2 t_1 t_5 - \mu_3\mu_5 t_2 t_4 - \mu_2\mu_3 t_4 t_5 - \mu_4 t_3 \mu_1 t_2 - \mu_2\mu_4 t_3 t_5) = \frac{\mu_2^2}{\mu_d^2} t_1^2 + [\frac{\mu_1^2}{\mu_d^2} + (\frac{\mu_u}{\mu_d^2})^2 \mu_5^2 - 2\frac{\mu_u}{\mu_d^3}\mu_1\mu_5] t_2^2 \\
&\quad + \frac{\mu_4^2}{\mu_d^2} t_3^2 + \frac{\mu_3^2}{\mu_d^2} t_4^2 + (\frac{\mu_u}{\mu_d^2})^2 \mu_2^2 t_5^2 + (\frac{2\mu_1\mu_2}{\mu_d^2} - 2\frac{\mu_u}{\mu_d^3}\mu_5\mu_2) t_1 t_2 - \frac{1}{\mu_d^2}(2\mu_2\mu_4) t_1 t_3 - 2\frac{1}{\mu_d^2}\mu_2\mu_3 t_1 t_4 \\
&\quad - 2\frac{\mu_u}{\mu_d^3}\mu_2^2 t_1 t_5 + (\frac{1}{\mu_d^2}(-2\mu_1\mu_4) + 2\frac{\mu_u}{\mu_d^3}\mu_1\mu_4) t_2 t_3 - (\frac{2\mu_3\mu_1}{\mu_d^2} - 2(\frac{\mu_u}{\mu_d^3})\mu_5\mu_3) t_2 t_4 \\
&\quad + [(\frac{\mu_u}{\mu_d^2})^2 2\mu_5\mu_2 - 2\frac{\mu_u}{\mu_d^3}\mu_1\mu_2] t_2 t_5 + \frac{1}{\mu_d^2} 2\mu_3\mu_4 t_3 t_4 + 2\frac{\mu_u}{\mu_d^3}\mu_4\mu_2 t_3 t_5 + 2\frac{\mu_u}{\mu_d^3}\mu_2\mu_3 t_4 t_5 + o_p(n^{-1}).
\end{aligned}
$$

Hence $E(\widehat{\beta}_1 - \beta_1)^2$, to order n^{-1}, is given by

$$
\begin{aligned}
E(\widehat{\beta}_1 - \beta_1)^2 &= \frac{\mu_2^2}{\mu_d^2} var(t_1) + \frac{1}{\mu_d^4}(\mu_1\mu_d - \mu_5\mu_u)^2 var(t_2) + \frac{\mu_4^2}{\mu_d^2} var(t_3) + \frac{\mu_3^2}{\mu_d^2} var(t_4) + \mu_2^2 (\frac{\mu_u}{\mu_d^2})^2 var(t_5) \\
&\quad + 2\mu_2(\frac{\mu_1}{\mu_d^2} - \frac{\mu_5\mu_u}{\mu_d^3}) Cov(t_1 t_2) - \frac{1}{\mu_d^2}(2\mu_2\mu_4) Cov(t_1 t_3) - 2\mu_2\mu_3 \frac{1}{\mu_d^2} Cov(t_1 t_4) \\
&\quad - 2\mu_2^2 \frac{\mu_u}{\mu_d^3} Cov(t_1 t_5) - 2\mu_1\mu_4(\frac{1}{\mu_d^2} - \frac{\mu_u}{\mu_d^3}) Cov(t_2 t_3) - 2\mu_3(\frac{\mu_1}{\mu_d^2} - \frac{\mu_5\mu_u}{\mu_d^3}) Cov(t_2 t_4) \\
&\quad + 2\mu_2(\frac{\mu_5\mu_u^2}{\mu_d^4} - \frac{\mu_u\mu_1}{\mu_d^3}) Cov(t_2 t_5) + 2\mu_3\mu_4 \frac{1}{\mu_d^2} Cov(t_3 t_4) + 2\mu_4\mu_2 \frac{\mu_u}{\mu_d^3} Cov(t_3 t_5) + 2\mu_2\mu_3 \frac{\mu_u}{\mu_d^3} Cov(t_4 t_5) \quad (11)
\end{aligned}
$$

It is now possible to find an approximation for the bias of $\hat{\kappa}$ to $o(n^{-1})$.

Expansion of $\widehat{\kappa}$ Finally, in order to transform $\widehat{\beta}_1$ back to $\widehat{\kappa}$ for a fixed δ, we first carry out a Taylor expansion up to the second order term

$$
\begin{aligned}
\widehat{\kappa} &= -\delta^{-1} \log\left(\widehat{\beta}_1\right) = -\delta^{-1}\left[\log(\beta_1) + \frac{1}{\beta_1}\left(\widehat{\beta}_1 - \beta_1\right) - \frac{1}{2\beta_1^2}\left(\widehat{\beta}_1 - \beta_1\right)^2\right] + O_p\left(n^{-2}\right) \\
&= \kappa - \frac{1}{\delta\beta_1}\left(\widehat{\beta}_1 - \beta_1\right) + \frac{1}{2\delta\beta_1^2}\left(\widehat{\beta}_1 - \beta_1\right)^2 + O\left(n^{-2}\right),
\end{aligned}
$$

and therefore from the asymptotic bias of $\widehat{\beta}_1$ and its second moment it is seen that

$$E\left(\widehat{\kappa}-\kappa\right)=-\frac{1}{\delta\beta_1}E\left(\widehat{\beta}_1-\beta_1\right)+\frac{1}{2\delta\beta_1^2}E\left(\widehat{\beta}_1-\beta_1\right)^2+O\left(n^{-2}\right),$$

where $E\left(\widehat{\beta}_1-\beta_1\right)$ is given at (9) and $E\left(\widehat{\beta}_1-\beta_1\right)^2$ is given at (11). Note that $\left(\widehat{\beta}_1-\beta_1\right)$ is $O_p\left(n^{-1/2}\right)$ and $\left(\widehat{\beta}_1-\beta_1\right)^2$ is $O_p\left(n^{-1}\right)$. Therefore we conclude that the bias of $\widehat{\kappa}$ is of order n^{-1}, for any non-negative integer ρ and also of order T^{-1} and it is given by

$$E(\hat{\kappa}-\kappa)=\frac{-1}{\delta\beta_1}E(\hat{\beta}_1-\beta_1)+\frac{1}{2\delta\beta_1^2}E(\hat{\beta}_1-\beta_1)^2+o(n^{-1}) \tag{12}$$

$$\begin{aligned}
&=\frac{-1}{\delta\beta_1}[(\frac{\mu_u}{\mu_d^3}\mu_5^2-\frac{1}{\mu_d^2}\mu_1\mu_5)var(t_2)+(\frac{\mu_u}{\mu_d^2}+\frac{\mu_u 4\mu_4^2}{\mu_d^3}-\frac{2\mu_3\mu_4}{\mu_d^2})var(t_4)+(\frac{\mu_u}{\mu_d^3}\mu_2^2)var(t_5)+(\frac{1}{\mu_d}-\mu_2\mu_5\frac{1}{\mu_d^2})Cov(t_1t_2)\\
&+2\mu_2\mu_4\frac{1}{\mu_d^2}Cov(t_1t_4)-\frac{1}{\mu_d^2}\mu_2^2Cov(t_1t_5)+\frac{1}{\mu_d^2}\mu_4\mu_5Cov(t_2t_3)-(4\mu_5\mu_4\frac{\mu_u}{\mu_d^3}-2\mu_1\mu_4\frac{1}{\mu_d^2}-\frac{1}{\mu_d^2}\mu_3\mu_5)Cov(t_2t_4)\\
&+(2\frac{\mu_u}{\mu_d^3}\mu_5\mu_2-\frac{1}{\mu_d^2}\mu_1\mu_2-\frac{\mu_u}{\mu_d^2})Cov(t_2t_5)+(-\frac{2\mu_4^2}{\mu_d^2}-\frac{1}{\mu_d})Cov(t_3t_4)+\frac{1}{\mu_d^2}\mu_4\mu_2Cov(t_3t_5)\\
&+(\frac{1}{\mu_d^2}\mu_2\mu_3-2\frac{\mu_u}{\mu_d^3}(\mu_4\mu_2+\mu_1\mu_2))Cov(t_4t_5)]+\frac{1}{2\delta\beta_1^2}[\frac{\mu_2^2}{\mu_d^2}var(t_1)+\frac{1}{\mu_d^4}\left(\mu_1\mu_d-\mu_5\mu_u\right)^2var(t_2)+\frac{\mu_4^2}{\mu_d^2}var(t_3)+\frac{\mu_3^2}{\mu_d^2}var(t_4)\\
&+\mu_2^2(\frac{\mu_u}{\mu_d^2})^2var(t_5)+2\mu_2(\frac{\mu_1}{\mu_d^2}-\frac{\mu_5\mu_u}{\mu_d^3})Cov(t_1t_2)-\frac{1}{\mu_d^2}(2\mu_2\mu_4)Cov(t_1t_3)-2\mu_1\mu_3\frac{1}{\mu_d^2}Cov(t_1t_4)\\
&-2\mu_2^2\frac{\mu_u}{\mu_d^3}Cov(t_1t_5)-2\mu_1\mu_4(\frac{1}{\mu_d^2}-\frac{\mu_u}{\mu_d^3})Cov(t_2t_3)-2\mu_3(\frac{\mu_1}{\mu_d^2}-\frac{\mu_5\mu_u}{\mu_d^3})Cov(t_2t_4)+2\mu_2(\frac{\mu_u^2\mu_5}{\mu_d^4}-\frac{\mu_1\mu_u}{\mu_d^3})Cov(t_2t_5)\\
&+2\mu_3\mu_4\frac{1}{\mu_d^2}Cov(t_3t_4)+2\mu_4\mu_2\frac{\mu_u}{\mu_d^3}Cov(t_3t_5)+2\mu_2\mu_3\frac{\mu_u}{\mu_d^3}Cov(t_4t_5)]\\
&=\frac{\mu_2^2var(t_1)}{2\delta\beta_1^2\mu_d^2}+(\mu_1\mu_5+\frac{(\mu_1\mu_d-\mu_5\mu_u)^2}{2\beta_1\mu_d^2}-\frac{\mu_u}{\mu_d}\mu_5^2)\frac{var(t_2)}{\delta\beta_1\mu_d^2}+\frac{\mu_4^2var(t_3)}{2\delta\beta_1^2\mu_d^2}+(2\mu_3\mu_4+\frac{\mu_3^2}{2\beta_1}-\mu_u-\frac{4\mu_u\mu_4^2}{\mu_d})\frac{var(t_4)}{\delta\beta_1\mu_d^2}\\
&+(\frac{\mu_u^2}{2\mu_d^2\beta_1}-\frac{\mu_u}{\mu_d})\frac{\mu_2^2var(t_5)}{\delta\beta_1\mu_d^2}-[1-\frac{\mu_2\mu_5}{\mu_d}-\frac{\mu_2}{\beta_1}(\frac{\mu_1}{\mu_d}-\frac{\mu_5\mu_u}{\mu_d^2})]\frac{Cov(t_1t_2)}{\mu_d\delta\beta_1}-\frac{\mu_2\mu_4Cov(t_1t_3)}{\delta\beta_1^2\mu_d^2}\\
&-(2\mu_2\mu_4+\frac{\mu_1\mu_3}{\beta_1})\frac{Cov(t_1t_4)}{\mu_d^2\delta\beta_1}+(\frac{\mu_2^2}{\mu_d}-\frac{\mu_2^2\mu_u}{\beta_1\mu_d^2})\frac{Cov(t_1t_5)}{\delta\beta_1\mu_d}\\
&-[\frac{\mu_5}{\mu_d}+\frac{\mu_1}{\beta_1}(\frac{1}{\mu_d}-\frac{\mu_u}{\mu_d^2})]\frac{\mu_4Cov(t_2t_3)}{\delta\beta_1\mu_d}+[\frac{4\mu_5\mu_4\mu_u}{\mu_d^2}-\frac{2\mu_1\mu_4+\mu_3\mu_5}{\mu_d}-\frac{\mu_3}{\beta_1}(\frac{\mu_1}{\mu_d}-\frac{\mu_5\mu_u}{\mu_d^2})]\frac{Cov(t_2t_4)}{\delta\beta_1\mu_d}\\
&+[\frac{\mu_1\mu_2}{\mu_d}+\frac{\mu_u}{\mu_d}-\frac{2\mu_5\mu_2\mu_u}{\mu_d^2}+\frac{\mu_2}{\beta_1}(\frac{\mu_5\mu_u^2}{\mu_d^3}-\frac{\mu_1\mu_u}{\mu_d^2})]\frac{Cov(t_2t_5)}{\delta\beta_1\mu_d}+(1+\frac{2\mu_4^2}{\mu_d}+\frac{\mu_3\mu_4}{\mu_d\beta_1})\frac{Cov(t_3t_4)}{\delta\beta_1\mu_d}\\
&+(\frac{\mu_u}{\mu_d^2\beta_1}-\frac{1}{\mu_d})\frac{Cov(t_3t_5)\mu_4\mu_2}{\delta\beta_1\mu_d}+(-\frac{\mu_2\mu_3}{\mu_d}+2\frac{(\mu_4\mu_2+\mu_1\mu_2)\mu_u}{\mu_d^2}+\frac{\mu_2\mu_3\mu_u}{\mu_d^2\beta_1})\frac{Cov(t_4t_5)}{\delta\beta_1\mu_d}+o(n^{-1}).
\end{aligned}$$

