# Finite sample properties of the maximum likelihood estimator in continuous time models

**Nancy Milena Hoyos Gómez**

A thesis submitted for the degree of
Doctor of Philosophy in Economics

Department of Economics
University of Essex
United Kingdom
January 2017

# Contents

# Acknowledgements

I would like to express my deep gratitude to my supervisor Prof. Marcus J. Chambers for his continuous guidance, helpful advice and constructive suggestions during my doctoral studies. His willingness to give his time so generously and to share his immense knowledge has been very much appreciated. My thanks are also extended to my professors, especially Prof. João Santos Silva, whose passion and dedication to teaching have inspired my academic career. I am grateful for the helpful comments of my examiners Dr. Gordon Kemp and Dr. Michael Thornton.

I gratefully acknowledge the study leave granted by National University of Colombia to pursue my PhD studies. I also acknowledge the financial support from Colciencias (Call 512, 2010) and from University of Essex.

I wish to thank my family. My parents, Cristina and Jorge, for their unconditional love and support throughout my life. Lastly and most importantly, my loving husband, Mario, for always believing in me and encouraging me to follow my dreams. I am profoundly grateful for his spiritual, emotional and intellectual support along this journey.

# Abstract

This dissertation consists of three papers on finite sample properties of the maximum likelihood (ML) estimator of parameters in continuous time dynamic models. In the first chapter, we obtain analytical expressions to approximate the bias and variance of the ML estimator in a univariate model with a known mean. We analyze two cases, when the variable of interest is a stock and when it is a flow. We also study the effect of the initial condition by considering both a fixed and a random initial value. A Monte Carlo study suggests that the performance of the formulae is reasonably good.

Analytical bias expressions are then used in the second chapter to compute bias corrected estimators. This chapter also explores other methods for bias reduction that have been employed in the literature, these being the bootstrap, jackknife, and indirect inference. A Monte Carlo experiment shows that all approaches deliver substantial bias reductions. We also explore the robustness of the results to model misspecifications, and provide an empirical application to the broad effective exchange rate series for euro area.

The third chapter derives the exact discrete representation corresponding to a cointegrated system of mixed first- and second-order stochastic differential equations with mixed sampling and observable stochastic trends. We also provide some formulae to implement the Gaussian estimation and conduct a Monte Carlo experiment to examine the finite sample properties of the Gaussian estimator. Monte Carlo simulations suggest that the bias and variance of the estimators of the short-run, long-run and adjustment coefficients as well as the variance of the intercepts are mainly determined by the data span, while the bias and variance of the covariance coefficients seem to depend on the sample size.

# Introduction

The advantages of continuous time models over those formulated in discrete time have been widely discussed in the literature (see for example, Bergstrom, 1996; Bergstrom and Nowman, 2007). Among them, is their invariance property, which implies that a model can be specified and analysed independently of the data observation interval to be used for estimation. This property however, does not hold for discrete time models since a model built on a particular data frequency, for instance monthly, will be different from one based on another frequency, for instance quarterly. Another advantage is the possibility of obtaining a continuous time path for each variable in the model, which can be then used, for example, to forecast at any point in time.

These models also have the advantage of allowing for separate treatment of stock and flow variables. It is explicitly recognised that stocks are observed at specific points of time, while flows are measured as the accumulation of the underlying rate over a time interval. In terms of modelling, the main difference of this last variable as compared to instantaneously observed ones is that the temporal aggregation occurring with flows induces serial correlation in the disturbances. Because of the distinction in the treatment of these two kind of variables, these models do not suffer from time aggregation bias whereas it may be a serious problem in their discrete time counterpart. Furthermore, both type of variables can be incorporated within the model simultaneously.

Although these models may be more appropriate for describing the dynamics of economic behaviour, they are more difficult to deal with econometrically than discrete time models. A difficulty arises since a continuous record of the variables over time is not available. It is, however, possible to obtain a discrete representation of the continuous time model and to estimate the parameters from discrete data. The approach adopted is to make a discrete approximation to the continuous model or alternatively to derive the exact discrete analogue, and then apply standard estimation methods on the discrete representation.

In the last decades much work has been done on estimating continuous time models based on the exact discrete analogue. The Gaussian methods proposed by Bergstrom (1983, 1985, 1986), which under appropriate conditions yield exact maximum likelihood estimates, has been commonly employed. Recently, it has been found that the maximum likelihood estimates of parameters in continuous time models may be seriously biased in finite samples, being especially pronounced when the continuous time process has a root near zero. Unfortunately, estimation bias may have important implications on hypothesis testing, analysis and prediction. Understanding the bias of parameter estimates is, therefore, of great practical importance and is the main concern of this dissertation.

Estimation bias has been widely studied in the context of discrete time models, and has begun to receive much attention in the continuous time framework. Some attempts along this line of research have been undertaken by Tang and Chen (2009) and Yu (2012), who derived analytical expressions to approximate the bias and variance of the mean reversion estimator. These works have discussed the problem when the variables of interest are stocks, but the finite sample behaviour of estimators in the presence of flow variables have not received attention in the literature. This research focuses in this last case.

The main contribution of the first chapter is the derivation of explicit expressions to approximate the bias of the maximum likelihood estimator in a univariate continuous time model when the variable is a flow. A secondary contribution is the use of a methodology for the derivation of these expressions, that had not been applied in the context of continuous time models. This methodology offers some advantages in terms of ease of implementation and produces results that are comparable to those available in the literature. The performance of the bias expressions is examined through Monte Carlo simulations.

Once an analytical expression has been derived, a simple bias correction can be computed by subtracting the leading term of the expression from the parameter estimate. This approach has the disadvantage that it is only applicable to models for which a bias expression is available. Furthermore, bias expressions may be complex and are often difficult to obtain, this being the case for models with correlated disturbances. In view of these difficulties, other approaches could be preferred in practice.

The aim of the second chapter is to compare four methods of bias reduction for stationary continuous time processes. The first one uses analytical expressions of the type derived in the first chapter, while the other three are based on methods which do not rely on the explicit form of an asymptotic expansion for the bias, those being the jackknife, bootstrap and indirect inference. We consider two scenarios, the former assumes that the model is correctly specified and the latter allows for model misspecifications, these being heteroskedasticity, non-normality and serial correlation in the disturbances. Special attention is given to models with correlated discrete time disturbances. The performance of the four bias corrected estimators is study through

Monte Carlo simulations and an empirical application to the broad effective exchange rate serie for euro area.

The finite sample properties of the maximum likelihood estimators have been studied in the context of stationary models, but have not received much attention in the cointegration framework. The third chapter explores the behaviour of the Gaussian estimator of parameters in cointegrated continuous time systems with mixed stock and flow data. Another contribution of the chapter is the derivation of the exact discrete representation corresponding to a system of mixed first- and second-order stochastic differential equations, which is required to implement the Gaussian estimation.

# Chapter 1

# Estimation bias in continuous time models with stock and flow variables

## 1.1 Introduction

Estimation bias has been widely studied in the context of discrete time models, and has begun to receive considerable attention in the continuous time framework in recent years. It has been shown that estimators can suffer from substantial bias in finite samples, being especially severe when the continuous time process has a root near zero and further exacerbated when intercepts and trends are estimated. Bias may have serious implications. Since parameter estimates serve as input for statistical inference and model analysis, we expect to have distortions not only in hypothesis testing, but also in confidence intervals and impulse response functions. In addition, estimation bias may affect the accuracy of the forecasts and may have a large impact on many practical applications. In finance, for example, it has been found that any bias in the mean reversion estimate is transmitted to the pricing formulae for bonds, bond options, and other derivative securities (see Phillips and Yu, 2005). Obtaining more accurate estimators is, therefore, of great practical importance and is the main concern of this paper.

Finite sample bias can be corrected if the bias function, which relates the bias of the estimates to the values of the parameters, is known. Unfortunately, this function is often unknown. In the continuous time framework, analytical expressions to approximate the bias function have been derived for the Vasicek model (Vasicek, 1977) and the CIR model (Cox et al., 1985). In the univariate case, Tang and Chen (2009) approximated the bias of the mean reversion estimator when the long run mean is unknown. It was shown that the bias in both models is of order $T^{-1}$, but not of order $n^{-1}$, where $T$ is the data span and $n$ is the number of observations, implying that the bias will not disappear unless the data span goes to infinity. This result suggests that the estimation bias in continuous time models is mainly determined by the data span, but not by the sample size.

In the context of Vasicek processes with a known long run mean, Yu (2012) obtained two analytical expressions to approximate the bias of the mean reversion estimator. The first is analogous to that of Marriott and Pope (1954) for the discrete time model and is equivalent to the bias formula in Tang and Chen (2009) of the Vasicek process with an unknown mean. The second expression includes a non-linear term in the bias formula, which was found to be particularly important when the mean reversion parameter is close to zero. The bias has been studied in the context of continuous time Lévy processes by Bao et al. (2015), who approximated the bias of the mean reversion estimator with known and unknown mean, and studied the effects of non-Gaussianity on the bias. In the multivariate case, an analytical expression was derived by Wang et al. (2011). Despite these works having made major contributions to the understanding of the bias in continuous time models, they have only discussed the problem when the variable of interest is a stock.

Economic models can comprise not only stocks, but also flows. Such flow variables as GDP, consumption and exports differ from stocks in that the former are measured as an integral over an interval of time while the latter are observed at points of time. Continuous time models have been used in economics, where several applications have appeared in the literature since the development of the first continuous time macroeconometric model by Bergstrom and Wymer (1976). Although much work on estimating continuous time models in the presence of flow data has been done, the finite sample properties of these estimators have been entirely unexplored. In particular, what is unknown is the magnitude of the bias arising from estimation when the variable is a flow. The main aim of this paper is, therefore, to extend existing bias formulae for the flow case.

Approaches based on approximate moments and distributions, such as Nagar's approximations and Edgeworth expansions, have been commonly employed to obtain analytical bias expressions. Several formulae for a large class of estimators are available in the literature (see for example, Cox and Snell, 1968; Bao and Ullah, 2007). Some of them could be applied to continuous time models with flows. Nevertheless, dealing with flow data is a considerably more complex problem. Because of this complexity, bias expressions may be difficult to compute explicitly or they may be analytically involved. In this paper we employ the formula obtained by Cox and Snell (1968), which is tractable and hence straightforward to implement. We believe that this approach may offer some advantages compared to other methodologies. To illustrate this, we derive an alternative expression for approximating the bias when the variable is a stock and the mean is known. Our results are comparable to those available in the literature. In particular, we show that our bias expressions for the stock case are analogous to those of Bao et al. (2015) and Yu (2012) for the Vasicek process.

The plan of this paper is as follows. Section 2 presents the basics of the continuous time model. Section 3 derives analytical expressions to approximate the bias under different cases: stock or flow variables, fixed or random initial condition. Explicit expressions for the asymptotic variance and the asymptotic bias that arises when the correlation of the disturbances induced by temporal aggregation is not taken into account are derived in Section 4 and 5, respectively. Section 6 reports some simulation results to examine the performance of the analytical expressions and section 7 concludes.

## 1.2 The continuous time model

We consider the continuous time model given by

$$dx(t) = ax(t)dt + \zeta(dt), \quad t > 0, \tag{1.1}$$

where $x(t)$ is a scalar continuous time process, $a$ is a scalar parameter, $\zeta(dt)$ is white noise with mean zero and variance $\sigma^2 dt$ and the initial condition $x(0)$ is defined later.[1] It is assumed that $a$ is negative in order for $x(t)$ to be stationary. Note that $x(t)$ has a root near zero when $a$ is close to zero.

Since $x(t)$ is only observed at discrete intervals of time, the parameters $a$ and $\sigma^2$ must be estimated from discrete data. For this purpose, it is necessary to obtain a discrete representation of the continuous time model. Assuming equispaced discrete data observed at $(h, 2h, ..., nh(= T))$, the exact discrete model is given by

$$x_{th} = e^{ah} x_{th-h} + \eta_{th}, \quad t = 2, ..., n, \tag{1.2}$$

where $h$ is the sampling interval (i.e. the period between observations), $n$ is the total number of observations and $T$ is the data span.[2] The exact discrete analogue, which is obtained from the solution to the differential equation (1.1), has the property that it satisfies data generated by the continuous time model exactly.

Observation $x_h$ takes different forms depending on the way in which the variable is sampled, as a stock or as a flow. When $x(t)$ is a stock, $x_h = e^{ah}x(0) + \eta_h$, and when $x(t)$ is a flow $x_h = a^{-1}(e^{ah} - 1)x(0) + \eta_h$. The properties of the error term $\eta_{th}$ also depend on

[1]The Vasicek model with a known mean setting to zero can be seen as a special case of (1.1) with $a = -k$ and $\zeta(dt) = \sigma dB(t)$, where $B(t)$ is a standard Brownian motion. The parameter $k$ reflects the rate of convergence towards its long term mean and is the parameter of major interest in the bias literature.

[2]For example, if quarterly data are recorded in a 20 year time interval, then $h = 1/4$, $T = 20$ and $n = 80$.

the nature of the variable of interest. When $x(t)$ is a stock variable, the error is given by

$$\eta_{th} = \int_{th-h}^{th} e^{a(th-r)}\zeta(dr), \quad t = 1, ..., n, \tag{1.3}$$

and satisfies the following properties

$$E(\eta_{th}^2) = \sigma_\eta^2 = \frac{\sigma^2}{2a}\left[e^{2ah} - 1\right], \tag{1.4}$$

$$E(\eta_{th}\eta_{th-kh}) = 0 \text{ for } k \geq 1. \tag{1.5}$$

If $x(t)$ is a flow variable, the observed data take the form of integrals $x_{th} = \int_{th-h}^{th} x(r)dr$. In this case, the error term is defined by

$$\eta_h = \frac{1}{a}\int_0^h \left(e^{a(h-r)} - 1\right)\zeta(dr), \tag{1.6}$$

$$\eta_{th} = \frac{1}{a}\left[\int_{th-h}^{th}\left(e^{a(th-r)} - 1\right)\zeta(dr) + \int_{th-2h}^{th-h}\left(e^{ah} - e^{a(th-h-r)}\right)\zeta(dr)\right], \tag{1.7}$$

for $t = 2, ..., n$, and satisfies

$$E(\eta_h^2) = \tilde{\gamma}_0 = \frac{\sigma^2}{2a^3}\left[3 + 2ah - 4e^{ah} + e^{2ah}\right], \tag{1.8}$$

$$E(\eta_{th}^2) = \gamma_0 = \frac{\sigma^2}{a^2}\left[h(e^{2ah} + 1) + \frac{1 - e^{2ah}}{a}\right], \quad t = 2, ..., n, \tag{1.9}$$

$$E(\eta_{th}\eta_{th-h}) = \gamma_1 = \frac{\sigma^2}{a^2}\left[\frac{e^{2ah} - 1}{2a} - he^{ah}\right], \tag{1.10}$$

$$E(\eta_{th}\eta_{th-kh}) = 0 \text{ for } k \geq 2. \tag{1.11}$$

We can see that the form of the exact discrete model differs depending on whether observations are stocks or flows. The exact discrete model is an AR(1) when the variable is observed at points of time while it is an ARMA(1,1) if the variable is measured as a flow over an interval of time. It is important to notice that although the autoregressive coefficient $e^{ah}$ is identical in the two cases, the error given in (1.7) differs from that given in (1.3) in that the former is correlated. It follows from the implied conditions on $E(\eta_{th}\eta_{th-h})$ that the error given in (1.7) is a first-order moving average.

Another important point to notice is that $a < 0$ implies $e^{ah} < 1$ and therefore stationarity. The discrete model has a root near unity when the parameter $a$ is close to zero[3] or when the sampling interval $h$ tends to zero. Note that $e^{ah} \to 1$ when $a \to 0$ or

[3]A root near zero in a continuous time process corresponds to a root near unity in a discrete time process.

$h \to 0$. The initial condition $x(0)$ can be assumed to be either fixed or random. We consider both a fixed value equal to zero and a random value obtained from a normal distribution. It should be noted that whereas the starting value of the stock variable is part of the sample, the starting value of the flow variable is unobservable.

After characterizing the continuous time model we turn now to the calculation of the bias expressions.

## 1.3 Bias formula

We start by briefly describing the approach employed in this paper. Cox and Snell (1968) derived a formula to approximate the bias of the maximum likelihood (ML) estimator in models with independent but not necessarily identically distributed samples. Cordeiro and Klein (1994) pointed out that this formula is also valid for non-independent observations, and rewrote the bias expression given by Cox and Snell for a single parameter estimator $\hat{\beta}_s, 1 \leq s \leq p$, as

$$bias_s = E(\hat{\beta}_s - \beta_s) = \sum_{i=1}^{p} k^{si} \sum_{j,l=1}^{p} \left( k_{ij}^{(l)} - \frac{1}{2} k_{ijl} \right) k^{jl} + \mathrm{O}(n^{-2}), \tag{1.12}$$

where $K^{-1} = \{-k^{ij}\}$ is the inverse of the Fisher Information matrix of dimension $p \times p$ evaluated at the parameter vector $\beta$ of dimension $p \times 1$, $k_{ij} = E(\partial^2 \ell / \partial \beta_i \partial \beta_j)$, $k_{ijl} = E(\partial^3 \ell / \partial \beta_i \partial \beta_j \partial \beta_l)$, $k_{ij}^{(l)} = \partial k_{ij} / \partial \beta_l$, for $i, j, l \leq p$, with $p$ being the total number of parameters and $\ell$ the log-likelihood function. The result in (1.12) is valid provided that all of the $k$ terms are $O(n)$ and $\ell(\beta)$ is regular with respect to all derivatives up to and including the third order.

In order to derive bias expressions of the ML estimators in continuous time models it is convenient to write the bias in matrix notation as follows

$$bias = E(\hat{\beta} - \beta) = K^{-1} F \ \mathrm{vec}(K^{-1}) + \mathrm{O}(n^{-2}), \tag{1.13}$$

where $F = \{F^{(1)} | \cdots | F^{(p)}\}$, with the $p \times p$ matrix $F^{(l)} = \{f_{ij}^{(l)}\}$ and $f_{ij}^{(l)} = k_{ij}^{(l)} - \frac{1}{2} k_{ijl}$ for $l = 1, ..., p$, and $\mathrm{vec}(K^{-1})$ is the vector obtained by stacking the columns of $K^{-1}$.

We now use formula (1.13) to derive explicit bias expressions for the parameter estimator $\hat{a}$.[4] This approach requires specifying a distribution for the error term. We assume that $\eta_{th}$ is a Gaussian process.[5] Let $x$ denote the observed time series and $\sigma^2 V$

[4]We only focus on the estimator $\hat{a}$ because of the little bias that has been found in $\hat{\sigma}^2$ (see Phillips and Yu, 2005; Tang and Chen, 2009).

[5]Note that under the assumption of a Gaussian distribution $\zeta(dt) = \sigma$ dB(t).

the covariance matrix of $x$. The log-likelihood is defined as

$$\ell(a,\sigma^2) = -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln|\sigma^2 V| - \frac{1}{2\sigma^2}x'V^{-1}x. \tag{1.14}$$

It is straightforward to show that the bias formula for $\hat{a}$ is

$$E(\hat{a}-a) = -\frac{\left(k_{111}-2k_{11}^{(1)}\right)k_{22}^2 + k_{12}\left(-3k_{112}k_{22}+4k_{12}^{(1)}k_{22}\right)+k_{12}^2k_{122}}{2\left(k_{12}^2-k_{11}k_{22}\right)^2} + \mathrm{O}(n^{-2}), \tag{1.15}$$

where $k_{11} = E(\partial^2\ell/\partial a^2),\ k_{12} = E(\partial^2\ell/\partial a\partial\sigma^2),\ k_{22} = E(\partial^2\ell/(\partial\sigma^2)^2),\ k_{111} = E(\partial^3\ell/\partial a^3)$, $k_{112} = E(\partial^3\ell/\partial a^2\partial\sigma^2),\ k_{122} = E(\partial^3\ell/\partial a(\partial\sigma^2)^2),\ k_{11}^{(1)} = \partial k_{11}/\partial a$ and $k_{12}^{(1)} = \partial k_{12}/\partial a$.

Formula (1.15) is equivalent to that proposed by Cordeiro and Klein (1994) for ARMA models. The difference is that our formula requires computing the first three derivatives of the likelihood function with respect to the parameters $a$ and $\sigma^2$ while the expression in Cordeiro and Klein (1994) requires calculating the covariance matrix of the data, its inverse and the first two partial derivatives of either one with respect to the parameters. The inverse of the covariance matrix could be, however, difficult to compute for some continuous time models such as those containing flow data. We believe that calculations can be greatly facilitated by using the likelihood function instead of the covariance matrix of the data. Furthermore, the likelihood function allows us to derive the bias expression under the assumption of both a fixed and a random initial condition.

We turn now to the computation of the bias expressions assuming that both type of variables are observed at $(0, h, 2h, ..., nh-h)$, hence the sample size is $n$. We start by considering the simplest case, this being when the variable is observed at points of time. Since calculations are straightforward, but somewhat tedious, we only present the resulting bias expressions. Details are given in the Appendix.

### 1.3.1 Stock variables

In order to compute a formula for the bias, it is convenient to use the predictive error decomposition of the log-likelihood function, which for the model given in (1.2)-(1.5) for $t = 1, ..., n-1$ results in

$$\ell(a,\sigma^2) = -\frac{n}{2}\ln(2\pi) - \frac{n-1}{2}\ln\sigma_\eta^2 - \frac{1}{2\sigma_\eta^2}\sum_{t=1}^{n-1}\left(x_{th}-e^{ah}x_{th-h}\right)^2 - \frac{1}{2}\ln\sigma_x^2 - \frac{x_0^2}{2\sigma_x^2}, \tag{1.16}$$

where $\sigma_\eta^2$ is the variance of the disturbances, $\sigma_x^2 = -\sigma^2/2a$ is the variance of $x$ and the initial condition $x_0 = x(0)$ is assumed to be $x_0 \sim N(0, -\sigma^2/2a)$. Taking the initial value $x_0$ as deterministic and equal to zero, the log-likelihood function can be calculated conditioning on the first observation. In this case, the log-likelihood function becomes

$$\ell(a,\sigma^2) = -\frac{n-1}{2}\ln(2\pi) - \frac{n-1}{2}\ln\sigma_\eta^2 - \frac{1}{2\sigma_\eta^2}\sum_{t=1}^{n-1}\left(x_{th} - e^{ah}x_{th-h}\right)^2. \tag{1.17}$$

The bias expressions are, therefore, found by computing the first three derivatives of the exact log-likelihood function (1.16) and the conditional log-likelihood function (1.17) with respect to $a$ and $\sigma^2$, taking expectations and then substituting these figures into (1.15). The first result corresponding to a random initial condition and the second to a fixed initial condition are presented in the following theorems.

**Theorem 1.3.1.** *Under model given in (1.2)-(1.5) for $t = 1, ..., n-1$ with a fixed initial condition $x_0 = 0$, the bias of $\hat{a}$ is given by*

$$E(\hat{a} - a) = -\frac{3 + e^{-2ah}}{2T} + \mathrm{o}(T^{-1}). \tag{1.18}$$

*When $a \to 0$ the bias reduces to*

$$E(\hat{a} - a) = -\frac{2}{T} + \mathrm{o}(T^{-1}), \tag{1.19}$$

*and when $h \to 0$ it becomes*

$$E(\hat{a} - a) = -\frac{2}{T} + \mathrm{o}(T^{-1}). \tag{1.20}$$

**Theorem 1.3.2.** *Under model given in (1.2)-(1.5) for $t = 1, ..., n-1$ with a random Gaussian initial condition $x_0$ with mean 0 and variance $-\sigma^2/2a$, the bias of $\hat{a}$ is given by*

$$E(\hat{a} - a) = -\frac{3 + e^{-2ah}}{2T} + \frac{4n - 2 + (5n-6)e^{2ah} - ne^{-2ah}}{2T(n-1)\left[n - (n-2)e^{2ah}\right]} + \mathrm{o}(T^{-1}). \tag{1.21}$$

*When $a \to 0$ the bias reduces to*

$$E(\hat{a} - a) = 0 + \mathrm{o}(T^{-1}), \tag{1.22}$$

*and when $h \to 0$ it becomes*

$$E(\hat{a} - a) = -\frac{2}{T} + \frac{2}{T(1 - aT)} + \mathrm{o}(T^{-1}). \tag{1.23}$$

The theoretical result indicates that the estimator is downward biased. The bias can be substantial, being especially large when the continuous time parameter $a$ is close to zero. For example, for the case of fixed $x_0$ and 10 years of monthly data (i.e. $T = 10$ and $h = 1/12$), the percentage bias is 200% when $a = -0.1$, while it is less than 8% when $a = -3$. Note that the leading term in equations (1.18) and (1.21), which is the same as that obtained by Yu (2012), is of order $\mathrm{O}(T^{-1})$. It suggests that the bias is mainly determined by the data span, implying that the bias will not disappear unless the data span goes to infinity.

There are two additional remarks worth noting. First, the bias is larger when $x_0$ is fixed than when $x_0$ is random, which is consistent with the result in Bao et al. (2015). To see this, we plot the bias expression for the sampling interval $h = 1/12$ and three values of data span, these being $T = 5, 10, 20$ (see Figure 1.1). These combinations correspond to 5, 10 and 20 years of monthly data. The discrepancy seems to reduce when we increase $n$ by increasing $T$, implying that the influence of the initial value on the bias function tends to be negligible as the data span goes to infinity. We note, however, that the initial value has an important impact on the bias function when $a$ is relatively small. As can be seen in Theorems 1.3.1 and 1.3.2, as $a$ goes to zero the bias converges to $-2/T$ when $x_0$ is fixed while the bias tends to zero when $x_0$ is random.

Second, the bias decreases when we increase $n$ by increasing the data frequency (i.e. when $h$ becomes smaller). This is illustrated in Figure 1.2, where we plot the bias expression for $T = 10$ and three values of $h$, these being $h = 1/4, 1/12, 1/252$. These combinations correspond to 10 years of quarterly, monthly and daily data. Results also indicate that the implications of high frequency data differ depending on whether the initial condition is fixed or random. Indeed, (1.20) suggests than the bias converges to $-2/T$ as $h$ goes to zero when $x_0$ is fixed, implying that the bias does not disappear unless $T$ goes to infinity. In contrast, (1.23) suggests that the bias tends to $-2/T \ + 2/T(1 - aT)$ as $h \to 0$ when $x_0$ is random, implying that the bias approaches zero when either $a \to 0$ or $T \to \infty$.

Third, the second term in equation (1.21) is of smaller order than $1/T$ and can be ignored when $a$ is far away from zero. However, this second term is not negligible when $a$ is close to zero, in which case it becomes of order $\mathrm{O}(T^{-1})$. This observation is consistent with the result in Yu (2012).

### 1.3.2 Flow variables

We now illustrate how to compute analytical bias expressions when the variable is measured as an integral. It should be noted that calculations for this case become more

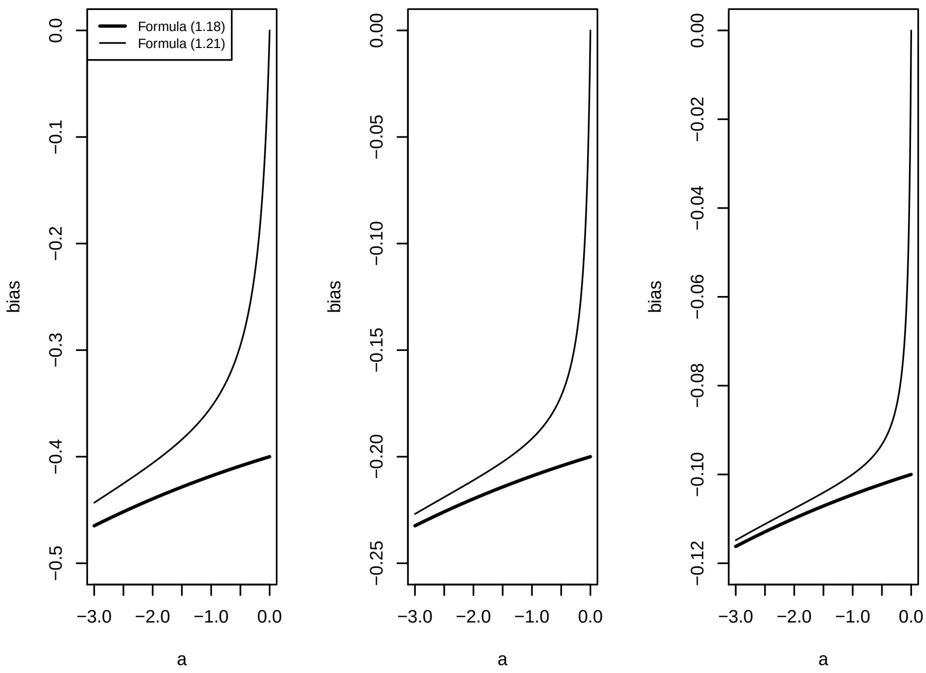


Figure 1.1: The bias of $\hat{a}$ for stock and monthly frequency data as a function of $a$. The graphs correspond to $T = 5, 10, 20$ (i.e. $n = 60, 120, 240$), respectively. The solid line is the approximate bias according to the formula (1.21) (i.e. $x_0$ random) and the dark solid line is the approximate bias according to the formula (1.18) (i.e. $x_0$ fixed at zero).

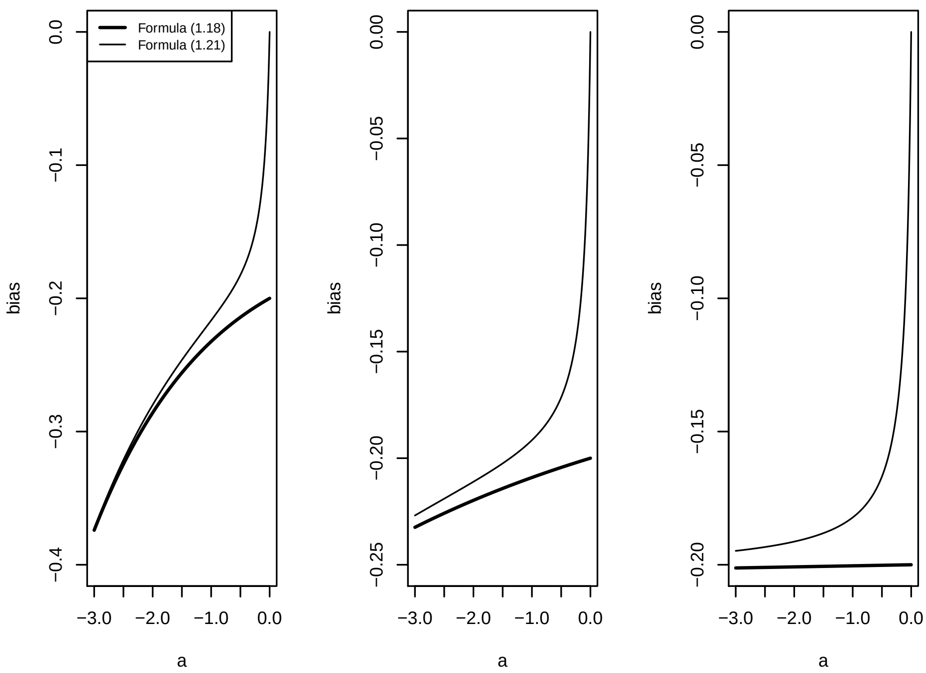


Figure 1.2: The bias of $\hat{a}$ for stock data and $T = 10$ as a function of $a$. The graphs correspond to quarterly, monthly and daily data (i.e. $n = 40, 120, 2520$), respectively. The solid line is the approximate bias according to the formula (1.21) (i.e. $x_0$ random) and the dark solid line is the approximate bias according to the formula (1.18) (i.e. $x_0$ fixed at zero).

complicated than those for stocks since the error term in the exact discrete model is correlated. In spite of the additional complications, the formula by Cox and Snell (1968) can be applied by using the ARMA representation of $x_{th}$, that is, by expressing the error $\eta_{th}$ as a first-order moving average. We present the MA representation of the disturbance process in the following lemma.

**Lemma 1.3.1.** *The disturbance process $\eta_{th}$ defined in (1.7) with variance $\gamma_0$ and autocovariance $\gamma_1$ given in (1.9) and (1.10) admits the representation*

$$\eta_{th} = \varepsilon_{th} + \theta\varepsilon_{th-h}, \quad t = 2, ..., n, \tag{1.24}$$

*where $\varepsilon_{th}$ is a white noise process with mean zero and variance $\sigma_\varepsilon^2 = \gamma_1/\theta = \gamma_0/(1+\theta^2)$, $\theta = (1-\sqrt{1-4\rho^2})/(2\rho)$, and $\rho = \gamma_1/\gamma_0$.*

It should be noted that the moving average coefficient $\theta$ is a function of $a$, being positive for $a < 0$ and reaching its maximum value of $2-\sqrt{3}$ when the continuous time process has a root near zero (see Figures 1.25 and 1.26). The ARMA representation leads us to consider the log-likelihood function

$$\ell(a,\sigma^2) = -\frac{n}{2}\ln(2\pi) - \frac{n-1}{2}\ln\sigma_\varepsilon^2 - \frac{1}{2\sigma_\varepsilon^2}\sum_{t=1}^{n-1}\left(x_{th} - e^{ah}x_{th-h} - \theta\varepsilon_{th-h}\right)^2 - \frac{1}{2}\ln\sigma_x^2 - \frac{\tilde{x}_0^2}{2\sigma_x^2}, \tag{1.25}$$

where $\sigma_x^2$ is the variance of $x$ given by[6]

$$\sigma_x^2 = -\frac{(1+2ahe^{ah}-e^{2ah})(1+\theta^2+2\theta e^{ah})\sigma^2}{2a^3\theta(1-e^{2ah})},$$

and $\tilde{x}_0$ is the starting value assumed to be $\tilde{x}_0 \sim N(0, \sigma^2(1+ah-e^{ah})/a^3)$. As it was shown for the stock case, the initial condition plays an important role when $a$ approaches zero. To show that this situation also occurs for the flow case, in the derivation of the bias formula it is assumed that the initial value of the flow variable is also part of the sample. Note that we have used a different letter to distinguish $\tilde{x}_0$ from $x_0$, this last value being unobservable when the variable is a flow. Since $\tilde{x}_0$ is used instead of $x_0$, the equation relating $x_h$ to $x(0) = x_0$ is not employed. Instead of this, we use $x_h = e^{ah}\tilde{x}_0 + \varepsilon_h + \theta\varepsilon_0$.

It is important to notice that equation (1.25) is only an approximation to the exact likelihood function. We could use the exact $\ell$ to derive the bias formula,[7] but the algebra is more involved and the resulting bias expression is, of course, more complicated than

[6]Using the representation (1.2) with disturbances given in (1.6)-(1.11), the variance of $x$ can alternatively be expressed as $\sigma_x^2 = (1+ah-e^{ah})\sigma^2/a^3$.

[7]See for example the innovation algorithm suggested by Brockwell and Davis (1991, pg. 254-256) to obtain the exact likelihood of an ARMA model. Also, Bergstrom (1990) for the exact likelihood function of a continuous time model with flow data.

the one in Theorem 1.3.4. We believe that the gain in accuracy achieved by using the exact likelihood function rather than an approximation does not justify the additional complications involved. Moreover, as can be seen below this approximation appears to work well.

Taking $\tilde{x}_0$ as deterministic and equal to zero, the log-likelihood function can be computed conditioning on the first observation as following

$$\ell(a,\sigma^2) = -\frac{n-1}{2}\ln(2\pi) - \frac{n-1}{2}\ln\sigma_\varepsilon^2 - \frac{1}{2\sigma_\varepsilon^2}\sum_{t=1}^{n-1}\left(x_{th} - e^{ah}x_{th-h} - \theta\varepsilon_{th-h}\right)^2. \quad (1.26)$$

The bias expressions are obtained by differentiating (1.25) and (1.26) with respect to $a$ and $\sigma^2$. The first result corresponding to a random initial value and the second to a fixed initial value presented in the following theorems are obtained by assuming that the derivative of $\theta$ with respect to $a$ is zero.[8] It should be noted that although these results provide a good approximation for a range of parameter values that would appear to be relevant in practice, the quality of the approximation deteriorates as $a$ moves away from zero.

**Theorem 1.3.3.** *Under model given in (1.2) and (1.24) for $t = 1, ..., n-1$ with a fixed initial condition $\tilde{x}_0 = 0$, the bias of $\hat{a}$ is*

$$E(\hat{a} - a) = -\frac{3 + e^{-2ah}}{2T} + \mathrm{o}(T^{-1}). \quad (1.27)$$

*When $a \to 0$ the bias reduces to*

$$E(\hat{a} - a) = -\frac{2}{T} + \mathrm{o}(T^{-1}), \quad (1.28)$$

*and when $h \to 0$ it becomes*

$$E(\hat{a} - a) = -\frac{2}{T} + \mathrm{o}(T^{-1}). \quad (1.29)$$

**Theorem 1.3.4.** *Under model given in (1.2) and (1.24) for $t = 1, ..., n-1$ with a random Gaussian initial condition $\tilde{x}_0$ with mean 0 and variance $\sigma^2(1 + ah - e^{ah})/a^3$, the bias of*

[8]From Lemma 1.3.1 we see that $\theta$ is a function of $a$, implying that $d\theta/da$ is non-zero. This derivative is, however, a complex function of $a$ and its inclusion considerably complicate calculations. Moreover, as can be seen in the Appendix $d\theta/da$ is negligible. To simplify calculations the results given in Theorem 1.3.3 and 1.3.4 are obtained by assuming $d\theta/da = 0$. Details are presented in Remarks 1 and 2.

*$\hat{a}$ is*

$$E(\hat{a}-a) = -\frac{3+e^{-2ah}}{2T} + \frac{e^{-2ah}C}{2TD^2} + \mathrm{o}(T^{-1}), \tag{1.30}$$

*where $C$ and $D$ are complicated functions of the continuous time parameter $a$ (see the Appendix). When $a \to 0$ the bias reduces to*

$$E(\hat{a}-a) = 0 + \mathrm{o}(T^{-1}), \tag{1.31}$$

*and when $h \to 0$ it becomes*

$$E(\hat{a}-a) = -\frac{2}{T} + \frac{2}{T(1-aT)} + \mathrm{o}(T^{-1}). \tag{1.32}$$

We see that (1.27) is identical to (1.18). Although the analytical bias expression for flows is more complicated than that for stocks when the initial condition is random, the implications seem to be similar. The leading term is the same in both expressions. Similar to the stock case, in the near unit root situation the bias converges to zero as $a \to 0$ and to $-2/T + 2/T(1-aT)$ as $h \to 0$. Evaluating the expressions at different values of $a$, $T$ and $h$ (see Figures 1.3 and 1.4) we see that the bias for flows is slightly smaller in magnitude than the bias for stocks, although the small gap vanishes when the sample size increases. These results suggest that for large samples the size of the bias does not depend on whether observations are stocks or flows. However, it is only true when the correlation in the discrete time disturbances induced by temporal aggregation is taken into account. As we will show in the next section the bias can be considerably large when correlated errors are ignored.

It should be noted that the approach of Cox and Snell (1968) is particularly convenient for deriving the bias formula when the variable of interest is a flow since it only requires computing the first three partial derivatives of the likelihood function with respect to the parameters. Alternatively, we could use other formulae available in the discrete time literature, for example that of Bao and Ullah (2007) (for its implementation in the continuous time framework see Yu, 2012). One possible advantage of the approach of Bao and Ullah (2007) as compared with the approach used in this paper is that the former does not require any distributional assumption, and therefore their analytical results are valid for both normal and non-normal samples of observations. It should be emphasized, however, that our bias expressions are expected to be robust to non-normality (for details see Bao et al., 2015).[9] Moreover, we believe that the approach of Cox and Snell (1968) is

[9]Bao et al. (2015) studied the bias of the estimator of the mean reversion parameter in the context of continuous time Lévy processes, and showed that the bias expression is robust to non-normality under

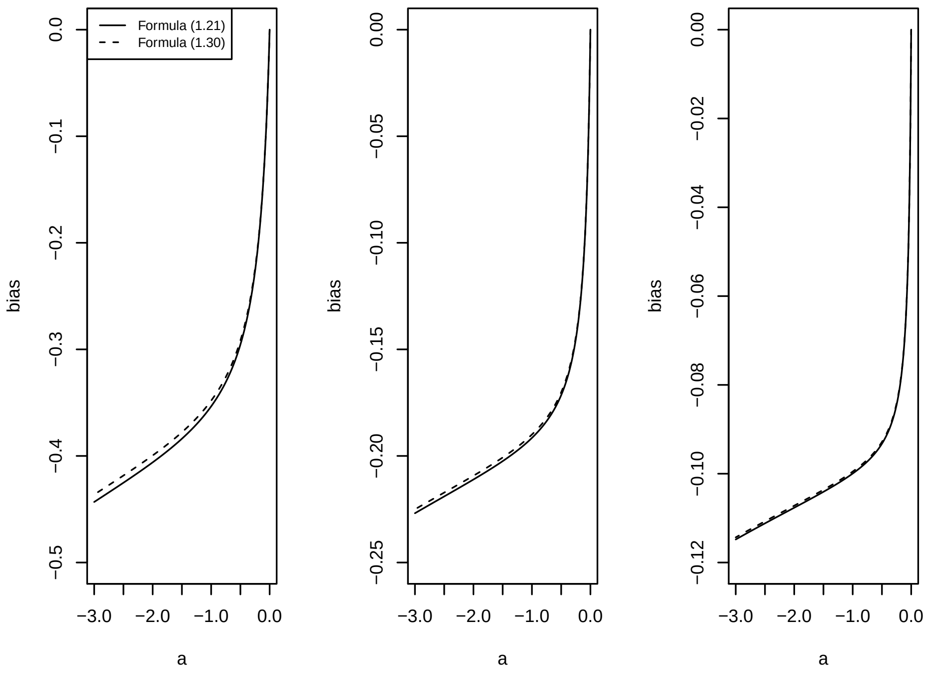


Figure 1.3: The bias of $\hat{a}$ for monthly data and a random initial condition as a function of $a$. The graphs correspond to $T = 5, 10, 20$ (i.e. $n = 60, 120, 240$), respectively. The solid line is the approximate bias according to the formula (1.21) (i.e. the variable is a stock) and the dashed line is the approximate bias according to the formula (1.30) (i.e. the variable is a flow).

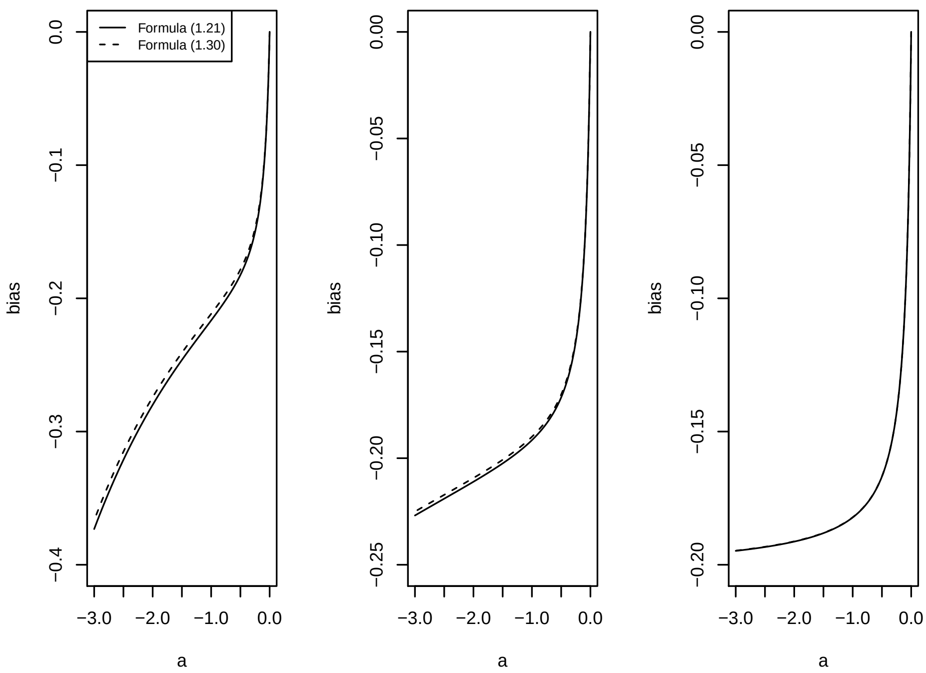


Figure 1.4: The bias of $\hat{a}$ for $T = 10$ and a random initial condition as a function of $a$. The graphs correspond to quarterly, monthly and daily data (i.e. $n = 40, 120, 2520$), respectively. The solid line is the approximate bias according to the formula (1.21) (i.e. the variable is a stock) and the dashed line is the approximate bias according to the formula (1.30) (i.e. the variable is a flow).

simpler and hence easier to implement.

## 1.4 Asymptotic variance

Another advantage of the Cox and Snell's approach is that an analytical expression for the asymptotic variance of the estimators can be obtained straightforwardly. The following theorem presents the corresponding formula for $\hat{a}$.

**Theorem 1.4.1.** *Under model (1.2) with disturbances given in (1.3)-(1.5) or (1.24) for $t = 1, ..., n-1$, the asymptotic variance of $\hat{a}$ is*

$$Var(\hat{a}) = -\frac{1 - e^{-2ah}}{Th} + \mathrm{o}(T^{-1}). \tag{1.33}$$

This result is identical to that obtained by Tang and Chen (2009) and Yu (2012) for the stock case. In the next section we derive the asymptotic bias that arises when the temporal aggregation implicit in the definition of a flow variable, which induces serial correlation in the disturbances of the exact discrete model, is not taken into account. The finite sample bias caused by serial correlation in the errors and the finite sample variance will be explored in section 1.6 through simulations.

## 1.5 Asymptotic bias

It is well known from the discrete time literature that ignoring correlated errors in dynamic models leads to inconsistent estimates. Maeshiro (1999) provides an intuitive explanation of this bias. He points out that since the disturbance term $\eta_{th}$ is correlated with the regressor $x_{th-h}$, part of the variation in the dependent variable $x_{th}$ that is caused by $\eta_{th}$ is credited to $x_{th-h}$, contaminating the autoregressive coefficient. The sign of this contaminating effect is the same as that of the moving average coefficient.

Maeshiro (1999) derived the following expression for the asymptotic bias of the least square (LS) estimator of the autoregressive coefficient $\phi = e^{ah}$ when the disturbances follow a MA(1) process[10]

$$\text{plim}(\hat{\phi} - \phi) = \frac{Cov(\eta_{th}, x_{th-h})}{\sigma_x^2} = \frac{(1-\phi^2)\theta}{1 + \theta^2 + 2\phi\theta}. \tag{1.34}$$

the case of a random initial condition. They also found that although the skewness parameter matters for the bias expression when the initial condition is fixed, the effect disappears when it is assumed to be zero.

[10]The asymptotic bias was also derived by Phillips and Wickens (1978) for a more general model that includes exogenous variables. Both MA(1) and AR(1) disturbances were considered.

Using the alternative expression for the variance of $x$ (i.e. $\sigma_x^2 = (1 + ah - e^{ah})\sigma^2/a^3$), the asymptotic bias can be written in terms of the continuous time parameter $a$ as following

$$\text{plim}(\hat{\phi} - \phi) = \frac{(e^{2ah} - 2ahe^{ah} - 1)}{2(1 + ah - e^{ah})}. \tag{1.35}$$

Since $\hat{a} = \ln(\hat{\phi})/h$ we can use the result given in (1.35) to derive the asymptotic bias of the estimator $\hat{a}$. The resulting bias of the conditional maximum likelihood estimator is presented in the following theorem.[11]

**Theorem 1.5.1.** *Under model given in (1.2) and (1.24) the asymptotic bias of the ML estimator $\hat{a}$ (conditional on the initial condition) is*

$$\text{plim}(\hat{a} - a) = \frac{1}{h}\ln\left[\frac{(e^{ah} - 1)^2}{2e^{ah}(-1 - ah + e^{ah})}\right]. \tag{1.36}$$

Since $\theta$ is positive the sign of the bias must be positive as noted by Maeshiro (1999) for the discrete model. This condition is equivalent to

$$\frac{(e^{ah} - 1)^2}{2e^{ah}(e^{ah} - 1 - ah)} > 1. \tag{1.37}$$

Simpliying (1.37) becomes $e^{2ah} - 2ahe^{ah} - 1 < 0$. Note that $e^{2ah} - 2ahe^{ah} - 1 = 2a^3\gamma_1$. Since $a < 0$, the condition is satisfied if $\gamma_1 > 0$. Evaluating $\gamma_1$ at different values of $a$ and $h$, we see that $\gamma_1 > 0$ for $a < 0$ (see also Figure 1.24 which shows $\rho = \gamma_1/\gamma_0$ for quarterly, monthly and daily data). The bias caused by the serial correlation can be much more serious than that induced by the presence of lagged dependent variables. Note that it does not disappear as the span goes to infinity.

## 1.6 Monte Carlo simulations

This section examines the performance of the analytical expressions using Monte Carlo simulations. To generate simulated data, we first compute the discrete time disturbances as follows

$$\eta_{th} = \sigma_\eta \vartheta_{th}, \quad t = 1, ..., n, \tag{1.38}$$

[11]Note that the ML estimator (conditional on fixed or random initial value) is equivalent to the LS estimator.

when the variable is a stock and according to

$$\eta_h = m_{11}\vartheta_h, \tag{1.39}$$

$$\eta_{2h} = m_{22}\vartheta_{2h} + m_{21}\vartheta_h, \tag{1.40}$$

$$\eta_{th} = m_{tt}\vartheta_{th} + m_{t,t-1}\vartheta_{th-h}, \quad t = 3, ..., n, \tag{1.41}$$

when the variable is a flow, where $\vartheta_{th}$ is a random draw from a standard normal distribution, $\sigma_\eta$ is the positive square root of the variance given in (1.4), the coefficients $m_{21}$, $m_{t,t-1}$ $(t = 3, ..., n)$ and the positive coefficients $m_{tt}$ $(t = 1, ..., n)$ are obtained recursively from the equations

$$\begin{aligned} m_{11}^2 &= \tilde{\gamma}_0, \\ m_{21} &= \gamma_1 / m_{11}, \\ m_{22}^2 &= \gamma_0 - m_{21}^2, \\ m_{t,t-1} &= \gamma_1 / m_{t-1,t-1}, \\ m_{tt}^2 &= \gamma_0 - m_{t,t-1}^2, \quad t = 3, ..., n, \end{aligned}$$

with $\tilde{\gamma}_0$, $\gamma_0$ and $\gamma_1$ given in (1.8)-(1.10).

Assuming that the initial condition is either $x(0) = 0$ or $x(0) \sim N(0, -\sigma^2/2a)$, simulated data are then computed by using (1.2), and the corresponding equation for observation $h$ (i.e. $x_h = e^{ah}x(0) + \eta_h$ for stocks and $x_h = a^{-1}(e^{ah} - 1)x(0) + \eta_h$ for flows). We consider different values for the continuous time parameter $a$ over the interval [-3,0) and five combinations of sampling interval and data span, these being $(T = 5, h = 1/12)$, $(T = 10, h = 1/12)$, $(T = 20, h = 1/12)$, $(T = 10, h = 1/4)$, $(T = 10, h = 1/252)$, which correspond to 5, 10 and 20 years of monthly data, 10 years of quarterly data, and 10 years of daily data, respectively. The experiment is replicated 10000 times for each of these combinations. The initial condition is included in the sample for stocks but not for flows. Thus, the number of observations used for estimation is $n + 1$ when $x(t)$ is a stock and $n$ when $x(t)$ is a flow. For simplicity, we assume that the discrete time disturbance corresponding to the first observation is zero and compute the conditional maximum likelihood estimates. For the flow case, the log-likelihood function is calculated based on the ARMA representation.[12]

In order to reduce the dimension of the optimization problem from two to only one dimension, it is convenient to concentrate out $\sigma^2$. The concentrated log-likelihood function is obtained by maximising $\ell(a, \sigma^2)$ with respect to $\sigma^2$, and then substituting $\hat{\sigma}^2$

[12] Alternatively, we can obtain exact Gaussian estimates. See the iterative estimation procedure proposed by Bergstrom (1990, pg. 113-114), which does not require expressing the disturbances as a moving average process.

into $\ell(a, \sigma^2)$. Conditioning on $x_0$, the concentrated log-likelihood function for stocks is (ignoring a constant)

$$\ell(a) = -\frac{n}{2}\ln\left[\sum_{t=1}^{n}(x_{th} - e^{ah}x_{th-h})^2\right], \tag{1.42}$$

and conditioning on $x_h$ and $\varepsilon_h = 0$, the concentrated log-likelihood function for flows becomes (ignoring a constant)

$$\ell(a) = -\frac{n-1}{2}\ln\left[\sum_{t=2}^{n}(x_{th} - e^{ah}x_{th-h} - \theta\varepsilon_{th-h})^2\right]. \tag{1.43}$$

For each simulation the ML estimate of $a$ is, therefore, the value of $a$ that maximizes the concentrated likelihood. The actual bias is calculated by subtracting $a$ from the average of the estimates obtained from the 10000 simulated samples.[13] Figures 1.5-1.12 plot the actual bias and the approximate bias according to the analytical bias expressions as a function of $a$. When $x(0)$ is random, the bias is approximated by using expression (1.21) for both stocks and flows. We observe from the figures that the bias expressions do a good job of capturing the actual bias of $\hat{a}$.

The implications of the analytical bias expressions seem to be similar to those suggested by the simulations. First, the bias is mainly determined by the data span. Increasing the data frequency reduces the bias, but it does not disappear unless the span goes to infinity. Second, the bias is larger when the initial condition is fixed than when it is random, though the gap seems to reduce as $T$ increases. Third, the initial condition has an important impact on the bias function when $a$ is relatively small. Results indicate that as $a \to 0$ the bias goes to zero for the case of random $x_0$ while it does not disappear for the case of fixed $x_0$. Fourth, the bias is smaller for flows than for stocks, although the discrepancy appears to reduce when the span increases or the sampling interval decreases. For example, if $a = -3$, $T = 10$ and $x_0$ is fixed, the bias is -0.34 for flows and -0.45 for stocks when $h = 1/4$ (i.e. quarterly data) while it is -0.19 for both flows and stocks when $h = 1/252$ (i.e. daily data). Results indicate, therefore, that for large samples the magnitude of the bias does not depend on whether observations are stocks or flows.

Fifth, there is a considerable discrepancy between the analytical bias function and that obtained by simulation for the near unit root situation for the case of random $x_0$, which is where the bias function is highly non-linear. This result is, however, not surprising since we have derived our bias expressions under the assumption of stationarity (i.e. $a < 0$). We can see this in Figure 1.13 which shows the bias as a function of the data span for

[13] Since the number of simulated samples is large, the bias obtained by simulations must closely match the true bias.

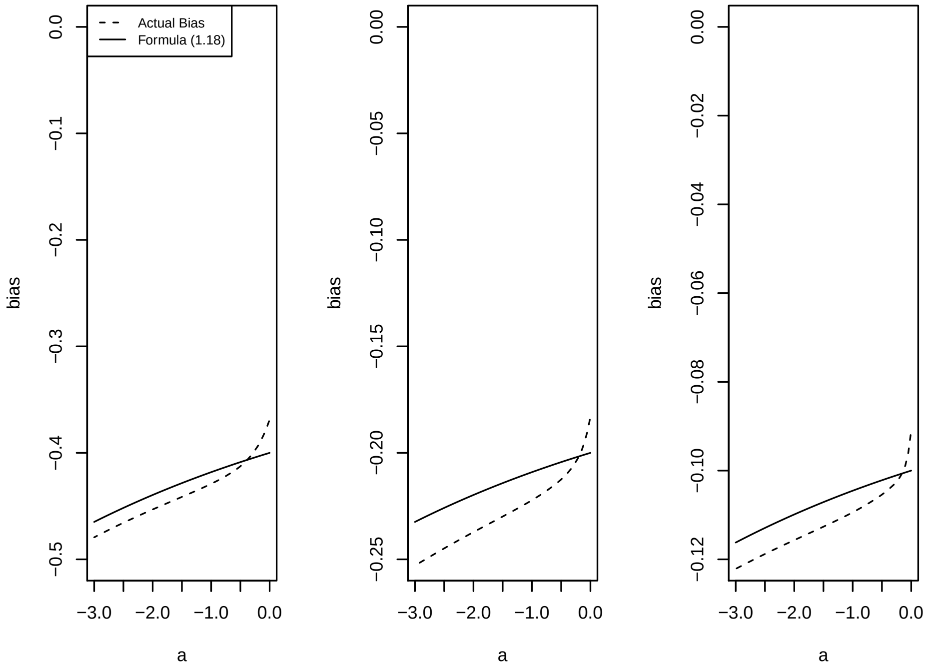


Figure 1.5: The bias of $\hat{a}$ for stock and monthly frequency data with $x_0$ fixed as a function of $a$. The graphs correspond to $T = 5, 10, 20$ (i.e. $n = 60, 120, 240$), respectively. The dashed line is the actual bias and the solid line is the approximate bias according to the formula (1.18).

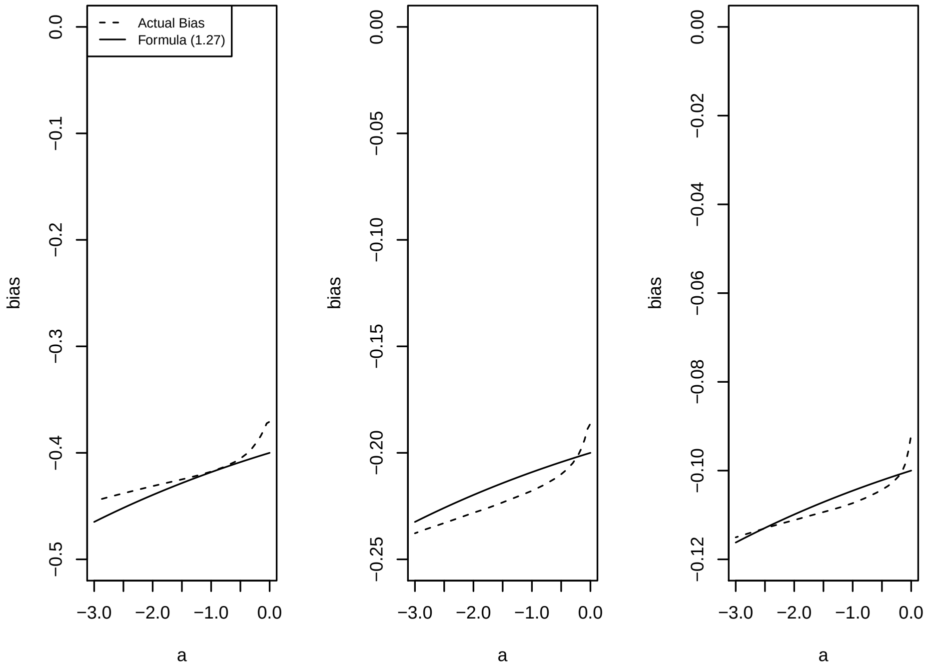


Figure 1.6: The bias of $\hat{a}$ for flow and monthly frequency data with $x_0$ fixed as a function of $a$. The graphs correspond to $T = 5, 10, 20$ (i.e. $n = 60, 120, 240$), respectively. The dashed line is the actual bias and the solid line is the approximate bias according to the formula (1.27).

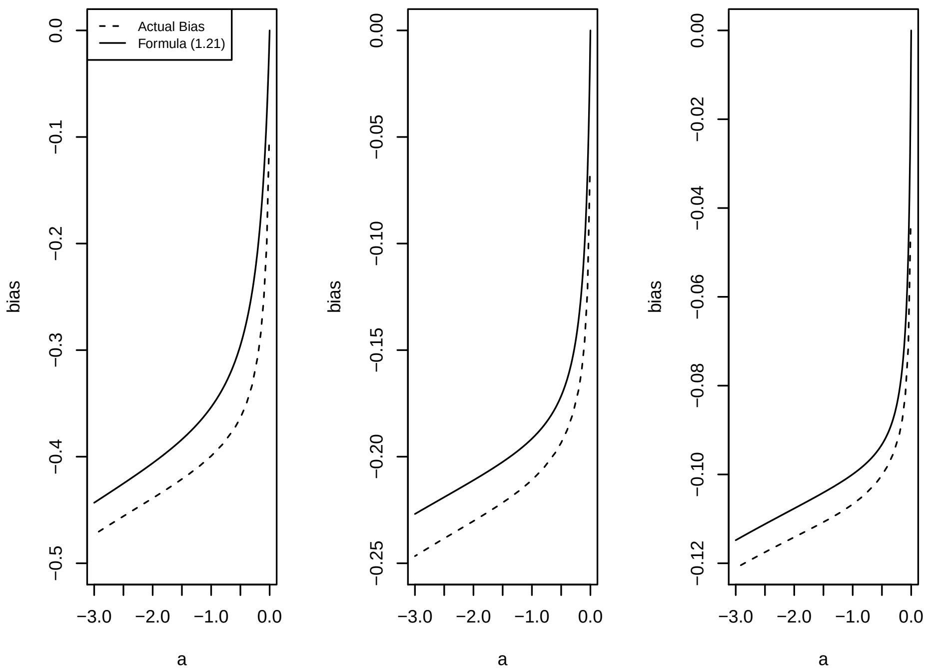


Figure 1.7: The bias of $\hat{a}$ for stock and monthly frequency data with $x_0$ random as a function of $a$. The graphs correspond to $T = 5, 10, 20$ (i.e. $n = 60, 120, 240$), respectively. The dashed line is the actual bias and the solid line is the approximate bias according to the formula (1.21).

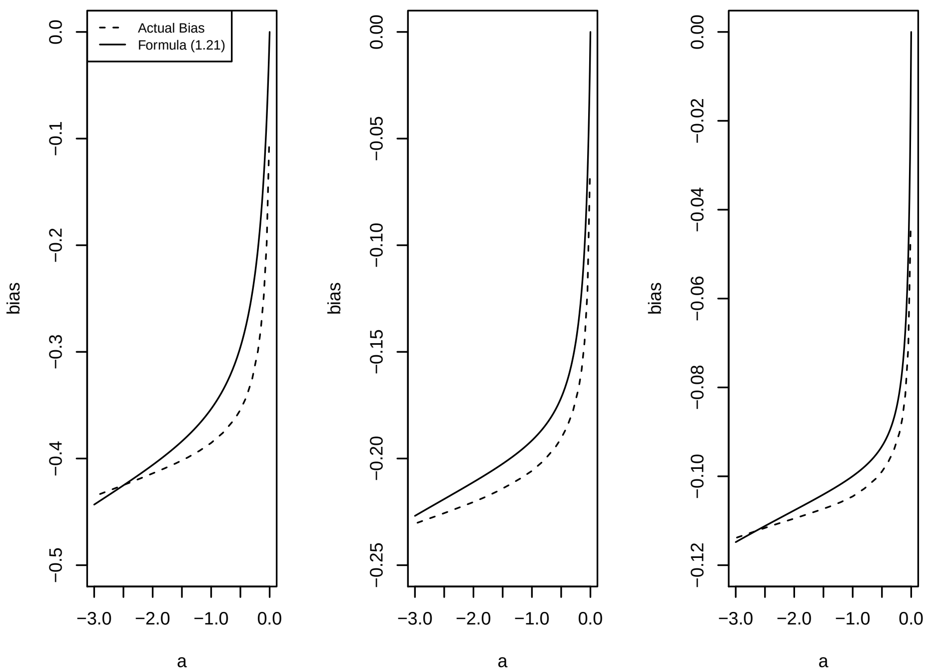


Figure 1.8: The bias of $\hat{a}$ for flow and monthly frequency data with $x_0$ random as a function of $a$. The graphs correspond to $T = 5, 10, 20$ (i.e. $n = 60, 120, 240$), respectively. The dashed line is the actual bias and the solid line is the approximate bias according to the formula (1.21).

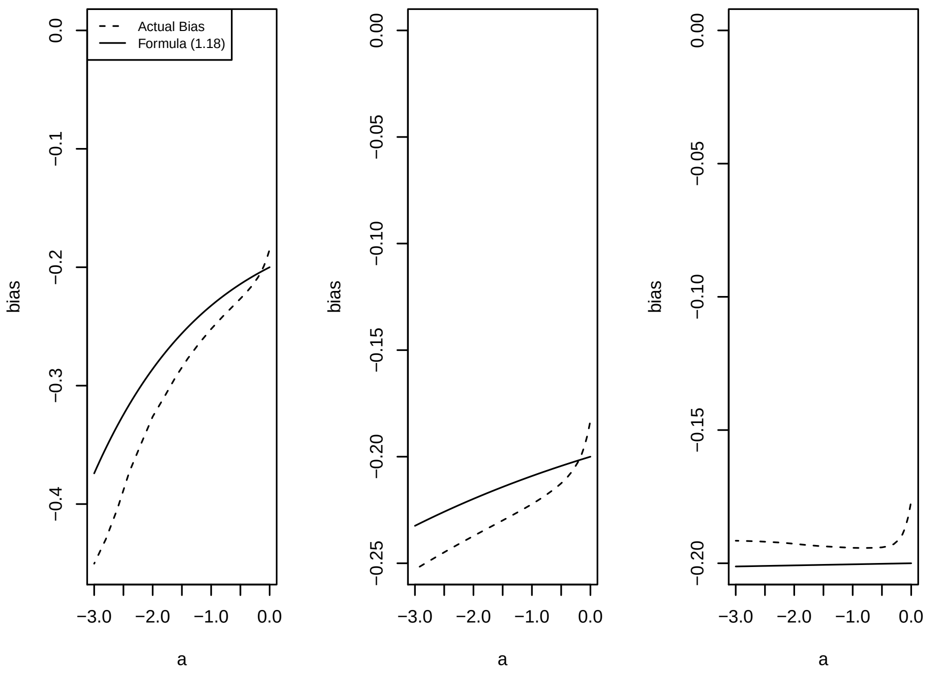


Figure 1.9: The bias of $\hat{a}$ for stock data and $T = 10$ with $x_0$ fixed as a function of $a$. The graphs correspond to quarterly, monthly and daily data (i.e. $n = 40, 120, 2520$), respectively. The dashed line is the actual bias and the solid line is the approximate bias according to the formula (1.18).

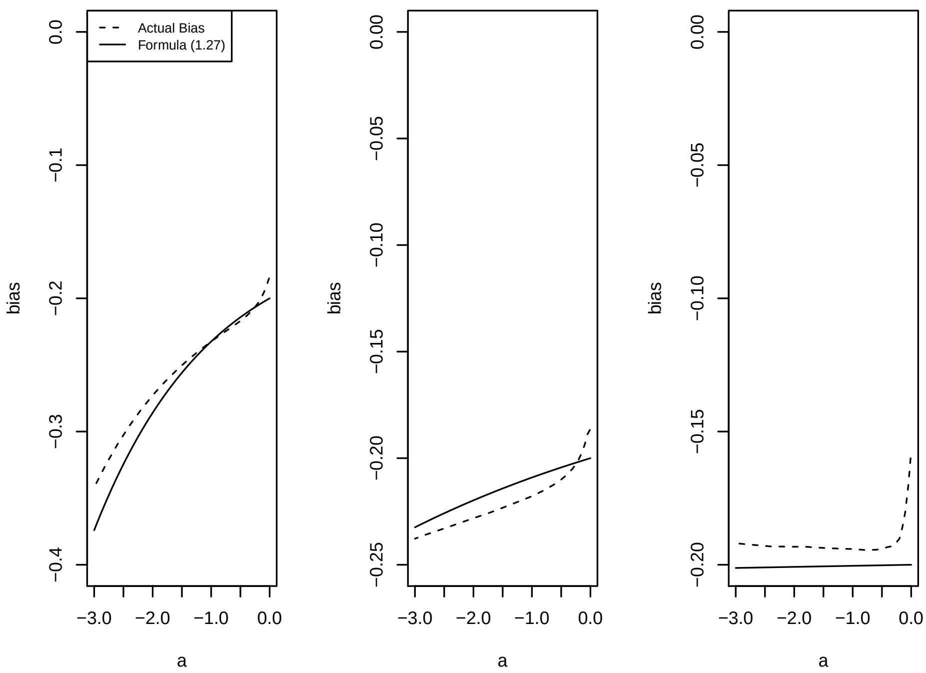


Figure 1.10: The bias of $\hat{a}$ for flow data and $T = 10$ with $x_0$ fixed as a function of $a$. The graphs correspond to quarterly, monthly and daily data (i.e. $n = 40, 120, 2520$), respectively. The dashed line is the actual bias and the solid line is the approximate bias according to the formula (1.27).

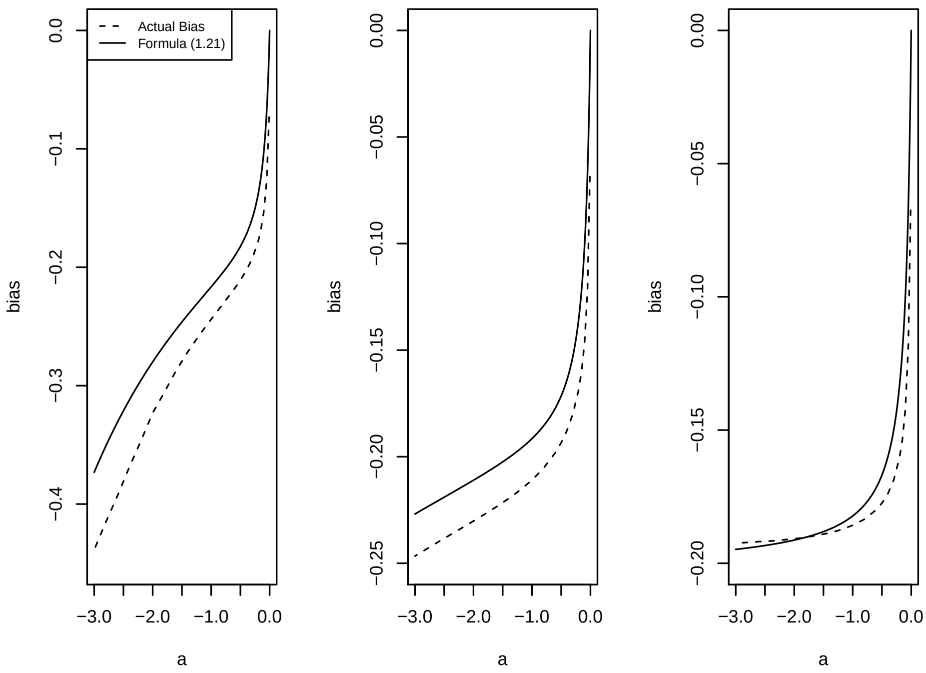


Figure 1.11: The bias of $\hat{a}$ for stock data and $T = 10$ with $x_0$ random as a function of $a$. The graphs correspond to quarterly, monthly and daily data (i.e. $n = 40, 120, 2520$), respectively. The dashed line is the actual bias and the solid line is the approximate bias according to the formula (1.21).

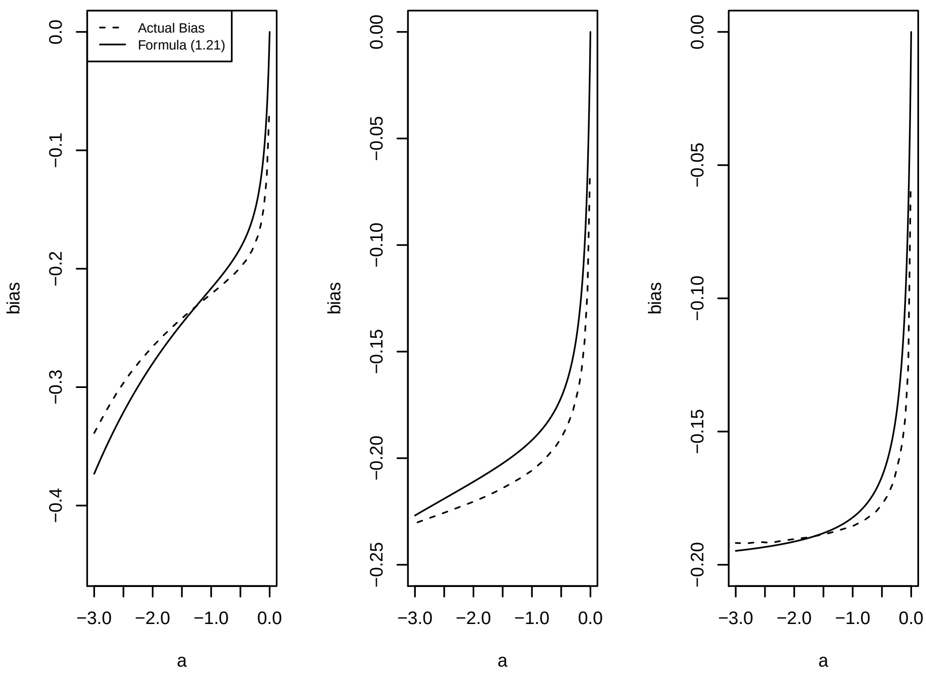


Figure 1.12: The bias of $\hat{a}$ for flow data and $T = 10$ with $x_0$ random as a function of $a$. The graphs correspond to quarterly, monthly and daily data (i.e. $n = 40, 120, 2520$), respectively. The dashed line is the actual bias and the solid line is the approximate bias according to the formula (1.21).

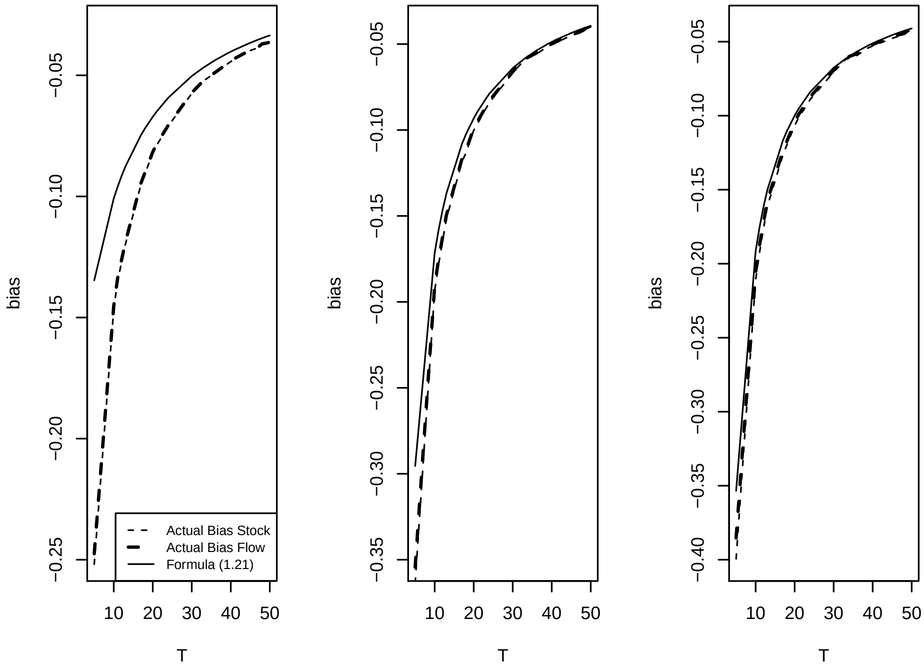


Figure 1.13: The bias of $\hat{a}$ for monthly frequency data with $x_0$ random as a function of $T$. The graphs correspond to $a = -0.1, -0.5, -1$, respectively. The dashed lines are the actual bias and the solid line is the approximate bias according to the formula (1.21).

monthly data and three values of $a$, these being -0.1, -0.5 and -1. Figure 1.13 suggests that the gap between the actual bias and the bias expressions is larger for small values of $a$ and $T$. The gap reduces, however, when the data span increases.

Figures 1.14-1.17 display the actual and asymptotic variance of $\hat{a}$. There is a considerable discrepancy between the analytical expression and that obtained by simulations, although the gap decreases as the data span increases. We also see that the variance is larger for stocks than for flows when $T$ is small or the frequency is low. The variance does not depend on the nature of the variable when the sample size is large. We now explore the bias caused by ignoring the serial correlation in the disturbances. Data are generated according to (1.2) and (1.39)-(1.41) (i.e. the variable is a flow), but the model is estimated assuming that disturbances satisfy the properties given in (1.4) and (1.5) (i.e. the variable is treated as a stock).

Following Maeshiro (1999) the bias is decomposed into the bias caused by the serial correlation and the bias caused by the lagged dependent variable. The former is called by Maeshiro (1999) the correlation effect and the latter the dynamic effect. We use the same terminology. The dynamic effect corresponds to the bias expressions derived for the stock case (i.e. equation (1.18)). The correlation effect is obtained by subtracting the dynamic effect from the bias. These two effects are shown in Figures 1.18-1.21. It can be seen from these figures that the bias can be substantial, being especially large when the parameter

$a$ moves away from zero. The correlation effect is positive as predicted by the theoretical results and is much larger than the dynamic effect when $a$ is away from zero. Since the dynamic effect is negative, the two effects are mutually offsetting. Figures 1.18-1.21 also suggest that the bias does not disappear as the data span goes to infinity nor as the sampling interval tends to zero. Figures 1.22 and 1.23 show the actual correlation bias and the asymptotic bias of the estimator $\hat{a}$. Results reveal that there is a considerable difference between the analytical expression and that obtained by simulations, although the gap decreases as the data span increases.

## 1.7 Conclusions

The main contribution of this paper has been the derivation of explicit expressions to approximate the bias in continuous time models for stock and flow variables. The implications of the bias formulae are consistent with what has been found in the literature. The bias depends mainly on the data span, implying that the bias will not disappear unless the data span goes to infinity. Findings also suggest that the bias is larger when the initial condition is fixed than when it is random, and smaller for flows than for stocks, although the discrepancy decreases when the data span and the sample size increases, respectively. A Monte Carlo study reveals that the performance of the bias formulae is reasonably good, although there is a considerable discrepancy between the analytical bias function and that obtained by simulation for the near unit root situation for the case of random $x_0$. These gaps reduce, however, when the data span increases.

The bias formulae derived in this paper have important practical applications for statistical inference and model analysis, and its application can be expected to lead to more accurate estimates of the parameters as compared with uncorrected estimates. Although our analytical results have been confined to univariate first-order differential equations, the approach used in this paper could be applied to more general models. It would be useful, for example, to measure the magnitude of the estimation bias in higher order differential equations or a system of equations with mixed stock and flow data.

## 1.8 Appendix

*Proof of Theorem 1.3.1.* Substituting (1.4) into (1.17) gives

$$\ell(a,\sigma^2) = -\frac{n-1}{2}\ln(2\pi) - \frac{n-1}{2}\ln\left[-\frac{\sigma^2(1-e^{2ah})}{2a}\right] + \frac{a\sum_{t=1}^{n-1}(x_{th}-e^{ah}x_{th-h})^2}{(1-e^{2ah})\sigma^2}. \quad (1.44)$$

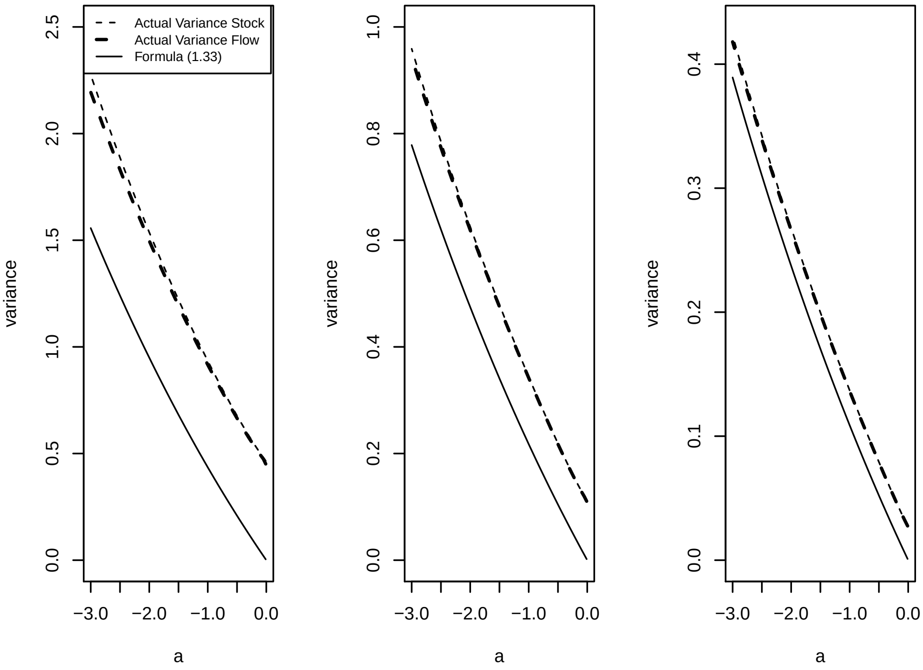


Figure 1.14: The variance of $\hat{a}$ for monthly frequency data with $x_0$ fixed as a function of $a$. The graphs correspond to $T = 5, 10, 20$ (i.e. $n = 60, 120, 240$), respectively. The dashed lines are actual variances and the solid line is the approximate variance according to the formula (1.33).

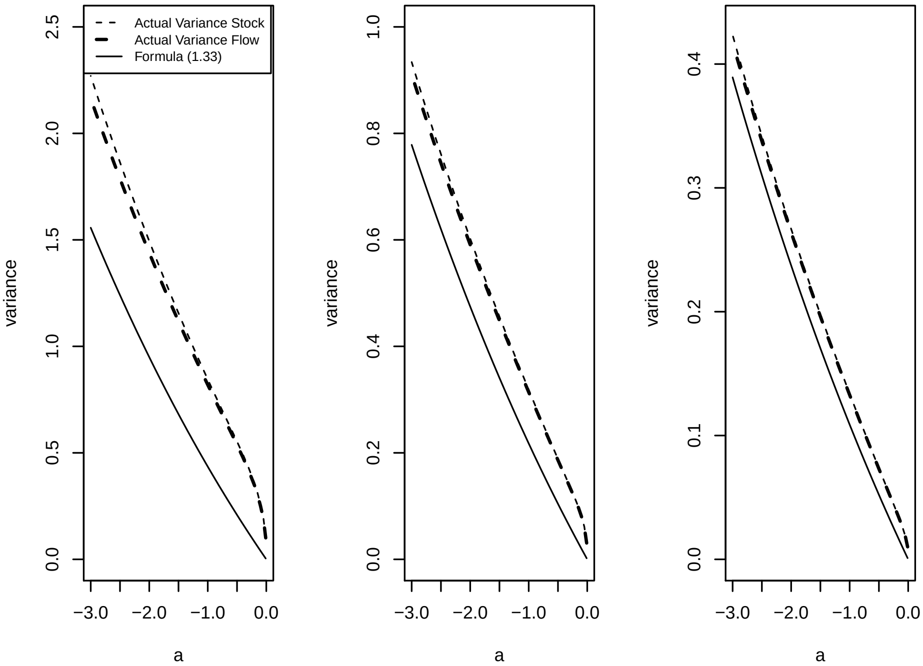


Figure 1.15: The variance of $\hat{a}$ for monthly frequency data with $x_0$ random as a function of $a$. The graphs correspond to $T = 5, 10, 20$ (i.e. $n = 60, 120, 240$), respectively. The dashed lines are actual variances and the solid line is the approximate variance according to the formula (1.33).

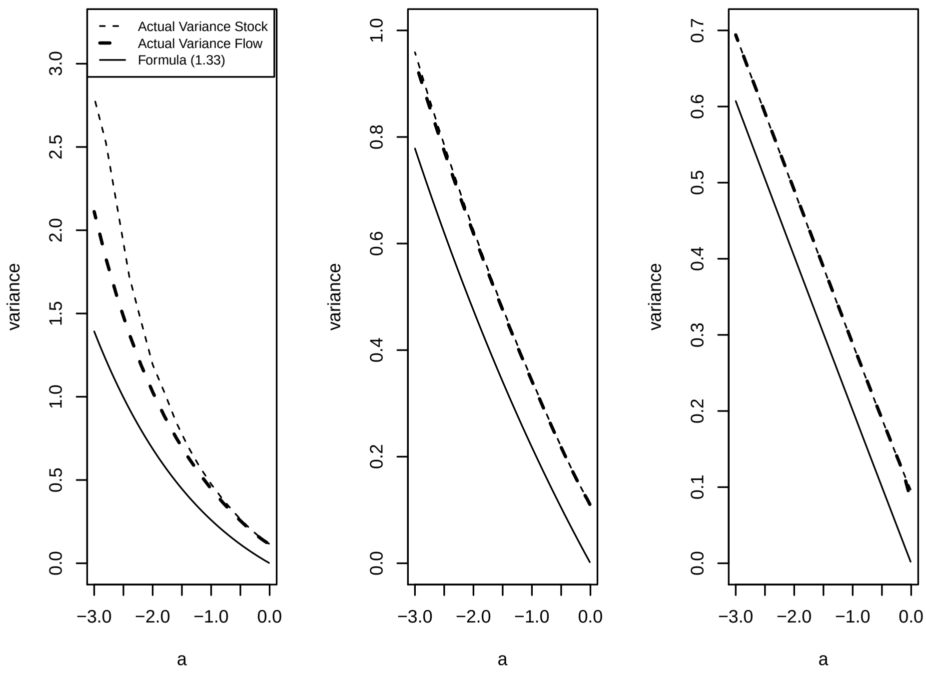


Figure 1.16: The variance of $\hat{a}$ for $T = 10$ with $x_0$ fixed as a function of $a$. The graphs correspond to quarterly, monthly and daily data (i.e. $n = 40, 120, 2520$), respectively. The dashed lines are actual variances and the solid line is the approximate variance according to the formula (1.33).

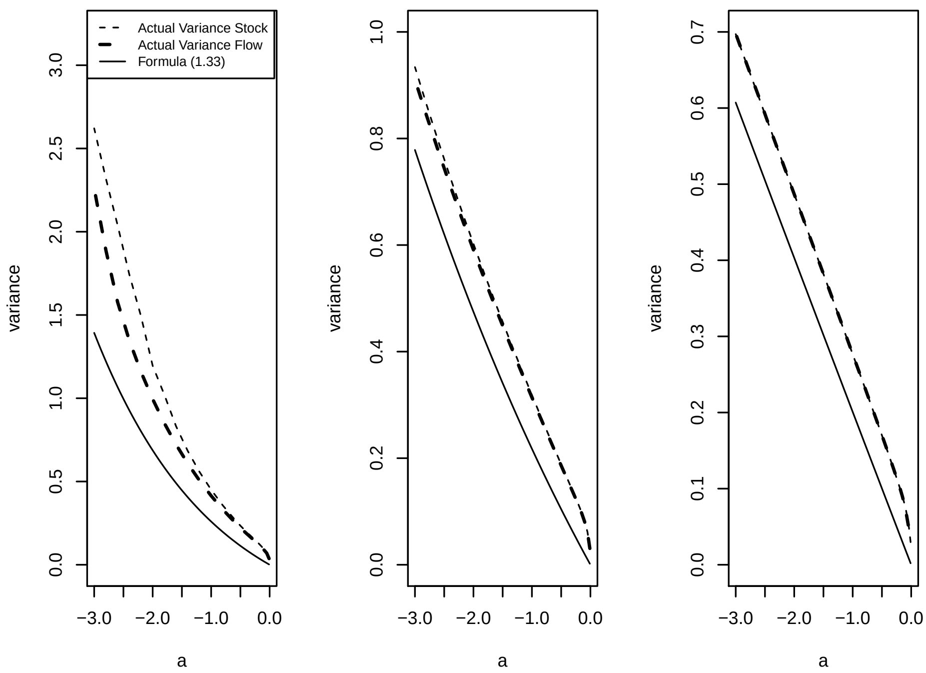


Figure 1.17: The variance of $\hat{a}$ for $T = 10$ with $x_0$ random as a function of $a$. The graphs correspond to quarterly, monthly and daily data (i.e. $n = 40, 120, 2520$), respectively. The dashed lines are actual variances and the solid line is the approximate variance according to the formula (1.33).

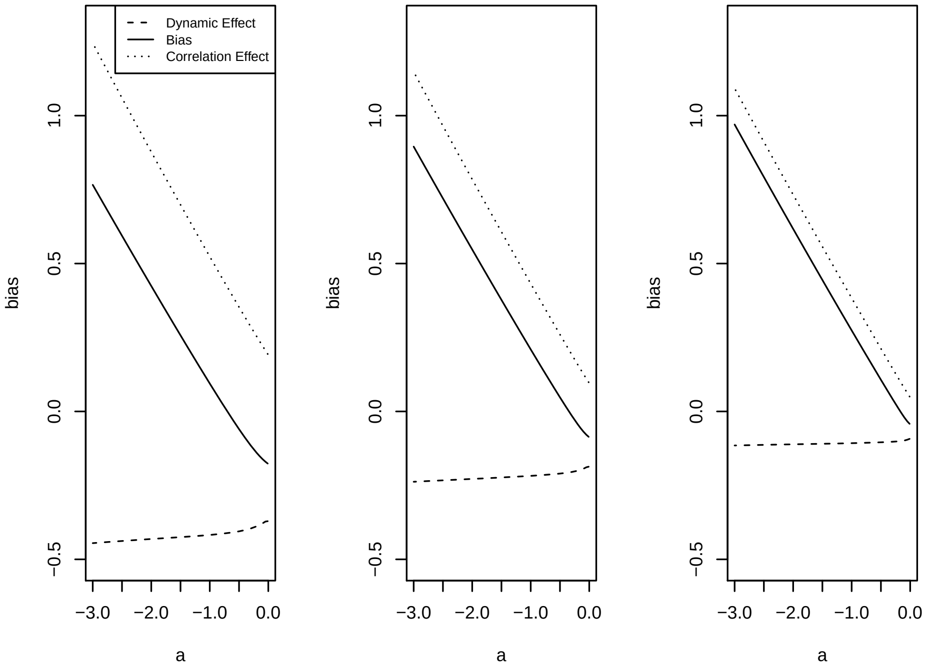


Figure 1.18: The bias of $\hat{a}$ for flow and monthly frequency data with $x_0$ fixed as a function of $a$. The graphs correspond to $T = 5, 10, 20$ (i.e. $n = 60, 120, 240$), respectively. The dashed line is the dynamic effect, the dotted line is the correlation effect and the solid line is the total bias.

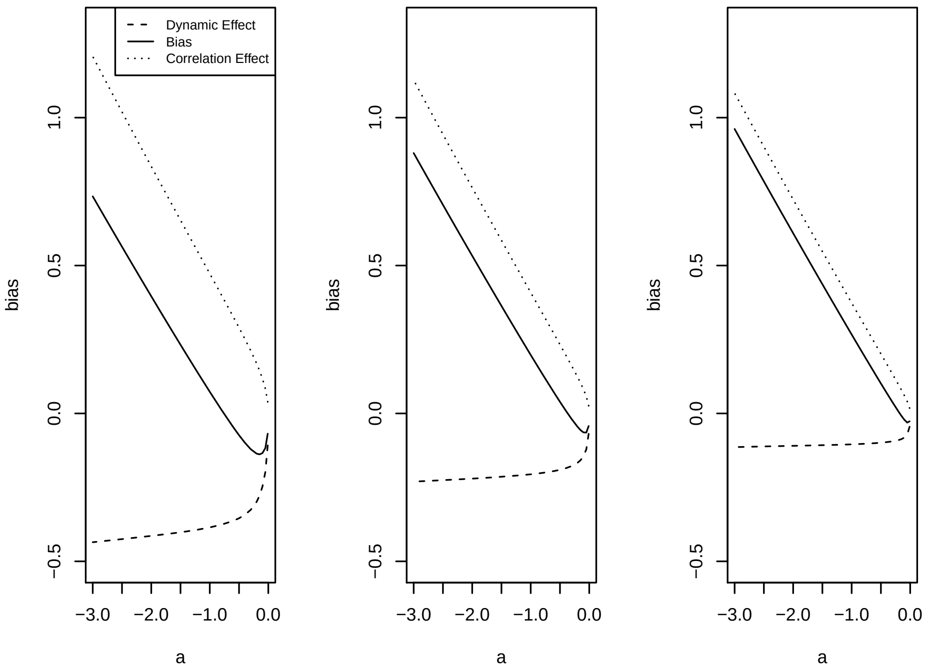


Figure 1.19: The bias of $\hat{a}$ for flow and monthly frequency data with $x_0$ random as a function of $a$. The graphs correspond to $T = 5, 10, 20$ (i.e. $n = 60, 120, 240$), respectively. The dashed line is the dynamic effect, the dotted line is the correlation effect and the solid line is the total bias.