3.2.2 BIAS EXPANSION OF $\widehat{\mu}$

From Proposition 1, we know that $\widehat{\mu} = \widehat{\beta}_2$. We first focus on the consistency of $\widehat{\mu}$.

Consistency of $\widehat{\mu}$ To prove that $\widehat{\beta}_2$ is a consistent estimator of μ for any non-negative integer ρ and where

$$E(x_j|x_i) = e^{-\delta_{ij}\kappa}x_i + \mu\left(1 - e^{-\delta_{ij}\kappa}\right) \text{ with } \delta_{ij} = \delta|j - i|. \tag{13}$$

Let $t_{1i} = x_i x_{i-1}^{(1-2\rho)} - \mu_1$, $t_{2i} = x_{i-1}^{-2\rho} - \mu_2$, $t_{3i} = x_i x_{i-1}^{-2\rho} - \mu_3$, $t_{4i} = x_{i-1}^{(1-2\rho)} - \mu_4$ and $t_{5i} = x_{i-1}^{(2-2\rho)} - \mu_5$, where from (13)

$$\begin{aligned}
\mu_1 &= E\left(x_i x_{i-1}^{(1-2\rho)}\right) = E\left(x_{t-1}^{(1-2\rho)} E(x_t|x_{t-1})\right) = E\left(x_{t-1}^{(1-2\rho)}\left(e^{-\delta\kappa}x_{t-1} + \mu\left(1 - e^{-\delta\kappa}\right)\right)\right) \\
&= e^{-\delta\kappa}E\left(x_{t-1}^{(2-2\rho)}\right) + \mu\left(1 - e^{-\delta\kappa}\right)E\left(x_{t-1}^{(1-2\rho)}\right) = e^{-\delta\kappa}\mu_5 + \mu\left(1 - e^{-\delta\kappa}\right)\mu_4, \\
\mu_3 &= E\left(x_t x_{t-1}^{-2\rho}\right) = E\left(x_{t-1}^{-2\rho}E(x_t|x_{t-1})\right) = E\left(x_{t-1}^{-2\rho}\left(e^{-\delta\kappa}x_{t-1} + \mu\left(1 - e^{-\delta\kappa}\right)\right)\right) \\
&= e^{-\delta\kappa}E\left(x_{t-1}^{(1-2\rho)}\right) + \mu\left(1 - e^{-\delta\kappa}\right)E\left(x_{t-1}^{-2\rho}\right) = e^{-\delta\kappa}\mu_4 + \mu\left(1 - e^{-\delta\kappa}\right)\mu_2,
\end{aligned}$$

$\mu_2 = E\left(x_{t-1}^{-2\rho}\right)$, $\mu_4 = E\left(x_{t-1}^{(1-2\rho)}\right)$ and $\mu_5 = E\left(x_{t-1}^{(2-2\rho)}\right)$. Also define $t_a = n^{-1}\sum_{i=1}^{n} t_{ai}$ which is $O_p(n^{\frac{-1}{2}})$ and let $\widetilde{t}_a = n^{-1}\sum_{i=1}^{n} t_{a(i-1)} = t_a + n^{-1}(t_{a0} - t_{an}) = t_a + O_p\left(n^{-1}\right)$, for $a = 1, .., 5$. In addition let $\mu_u = \mu_1\mu_2 - \mu_3\mu_4$, $\mu_d = \mu_5\mu_2 - \mu_4^2$. Then we notice that using the relevant expectations given above

$$\begin{aligned}
\widehat{\beta}_2 &= \frac{\sum_{t=1}^{n}\left(x_t x_{t-1}^{-2\rho} - \widehat{\beta}_1 x_{t-1}^{(1-2\rho)}\right)}{\left(1 - \widehat{\beta}_1\right)\sum_{t=1}^{n} x_{t-1}^{-2\rho}} \Longrightarrow p\lim\widehat{\beta}_2 = \frac{\mu_3 - e^{-\delta\kappa}\mu_4}{\left(1 - e^{-\delta\kappa}\right)\mu_2} \\
&= \frac{\left(e^{-\delta\kappa}\mu_4 + \mu\left(1 - e^{-\delta\kappa}\right)\mu_2\right) - e^{-\delta\kappa}\mu_4}{\left(1 - e^{-\delta\kappa}\right)\mu_2} = \frac{\mu\left(1 - e^{-\delta\kappa}\right)\mu_2}{\left(1 - e^{-\delta\kappa}\right)\mu_2} = \mu.
\end{aligned}$$

where we find that $\widehat{\beta}_2$ is a consistent estimator of μ for any non-negative integer ρ and also when $\rho = 0.5$.

Bias expansion of $\widehat{\mu}$ From (2), we can commence from

$$x_t = \beta_1 x_{t-1} + \mu(1 - \beta_1) + \varepsilon_t \tag{14}$$

since the estimator of β_1 will depend on β_2, we start from

$$x_t - \beta_1 x_{t-1} = (1 - \beta_1)\beta_2 + \varepsilon_t$$

i.e.

$$\frac{x_t - \beta_1 x_{t-1}}{1-\beta_1} = \beta_2 + \frac{\varepsilon_t}{1-\beta_1}$$

and multiply by x_{t-1}^{-2p} and sum over n to yield

$$\frac{\sum_{t=1}^n x_t x_{t-1}^{-2p} - \beta_1 \sum_{t=1}^n x_{t-1}^{1-2p}}{1-\beta_1} = \beta_2 \sum_{t=1}^n x_{t-1}^{-2p} + \frac{\sum_{t=1}^n x_{t-1}^{-2p}\varepsilon_t}{1-\beta_1}$$

and solving for β_2 yields

$$\beta_2 = \frac{\sum_{t=1}^n x_t x_{t-1}^{-2p} - \beta_1 \sum_{t=1}^n x_{t-1}^{1-2p}}{(1-\beta_1)\sum_{t=1}^n x_{t-1}^{-2p}} - \frac{\sum_{t=1}^n x_{t-1}^{-2p}\varepsilon_t}{(1-\beta_1)\sum_{t=1}^n x_{t-1}^{-2p}} \tag{15}$$

From (15), we deduce that

$$\hat{\beta}_2 = \frac{\sum_{t=1}^n x_t x_{t-1}^{-2p} - \hat{\beta}_1 \sum_{t=1}^n x_{t-1}^{1-2p}}{(1-\hat{\beta}_1)\sum_{t=1}^n x_{t-1}^{-2p}}$$

If β_1 is known then

$$\hat{\beta}_2^* = \frac{\sum_{t=1}^n x_t x_{t-1}^{-2p} - \beta_1 \sum_{t=1}^n x_{t-1}^{1-2p}}{(1-\beta_1)\sum_{t=1}^n x_{t-1}^{-2p}} = \beta_2 + \frac{\sum_{t=1}^n x_{t-1}^{-2p}\varepsilon_t}{(1-\beta_1)\sum_{t=1}^n x_{t-1}^{-2p}}$$

so that $\frac{\sum_{t=1}^n x_{t-1}^{-2p}\varepsilon_t}{(1-\beta_1)\sum_{t=1}^n x_{t-1}^{-2p}}$ is the estimation error when estimating β_2 if β_1 is known.

Now consider the estimator for β_2 when β_1 is unknown so that β_1 is replaced by $\hat{\beta}_1$

$$\hat{\beta}_2 = \frac{\sum_{t=1}^n x_t x_{t-1}^{-2p} - \hat{\beta}_1 \sum_{t=1}^n x_{t-1}^{1-2p}}{(1-\hat{\beta}_1)\sum_{t=1}^n x_{t-1}^{-2p}} = \frac{\sum_{t=1}^n x_t x_{t-1}^{-2p} - \beta_1 \sum_{t=1}^n x_{t-1}^{1-2p} - (\hat{\beta}_1-\beta_1)\sum_{t=1}^n x_{t-1}^{1-2p}}{(1-\beta_1)\sum_{t=1}^n x_{t-1}^{-2p} - (\hat{\beta}_1-\beta_1)\sum_{t=1}^n x_{t-1}^{-2p}}$$

where $(\hat{\beta}_1 - \beta_1)$ is $O_p(T^{\frac{1}{2}})$. Then

$$\begin{aligned}
\hat{\beta}_2 &= \frac{\sum_{t=1}^n x_t x_{t-1}^{-2p} - \beta_1 \sum_{t=1}^n x_{t-1}^{1-2p} - (\hat{\beta}_1-\beta_1)\sum_{t=1}^n x_{t-1}^{1-2p}}{(1-\beta_1)\sum_{t=1}^n x_{t-1}^{-2p}(1-\frac{(\hat{\beta}_1-\beta_1)}{(1-\beta_1)})} \\
&= \frac{\sum_{t=1}^n x_t x_{t-1}^{-2p} - \beta_1 \sum_{t=1}^n x_{t-1}^{1-2p}}{1-\beta_1)\sum_{t=1}^n x_{t-1}^{-2p}}(1-\frac{(\hat{\beta}_1-\beta_1)}{(1-\beta_1)})^{-1} - \frac{(\hat{\beta}_1-\beta_1)\sum_{t=1}^n x_{t-1}^{1-2p}}{(1-\beta_1)\sum_{t=1}^n x_{t-1}^{-2p}}(1-\frac{(\hat{\beta}_1-\beta_1)}{(1-\beta_1)})^{-1} \\
&= (\beta_2 + \frac{\sum_{t=1}^n x_{t-1}^{-2p}\varepsilon_t}{1-\beta_1)\sum_{t=1}^n x_{t-1}^{-2p}})(1+\frac{(\hat{\beta}_1-\beta_1)}{(1-\beta_1)} + (\frac{(\hat{\beta}_1-\beta_1)}{(1-\beta_1)})^2 + ...) \\
&\quad - \frac{(\hat{\beta}_1-\beta_1)\sum_{t=1}^n x_{t-1}^{1-2p}}{(1-\beta_1)\sum_{t=1}^n x_{t-1}^{-2p}}(1+\frac{(\hat{\beta}_1-\beta_1)}{(1-\beta_1)} + ...)
\end{aligned}$$