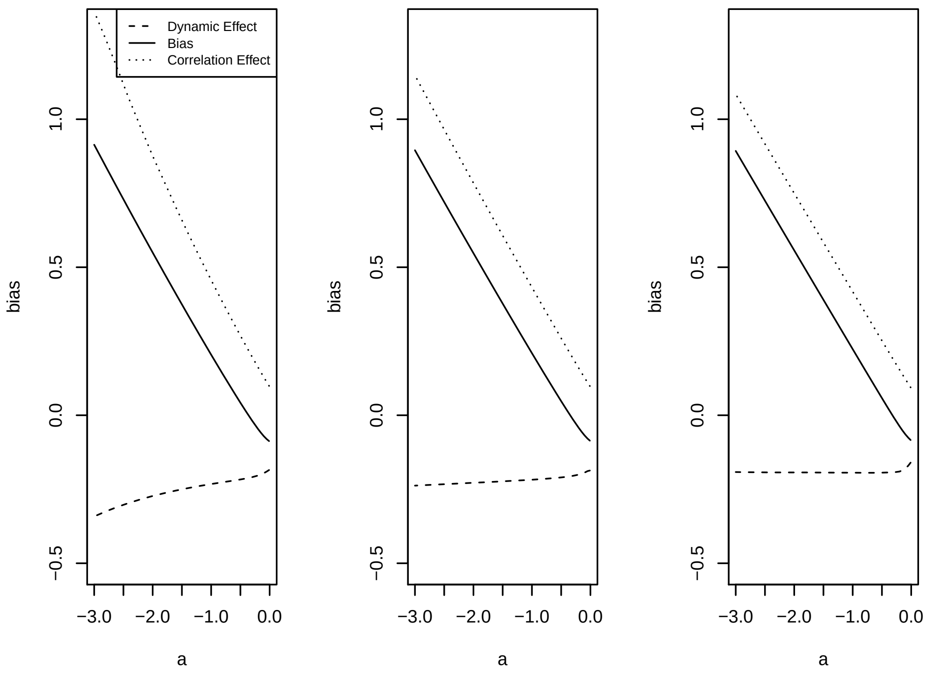


Figure 1.20: The bias of $\hat{a}$ for flow data and $T = 10$ with $x_0$ fixed as a function of $a$. The graphs correspond to quarterly, monthly and daily data (i.e. $n = 40, 120, 2520$), respectively. The dashed line is the dynamic effect, the dotted line is the correlation effect and the solid line is the total bias.

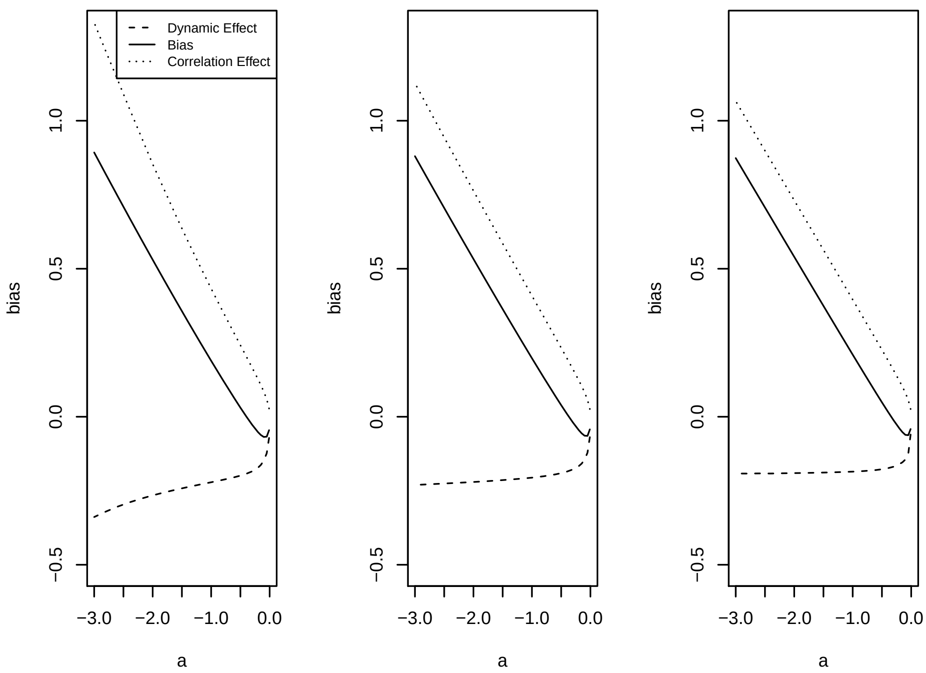


Figure 1.21: The bias of $\hat{a}$ for flow data and $T = 10$ with $x_0$ random as a function of $a$. The graphs correspond to quarterly, monthly and daily data (i.e. $n = 40, 120, 2520$), respectively. The dashed line is the dynamic effect, the dotted line is the correlation effect and the solid line is the total bias.

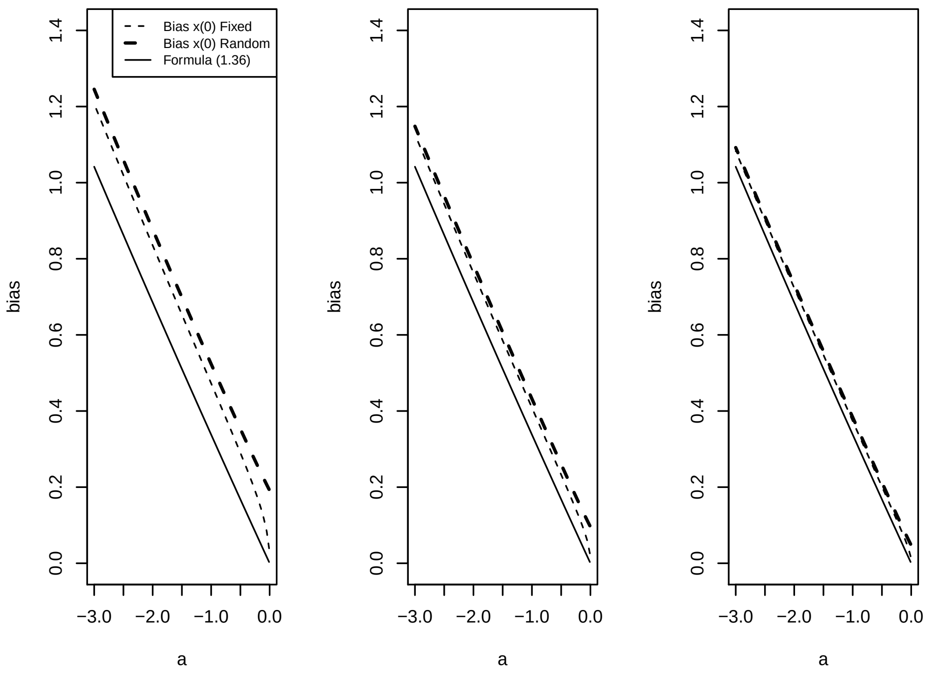


Figure 1.22: The correlation bias of $\hat{a}$ for flow and monthly frequency data as a function of $a$. The graphs correspond to $T = 5, 10, 20$ (i.e. $n = 60, 120, 240$), respectively. The dashed lines are actual correlation biases and the solid line is the asymptotic bias according to the formula (1.36).

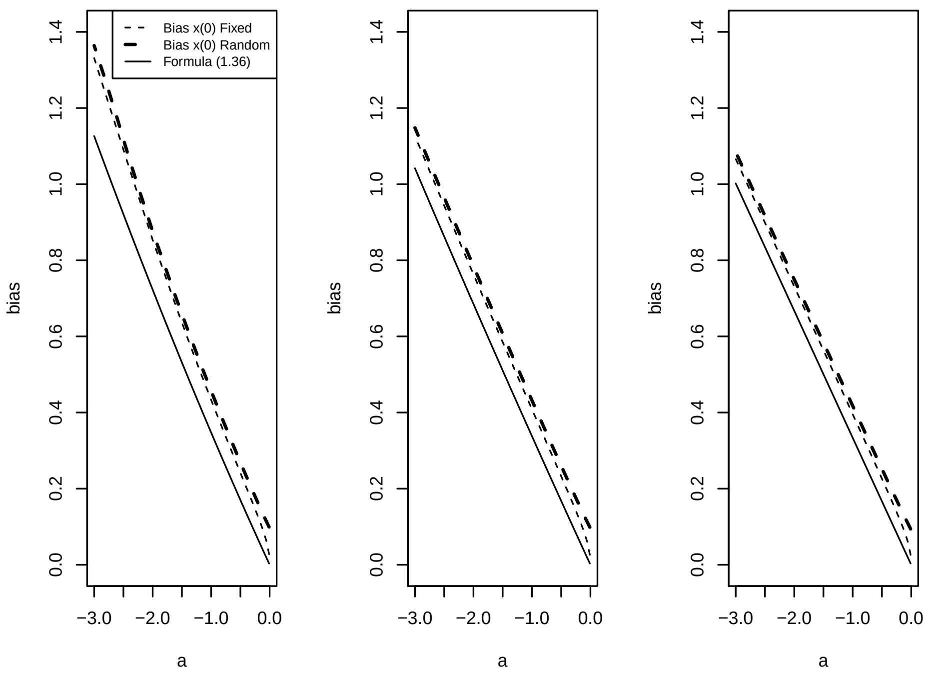


Figure 1.23: The correlation bias of $\hat{a}$ for flow data and $T = 10$ as a function of $a$. The graphs correspond to quarterly, monthly and daily data (i.e. $n = 40, 120, 2520$), respectively. The dashed lines are actual correlation biases and the solid line is the asymptotic bias according to the formula (1.36).

The partial derivatives of (1.44) with respect to $a$ and $\sigma^2$ are

$$\frac{\partial l}{\partial a} = \frac{\left(1-e^{2ah}+2ahe^{2ah}\right)(n-1)}{2a\left(1-e^{2ah}\right)} - \frac{2ahe^{ah}\sum_{t=1}^{n-1}\left(x_{th}-e^{ah}x_{th-h}\right)x_{th-h}}{\left(1-e^{2ah}\right)\sigma^2}$$
$$+\frac{\sum_{t=1}^{n-1}\left(x_{th}-e^{ah}x_{th-h}\right)^2}{\left(1-e^{2ah}\right)\sigma^2} + \frac{2ahe^{2ah}\sum_{t=1}^{n-1}\left(x_{th}-e^{ah}x_{th-h}\right)^2}{\left(1-e^{2ah}\right)^2\sigma^2}, \tag{1.45}$$

$$\frac{\partial l}{\partial \sigma^2} = -\frac{n-1}{2\sigma^2} - \frac{a\sum_{t=1}^{n-1}\left(x_{th}-e^{ah}x_{th-h}\right)^2}{\left(1-e^{2ah}\right)\sigma^4}. \tag{1.46}$$

Computing the second derivatives of the function (1.44), taking expectation, and simplifying, we have

$$k_{11} = -\frac{(n-1)+e^{2ah}\left[-2(n-1)+4(n-1)ah+2(n-2)a^2h^2\right]}{2a^2\left(1-e^{2ah}\right)^2}$$
$$-\frac{e^{4ah}\left[(n-1)-4(n-1)ah+2na^2h^2\right]}{2a^2\left(1-e^{2ah}\right)^2}, \tag{1.47}$$

$$k_{12} = \frac{\left(1-e^{2ah}+2ahe^{2ah}\right)(n-1)}{2a\left(1-e^{2ah}\right)\sigma^2}, \tag{1.48}$$

$$k_{22} = -\frac{n-1}{2\sigma^4}. \tag{1.49}$$

(1.47)-(1.49) are obtained by taking account the fact that $E\left[\left(x_{th}-e^{ah}x_{th-h}\right)x_{th-h}\right] = E\left[\eta_{th}x_{th-h}\right] = 0$, $E\left[\sum_{t=1}^{n-1}\left(x_{th}-e^{ah}x_{th-h}\right)^2\right] = E\left[\sum_{t=1}^{n-1}\eta_{th}^2\right] = (n-1)\sigma^2\left(e^{2ah}-1\right)/2a$, and $E\left[\sum_{t=1}^{n-1}x_{th-h}^2\right] = -(n-2)\sigma^2/2a$. Note that $E\left(x_0^2\right)=0$ since $x_0=0$.

Differentiating (1.47) and (1.48) with respect to $a$, we obtain

$$k_{11}^{(1)} = \frac{(n-1)-e^{2ah}\left[3(n-1)-2(n-1)ah+4(n-1)a^2h^2+2(n-2)a^3h^3\right]}{a^3\left(1-e^{2ah}\right)^3}$$
$$+\frac{e^{4ah}\left[3(n-1)-4(n-1)ah+4(n-1)a^2h^2-2(3n-2)a^3h^3\right]}{a^3\left(1-e^{2ah}\right)^3}$$
$$-\frac{e^{6ah}\left[(n-1)-2(n-1)ah\right]}{a^3\left(1-e^{2ah}\right)^3}, \tag{1.50}$$

$$k_{12}^{(1)} = -\frac{\left(1+e^{4ah}-2e^{2ah}-4a^2h^2e^{2ah}\right)(n-1)}{2a^2\left(1-e^{2ah}\right)^2\sigma^2}. \tag{1.51}$$

Computing the third derivatives of the function (1.17), taking expectation, and simplifying, we have

$$\begin{aligned} k_{111} &= \frac{(n-1) - 3e^{2ah}\left[(n-1) + (3n-4)a^2h^2 + (n-2)a^3h^3\right]}{a^3(1-e^{2ah})^3} \\ &+ \frac{3e^{4ah}\left[(n-1) + 2(n-2)a^2h^2 - 4(n-1)a^3h^3\right)}{a^3(1-e^{2ah})^3} \\ &- \frac{e^{6ah}\left[(n-1) - 3na^2h^2 + (n+2)a^3h^3\right]}{a^3(1-e^{2ah})^3}, \end{aligned} \tag{1.52}$$

$$k_{112} = \frac{e^{2ah}h\left[2(n-1) + (3n-4)ah\right] - e^{4ah}h\left[2(n-1) - nah\right]}{a\left(1-e^{2ah}\right)^2\sigma^2}, \tag{1.53}$$

$$k_{122} = -\frac{\left(1 - e^{2ah} + 2ahe^{2ah}\right)(n-1)}{a\left(1-e^{2ah}\right)\sigma^4}. \tag{1.54}$$

Substituting (1.47)-(1.54) into (1.15) gives

$$E(\hat{a} - a) = -\frac{3 + e^{-2ah}}{2T} - \frac{3 + e^{-2ah}}{T(n-2)} + \mathrm{O}(n^{-2}). \tag{1.55}$$

Ignoring terms of order less than $T^{-1}$ gives equation (1.18).

$\square$

*Proof of Theorem 1.3.2.* Substituting $\sigma_\eta^2$ and $\sigma_x^2$ into (1.16) gives

$$\begin{aligned} \ell(a, \sigma^2) &= -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln\left[-\frac{\sigma^2}{2a}\right] + \frac{a}{\sigma^2}x_0^2 - \frac{n-1}{2}\ln\left[-\frac{\sigma^2(1-e^{2ah})}{2a}\right] \\ &+ \frac{a\sum_{t=1}^{n-1}(x_{th} - e^{ah}x_{th-h})^2}{(1-e^{2ah})\sigma^2}. \end{aligned} \tag{1.56}$$

The partial derivatives of (1.56) with respect to $a$ and $\sigma^2$ are

$$\begin{aligned} \frac{\partial l}{\partial a} &= \frac{n\left(1-e^{2ah}\right) + 2(n-1)ahe^{2ah}}{2a\left(1-e^{2ah}\right)} + \frac{x_0^2}{\sigma^2} - \frac{2ahe^{ah}\sum_{t=1}^{n-1}\left(x_{th} - e^{ah}x_{th-h}\right)x_{th-h}}{\left(1-e^{2ah}\right)\sigma^2} \\ &+ \frac{\sum_{t=1}^{n-1}\left(x_{th} - e^{ah}x_{th-h}\right)^2}{\left(1-e^{2ah}\right)\sigma^2} + \frac{2ahe^{2ah}\sum_{t=1}^{n-1}\left(x_{th} - e^{ah}x_{th-h}\right)^2}{\left(1-e^{2ah}\right)^2\sigma^2}, \end{aligned} \tag{1.57}$$

$$\frac{\partial l}{\partial \sigma^2} = -\frac{n}{2\sigma^2} - \frac{ax_0^2}{\sigma^4} - \frac{a\sum_{t=1}^{n-1}\left(x_{th} - e^{ah}x_{th-h}\right)^2}{\left(1-e^{2ah}\right)\sigma^4}. \tag{1.58}$$

Computing the second derivatives of (1.56) and taking expectation, we have

$$k_{11} = -\frac{n + e^{2ah}\left[-2n + 4(n-1)ah + 2(n-1)a^2h^2\right]}{2a^2\left(1-e^{2ah}\right)^2} - \frac{e^{4ah}\left[n - 4(n-1)ah + 2(n-1)a^2h^2\right]}{2a^2\left(1-e^{2ah}\right)^2}, \tag{1.59}$$

$$k_{12} = \frac{n\left(1-e^{2ah}\right) + 2(n-1)ahe^{2ah}}{2a\left(1-e^{2ah}\right)\sigma^2}, \tag{1.60}$$

$$k_{22} = -\frac{n}{2\sigma^4}. \tag{1.61}$$

As before we use the fact that $E[\eta_{th}x_{th-h}] = 0$, $E\left[\sum_{t=1}^{n-1}\eta_{th}^2\right] = (n-1)\sigma^2\left(e^{2ah}-1\right)/2a$. Note that in this case $E\left[\sum_{t=1}^{n-1}x_{th-h}^2\right] = -(n-1)\sigma^2/2a$ since $E\left(x_0^2\right) = -\sigma^2/2a$.

Differentiating (1.59) and (1.60) with respect to $a$, we obtain

$$k_{11}^{(1)} = \frac{n - e^{2ah}\left[3n - 2(n-1)ah + 4(n-1)a^2h^2 + 2(n-1)a^3h^3\right]}{a^3\left(1-e^{2ah}\right)^3} + \frac{e^{4ah}\left[3n - 4(n-1)ah + 4(n-1)a^2h^2 - 6(n-1)a^3h^3\right] - e^{6ah}\left[n - 2(n-1)ah\right]}{a^3\left(1-e^{2ah}\right)^3}, \tag{1.62}$$

$$k_{12}^{(1)} = -\frac{n\left(1-e^{2ah}\right)^2 - 4(n-1)a^2h^2e^{2ah}}{2a^2\left(1-e^{2ah}\right)^2\sigma^2}. \tag{1.63}$$

Computing the third derivatives of (1.56) and taking expectation, we have

$$k_{111} = \frac{n - 3e^{2ah}\left[n + 3(n-1)a^2h^2 + (n-1)a^3h^3\right]}{a^3(1-e^{2ah})^3} + \frac{3e^{4ah}\left[n + 2(n-1)a^2h^2 - 4(n-1)a^3h^3\right)}{a^3(1-e^{2ah})^3} - \frac{e^{6ah}\left[n - 3(n-1)a^2h^2 + (n-1)a^3h^3\right]}{a^3(1-e^{2ah})^3}, \tag{1.64}$$

$$k_{112} = \frac{e^{2ah}h(n-1)\left[2+3ah\right] - e^{4ah}h(n-1)\left[2-ah\right]}{a\left(1-e^{2ah}\right)^2\sigma^2}, \tag{1.65}$$

$$k_{122} = -\frac{n\left(1-e^{2ah}\right) + 2(n-1)ahe^{2ah}}{a\left(1-e^{2ah}\right)\sigma^4}. \tag{1.66}$$

Substituting (1.59)-(1.66) into (1.15), we obtain equation (1.21).

□

*Proof of Lemma 1.3.1.* It follows from representation (1.24) that

$$E(\eta_{th}^2) = \gamma_0 = \sigma_\varepsilon^2(1+\theta^2), \tag{1.67}$$

$$E(\eta_{th}\eta_{th-h}) = \gamma_1 = \theta\sigma_\varepsilon^2. \tag{1.68}$$

Equation (1.67) implies $\sigma_\varepsilon^2 = \gamma_0/(1+\theta^2)$. Substituting into equation (1.68) we obtain $\gamma_1 = \gamma_0\theta/(1+\theta^2)$, which can be written as

$$\theta^2\rho - \theta + \rho = 0, \tag{1.69}$$

where

$$\rho = \frac{\gamma_1}{\gamma_0} = \frac{e^{2ah} - 2ahe^{ah} - 1}{2\left[1 + ah(e^{2ah}+1) - e^{2ah}\right]}. \tag{1.70}$$

The coefficient $\rho$ is positive for $a < 0$ and reaches its maximum value of 0.25 when the continuous time process has a root near zero. This last result is consistent with the findings in Working (1960) for the first differences of averages in a discrete time random walk. See also Figure 1.24, which shows $\rho$ as a function of $a$[14] for quartely (i.e. h=1/4), monthly (i.e. h=1/12) and daily (i.e. h=1/12) data.

The roots of the equation (1.69) are

$$\frac{1 \pm \sqrt{1-4\rho^2}}{2\rho}, \tag{1.71}$$

Note that the term into the radical is positive and the roots are reciprocals of each other. Assuming invertibility, $\theta$ is, therefore, the smallest root $(1 - \sqrt{1-4\rho^2})/(2\rho)$.

□

*Proof of Theorem 1.3.3.* Substituting $\sigma_\varepsilon^2$ into (1.26) gives

$$\begin{aligned}\ell(a,\sigma^2) = &-\frac{n-1}{2}\ln(2\pi) - \frac{n-1}{2}\ln\left[-\frac{\sigma^2(1+2ahe^{ah}-e^{2ah})}{2a^3\theta}\right] \\ &+ \frac{a^3\theta\sum_{t=1}^{n-1}\left[x_{th} - e^{ah}x_{th-h} - \theta\varepsilon_{th-h}\right]^2}{(1+2ahe^{ah}-e^{2ah})\sigma^2}.\end{aligned} \tag{1.72}$$

[14] For these graphs we use a wide range of possible values of $a$. The parameter $a$ takes values from the region of [-10, 0), [-30, 0) and [-630, 0) for quartely, monthly and daily data, respectively, which correspond to [0.08, 1) for the autoregressive coefficient in the discrete time model.

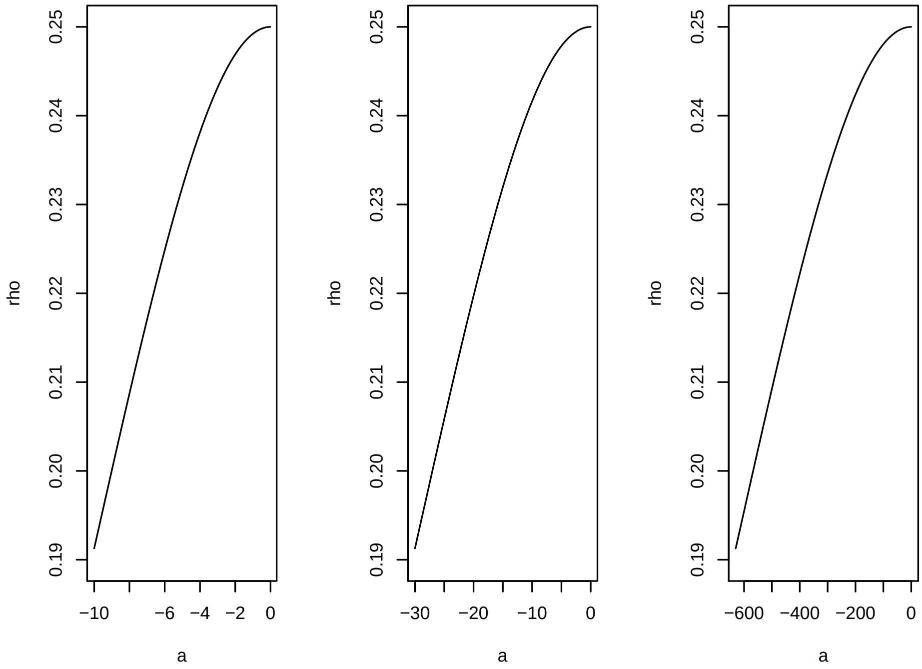


Figure 1.24: The parameter $\rho$ for quartely, monthly and daily data as a function of $a$.

It should be noted that the error $\varepsilon_{th}$ in (1.72) depends on $a$, implying that $d\varepsilon_{th}/da$ is different from zero. To see this, let us write $\varepsilon_{th}$ in terms of $x_{th}$ and its lagged values

$$\begin{aligned}\varepsilon_h &= x_h - e^{ah}x_0 - \theta\varepsilon_0,\\ \varepsilon_{th} &= x_{th} - e^{ah}\left[x_{th-h} - \theta x_{th-2h} + ... + (-1)^{t-1}\theta^{t-1}x_0\right]\\ &\quad - \theta\left[x_{th-h} - \theta x_{th-2h} + ... + (-1)^{t-2}\theta^{t-2}x_h + (-1)^{t-1}\theta^{t-1}\varepsilon_0\right],\ t = 2, ..., n-1, \quad (1.73)\end{aligned}$$

which can be written more compactly as

$$\varepsilon_{th} = x_{th} - e^{ah}\lambda_t(L^h)x_{th-h} - \theta\left[\lambda_{t-1}(L^h)x_{th-h} + (-1)^{t-1}\theta^{t-1}\varepsilon_0\right],\ t = 1, ..., n-1, \quad (1.74)$$

where $\lambda_t(L^h) = [1 - \theta L^h + ... + (-1)^{t-1}(\theta L^h)^{t-1}]$ and $\lambda_0(L^h) = 0$. In what follows we simply write $\lambda_t(L^h)$ as $\lambda_t$.

Note, too, that $\theta$ is a function of $a$, implying that $d\theta/da$ is non-zero. This derivative is, however, a complex function of $a$ and its inclusion considerably complicates calculations as will be shown in Remark 2, where we present a new bias expression for the case of fixed $x_0$. Moreover, as can be seen in Remark 1 the derivative of $\theta$ with respect to $a$ is negligible. For these reasons, in the rest of the proof we assume that $d\theta/da = 0$. Remark 2 also describes the implications of this assumption.

We now turn to the derivation of the likelihood function with respect to the parameters.

The partial derivatives of (1.72) with respect to $a$ and $\sigma^2$ are

$$\begin{aligned}
\frac{\partial l}{\partial a} =& \frac{\left(3-2ah(-2+ah)e^{ah}+(-3+2ah)e^{2ah}\right)(n-1)}{2a\left(1+2ahe^{ah}-e^{2ah}\right)} \\
&+ \frac{2a^3\theta\sum_{t=1}^{n-1}\left(h\theta e^{ah}\lambda_{t-1}x_{th-2h}-he^{ah}x_{th-h}\right)\left[x_{th}-e^{ah}x_{th-h}-\theta\varepsilon_{th-h}\right]}{\left(1+2ahe^{ah}-e^{2ah}\right)\sigma^2} \\
&+ \frac{3a^2\theta\sum_{t=1}^{n-1}\left(x_{th}-e^{ah}x_{th-h}-\theta\varepsilon_{th-h}\right)^2}{\left(1+2ahe^{ah}-e^{2ah}\right)\sigma^2} \\
&- \frac{a^3\theta(2he^{ah}-2he^{2ah}+2ah^2e^{ah})\sum_{t=1}^{n-1}\left(x_{th}-e^{ah}x_{th-h}-\theta\varepsilon_{th-h}\right)^2}{\left(1+2ahe^{ah}-e^{2ah}\right)^2\sigma^2},
\end{aligned} \tag{1.75}$$

$$\frac{\partial l}{\partial \sigma^2} = -\frac{n-1}{2\sigma^2} - \frac{a^3\theta\sum_{t=1}^{n-1}\left(x_{th}-e^{ah}x_{th-h}-\theta\varepsilon_{th-h}\right)^2}{\left(1+2ahe^{ah}-e^{2ah}\right)\sigma^4}. \tag{1.76}$$

Before computing second derivatives, we first note that the second term of $\partial l/\partial a$ depends on $\lambda_{t-1}x_{th-2h}$, which implies that two additional terms need to be calculated to obtain $k_{11}$. The first term is

$$\begin{aligned}
E[\sum_{t=1}^{n-1} x_{th-h}(\lambda_{t-1}x_{th-2h})] &= E\left[x_{2h}\lambda_2 x_h + x_{3h}\lambda_3 x_{2h} + \cdots + x_{nh-2h}\lambda_{n-2}x_{nh-3h}\right] \\
&= (n-3)\gamma_1^{(x)} - (n-4)\theta\gamma_2^{(x)} + \cdots + (-1)^{n-4}\theta^{n-4}\gamma_{n-3}^{(x)} \\
&= n\gamma_1^{(x)}[1-\theta e^{ah}+\cdots+(-1)^{n-4}\theta^{n-4}e^{ah(n-4)}] \\
&- \gamma_1^{(x)}[3-4\theta e^{ah}+\cdots+(-1)^{n-4}(n-1)\theta^{n-4}e^{ah(n-4)}] \\
&= (n-2)\gamma_1^{(x)}\left[\frac{1+(-1)^{n-4}(\theta e^{ah})^{n-3}}{1+\theta e^{ah}}\right] \\
&- (n-3)\gamma_1^{(x)}\left[\frac{(-1)^{n-4}(\theta e^{ah})^{n-3}}{1+\theta e^{ah}}\right] \\
&- \gamma_1^{(x)}\left[\frac{1+(-1)^{n-4}(\theta e^{ah})^{n-3}}{(1+\theta e^{ah})^2}\right],
\end{aligned} \tag{1.77}$$

where $\gamma_i^{(x)}$ denotes the $i$ autocovariance of $x$. Notice that $\theta^{n-2} \cong 0$. Approximating, equation (1.77) becomes

$$E[\sum_{t=1}^{n-2} x_{th-h}(\lambda_{t-1}x_{th-2h})] = \frac{(n-2)\gamma_1^{(x)}}{1+\theta e^{ah}} - \frac{\gamma_1^{(x)}}{(1+\theta e^{ah})^2}, \tag{1.78}$$

where the first autocovariance $\gamma_1^{(x)}$ is

$$\gamma_1^{(x)} = -\frac{(1+2ahe^{ah}-e^{2ah})(\theta+e^{ah})(1+\theta e^{ah})\sigma^2}{2a^3\theta(1-e^{2ah})}. \tag{1.79}$$

The second term that we have to calculate is

$$
\begin{aligned}
E[\sum_{t=1}^{n-1}(\lambda_{t-1}x_{th-2h})^2] &= E\left[(\lambda_2 x_h)^2 + (\lambda_3 x_{2h})^2 + \cdots + (\lambda_{n-2}x_{nh-3h})^2\right] \\
&= \sigma_x^2\left[(n-3) + (n-4)\theta^2 + \cdots + \theta^{2n-8}\right] \\
&- 2\gamma_1^{(x)}\left[(n-4)\theta + (n-5)\theta^3 + \cdots + \theta^{2n-9}\right] \\
&+ 2\gamma_2^{(x)}\left[(n-5)\theta^2 + (n-6)\theta^4 + \cdots + \theta^{2n-10}\right] \\
&+ \cdots + (-1)^{n-5}2\gamma_{n-5}^{(x)}\left[2\theta^{n-5} + \theta^{n-3}\right] + (-1)^{n-4}2\gamma_{n-4}^{(x)}\theta^{n-4} \\
&= (n-2)\sigma_x^2\left[\frac{1-\theta^{2n-6}}{1-\theta^2}\right] - \sigma_x^2\left[\frac{1-\theta^{2n-6}}{(1-\theta^2)^2} - (n-3)\frac{\theta^{2n-6}}{1-\theta^2}\right] \\
&- \frac{2\theta\gamma_1^{(x)}}{1-\theta^2}\left[\frac{1+(-1)^{n-5}(\theta e^{ah})^{n-4}}{1+\theta e^{ah}}\right]\left[(n-2) - \frac{1}{1+\theta e^{ah}} - \frac{1}{1-\theta^2}\right] \\
&+ \frac{2\theta^{2n-7}\gamma_1^{(x)}}{1-\theta^2}\left[\frac{1+(-1)^{n-5}(\theta^{-1}e^{ah})^{n-4}}{1+\theta^{-1}e^{ah}}\right]\left[(n-2) - \frac{1}{1-\theta^2} - (n-3)\right] \\
&+ \frac{2\theta(n-4)\gamma_1^{(x)}}{1-\theta^2}\left[\frac{(-1)^{n-5}(\theta e^{ah})^{n-4}}{1+\theta e^{ah}}\right]. \qquad (1.80)
\end{aligned}
$$

Approximating, (1.80) becomes

$$
\begin{aligned}
E[\sum_{t=1}^{n-1}(\lambda_{t-1}x_{th-2h})^2] &= \frac{(n-2)\sigma_x^2}{1-\theta^2} - \frac{\sigma_x^2}{(1-\theta^2)^2} - \frac{2\theta(n-2)\gamma_1^{(x)}}{(1-\theta^2)(1+\theta e^{ah})} + \frac{2\theta\gamma_1^{(x)}}{(1-\theta^2)(1+\theta e^{ah})^2} \\
&+ \frac{2\theta\gamma_1^{(x)}}{(1-\theta^2)^2(1+\theta e^{ah})}. \qquad (1.81)
\end{aligned}
$$

Computing the second derivatives of (1.72), taking expectation, and using (1.78) and (1.81), we obtain

$$
\begin{aligned}
k_{11} &= -\frac{\left[3 + 2ah(2-ah)e^{ah} - (3-2ah)e^{2ah}\right]^2(n-1)}{2a^2\left(1 + 2ahe^{ah} - e^{2ah}\right)^2} \\
&+ \frac{h^2e^{2ah}\left[2 - 3\theta^2 - \theta^3e^{ah} + n(1+\theta e^{ah})(-1+\theta^2)\right]}{(-1+e^{2ah})(1+\theta e^{ah})(-1+\theta^2)}, \qquad (1.82)
\end{aligned}
$$

$$
k_{12} = \frac{\left(3 + 2ah(2-ah)e^{ah} - (3-2ah)e^{2ah}\right)(n-1)}{2a\left(1 + 2ahe^{ah} - e^{2ah}\right)\sigma^2}, \qquad (1.83)
$$

$$
k_{22} = -\frac{n-1}{2\sigma^4}. \qquad (1.84)
$$

Differentiating (1.82) and (1.83) with respect to $a$ gives

$$\begin{aligned}
k_{11}^{(1)} &= \frac{2h^3e^{2ah}\left[-2+3\theta^2-n(\theta^2-1)(1+\theta e^{ah})^2+\theta(3\theta^2-1)e^{ah}+\theta^4e^{2ah}+\theta(\theta^2-1)e^{3ah}\right]}{(-1+e^{2ah})^2(1+e^{ah}\theta)^2(\theta^2-1)} \\
&+ \frac{(n-1)\left[9+6ah(8+ah+a^2h^2)e^{ah}+(-27+6ah+60a^2h^2-8a^3h^3-4a^5h^5)e^{2ah}\right]}{a^3(1+2ahe^{ah}-e^{2ah})^3} \\
&- \frac{(n-1)\left[4ah(24-9ah-3a^2h^2+a^3h^3)e^{3ah}+2ah(24-21ah+7a^2h^2-2a^3h^3)e^{5ah}\right]}{a^3(1+2ahe^{ah}-e^{2ah})^3} \\
&+ \frac{(n-1)\left[(27-12ah-60a^2h^2+48a^3h^3-16a^4h^4+4a^5h^5)e^{4ah}-(9-6ah)e^{6ah}\right]}{a^3(1+2ahe^{ah}-e^{2ah})^3},
\end{aligned} \tag{1.85}$$

$$\begin{aligned}
k_{12}^{(1)} &= -\frac{(n-1)\left[3+2ah(6+2ah+a^2h^2)e^{ah}-(6-4a^2h^2)e^{2ah}\right]}{2a^2\left(1+2ahe^{ah}-e^{2ah}\right)^2\sigma^2} \\
&- \frac{(n-1)\left[-2ah(6-2ah+a^2h^2)e^{3ah}+3e^{4ah}\right]}{2a^2\left(1+2ahe^{ah}-e^{2ah}\right)^2\sigma^2}.
\end{aligned} \tag{1.86}$$

Computing the third derivatives of (1.72), taking expectation and simplifying, we have

$$\begin{aligned}
k_{111} &= \frac{3h^2e^{2ah}\left[3+ah+4ahe^{ah}-(3-ah)e^{2ah}\right]\left[2-3\theta^2-\theta^3e^{ah}+n(1+\theta e^{ah})(\theta^2-1)\right]}{a(-1+e^{2ah})(1+\theta e^{ah})(\theta^2-1)(1+2ahe^{ah}-e^{2ah})} \\
&+ \frac{(n-1)he^{ah}\left[9(2+4ah+a^2h^2)-6(3-3ah-10a^2h^2+3a^3h^3+a^4h^4)e^{ah}\right]}{a^2(1+2ahe^{ah}-e^{2ah})^3} \\
&+ \frac{(n-1)he^{ah}\left[(-36-72ah+22a^2h^2+42a^3h^3-24a^4h^4+4a^5h^5)e^{2ah}\right]}{a^2(1+2ahe^{ah}-e^{2ah})^3} \\
&+ \frac{(n-1)he^{ah}\left[6(6-6ah-8a^2h^2+7a^3h^3-a^4h^4)e^{3ah}-2(9-9ah+2a^2h^2)e^{5ah}\right]}{a^2(1+2ahe^{ah}-e^{2ah})^3} \\
&+ \frac{(n-1)he^{ah}\left[3(6+12ah-13a^2h^2+2a^3h^3)e^{4ah}\right]}{a^2(1+2ahe^{ah}-e^{2ah})^3},
\end{aligned} \tag{1.87}$$

$$\begin{aligned}
k_{112} &= \frac{(n-1)\left[3+ah(6-8ah-a^2h^2)e^{ah}-2(3-3ah-3a^2h^2+4a^3h^3-a^4h^4)e^{2ah}\right]}{a^2\left(1+2ahe^{ah}-e^{2ah}\right)^2\sigma^2} \\
&+ \frac{(n-1)\left[-3ah(2-4ah+a^2h^2)e^{3ah}+(3-6ah+2a^2h^2)e^{4ah}\right]}{a^2\left(1+2ahe^{ah}-e^{2ah}\right)^2\sigma^2} \\
&- \frac{h^2e^{2ah}\left[2-3\theta^2-\theta^3e^{ah}+n(1+\theta e^{ah})(-1+\theta^2)\right]}{(-1+e^{2ah})(1+\theta e^{ah})(\theta^2-1)\sigma^2},
\end{aligned} \tag{1.88}$$

$$k_{122} = -\frac{(n-1)\left(3+2ah(2-ah)e^{ah}-(3-2ah)e^{2ah}\right)}{a\left(1+2ahe^{ah}-e^{2ah}\right)\sigma^4}. \tag{1.89}$$

Substituting (1.82)-(1.89) into (1.15) gives

$$E(\hat{a}-a) = -\frac{3+e^{-2ah}}{2T} - \frac{A}{2TB^2} + \mathrm{O}(n^{-2}), \tag{1.90}$$

where

$$\begin{aligned}A &= n(\theta^2-1)\left[(3+e^{-2ah})(-2+3\theta^2-\theta e^{ah}+3\theta^3e^{ah}+\theta^4e^{2ah})-(\theta^2-1)\theta(e^{ah}-5e^{-ah})\right]\\ &\quad -(3+e^{-2ah})(-2+3\theta^2+\theta^3e^{ah})^2,\\ B &= n(\theta^2-1)(1+\theta e^{ah})+2-3\theta^2-\theta^3e^{ah}.\end{aligned}$$

Ignoring terms of order less than $T^{-1}$ gives equation (1.27).

□

*Remark 1.* From Lemma 1.3.1 we observe that $\theta$ is a function of $a$, but not of $\sigma^2$. Figures 1.25 and 1.26 show $\theta$ for different values of $a$[15] for quartely (i.e. $h = 1/4$) and monthly (i.e. $h = 1/12$) data and the first two derivatives of $\theta$ with respect to $a$, which are given by

$$\frac{d\theta}{da} = \frac{\theta}{\rho\sqrt{1-4\rho^2}}\frac{d\rho}{da}, \tag{1.91}$$

$$\frac{d^2\theta}{da^2} = \frac{1}{\rho\sqrt{1-4\rho^2}}\left[\frac{d\theta}{da}\frac{d\rho}{da} - \frac{\theta}{\rho}\left(\frac{d\rho}{da}\right)^2 + \frac{4\theta\rho}{1-4\rho^2}\left(\frac{d\rho}{da}\right)^2 + \theta\frac{d^2\rho}{da^2}\right], \tag{1.92}$$

where

$$\frac{d\rho}{da} = -\frac{(e^{ah}-1)h\left[1-e^{ah}(1+2ah+2a^2h^2)+e^{2ah}(-1+2ah-2a^2h^2)+e^{3ah}\right]}{2\left[1+ah+e^{2ah}(-1+ah)\right]^2}, \tag{1.93}$$

$$\begin{aligned}\frac{d^2\rho}{da^2} &= -\frac{h^2\left[1+ahe^{ah}(3+2ah+a^2h^2)-e^{2ah}(3+4a^2h^2)-6ahe^{3ah}(1+a^2h^2)\right]}{\left[1+ah+e^{2ah}(-1+ah)\right]^3}\\ &\quad -\frac{h^2\left[e^{4ah}(3+4a^2h^2)+ahe^{5ah}(3-2ah+a^2h^2)-e^{6ah}\right]}{\left[1+ah+e^{2ah}(-1+ah)\right]^3}.\end{aligned} \tag{1.94}$$

We observe that the first two derivatives of $\theta$ with respect to $a$ are negligible, being smaller for high frequency data. In fact, applying L'Hôpital's rule it can be seen that $d\theta/da \to 0$ when $h \to 0$ or $a \to 0$ while $d^2\theta/da^2 \to 0$ when $h \to 0$ and $d^2\theta/da^2 \to (3-2\sqrt{3})h^2/15$ when $a \to 0$.

[15] For these graphs we use a wide range of possible values of $a$. The parameter $a$ takes values from the region of [-10, 0) and [-30, 0) for quartely and monthly data, respectively, which correspond to [0.08, 1) for the autoregressive coefficient in the discrete time model.

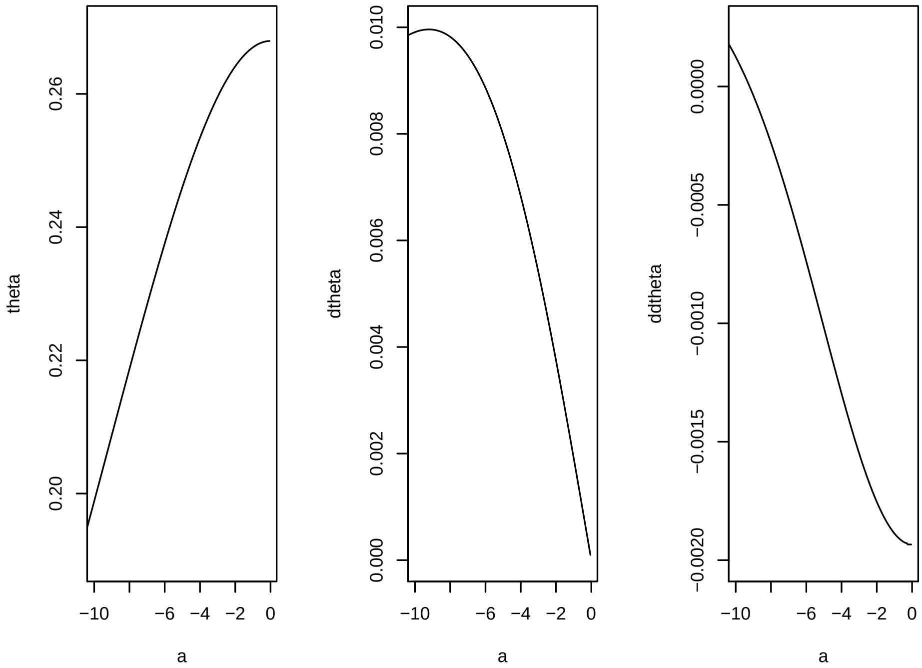


Figure 1.25: The parameter $\theta$, $d\theta/da$ and $d^2\theta/da^2$ for quartely data as a function of $a$.

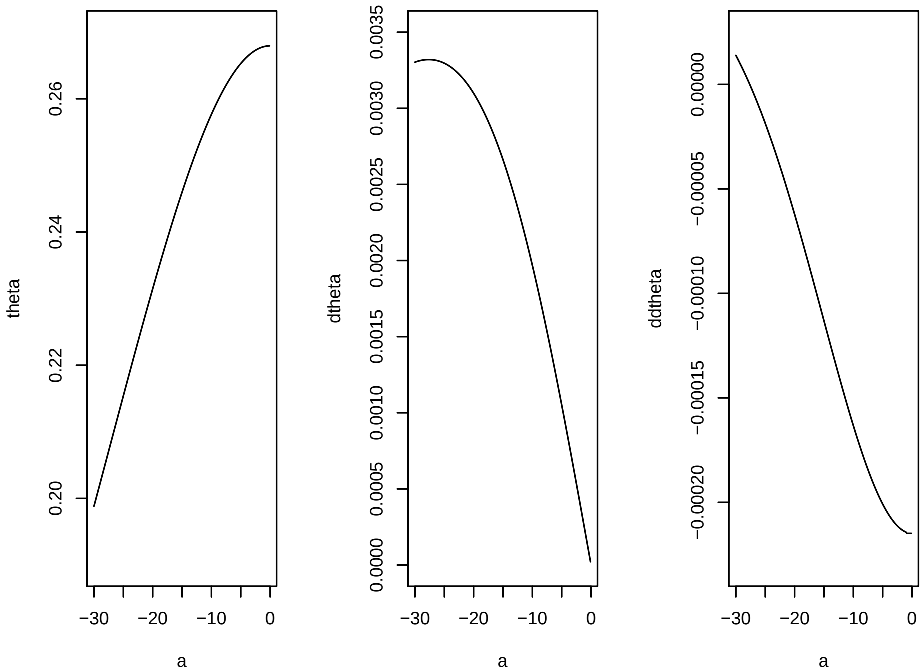


Figure 1.26: The parameter $\theta$, $d\theta/da$ and $d^2\theta/da^2$ for monthly data as a function of $a$.

*Remark 2: Bias expression without assuming $d\theta/da = 0$.* Under model given in (1.2) and (1.24) for $t = 1, ..., n-1$ with a fixed initial condition $\tilde{x}_0 = 0$, the bias of $\hat{a}$ is

$$E(\hat{a} - a) = -\frac{(\theta^2 - 1)E}{2TF^2} + \mathrm{O}(n^{-2}), \tag{1.95}$$

where

$$\begin{aligned}
E = {} & n^2 \left[ e^{2ah}(\theta^2 - 1)(1 + \theta e^{ah})^2(1 + 3e^{2ah}) \right] + n \left[ 2 - 3\theta^2 - 2\theta e^{ah} + (6 - 9\theta^2 - \theta^4)e^{2ah} \right. \\
& \left. + 2\theta(5 - 8\theta^2)e^{3ah} - 3\theta^4 e^{4ah} \right] \\
& + n^2 \left[ (1 + e^{ah}\theta)^2(\theta^2 - 1) \left( \frac{d\theta}{da} e^{ah}(1 - 5\theta e^{ah} + 4e^{2ah} - 3\theta e^{3ah} - 5e^{4ah}) \right. \right. \\
& + \left( \frac{d\theta}{da} \right)^2 2e^{ah}(-3\theta + (1 + 2\theta^2)e^{ah} + 3\theta e^{2ah} - e^{3ah}) + \left( \frac{d\theta}{da} \right)^3 2\theta(e^{2ah} - 1) \\
& \left. \left. + \frac{d^2\theta}{da}(e^{2ah} - 1)(-e^{ah} + e^{3ah} + \theta e^{2ah}) + \frac{d^2\theta}{da}\frac{d\theta}{da}(e^{2ah} - 1)(1 + \theta^2 - e^{2ah}) \right) \right]
\end{aligned}$$

and

$$\begin{aligned}
F = {} & n \left[ e^{2ah}(1 + e^{ah}\theta)(\theta^2 - 1) \right] + e^{2ah} \left[ 2 - 3\theta^2 - \theta^3 e^{ah} \right] \\
& + n \left[ (1 + e^{ah}\theta)(\theta^2 - 1) \left( -\frac{d\theta}{da} 2e^{ah}(-1 + \theta e^{ah} + e^{2ah}) + \left( \frac{d\theta}{da} \right)^2 (1 + \theta^2 - e^{2ah}) \right) \right].
\end{aligned}$$

When $a \to 0$ the bias reduces to

$$E(\hat{a} - a) = -\frac{2}{T} + \mathrm{o}(T^{-1}), \tag{1.96}$$

and when $h \to 0$ it becomes

$$E(\hat{a} - a) = -\frac{2}{T} + \mathrm{o}(T^{-1}). \tag{1.97}$$

$\square$

To save space we have ignored terms of order less than $O(n)$ in expressions $E$ and $F$. To see the impact of dropping the additional terms, which are of order $O(n^0) = O(1)$, we plot the bias expression (1.95) and that including the additional terms (see Figure 1.27). We observe that the two expressions are almost the same. For comparison purpose we also plot the bias expression (1.27). Figure 1.27 shows that there is a considerable discrepancy between (1.95) and (1.27) for large values of $a$. This gap reduces, however, when $a$ becomes smaller. Indeed, (1.28) and (1.96) suggest that the bias expressions are identical when $a \to 0$. Similarly, (1.29) and (1.97) indicate the bias expressions are identical when $h \to 0$. These results suggest that although the bias expressions (1.27) and

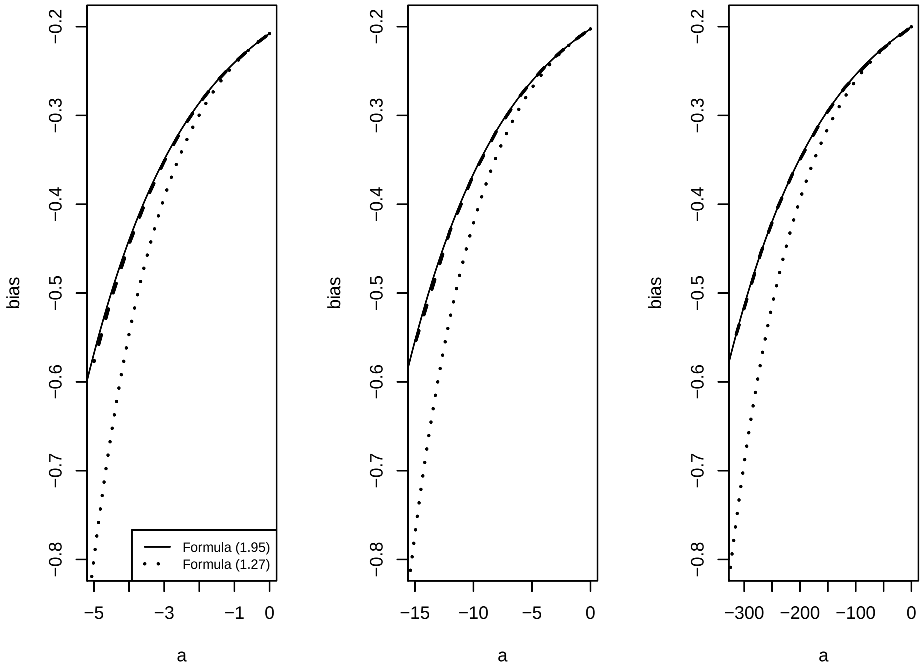


Figure 1.27: The bias of $\hat{a}$ for flow data and $T = 10$ with a fixed initial condition as a function of $a$. The graphs correspond to quartely, monthly and daily data (i.e. $n = 40, 120, 2520$), respectively. The solid line is the approximate bias according to the formula (1.95) and the dashed line is the bias expression that includes terms of order $O(1)$. The dotted line is the approximate bias according to the formula (1.27).

(1.30) provide a good approximation for a range of parameter values that would appear to be relevant in practice, the quality of the approximation deteriorates as $a$ increases.

$\square$

*Proof of Theorem 1.3.4.* Substituting $\sigma_\varepsilon^2$ and $\sigma_x^2$ into (1.25) gives

$$\begin{aligned}\ell(a,\sigma^2) = &-\frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln\left[-\frac{\sigma^2(1+2ahe^{ah}-e^{2ah})(1+2\theta e^{ah}+\theta^2)}{2a^3\theta(1-e^{2ah})}\right] \\ &+ \frac{a^3\theta(1-e^{2ah})x_0^2}{(1+2ahe^{ah}-e^{2ah})(1+2\theta e^{ah}+\theta^2)\sigma^2} - \frac{n-1}{2}\ln\left[-\frac{\sigma^2(1+2ahe^{ah}-e^{2ah})}{2a^3\theta}\right] \\ &+ \frac{a^3\theta\sum_{t=1}^{n-1}\left[x_{th}-e^{ah}x_{th-h}-\theta(x_{th-h}-e^{ah}\lambda_{t-1}x_{th-2h}-\theta\lambda_{t-1}x_{th-2h})\right]^2}{(1+2ahe^{ah}-e^{2ah})\sigma^2}. \end{aligned} \tag{1.98}$$

The rest of the proof is very similar to that of Theorem 1.3.3, but longer. To save space we only present the equations equivalent to (1.77) and (1.80) and the corresponding bias expression.

$$
\begin{aligned}
E[\sum_{t=1}^{n-1} x_{th-h}(\lambda_{t-1}x_{th-2h})] &= E\left[x_h\lambda_1 x_0 + x_{2h}\lambda_2 x_h + \cdots + x_{nh-2h}\lambda_{n-2}x_{nh-3h}\right] \\
&= (n-2)\gamma_1^{(x)} - (n-3)\theta\gamma_2^{(x)} + \cdots + (-1)^{n-3}\theta^{n-3}\gamma_{n-2}^{(x)} \\
&= n\gamma_1^{(x)}[1-\theta e^{ah} + \cdots + (-1)^{n-3}\theta^{n-3}e^{ah(n-3)}] \\
&- \gamma_1^{(x)}[2-3\theta e^{ah} + \cdots + (-1)^{n-3}(n-1)\theta^{n-3}e^{ah(n-3)}] \\
&= (n-1)\gamma_1^{(x)}\left[\frac{1+(-1)^{n-3}(\theta e^{ah})^{n-2}}{1+\theta e^{ah}}\right] \\
&- (n-2)\gamma_1^{(x)}\left[\frac{(-1)^{n-3}(\theta e^{ah})^{n-2}}{1+\theta e^{ah}}\right] \\
&- \gamma_1^{(x)}\left[\frac{1+(-1)^{n-3}(\theta e^{ah})^{n-2}}{(1+\theta e^{ah})^2}\right].
\end{aligned}
\tag{1.99}
$$

Approximating, equation (1.99) becomes

$$
E[\sum_{t=1}^{n-1} x_{th-h}(\lambda_{t-1}x_{th-2h})] = \frac{(n-1)\gamma_1^{(x)}}{1+\theta e^{ah}} - \frac{\gamma_1^{(x)}}{(1+\theta e^{ah})^2}. \tag{1.100}
$$

$$
\begin{aligned}
E[\sum_{t=1}^{n-1}(\lambda_{t-1}x_{th-2h})^2] &= E\left[(\lambda_1 x_0)^2 + (\lambda_2 x_h)^2 + \cdots + (\lambda_{n-2}x_{nh-3h})^2\right] \\
&= \sigma_x^2\left[(n-2) + (n-3)\theta^2 + \cdots + \theta^{2n-6}\right] \\
&- 2\gamma_1^{(x)}\left[(n-3)\theta + (n-4)\theta^3 + \cdots + \theta^{2n-7}\right] \\
&+ 2\gamma_2^{(x)}\left[(n-4)\theta^2 + (n-5)\theta^4 + \cdots + \theta^{2n-8}\right] \\
&+ \cdots + (-1)^{n-4}2\gamma_{n-4}^{(x)}\left[2\theta^{n-4} + \theta^{n-2}\right] + (-1)^{n-3}2\gamma_{n-3}^{(x)}\theta^{n-3} \\
&= (n-1)\sigma_x^2\left[\frac{1-\theta^{2n-4}}{1-\theta^2}\right] - \sigma_x^2\left[\frac{1-\theta^{2n-4}}{(1-\theta^2)^2} - (n-2)\frac{\theta^{2n-4}}{1-\theta^2}\right] \\
&- \frac{2\theta\gamma_1^{(x)}}{1-\theta^2}\left[\frac{1+(-1)^{n-4}(\theta e^{ah})^{n-3}}{1+\theta e^{ah}}\right]\left[(n-1) - \frac{1}{1+\theta e^{ah}} - \frac{1}{1-\theta^2}\right] \\
&+ \frac{2\theta^{2n-5}\gamma_1^{(x)}}{1-\theta^2}\left[\frac{1+(-1)^{n-4}(\theta^{-1}e^{ah})^{n-3}}{1+\theta^{-1}e^{ah}}\right]\left[(n-1) - \frac{1}{1-\theta^2} - (n-2)\right] \\
&+ \frac{2\theta(n-3)\gamma_1^{(x)}}{1-\theta^2}\left[\frac{(-1)^{n-4}(\theta e^{ah})^{n-3}}{1+\theta e^{ah}}\right].
\end{aligned}
\tag{1.101}
$$

Approximating, (1.101) becomes

$$E[\sum_{t=1}^{n-1}(\lambda_{t-1}x_{th-2h})^2] = \frac{(n-1)\sigma_x^2}{1-\theta^2} - \frac{\sigma_x^2}{(1-\theta^2)^2} - \frac{2\theta(n-1)\gamma_1^{(x)}}{(1-\theta^2)(1+\theta e^{ah})} + \frac{2\theta\gamma_1^{(x)}}{(1-\theta^2)(1+\theta e^{ah})^2} + \frac{2\theta\gamma_1^{(x)}}{(1-\theta^2)^2(1+\theta e^{ah})}. \tag{1.102}$$

The bias of $\hat{a}$ is

$$E(\hat{a}-a) = -\frac{3+e^{-2ah}}{2T} + \frac{e^{-2ah}C}{2TD^2} + \mathrm{O}(n^{-2}), \tag{1.103}$$

where

$$\begin{aligned}
C = {} & n^3[(\theta^2-1)(e^{2ah}-1)(1+\theta^2+2\theta e^{ah})^2(-1+2\theta^2+\theta^4+2\theta^6+2\theta^3(3+2\theta^2+3\theta^4)e^{ah} \\
& + (4+21\theta^2+2\theta^4+5\theta^6)e^{2ah} + 4\theta(9+16\theta^2-7\theta^4-6\theta^6)e^{3ah} + (5+96\theta^2+39\theta^4-94\theta^6 \\
& - 6\theta^8)e^{4ah} + 2\theta(6+37\theta^2-36\theta^4-15\theta^6)e^{5ah} - \theta^2(7+28\theta^2+27\theta^4+2\theta^6)e^{6ah} - 16\theta^3(2 \\
& + \theta^2)e^{7ah} + 2\theta^4(-7+\theta^2)e^{8ah})] + n^2[1-6\theta^2+2\theta^4+4\theta^6+9\theta^8+2\theta^{10}+4\theta^{12}-2\theta(5 \\
& + 9\theta^2-28\theta^4-28\theta^6-9\theta^8+3\theta^{10})e^{ah} - (9+53\theta^2-136\theta^4-266\theta^6-67\theta^8+101\theta^{10} \\
& + 50\theta^{12})e^{2ah} - 2\theta(17-41\theta^2-335\theta^4-122\theta^6+147\theta^8+115\theta^{10}+11\theta^{12})e^{3ah} + \theta(3 \\
& + 79\theta^2+680\theta^4+730\theta^6-597\theta^8-433\theta^{10}-62\theta^{12})e^{4ah} + 2\theta(37+277\theta^2+389\theta^4-374\theta^6 \\
& - 425\theta^8+41\theta^{10}+23\theta^{12})e^{5ah} + \theta(21+357\theta^2+706\theta^4-1042\theta^6-1623\theta^8+297\theta^{10} \\
& + 384\theta^{12}+4\theta^{14})e^{6ah} + 2\theta(65+311\theta^2+367\theta^4-1294\theta^6+63\theta^8+575\theta^{10}+39\theta^{12})e^{7ah} \\
& + \theta^2(311+174\theta^2-2294\theta^4-760\theta^6+1795\theta^8+298\theta^{10}+12\theta^{12})e^{8ah} + 2\theta^3(180-285\theta^2 \\
& - 614\theta^4+680\theta^6+250\theta^8+29\theta^{10})e^{9ah} + 2\theta^4(119-194\theta^2+168\theta^4+170\theta^6+57\theta^8)e^{10ah} \\
& + 8\theta^5(19+5\theta^2-\theta^4+13\theta^6)e^{11ah} + 12\theta^6(7-6\theta^2+3\theta^4)e^{12ah} - 4n(e^{ah}+\theta)^2(1+e^{ah}\theta)^3 \\
& \times (\theta^2-1)[1-2\theta^2-\theta^4-2\theta^6-\theta(3+8\theta^2+\theta^4)e^{ah} + (-2-24\theta^2+6\theta^6)e^{2ah} - 2\theta(9 \\
& + 17\theta^2-10\theta^4-2\theta^6)e^{3ah} + (-7-38\theta^2+3\theta^4+30\theta^6)e^{4ah} + \theta(-19-8\theta^2+51\theta^4+4\theta^6) \\
& \times e^{5ah} + 2\theta^2(-4+19\theta^2+3\theta^4)e^{6ah} + 2\theta^3(5+\theta^2)e^{7ah}] + 4(1+3e^{2ah})(e^{ah}+\theta)^4 \\
& \times (1+e^{ah}\theta)^6(\theta^2-1)^2,
\end{aligned}$$

$$\begin{aligned}
D = {} & n^2[(e^{2ah}-1)(\theta^2-1)(1+\theta e^{ah})(1+2\theta e^{ah}+\theta^2)^2] + n[-1+2\theta^2+\theta^4+2\theta^6 \\
& + (\theta+8\theta^3+3\theta^5)e^{ah} + (3+10\theta^2+3\theta^4-8\theta^6)e^{2ah} + (9\theta+4\theta^3-19\theta^5-2\theta^7)e^{3ah} \\
& - 2\theta^2(-3+7\theta^2+2\theta^4)e^{4ah} - 2\theta^3(1+\theta^2)e^{5ah} + 2(\theta^2-1)(\theta+e^{ah})^2(1+\theta e^{ah})^3.
\end{aligned}$$

□

*Proof of Theorem 1.4.1.* Computing the inverse of the Fisher Information matrix for the continuous time parameters $a$ and $\sigma^2$ and extracting the entry (1,1), we have

Under model given in (1.2)-(1.5) for $t = 1, ..., n-1$ with a fixed initial condition $x_0 = 0$, the asymptotic variance of $\hat{a}$ is

$$Var(\hat{a}) = -\frac{1 - e^{-2ah}}{Th} - \frac{2(1 - e^{-2ah})}{T(n-2)h}. \tag{1.104}$$

Under model given in (1.2)-(1.5) for $t = 1, ..., n-1$ with a random Gaussian initial condition $x_0$ with mean 0 and variance $-\sigma^2/2a$, the asymptotic variance of $\hat{a}$ is

$$Var(\hat{a}) = -\frac{1 - e^{-2ah}}{Th} - \frac{(1 - e^{-2ah})[(3n-2)e^{2ah} - n]}{T(n-1)h[(n-2)e^{2ah} - n]}. \tag{1.105}$$

Under model given in (1.2) and (1.24) for $t = 1, ..., n-1$ with a fixed initial condition $x_0 = 0$, the asymptotic variance of $\hat{a}$ is

$$Var(\hat{a}) = -\frac{1 - e^{-2ah}}{Th} - \frac{(1 - e^{-2ah})[-2 + 3\theta^2 + \theta^3 e^{ah}]}{Th[n(\theta^2 - 1)(\theta e^{ah} + 1) + 2 - 3\theta^2 - \theta^3 e^{ah}]}. \tag{1.106}$$

Under model given in (1.2) and (1.24) for $t = 1, ..., n-1$ with a random Gaussian initial condition $x_0 = 0$ with mean 0 and variance $\sigma^2(1 + ah - e^{ah})/a^3$, the asymptotic variance of $\hat{a}$ is

$$Var(\hat{a}) = -\frac{1 - e^{-2ah}}{Th} - \frac{(1 - e^{-2ah})(nF + G)}{Th(n^2 H + nF + G)}, \tag{1.107}$$

where

$$\begin{aligned}
F &= 1 - 2\theta^2 - \theta^4 - 2\theta^6 - \theta(1 + 8\theta^2 + 3\theta^4)e^{ah} - (3 + 10\theta^2 + 3\theta^4 - 8\theta^6)e^{2ah} - \theta(9 + 4\theta^2 \\
&\quad - 19\theta^4 - 2\theta^6)e^{3ah} - 2\theta^2(3 - 7\theta^2 - 2\theta^4)e^{4ah} + 2\theta^3(1 + \theta^2)e^{5ah}, \\
G &= -2(\theta + eah)^2(1 + \theta e^{ah})^3(\theta^2 - 1), \\
H &= (e^{2ah} - 1)(1 + \theta e^{ah})(\theta^2 - 1)(1 + 2\theta e^{ah} + \theta^2)^2.
\end{aligned}$$

Ignoring terms of order less than $T^{-1}$ gives equation (1.33).

$\square$

*Proof of Theorem 1.5.1.* Applying natural logarithm to $\hat{\phi}$ and using the preservation of convergence for continuous transformations (i.e. plim $\ln(\hat{\phi}) = \ln(\text{plim } \hat{\phi})$), we have

$$\begin{aligned}
\text{plim } \hat{a} &= \frac{1}{h}\ln\left(\text{plim } e^{\hat{a}h}\right) \\
&= \frac{1}{h}\ln\left[e^{ah} + \frac{-1 - 2ahe^{ah} + e^{2ah}}{2(1 + ah - e^{ah})}\right] \\
&= \frac{1}{h}\ln\left[\frac{-(e^{ah} - 1)^2}{2(1 + ah - e^{ah})}\right]. \qquad (1.108)
\end{aligned}$$

Subtracting $a$ in both sides of the equation and noting that $a$ can be written as $h^{-1}\ln e^{ah}$, we obtain (1.36).

□

# Chapter 2

# Bias reduction methods in continuous time models with stock and flow variables

## 2.1 Introduction

It is well known that the maximum likelihood (ML) estimator in continuous time models may be seriously biased in finite samples, being especially severe when the continuous time process is nearly a zero root process. For example, Yu (2012) showed that the bias of the ML estimator for the mean reversion parameter in the Vasicek model is about 250% when the speed of mean reversion is slow and a three-year time interval is considered, regardless of the sample frequency. The bias can be corrected if an explicit expression for the bias function or the first terms in an asymptotic expansion of the bias can be obtained analytically. For standard models, a bias expansion is often available. For instance, Tang and Chen (2009) developed expansions for the bias of parameter estimators for the Vasicek (Vasicek, 1977), CIR (Cox et al., 1985) and general linear drift diffusion processes. They showed that the first term in the bias expansion of the mean reversion estimator is of order $1/T$, where $T$ is the data span.

A simple bias correction can be computed by subtracting the first order term of the bias expansion from the parameter estimate. The approach has the advantage of removing the first order bias without altering the variance. A drawback of this correction method is that it is not applicable to models for which an asymptotic expansion is not available. In many cases, even the first terms of a bias expansion are difficult to compute explicitly as in the case of complex models containing flow variables, or the leading terms may be analytically involved. Because of these difficulties, alternative methods, which do not rely on the explicit form of an asymptotic expansion, could be preferred to implement in practice. The main purpose of this paper is, therefore, to compare some bias reduction techniques in the context of continuous time models.

Among these bias reduction techniques is the jackknife, which combines estimates obtained by application of a given method to the whole sample and sub-samples, in a way that eliminates the first order bias. One drawback of the jackknife is that it tends to inflate the variance of the estimator. Various studies have shown, however, that a carefully designed method may reduce the finite sample bias without any loss of asymptotic efficiency. For example, in the context of continuous time models, the jackknife has been applied to maximum likelihood estimators and directly to option price by Phillips and Yu (2005), who showed that this technique not only leads to substantial bias reductions, but also these reductions may be accomplished without compromising the gains by much larger variability. The main attraction of the jackknife is that it is computationally cheap to implement.

Another resampling method commonly used for bias reduction is the bootstrap. As opposed to the jackknife, it has been found that the bootstrap can deliver substantial bias reductions without inflating the variance (see for example, Tang and Chen, 2009). Because it depends on new samples drawn from the original sample, this technique can take into account different features of the data such as dependent observations or heteroskedasticity without making specific assumptions on the structure of the data generating process.

A third method, which can be effective for bias reduction, is indirect inference. Similar to the bootstrap, indirect inference is computer intensive because it makes use of Monte Carlo simulations. Its advantages not only for estimating models with intractable likelihood functions, but also for correcting the bias have been highlighted in the literature. For example, in the context of CIR models, it was shown by Phillips and Yu (2009) that indirect inference estimation can be very successful in removing bias without increasing the variance. Indirect inference uses an auxiliary model for the estimation stage, which does not need to be correctly specified. This auxiliary model becomes particularly convenient when the initial model is difficult to estimate as in the case of continuous time models with correlated discrete time disturbances.

Some comparisons of these bias correction techniques for continuous time processes have been conducted in recent years. For example, Tang and Chen (2009) found that indirect inference and bootstrap exhibit a similar performance, although the first method tends to have smaller variance but larger bias than the second one. These authors also provide evidence to support the superiority of these two methods in terms of root mean square error (RMSE) as compared to the jackknife. A Monte Carlo experiment conducted in Phillips and Yu (2009) reveals that indirect inference is more effective on removing the bias than the jackknife. These bias reduction techniques have been implemented on continuous time models in finance, with diffusion processes being the main focus. The most important diffusion processes, which have been commonly used to describe the dynamics of asset prices, including stock prices, interest rates and exchange rates, allow for instantaneously observed variables and are formulated as a first order stochastic differential equation.

Many economic variables such as GDP, consumption and exports are measured as the accumulation of the underlying rate of flow over a given time interval. The main difference of this type of variable as compared to instantaneously observed ones is that the temporal aggregation occurring with flows induces serial correlation in the discrete time disturbances. In addition, economic variables often exhibit a more complex correlation structure, which could be more accurately described by using higher order differential equations. Orders greater than one also induce correlation in the discrete time errors. Contrary to diffusion processes, bias reduction methods in the presence of correlated discrete time disturbances have not received attention in the continuous time literature. This paper considers two cases where this situation occurs, the former is a continuous time AR(1) with the variable of interest assumed to be a flow, and the latter is a continuous time AR(2) with the variable assumed to be a stock.

In the last decades much work has been done on estimating continuous time models in the presence of correlated errors (see for example, Bergstrom, 1990). Working with serially correlated disturbances is, however, more difficult econometrically than working with white noise disturbances. In particular, the likelihood function becomes more complex. In this paper, we show that indirect inference is very convenient for estimating continuous time models because it not only corrects the finite sample bias arising from estimation, but also can be successfully used to avoid the complications of working with correlated errors. We show that the indirect inference estimator based on the likelihood function of an approximate model with white noise disturbances has little bias. We also show that unlike indirect inference, the other bias reduction methods are not able to deliver substantial bias reductions when the serial correlation in the disturbances is not taken into account.

The paper is organised as follows. Section 2 briefly presents the continuous time model for the univariate and multivariate cases and the bias corrected estimators computed by using analytical bias expressions. Section 3 describes the jacknnife, indirect inference and bootstrap methods of bias reduction, and examines the finite sample properties of these three estimators by providing stochastic expansions. Section 4 explores the finite sample performance of the four bias corrected estimators using a Monte Carlo experiment. The robustness of the results to departures from normal, homoskedastic and white noise errors are also explored. Section 5 considers an empirical application to the broad effective exchange rate series for euro area and Section 6 concludes.

## 2.2 Continuous time model

### 2.2.1 Univariate model

We consider the model defined by,

$$dx(t) = [ax(t) + b]dt + \zeta(dt),\ t > 0, \tag{2.1}$$

where $x(t)$ is a scalar continuous time process, $x(0)$ is given, $a$ is a negative scalar parameter and $\zeta(dt)$ is white noise with mean zero and variance $\sigma^2 dt$. Since $x(t)$ is observable only at discrete time points, the parameters of the continuous time model must be estimated from discrete data. Assuming equispaced discrete data observed at $(h, 2h, ..., nh)$,[1] the exact discrete time model corresponding to (2.1) is

$$x_{th} = e^{ah}x_{th-h} + c + \eta_{th},\ t = 2, ..., n. \tag{2.2}$$

The constant $c$, the first observation $x_h$ and the discrete time disturbance $\eta_{th}$ take different forms depending on the way in which the variable is sampled, as a stock or as a flow. When it is a stock, so that the sequence $\{x_h, x_{2h}, ..., x_{nh}\} = \{x(h), x(2h), ..., x(nh)\}$ is observed at points of time, the constant is $c = (1 - e^{ah})\mu$ with $\mu = -b/a$, the first observation is

$$x_h = e^{ah}x(0) - \frac{b}{a}(1 - e^{ah}) + \eta_h, \tag{2.3}$$

and the discrete time disturbance given by

$$\eta_{th} = \int_{th-h}^{th} e^{a(th-r)}\zeta(dr), \quad t = 1, ..., n, \tag{2.4}$$

is white noise with zero mean and variance

$$E(\eta_{th}^2) = \sigma_\eta^2 = \frac{\sigma^2}{2a}\left[e^{2ah} - 1\right]. \tag{2.5}$$

If $x(t)$ is a flow variable, the observed data take the form of integrals $x_{th} = \int_{th-h}^{th} x(r)dr$. The observed vector is $\{x_h, x_{2h}, ..., x_{nh}\} = \{\int_0^h x(r)dr, \int_h^{2h} x(r)dr, ..., \int_{nh-h}^{nh} x(r)dr\}$, the

[1] $h$ is the sampling interval and $n$ is the total number of observations. Note that the time span of the data is $T = nh$. For example, if monthly data are recorded in a 20 year time interval, then $h = \frac{1}{12}$, $T = 20$ and $n = 240$.

constant is $c = (1 - e^{ah})\mu h$ and the first observation is

$$x_h = \frac{1}{a}\left[(e^{ah} - 1)\left(x(0) + \frac{b}{a}\right) - bh\right] + \eta_h. \tag{2.6}$$

The disturbance defined by

$$\eta_h = \frac{1}{a}\int_0^h \left(e^{a(h-r)} - 1\right)\zeta(dr), \tag{2.7}$$

$$\eta_{th} = \frac{1}{a}\left[\int_{th-h}^{th} \left(e^{a(th-r)} - 1\right)\zeta(dr) + \int_{th-2h}^{th-h} \left(e^{ah} - e^{a(th-h-r)}\right)\zeta(dr)\right], \tag{2.8}$$

for $t = 2, ..., n$, is a first-order moving average with variance

$$E(\eta_h^2) = \tilde{\gamma}_0 = \frac{\sigma^2}{2a^3}\left[3 + 2ah - 4e^{ah} + e^{2ah}\right], \tag{2.9}$$

$$E(\eta_{th}^2) = \gamma_0 = \frac{\sigma^2}{a^2}\left[h(e^{2ah} + 1) + \frac{1 - e^{2ah}}{a}\right], \quad t = 2, ..., n, \tag{2.10}$$

and first-order autocovariance

$$E(\eta_{th}\eta_{th-h}) = \gamma_1 = \frac{\sigma^2}{a^2}\left[\frac{e^{2ah} - 1}{2a} - he^{ah}\right]. \tag{2.11}$$

The exact discrete time model for this case of flow data can alternatively be expressed as

$$x_{th} = e^{ah}x_{th-h} + c + \varepsilon_{th} + \theta\varepsilon_{th-h},\ t = 2, ..., n, \tag{2.12}$$

where $\theta = (1 - \sqrt{1 - 4\rho^2})/(2\rho)$, $\rho = \gamma_1/\gamma_0$ and $\varepsilon_{th}$ is a white noise process with mean zero and variance $\sigma_\varepsilon^2 = \gamma_1/\theta = \gamma_0/(1 + \theta^2)$ (see Lemma 1.3.1).

One field where the literature on continuous time models driven by Brownian motions and more recently by Lévy processes is well developed is finance. Much of the focus in this area has been on the univariate stochastic differential equation

$$dx(t) = k(\mu - x(t))dt + \sigma dB(t),\ t > 0, \tag{2.13}$$

where $B(t)$ is a standard Brownian motion. The parameter of major interest being $k$, reflects the rate of convergence towards the long run mean, $\mu$. This model proposed by (Vasicek, 1977) for interest rates, can be seen as a special case of (2.1) with $a = -k$, $b = k\mu$ and $\zeta(dt) = \sigma dB(t)$.

From the financial literature it is well known that the estimator of the mean reversion parameter $k$, as opposed to the estimators of $\mu$ and $\sigma^2$, can suffer from substantial bias in finite samples, being especially acute when the continuous time process has a root near zero[2]. Yu (2012) derived the following bias expression for the case of known long run mean

$$E(\hat{a} - a) = -\frac{3 + e^{-2ah}}{2T} + o(T^{-1}). \tag{2.14}$$

Tang and Chen (2009) approximated the bias when the long run mean is unknown, and showed that the bias expression for this case becomes

$$E(\hat{a} - a) = -\frac{5 + 2e^{-ah} + e^{-2ah}}{2T} + o(T^{-1}). \tag{2.15}$$

These bias expressions were derived by assuming that the variable is observed at specific points in time. This assumption is reasonable for financial variables, but not for some economic variables such as consumption, income and exports, which are observed as the accumulation of the underlying rate over the sampling interval. Using the general formula obtained by Cox and Snell (1968), a bias expression for this type of variable when the mean is known (assumed to be 0) was derived in the first chapter. It was shown that the bias is smaller for flows than for stocks, although the leading term in both expressions is the same.

We could apply the Cox and Snell's formula to approximate the bias for the flow case with unknown mean. Since calculations are more complicated due to the inclusion of an additional parameter, the mean, and the resulting bias expression is cumbersome, we do not derive these results in this paper. We will show through simulations that similar to the known mean case, the bias is smaller for flows than for stocks when the mean is unknown. The leading term is expected to be identical in both formulae, for flows and stocks.

We can use the analytical bias expression evaluated at $\hat{a}$ to compute a bias corrected estimator (BC). The BC estimator for the case of known mean (equal to zero) is given by

$$\tilde{a} = \hat{a} + \frac{3 + e^{-2\hat{a}h}}{2T}, \tag{2.16}$$

and for the case of unknown mean by

$$\tilde{a} = \hat{a} + \frac{5 + 2e^{-\hat{a}h} + e^{-2\hat{a}h}}{2T}. \tag{2.17}$$

The BC estimators can be used in both cases, when the variable of interest is a stock or when it is a flow. Since we subtract the first order term of the bias expansion from the

[2] It implies a root near unity in the discrete time process.

parameter estimate, the BC estimator satisfies

$$E(\tilde{a} - a) = \mathrm{o}(T^{-1}). \tag{2.18}$$

### 2.2.2 Multivariate model

We consider the multivariate continuous time model given by

$$dX(t) = [A(\Theta)X(t) + B(\Theta)]\, dt + \zeta(dt),\ t > 0, \tag{2.19}$$

where $X(t) = (X_1(t), ..., X_g(t))'$ is a $g$-dimensional continuous time random process, $X(0)$ is given, the $g \times g$ matrix $A$ and $g \times 1$ vector $B$ are known functions of an unknown $p \times 1$ parameter vector $\Theta$ ($p \leq g(g+1)$) and $\zeta(dt)$ is vector white noise with covariance matrix $\Sigma(\omega)dt$ with $\Sigma(\omega)$ a function of an unknown $q \times 1$ parameter vector $\omega$ ($q \leq g(g+1)/2$). It is also assumed that all characteristic roots of the coefficient matrix $A$ have negative real parts to ensure stationarity and a priori restrictions on the structure of the continuous system ensure identifiability.[3]

The exact discrete time model corresponding to (2.19) is given by

$$X_h = e^{A(\Theta)h} X(0) + (e^{A(\Theta)h} - I)A^{-1}(\Theta)B(\Theta) + \epsilon_h, \tag{2.20}$$

$$X_{th} = e^{A(\Theta)h} X_{th-h} + (e^{A(\Theta)h} - I)A^{-1}(\Theta)B(\Theta) + \epsilon_{th}, \quad t = 2, ..., n, \tag{2.21}$$

where the matrix exponential is defined by $e^{Ah} = \sum_{i=0}^{\infty}(Ah)^i/i!$. Under the assumption that the vector $X(t)$ comprises only stock variables,[4] the discrete time disturbance vector $\epsilon_{th}$ takes the form

$$\epsilon_{th} = \int_{th-h}^{th} e^{(th-r)A}\zeta(dr), \tag{2.22}$$

and satisfies the following properties

$$E(\epsilon_{th}\epsilon_{th}') = \Omega = \int_0^h e^{rA}\Sigma e^{rA'} dr, \tag{2.23}$$

$$E(\epsilon_{sh}\epsilon_{th}') = 0, s \neq t. \tag{2.24}$$

Wang et al. (2011) derived two analytical expressions to approximate the bias in linear systems when the vector $X(t)$ comprises stock variables. The first expression corresponds to the estimator based on the Euler approximation and the second one to that based on

[3]One complication arises when a continuous system is estimated from the exact discrete model. This complication is the aliasing phenomenon, which in a system of differential equations leads to an identification problem.

[4]We focus on stock variables, but flows or a mixture of stocks and flows may be considered.

the trapezoidal approximation. We only consider the second bias expression to compute the bias corrected estimator since the trapezoidal rule leads to a smaller discretization bias. The estimation bias is given by

$$E(\hat{A}-A) = -\frac{4}{T}(I+e^{Ah})^{-1}K(I+e^{Ah})^{-1} - \frac{4}{h}L(I+e^{Ah})^{-1} + \mathrm{o}(T^{-1}), \tag{2.25}$$

where $L$ is a $g \times g$ matrix whose $ij^{\text{th}}$ element is

$$L_{ij} = \frac{1}{n}\sum_{s=1}^{g} e'_{g(s-1)+i}\Delta e_{g(j-1)+s}, \tag{2.26}$$

with $\Delta = [I_g \otimes (I+e^{Ah})^{-1}]\Gamma(0)^{-1} \otimes \Omega[I_g \otimes (I+e^{Ah})^{-1}]'$, and $e_i$ being a column vector of dimension $g^2$ whose $i^{\text{th}}$ element is 1 and other elements are 0. When $B(\Theta)$ is unknown, $K$ is given by

$$K = \Omega\left[(I-e^{A'h})^{-1} + e^{A'h}(I-e^{2A'h})^{-1} + \sum_{\lambda\in Spec(C)}\lambda(I-\lambda e^{A'h})^{-1}\right]\Gamma(0)^{-1}, \tag{2.27}$$

where $\Gamma(0) = \sum_{i=0}^{\infty} e^{iAh}\Omega e^{iA'h}$ and $Spec(C)$ is the set of eigenvalues of $e^{A'h}$. If $B(\Theta)$ is known, $K$ becomes

$$K = \Omega\left[e^{A'h}(I-e^{2A'h})^{-1} + \sum_{\lambda\in Spec(C)}\lambda(I-\lambda e^{A'h})^{-1}\right]\Gamma(0)^{-1}. \tag{2.28}$$

The BC estimator can be computed as

$$\tilde{A} = \hat{A} + \frac{4}{T}(I+e^{\hat{A}h})^{-1}\hat{K}(I+e^{\hat{A}h})^{-1} + \frac{4}{h}L(I+e^{\hat{A}h})^{-1}, \tag{2.29}$$

and satisfies

$$E(\tilde{A}-A) = \mathrm{o}(T^{-1}). \tag{2.30}$$

Computing a bias corrected estimator is straightforward once an explicit expression for the bias function or the first terms in an asymptotic expansion of the bias have been derived. In many cases, even the first terms of a bias expansion are difficult to compute explicitly, or they may be analytically involved. Because of these difficulties, other approaches could be preferred to implement in practice.

## 2.3 Bias reduction methods

In this section we consider three alternative approaches for bias reduction which do not rely on the explicit form of an asymptotic expansion of the bias, these being the jackknife, indirect inference and bootstrap methods. We start by describing the bias correction procedures. The finite sample properties of these three alternative estimators are then examined by providing stochastic expansions. We first assume that the estimator is consistent. We then consider the case of inconsistent estimators and show that the methods may fail to reduce the bias as intended. Let $\Phi$ denote a $d$-dimensional parameter vector, whose true value is equal to $\Phi_0$.

### 2.3.1 Jackknife

The jackknife method for reducing the bias was introduced by Quenouille (1949). Based on Quenouille's jackknife, Phillips and Yu (2005) proposed a bias reduction method in the time series context. Instead of deleting observation $i$ or a block of $d$ observations from the full sample, which is applicable for i.i.d observations, the Phillips and Yu's method uses non-overlapping sub-samples in order to preserve the dynamic dependence of the data. The procedure requires dividing the sample of $n$ observations into $m$ non-overlapping sub-samples of size $l$ such that $n = ml$. Using a given method, such as ML, the parameter vector is estimated for each sub-sample and the entire sample. Let $Y_i$ $(i = 1, ..., m)$ denote the set of $m$ sub-samples, $\hat{\Phi}_i$ $(i = 1, ..., m)$ the set of $m$ sub-sample estimators and $\hat{\Phi}$ the estimator that uses the whole sample. The jackknife estimator is then computed as the linear combination of the whole-sample and sub-sample estimators

$$\hat{\Phi}_{J,m} = w\hat{\Phi} + w_m \frac{1}{m} \sum_{i=1}^{m} \hat{\Phi}_i, \tag{2.31}$$

where the weights are given by

$$w = \frac{m}{m-1} \quad \text{and} \quad w_m = -\frac{1}{m-1}.$$

Alternative sampling schemes could be used to obtain the jackknife estimator (see Chambers, 2013). The weights have to be, however, modified to take into account the type of employed sub-samples. Weights need to be also modified when the sub-sample lengths are not all equal. For the case of two different sets of sub-samples $Y_{1,i}$ $(i = 1, ..., m_1)$ and $Y_{2,i}$ $(i = m_1 + 1, ..., m_1 + m_2)$, Chambers (2013) showed that the weights become

$$w = \frac{m_1 n(l_1 - l_2) - mnl_1}{m_1 n(l_1 - l_2) - ml_1(n - l_2)} \quad \text{and} \quad w_m = \frac{ml_1 l_2}{m_1 n(l_1 - l_2) - ml_1(n - l_2)},$$

where $l_j$ $(j = 1, 2)$ is the length of sub-sample $Y_{j,i}$ and $m_1 = m - m_2$.

The jackknife estimator defined above has the property that it eliminates the bias term of order $n^{-1}$. Since the jackknife method based on non-overlapping sub-samples only requires estimating the model $m + 1$ times, it is computationally much cheaper to implement than other bias reduction methods such as bootstrapping or indirect inference, which usually require a large number of repetitions. One drawback of jackknife is that it tends to inflate the variance of the estimator. In the next section it will be shown that despite increasing the variance, the jackknife can make substantial reductions in the bias that offset the inflated variance, leading to a decrease in the RMSE. Another limitation of the jackknife is that it does not completely remove the first order bias when the process has a unit root. The reason is that the limit distribution of the estimator depends on the initial condition, implying different functional forms in different sub-samples. We expect a similar behaviour for the nearly unit root situation.

To see if this issue occurs in a near unit root process, we simulate 240 monthly observations from the univariate model (2.2)-(2.5) for two values of the continuous time parameter $a$, these being -3 and -0.1 and corresponding to a stationary and a nearly unit root process, respectively. The variable is assumed to be a stock and $b$ is set to 0. Figure 2.1 shows the distribution of the sub-sample estimates using $m = 3$ and 10000 replications. As opposed to the stationary case, when $a = -0.1$ the distribution of $\hat{\Phi}_i$ appears to be different for one of the sub-samples.

In order to improve the performance of the jackknife in unit root models, a modified estimator with new weights was proposed by Chambers and Kyriacou (2013). Optimal weights that minimize the variance of this modified jackknife were suggested by Chen and Yu (2015).

### 2.3.2 Indirect inference

The indirect inference method was originally proposed by Smith (1993) and generalized independently by Gallant and Tauchen (1996) and Gourieroux et al. (1993). It is a simulation-based method that can be used to estimate the parameters of a model. Its advantages not only for estimating models with complex likelihood functions, but also for correcting the bias arising from estimation, have been highlighted in the literature (see for example, Gourieroux et al., 1993; Phillips and Yu, 2009). As pointed out by

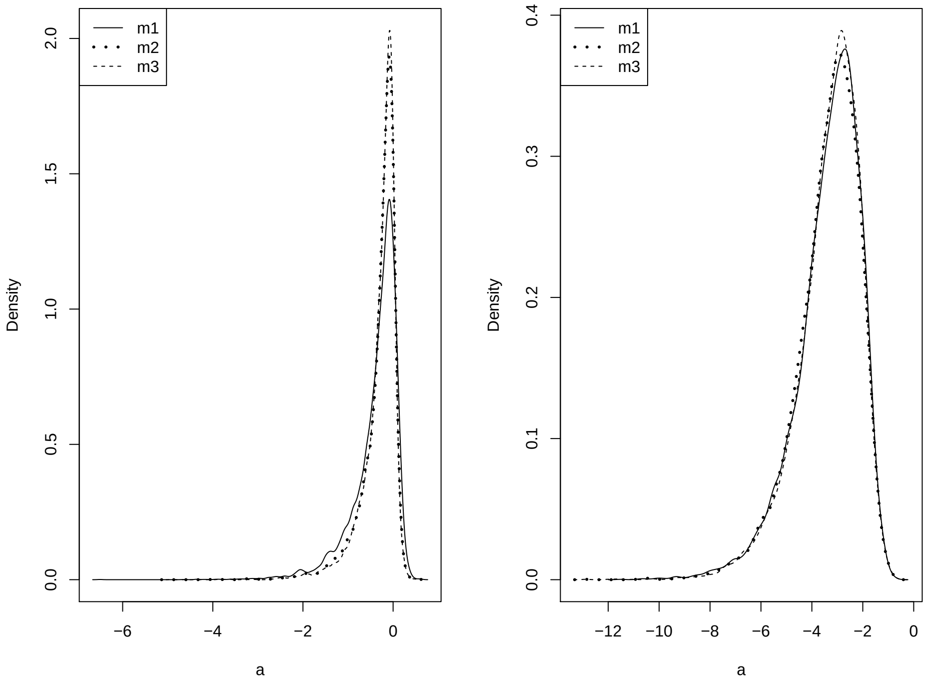


Figure 2.1: Distribution of the sub-sample estimates using $m = 3$. The solid line corresponds to the $1^{\text{st}}$ sub-sample and the dotted and dashed lines to the $2^{\text{nd}}$ and $3^{\text{rd}}$ sub-samples, respectively. The graph in the left corresponds to $a = -0.1$ and the one in the right to $a = -3$.

Gourieroux et al. (1993), indirect inference can be also used to remove the discretization bias that is induced by discretizing continuous time models.[5] This method also has the valuable advantage of its generality. It is applicable in a wide range of economic and statistical models, including DSGE models, dynamic panel models, discrete choice models and continuous time models. It can also be implemented with different estimation methods and may inherit some of the good properties of the original estimators.

The method requires generating samples from the original model. Given a parameter choice $\Phi$, let $y^s(\Phi) = \{y^s_h, y^s_{2h}, ..., y^s_{nh}\}$[6] be the artificial data, where $s = 1, ..., S$, and $S$ being the number of simulated paths. Indirect inference uses these artificial data to estimate an auxiliary model, which is easy to estimate but is not necessarily correctly specified. Let $\{\hat{\Phi}^s(\Phi)\}_{s=1}^S$ denote the estimates obtained from simulated data. The auxiliary model is also estimated using the observed data. Let $\hat{\Phi}$ denote the estimates obtained from observed data. The estimate that uses the observed data is then compared to the average of the $S$ estimates based on the artificial data by employing some measure of distance, for example the absolute value of its numerical difference (i.e.

[5]In finance, for example, where many models involve non-linearities, some form of approximation such as Euler scheme or trapezoidal rule is often used to obtain a discrete representation of the continuous time model. The estimates suffer, however, from bias that is induced by the approximation.

[6]For calibration purposes, the number of observations in $y^s(\Phi)$ has to be the same as that in the observed data.

$|\hat{\Phi} - \frac{1}{S}\sum_{s=1}^{S}\hat{\Phi}^s(\Phi)|)$. Let us call it the distance estimate. The procedure is repeated for various (say $i$) possible values of the parameter vector $\Phi$, leading to a set of $i$ distance estimates. The indirect inference estimate is the minimum distance estimate (see section 2.4 for a more detailed description of the method applied to continuous time models).