Then

$$\begin{aligned}
\hat{\beta}_2-\beta_2 &= \frac{\sum_{t=1}^{n} x_{t-1}^{-2p}\varepsilon_t}{(1-\beta_1)\sum_{t=1}^{n} x_{t-1}^{-2p}}+(\beta_2+\frac{\sum_{t=1}^{n} x_{t-1}^{-2p}\varepsilon_t}{(1-\beta_1)\sum_{t=1}^{n} x_{t-1}^{-2p}})\;(\frac{(\hat{\beta}_1-\beta_1)}{(1-\beta_1)}+(\frac{(\hat{\beta}_1-\beta_1)}{(1-\beta_1)})^2+ \\
&\quad -\frac{(\hat{\beta}_1-\beta_1)\sum_{t=1}^{n} x_{t-1}^{1-2p}}{(1-\beta_1)\sum_{t=1}^{n} x_{t-1}^{-2p}}(1+\frac{(\hat{\beta}_1-\beta_1)}{(1-\beta_1)}+...) \\
&= \frac{\sum_{t=1}^{n} x_{t-1}^{-2p}\varepsilon_t}{(1-\beta_1)\sum_{t=1}^{n} x_{t-1}^{-2p}}+\beta_2(\frac{(\hat{\beta}_1-\beta_1)}{(1-\beta_1)}+(\frac{(\hat{\beta}_1-\beta_1)}{(1-\beta_1)})^2)+\frac{\sum_{t=1}^{n} x_{t-1}^{-2p}\varepsilon_t}{(1-\beta_1)\sum_{t=1}^{n} x_{t-1}^{-2p}}(\frac{(\hat{\beta}_1-\beta_1)}{(1-\beta_1)}) \\
&\quad -\frac{(\hat{\beta}_1-\beta_1)\sum_{t=1}^{n} x_{t-1}^{1-2p}}{(1-\beta_1)\sum_{t=1}^{n} x_{t-1}^{-2p}}- -\frac{(\hat{\beta}_1-\beta_1)^2\sum_{t=1}^{n} x_{t-1}^{1-2p}}{(1-\beta_1)^2\sum_{t=1}^{n} x_{t-1}^{-2p}} \\
&\quad +o_p(n^{-1}).
\end{aligned}$$

And therefore $E\left(\widehat{\beta}_2-\beta_2\right)$ can be decomposed in six terms

$$\begin{aligned}
E\left(\widehat{\beta}_2-\beta_2\right) &= \frac{\beta_2 E\left(\widehat{\beta}_1-\beta_1\right)}{(1-\beta_1)}+\frac{\beta_2 E\left(\widehat{\beta}_1-\beta_1\right)^2}{(1-\beta_1)^2}+E(\frac{\sum_{t=1}^{n} x_{t-1}^{-2p}\varepsilon_t}{(1-\beta_1)\sum_{t=1}^{n} x_{t-1}^{-2p}}) \\
&\quad +E(\frac{\left(\widehat{\beta}_1-\beta_1\right)\sum_{t=1}^{n} x_{t-1}^{-2p}\varepsilon_t}{(1-\beta_1)\sum_{t=1}^{n} x_{t-1}^{-2p}})-E(\frac{(\hat{\beta}_1-\beta_1)\sum_{t=1}^{n} x_{t-1}^{1-2p}}{(1-\beta_1)\sum_{t=1}^{n} x_{t-1}^{-2p}})-E(\frac{(\hat{\beta}_1-\beta_1)^2\sum_{t=1}^{n} x_{t-1}^{1-2p}}{(1-\beta_1)^2\sum_{t=1}^{n} x_{t-1}^{-2p}}) \\
&\quad +o\left(n^{-1}\right),
\end{aligned} \tag{16}$$

where since

$$\begin{aligned}
\frac{\sum x_{t-i}^{1-2\rho}}{\sum x_{t-i}^{-2\rho}} &= \frac{\sum x_{t-i}x_{t-i}^{-2\rho}}{\sum x_{t-i}^{-2\rho}}=\frac{\sum E(x_{t-i})x_{t-i}^{-2\rho}}{\sum x_{t-i}^{-2\rho}}+\frac{\sum(x_{t-i}-E(x_{t-i}))x_{t-i}^{-2\rho}}{\sum x_{t-i}^{-2\rho}} \\
&= \frac{\sum \mu x_{t-i}^{-2\rho}}{\sum x_{t-i}^{-2\rho}}+\frac{\sum(x_{t-i}-E(x_{t-i}))x_{t-i}^{-2\rho}}{\sum x_{t-i}^{-2\rho}}=\mu\frac{\sum x_{t-i}^{-2\rho}}{\sum x_{t-i}^{-2\rho}}+\frac{\sum(x_{t-i}-E(x_{t-i}))x_{t-i}^{-2\rho}}{\sum x_{t-i}^{-2\rho}}=\beta_2+0_p(n^{-\frac{1}{2}}),
\end{aligned}$$

the first term in (16) can be cancelled up to the desired order with the fifth term, and the second term can be cancelled up to the desired order with the sixth term, leaving only the third and the fourth terms. Then the bias of $\hat{\beta}_2$ when β_1 is known, is the determining factor since

$$E\left(\widehat{\beta}_2-\beta_2\right)=E(\frac{\sum_{t=1}^{n} x_{t-1}^{-2p}\varepsilon_t}{(1-\beta_1)\sum_{t=1}^{n} x_{t-1}^{-2p}})+E(\frac{\left(\widehat{\beta}_1-\beta_1\right)\sum_{t=1}^{n} x_{t-1}^{-2p}\varepsilon_t}{(1-\beta_1)\sum_{t=1}^{n} x_{t-1}^{-2p}})+o\left(n^{-1}\right). \tag{17}$$

Recall that $E(\frac{\sum_{t=1}^{n} x_{t-1}^{-2p}\varepsilon_t}{(1-\beta_1)\sum_{t=1}^{n} x_{t-1}^{-2p}})$ is the bias in estimating β_2 when β_1 is known. The natural way to expand this is to set $E(\sum_{t=1}^{n} x_{t-1}^{-2p}) = n\mu^*$ and

$$\begin{aligned}
\frac{\sum_{t=1}^{n} x_{t-1}^{-2p}\varepsilon_t}{(1-\beta_1)\sum_{t=1}^{n} x_{t-1}^{-2p}} &= \frac{\sum_{t=1}^{n} x_{t-1}^{-2p}\varepsilon_t}{(1-\beta_1)n\mu^*}(1+\frac{\sum_{t=1}^{n} x_{t-1}^{-2p}-n\mu^*}{n\mu^*})^{-1} \\
&= \frac{\sum_{t=1}^{n} x_{t-1}^{-2p}\varepsilon_t}{(1-\beta_1)n\mu^*}(1-(\frac{\sum_{t=1}^{n} x_{t-1}^{-2p}-n\mu^*}{n\mu^*})+(\frac{\sum_{t=1}^{n} x_{t-1}^{-2p}-n\mu^*}{n\mu^*})^2-...) \\
&= \frac{\sum_{t=1}^{n} x_{t-1}^{-2p}\varepsilon_t}{(1-\beta_1)n\mu^*}(1-\frac{\sum_{t=1}^{n} x_{t-1}^{-2p}-n\mu^*}{n\mu^*})+o_p(n^{-1}),
\end{aligned}$$

and so the bias to order n^{-1} is obtained by taking the expectation. In effect we need to find

$$\begin{aligned}
& -E(\frac{\sum_{t=1}^{n} x_{t-1}^{-2p}\varepsilon_t}{(1-\beta_1)n\mu^*}\frac{\sum_{t=1}^{n} x_{t-1}^{-2p})}{n\mu^*}) \\
= & -E[(\frac{x_0^{-2\rho}\varepsilon_1+x_1^{-2\rho}\varepsilon_2+x_2^{-2\rho}\varepsilon_3+...+x_{n-1}^{-2\rho}\varepsilon_n}{(1-\beta_1)n^2(\mu^*)^2})(x_0^{-2\rho}+x_1^{-2\rho}+x_2^{-2\rho}+...+x_{n-1}^{-2\rho})].
\end{aligned}$$

Note that ε_t is correlated with $x_t^{-2\rho}$ and later values $x_{t+1}^{-2\rho}, ...$, but not with earlier values. Hence the above expectation requires the evaluation of the terms

$$\begin{aligned}
& E\left(x_0^{-2\rho}\varepsilon_1(x_1^{-2\rho}+x_2^{-2\rho}+...+x_{n-1}^{-2\rho})\right)+E\left(x_1^{-2\rho}\varepsilon_2(x_2^{-2\rho}+...+x_{n-1}^{-2\rho})\right) \\
& +E\left(x_2^{-2\rho}\varepsilon_3(x_3^{-2\rho}+...+x_{n-1}^{-2\rho})\right)+...+E\left(x_{n-2}^{-2\rho}\varepsilon_{n-1}(x_{n-1}^{-2\rho})\right).
\end{aligned}$$

The dependence between ε_t and $x_{t+i}^{-2\rho}$, $i = 1, 2, 3, ...$ decreases rapidly as is found in stationary $AR(1)$ models (see Kendall (1954)), so that each of these $(n-1)$ terms is $O(1)$ and their sum is $O(n)$. To see this, note that in the above there are only $(n-1)$ distinct terms of interest since, for example, $E(x_i^{-2\rho}\varepsilon_{i+1}x_{i+1}^{-2\rho})$ is the same for $i = 0, 1, .., n-2$, while $E(x_i^{-2\rho}\varepsilon_{i+1}x_{i+2}^{-2\rho})$ is the same for $i = 0, 1, .., n-3$,and so on. Hence to evaluate the above we shall need to find $(n-1)$ times $E(x_0^{-2\rho}\varepsilon_1 x_1^{-2\rho})$, $(n-2)$ times $E(x_0^{-2\rho}\varepsilon_1 x_2^{-2\rho})$, $(n-3)$ times $E(x_0^{-2\rho}\varepsilon_1 x_3^{-2\rho})$,..., and finally, $E(x_0^{-2\rho}\varepsilon_1 x_{n-1}^{-2\rho})$.

To find a suitable approximation to the expectations we have proceeded by first finding a Taylor series expansion for $x_i^{-2\rho}$, for $i = 1, ...n-1$ about $E(x_i) = \mu$, as follows

$$\begin{aligned}
x_i^{-2\rho} &= \frac{1}{\mu^{2\rho}} - 2\rho\mu^{-2\rho-1}(x_i-\mu) + 2\rho(2\rho+1)\mu^{-2\rho-2}\frac{(x_i-\mu)^2}{1.2} \\
& -2\rho(2\rho+1)(2\rho+1)\mu^{-2\rho-3}\frac{(x_i-\mu)^3}{1.2.3} + o(\mu^{-2\rho-3}) \\
&= \frac{1}{\mu^{2\rho}}[1-2\rho\frac{(x_i-\mu)}{\mu}+2\rho(2\rho+1)\frac{1}{\mu^2}\frac{(x_i-\mu)^2}{1.2}-2\rho(2\rho+1)(2\rho+1)\frac{1}{\mu^3}\frac{(x_i-\mu)^3}{1.2.3}+o(\frac{1}{\mu^{2\rho+3}}).
\end{aligned}$$

Then it is possible to find an approximation, for example, to $E(x_0^{-2\rho}\varepsilon_1 x_1^{-2\rho})$,by first replacing $x_1^{-2\rho}$ with the associated Taylor expansion and initially taking expectations conditional on x_0 before obtaining the unconditional expectation noting that the ε_t are assumed normal, uncorrelated, with conditional mean $E(\varepsilon_t|x_{t-1}) = 0$ and conditional variance $Var\left(\varepsilon_t|x_{t-1}\right) = 0.5\sigma^2\kappa^{-1}\left(1 - e^{-2\delta\kappa}\right)x_{t-1}^{2\rho}$.

The resulting analysis is complex but it is found that the bias in estimating β_2 when β_1 is known, given by $E(\frac{\sum_{t=1}^{n} x_{t-1}^{-2p}\varepsilon_t}{(1-\beta_1)\sum_{t=1}^{n} x_{t-1}^{-2p}})$, is well approximated by

$$\begin{aligned} & E(\frac{\sum_{t=1}^{n} x_{t-1}^{-2p}\varepsilon_t}{(1-\beta_1)\sum_{t=1}^{n} x_{t-1}^{-2p}}) \\ = \; & -\frac{1}{n}[\frac{2\rho(\frac{1}{2}\sigma^2\kappa^{-1}(1-e^{-2\kappa\delta}))}{(1-\beta_1)^2\mu^{(2\rho+3)}} + \frac{2\rho(2\rho+1)(2\rho+2)3(\frac{1}{2}\sigma^2\kappa^{-1}(1-e^{-2\kappa\delta}))^2}{6(1-\beta_1)^2\mu^{(2\rho+4)}(1-\beta_1^3)} \\ & + \frac{2\rho(2\rho+1)(2\rho+2)3\beta_1^2 E\left(x_t-\mu\right)^2(\frac{1}{2}\sigma^2\kappa^{-1}(1-e^{-2\kappa\delta}))}{6(1-\beta_1)\mu^{(2\rho+5)}(1-\beta_1^3)}], \end{aligned}$$

showing then that the bias of the long term mean parameter estimator is $O\left(n^{-1}\right)$. For the case $\rho = \frac{1}{2}$, the above reduces to

$$\begin{aligned} E(\frac{\sum_{t=1}^{n} x_{t-1}^{-1}\varepsilon_t}{(1-\beta_1)\sum_{t=1}^{n} x_{t-1}^{-1}}) \; = \; & -\frac{1}{n}[\frac{(\frac{1}{2}\sigma^2\kappa^{-1}(1-e^{-2\kappa\delta}))}{(1-\beta_1)^2\mu^4} + \frac{3(\frac{1}{2}\sigma^2\kappa^{-1}(1-e^{-2\kappa\delta}))^2}{(1-\beta_1^3)(1-\beta_1)^2\mu^5} \\ & + \frac{3\beta_1^2 E\left(x_t-\mu\right)^2(\frac{1}{2}\sigma^2\kappa^{-1}(1-e^{-2\kappa\delta}))}{(1-\beta_1)\mu^6(1-\beta_1^3)}]. \end{aligned}$$

■

4 APPENDIX D. Proof of Theorem 1 and Corollaries 1 and 2.

4.1 Proof of Theorem 1

$$LogL\left(\theta\right)=-\frac{1}{2}\sum_{t=1}^{n}\left(\ln\left(0,5\sigma^{2}\kappa^{-1}\left(1-e^{-2\delta\kappa}\right)x_{t-1}^{2\rho}\right)+\frac{\left(x_{t}-e^{-\delta\kappa}x_{t-1}-\mu\left(1-e^{-\delta\kappa}\right)\right)^{2}}{0,5\sigma^{2}\kappa^{-1}\left(1-e^{-2\delta\kappa}\right)x_{t-1}^{2\rho}}\right),$$

$$\frac{\partial LogL\left(\theta\right)}{\partial\kappa}=\sum_{t=1}^{n}\left(\left(2\kappa\right)^{-1}-\frac{\delta e^{-2\delta\kappa}}{\left(1-e^{-2\delta\kappa}\right)}-\frac{\left(x_{t}-e^{-\delta\kappa}x_{t-1}-\mu\left(1-e^{-\delta\kappa}\right)\right)\delta e^{-\delta\kappa}\left(x_{t-1}-\mu\right)}{0.5\sigma^{2}x_{t-1}^{2\rho}\kappa^{-1}\left(1-e^{-2\delta\kappa}\right)}\right)$$
$$+\sum_{t=1}^{n}\left(\frac{2\delta e^{-2\delta\kappa}\left(x_{t}-e^{-\delta\kappa}x_{t-1}-\mu\left(1-e^{-\delta\kappa}\right)\right)^{2}}{\sigma^{2}x_{t-1}^{2\rho}\kappa^{-1}\left(1-e^{-2\delta\kappa}\right)^{2}}-\frac{\left(x_{t}-e^{-\delta\kappa}x_{t-1}-\mu\left(1-e^{-\delta\kappa}\right)\right)^{2}}{\sigma^{2}x_{t-1}^{2\rho}\left(1-e^{-2\delta\kappa}\right)}\right),$$

$$\frac{\partial LogL\left(\theta\right)}{\partial\mu}=\sum_{t=1}^{n}\frac{\left(\left(x_{t}-e^{-\delta\kappa}x_{t-1}\right)-\mu\left(1-e^{-\delta\kappa}\right)\right)\left(1-e^{-\delta\kappa}\right)}{0,5\sigma^{2}\kappa^{-1}\left(1-e^{-2\delta\kappa}\right)x_{t-1}^{2\rho}},$$

$$\frac{\partial LogL\left(\theta\right)}{\partial\sigma^{2}}=-\frac{1}{2\sigma^{2}}\sum_{t=1}^{n}\left(1-\frac{\left(x_{t}-e^{-\delta\kappa}x_{t-1}-\mu\left(1-e^{-\delta\kappa}\right)\right)^{2}}{\sigma^{2}0,5\kappa^{-1}\left(1-e^{-2\delta\kappa}\right)x_{t-1}^{2\rho}}\right),$$

$$\frac{\partial^{2}LogL\left(\theta\right)}{\partial\kappa^{2}}=$$
$$\sum_{t=1}^{n}\left(\begin{array}{c}
-\frac{1}{2\kappa^{2}}-\frac{\left(1-e^{-2\delta\kappa}\right)\left[-2\delta^{2}e^{-2\delta\kappa}\right]-\delta e^{-2\delta\kappa}\left[2\delta e^{-2\delta\kappa}\right]}{\left(1-e^{-2\delta\kappa}\right)^{2}}\\
-\frac{\begin{array}{c}\left[\kappa^{-1}\left(1-e^{-2\delta\kappa}\right)\right]\left[e^{-\delta\kappa}\left(\delta e^{-\delta\kappa}x_{t-1}-\mu\delta e^{-\delta\kappa}\right)+\left(x_{t}-e^{-\delta\kappa}x_{t-1}-\mu\left(1-e^{-\delta\kappa}\right)\right)\delta e^{-\delta\kappa}\right]\\
-\left(x_{t}-e^{-\delta\kappa}x_{t-1}-\mu\left(1-e^{-\delta\kappa}\right)\right)e^{-\delta\kappa}\left[-\kappa^{-2}\left(1-e^{-2\delta\kappa}\right)+\kappa^{-1}2\delta e^{-2\delta\kappa}\right]\end{array}}{0.5\left(x_{t-1}-\mu\right)^{-1}\delta^{-1}\sigma^{2}x_{t-1}^{2\rho}\left[\kappa^{-2}\left(1-e^{-2\delta\kappa}\right)^{2}\right]}\\
\kappa^{-1}\left(1-e^{-2\delta\kappa}\right)^{2}\left[\begin{array}{c}-2\delta e^{-2\delta\kappa}\left(x_{t}-e^{-\delta\kappa}x_{t-1}-\mu\left(1-e^{-\delta\kappa}\right)\right)^{2}\\
+e^{-2\delta\kappa}2\left(x_{t}-e^{-\delta\kappa}x_{t-1}-\mu\left(1-e^{-\delta\kappa}\right)\right)\left(\delta e^{-\delta\kappa}x_{t-1}-\mu\delta e^{-\delta\kappa}\right)\end{array}\right]\\
-\left[e^{-2\delta\kappa}\left(x_{t}-e^{-\delta\kappa}x_{t-1}-\mu\left(1-e^{-\delta\kappa}\right)\right)^{2}\right]\\
+\frac{\left[-\kappa^{-2}\left(1-e^{-2\delta\kappa}\right)^{2}+\kappa^{-1}2\left(1-e^{-2\delta\kappa}\right)2\delta e^{-2\delta\kappa}\right]}{\left(2\delta\right)^{-1}\sigma^{2}x_{t-1}^{2\rho}\kappa^{-2}\left(1-e^{-2\delta\kappa}\right)^{4}}\\
-\frac{\left[1-e^{-2\delta\kappa}\right]\left[2\left(x_{t}-e^{-\delta\kappa}x_{t-1}-\mu\left(1-e^{-\delta\kappa}\right)\right)\left(\delta e^{-\delta\kappa}x_{t-1}-\mu\delta e^{-\delta\kappa}\right)\right]-\left(x_{t}-e^{-\delta\kappa}x_{t-1}-\mu\left(1-e^{-\delta\kappa}\right)\right)^{2}\left[2\delta e^{-2\delta\kappa}\right]}{\sigma^{2}x_{t-1}^{2\rho}\left(1-e^{-2\delta\kappa}\right)^{2}}
\end{array}\right).$$

$$\frac{\partial^2 LogL(\theta)}{\partial\mu^2} = \sum_{t=1}^{n} \frac{-\left(1-e^{-\delta\kappa}\right)^2}{0,5\sigma^2\kappa^{-1}\left(1-e^{-2\delta\kappa}\right)x_{t-1}^{2\rho}},$$

$$\frac{\partial^2 LogL(\theta)}{\partial\sigma^4} = \frac{n}{2\sigma^4} - \sum_{t=1}^{n} \frac{2\left(x_t - e^{-\delta\kappa}x_{t-1} - \mu\left(1-e^{-\delta\kappa}\right)\right)^2}{\sigma^6\kappa^{-1}\left(1-e^{-2\delta\kappa}\right)x_{t-1}^{2\rho}},$$

$$\frac{\partial^2 LogL(\theta)}{\partial\mu\partial\sigma^2} = \sum_{t=1}^{n} \frac{-\left(x_t - e^{-\delta\kappa}x_{t-1} - \mu\left(1-e^{-\delta\kappa}\right)\right)\left(1-e^{-\delta\kappa}\right)}{0,5\sigma^4\kappa^{-1}\left(1-e^{-2\delta\kappa}\right)x_{t-1}^{2\rho}},$$

$$\frac{\partial^2 LogL(\theta)}{\partial\kappa\partial\sigma^2} = \frac{1}{\sigma^2}\sum_{t=1}^{n}\left(\frac{\begin{array}{c}\left[\kappa^{-1}\left(1-e^{-2\delta\kappa}\right)\right]2\left(x_t - e^{-\delta\kappa}x_{t-1} - \mu\left(1-e^{-\delta\kappa}\right)\right)\left(\delta e^{-\delta\kappa}x_{t-1} - \mu\delta e^{-\delta\kappa}\right)\\ -\left(\left(x_t - e^{-\delta\kappa}x_{t-1} - \mu\left(1-e^{-\delta\kappa}\right)\right)^2\right)\left[\kappa^{-1}2\delta e^{-2\delta\kappa} - \kappa^{-2}\left(1-e^{-2\delta\kappa}\right)\right]\end{array}}{\sigma^2 x_{t-1}^{2\rho}\kappa^{-2}\left(1-e^{-2\delta\kappa}\right)^2}\right),$$

$$\frac{\partial^2 LogL(\theta)}{\partial\mu\partial\kappa} = \sum_{t=1}^{n}\frac{\begin{array}{c}\kappa^{-1}\left(1-e^{-2\delta\kappa}\right)\left[(x_t - e^{-\delta\kappa}x_{t-1} - \mu\left(1-e^{-\delta\kappa}\right))\delta e^{-\delta\kappa} + \left(1-e^{-\delta\kappa}\right)\left(\delta e^{-\delta\kappa}x_{t-1} - \delta\mu e^{-\delta\kappa}\right)\right]\\ +(x_t - e^{-\delta\kappa}x_{t-1} - \mu\left(1-e^{-\delta\kappa}\right))\left(1-e^{-\delta\kappa}\right)\left[\kappa^{-1}2\delta e^{-2\delta\kappa} - \kappa^{-2}\left(1-e^{-2\delta\kappa}\right)\right]\end{array}}{0,5\sigma^2 x_{t-1}^{2\rho}\kappa^{-2}\left(1-e^{-2\delta\kappa}\right)^2}.$$