More formally, the estimator is defined by[7]

$$\hat{\Phi}_{II,S} = \underset{\Phi}{\operatorname{argmin}} \left\| \hat{\Phi} - \frac{1}{S}\sum_{s=1}^{S}\hat{\Phi}^s(\Phi) \right\|, \tag{2.32}$$

where $\|.\|$ is some finite-dimensional distance metric. As the number of simulated paths $S$ tends to infinity, the II estimator becomes

$$\hat{\Phi}_{II} = \underset{\Phi}{\operatorname{argmin}} \left\| \hat{\Phi} - E\left(\hat{\Phi}^s(\Phi)\right) \right\|. \tag{2.33}$$

### 2.3.3 Bootstrap

The bootstrap, due to Efron (1979, 1982), is a resampling method that uses the original sample to generate new samples, each of which is employed to calculate the statistic of interest. Unlike jackknife, every resample has the same number of observations as the original sample. There are different ways to obtain bootstrap samples. In the residual bootstrap, for example, $S$ resamples are drawn from the residuals with replacement, which are then employed to generate bootstrap samples of the variable of interest. Given a parameter estimate $\hat{\Phi}$, let $y^s(\hat{\Phi}) = \{y_h^s, y_{2h}^s, ..., y_{nh}^s\}$ be the $s^{\text{th}}$ resample, where $s = 1, ..., S$, and $S$ being the number of bootstrap simulated paths. This leads to $S$ data sets and hence $S$ estimates of $\Phi$. Let $\{\hat{\Phi}^s(\hat{\Phi})\}_{s=1}^{S}$ denote the estimates obtained from the artificial data and $\widetilde{\Phi}$ be the sample average of $\{\hat{\Phi}^s(\hat{\Phi})\}_{s=1}^{S}$, that is,

$$\widetilde{\Phi} = \frac{1}{S}\sum_{s=1}^{S}\hat{\Phi}^s(\hat{\Phi}). \tag{2.34}$$

The bootstrap bias corrected estimator of $\Phi$ is

$$\hat{\Phi}_{B,S} = 2\hat{\Phi} - \widetilde{\Phi}. \tag{2.35}$$

[7]The number of paramaters in the auxiliary model must be at least as large as the number of parameters in the original model. Note that we are assuming that the dimension of the auxiliary paramenter is the same as that of the initial parameter. When the dimension of the auxiliary parameter is larger, the indirect inference estimator becomes $\hat{\Phi}_{II,S} = \underset{\Phi}{\operatorname{argmin}} \left\| \hat{\Phi} - \frac{1}{S}\sum_{s=1}^{S}\hat{\Phi}^s(\Phi) \right\|_{\hat{\Omega}}^2$, where $\hat{\Omega}$ is a positive definite matrix converging to a deterministic positive definite matrix $\Omega$.

### 2.3.4 Stochastic expansions of the bias corrected estimators

We now present some stochastic expansions of the bias corrected estimators in continuous time models. Let us first consider the ML estimator of $\Phi$, which is given by

$$\hat{\Phi} = \underset{\Phi}{\operatorname{argmax}} \ \ell(\Phi), \tag{2.36}$$

where $\ell(\Phi)$ is the log-likelihood function. If the function $\ell(\Phi)$ is correctly specified, $\hat{\Phi}$ is consistent. In such situations, jackknife, indirect inference and bootstrap are useful methods for correcting the finite sample bias.

We assume that the estimator $\hat{\Phi}$ admits the stochastic expansion

$$\hat{\Phi} = \Phi_0 + \frac{P(v, \Phi_0)}{T^{1/2}} + \frac{Q(v, \Phi_0)}{T} + \mathrm{o}(T^{-1}), \tag{2.37}$$

where $P(v, \Phi_0)$ and $Q(v, \Phi_0)$ are $d$-dimensional random vectors, which depend on some asymptotic random term $v$. Note that the stochastic expansion is given in powers of $T^{-1/2}$, where $T$ is the data span. The reason for expanding the estimator in powers of $T^{-1/2}$ rather than of $n^{-1/2}$ is that the bias in continuous time models is mainly determined by the data span.

It is straightforward to show that the jackknife estimator satisfies the following expansion

$$\hat{\Phi}_{J,m} = (w + w_m)\Phi_0 + \left(\frac{w}{T^{1/2}} + \frac{w_m}{(lh)^{1/2}}\right) P(v, \Phi_0) + \left(\frac{w}{T} + \frac{w_m}{lh}\right) Q(v, \Phi_0) + \mathrm{o}(T^{-1}). \tag{2.38}$$

Chambers (2013) shows that the jackknife estimator $\hat{\Phi}_{J,m}$ with weights

$$w = \frac{m}{m-1} \quad \text{and} \quad w_m = -\frac{1}{m-1}$$

for equal sub-sample lengths or weights equal to

$$w = \frac{m_1 T(l_1 - l_2) - mTl_1}{m_1 T(l_1 - l_2) - ml_1(T - hl_2)} \quad \text{and} \quad w_m = \frac{mhl_1 l_2}{m_1 T(l_1 - l_2) - ml_1(T - hl_2)}$$

for unequal sub-sample lengths, satisfies

$$E(\hat{\Phi}_{J,m} - \Phi_0) = \mathrm{o}(T^{-1}). \tag{2.39}$$

To obtain the last result we have used the fact that $E[P(v, \Phi_0)] = 0$.[8] The jackknife

[8]Note that the first order term of the bias expressions presented in section two is of order $T^{-1}$ not $T^{-1/2}$. See also the results obtained by Bao and Ullah (2007) for a general class of estimators in linear

estimator has, therefore, the property that it eliminates the bias term of order $T^{-1}$.

Gouriéroux et al. (2000) show that under some regularity conditions, the indirect inference estimator admits the following stochastic expansion[9]

$$\hat{\Phi}_{II,S} = \Phi_0 + \frac{P^{II}}{T^{1/2}} + \frac{Q^{II}}{T} + \mathrm{o}(T^{-1}), \tag{2.40}$$

where the coefficients $P^{II}$ and $Q^{II}$ are given by

$$P^{II} = P(v, \Phi_0) - \frac{1}{S}\sum_{s=1}^{S} P(v_s, \Phi_0), \tag{2.41}$$

$$Q^{II} = Q(v, \Phi_0) - \frac{1}{S}\sum_{s=1}^{S} Q(v_s, \Phi_0) - \left[\frac{1}{S}\sum_{s=1}^{S} \frac{\partial P}{\partial \Phi'}(v_s, \Phi_0)\right] P^{II}. \tag{2.42}$$

The random vectors $P(v, \Phi_0)$ and $Q(v, \Phi_0)$ are the first and second order terms of the expansion applied to the estimator $\hat{\Phi}$ (i.e. that based on the observed data). The random vectors $P(v_s, \Phi_0)$ and $Q(v_s, \Phi_0)$ are the corresponding first and second order terms of the expansion applied to the estimator $\hat{\Phi}^s(\Phi)$ (i.e. that based on the simulated values). It is also assumed that the random variables $v$ and $v_s$, $s = 1, ..., S$ are independent and identically distributed.

For an infinite number of replications $S$, the terms of order $T^{-1/2}$ and $T^{-1}$ in the expansion satisfy $E\left(\lim_{S\to\infty} P^{II}\right) = E\left(\lim_{S\to\infty} Q^{II}\right) = 0$. The indirect inference estimator has the property that it eliminates the $T^{-1/2}$ and $T^{-1}$ order bias terms, satisfying, therefore,

$$E(\hat{\Phi}_{II} - \Phi_0) = \mathrm{o}(T^{-1}). \tag{2.43}$$

For a finite number of replications $S$, the last result is not valid. Although $E(P^{II}) = 0$, $E(Q^{II})$ is no longer zero. It is now given by

$$E(Q^{II}) = \frac{1}{S}\sum_{j=1}^{d} \mathrm{cov}\left[\frac{\partial P}{\partial \Phi_j}(v, \Phi_0), P_j(v, \Phi_0)\right]. \tag{2.44}$$

The $T^{-1}$ order bias term of the indirect inference estimator $\hat{\Phi}_{II,S}$ is smaller than that of the estimator $\hat{\Phi}$ as soon as

$$\frac{1}{S}\left|\sum_{j=1}^{d} \mathrm{cov}\left[\frac{\partial P}{\partial \Phi_j}(v, \Phi_0), P_j(v, \Phi_0)\right]\right| \leq \left|E\left[Q(v, \Phi_0)\right]\right| \tag{2.45}$$

and non-linear time series models.

[9]See also Gourieroux et al. (1993) and Smith (1993) for the proof of consistency of the indirect inference estimator.

(for further details see Gouriéroux et al., 2000, pg. 342).

The bootstrap estimator also satisfies the stochastic expansion

$$\hat{\Phi}_{B,S} = \Phi_0 + \frac{P^B}{T^{1/2}} + \frac{Q^B}{T} + \mathrm{o}(T^{-1}), \tag{2.46}$$

where the coefficients $P^B$ and $Q^B$ are

$$P^B = P(v, \Phi_0) - \frac{1}{S}\sum_{s=1}^{S} P(v_s, \Phi_0), \tag{2.47}$$

$$Q^B = Q(v, \Phi_0) - \frac{1}{S}\sum_{s=1}^{S} Q(v_s, \Phi_0) - \left[\frac{1}{S}\sum_{s=1}^{S} \frac{\partial P}{\partial \Phi'}(v_s, \Phi_0)\right] P(v, \Phi_0). \tag{2.48}$$

Given independence between the random variables $v$ and $v_s$, $s = 1, ..., S$, and the fact that $E[P(v, \Phi_0)] = 0$, we obtain $E(P^B) = E(Q^B) = 0$. The bootstrap estimator eliminates the $T^{-1/2}$ and $T^{-1}$ order bias terms, satisfying, therefore,

$$E(\hat{\Phi}_{B,S} - \Phi_0) = \mathrm{o}(T^{-1}). \tag{2.49}$$

The expansions indicate that the $T^{-1}$ order bias of the jackknife and bootstrap estimators vanishes, while that of the indirect inference estimator only disappears for an infinite number of simulations. Note that since $S$ cannot be made infinite in practice, the number of indirect inference replications has to be large to ensure accuracy in bias corrections.

We now present the corresponding asymptotic expansions when the estimators are inconsistent and show that the methods may fail to correct the bias in this scenario. A typical case in which this situation can occur is when an approximate discrete model is used for estimation. In this case the estimator is inconsistent due to the discretization bias, although this asymptotic bias reduces as the data frequency increases. If the variable is observed as a flow or the order of the differential equation is higher than one, there is an additional source of asymptotic bias. The temporal aggregation occurring with flows as well as orders greater than one induce serial correlation in the discrete time disturbances, which is not taken into account by an approximate discrete model obtained, for example, from the Euler scheme.

Let the ML estimator calculated on this approximate model be

$$\hat{\Phi} = \underset{\Phi}{\operatorname{argmax}} \;\; \ell^a(\Phi), \tag{2.50}$$

where $\ell^a(\Phi)$ denotes the log-likelihood function of the approximate model. Since $\ell^a(\Phi)$ is misspecified, $\hat{\Phi}$ is in general inconsistent. Suppose that the estimator $\hat{\Phi}$ admits the following stochastic expansion

$$\hat{\Phi} = b(\Phi_0) + \frac{P(v,\Phi_0)}{T^{1/2}} + \frac{Q(v,\Phi_0)}{T} + \mathrm{o}(T^{-1}), \tag{2.51}$$

where $b(\Phi_0) = \Phi_0 + \text{bias}(\Phi_0)$, with $\text{bias}(\Phi_0)$ being the asymptotic bias of $\hat{\Phi}$.

The expansion of the jackknife estimator is now given by

$$\hat{\Phi}_{J,m} = (w + w_m)b(\Phi_0) + \left(\frac{w}{T^{1/2}} + \frac{w_m}{(lh)^{1/2}}\right) P(v,\Phi_0) + \left(\frac{w}{T} + \frac{w_m}{lh}\right) Q(v,\Phi_0) + \mathrm{o}(T^{-1}), \tag{2.52}$$

and the bias of $\hat{\Phi}_{J,m}$ by

$$E(\hat{\Phi}_{J,m} - \Phi_0) = \text{bias}(\Phi_0) + \mathrm{o}(T^{-1}). \tag{2.53}$$

Although the jackknife estimator eliminates the $T^{-1}$ order bias term, it is not able to remove the asymptotic bias. The estimator $\hat{\Phi}_{J,m}$ is, therefore, inconsistent.

Gourieroux and Monfort (1997, pg. 78) show that when the first step estimator $\hat{\Phi}$ is inconsistent, the terms $P^{II}$ and $Q^{II}$ in the stochastic expansion given in (2.40) become

$$P^{II} = \left[\frac{\partial b}{\partial \Phi'}(\Phi_0)\right]^{-1} \left[P(v,\Phi_0) - \frac{1}{S}\sum_{s=1}^{S} P(v_s,\Phi_0)\right], \tag{2.54}$$

$$\begin{aligned} Q^{II} = & \left[\frac{\partial b}{\partial \Phi'}(\Phi_0)\right]^{-1} \left[Q(v,\Phi_0) - \frac{1}{S}\sum_{s=1}^{S} Q(v_s,\Phi_0)\right] \\ & - \left[\frac{\partial b}{\partial \Phi'}(\Phi_0)\right]^{-1} \left[\frac{1}{S}\sum_{s=1}^{S} \frac{\partial P}{\partial \Phi'}(v_s,\Phi_0)P^{II} + \frac{1}{2}\left(P^{II\prime}\frac{\partial^2 b_j(\Phi_0)}{\partial\Phi\partial\Phi'}P^{II}\right)_{j=1,\dots,d}\right]. \end{aligned} \tag{2.55}$$

The term of order $T^{-1/2}$ satisfies $E(P^{II}) = 0$, while the term of order $T^{-1}$ satisfies

$$\begin{aligned} E(Q^{II}) = & -\left[\frac{\partial b}{\partial \Phi'}(\Phi_0)\right]^{-1} \Bigg\{ -\frac{1}{S}\sum_{j=1}^{d} \operatorname{cov}\left[\frac{\partial P}{\partial \Phi_j}(v,\Phi_0), \left(\left[\frac{\partial b}{\partial \Phi'}(\Phi_0)\right]^{-1} P(v,\Phi_0)\right)_j\right] \\ & + \frac{1}{2}\left(1 + \frac{1}{S}\right)\left[\operatorname{Tr}\left(\frac{\partial^2 b_j}{\partial\Phi\partial\Phi'}(\Phi_0)\left[\frac{\partial b}{\partial \Phi'}(\Phi_0)\right]^{-1} V(P(v,\Phi_0))\left[\frac{\partial b'}{\partial \Phi}(\Phi_0)\right]^{-1}\right)\right]_{j=1,\dots,d}\Bigg\} \end{aligned} \tag{2.56}$$

(for details the interested reader is referred to Gourieroux and Monfort, 1997, pg. 79).

Although not eliminating the $T^{-1}$ order bias term in its entirety, the indirect inference estimator is able to remove the asymptotic bias. The bias of $\hat{\Phi}_{II,S}$ is given by

$$E(\hat{\Phi}_{II,S} - \Phi_0) = \frac{1}{T}E(Q^{II}) + \mathrm{o}(T^{-1}). \tag{2.57}$$

Note that the bias of the indirect inference estimator vanishes when the span goes to infinity. As opposed to the jackknife estimator, the indirect inference estimator appears to enjoy good properties when the ML estimator $\hat{\Phi}$ is inconsistent.

The expansion of the bootstrap estimator becomes

$$\hat{\Phi}_{B,S} = b(\Phi_0) - [b(\hat{\Phi}) - \hat{\Phi}] + \frac{P^B}{T^{1/2}} + \frac{Q^B}{T} + \mathrm{o}(T^{-1}), \tag{2.58}$$

and the bias is now given by

$$E(\hat{\Phi}_{B,S} - \Phi_0) = \mathrm{bias}(\Phi_0) - \mathrm{bias}(\hat{\Phi}) + \mathrm{o}(T^{-1}). \tag{2.59}$$

The last expression is obtained by taking account the fact that $b(\hat{\Phi}) = \hat{\Phi} + \mathrm{bias}(\hat{\Phi})$, with $\mathrm{bias}(\hat{\Phi})$ being the asymptotic bias of the estimator $\hat{\Phi}^s(\hat{\Phi})$ (i.e. that based on simulated data). This formula indicates that the bias of the bootstrap estimator is is small when $\hat{\Phi}$ is close to $\Phi_0$ and smaller in magnitude than that of the jackknife estimator, when the terms $\mathrm{bias}(\Phi_0)$ and $\mathrm{bias}(\hat{\Phi})$ have the same sign and $\mathrm{bias}(\Phi_0) > 0.5\ \mathrm{bias}(\hat{\Phi})$, in which case the two asymptotic biases are mutually offsetting. However, since these biases do not disappear as the data span goes to infinity, the indirect inference seems to be preferred in general when the ML estimator is inconsistent.

The next section explores the implementation of the bias reduction methods in continuous time models using simulations.

## 2.4 Monte Carlo simulations

### 2.4.1 Univariate model

We begin by considering the univariate continuous time model (2.1) with $\zeta(dt) = \sigma dB(t)$ and $B(t)$ a standard Brownian motion. Data are generated from the exact discrete representation with $h = 1/12$ and $T = 20$, corresponding to 20 years of monthly data. We first generate the discrete time disturbances as

$$\eta_{th} = \sigma_\eta \vartheta_{th}, \quad t = 1, ..., n, \tag{2.60}$$

when the variable is a stock and according to

$$\eta_h = m_{11}\vartheta_h, \tag{2.61}$$

$$\eta_{2h} = m_{22}\vartheta_{2h} + m_{21}\vartheta_h, \tag{2.62}$$

$$\eta_{th} = m_{tt}\vartheta_{th} + m_{t,t-1}\vartheta_{th-h}, \quad t = 3, ..., n, \tag{2.63}$$

when the variable is a flow, where $\vartheta_{th}$ is a random draw from a standard normal distribution, $\sigma_\eta$ is the positive square root of the variance given in (2.5), the coefficients $m_{21}$, $m_{t,t-1}$ $(t = 3, ..., n)$ and the positive coefficients $m_{tt}$ $(t = 1, ..., n)$ are obtained recursively from the equations

$$\begin{aligned} m_{11}^2 &= \tilde{\gamma_0}, \\ m_{21} &= \gamma_1/m_{11}, \\ m_{22}^2 &= \gamma_0 - m_{21}^2, \\ m_{t,t-1} &= \gamma_1/m_{t-1,t-1}, \\ m_{tt}^2 &= \gamma_0 - m_{t,t-1}^2, \quad t = 3, ..., n, \end{aligned}$$

with $\tilde{\gamma_0}$, $\gamma_0$ and $\gamma_1$ given in (2.9)-(2.11). Assuming $x(0) = -b/a$, simulated data are then computed by using (2.2), and the corresponding equation for observation $x_h$ (i.e. (2.3) for stocks and (2.6) for flows). We consider different values of the continuous time parameter $a$ over the interval $[-3, 0)$, $\sigma = 0.1$, $\mu = 0.1$ when the mean is unknown and $\mu = 0$ when the mean is known.[10] The experiment is replicated 10000 times.

For simplicity, we assume that the discrete time disturbance corresponding to the first observation is zero and compute the conditional maximum likelihood estimates. For the flow case, the log-likelihood function is calculated based on the ARMA representation given in (2.12).[11] Since we are conducting a large number of simulations it is convenient to reduce the dimension of the optimisation problem by concentrating the likelihood function with respect to $\sigma^2$. Conditioning on the first observation $x_h$, the concentrated log-likelihood function for stocks is (ignoring a constant)

$$\ell = -\frac{n-1}{2}\ln\left[\sum_{t=2}^{n}(x_{th} - e^{ah}x_{th-h} - (1-e^{ah})\mu)^2\right], \tag{2.64}$$

and conditioning on $x_h$ and $\varepsilon_h = 0$, the concentrated log-likelihood function for flows is

[10] Results are invariant to other choices of $\sigma$ and $\mu$.

[11] Alternatively, we can obtain exact Gaussian estimates. See the iterative estimation procedure proposed by Bergstrom (1990, pg. 113-114), which does not require expressing the disturbances as a moving average process.

(ignoring a constant)

$$\ell = -\frac{n-1}{2}\ln\left[\sum_{t=2}^{n}(x_{th} - e^{ah}x_{th-h} - (1-e^{ah})\mu h - \theta\varepsilon_{th-h})^2\right], \tag{2.65}$$

where $\theta = (1-\sqrt{1-4\rho^2})/(2\rho)$, $\rho = \gamma_1/\gamma_0$. For each simulation, we employ maximum likelihood (ML), jackknife (JA), indirect inference (II) and bootstrap (BOOT) to estimate $a$. We use the bias of the ML estimator as the benchmark.

Four values of $m = \{2, 3, 4, 8\}$, the number of non-overlapping sub-samples, are employed to compute the JA estimator. We consider three values of the number of simulated paths for indirect inference, these being $S = \{100, 1000, 10000\}$. For a stock variable and a given value of $a$, each indirect inference sample is computed recursively according to

$$x_{th}^s = e^{ah}x_{th-h}^s + (1-e^{ah})\mu + \sigma_\eta u_{th}^s, \tag{2.66}$$

where $u_{th}^s \sim i.i.dN(0,1)$ and $\sigma_\eta$ is the positive square root of the variance given in (2.5).

For a flow, each sample is obtained as

$$x_{th}^s = e^{ah}x_{th-h}^s + (1-e^{ah})\mu h + \sigma_\varepsilon w_{th}^s + \theta\sigma_\varepsilon w_{th-h}^s, \tag{2.67}$$

where $w_{th}^s \sim i.i.dN(0,1)$ and $\sigma_\varepsilon = \sqrt{\gamma_1/\theta}$.

The indirect inference samples are obtained by assuming $\sigma_\eta$, $\sigma_\varepsilon$ and $\mu$ known.[12] We select a grid of possible values for the parameter $a$, these being $a = 0.5 - 0.05i, i = 0, ..., 150,$[13] and compute the sample average of the $S$ estimates obtained from the artificial data for each value of $a$ (i.e. $\frac{1}{S}\sum_{s=1}^{S}\hat{a}^s(a)$). We keep the simulated disturbance paths constant over $a$ to allow $x_{th}^s$ to change only because of the different values of $a$. The indirect inference estimator is the value of $a$ that minimize the distance between $\hat{a}$ and $\frac{1}{S}\sum_{s=1}^{S}\hat{a}^s(a)$.

The number of bootstrap repetitions $S$ is set to 100. For stocks, each bootstrap sample is obtained according to

$$x_{th}^s = e^{\hat{a}h}x_{th-h}^s + (1-e^{\hat{a}h})\hat{\mu} + u_{th}^s, \tag{2.68}$$

where $u_{th}^s$ is a random draw with replacement from the centered residuals $\hat{\eta}_{th}$, with $\hat{\eta}_{th} =$

[12]Results are robust to other choices of $\sigma_\eta$, $\sigma_\varepsilon$ and $\mu$.

[13]Gouriéroux et al. (2000) applied indirect inference in a discrete time AR(1) model using $S = 15000$ and a finer grid of possible values for the autoregressive parameter. Their results indicate that the indirect inference estimator works as well as the median unbiased estimator of Andrews (1993). We also consider a finer grid of possible values for $a$. Results do not change much with respect to the ones presented in this paper.

$x_{th} - e^{\hat{a}h}x_{th-h} - (1 - e^{\hat{a}h})\hat{\mu}$. For flows, each bootstrap sample is computed using

$$x^s_{th} = e^{\hat{a}h}x^s_{th-h} + (1 - e^{\hat{a}h})\hat{\mu}h + w^s_{th} + \hat{\theta}w^s_{th-h}, \tag{2.69}$$

where $w^s_{th}$ is a random draw with replacement from the centered residuals $\hat{\varepsilon}_{th}$, with $\hat{\varepsilon}_{th} = x_{th} - e^{\hat{a}h}x_{th-h} - (1 - e^{\hat{a}h})\hat{\mu}h - \hat{\theta}\varepsilon_{th-h}$ and $\hat{\theta} = (1 - \sqrt{1 - 4\hat{\rho}^2})/(2\hat{\rho})$ and $\hat{\rho} = \hat{\gamma}_1/\hat{\gamma}_0$.[14]

Figures 2.2 and 2.3 plot the bias and root mean square error (RMSE) of the ML estimator in the model with known mean and the BC estimator given in (2.16) for different values of the continuous time parameter $a$ over the interval $[-3, 0)$. Results shown in Figure 2.2 are obtained under the assumption that the variable is a stock while those contained in Figure 2.3 assume that the variable is a flow.

We observe that the curvature of the bias function of the BC estimator is similar to that of the ML estimator for the near unit root situation. The reason for this is that the analytical bias expression is not able to reproduce the curvature of the actual bias function. Note that when $a$ approaches zero the bias function obtained from simulations is highly nonlinear, while the first order term in the analytical bias expression is linear. More importantly, we see from the figures that the BC estimator exhibits only a small bias and has lower RMSE than the ML estimator for both stock and flow variables.

Table 2.2 in the Appendix reports the bias and RMSE of the JA estimators of $a$ for the values of $m$ considered. It is clear from Table 2.2 that the bias increases with $m$ for the majority of parameter values while the RMSE decreases with $m$. As opposed to the JA estimator based on 2 sub-samples, the RMSE of the JA estimator based on 3, 4 or 8 sub-samples is smaller than that of the original estimator. The bias and RMSE of the II estimator for the three values of $S$ are reported in Table 2.3 in the Appendix. We see that the bias reduces with $S$ as suggested by equation (2.44). The RMSE also decreases as $S$ becomes larger.

Figures 2.2 and 2.3 show the bias and RMSE of the JA estimator that minimises bias and performs better than the original estimator in terms of RMSE, this being the one based on 3 sub-samples. We see that $\hat{a}_{J,3}$ produces considerable bias reductions, despite not removing completely the first order bias when $a$ approaches zero. The bias and RMSE of $\hat{a}_{II,10000}$ and $\hat{a}_{B,100}$ are also reported in Figures 2.2 and 2.3. Related to the II estimator, the graphs indicate that it performs as well as the BC estimator. A similar picture emerges for the BOOT estimator. There are no important differences between II, BOOT and BC estimators in terms of bias or RMSE. We also see that BC, II and BOOT estimators produce not only substantial bias reductions for all values of $a$, even those near the unit root, but also smaller RMSEs than the ML and JA estimators (see

[14]Note that we have used the ARMA representation to simulate the indirect inference and bootstrap samples. We could alternatively use the approach employed to generate the original data.

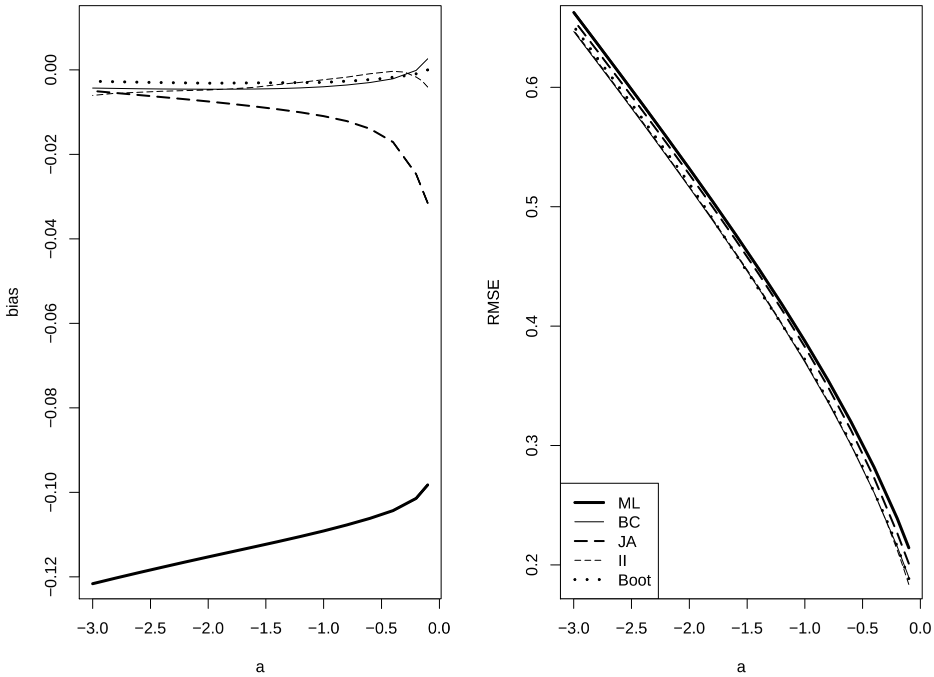


Figure 2.2: Bias and RMSE of the ML, BC, JA, II and BOOT estimators of $a$ in a univariate model with known mean and stock data as a function of $a$. Data are generated using $T = 20$, $h = 1/12$ (i.e. $n = 240$) and $\mu = 0$. The dark solid line corresponds to the ML estimator and the solid line to the BC estimator. The dark dashed line corresponds to the JA estimator, the dashed line to the II estimator and the dotted line to the BOOT estimator.

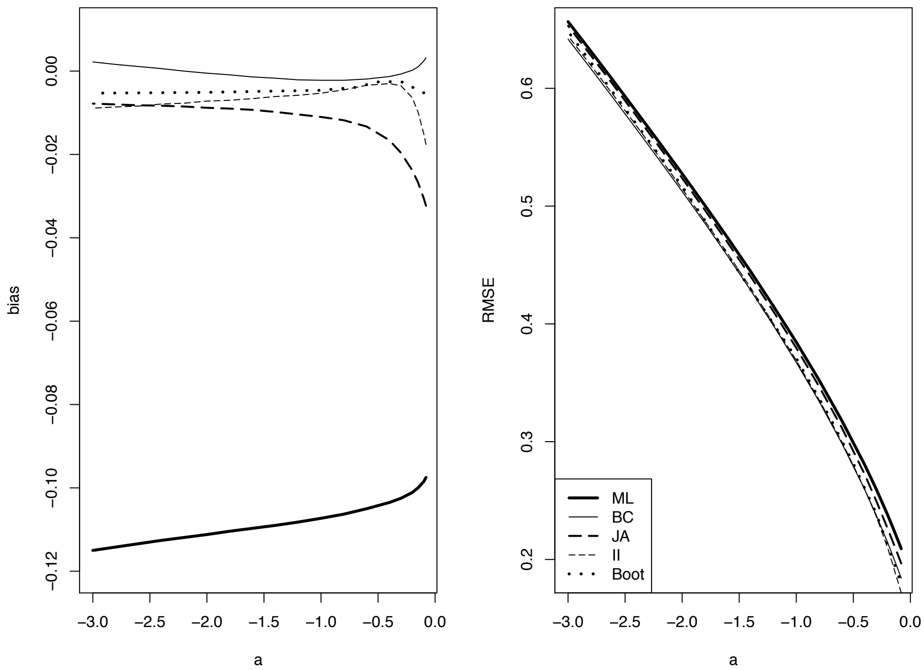


Figure 2.3: Bias and RMSE of the ML, BC, JA, II and BOOT estimators of $a$ in a univariate model with known mean and flow data as a function of $a$. Data are generated using $T = 20$, $h = 1/12$ (i.e. $n = 240$) and $\mu = 0$. The dark solid line corresponds to the ML estimator and the solid line to the BC estimator. The dark dashed line corresponds to the JA estimator, the dashed line to the II estimator and the dotted line to the BOOT estimator.

also Table 2.4 in the Appendix, which summarises these results for some values of $a$).

Tables 2.5, 2.6 and 2.7 in the Appendix, and Figures 2.4 and 2.5 report the corresponding biases and RMSEs of $\hat{a}$ in the model with unknown mean. Results for the JA estimators are similar to those for the known mean case, reductions in the RMSE come at the cost of increased bias. Table 2.6 indicates that the bias of the II estimator reduces with $S$. From the figures we observe that the JA estimator seem to remove more than the first order bias. The graphs also suggest that the BC, II and BOOT estimators perform better than the JA estimator in terms of RMSE, even when larger values of the number of sub-samples $m$ are used (see also Tables 2.5 and 2.7).

### 2.4.2 Multivariate model

We now turn to the multivariate case. We consider the bivariate model with known mean used by Wang et al. (2011)

$$dX(t) = AX(t)dt + \Sigma dB(t), \tag{2.70}$$

where $X(t) = (X_1(t), X_2(t))'$, $B(t)$ is the standard bivariate Brownian motion whose components are independent and the matrices $A$ and $\Sigma$ are given by

$$A = \begin{pmatrix} a_{11} & 0 \\ a_{21} & a_{22} \end{pmatrix} \quad \text{and} \quad \Sigma = \sigma^2 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

with $a_{11} = -0.7$, $a_{21} = -0.5$, $a_{22}$ taking different values over the interval $[-3, 0)$, and $\sigma^2 = 1$.[15] We generate data from the exact discrete model

$$X_h = FX(0) + \epsilon_h, \tag{2.71}$$

$$X_{th} = FX_{th-h} + \epsilon_{th}, \quad t = 2, ..., n, \tag{2.72}$$

where the parameter matrix $F$ is given by

$$F = e^{Ah} = \begin{pmatrix} f_{11} & 0 \\ f_{21} & f_{22} \end{pmatrix},$$

with $f_{11} = e^{a_{11}h}$, $f_{22} = e^{a_{22}h}$ and $f_{21}$ a function of the continuous time parameters. The vector $\epsilon_{th}$ is white noise with covariance matrix $\Omega = \sigma^2 V$ with $V = \int_0^h e^{rA} e^{rA'} dr$.

The experiment is replicated 10000 times, using $X(0) = 0$, $h = 1/12$ and $T = 20$. The restrictions on the exact discrete representation implied by the triangular form of

[15] Results are invariant to other choices of $\sigma^2$.

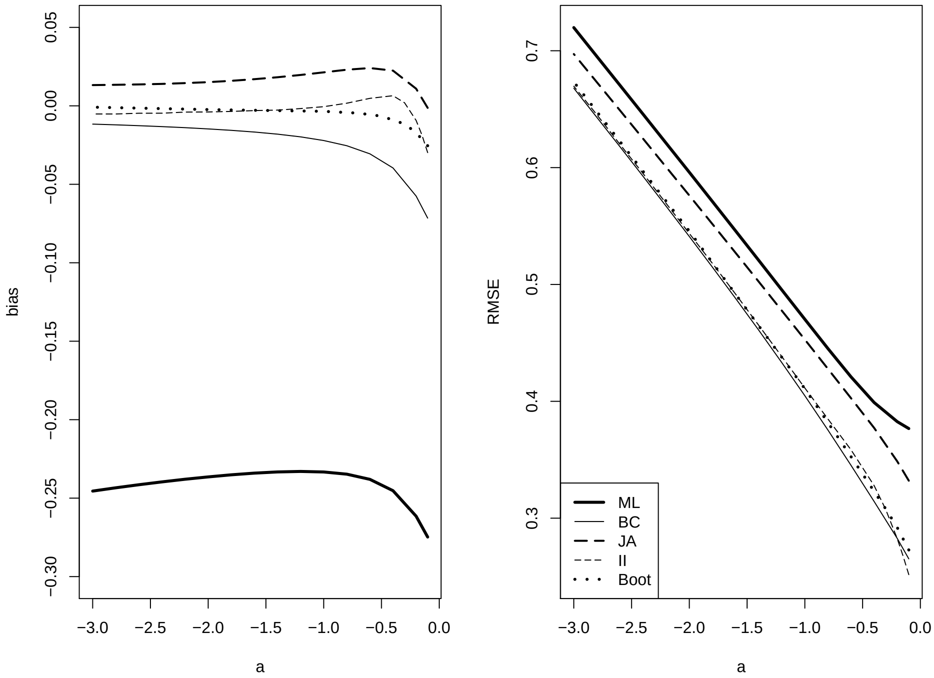


Figure 2.4: Bias and RMSE of the ML, BC, JA, II and BOOT estimators of $a$ in a univariate model with unknown mean and stock data as a function of $a$. Data are generated using $T = 20$, $h = 1/12$ (i.e. $n = 240$) and $\mu = -0.1$. The dark solid line corresponds to the ML estimator and the solid line to the BC estimator. The dark dashed line corresponds to the JA estimator, the dashed line to the II estimator and the dotted line to the BOOT estimator.

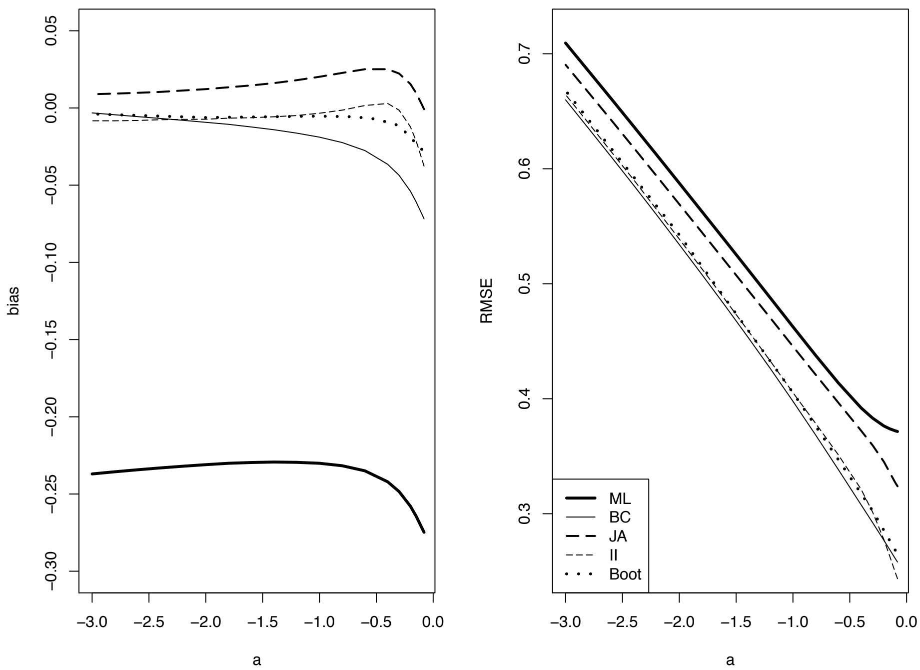


Figure 2.5: Bias and RMSE of the ML, BC, JA, II and BOOT estimators of $a$ in a univariate model with unknown mean and flow data as a function of $a$. Data are generated using $T = 20$, $h = 1/12$ (i.e. $n = 240$), and $\mu = -0.1$. The dark solid line corresponds to the ML estimator and the solid line to the BC estimator. The dark dashed line corresponds to the JA estimator, the dashed line to the II estimator and the dotted line to the BOOT estimator.

the matrix $A$ are taken into account for both, simulation and estimation. As for the univariate case, the parameter vector is estimated by maximum likelihood. To reduce the dimension of the optimisation problem we concentrate the likelihood function with respect to $\Sigma$. It is straightforward to show that the concentrated conditional log-likelihood function is (ignoring a constant)

$$l = -\frac{n-1}{2} \ln \left| \sum_{t=2}^{n} \epsilon_{th} \epsilon_{th}' \right|. \tag{2.73}$$

We also compute the BC, JA, II, and BOOT estimators, and compare their bias and RMSE with those of the ML estimator. To implement indirect inference in the multivariate context we use the following algorithm. First, use ML estimates obtained from observed data $\hat{\Phi} = (\hat{a}_{11}, \hat{a}_{21}, \hat{a}_{22})'$ to generate $S$ simulated paths for $X_{th}^s$, $s = 1, ..., S$. Each path is generated recursively according to

$$X_{th}^s = e^{\hat{A}h} X_{th-h}^s + \hat{\Omega}^{\frac{1}{2}} U_{th}^s, \tag{2.74}$$

where $U_{th}^s \sim i.i.dN(0, I)$. The estimates $\{\hat{\Phi}^s(\hat{\Phi})\}_{s=1}^S$ obtained from the artificial data are then employed to compute a new estimate of $\Phi$

$$\hat{\Phi}^{(1)} = \hat{\Phi} + \lambda \left( \hat{\Phi} - \frac{1}{S} \sum_{s=1}^{S} \hat{\Phi}^s(\hat{\Phi}) \right), \tag{2.75}$$

with $\lambda$ a given scalar between 0 and 1. Second, using $\hat{\Phi}^{(1)}$ and the same pseudo-observations for $\{U_{th}^s\}_{s=1}^S$ generate $S$ new simulated paths for $X_{th}^s$, and update the estimate of $\Phi$ by

$$\hat{\Phi}^{(2)} = \hat{\Phi}^{(1)} + \lambda \left( \hat{\Phi} - \frac{1}{S} \sum_{s=1}^{S} \hat{\Phi}^s(\hat{\Phi}^{(1)}) \right). \tag{2.76}$$

Third, repeat step two until convergence, that is, until the change in $\hat{\Phi}^{(q)}$ with respect to the estimate obtained in the previous step (i.e. $\hat{\Phi}^{(q-1)}$) be sufficiently small.

This algorithm leads to a sequence of $Q$ estimates $\hat{\Phi}^{(q)}$, $q = 1, ..., Q$, with the last value $\hat{\Phi}^{(Q)}$ being the indirect inference estimate.[16] Note that if this sequence converges, $\hat{\Phi}^{(Q)}$ is the value of $\hat{\Phi}^{(q)}$ that minimizes the objective function $\left\| \hat{\Phi} - \frac{1}{S} \sum_{s=1}^{S} \hat{\Phi}^s(\hat{\Phi}^{(q)}) \right\|$. We use $\lambda = 1$ and $\lambda = 0.2$ for those cases in which the algorithm does not seem to

[16] Gouriéroux et al. (2000) uses this algorithm to implement the indirect inference method in a discrete time AR(2) model. Results obtained by the authors suggest that the indirect inference estimator computed with this numerical algorithm performs as well as the approximately median unbiased estimators of Rudebusch (1992) and Andrews and Chen (1994). A similar algorithm has been also used by MacKinnon and Smith (1998) and Smith et al. (1997).

be converging.[17] The convergence criterion used is $|\hat{\Phi}_j - \frac{1}{S}\sum_{s=1}^{S}\hat{\Phi}_j^s(\hat{\Phi}_j^{(q)})| \leq 0.0002$ for $j = 1, 2, 3$. The number of simulated paths is set to $S = 100$. We do not consider larger values of $S$ because the implementation of the algorithm is highly computationally expensive. Note that a different sequence $\hat{\Phi}^{(q)}$ has to be obtained for each Monte Carlo simulation. Moreover, the algorithm seems to work well with this chosen value of $S$, although it may be desirable to employ larger values of $S$ for empirical applications.

Alternatively, a grid search procedure could be employed to obtain the indirect inference estimates of $\Phi = (a_{11}, a_{21}, a_{22})$. The key advantage of the grid search approach is that $\frac{1}{S}\sum_{s=1}^{S}\hat{\Phi}^s(\Phi)$ is computed only once for each of the possible values of $\Phi$, and this quantity can be then used to calculate the objective function for each Monte Carlo simulation without any additional simulation or resampling. However, different from the univariate case the implementation of the grid search method may be computationally expensive in the multivariate context, since different combinations of the elements of $\Phi$ would have to be considered. For example, if 10 possible values for each parameter are chosen, the grid search method requires evaluating the function 10000 times. The number of evaluations increases exponentially as the grid becomes finer. This would be 8000 if 20 values of each parameter are considered and 125000 when the grid evaluates each parameter at 50 points.

The bootstrap samples are obtained using

$$X_{th}^s = e^{\hat{A}h}X_{th-h}^s + U_{th}^s, \tag{2.77}$$

where $U_{th}^s$ is a random draw with replacement from the centered residuals $\hat{\epsilon}_{th}$. The number of bootstrap repetitions is $S = 100$. Figures 2.6-2.8 plot the bias and RMSE of the ML, BC, JA, II and BOOT estimators of the parameter matrix $A$ as a function of the continuous time parameter $a_{22}$ using 10000 replications. The figures only show the JA estimator based on 3 sub-samples, results for other values of $m$ are reported in Table 2.8 in the Appendix (see also Table 2.9 which summarizes the results for the multivariate case).

We observe that the bias of the BC estimator of $a_{11}$ is positive, which is opposite to that of the ML estimator, and larger than those of its rivals JA, II and BOOT estimators. Nonetheless, the BC estimator exhibits the smallest RMSE. As opposed to the JA estimator based on 2 sub-samples, all estimators have smaller RMSE than the benchmark. The BC, II, BOOT and JA estimators of the parameters $a_{21}$ and $a_{22}$ produce substantial bias reductions, being the first three ones which offer the best

[17] As pointed out by MacKinnon and Smith (1998) larger values of $\lambda$ lead to a faster convergence but a lower probability that the sequence will converge. The authors recommend to start with $\lambda = 1$ and then use lower values of $\lambda$ for thoses cases in which the algorithm does not seem to be converging.

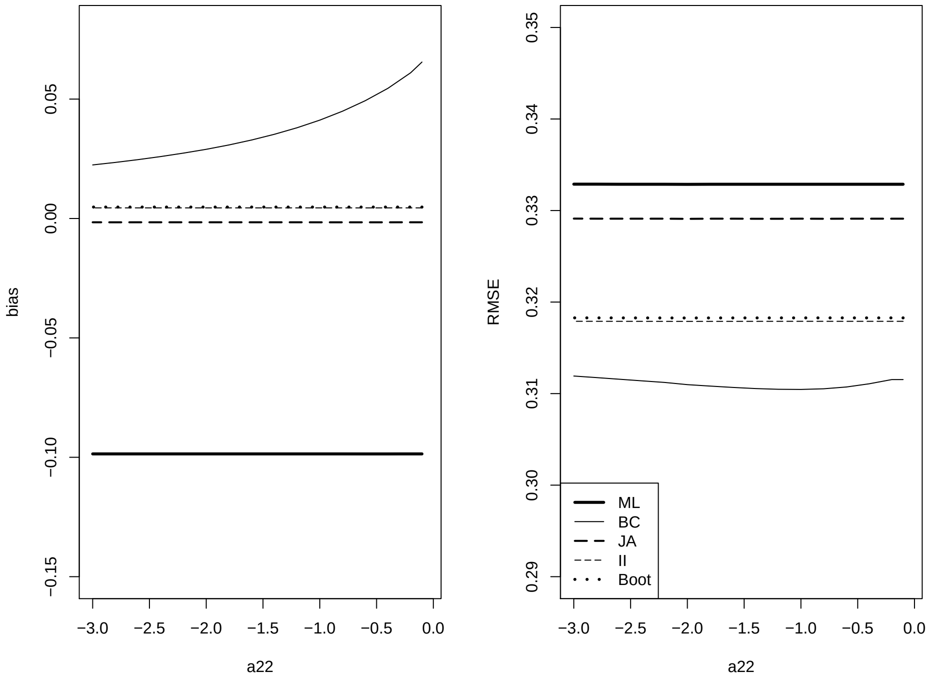


Figure 2.6: Bias and RMSE of the ML, BC, JA, II and BOOT estimators of the parameter $a_{11}$ in a multivariate model with stock data as a function of $a_{22}$. The graphs correspond to $T = 20$, $h = 1/12$ (i.e. $n = 240$) and $B = 0$. The dark solid line corresponds to the ML estimator and the solid line to the BC estimator. The dark dashed line corresponds to the JA estimator, the dashed line to the II estimator and the dotted line to the BOOT estimator.

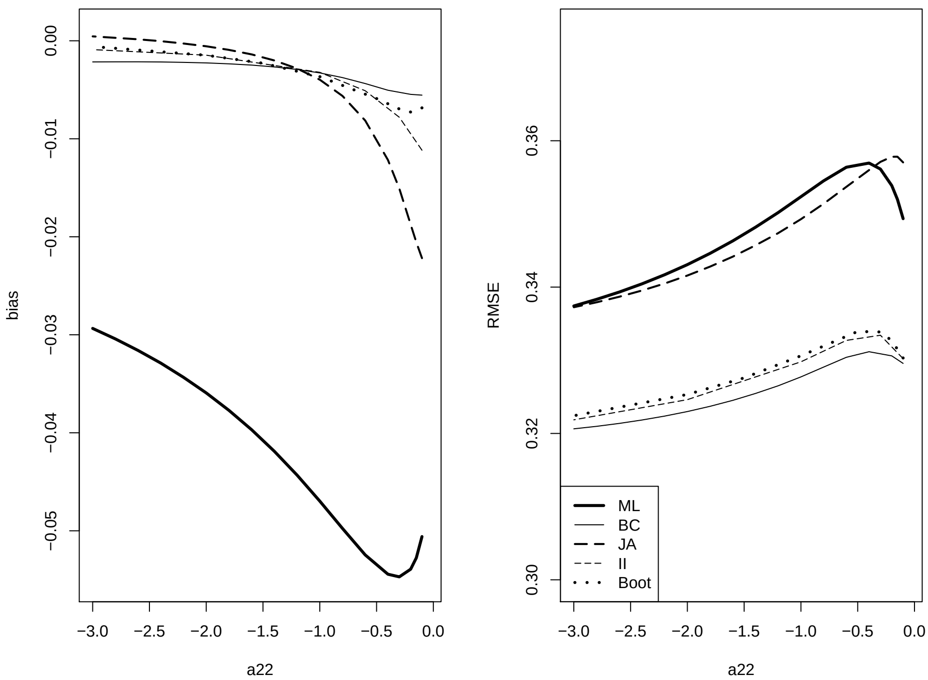


Figure 2.7: Bias and RMSE of the ML, BC, JA, II and BOOT estimators of the parameter $a_{21}$ in a multivariate model with stock data as a function of $a_{22}$. The graphs correspond to $T = 20$, $h = 1/12$ (i.e. $n = 240$) and $B = 0$. The dark solid line corresponds to the ML estimator and the solid line to the BC estimator. The dark dashed line corresponds to the JA estimator, the dashed line to the II estimator and the dotted line to the BOOT estimator.

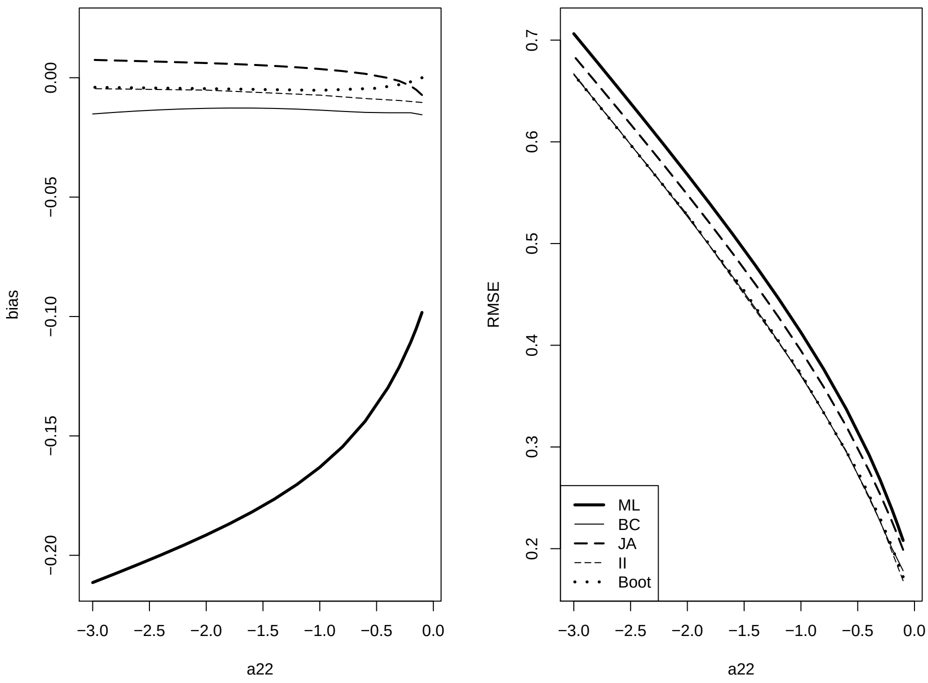


Figure 2.8: Bias and RMSE of the ML, BC, JA, II and BOOT estimators of the parameter $a_{22}$ in a multivariate model with stock data as a function of $a_{22}$. The graphs correspond to $T = 20$, $h = 1/12$ (i.e. $n = 240$) and $B = 0$. The dark solid line corresponds to the ML estimator and the solid line to the BC estimator. The dark dashed line corresponds to the JA estimator, the dashed line to the II estimator and the dotted line to the BOOT estimator.

performance in terms of RMSE, even when larger values of $m$ are used to compute the JA estimator. From Table 2.8 we also see that similar to the univariate case reductions in the RMSE of the JA estimator come at the cost of increased bias.

Although not producing the smallest RMSE, the JA estimator is much cheaper to implement that those based on simulations. The jackknife method requires estimating the model $m + 1$ times for each Monte Carlo simulation, while bootstrap requires generating $S$ new pseudo-data sets and estimating the model $S + 1$ times. Indirect inference is even more expensive. As pointed out by Gouriéroux et al. (2000), the bootstrap method performs only the first step of the algorithm employed to compute the indirect inference estimator for the choice $\lambda = 1$. If the number of simulated paths are the same in both methods, computing $\hat{\Phi}^{(1)}$ is as expensive as computing $\hat{\Phi}_{B,S}$. The iterated algorithm requires estimating the model $S$ times again for each additional iteration, making the procedure highly expensive. This additional computational cost is, however, offset by the gains derived from a better evaluation of the bias used to correct the initial estimator.[18] These gains are clearly visible when the ML estimator is inconsistent as will be shown later.

[18] Note that the $q$ step estimator can be written as $\hat{\Phi}^{(q)} = (1-\lambda)\hat{\Phi}^{(q-1)} + \lambda\left(\hat{\Phi} - \text{bias}\right)$, where bias $= \frac{1}{S}\sum_{s=1}^{S}\hat{\Phi}^{s}(\hat{\Phi}^{(q-1)}) - \hat{\Phi}^{(q-1)}$. The $q$ step estimator is, therefore, a combination of the previous step estimator and the bias corrected estimator based on observed data.

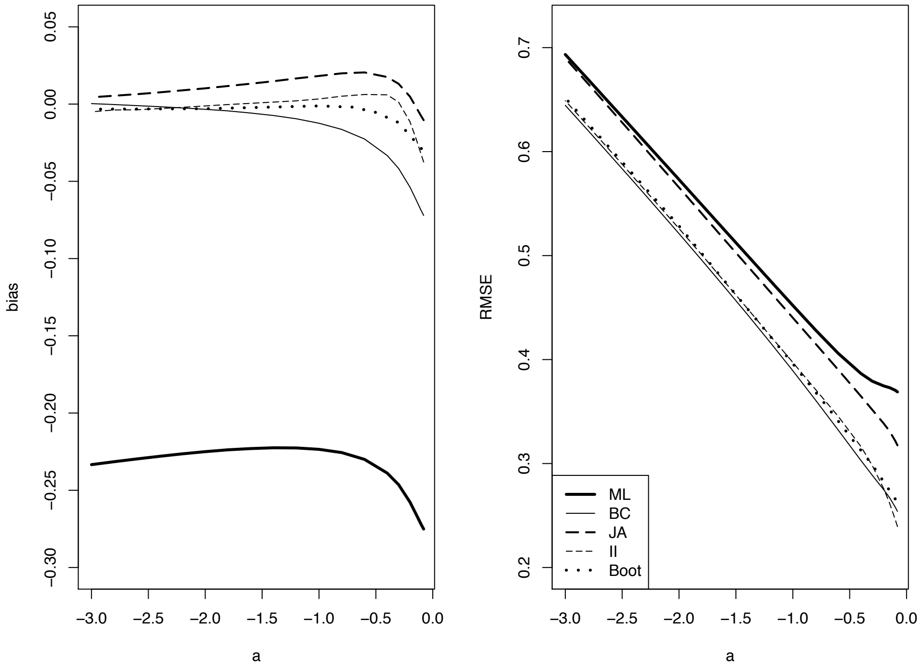


Figure 2.9: Bias and RMSE of the ML, BC, JA, II and BOOT estimators of $a$ in a univariate model with unknown mean, flow data and Gamma disturbances as a function of $a$. Data are generated using $T = 20$, $h = 1/12$ (i.e. $n = 240$), and $\mu = -0.1$. The dark solid line corresponds to the ML estimator and the solid line to the BC estimator. The dark dashed line corresponds to the JA estimator, the dashed line to the II estimator and the dotted line to the BOOT estimator.

We now study the performance of the bias reduction methods under model misspecifications. We only present results for one of the models considered in this section, this being the univariate continuous time model with unknown mean and the variable of interest assumed to be a flow.

### 2.4.3 Non-normal and heteroskedastic disturbances

So far we have assumed that the disturbances are homoskedastic and follow a normal distribution. In this section we explore the robustness of the results to departures from these assumptions. To investigate the effect of non-normality, we generate the discrete time disturbances according to (2.61)-(2.63) with $\vartheta_{th} \sim \Gamma(1,1)-1$. It is assumed therefore that $\vartheta_{th}$ follows a Gamma distribution with raw moments $m_1 = 0$, $m_2 = 1$, $m_3 = 2$ and $m_4 = 9$. Note that under the assumption of Gamma discrete time disturbances $\zeta(dt) = \sigma dL(t)$, where $L(t)$ is a Lévy process. The simulated bias and RMSE reported in Figure 2.9 are smaller in magnitude than the corresponding values under normal disturbances. Similar to the case of normal errors, BC, BOOT and II estimators dominate in terms of RMSE.

To investigate the effect of heteroskedasticity we consider the model

$$dx(t) = [ax(t) + b]dt + |x(t)|^{\alpha} dB(t). \tag{2.78}$$

The Euler scheme is used to obtain an approximate discrete model, from which data are generated. It is desirable to use a sufficiently small step size in order to increase the accuracy of the approximation. To do this, we partition each original interval $[(j-1)h, jh]$ $j = 1, ..., n$, into 100 new sampling intervals. Let $\delta = h/100$ denote the new sampling interval. It should be noted that as the partition becomes finer, the accuracy of the approximation increases. However, the amount of computation required is inevitably higher as well. The simulated data from the Euler scheme are obtained as following

$$x_{i\delta} = (a\delta + 1)x_{(i-1)\delta} - a\mu\delta + |x_{(i-1)\delta}|^{\alpha}\delta^{1/2}\upsilon_{i\delta}, \tag{2.79}$$

where $i = 1, ..., \frac{h}{\delta}, \frac{h}{\delta} + 1, ..., \frac{2h}{\delta},\ ... \frac{h}{\delta}(n-1) + 1, ..., \frac{h}{\delta}n$ and $\upsilon_{i\delta} \sim i.i.dN(0, 1)$.

This simulated trajectory can be seen as a nearly exact continuous sample path of the process $x(t)$. Observations from $(t-1)h + \delta$ to $th$, $t = 1, ..., n$, are then used to compute $x^s_{th}$ as following

$$x^s_{th} = \sum_{j=1}^{h/\delta} \delta x_{(t-1)h+j\delta}, \quad t = 1, ..., n. \tag{2.80}$$

Note that the integral $\int_{th-h}^{th} x(r)dr$ is approximated by the sum of the observations over the interval $(t-1)h + \delta$ to $th$.

To take into account heteroskedastic disturbances, a recursive-design wild bootstrap estimator of the type proposed by Gonçalves and Kilian (2004) is used. Each bootstrap sample is computed recursively according to

$$x^s_{th} = e^{\hat{a}h}x^s_{th-h} + \hat{c} + w^s_{th} + \hat{\theta}w^s_{th-h}, \tag{2.81}$$

where $w^s_{th} = \hat{\varepsilon}_{th}\upsilon_{th}$, with $\hat{\varepsilon}_{th} = x_{th} - e^{\hat{a}h}x_{th-h} - \hat{c} - \hat{\theta}\varepsilon_{th-h}$ and $\upsilon_{th} \sim i.i.dN(0, 1)$.

Figures 2.10 and 2.11 show the bias and RMSE of the ML, BC, BOOT, II and JA estimators of $a$ in the univariate model with $\alpha = \{1.5, 1\}$. The second value of $\alpha$, as well as small values of $a$ (i.e. those close to zero), generate larger departures from homoskedasticity. As can be seen, the bias is bigger than the corresponding values under homoskedastic errors. We observe that heteroskedasticity changes the curvature of the bias function, which is now much more negative as $a$ approaches zero. The BOOT and JA estimators are more robust to heteroskedasticity than the other estimators. Heteroskedasticity also has a big impact on RMSE, with this figure being larger in magnitude. Although all methods still offer considerable bias reductions, the remaining

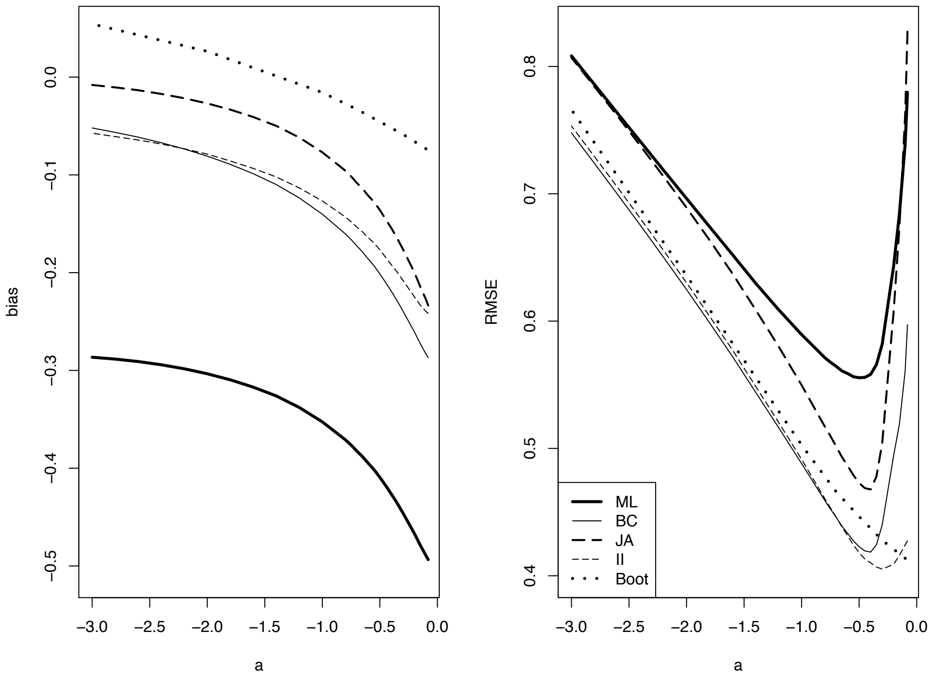


Figure 2.10: Bias and RMSE of the ML, BC, JA, II and BOOT estimators of $a$ in a univariate model with unknown mean, flow data and heteroskedastic disturbances as a function of $a$. Data are generated using $T = 20$, $h = 1/12$ (i.e. $n = 240$), $\mu = -0.1$ and $\alpha = 1.5$. The dark solid line corresponds to the ML estimator and the solid line to the BC estimator. The dark dashed line corresponds to the JA estimator, the dashed line to the II estimator and the dotted line to the BOOT estimator.

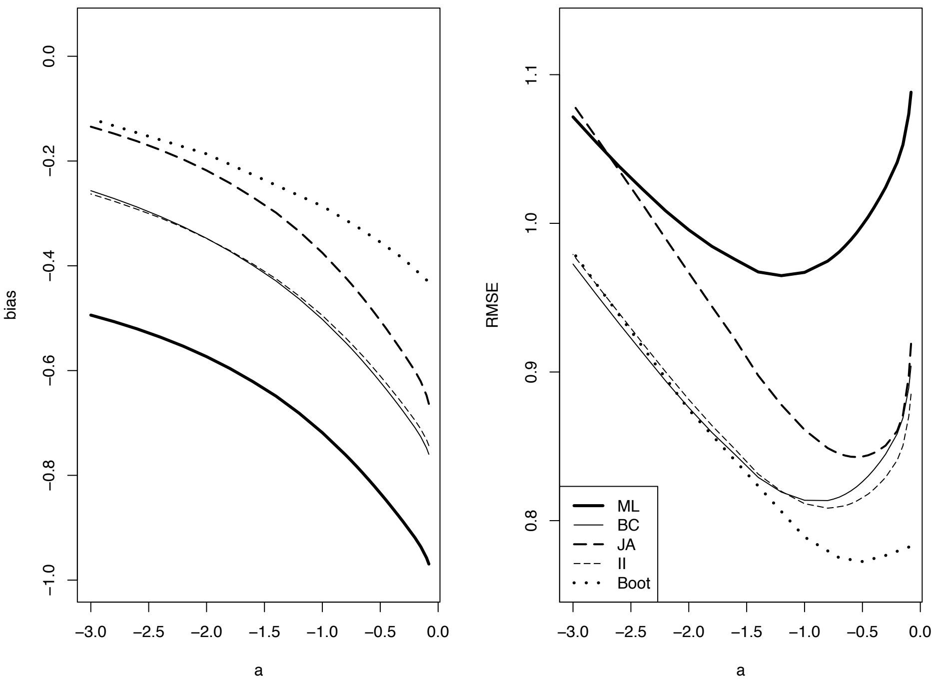


Figure 2.11: Bias and RMSE of the ML, BC, JA, II and BOOT estimators of $a$ in a univariate model with unknown mean, flow data and heteroskedastic disturbances as a function of $a$. Data are generated using $T = 20$, $h = 1/12$ (i.e. $n = 240$), $\mu = -0.1$ and $\alpha = 1$. The dark solid line corresponds to the ML estimator and the solid line to the BC estimator. The dark dashed line corresponds to the JA estimator, the dashed line to the II estimator and the dotted line to the BOOT estimator.

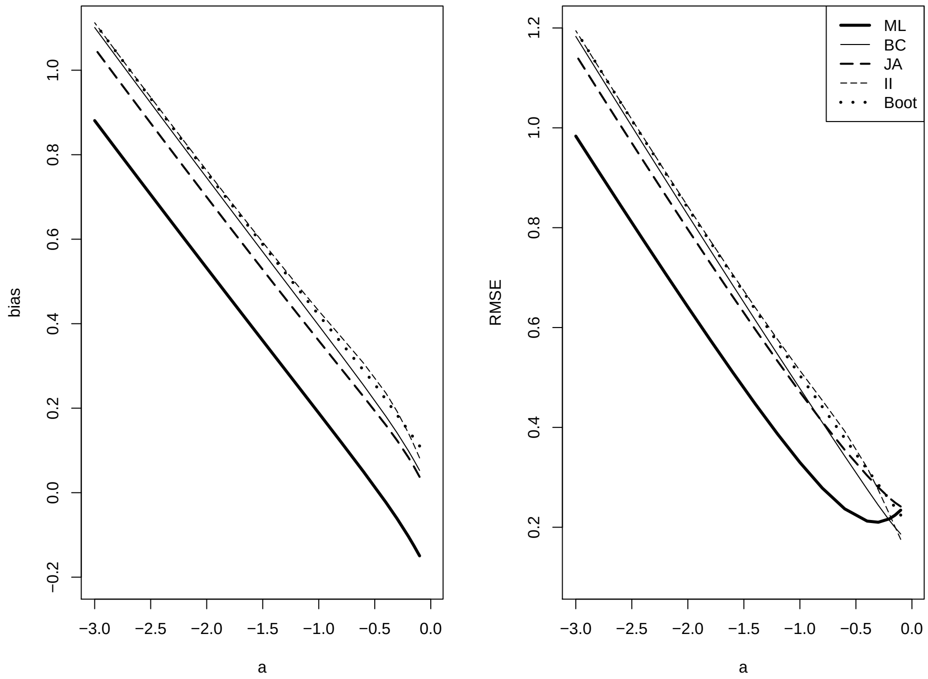


Figure 2.12: Bias and RMSE of the ML, BC, JA, II and BOOT estimators of $a$ in a univariate model with unknown mean and flow data when correlated errors are ignored. Data are generated using $T = 20$, $h = 1/12$ (i.e. $n = 240$), and $\mu = -0.1$. The dark solid line corresponds to the ML estimator and the solid line to the BC estimator. The dark dashed line corresponds to the JA estimator, the dashed line to the II estimator and the dotted line to the BOOT estimator.

bias is larger than when the disturbances are homoskedastic (normal or non-normal).

### 2.4.4 Misspecified model

In this section we explore the robustness of the results to departures from white noise disturbances. The bias caused by correlated errors can be much more serious than that induced by the presence of lagged dependent variables because it can be considerably large, and it does not disappear as the span goes to infinity nor as the sampling interval tends to zero. Data are generated from the exact discrete analogue (2.2) with Gaussian errors with variance and first-order autocovariance given by equations (2.9)-(2.11) (i.e. the variable is a flow). The model is estimated by using the concentrated log-likelihood function for a stock variable, that is ignoring the correlation in the disturbances. The simulated data used to compute the indirect inference and bootstrap estimators are generated by using equations (2.66) and (2.68), respectively. Note that the indirect inference and bootstrap samples are obtained by assuming that the data generating process (DGP) is a continuous time AR(1) with stock data, which is different from the DGP used to generate the original data. Figure 2.12 shows the bias and RMSE of the ML, BC, JA, II and BOOT estimators of $a$.

We see that the bias of the ML estimator can be substantial, being especially large when the parameter $a$ moves away from zero. Not surprisingly, the four bias corrected estimators are unable to produce considerable bias reductions. The BC estimator given in (2.17) is not designed to eliminate the bias caused by the serial correlation in the disturbances. Similarly, the JA estimator is unable to remove the asymptotic bias as suggested by the stochastic expansion when the original estimator is inconsistent. Related to corrections based on indirect inference and bootstrap methods, note that the synthetic samples do not mimic the dependence structure of the original data, and hence the estimators $\hat{a}$ and $\hat{a}^s$ enjoy different properties. In particular, $\hat{a}$ is inconsistent while $\hat{a}^s$ is consistent. Results suggest, therefore, that both methods are dependent on the validity of the DGP used to produce the synthetic samples, although some robustness may be expected, as it was shown above for the case of non-normal disturbances.

We also observe that the bias of the four alternative estimators is larger than that of the ML estimator for most of the parameter values. The reason for this is that the bias due to correlated errors has sign opposite to that arising from estimation. The bias of the ML estimator reported in Figure 2.12 is, therefore, the total bias caused by these two mutually offsetting effects. Figure 2.12 also suggests that the bias caused by the serial correlation can be more serious than that induced by the presence of a lagged dependent variable, being the first one which dominates for most of the parameter values. We also see that the bias and RMSE of all estimators are larger in magnitude than the corresponding values when the model is correctly specified (compare Figures 2.5 and 2.12), although the difference reduces as $a$ approaches zero, leading in some cases to the opposite situation.

As suggested by the stochastic expansions, corrections based on indirect inference and bootstrap methods should be able to produce large bias reductions when the original estimator is inconsistent. It is clear from the results obtained in this section that both methods fail to correct the bias because the synthetic samples do not resemble the original data. In particular, they do not preserve the dependence structure of the original observations. In the next section we show that both estimators are capable of considerable bias reductions when the synthetic samples are generated from the true DGP.

### 2.4.5 Bias correction for inconsistent estimators

We now assume that the DGP is known, and explore two modified versions of the bias correction methods reviewed in the previous section, which are based on the misspecified likelihood that ignores the serial correlation in the discrete time disturbances. There are a number of reasons why it would be convenient to use a misspecified likelihood. First of all, it should be noted that working with serially correlated disturbances is more difficult econometrically than working with white noise disturbances. The reason is that the likelihood function becomes more complex. The approach considered in this section offers, therefore, some benefits in terms of computational cost since the likelihood function is not only easier to derive, but also easier to evaluate.

Second, different to the approach based on the correctly specified likelihood function, the first modified version does not require to fully characterize the properties of the discrete time disturbances. Note that in the case of the AR(1) model, the approach based on the correctly specified likelihood function requires knowing the form of the variance and first-order autocovariance of $\eta_{th}$, while for the misspecified likelihood case only the precise form of the variance of $\eta_{th}$ needs to be known in order to compute an estimate for the continuous time parameter $\sigma$. The second modified version does not require deriving the exact discrete representation, and could be applied to more general models, for example those involving non-linearities.

We illustrate both approaches by considering two univariate continuous time models whose discrete time representations contain a moving average component. The former is the same model studied above, a continuous time AR(1) with unknown mean and the variable of interest assumed to be a flow, and the latter is a continuous time AR(2) with known mean and the variable assumed to be a stock. Note that in both models the discrete time disturbances admit the representation $\eta_{th} = \varepsilon_{th} + \theta\varepsilon_{th-h}$, where $\theta$ is a function of the continuous time autoregressive parameters.

For the AR(1) model, original data are generated as described in last section. The BC estimator can be easily modified to reduce not only the bias arising from estimation, but also the bias induced by the correlation in the disturbances. The asymptotic bias caused by this second effect is presented in the following theorem.

**Theorem 2.4.1.** *Under model $x_{th} = \phi x_{th-h} + c + \eta_{th}$ where $\eta_{th}$ is a first-order moving average, the asymptotic bias of the ML estimator $\hat{a} = \ln\hat{\phi}/h$ is*

$$\text{plim}(\hat{a} - a) = \frac{1}{h}\ln\left[\frac{\gamma_1}{e^{ah}\gamma_0^{(x)}} + 1\right], \tag{2.82}$$

*where $\gamma_1$ is the first-order autocovariance of $\eta_{th}$ and $\gamma_0^{(x)}$ is the variance of $x_{th}$.*

A bias corrected (BC) estimator can be, therefore, computed as[19]

$$\tilde{a} = \hat{a} + \frac{5 + 2e^{-\hat{a}h} + e^{-2\hat{a}h}}{2T} - \frac{1}{h}\ln\left[\frac{\hat{\gamma}_1}{e^{\hat{a}h}\hat{\gamma}_0^{(x)}} + 1\right]. \tag{2.83}$$

[19]Note that we have subtracted the asymptotic correlation bias. A more accurate estimator could be calculated by subtracting the $T^{-1}$ order term of an expansion for the correlation bias as well. This derivation is, however, more difficult to obtain, and it is not considered in this paper.

To avoid the derivation of explicit expressions for $\gamma_1$ and $\gamma_0^{(x)}$, in this section we employ the sample autocovariance of $\hat{\eta}_{th}$ and the sample variance of $x_{th}$.

As suggested by the stochastic expansion, the JA estimator is not able to eliminate the asymptotic bias when the estimator is inconsistent. In order to improve the corrections offered by this method we subtract the asymptotic bias from the estimate based on the whole sample as well as from the estimates based on the sub-samples. We then compute the JA estimator in the usual way, that is as

$$\hat{\Phi}_{J,m} = w\ddot{a} + w_m \frac{1}{m}\sum_{i=1}^{m} \ddot{a}_i, \tag{2.84}$$

where the estimates that uses the whole sample and sub-samples are given by

$$\ddot{a} = \hat{a} - \frac{1}{h}\ln\left[\frac{\hat{\gamma}_1}{e^{\hat{a}h}\hat{\gamma}_0^{(x)}} + 1\right] \quad \text{and} \quad \ddot{a}_i = \hat{a}_i - \frac{1}{h}\ln\left[\frac{\hat{\gamma}_1}{e^{\hat{a}h}\hat{\gamma}_0^{(x)}} + 1\right].$$

So far we have used the exact discrete model to simulate the artificial data employed to compute the indirect inference and bootstrap estimators. We now explore an alternative approach based on an approximate discrete model obtained from the Euler scheme. Simulated data are generated as follows

$$x_{i\delta} = (a\delta + 1)x_{i\delta-\delta} - a\mu\delta + \sigma\delta^{1/2}\upsilon_{i\delta}, \tag{2.85}$$

where $i = 1, ..., \frac{h}{\delta}, \frac{h}{\delta} + 1, ..., \frac{2h}{\delta}, \ \ldots \ , \frac{h}{\delta}(n-1) + 1, ..., \frac{h}{\delta}n$ and $\delta = h/20$. For bootstrap $\upsilon_{i\delta}$ is a random draw with replacement from the standardized residuals and for indirect inference $\upsilon_{i\delta} \sim i.i.dN(0,1)$. Observations from $(t-1)h + \delta$ to $th$, $t = 1, ..., n$, are then used to compute $x^s_{th}$ as following

$$x^s_{th} = \sum_{j=1}^{h/\delta} \delta x_{(t-1)h+j\delta}, \quad t = 1, ..., n. \tag{2.86}$$

The advantage of using this approach to generate the synthetic samples instead of the one described in section 2.4.1 is that it does not require knowing the form of either $\theta$, $\gamma_0$, or $\gamma_1$. The II estimator is computed by using the grid search reviewed in section 2.4.1. As before, for the estimation stage we use the log-likelihood function for a stock variable as the auxiliary misspecified model. The graph in the left of Figure 2.13 reports the ML and the four bias corrected estimators. We see that the BC, JA and BOOT estimators correct only part of the bias caused by the serial correlation. In contrast, the bias of the ML estimator is succesfully corrected by the II estimator.

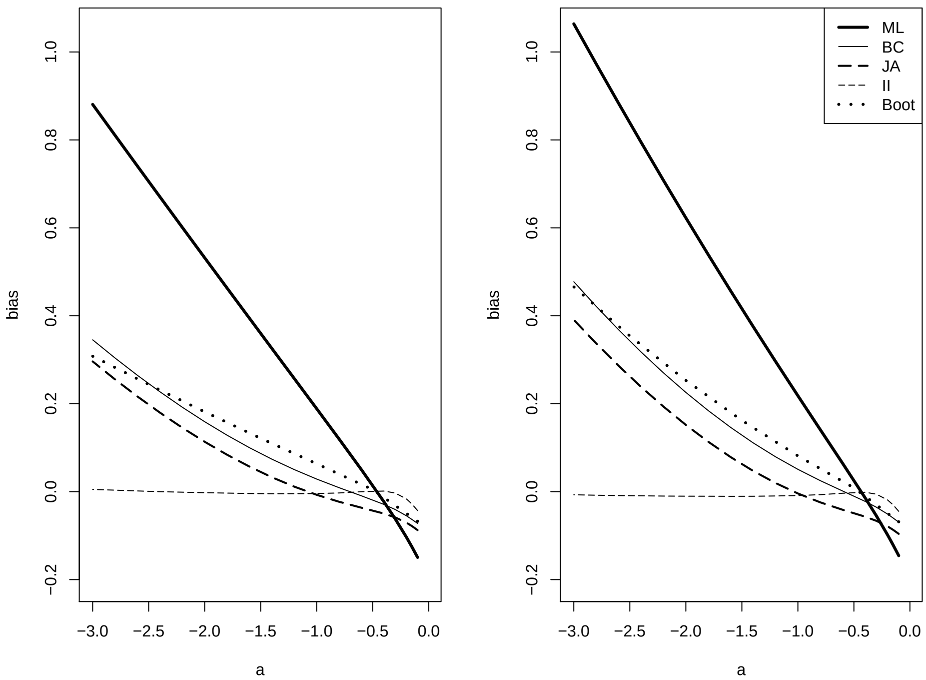


Figure 2.13: Bias of the ML, BC, JA, II and BOOT estimators of $a$ in a univariate model with unknown mean and flow data when correlated errors are ignored and the DGP is assumed to be known. Data are generated using $T = 20$, $h = 1/12$ (i.e. $n = 240$) and $\mu = -0.1$. The estimator in the left is $\hat{a} = \frac{1}{h}\ln\hat{\phi}$, while the one in the right is $\hat{a} = \frac{1}{h}(\hat{\phi} - 1)$. The dark solid line corresponds to the ML estimator and the solid line to the BC estimator. The dark dashed line corresponds to the JA estimator, the dashed line to the II estimator and the dotted line to the BOOT estimator.

Until now we have used the exact discrete representation to estimate the continuous time parameters. Although formulae for the exact discrete analogue corresponding to a general linear system of stochastic differential equations are available in the literature (see Bergstrom, 1990; Chambers, 1999), the precise form of the coefficient matrices and especially the covariance matrices of the disturbances are increasingly complicated as the continuous time model becomes more complex. An alternative estimator of the continuous time parameters can be obtained from the Euler approximate discrete model. The asymptotic bias of this estimator is given in Theorem 2.4.2.

**Theorem 2.4.2.** *Under model $x_{th} = \phi x_{th-h} + c + \eta_{th}$ where $\eta_{th}$ is a first-order moving average, the asymptotic bias of the ML estimator $\hat{a} = (\hat{\phi} - 1)/h$ is*

$$\mathrm{plim}(\hat{a} - a) = \frac{1}{h}\left[\frac{\gamma_1}{\gamma_0^{(x)}} + H\right], \tag{2.87}$$

*where $H = e^{ah} - 1 - ah$.*

According to Theorem 2.4.2, the asymptotic bias can be decomposed into two parts, the correlation bias and the discretization bias, which are given by $\gamma_1/h\gamma_0^{(x)}$ and $H/h$, respectively. We can use this result to compute a new BC estimator

$$\tilde{a} = \hat{a} + \frac{5 + 2e^{-\hat{a}h} + e^{-2\hat{a}h}}{2T} - \frac{1}{h}\left[\frac{\hat{\gamma}_1}{\hat{\gamma}_0^{(x)}} - e^{\hat{a}h} + 1 + \hat{a}h\right], \tag{2.88}$$

and a new JA estimator, where the estimates $\ddot{a}$ and $\ddot{a}_i$ are now given by

$$\ddot{a} = \hat{a} - \frac{1}{h}\left[\frac{\hat{\gamma}_1}{\hat{\gamma}_0^{(x)}} - e^{\hat{a}h} + 1 + \hat{a}h\right]$$

and

$$\ddot{a}_i = \hat{a}_i - \frac{1}{h}\left[\frac{\hat{\gamma}_1}{\hat{\gamma}_0^{(x)}} - e^{\hat{a}h} + 1 + \hat{a}h\right].$$

The graph in the right of Figure 2.13 reports the ML and the four bias corrected estimators based on the Euler approximate discrete model. We see from this figure that the bias of the estimators is larger than that based on the exact discrete analogue, implying that the sign of the discretization bias is positive. As shown by Wang et al. (2011), the signs of the discretization bias and estimation bias are opposite to each other. Similar to the estimator based on the exact discrete model, the BC, JA and BOOT estimators correct only part of the correlation bias. The reason is that the bias expressions depend on $a$, but we are employing the estimates, which are very different from the true value of the parameter, especially when $a$ moves away from zero. For example, when $a = -3$ the discretization bias $H/h$ is equal to 0.35, while it is only 0.15 if $\hat{a}$ is used instead of $a$. In contrast, the II estimator appears to enjoy good properties as suggested by the stochastic expansions when the original estimator is inconsistent.

For the second case, we consider the continuous time AR(2) model with known mean

$$d[Dx(t)] = [a_1 Dx(t) + a_2 x(t)]dt + \sigma dB(t),\ t > 0, \tag{2.89}$$

where $D$ is the mean square differential operator. Bergstrom (1983) showed that the exact discrete analogue corresponding to (2.89) is

$$x_h = [e^{h\bar{A}}]_{11}x(0) + [e^{h\bar{A}}]_{12}Dx(0) + \eta_h, \tag{2.90}$$

$$x_{th} = \left([e^{h\bar{A}}]_{22} + [e^{h\bar{A}}]_{11}\right)x_{th-h} + \left([e^{h\bar{A}}]_{12}[e^{h\bar{A}}]_{21} - [e^{h\bar{A}}]_{22}[e^{h\bar{A}}]_{11}\right)x_{th-2h} + \eta_{th}, \tag{2.91}$$

for $t = 2, ..., n$, where $[e^{h\bar{A}}]_{ij}$ $(i, j = 1, 2)$ is the $ij^{\text{th}}$ element of the matrix

$$\begin{bmatrix} [e^{h\bar{A}}]_{11} & [e^{h\bar{A}}]_{12} \\ [e^{h\bar{A}}]_{21} & [e^{h\bar{A}}]_{22} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 0 & 1 \\ a_2 & a_1 \end{bmatrix} h + \frac{1}{2}\begin{bmatrix} 0 & 1 \\ a_2 & a_1 \end{bmatrix}^2 h^2 + \frac{1}{3!}\begin{bmatrix} 0 & 1 \\ a_2 & a_1 \end{bmatrix}^3 h^3 + ...$$
$$= I + \bar{A}h + \frac{1}{2}\bar{A}^2h^2 + \frac{1}{3!}\bar{A}^3h^3 + ...,$$

and the error $\eta_{th}$ is a first-order moving average with variance

$$E(\eta_h^2) = \sigma^2 \int_0^h [e^{(h-s)\bar{A}}]_{12}^2 ds, \tag{2.92}$$
$$E(\eta_{th}^2) = \sigma^2 \left( \int_0^h [e^{s\bar{A}}]_{12}^2 ds + [e^{h\bar{A}}]_{22}^2 \int_0^h [e^{s\bar{A}}]_{12}^2 ds + [e^{h\bar{A}}]_{12}^2 \int_0^h [e^{s\bar{A}}]_{22}^2 ds \right.$$
$$\left. - 2[e^{h\bar{A}}]_{12}[e^{h\bar{A}}]_{22} \int_0^h [e^{s\bar{A}}]_{22}[e^{s\bar{A}}]_{12} ds \right), \tag{2.93}$$

for $t = 2, ..., n$, and first-order autocovariance

$$E(\eta_{th}\eta_{th-h}) = \sigma^2 \left( [e^{h\bar{A}}]_{12} \int_0^h [e^{s\bar{A}}]_{12}[e^{s\bar{A}}]_{22} ds - [e^{h\bar{A}}]_{22} \int_0^h [e^{s\bar{A}}]_{12}^2 ds \right). \tag{2.94}$$

Disturbances $\eta_{th}$ are generated by using the Cholesky decomposition of the covariance matrix of $\eta = (\eta_h, ..., \eta_{nh})$, that is as $\eta = M\varepsilon$, where $\varepsilon \sim N(0, I)$ and $M$ is a lower triangular real matrix with positive elements on the diagonal such that $MM' = \Omega$, with $\Omega$ being the covariance matrix. It is assumed that $x(0) = Dx(0) = 0$. We fix the parameter $a_1 = -3$ and consider different values for $a_2$ over the interval $[-3, 0)$.[20] We use $h = 1/12$ and $T = 20$, corresponding to 20 years of monthly data. Similar to the AR(1) model, the synthetic samples used to compute the II and BOOT estimators are generated from the approximate discrete model obtained by applying the Euler scheme[21]

$$x_{i\delta} = (a_1\delta + 2)x_{i\delta-\delta} + (a_2\delta^2 - a_1\delta - 1)x_{i\delta-2\delta} + \sigma\delta^{3/2}\upsilon_{i\delta}, \tag{2.95}$$

where $i = 1, ..., \frac{h}{\delta}, \frac{h}{\delta}+1, ..., \frac{2h}{\delta}, \ ... \ , \frac{h}{\delta}(n-1)+1, ..., \frac{h}{\delta}n$ and $\delta = h/20$, $\upsilon_{i\delta}$ is a random draw with replacement from the standardized residuals for bootstrap and $\upsilon_{i\delta} \sim i.i.dN(0, 1)$ for indirect inference. Every $(h/\delta)^{\text{th}}$ observation is then choosen to form the sequence $\{x_h^s, x_{2h}^s, ..., x_{nh}^s\}$. The II estimator is calculated by using the algorithm described in section 2.4.2. The number of simulated paths $S$ is set to 100.[22]

[20]Note that $a_2 < 0$ implies stationarity, the process having a root near unity when $a_2$ approaches zero. Note too that by allowing the paramater to vary over the range $[-3, 0)$ we are considering different scenarios, equal real roots when $a_2 = -2.25$ and complex or different real roots for other values of $a_2$.

[21]The Euler approximate discrete model corresponding to the continuous time AR(2) model is derived in the Appendix.

[22]As can be seen below the procedure with the chosen value of $S$ seems to work well for the Monte Carlo simulations of the continuous time AR(2) model considered here. In empirical applications it would

Because of the difficulty of solving $([e^{h\bar{A}}]_{22}+[e^{h\bar{A}}]_{11})$ and $([e^{h\bar{A}}]_{12}[e^{h\bar{A}}]_{21}-[e^{h\bar{A}}]_{22}[e^{h\bar{A}}]_{11})$ for $a_1$ and $a_2$, the asymptotic biases of $\hat{a}_1$ and $\hat{a}_2$ are only derived by using the approximate discrete model obtained by the Euler scheme. The results are presented in the following theorem.

**Theorem 2.4.3.** *Under model $x_{th} = \phi_1 x_{th-h} + \phi_2 x_{th-2h} + \eta_{th}$ where $\eta_{th}$ is a first-order moving average, the asymptotic biases of the ML estimators $\hat{a}_1 = (\hat{\phi}_1 - 2)/h$ and $\hat{a}_2 = (\hat{\phi}_2 + 1 + \hat{a}_1 h)/h^2$ are*

$$\text{plim}(\hat{a}_1 - a_1) = \frac{1}{h}\left[\frac{\gamma_1\gamma_0^{(x)}}{(\gamma_0^{(x)})^2 - (\gamma_1^{(x)})^2}\right] + \frac{H_1}{h}, \tag{2.96}$$

$$\text{plim}(\hat{a}_2 - a_2) = \frac{1}{h^2}\left[\frac{\gamma_1(\gamma_0^{(x)} - \gamma_1^{(x)})}{(\gamma_0^{(x)})^2 - (\gamma_1^{(x)})^2}\right] + \frac{H_1 + H_2}{h^2}, \tag{2.97}$$

*where $H_1 = [e^{h\bar{A}}]_{22}+[e^{h\bar{A}}]_{11}-2-a_1h$ and $H_2 = [e^{h\bar{A}}]_{12}[e^{h\bar{A}}]_{21}-[e^{h\bar{A}}]_{22}[e^{h\bar{A}}]_{11}-a_2h^2+1+a_1h$.*

The asymptotic bias can be decomposed into two parts, the correlation bias, which is given by the first term of each expression, and the discretization bias, which is given by the second term. Similar to the AR(1) case, the JA estimator can be computed by subtracting the asymptotic bias from the estimate based on the whole sample as well as from the estimates based on the sub-samples.[23] The model is estimated by using the concentrated log-likelihood function corresponding to a discrete time AR(2), that is ignoring the correlation in the disturbances. Figure 2.14 shows the bias of the ML, II and BOOT estimators of the parameters $a_1$ and $a_2$ as a function of $a_2$ based on the exact discrete analogue, while Figure 2.15 reports the bias of the ML, II, JA and BOOT estimators of the continuous time parameters based on the Euler approximate discrete model. A similar pictures emerges from these graphs. The II estimator can deliver substantial bias reductions, outperforming its rivals.

Overall, the simulation study shows that the indirect inference method could be successfully used to avoid the complications of working with correlated errors. Instead of maximizing the exact log-likelihood function, which is of a more complicated form, we can maximize that corresponding to an approximate model with white noise disturbances, that is, ignoring the serial correlation in the disturbances. The estimator

be desirable to use a large number of simulated paths to ensure a good finite sample performance of the indirect inference estimator.

[23]Since the asymptotic bias of the ML estimators has been only derived for the discrete time representation obtained by the Euler scheme, we only compute the JA estimator based on this approximate model.

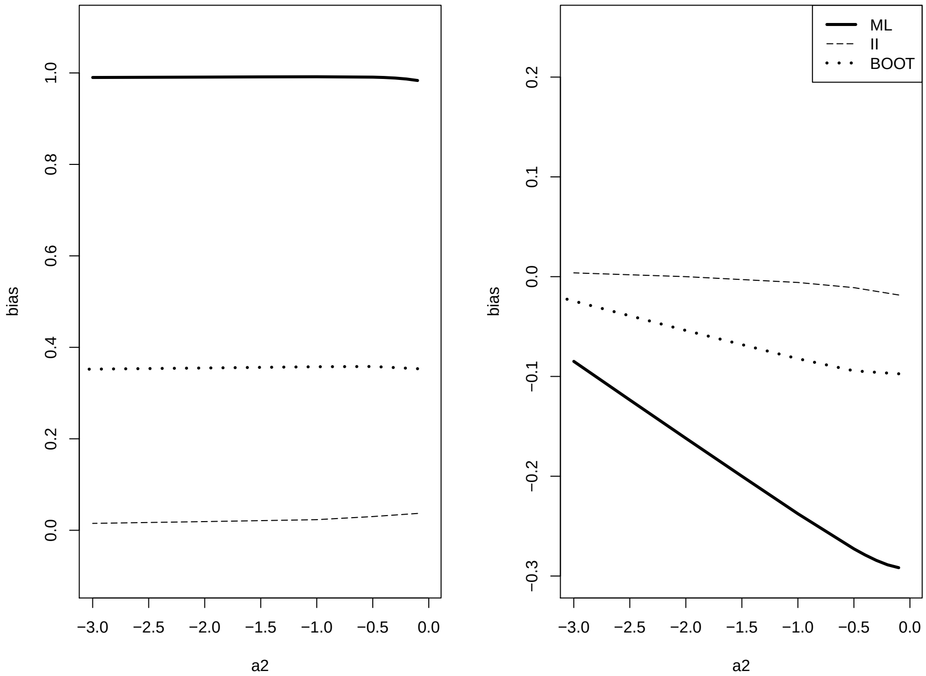


Figure 2.14: Bias of the ML, II and BOOT estimators of $a_1$ and $a_2$ in a univariate AR(2) model with stock data when correlated errors are ignored and the DGP is assumed to be known. Data are generated using $T = 20$, $h = 1/12$ (i.e. $n = 240$) and $\mu = 0$. The graph in the left corresponds to $\hat{a}_1$, while the one in the right to $\hat{a}_2$, based on the exact discrete model. The dark solid line is the bias of the ML estimator. The dashed line to the II estimator and the dotted line to the BOOT estimator.