We start to show each of the components as a function of ρ

$$E\left(-\frac{1}{n}\frac{\partial^2 LogL(\theta)}{\partial\kappa^2}\right) = \frac{\left(e^{-2\kappa\delta} + 2\kappa\delta e^{-2\kappa\delta} - 1\right)^2}{2\kappa^2\left(e^{-2\kappa\delta} - 1\right)^2} + \frac{2\delta^2 e^{-2\delta\kappa}}{\sigma^2\kappa^{-1}\left(1-e^{-2\delta\kappa}\right)}E\left(\left(x_{t-1}^{(1-\rho)} - \mu x_{t-1}^{-\rho}\right)^2\right).$$

$$E\left(-\frac{1}{n}\frac{\partial^2 LogL(\theta)}{\partial\mu^2}\right) = \sigma^{-2}2\kappa\left(e^{\delta\kappa} - 1\right)\left(e^{\delta\kappa} + 1\right)^{-1}E\left(\frac{1}{x_{t-1}^{2\rho}}\right).$$

$$E\left(-\frac{1}{n}\frac{\partial^2 LogL(\theta)}{\partial\sigma^4}\right) = -\left(\frac{1}{2\sigma^4} - \frac{1}{\sigma^4}\right) = \frac{1}{2\sigma^4}.$$

$$E\left(-\frac{1}{n}\frac{\partial^2 LogL(\theta)}{\partial\mu\partial\kappa}\right) = -\frac{2\kappa\delta e^{-\delta\kappa}}{\sigma^2\left(1+e^{-\delta\kappa}\right)}E\left(x_{t-1}^{(1-2\rho)} - \mu x_{t-1}^{-2\rho}\right).$$

$$E\left(-\frac{1}{n}\frac{\partial^2 LogL(\theta)}{\partial\mu\partial\sigma^2}\right) = E\left(\frac{\left(x_t - e^{-\delta\kappa}x_{t-1} - \mu\left(1-e^{-\delta\kappa}\right)\right)\left(1-e^{-\delta\kappa}\right)}{0,5\sigma^4\kappa^{-1}\left(1-e^{-2\delta\kappa}\right)x_{t-1}^{2\rho}}\right) = 0$$

due to the assumption that $E(\varepsilon_t|x_{t-1}) = 0$.

$$E\left(-\frac{1}{n}\frac{\partial^2 LogL(\theta)}{\partial\kappa\partial\sigma^2}\right) = \frac{\delta e^{-2\delta\kappa}}{\sigma^2\left(1-e^{-2\delta\kappa}\right)} - \frac{1}{2\kappa\sigma^2}.$$

So we get that, as a function of ρ

$$\Lambda = E\left(-\frac{1}{n}\frac{\partial^2 LogL(\theta)}{\partial\theta\partial\theta'}\right) = \left(\begin{array}{ccc} \overline{A} & \overline{B} & \overline{C} \\ \overline{B} & \overline{D} & 0 \\ \overline{C} & 0 & \overline{E} \end{array}\right),$$

with

$$\begin{aligned} \overline{A} &= \frac{\left(e^{-2\kappa\delta} + 2\kappa\delta e^{-2\kappa\delta} - 1\right)^2}{2\kappa^2\left(e^{-2\kappa\delta} - 1\right)^2} + \frac{2\delta^2 e^{-2\delta\kappa} E\left(\left(x_{t-1}^{(1-\rho)} - \mu x_{t-1}^{-\rho}\right)^2\right)}{\sigma^2\kappa^{-1}\left(1 - e^{-2\delta\kappa}\right)}; \ \overline{B} = -\frac{2\kappa\delta e^{-\delta\kappa} E\left(x_{t-1}^{(1-2\rho)} - \mu x_{t-1}^{-2\rho}\right)}{\sigma^2\left(1 + e^{-\delta\kappa}\right)}, \\ \overline{C} &= \frac{\delta e^{-2\delta\kappa}}{\sigma^2\left(1 - e^{-2\delta\kappa}\right)} - \frac{1}{2\kappa\sigma^2}; \ \overline{D} = \frac{2\kappa\left(e^{\delta\kappa} - 1\right)}{\sigma^2\left(e^{\delta\kappa} + 1\right)} E\left(\frac{1}{x_{t-1}^{2\rho}}\right); \ \overline{E} = \frac{1}{2\sigma^4}. \end{aligned}$$

We can show that $\Lambda = E\left(-\frac{1}{n}\frac{\partial^2 LogL(\theta)}{\partial\theta\partial\theta'}\right) > 0$ is positive definite since for any non-zero column vector z with entries a, b and c, we have $z'\Lambda z > 0$.■

4.2 PROOF of Corollary 1 when $\rho = 0$

Now, we check for $\rho = 0$

$$E\left(-\frac{1}{n}\frac{\partial^2 LogL(\theta)}{\partial\kappa^2}\right) = \frac{\left(e^{-2\kappa\delta} + 2\kappa\delta e^{-2\kappa\delta} - 1\right)^2 + 2\kappa^2\delta^2 e^{-2\delta\kappa}\left(1 - e^{-2\delta\kappa}\right)}{2\kappa^2\left(1 - e^{-2\kappa\delta}\right)^2}$$

since for $\rho = 0$

$$\begin{aligned} x_t - e^{-\delta\kappa}x_{t-1} - \mu\left(1 - e^{-\delta\kappa}\right) &= \varepsilon_t, \\ x_t - \mu &= e^{-\delta\kappa}\left(x_{t-1} - \mu\right) + \varepsilon_t, \\ E\left(x_t - \mu\right) &= e^{-\delta\kappa}E\left(x_{t-1} - \mu\right) + E\left(\varepsilon_t\right), \\ \left(x_t - \mu\right)^2 &= \left(e^{-\delta\kappa}\left(x_{t-1} - \mu\right) + \varepsilon_t\right)^2 \\ &= e^{-2\delta\kappa}\left(x_{t-1} - \mu\right)^2 + 2e^{-\delta\kappa}\left(x_{t-1} - \mu\right)\varepsilon_t + \varepsilon_t^2, \\ E\left(x_t - \mu\right)^2 &= e^{-2\delta\kappa}E\left(x_{t-1} - \mu\right)^2 + 2e^{-\delta\kappa}E\left(\left(x_{t-1} - \mu\right)\varepsilon_t\right) + E\left(\varepsilon_t^2\right), \\ \left(1 - e^{-2\delta\kappa}\right)E\left(x_t - \mu\right)^2 &= 0,5\sigma^2\kappa^{-1}\left(1 - e^{-2\delta\kappa}\right) \Longrightarrow E\left(x_t - \mu\right)^2 = 0,5\sigma^2\kappa^{-1}, \end{aligned}$$

from (2). Also

$$E\left(-\frac{1}{n}\frac{\partial^2 LogL(\theta)}{\partial\mu^2}\right) = \sigma^{-2}2\kappa\left(e^{\delta\kappa} - 1\right)\left(e^{\delta\kappa} + 1\right)^{-1},$$

Moreover

$$E\left(-\frac{1}{n}\frac{\partial^2 LogL(\theta)}{\partial\sigma^4}\right)=\frac{1}{2\sigma^4}.$$

and

$$E\left(-\frac{1}{n}\frac{\partial^2 LogL(\theta)}{\partial\mu\partial\kappa}\right)=0$$

since $E(x_t-\mu)=e^{-\delta\kappa}E(x_{t-1}-\mu)+E(\varepsilon_t)$ from (2); $E(x_t)=e^{-\delta\kappa}x_0+\mu\left(1-e^{-\delta\kappa}\right)$ and $E(x_t-\mu)=e^{-\delta\kappa}x_0-\mu e^{-\delta\kappa}=0$ since we assume that the initial condition $x_0=\mu$. Also

$$E\left(-\frac{1}{n}\frac{\partial^2 LogL(\theta)}{\partial\kappa\partial\sigma^2}\right)=\frac{\delta e^{-2\delta\kappa}}{\sigma^2\left(1-e^{-2\delta\kappa}\right)}-\frac{1}{2\kappa\sigma^2}.$$

Finally

$$E\left(-\frac{1}{n}\frac{\partial^2 LogL(\theta)}{\partial\mu\partial\sigma^2}\right)=0$$

Finding now the inverse of

$$\begin{pmatrix} \frac{\left(e^{-2\kappa\delta}+2\kappa\delta e^{-2\kappa\delta}-1\right)^2+2\kappa^2\delta^2e^{-2\delta\kappa}\left(1-e^{-2\delta\kappa}\right)}{2\kappa^2\left(1-e^{-2\kappa\delta}\right)^2} & 0 & \frac{\delta e^{-2\delta\kappa}}{\sigma^2\left(1-e^{-2\delta\kappa}\right)}-\frac{1}{2\kappa\sigma^2} \\ 0 & \frac{2\kappa\left(e^{\delta\kappa}-1\right)}{\sigma^2\left(e^{\delta\kappa}+1\right)} & 0 \\ \frac{\delta e^{-2\delta\kappa}}{\sigma^2\left(1-e^{-2\delta\kappa}\right)}-\frac{1}{2\kappa\sigma^2} & 0 & \frac{1}{2\sigma^4} \end{pmatrix}$$

we obtain Ω_1

$$\Omega_1=\begin{pmatrix} \delta^{-2}\left(e^{2\kappa\delta}-1\right) & 0 & -\frac{\sigma^2\left(1-e^{2\kappa\delta}+2\kappa\delta\right)}{\kappa\delta^2} \\ 0 & \frac{\sigma^2\left(1+e^{\kappa\delta}\right)}{2\kappa\left(e^{\delta\kappa}-1\right)} & 0 \\ -\frac{\sigma^2\left(1-e^{2\kappa\delta}+2\kappa\delta\right)}{\kappa\delta^2} & 0 & \frac{\sigma^4\left(2\kappa^2\delta^2\left(1+e^{2\kappa\delta}\right)+4\kappa\delta\left(1-e^{2\kappa\delta}\right)+e^{4\kappa\delta}-2e^{2\kappa\delta}+1\right)}{\delta^2\kappa^2\left(e^{2\kappa\delta}-1\right)} \end{pmatrix}.$$

■

4.3 STATEMENT AND PROOF of Corollary 2

Corollary 2. *For a stationary Vasicek (1977) process, as* $n\longrightarrow\infty$, $\delta\to 0, T=n\delta\to\infty$, for some $k>2$, $T\delta^{\frac{1}{k}}\to\infty$ *and for* $\rho=0$*, let* $\widehat{\theta}=(\hat{\kappa},\widehat{\mu},\widehat{\sigma}^2)'$*, and* $\widetilde{\theta}=(\kappa,\mu,\sigma^2)'$*, then*

$$R_{n,\delta}\left(\widehat{\theta}-\widetilde{\theta}\right)\xrightarrow{d}N(0,\Omega_2)\ where$$

$$R_{n,\delta}=diag\left(\sqrt{T},\sqrt{T},\sqrt{n}\right),\Omega_2=diag\left\{2\kappa,\sigma^2\kappa^{-2},2\sigma^4\right\}.$$

Proof of Corollary 2. For $\rho = 0$ and when $\delta \to 0$, applying l'Hôpital´s rule, $n\delta = T$, and pre-multiplying by δ (justifying the lower convergence rate because otherwise the term will explode)

$$E\left(-\frac{1}{T}\frac{\partial^2 LogL(\theta)}{\partial\kappa^2}\right)^{-1}_{when\delta\longrightarrow 0} = 2\kappa.$$