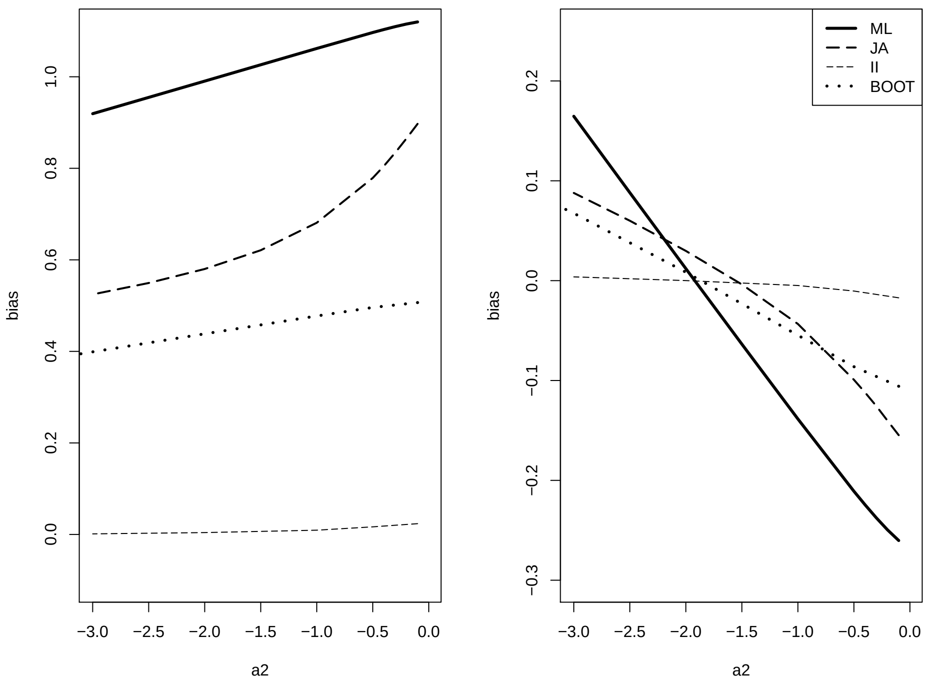


Figure 2.15: Bias of the ML, II, JA and BOOT estimators of $a_1$ and $a_2$ in a univariate AR(2) model with stock data when correlated errors are ignored and the DGP is assumed to be known. Data are generated using $T = 20$, $h = 1/12$ (i.e. $n = 240$) and $\mu = 0$. The graph in the left corresponds to $\hat{a}_1$, while the one in the right to $\hat{a}_2$, based on the Euler approximate discrete model. The dark solid line is the bias of the ML estimator. The dotted line corresponds to the JA estimator with $m = 3$, the dashed line to the II estimator and the dark dashed line to the BOOT estimator.

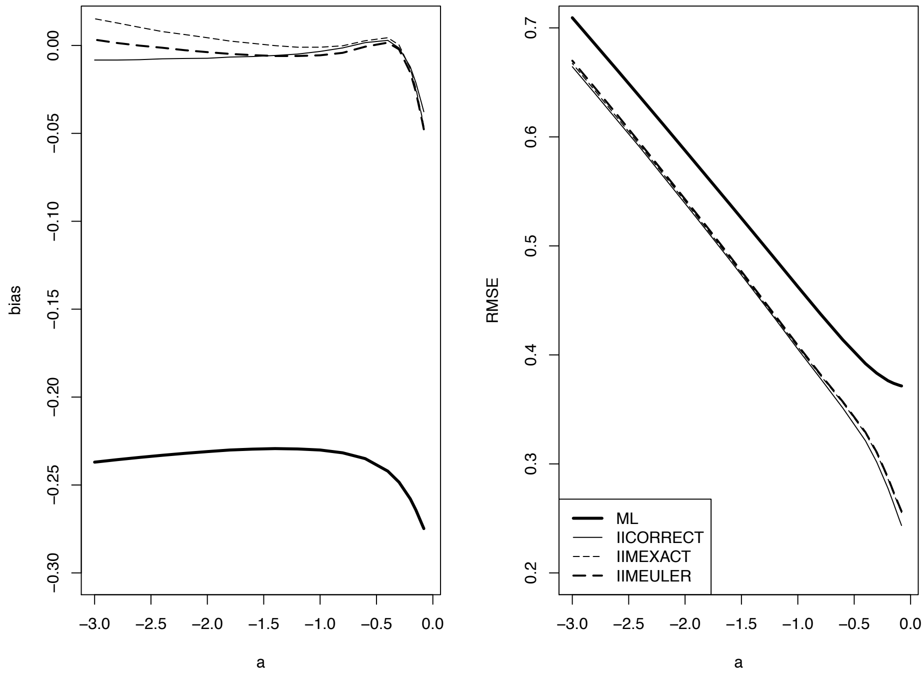


Figure 2.16: Bias and RMSE of the ML and II estimators of $a$ in a univariate model with unknown mean and flow data as a function of $a$. Data are generated using $T = 20$, $h = 1/12$ (i.e. $n = 240$), and $\mu = -0.1$. The dark solid line corresponds to the ML estimator, the solid line to the II estimator based on the correct likelihood function, the dashed lines to the II estimators based on the misspecified likelihood function.

based on this misspecified log-likelihood function is inconsistent. However, the second step of this bias reduction method, which is based on $S$ simulated paths of the initial model, provides a consistent estimator. The alternative bias corrected estimators also produce some important bias reductions, but they are not as great as the ones offered by the II estimator.

We now compare the II estimator based on the correct likelihood function to those based on the misspecified likelihood function. We only present results for the continuous time AR(1) model with a flow variable. Figure 2.16 shows that there are not important differences between the estimator based on the correct likelihood function (i.e. IICORRECT) and those based on the misspecified function (i.e. IIMEXACT and IIMEULER) in terms of bias and RMSE.

## 2.5 Empirical application

The bias reduction methods are applied to the broad effective exchange rate for euro area, obtained from the Bank for International Settlements (BIS). The effective exchange rate, which is calculated as geometric weighted averages of bilateral exchange rates, measures how the euro changes over time, relative to a bundle of other currencies (for an explanation of the methodology used to compute the index see Klau and Fung, 2006). The exchange rate is available as monthly averages and consists of 256 observations, corresponding to the period 1994:1-2015:4.[24] The observed series is averaged to form a monthly series, and hence it suffers from temporal aggregation. We treat it as a flow variable.

We begin by considering an unrestricted discrete ARMA(1,1) model. The estimated equation (with standard errors in parenthesis) is

$$\underset{}{x_{th}} = \underset{(0.011)}{0.972}x_{th-h} + \underset{(1.029)}{2.643} + \hat{\varepsilon}_{th} + \underset{(0.059)}{0.338}\hat{\varepsilon}_{th-h}, \quad \hat{\sigma}^2_{\varepsilon} = 1.579. \tag{2.98}$$

As can be seen all coefficients are statistically significant at the 5% level. We now estimate the discrete time analogue corresponding to a continuous time AR(1) model with flow data, which is an ARMA(1,1) with its three parameters (autoregressive, moving average and variance) being functions of the continuous time parameters $a$ and $\sigma^2$. The estimated equation of the continuous time model (with standard errors in parenthesis)[25] is

$$dx(t) = [\underset{(0.128)}{-0.316}x(t) + \underset{(143.928)}{366.021}]dt + 67.285\, d\hat{B}(t), \tag{2.99}$$

and the corresponding exact discrete analogue is

$$x_{th} = 0.974x_{th-h} + 2.509 + \hat{\varepsilon}_{th} + 0.268\hat{\varepsilon}_{th-h}, \quad \hat{\sigma}^2_{\varepsilon} = 1.587. \tag{2.100}$$

We observe that there are no big differences between the parameter estimates in (2.98) and those in (2.100), except for $\hat{\theta}$.

We also compute the bias corrected estimators considered in this paper. The first row of Table 2.1 shows the estimates of the continuous time parameter $a$ when the likelihood function is correctly specified with their corresponding standard errors. Figures in brackets are the asymptotic standard errors and figures in parenthesis are the standard errors based on the Hessian matrix computed numerically, II and BOOT resamples, respectively. The asymptotic standard error is the square root of the leading term of an analytical variance expression evaluated at $\hat{a}$ (see Section 1.4, also Tang and Chen, 2009). The number of simulated paths used is S = 100 for bootstrap and S = 10000 for indirect inference.[26]

[24]See the Appendix for some additional information about the time series under consideration.

[25]Standard errors are based on the Hessian matrix computed numerically. See also the Appendix for some diagnostic tests.

[26]We also consider $S = 1000$ for indirect inference, but the results do not change much to the ones presented in Table 2.1.

| ML | BC | JA | II | BOOT |
|---|---|---|---|---|
| Correctly specified likelihood | | | | |
| -0.316 | -0.126 | -0.084 | -0.055 | -0.086 |
| [0.174] | [0.109] | [0.089] | [0.072] | [0.090] |
| (0.128) | | | (0.201) | (0.237) |
| Misspecified likelihood - Exact | | | | |
| -0.265 | -0.147 | -0.173 | -0.162 | -0.196 |
| [0.159] | [0.118] | [0.128] | [0.124] | [0.137] |
| (0.106) | | | (0.172) | (0.161) |
| Misspecified likelihood - Euler | | | | |
| -0.262 | -0.145 | -0.183 | -0.163 | -0.198 |
| [0.158] | [0.117] | [0.132] | [0.124] | [0.137] |
| (0.104) | | | (0.167) | (0.156) |

Figures in brackets and parenthesis are standard errors.

Table 2.1: Parameter estimates of $a$ in the continuous time AR(1) model fitted to the effective exchange rate for euro area.

We see from the table that all bias corrected estimates are smaller (in absolute value) than the ML estimate, which is consistent with the Monte Carlo results. We also compute the estimates of $a$ in the discrete time AR(1) model that ignores the serial correlation in the disturbances. As can be seen, the ML estimator is smaller (in absolute value) than the corresponding value in the ARMA(1,1) model, this being consistent with the results obtained in the Monte Carlo simulations. Similar to the case of correctly specified likelihood, the bias corrected estimates are smaller in magnitude (in absolute value) than the ML estimate. Bias corrections differ significantly between approaches. We also observe that there is a substantial discrepancy between asymptotic standard errors and those based on simulations (i.e. bootstrap or indirect inference samples). In spite of these differences, all bias corrected coefficients are not statistically significant at conventional significance levels as opposed to the original estimate.

## 2.6 Conclusions

This paper has conducted a comparison of some bias reduction methods in continuous time models. All approaches deliver substantial bias reductions when the original estimator is consistent. There are not important differences between the bias corrected estimator based on the asymptotic expansion and those based on indirect inference or bootstrap methods. The jackknife tends to inflate the variance of the estimator, leading to larger values of RMSE compared to its competitors. Although not producing the smallest RMSE, the jackknife estimator appears to be very useful in practice since it is computationally much cheaper to implement compared to alternative bias reduction methods. We also explore the robustness of the results to departures from normal, homoskedastic and white noise errors. As opposed to non-normality, heteroskedasticity has a significant impact on all estimators, the bootstrap and jackknife being the most robust estimators.

The bias correction based on indirect inference methods is found to work particularly well when the original estimator is inconsistent, a case ocurring when the serial correlation in the discrete time disturbances is ignored. The bootstrap also produces some important bias reductions, but one iteration seems not to be enough to remove the bias caused by the serial correlation in the disturbances. It would be revelant to explore the implementation of the iterated bootstrap or at least the fast double bootstrap in this case, in which the ML is inconsistent. An empirical application to the broad effective exchange rate data for euro area indicates that there are considerable differences between the bias reduction approaches considered in this paper, although the bias corrected estimates are smaller (in absolute value) than the ML estimate, which is consistent with the results obtained in the Monte Carlo simulations.

## 2.7 Appendix: Tables

### 2.7.1 Results for the univariate model with known mean

| | | Bias | | | | RMSE | | | |
|---|---|---|---|---|---|---|---|---|---|
| | $a$ | $\hat{a}_{J,2}$ | $\hat{a}_{J,3}$ | $\hat{a}_{J,4}$ | $\hat{a}_{J,8}$ | $\hat{a}_{J,2}$ | $\hat{a}_{J,3}$ | $\hat{a}_{J,4}$ | $\hat{a}_{J,8}$ |
| Stock | -0.1 | -0.0255 | -0.0315 | -0.0350 | -0.0443 | 0.2219 | 0.2011 | 0.1942 | 0.1887 |
| | -1 | -0.0082 | -0.0109 | -0.0126 | -0.0164 | 0.3974 | 0.3826 | 0.3765 | 0.3708 |
| | -2 | -0.0065 | -0.0075 | -0.0081 | -0.0082 | 0.5392 | 0.5271 | 0.5224 | 0.5172 |
| | -3 | -0.0054 | -0.0050 | -0.0054 | -0.0032 | 0.6677 | 0.6563 | 0.6525 | 0.6483 |
| Flow | -0.1 | -0.0254 | -0.0305 | -0.0360 | -0.0458 | 0.2257 | 0.2017 | 0.1948 | 0.1887 |
| | -1 | -0.0093 | -0.0110 | -0.0143 | -0.0208 | 0.3953 | 0.3796 | 0.3751 | 0.3690 |
| | -2 | -0.0077 | -0.0088 | -0.0102 | -0.0136 | 0.5348 | 0.5237 | 0.5197 | 0.5169 |
| | -3 | -0.0072 | -0.0078 | -0.0083 | -0.0101 | 0.6623 | 0.6531 | 0.6493 | 0.6448 |

Table 2.2: Bias and RMSE of the JA estimator of $a$ in a univariate model with known mean for different values of $m$.

| | | Bias | | | RMSE | | |
|---|---|---|---|---|---|---|---|
| | $a$ | $\hat{a}_{II,100}$ | $\hat{a}_{II,1000}$ | $\hat{a}_{II,10000}$ | $\hat{a}_{II,100}$ | $\hat{a}_{II,1000}$ | $\hat{a}_{II,10000}$ |
| Stock | -0.1 | -0.0377 | -0.0140 | -0.0040 | 0.1990 | 0.1883 | 0.1836 |
| | -1 | -0.0605 | -0.0244 | -0.0023 | 0.3751 | 0.3733 | 0.3708 |
| | -2 | -0.0616 | -0.0294 | -0.0048 | 0.5212 | 0.5193 | 0.5168 |
| | -3 | -0.0646 | -0.0330 | -0.0060 | 0.6527 | 0.6499 | 0.6478 |
| Flow | -0.1 | -0.0275 | -0.0197 | -0.0150 | 0.1856 | 0.1826 | 0.1781 |
| | -1 | -0.0327 | -0.0232 | -0.0053 | 0.3711 | 0.3710 | 0.3690 |
| | -2 | -0.0367 | -0.0283 | -0.0072 | 0.5175 | 0.5167 | 0.5148 |
| | -3 | -0.0406 | -0.0318 | -0.0089 | 0.6480 | 0.6468 | 0.6446 |

Table 2.3: Bias and RMSE of the II estimator of $a$ in a univariate model with known mean for different values of $S$.

| | | Bias | | | | | RMSE | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | $a$ | $\hat{a}$ | $\tilde{a}$ | $\hat{a}_{J,3}$ | $\hat{a}_{II,10000}$ | $\hat{a}_{B,100}$ | $\hat{a}$ | $\tilde{a}$ | $\hat{a}_{J,3}$ | $\hat{a}_{II,10000}$ | $\hat{a}_{B,100}$ |
| Stock | -0.1 | -0.0982 | 0.0026 | -0.0315 | -0.0040 | -0.0000 | 0.2143 | 0.1897 | 0.2011 | 0.1836 | 0.1885 |
| | -1 | -0.1091 | -0.0040 | -0.0109 | -0.0023 | -0.0030 | 0.3875 | 0.3699 | 0.3826 | 0.3708 | 0.3720 |
| | -2 | -0.1153 | -0.0046 | -0.0075 | -0.0048 | -0.0032 | 0.5319 | 0.5161 | 0.5271 | 0.5168 | 0.5194 |
| | -3 | -0.1216 | -0.0043 | -0.0050 | -0.0060 | -0.0027 | 0.6627 | 0.6467 | 0.6563 | 0.6478 | 0.6509 |
| Flow | -0.1 | -0.0985 | 0.0024 | -0.0305 | -0.0150 | -0.0052 | 0.2141 | 0.1893 | 0.2017 | 0.1781 | 0.1852 |
| | -1 | -0.1073 | -0.0022 | -0.0110 | -0.0053 | -0.0046 | 0.3845 | 0.3673 | 0.3796 | 0.3690 | 0.3708 |
| | -2 | -0.1112 | -0.0005 | -0.0088 | -0.0072 | -0.0051 | 0.5275 | 0.5125 | 0.5237 | 0.5148 | 0.5168 |
| | -3 | -0.1150 | 0.0022 | -0.0078 | -0.0089 | -0.0053 | 0.6566 | 0.6418 | 0.6531 | 0.6446 | 0.6483 |

Table 2.4: Bias and RMSE of the ML, BC, JA, II and BOOT estimators of $a$ in a univariate model with known mean.

### 2.7.2 Results for the univariate model with unknown mean

| | | Bias | | | | RMSE | | | |
|---|---|---|---|---|---|---|---|---|---|
| | $a$ | $\hat{a}_{J,2}$ | $\hat{a}_{J,3}$ | $\hat{a}_{J,4}$ | $\hat{a}_{J,8}$ | $\hat{a}_{J,2}$ | $\hat{a}_{J,3}$ | $\hat{a}_{J,4}$ | $\hat{a}_{J,8}$ |
| Stock | -0.1 | -0.0031 | -0.0013 | 0.0035 | 0.0176 | 0.3851 | 0.3321 | 0.3081 | 0.2851 |
| | -1 | 0.0150 | 0.0213 | 0.0297 | 0.0557 | 0.4904 | 0.4528 | 0.4411 | 0.4305 |
| | -2 | 0.0093 | 0.0151 | 0.0233 | 0.0548 | 0.6051 | 0.5762 | 0.5675 | 0.5605 |
| | -3 | 0.0076 | 0.0132 | 0.0212 | 0.0578 | 0.7229 | 0.6972 | 0.6895 | 0.6829 |
| Flow | -0.1 | -0.0037 | 0.0023 | 0.0000 | 0.0061 | 0.3832 | 0.3271 | 0.3049 | 0.2789 |
| | -1 | 0.0146 | 0.0202 | 0.0242 | 0.0404 | 0.4812 | 0.4455 | 0.4359 | 0.4211 |
| | -2 | 0.0051 | 0.0122 | 0.0145 | 0.0339 | 0.5976 | 0.5692 | 0.5644 | 0.5531 |
| | -3 | 0.0067 | 0.0089 | 0.0138 | 0.0349 | 0.7123 | 0.6904 | 0.6842 | 0.6751 |

Table 2.5: Bias and RMSE of the JA estimator of $a$ in a univariate model with unknown mean for different values of $m$.

| | | Bias | | | RMSE | | |
|---|---|---|---|---|---|---|---|
| | $a$ | $\hat{a}_{II,100}$ | $\hat{a}_{II,1000}$ | $\hat{a}_{II,10000}$ | $\hat{a}_{II,100}$ | $\hat{a}_{II,1000}$ | $\hat{a}_{II,10000}$ |
| Stock | -0.1 | -0.0633 | -0.0415 | -0.0298 | 0.2671 | 0.2578 | 0.2517 |
| | -1 | -0.0384 | -0.0184 | -0.0005 | 0.4061 | 0.4118 | 0.4114 |
| | -2 | -0.0291 | -0.0214 | -0.0039 | 0.5415 | 0.5448 | 0.5437 |
| | -3 | -0.0266 | -0.0238 | -0.0052 | 0.6693 | 0.6707 | 0.6696 |
| Flow | -0.1 | -0.0519 | -0.0387 | -0.0329 | 0.2614 | 0.2535 | 0.2490 |
| | -1 | -0.0306 | -0.0150 | -0.0034 | 0.4023 | 0.4055 | 0.4053 |
| | -2 | -0.0265 | -0.0184 | -0.0073 | 0.5376 | 0.5394 | 0.5389 |
| | -3 | -0.0265 | -0.0209 | -0.0083 | 0.6645 | 0.6656 | 0.6644 |

Table 2.6: Bias and RMSE of the II estimator of $a$ in a univariate model with unknown mean for different values of $S$.

| | | Bias | | | | | RMSE | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | $a$ | $\hat{a}$ | $\tilde{a}$ | $\hat{a}_{J,3}$ | $\hat{a}_{II,10000}$ | $\hat{a}_{B,100}$ | $\hat{a}$ | $\tilde{a}$ | $\hat{a}_{J,3}$ | $\hat{a}_{II,10000}$ | $\hat{a}_{B,100}$ |
| Stock | -0.1 | -0.2748 | -0.0715 | -0.0013 | -0.0298 | -0.0254 | 0.3767 | 0.2652 | 0.3321 | 0.2517 | 0.2728 |
| | -1 | -0.2333 | -0.0221 | 0.0213 | -0.0005 | -0.0035 | 0.4702 | 0.4048 | 0.4528 | 0.4114 | 0.4107 |
| | -2 | -0.2365 | -0.0147 | 0.0151 | -0.0039 | -0.0024 | 0.5960 | 0.5411 | 0.5762 | 0.5437 | 0.5457 |
| | -3 | -0.2455 | -0.0116 | 0.0132 | -0.0052 | -0.0008 | 0.7198 | 0.6680 | 0.6972 | 0.6696 | 0.6733 |
| Flow | -0.1 | -0.2718 | -0.0686 | 0.0023 | -0.0329 | -0.0268 | 0.3721 | 0.2610 | 0.3271 | 0.2490 | 0.2683 |
| | -1 | -0.2301 | -0.0189 | 0.0202 | -0.0034 | -0.0053 | 0.4626 | 0.3978 | 0.4455 | 0.4053 | 0.4053 |
| | -2 | -0.2310 | -0.0093 | 0.0122 | -0.0073 | -0.0062 | 0.5875 | 0.5341 | 0.5692 | 0.5389 | 0.5426 |
| | -3 | -0.2370 | -0.0032 | 0.0089 | -0.0083 | -0.0039 | 0.7093 | 0.6599 | 0.6904 | 0.6644 | 0.6680 |

Table 2.7: Bias and RMSE of the ML, BC, JA, II and BOOT estimators of $a$ in a univariate model with unknown mean.

### 2.7.3 Results for the multivariate model

| | | Bias | | | | RMSE | | | |
|---|---|---|---|---|---|---|---|---|---|
| | $a_{22}$ | $\hat{A}_{J,2}$ | $\hat{A}_{J,3}$ | $\hat{A}_{J,4}$ | $\hat{A}_{J,8}$ | $\hat{A}_{J,2}$ | $\hat{A}_{J,3}$ | $\hat{A}_{J,4}$ | $\hat{A}_{J,8}$ |
| $a_{11}$ | -0.1 | -0.0001 | -0.0016 | -0.0043 | -0.0126 | 0.3463 | 0.3291 | 0.3236 | 0.3163 |
| | -1 | -0.0001 | -0.0016 | -0.0043 | -0.0126 | 0.3463 | 0.3291 | 0.3236 | 0.3163 |
| | -2 | -0.0001 | -0.0016 | -0.0043 | -0.0126 | 0.3463 | 0.3291 | 0.3235 | 0.3163 |
| | -3 | -0.0001 | -0.0016 | -0.0043 | -0.0126 | 0.3463 | 0.3291 | 0.3236 | 0.3163 |
| $a_{21}$ | -0.1 | -0.0195 | -0.0222 | -0.0254 | -0.0290 | 0.3738 | 0.3570 | 0.3506 | 0.3449 |
| | -1 | -0.0026 | -0.0040 | -0.0060 | -0.0078 | 0.3625 | 0.3493 | 0.3432 | 0.3378 |
| | -2 | -0.0001 | -0.0006 | -0.0020 | -0.0017 | 0.3545 | 0.3416 | 0.3360 | 0.3300 |
| | -3 | 0.0006 | 0.0004 | -0.0006 | 0.0008 | 0.3502 | 0.3372 | 0.3322 | 0.3261 |
| $a_{22}$ | -0.1 | -0.0081 | -0.0072 | -0.0073 | -0.0047 | 0.2227 | 0.1989 | 0.1893 | 0.1784 |
| | -1 | 0.0019 | 0.0036 | 0.0033 | 0.0129 | 0.4179 | 0.3947 | 0.3859 | 0.3773 |
| | -2 | 0.0041 | 0.0061 | 0.0075 | 0.0226 | 0.5687 | 0.5481 | 0.5401 | 0.5314 |
| | -3 | 0.0052 | 0.0075 | 0.0106 | 0.0324 | 0.7042 | 0.6845 | 0.6768 | 0.6688 |

Table 2.8: Bias and RMSE of the JA estimators of the elements of $A$ in a multivariate model for different values of $m$.

| | | Bias | | | | | RMSE | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | $a_{22}$ | $\hat{A}$ | $\tilde{A}$ | $\hat{A}_{J,3}$ | $\hat{A}_{II,100}$ | $\hat{A}_{B,100}$ | $\hat{A}$ | $\tilde{A}$ | $\hat{A}_{J,3}$ | $\hat{A}_{II,100}$ | $\hat{A}_{B,100}$ |
| $a_{11}$ | -0.1 | -0.0986 | 0.0655 | -0.0016 | 0.0044 | 0.0048 | 0.3329 | 0.3115 | 0.3291 | 0.3179 | 0.3183 |
| | -1 | -0.0986 | 0.0412 | -0.0016 | 0.0044 | 0.0048 | 0.3329 | 0.3104 | 0.3291 | 0.3179 | 0.3183 |
| | -2 | -0.0986 | 0.0290 | -0.0016 | 0.0044 | 0.0048 | 0.3329 | 0.3110 | 0.3291 | 0.3179 | 0.3183 |
| | -3 | -0.0986 | 0.0224 | -0.0016 | 0.0044 | 0.0048 | 0.3329 | 0.3119 | 0.3291 | 0.3179 | 0.3183 |
| $a_{21}$ | -0.1 | -0.0506 | -0.0055 | -0.0222 | -0.0112 | -0.0068 | 0.3493 | 0.3296 | 0.3570 | 0.3302 | 0.3303 |
| | -1 | -0.0470 | -0.0033 | -0.0040 | -0.0032 | -0.0037 | 0.3523 | 0.3277 | 0.3493 | 0.3298 | 0.3306 |
| | -2 | -0.0359 | -0.0022 | -0.0006 | -0.0015 | -0.0015 | 0.3431 | 0.3230 | 0.3416 | 0.3246 | 0.3253 |
| | -3 | -0.0293 | -0.0021 | 0.0004 | -0.0009 | -0.0006 | 0.3374 | 0.3206 | 0.3372 | 0.3219 | 0.3224 |
| $a_{22}$ | -0.1 | -0.0983 | -0.0155 | -0.0072 | -0.0104 | -0.0000 | 0.2081 | 0.1783 | 0.1989 | 0.1685 | 0.1724 |
| | -1 | -0.1631 | -0.0136 | 0.0036 | -0.0073 | -0.0053 | 0.4127 | 0.3701 | 0.3947 | 0.3716 | 0.3721 |
| | -2 | -0.1915 | -0.0128 | 0.0061 | -0.0052 | -0.0046 | 0.5678 | 0.5266 | 0.5481 | 0.5283 | 0.5279 |
| | -3 | -0.2114 | -0.0152 | 0.0075 | -0.0046 | -0.0040 | 0.7063 | 0.6660 | 0.6845 | 0.6669 | 0.6662 |

Table 2.9: Bias and RMSE of the ML, BC, JA, II and BOOT estimators of the elements of $A$ in a multivariate model.

## 2.8 Appendix: Proofs

*Proof of Theorem 2.4.1.* Maeshiro (1999) showed that the asymptotic correlation effect of the estimator of the parameter $\phi$ in the model $x_{th} = c + \phi x_{th-h} + \eta_{th}$ is given by the term $\text{cov}(\eta_{th}, x_{th-h})/\text{var}(x_{th-h})$, where cov(.) and var(.) denote covariance and variance, respectively. Applying the plim operator to $\hat{a} = \ln(\hat{\phi})/h$ and using the preservation of convergence for continuous transformations (i.e. plim $\ln(\hat{\phi}) = \ln(\text{plim } \hat{\phi})$), we have

$$\begin{aligned}
\text{plim } \hat{a} &= \frac{1}{h}\ln\left[\text{plim } \hat{\phi}\right] \\
&= \frac{1}{h}\ln\left[\frac{\gamma_1}{\gamma_0^{(x)}} + \phi\right] \\
&= \frac{1}{h}\ln\left[\frac{\gamma_1}{\gamma_0^{(x)}} + e^{ah}\right] \qquad (2.101)
\end{aligned}$$

□

*Proof of Theorem 2.4.2.* The coefficient $\phi$ can be written as $\phi = e^{ah} = 1 + ah + H$. Applying the plim operator to $\hat{a} = (\hat{\phi} - 1)/h$ and then substituting $\phi$ by $1 + ah + H$ gives

$$\begin{aligned} \text{plim } \hat{a} &= \frac{1}{h}\left[\text{plim } \hat{\phi}\right] - \frac{1}{h} \\ &= \frac{1}{h}\left[\frac{\gamma_1}{\gamma_0^{(x)}} + \phi - 1\right] \\ &= \frac{1}{h}\left[\frac{\gamma_1}{\gamma_0^{(x)}} + H\right] + a \end{aligned} \tag{2.102}$$

□

*Deriving the Euler approximate discrete model.* The continuous time AR(2) model given in (2.89) can be written in state space form as

$$dy(t) = \bar{A}y(t)dt + \sigma dW(t), \tag{2.103}$$

where

$$y(t) = \begin{bmatrix} x(t) \\ Dx(t) \end{bmatrix}, \quad \bar{A} = \begin{bmatrix} 0 & 1 \\ a_2 & a_1 \end{bmatrix}, \quad dW(t) = \begin{bmatrix} 0 \\ dB(t) \end{bmatrix}.$$

Applying the Euler scheme to (2.103), we obtain

$$y_{th} = (\bar{A}h + I_2)y_{th-h} + \sigma h^{1/2}\omega^*_{th},$$

where

$$y_{th} = \begin{bmatrix} x(th) \\ Dx(th) \end{bmatrix}, \quad \omega^*_{th} = \begin{bmatrix} 0 \\ \omega(th) \end{bmatrix},$$

and $\omega(th) \sim i.i.dN(0,1)$. From the first equation of the system we obtain

$$Dx(th-h) = \frac{1}{h}\left[x(th) - x(th-h)\right] \tag{2.104}$$

$$Dx(th-2h) = \frac{1}{h}\left[x(th-h) - x(th-2h)\right]. \tag{2.105}$$

Lagging the second equation of the system gives

$$Dx(th-h) = a_2 hx(th-2h) + (a_1 h + 1)Dx(th-2h) + \sigma h^{1/2}w(th-h). \tag{2.106}$$

Substituting (2.104) and (2.105) into (2.106) we obtain

$$x(th) = (a_1h + 2)x(th - h) + (a_2h^2 - a_1h - 1)x(th - 2h) + \sigma h^{3/2}w(th - h). \tag{2.107}$$

$\square$

*Proof of Theorem 2.4.3.* The model $x_{th} = \phi_1 x_{th-h} + \phi_2 x_{th-2h} + \eta_{th}$ can be written in state space form as

$$y_{th} = \Pi y_{th-h} + \eta^*_{th}, \tag{2.108}$$

where

$$y_{th} = \begin{bmatrix} x_{th} \\ x_{th-h} \end{bmatrix}, \quad \Pi = \begin{bmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{bmatrix}, \quad \eta^*_{th} = \begin{bmatrix} \eta_{th} \\ 0 \end{bmatrix}.$$

The asymptotic correlation effect of the estimator of $\Pi$ is given by

$$\operatorname{cov}(\eta^*_{th}, y'_{th-h})[\operatorname{var}(y_{th-h})]^{-1} = \begin{bmatrix} \gamma_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \gamma_0^{(x)} & \gamma_1^{(x)} \\ \gamma_1^{(x)} & \gamma_0^{(x)} \end{bmatrix}^{-1} \tag{2.109}$$

$$= \frac{1}{(\gamma_0^{(x)})^2 - (\gamma_1^{(x)})^2} \begin{bmatrix} \gamma_1\gamma_0^{(x)} & -\gamma_1\gamma_1^{(x)} \\ 0 & 0 \end{bmatrix}. \tag{2.110}$$

The coefficients $\phi_1$ and $\phi_2$ can be written as $\phi_1 = [e^{h\bar{A}}]_{22} + [e^{h\bar{A}}]_{11} = 2 + a_1h + H_1$ and $\phi_2 = [e^{h\bar{A}}]_{12}[e^{h\bar{A}}]_{21} - [e^{h\bar{A}}]_{22}[e^{h\bar{A}}]_{11} = a_2h^2 - 1 - a_1h + H_2$. Applying the plim operator to $\hat{a}_1 = (\hat{\phi}_1 - 2)/h$ and to $\hat{a}_2 = (\hat{\phi}_2 + 1 + \hat{a}_1h)/h^2$, using the results given in (2.110) and substituting $\phi_1$ by $2 + a_1h + H_1$ and $\phi_2$ by $a_2h^2 - 1 - a_1h + H_2$, we obtain

$$\begin{aligned}
\text{plim } \hat{a}_1 &= \frac{1}{h}\left[\text{plim}\hat{\phi}_1\right] - \frac{2}{h} \\
&= \frac{1}{h}\left[\frac{\gamma_1\gamma_0^{(x)}}{(\gamma_0^{(x)})^2 - (\gamma_1^{(x)})^2} + \phi_1 - 2\right] \\
&= \frac{1}{h}\left[\frac{\gamma_1\gamma_0^{(x)}}{(\gamma_0^{(x)})^2 - (\gamma_1^{(x)})^2} + H_1\right] + a_1
\end{aligned} \tag{2.111}$$

$$
\begin{aligned}
\text{plim } \hat{a}_2 &= \frac{1}{h^2}\left[\text{plim}\hat{\phi}_2 + h\text{plim}\hat{a}_1\right] + \frac{1}{h^2} \\
&= \frac{1}{h^2}\left[-\frac{\gamma_1\gamma_1^{(x)}}{(\gamma_0^{(x)})^2-(\gamma_1^{(x)})^2} + \phi_2 + 1 + h\text{plim}\hat{a}_1\right] \\
&= \frac{1}{h^2}\left[-\frac{\gamma_1\gamma_1^{(x)}}{(\gamma_0^{(x)})^2-(\gamma_1^{(x)})^2}\right] + a_2 + \frac{1}{h}\text{plim}(\hat{a}_1 - a_1) + \frac{H_2}{h^2} \\
&= \frac{1}{h^2}\left[\frac{\gamma_1\gamma_0^{(x)} - \gamma_1\gamma_1^{(x)}}{(\gamma_0^{(x)})^2-(\gamma_1^{(x)})^2} + H_1 + H_2\right] + a_2 \qquad (2.112)
\end{aligned}
$$

□

## 2.9 Appendix: Additional information for the empirical application

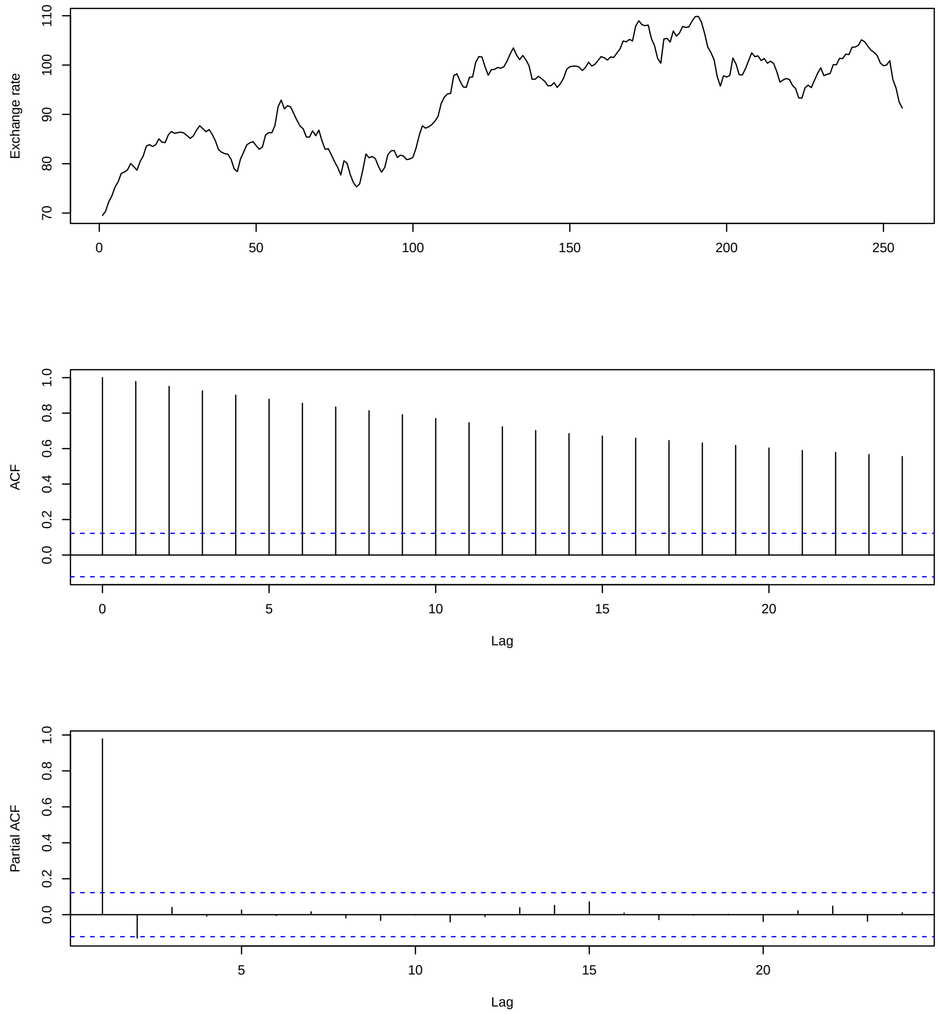


Figure 2.17: Graph, ACF and PACF of the effective exchange rate for euro area.

| | Normality | ARCH effects | | | Serial correlation | | |
|---|---|---|---|---|---|---|---|
| | JB | $A_{(1)}$ | $A_{(4)}$ | $A_{(8)}$ | $LM_{(8)}$ | $LM_{(12)}$ | $LM_{(24)}$ |
| p-value | 0.000 | 0.258 | 0.144 | 0.202 | 0.874 | 0.973 | 0.868 |

JB is the Jarque-Bera test, $A_{(i)}$ is the $i^{\text{th}}$ order ARCH test, $LM_{(i)}$ denotes the $i^{\text{th}}$ order Breusch-Godfrey LM test.

Table 2.10: Diagnostic tests for residuals. We assume that the conditional mean of the time series is described by a continuous time AR(1) model.

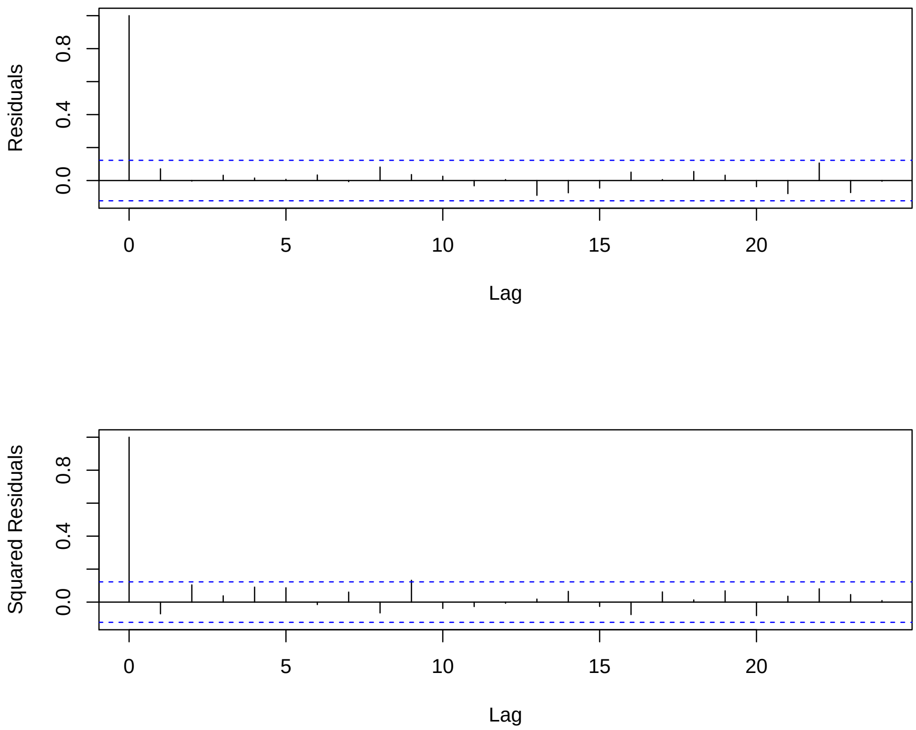


Figure 2.18: ACF of residuals and squared residuals. We assume that the conditional mean of the time series is described by a continuous time AR(1) model.

# Chapter 3

# Cointegrated continuous time models with mixed sample data

## 3.1 Introduction

Continuous time models have been commonly used to describe the dynamics of economic behaviour. Although information is collected at certain frequencies such as quarters, months or years, the movements in the variables do not necessarily coincide with the observational interval, this being the case of macroeconomic information, which is collected at low frequencies (usually quarterly or annually). Continuous time modelling may, therefore, provide a more realistic approximation to the actual dynamics of the economy, which involve a large number of economic agents making decisions at different points of time.

The advantages of this type of models over those formulated in discrete time have been widely discussed in the literature (see for example, Bergstrom, 1996; Bergstrom and Nowman, 2007). Among them is the separate treatment of stock and flow variables. It is explicitly recognised that stocks are observed at specific points of time, while flows are measured as the accumulation of the underlying rate over a time interval. Because of the distinction in the treatment of these two type of variables, these models do not suffer from time aggregation bias whereas it may be a serious problem in their discrete time counterpart.

Continuous time models also have the advantage of allowing for a mixture of both stock and flow variables. Mixed sampling often arises in econometric modelling. A typical money demand model, for example, comprises variables that are measured instantaneously such as prices and exchange rates, and variables that are observed as integrals such as income. Another example is the Fisher effect, this hypothesis being tested by using interest rates (stocks), and expected or actual inflation rate (a flow).

Different approaches have been considered to estimate parameters of continuous time systems with mixed sample data. Among them are Kalman filtering techniques based on state space representations (see Harvey and Stock, 1985, 1988; Zadrozny, 1988), and methods based on spectral representations (see Phillips, 1991a; Robinson, 1993). Another important contribution to estimation methodology is the exact Gaussian estimation method proposed by Bergstrom (1983, 1985, 1986), who pointed out the benefits of this last approach in terms of computational efficiency of estimators. The Gaussian method requires, however, the derivation of the exact discrete analogue that is induced by the continuous system.

In the last decades much work has been done on estimating continuous time models based on the exact discrete analogue. In the context of non-stationary continuous systems, Bergstrom (1997) developed an algorithm for the Gaussian estimation of a mixed first- and second-order stochastic differential equation system with mixed sampling and unobservable stochastic trends, which allows for the possibility of cointegration. However, the exact discrete analogue obtained by the author exhibits some excess differencing since it is written entirely in terms of first differences instead of lagged levels along with lagged differences as it is specified in cointegrated systems formulated directly in discrete time. As noted by Bergstrom (1997, 2009), the Gaussian likelihood remains invariant under differencing, but the discrete time representation is not adequate to investigate the sampling properties of the estimates due to the presence of unit roots in the moving average part of the discrete model. A more precise specification of the discrete analogue is, therefore, required.

Chambers (2009) proposed an alternative approach to derive the exact discrete analogue, which is based on the idea of replacing unobservable components with their observable counterpart, whose difference is then assigned to the disturbance term (see also Phillips, 1991a, who first proposed this idea). The model considered by the author is a triangular system of first-order stochastic differential equations with mixed sampling and observable stochastic trends. The resulting discrete analogue retains the triangular form of the continuous time system, and does not suffer from the excess differencing.

Many economic variables often exhibit a complex correlation structure, which requires the use of higher order differential equations. It is, therefore, relevant to extend the methods used by Chambers (2009) to more complex continuous time models, and this is the main purpose of this paper. We consider a system of mixed first- and second-order stochastic differential equations with mixed sampling and observable stochastic trends, and derive the corresponding exact discrete representation. Some formulae to implement the Gaussian estimation are also provided.

It is well known that parameter estimators in continuous time diffusion processes can suffer from substantial bias in finite samples (see for example Tang and Chen, 2009; Wang et al., 2011). The estimation bias has been widely studied in the context of stationary models, but has not received much attention in the cointegration framework. Another aim of this paper is to explore the finite sample properties of the Gaussian estimator of parameters in cointegrated continuous time systems with mixed sampling. For this purpose, we conduct a Monte Carlo experiment using two bivariate models, a system of first-order stochastic differential equations and a system of first- and second-order stochastic differential equations.

The remainder of this paper is organised as follows. Section 2 briefly describes the first-order model. In Section 3 the mixed-order system is formulated and the exact discrete representation is derived. Section 4 describes the computation of the Gaussian likelihood. Section 5 reports some simulations results to examine the finite sample properties of the Gaussian estimator and Section 6 concludes.

## 3.2 First-order model

We consider a $p$-dimensional continuous time random process $y(\tau)$ that is partitioned into two subvectors $y_1(\tau)$ and $y_2(\tau)$ of dimensions $p_1 \times 1$ and $p_2 \times 1$, respectively, where $p_1 + p_2 = p$. It is assumed that the vector $y(\tau)$ is $I(1)$, that is non-stationary with zero root, but that there exist an $p_1 \times p$ matrix $B'$ whose rows are linearly independent such that $B'y(\tau)$ is stationary. It is also assumed that $y(\tau)$ satisfies the stochastic differential equation system

$$dy(\tau) = [a + b\tau + AB'\ y(\tau)]\, d\tau + u(\tau)d\tau,\ \ \tau > 0, \tag{3.1}$$

where $u(\tau)$ is a $p \times 1$ vector of white noise with covariance matrix $\Sigma d\tau$, $a$ and $b$ are $p \times 1$ vectors of constants, and $A = (C', 0_{p_1 \times p_2})'$ and $B' = (I_{p_1}, -\Pi)$ are coefficient matrices. The matrix $C$, known as the matrix of adjustment coefficients, describes the responsiveness of the system variables to deviations from equilibrium, represented by $y_1(\tau) - \Pi y_2(\tau)$. All eigenvalues of $C$ are assumed to have negative real parts to ensure that $y_1(\tau)$ is $I(1)$ and cointegrated.[1] The matrix $B$ is called a cointegration matrix or cointegration vectors. Cointegration is characterized by the singularity of $AB'$, where $A$ and $B$ are $p \times p_1$ matrices, both of rank $p_1$. Note that the first $p_1$ components of $B$ have been normalized to the identity matrix to ensure the uniqueness of the cointegration matrix.