Also, applying l'Hôpital´s rule and pre-multiplying by δ

$$\left(\frac{\sigma^2\delta\left(e^{\kappa\delta}+1\right)}{2\kappa\left(e^{\kappa\delta}-1\right)}\right)_{when\delta\longrightarrow 0} = \sigma^2\kappa^{-2}.$$

Moreover, applying l'Hôpital´s rule three times, we find that we do not need to premultiply by δ the last term of the diagonal

$$\left(\frac{\sigma^4\left(2\kappa^2\delta^2\left(1+e^{2\kappa\delta}\right)+4\kappa\delta\left(1-e^{2\kappa\delta}\right)+e^{4\kappa\delta}-2e^{2\kappa\delta}+1\right)}{\delta^2\kappa^2\left(e^{2\kappa\delta}-1\right)}\right)_{when\delta\longrightarrow 0} = 2\sigma^4.$$

and all the off-diagonal terms are zero since for a fixed δ, the only one that was different from zero in Corollary 1, when $\delta \to 0$

$$\left(-\frac{\sigma^2\left(1-e^{2\kappa\delta}+2\kappa\delta\right)}{\kappa\delta^2}\right)_{when\delta\longrightarrow 0} = 0.$$

■

5 APPENDIX E. Expansions for the bias parameter estimators in the general CKLS model when $n \to \infty$ and δ is fixed. PMLEs interpreted as Instrumental Variable (IV) estimators

5.1 The estimator of β_1

5.1.1 For $\rho = 0$

From (2), we can commence from

$$x_t = \beta_1 x_{t-1} + \mu\left(1-\beta_1\right) + \varepsilon_t \tag{18}$$

Then the regression of x_t on x_{t-1} yields the Vasicek (1977) estimator of β_1

$$\widehat{\beta}_1 = \frac{\sum_{t=1}^n x_t x_{t-1} - n\overline{x}_t\overline{x}_{t-1}}{\sum_{t=1}^n x_{t-1}^2 - n\left(\overline{x}_{t-1}\right)^2} = \beta_1 + \frac{\sum_{t=1}^n \left(x_{t-1}-\overline{x}_{t-1}\right)\varepsilon_t}{\sum_{t=1}^n \left(x_{t-1}-\overline{x}_{t-1}\right)^2},$$

where $\sum_{t=1}^{n}(x_t/n) = \overline{x}_t$ is an unbiased estimator for μ. Then $\widehat{\beta}_1$ is biased as would be expected since the numerator is correlated with the denominator and the bias is of order n^{-1}. In fact, the expression (2.6) in Tang and Chen (2009, page 67) indicates that this is true. The estimator can be interpreted as an IV estimator where the instrumental variable is $(x_{t-1} - \overline{x}_{t-1})$. Remember the results from Proposition 1 (see (4) for $\rho = 0$).

5.1.2 For $\rho = 0.5$

The Cox, Ingersoll and Ross (1985) (CIR) estimator is also an IV estimator with instrumental variable $\left(x_{t-1}^{-1} - \overline{x_{t-1}^{-1}}\right)$. To see this, write (18) as

$$\sum_{t=1}^{n} x_t \left(x_{t-1}^{-1} - \overline{x_{t-1}^{-1}}\right) = \beta_1 \sum_{t=1}^{n} x_{t-1} \left(x_{t-1}^{-1} - \overline{x_{t-1}^{-1}}\right) + \mu(1-\beta_1) \sum_{t=1}^{n} \left(x_{t-1}^{-1} - \overline{x_{t-1}^{-1}}\right) + \sum_{t=1}^{n} \left(x_{t-1}^{-1} - \overline{x_{t-1}^{-1}}\right) \varepsilon_t,$$

where, since $\sum_{t=1}^{n} \left(x_{t-1}^{-1} - \overline{x_{t-1}^{-1}}\right) = 0$,

$$\sum_{t=1}^{n} x_t \left(x_{t-1}^{-1} - \overline{x_{t-1}^{-1}}\right) = \beta_1 \sum_{t=1}^{n} x_{t-1} \left(x_{t-1}^{-1} - \overline{x_{t-1}^{-1}}\right) + \sum_{t=1}^{n} \left(x_{t-1}^{-1} - \overline{x_{t-1}^{-1}}\right) \varepsilon_t$$

from which

$$\beta_1 = \frac{\sum_{t=1}^{n} x_t x_{t-1}^{-1} - n\overline{x}_t \overline{x_{t-1}^{-1}}}{\sum_{t=1}^{n} x_{t-1} \left(x_{t-1}^{-1} - \overline{x_{t-1}^{-1}}\right)} - \frac{\sum_{t=1}^{n} \left(x_{t-1}^{-1} - \overline{x_{t-1}^{-1}}\right) \varepsilon_t}{\sum_{t=1}^{n} x_{t-1} \left(x_{t-1}^{-1} - \overline{x_{t-1}^{-1}}\right)}$$

and

$$\hat{\beta}_1 = \frac{\sum_{t=1}^{n} x_t x_{t-1}^{-1} - n\overline{x}_t \overline{x_{t-1}^{-1}}}{\sum_{t=1}^{n} x_{t-1} \left(x_{t-1}^{-1} - \overline{x_{t-1}^{-1}}\right)} = \frac{\sum_{t=1}^{n} x_t \sum_{t=1}^{n} x_{t-1}^{-1} - n \sum_{t=1}^{n} x_t x_{t-1}^{-1}}{\sum_{t=1}^{n} x_{t-1} \sum_{t=1}^{n} x_{t-1}^{-1} - n^2},$$

which is the result in (4) for $\rho = 0.5$. Note that

$$\widehat{\beta}_1 = \beta_1 + \frac{\sum_{t=1}^{n} \left(x_{t-1}^{-1} - \overline{x_{t-1}^{-1}}\right) \varepsilon_t}{\sum_{t=1}^{n} x_{t-1} \left(x_{t-1}^{-1} - \overline{x_{t-1}^{-1}}\right)},$$

and also that

$$(1 - \bar{x}_{t-1}\overline{x_{t-1}^{-1}}) = E(1 - \bar{x}_{t-1}\overline{x_{t-1}^{-1}})[1 + \frac{((1 - \bar{x}_{t-1}\overline{x_{t-1}^{-1}}) - E(1 - \bar{x}_{t-1}\overline{x_{t-1}^{-1}}))}{E(1 - \bar{x}_{t-1}\overline{x_{t-1}^{-1}})}].$$

Then

$$\frac{\sum_{t=1}^{n}\left(x_{t-1}^{-1}-\overline{x_{t-1}^{-1}}\right)\varepsilon_t}{n(1-\bar{x}_{t-1}\overline{x_{t-1}^{-1}})}=\frac{\sum_{t=1}^{n}\left(x_{t-1}^{-1}-\overline{x_{t-1}^{-1}}\right)\varepsilon_t}{nE(1-\bar{x}_{t-1}\overline{x_{t-1}^{-1}})}[1-\frac{((\bar{x}_{t-1}\overline{x_{t-1}^{-1}})-E(\bar{x}_{t-1}\overline{x_{t-1}^{-1}}))}{E(1-\bar{x}_{t-1}\overline{x_{t-1}^{-1}})}]^{-1}$$

$$=\frac{\sum_{t=1}^{n}\left(x_{t-1}^{-1}-\overline{x_{t-1}^{-1}}\right)\varepsilon_t}{nE(1-\bar{x}_{t-1}\overline{x_{t-1}^{-1}})}[1+\frac{((\bar{x}_{t-1}\overline{x_{t-1}^{-1}})-E(\bar{x}_{t-1}\overline{x_{t-1}^{-1}}))}{E(1-\bar{x}_{t-1}\overline{x_{t-1}^{-1}})}+(\frac{((\bar{x}_{t-1}\overline{x_{t-1}^{-1}})-E(\bar{x}_{t-1}\overline{x_{t-1}^{-1}}))}{E(1-\bar{x}_{t-1}\overline{x_{t-1}^{-1}})})^2+...].$$

The first of these terms is

$$\frac{\sum_{t=1}^{n}\left(x_{t-1}^{-1}-\overline{x_{t-1}^{-1}}\right)\varepsilon_t}{nE(1-\bar{x}_{t-1}\overline{x_{t-1}^{-1}})}=\frac{\sum_{t=1}^{n}x_{t-1}^{-1}\varepsilon_t}{nE(1-\bar{x}_{t-1}\overline{x_{t-1}^{-1}})}-\frac{\overline{x_{t-1}^{-1}}\sum_{t=1}^{n}\varepsilon_t}{nE(1-\bar{x}_{t-1}\overline{x_{t-1}^{-1}})}$$

where the leading term has expectation zero and can be ignored. The first bias term is

$$\begin{aligned}-E[\frac{\overline{x_{t-1}^{-1}}\sum_{t=1}^{n}\varepsilon_t}{nE(1-\bar{x}_{t-1}\overline{x_{t-1}^{-1}})}] &= -E[\frac{\frac{1}{n}\sum_{t=1}^{n}x_{t-1}^{-1}\sum_{t=1}^{n}\varepsilon_t}{nE(1-\bar{x}_{t-1}\overline{x_{t-1}^{-1}})}]=-\frac{1}{n}\frac{E(\sum_{t=1}^{n}x_{t-1}^{-1}\sum_{t=1}^{n}\varepsilon_t)}{nE(1-\bar{x}_{t-1}\overline{x_{t-1}^{-1}})}\\ &= -\frac{1}{n}[\frac{E(\varepsilon_1(x_0^{-1}+x_1^{-1}+x_2^{-1}+...+x_{n-1}^{-1})}{nE(1-\bar{x}_{t-1}\overline{x_{t-1}^{-1}})}+\frac{E(\varepsilon_2(x_0^{-1}+x_1^{-1}+x_2^{-1}+...+x_{n-1}^{-1})}{nE(1-\bar{x}_{t-1}\overline{x_{t-1}^{-1}})}\\ &\quad +...+\frac{E(\varepsilon_n(x_0^{-1}+x_1^{-1}+x_2^{-1}+.........+x_{n-1}^{-1})}{nE(1-\bar{x}_{t-1}\overline{x_{t-1}^{-1}})}]\end{aligned}$$

where ε_j is correlated with $x_j^{-1},x_{j+1}^{-1},...,x_{n-1}^{-1}$ but is uncorrelated with $x_{j-i}^{-1}, i=1,2,...$ Hence we need to evaluate $n-1$ terms as follows

$$-E[\frac{\frac{1}{n}\sum_{t=1}^{n}x_{t-1}^{-1}\sum_{t=1}^{n}\varepsilon_t}{nE(1-\bar{x}_{t-1}\overline{x_{t-1}^{-1}})}]=-\frac{1}{n}[\frac{E(\varepsilon_1(x_1^{-1}+x_2^{-1}+...+x_{n-1}^{-1})}{nE(1-\bar{x}_{t-1}\overline{x_{t-1}^{-1}})}+\frac{E(\varepsilon_2(x_2^{-1}+...+x_{n-1}^{-1})}{nE(1-\bar{x}_{t-1}\overline{x_{t-1}^{-1}})}+...+\frac{E(\varepsilon_{n-1}x_{n-1}^{-1})}{nE(1-\bar{x}_{t-1}\overline{x_{t-1}^{-1}})}]. \tag{19}$$

The correlations between ε_t and future values x_{t+j}^{-1}, $j=1,2,...$, will quickly diminish so that all the expectations in (19) are $O(1)$ and their sum will be $O(n)$. Finally we may conclude that the leading bias term

$$-E[\frac{\frac{1}{n}\sum_{t=1}^{n}x_{t-1}^{-1}\sum_{t=1}^{n}\varepsilon_t}{nE(1-\bar{x}_{t-1}\overline{x_{t-1}^{-1}})}]$$

is $O(n^{-1})$.

The above analysis has much in common with the corresponding result for the least squares estimator of the slope coefficient in stationary AR(1) model where the bias is well known to be of order n^{-1} (see Kendall (1954)). Therefore we conclude that the bias of $\hat{\beta}_1$ is of $O\left(n^{-1}\right)$. This result agrees with the bias of $\widehat{\beta}_1$ given in expression (A.12) in Tang and Chen (2009, page 77).

5.2 The estimator of β_2

5.2.1 For $\rho = 0$

In the Vasicek (1977) case, we commence again from (18). Simple rearrangement yields

$$\frac{\sum_{t=1}^{n}(x_t - \beta_1 x_{t-1})}{n(1-\beta_1)} = \beta_2 + \frac{\sum_{t=1}^{n}\varepsilon_t}{n(1-\beta_1)} \tag{20}$$

If β_1 is known then

$$\hat{\beta}_2^* = \frac{\sum_{t=1}^{n}(x_t - \beta_1 x_{t-1})}{n(1-\beta_1)} \tag{21}$$

yields and unbiased estimator since

$$\hat{\beta}_2^* = \beta_2 + \frac{\sum_{t=1}^{n}\varepsilon_t}{n(1-\beta_1)} \tag{22}$$

and $E\left(\sum_{t=1}^{n}\varepsilon_t\right) = 0$. However, if β_1 is not known, we can replace it with $\widehat{\beta}_1$ which is biased to order n^{-1} so that now the PMLE for β_2 is

$$\widehat{\beta}_2 = \frac{\sum_{t=1}^{n}\left(x_t - \widehat{\beta}_1 x_{t-1}\right)}{n\left(1-\hat{\beta}_1\right)} \tag{23}$$

It is now straightforward to show that this estimator is unbiased to order n^{-1} supporting Theorem 3.1.1 in Tang and Chen (2009, page 68).

5.2.2 For $\rho = 0.5$

For the CIR process multiply (18) by x_{t-1}^{-1} so that

$$\sum_{t=1}^{n} x_t x_{t-1}^{-1} = \beta_1 \sum_{t=1}^{n} x_{t-1}x_{t-1}^{-1} + \beta_2(1-\beta_1)\sum_{t=1}^{n} x_{t-1}^{-1} + \sum_{t=1}^{n} x_{t-1}^{-1}\varepsilon_t = n\beta_1 + \beta_2(1-\beta_1)\sum_{t=1}^{n} x_{t-1}^{-1} + \sum_{t=1}^{n} x_{t-1}^{-1}\varepsilon_t$$

and

$$\frac{\sum_{t=1}^{n} x_t x_{t-1}^{-1} - n\beta_1}{(1-\beta_1)\sum_{t=1}^{n} x_{t-1}^{-1}} = \beta_2 + \frac{\sum_{t=1}^{n} x_{t-1}^{-1}\varepsilon_t}{(1-\beta_1)\sum_{t=1}^{n} x_{t-1}^{-1}}.$$

If β_1 is known then

$$\frac{\sum_{t=1}^{n} x_t x_{t-1}^{-1} - n\beta_1}{(1-\beta_1)\sum_{t=1}^{n} x_{t-1}^{-1}} = \hat{\beta}_2^*.$$

In this case

$$E\left(\hat{\beta}_2^* - \beta_2\right) = E\left(\frac{\sum_{t=1}^{n} x_{t-1}^{-1}\varepsilon_t}{(1-\beta_1)\sum_{t=1}^{n} x_{t-1}^{-1}}\right)$$

which is $O\left(n^{-1}\right)$ and so $\hat{\beta}_2^*$ has a bias of order n^{-1} since numerator and denominator are correlated. However since β_1 is unknown it can be replaced with a PML estimate $\widehat{\beta}_1$ and then an operational estimator is

$$\widehat{\beta}_2 = \frac{\sum_{t=1}^{n} x_t x_{t-1}^{-1} - n\widehat{\beta}_1}{\left(1-\widehat{\beta}_1\right)\sum_{t=1}^{n} x_{t-1}^{-1}}$$

Since the estimator has a bias of $O\left(n^{-1}\right)$ when β_1 is known, it will have a bias at least as large when β_1 is replaced by $\widehat{\beta}_1$. However Tang and Chen (2009, Theorem 3.1.3) state that $E\left(\widehat{\beta}_2 - \beta_2\right) = O\left(n^{-2}\right)$ and our analysis indicates that this is not correct. More specifically, write

$$\begin{aligned}
\widehat{\beta}_2 &= \left(\frac{n^{-1}\sum_{t=1}^{n} x_t x_{t-1}^{-1} - \beta_1}{(1-\beta_1)\, n^{-1}\sum_{t=1}^{n} x_{t-1}^{-1}} - \frac{\left(\widehat{\beta}_1-\beta_1\right)}{(1-\beta_1)\, n^{-1}\sum_{t=1}^{n} x_{t-1}^{-1}}\right)\left(1+\frac{\left(\widehat{\beta}_1-\beta_1\right)}{(1-\beta_1)}+\frac{\left(\widehat{\beta}_1-\beta_1\right)^2}{(1-\beta_1)^2}+...\right)\\
&= \left(\beta_2 + \frac{\sum_{t=1}^{n} x_{t-1}^{-1}\varepsilon_t}{(1-\beta_1)\sum_{t=1}^{n} x_{t-1}^{-1}} - \frac{\left(\widehat{\beta}_1-\beta_1\right)}{(1-\beta_1)\, n^{-1}\sum_{t=1}^{n} x_{t-1}^{-1}}\right)\left(1+\frac{\left(\widehat{\beta}_1-\beta_1\right)}{(1-\beta_1)}+\frac{\left(\widehat{\beta}_1-\beta_1\right)^2}{(1-\beta_1)^2}+...\right)\\
&= \beta_2 + \frac{\beta_2\left(\widehat{\beta}_1-\beta_1\right)}{(1-\beta_1)} + \frac{\beta_2\left(\widehat{\beta}_1-\beta_1\right)^2}{(1-\beta_1)^2} + ... + \frac{\sum_{t=1}^{n} x_{t-1}^{-1}\varepsilon_t}{(1-\beta_1)\sum_{t=1}^{n} x_{t-1}^{-1}} + \frac{\left(\widehat{\beta}_1-\beta_1\right)\sum_{t=1}^{n} x_{t-1}^{-1}\varepsilon_t}{(1-\beta_1)^2\sum_{t=1}^{n} x_{t-1}^{-1}}\\
&\quad - \frac{\left(\widehat{\beta}_1-\beta_1\right)}{(1-\beta_1)\, n^{-1}\sum_{t=1}^{n} x_{t-1}^{-1}} - \frac{\left(\widehat{\beta}_1-\beta_1\right)^2}{(1-\beta_1)^2\, n^{-1}\sum_{t=1}^{n} x_{t-1}^{-1}} + o_p\left(n^{-1}\right).
\end{aligned}$$

Then

$$\begin{aligned}
E\left(\widehat{\beta}_2-\beta_2\right) &= \frac{\beta_2 E\left(\widehat{\beta}_1-\beta_1\right)}{(1-\beta_1)} + \frac{\beta_2 E\left(\widehat{\beta}_1-\beta_1\right)^2}{(1-\beta_1)^2} + ...\\
&\quad + E\left(\frac{\sum_{t=1}^{n} x_{t-1}^{-1}\varepsilon_t}{(1-\beta_1)\sum_{t=1}^{n} x_{t-1}^{-1}}\right) + E\left(\frac{\left(\widehat{\beta}_1-\beta_1\right)\sum_{t=1}^{n} x_{t-1}^{-1}\varepsilon_t}{(1-\beta_1)^2\sum_{t=1}^{n} x_{t-1}^{-1}}\right)\\
&\quad - E\left(\frac{\left(\widehat{\beta}_1-\beta_1\right)}{(1-\beta_1)\, n^{-1}\sum_{t=1}^{n} x_{t-1}^{-1}}\right) - E\left(\frac{\left(\widehat{\beta}_1-\beta_1\right)^2}{(1-\beta_1)^2\, n^{-1}\sum_{t=1}^{n} x_{t-1}^{-1}}\right) + o\left(n^{-1}\right).
\end{aligned}$$

which contains several terms of order n^{-1}. Therefore our analysis indicates that the result in Tang and Chen (2009, Theorem 3.1.3), where it is stated that $E\left(\widehat{\beta}_2-\beta_2\right) = O\left(n^{-2}\right)$, is incorrect for $\rho = 0.5$.

6 APPENDIX F. Proofs of the results of the IV approach

6.1 Bias expansion of $\widehat{\beta}_1$. Using an IV approach

Now we proceed to analyze the bias $\widehat{\beta}_1$ of the case of non-negative integer ρ by using an IV approach to show things in a very clear way. Commencing from (18)

$$x_t = \beta_1 x_{t-1} + \mu(1-\beta_1) + \varepsilon_t, \tag{24}$$

since the estimator of β_1 will depend on β_2, we start from

$$x_t - \beta_1 x_{t-1} = (1-\beta_1)\beta_2 + \varepsilon_t$$

i.e.

$$\frac{x_t - \beta_1 x_{t-1}}{1-\beta_1} = \beta_2 + \frac{\varepsilon_t}{1-\beta_1}$$

and multiply by x_{t-1}^{-2p} and sum over n to yield

$$\frac{\sum_{t=1}^n x_t x_{t-1}^{-2p} - \beta_1 \sum_{t=1}^n x_{t-1}^{1-2p}}{1-\beta_1} = \beta_2 \sum_{t=1}^n x_{t-1}^{-2p} + \frac{\sum_{t=1}^n x_{t-1}^{-2p}\varepsilon_t}{1-\beta_1}$$

and solve for β_2 to yield

$$\beta_2 = \frac{\sum_{t=1}^n x_t x_{t-1}^{-2p} - \beta_1 \sum_{t=1}^n x_{t-1}^{1-2p}}{(1-\beta_1)\sum_{t=1}^n x_{t-1}^{-2p}} - \frac{\sum_{t=1}^n x_{t-1}^{-2p}\varepsilon_t}{(1-\beta_1)\sum_{t=1}^n x_{t-1}^{-2p}} \tag{25}$$

from (5). Now write

$$x_t - \beta_1 x_{t-1} = \frac{\sum_{t=1}^n x_t x_{t-1}^{-2p} - \beta_1 \sum_{t=1}^n x_{t-1}^{1-2p}}{\sum_{t=1}^n x_{t-1}^{-2p}} - \frac{\sum_{t=1}^n x_{t-1}^{-2p}\varepsilon_t}{\sum_{t=1}^n x_{t-1}^{-2p}} + \varepsilon_t$$

Now use $x_{t-1}^{(1-2\rho)}$ as an IV variable and sum over n, so that

$$\begin{aligned}\sum_{t=1}^n x_t x_{t-1}^{(1-2\rho)} - \beta_1 \sum_{t=1}^n x_{t-1}^{(2-2\rho)} &= \sum_{t=1}^n x_{t-1}^{1-2p}\left[\frac{\sum_{t=1}^n x_t x_{t-1}^{-2p} - \beta_1 \sum_{t=1}^n x_{t-1}^{1-2p}}{\sum_{t=1}^n x_{t-1}^{-2p}}\right] \\ &\quad - \frac{\sum_{t=1}^n x_{t-1}^{1-2p}}{\sum_{t=1}^n x_{t-1}^{-2p}} \sum_{t=1}^n x_{t-1}^{-2p}\varepsilon_t + \sum_{t=1}^n x_{t-1}^{1-2p}\varepsilon_t\end{aligned}$$

and then

$$\begin{aligned}\beta_1 &= \frac{\sum_{t=1}^n x_t x_{t-1}^{(1-2\rho)} \sum_{t=1}^n x_{t-1}^{-2\rho} - \sum_{t=1}^n x_t x_{t-1}^{-2\rho} \sum_{t=1}^n x_{t-1}^{(1-2\rho)}}{\sum_{t=1}^n x_{t-1}^{(2-2\rho)} \sum_{t=1}^n x_{t-1}^{-2\rho} - \left(\sum_{t=1}^n x_{t-1}^{(1-2\rho)}\right)^2} \\ &\quad + \frac{\sum_{t=1}^n x_{t-1}^{(1-2p)} \sum_{t=1}^n x_{t-1}^{-2p}\varepsilon_t - \sum_{t=1}^n x_{t-1}^{-2\rho} \sum_{t=1}^n x_{t-1}^{(1-2\rho)}\varepsilon_t}{\sum_{t=1}^n x_{t-1}^{(2-2\rho)} \sum_{t=1}^n x_{t-1}^{-2\rho} - \left(\sum_{t=1}^n x_{t-1}^{(1-2\rho)}\right)^2}\end{aligned}$$

Finally

$$\hat{\beta}_1 = \frac{\sum_{t=1}^n x_t x_{t-1}^{(1-2\rho)} \sum_{t=1}^n x_{t-1}^{-2\rho} - \sum_{t=1}^n x_t x_{t-1}^{-2\rho} \sum_{t=1}^n x_{t-1}^{(1-2\rho)}}{\sum_{t=1}^n x_{t-1}^{(2-2\rho)} \sum_{t=1}^n x_{t-1}^{-2\rho} - \left(\sum_{t=1}^n x_{t-1}^{(1-2\rho)}\right)^2}$$

i.e.

$$\widehat{\beta}_1 = \beta_1 - \frac{\sum_{t=1}^n x_{t-1}^{1-2p} \sum_{t=1}^n x_{t-1}^{-2p}\varepsilon_t - \sum_{t=1}^n x_{t-1}^{-2\rho} \sum_{t=1}^n x_{t-1}^{(1-2\rho)}\varepsilon_t}{\sum_{t=1}^n x_{t-1}^{(2-2\rho)} \sum_{t=1}^n x_{t-1}^{-2\rho} - \left(\sum_{t=1}^n x_{t-1}^{(1-2\rho)}\right)^2} \tag{26}$$

where numerator and denominator are correlated so that the bias should be $O\left(n^{-1}\right)$ for any non-negative integer ρ .

6.2 Bias expansion of $\widehat{\mu}$. Using an IV approach

For $\rho = 0$, from (23), it is straightforward to show that this estimator is unbiased to order n^{-1} supporting Theorem 3.1.1 in Tang and Chen (2009, page 68). We proceed now to expand $\widehat{\beta}_2$ for any ρ being a positive integer and also when $\rho = 0.5$

$$\widehat{\beta}_2 = \frac{\sum_{t=1}^n \left(x_t x_{t-1}^{-2\rho} - \widehat{\beta}_1 x_{t-1}^{(1-2\rho)}\right)}{\left(1-\widehat{\beta}_1\right)\sum_{t=1}^n x_{t-1}^{-2\rho}} = \frac{\sum_{t=1}^n \left(x_t x_{t-1}^{-2\rho^o} - \widehat{\beta}_1 x_{t-1}^{(1-2\rho)}\right)\left(1-\widehat{\beta}_1\right)^{-1}}{E\left(\sum_{t=1}^n x_{t-1}^{-2\rho}\right)\left[1 + \frac{\sum_{t=1}^n x_{t-1}^{-2\rho} - E\left(\sum_{t=1}^n x_{t-1}^{-2\rho}\right)}{E\left(\sum_{t=1}^n x_{t-1}^{-2\rho}\right)}\right]},$$

where $\frac{\sum_{t=1}^n x_{t-1}^{-2\rho} - E\left(\sum_{t=1}^n x_{t-1}^{-2\rho}\right)}{E\left(\sum_{t=1}^n x_{t-1}^{-2\rho}\right)}$ is $O_p\left(n^{-1/2}\right)$. Hence

$$\widehat{\beta}_2 = \frac{\sum_{t=1}^n \left(x_t x_{t-1}^{-2\rho} - \widehat{\beta}_1 x_{t-1}^{(1-2\rho)}\right)\left(1-\widehat{\beta}_1\right)^{-1}\left[1 + \frac{\sum_{t=1}^n x_{t-1}^{-2\rho} - E\left(\sum_{t=1}^n x_{t-1}^{-2\rho}\right)}{E\left(\sum_{t=1}^n x_{t-1}^{-2\rho}\right)}\right]^{-1}}{E\left(\sum_{t=1}^n x_{t-1}^{-2\rho}\right)}$$

where we may find an asymptotic expansion. Notice that $E\left(\sum_{t=1}^n x_{t-1}^{-2\rho}\right) = n\mu_2$ (so $\left(\sum_{t=1}^n x_{t-1}^{-2\rho}\right)/n$ is an unbiased estimator for μ_2) and $\left(1-\widehat{\beta}_1\right)^{-1} = (1-\beta_1)^{-1}\left(1 - \left(\left(\widehat{\beta}_1 - \beta_1\right)/(1-\beta_1)\right)^{-1}\right)$, which can be easily expanded noting that $\left(\widehat{\beta}_1 - \beta_1\right)$ is $O_p\left(n^{-1/2}\right)$.

Therefore for $\beta_1 \neq 1$

$$\begin{aligned}
\widehat{\beta}_2 &= \frac{\sum_{t=1}^{n}\left(x_t x_{t-1}^{-2\rho} - \beta_1 x_{t-1}^{(1-2\rho)}\right)(1-\beta_1)^{-1}\left(1 - \left(\left(\widehat{\beta}_1 - \beta_1\right)/(1-\beta_1)\right)^{-1}\right)\left[1 + \frac{\sum_{t=1}^{n} x_{t-1}^{-2\rho} - n\mu_2}{n\mu_2}\right]^{-1}}{n\mu_2} \\
&= (1-\beta_1)^{-1}\frac{\sum_{t=1}^{n}\left(x_t x_{t-1}^{-2\rho} - \beta_1 x_{t-1}^{(1-2\rho)}\right)}{n\mu_2}\left[1 + \frac{\sum_{t=1}^{n} x_{t-1}^{-2\rho} - n\mu_2}{n\mu_2}\right]^{-1} \\
&\quad + (1-\beta_1)^{-1}\left(\left(\widehat{\beta}_1 - \beta_1\right)/(1-\beta_1)\right)\frac{\sum_{t=1}^{n}\left(x_t x_{t-1}^{-2\rho} - \beta_1 x_{t-1}^{(1-2\rho)}\right)}{n\mu_2}\left[1 + \frac{\sum_{t=1}^{n} x_{t-1}^{-2\rho} - n\mu_2}{n\mu_2}\right]^{-1} \\
&\quad + (1-\beta_1)^{-1}\left(\left(\widehat{\beta}_1 - \beta_1\right)/(1-\beta_1)\right)^2\frac{\sum_{t=1}^{n}\left(x_t x_{t-1}^{-2\rho} - \beta_1 x_{t-1}^{(1-2\rho)}\right)}{n\mu_2}\left[1 + \frac{\sum_{t=1}^{n} x_{t-1}^{-2\rho} - n\mu_2}{n\mu_2}\right]^{-1},
\end{aligned}$$

and then to order n^{-1} we have for $\beta_1 \neq 1$

$$\begin{aligned}
\widehat{\beta}_2 &= (1-\beta_1)^{-1}\frac{\sum_{t=1}^{n}\left(x_t x_{t-1}^{-2\rho} - \beta_1 x_{t-1}^{(1-2\rho)}\right)}{n\mu_2}\left[1 + \frac{\sum_{t=1}^{n} x_{t-1}^{-2\rho} - n\mu_2}{n\mu_2} + \left(\frac{\sum_{t=1}^{n} x_{t-1}^{-2\rho} - n\mu_2}{n\mu_2}\right)^2\right] \\
&\quad + (1-\beta_1)^{-1}\left(\left(\widehat{\beta}_1 - \beta_1\right)/(1-\beta_1)\right)\frac{\sum_{t=1}^{n}\left(x_t x_{t-1}^{-2\rho} - \beta_1 x_{t-1}^{(1-2\rho)}\right)}{n\mu_2}\left[1 + \frac{\sum_{t=1}^{n} x_{t-1}^{-2\rho} - n\mu_2}{n\mu_2}\right] \\
&\quad + (1-\beta_1)^{-1}\left(\left(\widehat{\beta}_1 - \beta_1\right)/(1-\beta_1)\right)^2\frac{\sum_{t=1}^{n}\left(x_t x_{t-1}^{-2\rho} - \beta_1 x_{t-1}^{(1-2\rho)}\right)}{n\mu_2}
\end{aligned}$$

where $\left(\widehat{\beta}_1 - \beta_1\right)$ is $O_p\left(n^{-1/2}\right)$ and $\left(\widehat{\beta}_1 - \beta_1\right)^2$ is $O_p\left(n^{-1}\right)$.

7 APPENDIX G

7.1 Evidence of the order of the bias expressions

In relation to the order of the biases, our Theorem 2 indicates that the result in Tang and Chen (2009, Theorem 3.1.3), where it is stated that $E(\widehat{\mu} - \mu) = O\left(n^{-2}\right)$, is incorrect for $\rho = 0.5$. We can check the simulation results in Tang and Chen (2009) to see if their simulations shed any light on the biases of the long term mean parameter estimator μ. Examining the simulation results for the CIR model ($\rho = 0.5$) which appear on page 73 (Table 1) in the α column, simulated biases are given for sample sizes $n = 120$, 300, 500 and 2000. Also approximations to the bias are given in parenthesis which are described as being predicted from the theoretical expansions. If the bias is of order n^{-1}, then doubling the sample size would be expected to yield a ratio of the new bias to the original of 0.5 whereas if the bias is of order n^{-2}, the corresponding bias ratio is 0.25. When the sample size is increased from 120 to 300 the corresponding ratios are 0.4 and 0.16 and if the sample size increases from 500 to 2000 the ratios are 0.25 and 0.063 respectively. We shall use these ratios to check whether the results favour a suggested bias of order n^{-1} as opposed to order n^{-2}.

Examining the results presented for CIR Model 1 in Tang and Chen (2009), it is seen that when the sample increases from 120 to 300 the bias ratio based of predicted biases is 0.395 whereas the ratio for a bias of order n^{-1} is very close at 0.4. The ratio for the simulated biases however, does not show any pattern since the bias is shown as hardly changing when n is increased from 120 to 300. A clearer picture emerges when the sample size increases from 500 to 2000 so that the bias ratio for a bias of order n^{-1} is 0.25. The ratio of simulated biases is 0.33 while the ratio of the predicted biases is 0.547. On the other hand if the bias is of order n^{-2}, the suggested bias ratio is 0.063; hence a bias of order n^{-1} is supported.

Turning to the results for CIR Model 2 in Tang and Chen (2009), the evidence for a bias of order n^{-1} is even stronger. Increasing the sample size form 120 to 300 the simulated bias ratio is 0.745 and the predicted bias ratio is 0.412 while for a bias of order n^{-1} a ratio of 0.4 is anticipated. This is not especially close to the simulated bias ratio but if the bias is of order n^{-2} the anticipated ratio is 0.16. Finally if the sample size is increased from 500 to 2000, suggesting a bias ratio of 0.25 for a bias of order n^{-1}, it is seen that the ratio of simulated biases is 0.249 while the ratio of predicted biases is 0.244. Hence a very close correspondence indeed. For a bias of order n^{-2} the suggested bias ratio is 0.063. Hence these results clearly indicate that the bias is of order n^{-1}.

References

[1] Aït-Sahalia, Y. (1996), Nonparametric pricing of interest rate derivative securities, *Econometrica* 64, 527-560.

[2] Merton, R. C. (1990), *Continuous-Time Finance*. Massachusetts, Blackwell.