[1]Note that if some eigenvalues have non-negative real parts $y_1(\tau)$ may be $I(q)$, with $q > 1$, or $y_1(\tau)$ may be $I(1)$ but not cointegrated.

Equation (3.1) can be also written as following

$$dy_1(\tau) = [a_1 + b_1\tau + C(y_1(\tau) - \Pi y_2(\tau))]\, d\tau + u_1(\tau)d\tau,\ \tau > 0, \tag{3.2}$$

$$dy_2(\tau) = [a_2 + b_2\tau]\, d\tau + u_2(\tau)d\tau,\ \tau > 0. \tag{3.3}$$

in which the vectors $a$, $b$ and $u(\tau)$ have been partitioned conformably with $y_1$ and $y_2$. From the last representation it is clear that the stationary increments in $y_1(\tau)$ depend on the deviations from the equilibrium relation. The second equation specifies the $p_2$ common stochastic trends.

The system (3.2)-(3.3) can be seen as a special case of the Phillips' triangular representation[2] augmented with deterministic components

$$y_1(\tau) = \breve{a}_1 + \breve{b}_1\tau + \Pi y_2(\tau) + \nu_1(\tau),\ \tau > 0, \tag{3.4}$$

$$Dy_2(\tau) = a_2 + b_2\tau + u_2(\tau),\ \tau > 0, \tag{3.5}$$

where $D$ is the mean square differential operator and $\nu_1$ is a stationary process satisfying

$$d\nu_1(\tau) = [C\nu_1(\tau) + \theta u_2(\tau)]d\tau + \breve{u}_1(\tau)d\tau. \tag{3.6}$$

It is straightforward to show that the deterministic components of (3.2) are related to those of (3.4) by $a_1 = -C\breve{a}_1 + \breve{b}_1 + \Pi a_2$ and $b_1 = -C\breve{b}_1 + \Pi b_2$. Similarly, the error terms of both systems are related as follows: $u_1(\tau) = (\theta + \Pi)u_2(\tau) + \breve{u}_1(\tau)$.

System (3.1) was considered by Chambers (2009) for the general case in which each of the vectors $y_i(\tau)$ $(i = 1, 2)$ comprises $p_i^s$ stock variables observed at the points of time $0, h, 2h, ..., nh$, where $h$ is the sampling interval (i.e. the period between observations) and $n$ is the total number of observations, and $p_i^f$ flow variables observed as integrals over the intervals $[0, h], [h, 2h], ..., [(n-1)h, nh]$, where $p_i^s + p_i^f = p_i$. It was shown by the author that the exact discrete analogue corresponding to (3.1) is given by[3]

$$y_h - y(0) = \Phi_{10}B'y(0) + \beta_h + \xi_h, \tag{3.7}$$

$$\Delta y_{2h} = \Phi_{21}B'y_h + \Phi_{20}B'y(0) + \beta_{2h} + \xi_{2h}, \tag{3.8}$$

$$\Delta y_{th} = \Phi_0 B'y_{th-h} + \Phi_1\Delta y_{th-h} + \beta_{th} + \xi_{th},\ t = 3, ..., n, \tag{3.9}$$

where the first two equations corresponding to periods $h$ and $2h$, respectively, relate the first two observations (i.e. $y_h$ and $y_{2h}$) to the initial state value $y(0)$, with $\Delta = 1 - L^h$, $L$ being the lag operator such that $L^h y_{th} = y_{th-h}$, and $t$ being the time index

[2]The triangular representation of a cointegrated system was first proposed by Phillips (1991b) for discrete time models and Phillips (1991a) for continuous time models.

[3]System (3.7)-(3.9) differs from that given by Chambers (2009) in that the sampling interval $h$ is normalised to unity in the original paper.

of the observations. The matrices Φ's and the vector $\beta_{th}$ are complicated functions of the parameters of the continuous time system, and the disturbance vector $\xi_{th}$ is a MA(1) process (see Appendix for the definition of these terms).

The exact discrete model is a reduced rank VARMA(2,1) system when $y_1(\tau)$ includes both stocks and flows, and a reduced rank VARMA(1,1) when $y_1(\tau)$ includes one type of variable only, that is either $p_1$ stocks or $p_1$ flows. It was also shown by Chambers (2009) that the last $p_2$ elements of the matrices Φ's are zeros, implying a triangular structure in the exact discrete model.

## 3.3 Mixed-order model

### 3.3.1 The continuous time model and its solution

The disturbances of (3.2) are now assumed to satisfy

$$du_1(\tau) = [\Psi_1 u_1(\tau) + \Psi_2 u_2(\tau)]d\tau + \breve{n}_1(\tau)d\tau, \quad -\infty < \tau < \infty, \tag{3.10}$$

where $\Psi_1$ and $\Psi_2$ are coefficient matrices, and $\breve{n}_1(\tau)$ is a white noise process. It is also assumed that all characteristic roots of the coefficient matrix $\Psi_1$ have negative real parts in order for the process $u_1(\tau)$ to be stationary. Substituting (3.10) into (3.2), we obtain a system comprising a mixture of first- and second-order differential equations with white noise disturbances, which is given by

$$d[Dy_1(\tau)] = \left[\mu_1 + \gamma_1\tau + \Psi Dy_1(\tau) + \tilde{C}\left(y_1(\tau) - \Pi y_2(\tau)\right)\right] d\tau + n_1(\tau)d\tau, \ \tau > 0 \tag{3.11}$$

$$dy_2(\tau) = [a_2 + b_2\tau]\, d\tau + u_2(\tau)d\tau, \ \tau > 0. \tag{3.12}$$

The parameters of the system (3.2)-(3.3) with disturbances given in (3.10) are related to those of (3.11)-(3.12) as follows: $\tilde{C} = -\Psi_1 C$, $\Psi = C + \Psi_1$, $\mu_1 = b_1 - \Psi_1 a_1 - \Pi a_2$, $\gamma_1 = -\Psi_1 b_1 - \Pi b_2$, and $n_1(\tau) = (\Psi_2 - \Pi)u_2(\tau) + \breve{n}_1(\tau)$. It should be noted that only in certain regions of the parameter space there is one to one correspondence between the two parameter sets.[4] To avoid possible identification problems, we shall concentrate on the parameters of the system (3.11)-(3.12) rather than on the original parameter vector. Let $\Theta = (\Theta_1, \Theta_2)$ denote the vector of parameters, where $\Theta_1 = (\mu, \gamma, \Psi, \tilde{C}, \Pi)$, $\mu = (\mu_1', a_2')'$, $\gamma = (\gamma_1', b_2')'$ and $\Theta_2$ contains the elements of the covariance matrix of the disturbance vector $n(\tau) = (n_1(\tau)', u_2(\tau)')'$.

[4] For example, the vector $(\tilde{C} = -1, \Psi = -2))$ corresponds to $(C = -1, \Psi = -1)$. However, the vector $(\tilde{C} = -2, \Psi = -3))$ corresponds to both $(C = -2, \Psi = -1)$ and $(C = -1, \Psi = -2)$. In this last case the parameters of the first system are not identified from the second system.

In this section we shall deal with the problem of estimating $\Theta$ from a sample of discrete observations generated by the system (3.11)-(3.12). The approach adopted here is to derive the exact discrete analogue, which holds exactly for data generated by the continuous time system, and apply the Gaussian methods proposed by Bergstrom (1983, 1985, 1986) on the discrete representation. In order to derive the exact discrete analogue it is convenient to write (3.11)-(3.12) as a first-order system

$$dx(\tau) = \left[\mu^* + \gamma^*\tau + \bar{A}x(\tau)\right] d\tau + n^*(\tau)d\tau, \tag{3.13}$$

where the $(p + p_1) \times 1$ vector $x(\tau)$ is defined by $x(\tau) = [y_1(\tau)', y_2(\tau)', Dy_1(\tau)']'$, and

$$\bar{A} = \begin{pmatrix} 0 & 0 & I_{p_1} \\ 0 & 0 & 0 \\ \tilde{C} & -\tilde{C}\Pi & \Psi \end{pmatrix}, \quad \mu^* = \begin{pmatrix} 0 \\ a_2 \\ \mu_1 \end{pmatrix}, \quad \gamma^* = \begin{pmatrix} 0 \\ b_2 \\ \gamma_1 \end{pmatrix}, \quad n^*(\tau) = \begin{pmatrix} 0 \\ u_2(\tau) \\ n_1(\tau) \end{pmatrix}.$$

From the first-order system representation it is easy to see that cointegration is characterized by the singularity of the matrix $\bar{A}$. We know that subject to the boundary conditions $x(0) = (y_1(0)', y_2(0)', Dy_1(0)')'$, equation (3.13) has a solution given by

$$x(\tau) = e^{\tau\bar{A}}x(0) + \int_0^\tau e^{(\tau-s)\bar{A}}(\mu^* + \gamma^* s)ds + \int_0^\tau e^{(\tau-s)\bar{A}}n^*(s)ds, \ \tau > 0, \tag{3.14}$$

where, for any square matrix A, $e^A = \sum_{i=0}^{\infty} A^i/i!$ defines the matrix exponential.

We turn now to the assumptions about the observability of the continuous time processes $y_1(\tau)$ and $y_2(\tau)$. It is assumed that $y_2(\tau)$ includes $p_2^s$ stock variables and $p_2^f$ flow variables, where $p_2^s + p_2^f = p_2$. We also assume that $y_1(\tau)$ includes one type of variable only, that is, either $p_1$ stocks or $p_1$ flows. The case considered here is less general than that considered in Chambers (2009) since the vector $y_1(\tau)$ is not allowed to include both stocks and flows. The extension of our results to allow for that general case would require additional complications and will not be dealt with in this paper.[5]

We assume that the stocks are observed at the points of time $0, h, 2h, ..., nh$, and the flows as integrals over the intervals $[0, h], [h, 2h], ..., [(n-1)h, nh]$. Without loss of generality, let the elements of the vector $y_2(\tau)$ be ordered as following

$$y_2(\tau) = \begin{pmatrix} y_2^s(\tau) \\ y_2^f(\tau) \end{pmatrix},$$

[5]It should be noted that the method used by Chambers (2009) to eliminate the unobservable components associated with the variable $y_1$ can not be straightforwardly implemented in our context.

and the observations of this vector as

$$y_{2,th} = \begin{pmatrix} y^s_{2,th} \\ y^f_{2,th} \end{pmatrix} = \begin{pmatrix} y^s_2(th) \\ \dfrac{1}{h}\displaystyle\int_{th-h}^{th} y^f_2(r)dr \end{pmatrix},$$

for $t = 1, ..., n$, with $t$ being the time index of the observations, and where the superscripts $s$ and $f$ denote the stock and flow components, respectively. The observable vector for $y_1(\tau)$ is $y_{1,th} = y^s_1(th)$ if $y_1(\tau)$ is composed entirely of stocks and $y_{1,th} = h^{-1}\int_{th-h}^{th} y^f_1(r)dr$ if it is composed entirely of flows. Note that the flow variables have been normalised by dividing by $h$. As noted by Chambers (2011) this normalisation is required in nonstationary/cointegrated systems to ensure the stationarity of all terms in the discrete time representation.

Our interest is to estimate the parameters of the continuous time model from the observable vectors $y_{1,th}$ and $y_{2,th}$. It should be noticed that whereas the vectors of the initial values of the stock variables $y^s_1(0)$ and $y^s_2(0)$ are part of the sample, the vectors of the initial values of the flow variables $y^f_1(0)$ and $y^f_2(0)$ and those of the mean square derivatives of $y_1$ (i.e. $Dy_1(0)$ ) are unobservable. Thus, in addition to the continuous time parameters we could estimate the part of $x(0)$ that is unobservable by using, for example, the iterative estimation procedure proposed by Bergstrom (1990, pg. 113-114).[6]

The derivation of the exact discrete analogue with mixed sampling requires solving out the unobservable vectors $w_{2,th}$ defined by

$$w_{2,th} = \begin{pmatrix} w^s_{2,th} \\ w^f_{2,th} \end{pmatrix} = \begin{pmatrix} \dfrac{1}{h}\displaystyle\int_{th-h}^{th} y^s_2(r)dr \\ y^f_2(th) \end{pmatrix},$$

for $t = 1, ..., n$.

Next section provides some preliminary results that will be then used to derive the exact discrete model with mixed data.

### 3.3.2 The exact discrete model for common data sampling

To derive the exact discrete model satisfied by mixed data we shall make use of two sets of equations. The former is the exact discrete representation that would result if the sample was comprised entirely of stocks, and the latter is the exact discrete model if the

[6]However, as pointed out by Bergstrom (1990) there is no way of obtaining a consistent estimate of $x(0)$. Furthermore, the contribution of the initial condition to the likelihood function tends to be negligible as the sample size increases, implying that for large samples there is little to be gained from estimating $x(0)$.

sample consisted entirely of flows. Before presenting the exact discrete models, we shall make the following assumption.

Assumption 1. The $p_1 \times p_1$ sub-matrix $[e^{h\bar{A}}]_{13}$ is non-singular.[7]

Agbeyegbe (1984) showed that the exact discrete model for the case of stock variables is

$$y(th) = F_1 y(th-h) + F_2 y(th-2h) + \kappa_{th} + \vartheta_{th}, \; t = 2, ..., n, \tag{3.15}$$

where $F_1$ and $F_2$ are $p \times p$ matrices given by

$$F_1 = \begin{pmatrix} F_{1,11} & F_{1,12} \\ F_{1,21} & F_{1,22} \end{pmatrix}, \quad F_2 = \begin{pmatrix} F_{2,11} & F_{2,12} \\ F_{2,21} & F_{2,22} \end{pmatrix},$$

$$\begin{aligned}
F_{1,11} &= [e^{h\bar{A}}]_{11} + [e^{h\bar{A}}]_{13}[e^{h\bar{A}}]_{33}[e^{h\bar{A}}]_{13}^{-1}, \\
F_{1,12} &= \left(I_{p_1} - [e^{h\bar{A}}]_{11}\right)\Pi, \\
F_{1,21} &= F_{2,21} = F_{2,22} = 0, \\
F_{1,22} &= I_{p_2}, \\
F_{2,11} &= [e^{h\bar{A}}]_{13}[e^{h\bar{A}}]_{31} - [e^{h\bar{A}}]_{13}[e^{h\bar{A}}]_{33}[e^{h\bar{A}}]_{13}^{-1}[e^{h\bar{A}}]_{11}, \\
F_{2,12} &= -\left([e^{h\bar{A}}]_{13}[e^{h\bar{A}}]_{31} + [e^{h\bar{A}}]_{13}[e^{h\bar{A}}]_{33}[e^{h\bar{A}}]_{13}^{-1}(I_{p_1} - [e^{h\bar{A}}]_{11})\right)\Pi,
\end{aligned}$$

and $\kappa_{th}$ and $\vartheta_{th}$ are $p \times 1$ vectors given by

$$\begin{aligned}
\kappa_{th} &= \tilde{S}_1 \int_{th-h}^{th} e^{(th-s)\bar{A}}(\mu^* + \gamma^* s)ds + \tilde{S}_2 \int_{th-2h}^{th-h} e^{(th-h-s)\bar{A}}(\mu^* + \gamma^* s)ds, \\
\vartheta_{th} &= \tilde{S}_1 \int_{th-h}^{th} e^{(th-s)\bar{A}} n^*(s)ds + \tilde{S}_2 \int_{th-2h}^{th-h} e^{(th-h-s)\bar{A}} n^*(s)ds,
\end{aligned}$$

with $[e^{h\bar{A}}]_{ij}$ denoting the entry $(i, j)$ of the matrix exponential $e^{h\bar{A}}$ (see the Appendix)

$$e^{h\bar{A}} = \begin{pmatrix} [e^{h\bar{A}}]_{11} & [e^{h\bar{A}}]_{12} & [e^{h\bar{A}}]_{13} \\ [e^{h\bar{A}}]_{21} & [e^{h\bar{A}}]_{22} & [e^{h\bar{A}}]_{23} \\ [e^{h\bar{A}}]_{31} & [e^{h\bar{A}}]_{32} & [e^{h\bar{A}}]_{33} \end{pmatrix} = \begin{pmatrix} [e^{h\bar{A}}]_{11} & (I_{p_1} - [e^{h\bar{A}}]_{11})\Pi & [e^{h\bar{A}}]_{13} \\ 0 & I_{p_2} & 0 \\ [e^{h\bar{A}}]_{31} & -[e^{h\bar{A}}]_{31}\Pi & [e^{h\bar{A}}]_{33} \end{pmatrix},$$

and

$$\tilde{S}_1 = \begin{pmatrix} I_{p_1} & 0 & 0 \\ 0 & I_{p_2} & 0 \end{pmatrix}, \quad \tilde{S}_2 = \begin{pmatrix} -[e^{h\bar{A}}]_{13}[e^{h\bar{A}}]_{33}[e^{h\bar{A}}]_{13}^{-1} & 0 & [e^{h\bar{A}}]_{13} \\ 0 & 0 & 0 \end{pmatrix}.$$

Note that the matrices $F_1$ and $F_2$ have been partitioned conformably with the elements

[7] This sub-matrix is the entry (1,3) of the matrix exponential $e^{h\bar{A}}$ defined below.

of $y(\tau)$ (i.e. $y_1(\tau)$ and $y_2(\tau)$), and the matrix $e^{h\bar{A}}$ conformably with the elements of $x(\tau)$ (i.e. $y_1(\tau)$, $y_2(\tau)$ and $Dy_1(\tau)$). Equation (3.15) differs from that given by Agbeyegbe (1984) in that the matrix $\bar{A}$ is non-singular in his paper.[8]

Subtracting $y(th-h)$ on both sides of equation (3.15) and adding the term $(F_2 - F_2)y(th-h)$ to the right-hand side of (3.15) yields

$$\Delta y(th) = -(I_p - F_1 - F_2)y(th-h) - F_2\Delta y(th-h) + \kappa_{th} + \vartheta_{th}. \tag{3.16}$$

We can alternatively write (3.16) as

$$\Delta y(th) = GQB'y(th-h) - F_2\Delta y(th-h) + \kappa_{th} + \vartheta_{th},\ t = 2, ..., n, \tag{3.17}$$

where $G = (I_{p_1}, 0_{p_1 \times p_2})'$ and $Q = ([e^{h\bar{A}}]_{13}[e^{h\bar{A}}]_{33}[e^{h\bar{A}}]_{13}^{-1} - I_{p_1})(I_{p_1} - [e^{h\bar{A}}]_{11}) + [e^{h\bar{A}}]_{13}[e^{h\bar{A}}]_{31}$. Note that since $B'y(th-h)$ is stationary, $\Delta y(th)$ is stationary too. Note, too, that the coefficient matrix on $y(th-h)$ is of reduced rank.

The model given in (3.17) holds for periods $2h, ..., nh$. To complete the exact discrete representation for the case of stock variables, we must derive a supplementary equation relating the first observation $y(h)$ to the initial state vector $x(0) = (y(0)', Dy_1(0)')'$. This equation, which is obtained from the system of difference equations implied by (3.14), is given by (see the Appendix for details)

$$y(h) - y(0) = GR_1B'y(0) + GR_2Dy_1(0) + \tilde{\kappa}_h + \tilde{\vartheta}_h, \tag{3.18}$$

where $R_1 = [e^{h\bar{A}}]_{11} - I_{p_1}$, $R_2 = [e^{h\bar{A}}]_{13}$, and the vectors $\tilde{\kappa}_h$ and $\tilde{\vartheta}_h$ are

$$\tilde{\kappa}_h = \tilde{S}_1 \int_0^h e^{(h-s)\bar{A}}(\mu^* + \gamma^* s)ds,$$
$$\tilde{\vartheta}_h = \tilde{S}_1 \int_0^h e^{(h-s)\bar{A}} n^*(s)ds.$$

We turn now to the derivation of the exact discrete model assuming that all variables are flows. Integrating (3.17) over $(th-h, th)$ and defining $\bar{y}(th) = h^{-1}\int_{th-h}^{th} y(r)dr$ yields

$$\Delta\bar{y}(th) = GQB'\bar{y}(th-h) - F_2\Delta\bar{y}(th-h) + \iota_{th} + \varsigma_{th},\ t = 3, ..., n, \tag{3.19}$$

[8]See Assumption 2 in Agbeyegbe (1984) which implies that $\bar{A}$ is non-singular. We have also change the order of the elements of $y(th)$, that being $y(th) = (y_2(th)', y_1(th)')'$ in the original paper. The elements of the matrix exponential have been replaced by those defined before.

where

$$\iota_{th} = \frac{1}{h}\left[\tilde{S}_1\int_{th-h}^{th}\int_{r-h}^{r} e^{(r-s)\bar{A}}(\mu^* + \gamma^* s)dsdr + \tilde{S}_2\int_{th-h}^{th}\int_{r-2h}^{r-h} e^{(r-h-s)\bar{A}}(\mu^* + \gamma^* s)dsdr\right],$$
$$\varsigma_{th} = \frac{1}{h}\left[\tilde{S}_1\int_{th-h}^{th}\int_{r-h}^{r} e^{(r-s)\bar{A}}n^*(s)dsdr + \tilde{S}_2\int_{th-h}^{th}\int_{r-2h}^{r-h} e^{(r-h-s)\bar{A}}n^*(s)dsdr\right].$$

The following two supplementary equations relating the first two observations (i.e. $\bar{y}(h)$ and $\bar{y}(2h)$) to the initial state vector $x(0)$ complete the exact discrete representation for the case of flow variables (see the Appendix for details)

$$\bar{y}(h) - y(0) = GU_1B'y(0) + GU_2Dy_1(0) + \tilde{\iota}_h + \tilde{\varsigma}_h, \tag{3.20}$$
$$\Delta\bar{y}(2h) = GR_1B'\bar{y}(h) + GR_2U_3B'y(0) + GR_2U_4Dy_1(0) + \tilde{\iota}_{2h} + \tilde{\varsigma}_{2h}, \tag{3.21}$$

where $U_1 = h^{-1}\int_0^h([e^{r\bar{A}}]_{11} - I_{p_1})dr$, $U_2 = h^{-1}\int_0^h[e^{r\bar{A}}]_{13}dr$, $U_3 = h^{-1}\int_0^h[e^{r\bar{A}}]_{31}dr$, $U_4 = h^{-1}\int_0^h[e^{r\bar{A}}]_{33}dr$, and the vectors $\tilde{\iota}_h$, $\tilde{\varsigma}_h$, $\tilde{\iota}_{2h}$ and $\tilde{\varsigma}_{2h}$ are

$$\tilde{\iota}_h = \frac{1}{h}\left[\tilde{S}_1\int_0^h\int_0^r e^{(r-s)\bar{A}}(\mu^* + \gamma^* s)dsdr\right],$$
$$\tilde{\varsigma}_h = \frac{1}{h}\left[\tilde{S}_1\int_0^h\int_0^r e^{(r-s)\bar{A}}n^*(s)dsdr\right],$$
$$\tilde{\iota}_{2h} = \frac{1}{h}\left[\tilde{S}_1\int_h^{2h}\int_{r-h}^{r} e^{(r-s)\bar{A}}(\mu^* + \gamma^* s)dsdr + \tilde{S}_3\int_0^h\int_0^r e^{(r-s)\bar{A}}(\mu^* + \gamma^* s)dsdr\right],$$
$$\tilde{\varsigma}_{2h} = \frac{1}{h}\left[\tilde{S}_1\int_h^{2h}\int_{r-h}^{r} e^{(r-s)\bar{A}}n^*(s)dsdr + \tilde{S}_3\int_0^h\int_0^r e^{(r-s)\bar{A}}n^*(s)dsdr\right],$$

with $\int_0^h[e^{r\bar{A}}]_{ij}dr$ being the entry $(i,j)$ of the matrix $\int_0^h e^{r\bar{A}}dr$, and $\tilde{S}_3$ given by

$$\tilde{S}_3 = \begin{pmatrix} 0 & 0 & [e^{h\bar{A}}]_{13} \\ 0 & 0 & 0 \end{pmatrix}.$$

We shall now use equations (3.17)-(3.21) to derive the exact discrete model satisfied by mixed data.

### 3.3.3 The exact discrete time model for mixed data sampling

The exact discrete representation corresponding to the continuous time system (3.11)-(3.12) with mixed sampling can be obtained by combining the observable stock and flow components of equations (3.17)-(3.21). It should be noted, however, that the equation determining $y_1(\tau)$ contains some unobservable variables, which must be eliminated and replaced with their observable counterpart. The exact discrete model is presented in

Theorems 3.3.1 and 3.3.2, in which the unobservable components are eliminated by using the results contained in Lemma 1 in Chambers (2009) (see the Appendix for more details).

**Theorem 3.3.1.** *Let $y(\tau)$ be generated by the system (3.11)- (3.12) and the elements of $y_1(\tau)$ be stock variables. Then, under Assumption 1, the observable vector $y_{th}$ satisfies*

$$y_h - y(0) = GR_1B'y(0) + GR_2Dy_1(0) + \alpha_h + \xi_h, \tag{3.22}$$

$$\Delta y_{2h} = GQB'y_h - F_2\Delta y_h + \alpha_{2h} + \xi_{2h}, \tag{3.23}$$

$$\Delta y_{th} = GQB'y_{th-h} - F_2\Delta y_{th-h} + \alpha_{th} + \xi_{th}, \; t = 3, ..., n, \tag{3.24}$$

*where*

$$\begin{aligned} \alpha_h &= N_1\tilde{\kappa}_h + N_2\tilde{\iota}_h, & \xi_h &= N_1\tilde{\vartheta}_h + N_2\tilde{\varsigma}_h, \\ \alpha_{2h} &= N_1\kappa_{2h} + N_2\tilde{\iota}_{2h} + N_3\tilde{\kappa}_h - N_3\tilde{\iota}_h, & \xi_{2h} &= N_1\vartheta_{2h} + N_2\tilde{\varsigma}_{2h} + N_3\tilde{\vartheta}_h - N_3\tilde{\varsigma}_h, \\ \alpha_{th} &= N_1\kappa_{th} + N_2\iota_{th} + N_3\phi_{th-h} + N_4\phi_{th-2h}, & \xi_{th} &= N_1\vartheta_{th} + N_2\varsigma_{th} + N_3\epsilon_{th-h} + N_4\epsilon_{th-2h}, \end{aligned}$$

*with $\phi_{th-h} = \left[(\frac{1}{2}\mu - \frac{1}{6}\gamma h) + \frac{1}{2}\gamma(th-h)\right]h$, $\epsilon_{th-h} = -h^{-1}\int_{th-h}^{th}(th-h-s)n(s)ds$, and the selection matrices $N_j$ $(j = 1, 2, 3, 4)$ given by*

$$N_1 = \begin{pmatrix} I_{p_1} & 0 & 0 \\ 0 & I_{p_2^s} & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad N_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & I_{p_2^f} \end{pmatrix},$$

$$N_3 = \begin{pmatrix} 0 & 0 & -Q\Pi_f - F_{2,12}^f \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad N_4 = \begin{pmatrix} 0 & 0 & F_{2,12}^f \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

**Theorem 3.3.2.** *Let $y(\tau)$ be generated by the system (3.11)- (3.12) and the elements of $y_1(\tau)$ be flow variables. Then, under Assumption 1, the observable vector $y_{th}$ satisfies*

$$y_h - y(0) = GU_1B'y(0) + GU_2Dy_1(0) + \alpha_h + \xi_h, \tag{3.25}$$

$$\Delta y_{2h} = GR_1B'y_h + GR_2U_3B'y(0) + R_2U_4Dy_1(0) + \alpha_{2h} + \xi_{2h}, \tag{3.26}$$

$$\Delta y_{th} = GQB'y_{th-h} - F_2\Delta y_{th-h} + \alpha_{th} + \xi_{th}, \; t = 3, ..., n, \tag{3.27}$$

*where*

$$\begin{aligned} \alpha_h &= N_1\tilde{\kappa}_h + N_2\tilde{\iota}_h, & \xi_h &= N_1\tilde{\vartheta}_h + N_2\tilde{\varsigma}_h, \\ \alpha_{2h} &= N_1\kappa_{2h} + N_2\tilde{\iota}_{2h} - \tilde{N}_3\tilde{\kappa}_h + \tilde{N}_3\tilde{\iota}_h, & \xi_{2h} &= N_1\vartheta_{2h} + N_2\tilde{\varsigma}_{2h} - \tilde{N}_3\tilde{\vartheta}_h + \tilde{N}_3\tilde{\varsigma}_h, \\ \alpha_{th} &= N_1\kappa_{th} + N_2\iota_{th} + N_3\phi_{th-h} + N_4\phi_{th-2h}, & \xi_{th} &= N_1\vartheta_{th} + N_2\varsigma_{th} + N_3\epsilon_{th-h} + N_4\epsilon_{th-2h}, \end{aligned}$$

*with* $\phi_{th-h} = \left[(\frac{1}{2}\mu - \frac{1}{6}\gamma h) + \frac{1}{2}\gamma(th-h)\right]h$, $\epsilon_{th-h} = -h^{-1}\int_{th-h}^{th}(th-h-s)n(s)ds$, *and the selection matrices* $N_j$ $(j = 1,2,3,4)$ *and* $\tilde{N}_3$ *given by*

$$N_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & I_{p_2^s} & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad N_2 = \begin{pmatrix} I_{p_1} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & I_{p_2^f} \end{pmatrix}, \quad \tilde{N}_3 = \begin{pmatrix} 0 & -R_1\Pi^s & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

$$N_3 = \begin{pmatrix} 0 & Q\Pi^s + F_{2,12}^s & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad N_4 = \begin{pmatrix} 0 & -F_{2,12}^s & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Note that the matrices $N_j$ $(j = 1,2,3,4)$ and $\tilde{N}_3$ have been partitioned in accordance with the elements of $y_{th}$ (i.e. $y_{1,th}$, $y_{2,th}^s$, $y_{2,th}^f$). We turn now to the characterization of the properties of $\xi_{th}$.

### 3.3.4 Properties of the discrete time disturbances

For estimation purposes it is necessary to derive the precise form of the autocovariances of the discrete time disturbances. To do this, it is convenient to first reduce the double integrals appearing in the vectors $\tilde{\varsigma}_h$, $\tilde{\varsigma}_{2h}$, $\varsigma_{th}$ $(t = 3, ..., n)$ to single integrals as follows[9]

$$\int_0^h \int_0^r e^{(r-s)\bar{A}} n^*(s)dsdr = \int_0^h \Theta(h-s)n^*(s)ds, \tag{3.28}$$

$$\int_{th-h}^{th} \int_{r-h}^{r} e^{(r-s)\bar{A}} n^*(s)dsdr = \int_{th-2h}^{th-h} [\Theta(h) - \Theta(th-h-s)]n^*(s)ds + \int_{th-h}^{th} \Theta(th-s)n^*(s)ds, \quad t = 2, ..., n, \tag{3.29}$$

$$\int_{th-h}^{th} \int_{r-2h}^{r-h} e^{(r-h-s)\bar{A}} n^*(s)dsdr = \int_{th-3h}^{th-2h} [\Theta(h) - \Theta(th-2h-s)]n^*(s)ds + \int_{th-2h}^{th-h} \Theta(th-h-s)n^*(s)ds, \quad t = 3, ..., n, \tag{3.30}$$

where $\Theta(z) = \sum_{j=0}^{\infty} z^{j+1}\bar{A}^j/(j+1)!$.

Now, let $\zeta_{th} = (\zeta_{1,th}', \zeta_{2,th}', \zeta_{3,th}', \zeta_{4,th}')'$ $(t = 1, ..., n)$ be an $4(p + p_1) \times 1$ vector, whose

[9] These expressions are obtained by using the results of Bergstrom (1997), McCrorie (2000) and Lemma 1 in Chambers (1999). See also Appendix A.2 in McCrorie and Chambers (2006).

elements are given by

$$
\begin{aligned}
\zeta_{1,th} &= \int_{th-h}^{th} \digamma(th-s)n^*(s)ds, \quad \digamma(s) = e^{s\bar{A}},\\
\zeta_{2,th} &= \frac{1}{h}\int_{th-h}^{th} \Theta(th-s)n^*(s)ds, \quad \Theta(s) = \int_{th-s}^{th} \digamma(r-th+s)dr,\\
\zeta_{3,th} &= \frac{1}{h}\int_{th-h}^{th} \Xi(th-s)n^*(s)ds, \quad \Xi(s) = \Theta(h)-\Theta(s),\\
\zeta_{4,th} &= -\frac{1}{h}\int_{th-h}^{th} (th-h-s)n^*(s)ds.
\end{aligned}
$$

It can be shown that the vectors determining the discrete time disturbances can be written in terms of $\zeta_{th}$ as following

$$
\begin{aligned}
\tilde{\vartheta}_h &= \tilde{S}_1\zeta_{1,h}, && (3.31)\\
\vartheta_{th} &= \tilde{S}_1\zeta_{1,th} + \tilde{S}_2\zeta_{1,th-h},\ t = 2, ..., n, && (3.32)\\
\tilde{\varsigma}_h &= \tilde{S}_1\zeta_{2,h}, && (3.33)\\
\tilde{\varsigma}_{2h} &= \tilde{S}_1\zeta_{2,2h} + \tilde{S}_1\zeta_{3,h} + \tilde{S}_3\zeta_{2,h}, && (3.34)\\
\tilde{\varsigma}_{th} &= \tilde{S}_1\zeta_{2,th} + \tilde{S}_1\zeta_{3,th-h} + \tilde{S}_2\zeta_{2,th-h} + \tilde{S}_2\zeta_{3,th-2h},\ t = 3, ..., n, && (3.35)\\
\epsilon_{th} &= \tilde{S}_4\zeta_{4,th},\ t = 1, ..., n, && (3.36)
\end{aligned}
$$

where the selection matrix $\tilde{S}_4$ is

$$
\tilde{S}_4 = \begin{pmatrix} 0 & 0 & I_{p_1} \\ 0 & I_{p_2} & 0 \end{pmatrix}.
$$

It follows from the above equations that the discrete time disturbance vectors $\xi_h, ..., \xi_{nh}$ have the moving average representation

$$
\begin{aligned}
\xi_h &= B_0\zeta_h, && (3.37)\\
\xi_{2h} &= B_0\zeta_{2h} + \tilde{B}_1\zeta_h, && (3.38)\\
\xi_{th} &= B_0\zeta_{th} + B_1\zeta_{th-h} + B_2\zeta_{th-2h},\ t = 3, ..., n, && (3.39)
\end{aligned}
$$

where

$$
B_0 = \begin{pmatrix} N_1\tilde{S}_1 & N_2\tilde{S}_1 & 0 & 0 \end{pmatrix},\ B_1 = \begin{pmatrix} N_1\tilde{S}_2 & N_2\tilde{S}_2 & N_2\tilde{S}_1 & N_3\tilde{S}_4 \end{pmatrix},\ B_2 = \begin{pmatrix} 0 & 0 & N_2\tilde{S}_2 & N_4\tilde{S}_4 \end{pmatrix}.
$$

The matrix $\tilde{B}_1$ takes different forms depending on the nature of the variable $y_1$. When it is composed of stock variables, $\tilde{B}_1$ is

$$\tilde{B}_1 = \begin{pmatrix} N_1\tilde{S}_2 + N_3\tilde{S}_1 & N_2\tilde{S}_3 - N_3\tilde{S}_1 & N_2\tilde{S}_1 & 0 \end{pmatrix},$$

and when $y_1$ is composed of flow variables, the matrix becomes

$$\tilde{B}_1 = \begin{pmatrix} N_1\tilde{S}_2 - \tilde{N}_3\tilde{S}_1 & N_2\tilde{S}_3 + \tilde{N}_3\tilde{S}_1 & N_2\tilde{S}_1 & 0 \end{pmatrix}.$$

From the moving average representation, the autocovariances of the discrete time disturbances $\xi_{th}$ follow immediately. Theorem 3.3.3 provides the results.

**Theorem 3.3.3.** *The discrete time disturbance vectors $\xi_h, ..., \xi_{nh}$ satisfy the following properties*

$$\begin{aligned}
E(\xi_h\xi_h') &= \Omega_{11} = B_0\Omega_\zeta B_0', \\
E(\xi_{2h}\xi_{2h}') &= \Omega_{22} = B_0\Omega_\zeta B_0' + \tilde{B}_1\Omega_\zeta\tilde{B}_1', \\
E(\xi_{th}\xi_{th}') &= \Omega_0 = B_0\Omega_\zeta B_0' + B_1\Omega_\zeta B_1' + B_2\Omega_\zeta B_2',\ t = 3, ..., n, \\
E(\xi_{2h}\xi_h') &= \Omega_{21} = \tilde{B}_1\Omega_\zeta B_0', \\
E(\xi_{3h}\xi_{2h}') &= \Omega_{32} = B_1\Omega_\zeta B_0' + B_2\Omega_\zeta\tilde{B}_1', \\
E(\xi_{th}\xi_{th-h}') &= \Omega_1 = B_1\Omega_\zeta B_0' + B_2\Omega_\zeta B_1',\ t = 4, ..., n, \\
E(\xi_{th}\xi_{th-2h}') &= \Omega_2 = B_2\Omega_\zeta B_0',\ t = 3, ..., n, \\
E(\xi_{th}\xi_{th-jh}') &= 0,\ j > 2,
\end{aligned}$$

*where $\Omega_\zeta = E(\zeta_{th}\zeta_{th}')$ is*

$$\frac{1}{h^2}\begin{pmatrix}
h^2\int_0^h F(s)\Sigma^*F(s)'ds & h\int_0^h F(s)\Sigma^*\Theta(s)'ds & h\int_0^h F(s)\Sigma^*\Xi(s)'ds & h\int_0^h F(s)\Sigma^*(h-s)ds \\
h\int_0^h \Theta(s)\Sigma^*F(s)'ds & \int_0^h \Theta(s)\Sigma^*\Theta(s)'ds & \int_0^h \Theta(s)\Sigma^*\Xi(s)'ds & \int_0^h \Theta(s)\Sigma^*(h-s)ds \\
h\int_0^h \Xi(s)\Sigma^*F(s)'ds & \int_0^h \Xi(s)\Sigma^*\Theta(s)'ds & \int_0^h \Xi(s)\Sigma^*\Xi(s)'ds & \int_0^h \Xi(s)\Sigma^*(h-s)ds \\
h\int_0^h (h-s)\Sigma^*F(s)'ds & \int_0^h (h-s)\Sigma^*\Theta(s)'ds & \int_0^h (h-s)\Sigma^*\Xi(s)'ds & (h^3/3)\Sigma^*
\end{pmatrix}$$

*and $\Sigma^* = E[n^*(\tau)n^*(\tau)']$ is*

$$\Sigma^* = \begin{pmatrix} 0 & 0 \\ 0 & \tilde{\Sigma} \end{pmatrix}, \quad \tilde{\Sigma} = \begin{pmatrix} E[u_2(\tau)u_2(\tau)'] & E[u_2(\tau)n_1(\tau)'] \\ E[n_1(\tau)u_2(\tau)'] & E[n_1(\tau)n_1(\tau)'] \end{pmatrix}.$$

### 3.3.5 Computation

As can be seen from Theorems 3.3.1 and 3.3.2, the deterministic terms $\alpha_{th}$ depend on $\kappa_{th}$ and $\iota_{th}$, which involve integrals of the exponential of the matrix $h\bar{A}$. The covariance matrix $\Omega_\zeta$ also involves integrals of $e^{h\bar{A}}$. Some formulae to compute these terms are given in Theorem 3.3.4, which requires the following assumption.

Assumption 2. The $2p_1 \times 2p_1$ matrix $V$ defined in Theorem 3.3.4 is non-singular.

**Theorem 3.3.4.** *Under Assumption 2, the vectors $\kappa_h$, $\kappa_{th}$, $\iota_h$, $\iota_{2h}$ and $\iota_{th}$ have the following representations*

$$\tilde{\kappa}_h = \tilde{S}_1\Theta(h)\mu^* - \tilde{S}_1\daleth(h)\gamma^* + \tilde{S}_1\Theta(h)\gamma^* h, \tag{3.40}$$

$$\kappa_{th} = (\tilde{S}_1 + \tilde{S}_2)\Theta(h)\mu^* - (\tilde{S}_1 + \tilde{S}_2)\daleth(h)\gamma^* + \tilde{S}_1\Theta(h)\gamma^* th + \tilde{S}_2\Theta(h)\gamma^*(th - h), \tag{3.41}$$

$$\tilde{\iota}_h = \frac{1}{h}\tilde{S}_1\Upsilon(h)\mu^* + \tilde{S}_1\left[\Upsilon(h) - \frac{1}{h}\aleph(h)\right]\gamma^*, \tag{3.42}$$

$$\tilde{\iota}_{2h} = \left[\tilde{S}_1\Theta(h) + \frac{1}{h}\tilde{S}_3\Upsilon(h)\right]\mu^* + \left[\frac{3}{2}\tilde{S}_1\Theta(h)h - \tilde{S}_1\daleth(h) + \tilde{S}_3\Upsilon(h) - \frac{1}{h}\tilde{S}_3\aleph(h)\right]\gamma^*, \tag{3.43}$$

$$\begin{aligned}\iota_{th} &= (\tilde{S}_1 + \tilde{S}_2)\Theta(h)\mu^* - (\tilde{S}_1 + \tilde{S}_2)\left[\daleth(h) + \frac{1}{2}\Theta(h)h\right]\gamma^* + \tilde{S}_1\Theta(h)\gamma^* th \\ &\quad + \tilde{S}_2\Theta(h)\gamma^*(th - h),\end{aligned} \tag{3.44}$$

*where*

$$\Theta(h) = J_0 + J_1 h + J_2 e^{hV} A_2', \qquad \Upsilon(h) = J_0 h + \frac{1}{2}J_1 h^2 + J_2 \bar{V} A_2',$$
$$\daleth(h) = \frac{1}{2}J_1 h^2 + J_2 V \bar{V}_1 A_2', \qquad \aleph(h) = \frac{1}{2}J_0 h^2 + \frac{1}{3}J_1 h^3 + J_2 \bar{V}_1 A_2',$$

*with the matrices $J_i$ $(i = 0, 1, 2)$, $V$, $\bar{V}$, $\bar{V}_1$, $A_1$ and $A_2'$ being*

$$J_0 = -A_1 V^{-2} A_2', \quad J_1 = I_{p+p_1} - A_1 V^{-1} A_2', \quad J_2 = A_1 V^{-2},$$
$$V = A_2' A_1, \quad \bar{V} = V^{-1}(e^{hV} - I_{2p_1}), \quad \bar{V}_1 = V^{-1}(he^{hV} - \bar{V}),$$

$$A_1 = \begin{pmatrix} I_{p_1} & 0 \\ 0 & 0 \\ 0 & I_{p_1} \end{pmatrix}, \quad A_2' = \begin{pmatrix} 0 & 0 & I_{p_1} \\ \tilde{C} & -\tilde{C}\Pi & \Psi \end{pmatrix}.$$

*The integrals determining $\Omega_\zeta$ are given by*

$$\int_0^h F(s)\Sigma^* F(s)' ds = J_1 \Sigma^* J_1' h + J_1 \Sigma^* A_2 \bar{V}' V' J_2' + J_2 V \bar{V} A_2' \Sigma^* J_1' + J_2 V \bar{V}_2 V' J_2', \tag{3.45}$$

$$\begin{aligned}\int_0^h \Theta(s)\Sigma^*\digamma(s)'ds &= J_0\Sigma^*J_1'h + J_0\Sigma^*A_2\bar{V}'V'J_2' + \frac{1}{2}J_1\Sigma^*J_1'h^2 + J_1\Sigma^*A_2\bar{V}_1'V'J_2' \\ &\quad + J_2\bar{V}A_2'\Sigma^*J_1' + J_2\bar{V}_2V'J_2', \end{aligned} \tag{3.46}$$

$$\begin{aligned}\int_0^h \Theta(s)\Sigma^*\Theta(s)'ds &= J_0\Sigma^*J_0'h + \frac{1}{2}J_0\Sigma^*J_1'h^2 + J_0\Sigma^*A_2\bar{V}'J_2' + \frac{1}{2}J_1\Sigma^*J_0'h^2 + \frac{1}{3}J_1\Sigma^*J_1'h^3 \\ &\quad + J_1\Sigma^*A_2\bar{V}_1'J_2' + J_2\bar{V}A_2'\Sigma^*J_0' + J_2\bar{V}_1A_2'\Sigma^*J_1' + J_2\bar{V}_2J_2', \end{aligned} \tag{3.47}$$

$$\int_0^h \Xi(s)\Sigma^*\digamma(s)'ds = \Theta(h)\Sigma^*\Theta(h)' - \int_0^h \Theta(s)\Sigma^*\digamma(s)'ds, \tag{3.48}$$

$$\int_0^h \Xi(s)\Sigma^*\Theta(s)'ds = \Theta(h)\Sigma^*\Upsilon(h)' - \int_0^h \Theta(s)\Sigma^*\Theta(s)'ds, \tag{3.49}$$

$$\int_0^h \Xi(s)\Sigma^*\Xi(s)'ds = \Theta(h)\Sigma^*\Theta(h)'h - \Upsilon(h)\Sigma^*\Theta(h)' - \int_0^h \Xi(s)\Sigma^*\Theta(s)'ds, \tag{3.50}$$

$$\int_0^h (h-s)\Sigma^*\digamma(s)'ds = -\frac{1}{2}\Sigma^*J_1'h^2 - \Sigma^*A_2\bar{V}_1V'J_2' + h\Sigma^*\Theta(h)', \tag{3.51}$$

$$\int_0^h (h-s)\Sigma^*\Theta(s)'ds = -\frac{1}{2}\Sigma^*J_0'h^2 - \frac{1}{3}\Sigma^*J_1'h^3 - \Sigma^*A_2\bar{V}_1'J_2' + h\Sigma^*\Upsilon(h)', \tag{3.52}$$

$$\int_0^h (h-s)\Sigma^*\Xi(s)'ds = -\frac{1}{2}\Sigma^*\Theta(h)'h^2 - \int_0^h (h-s)\Sigma^*\Theta(s)'ds + \Sigma^*\Theta(h)', \tag{3.53}$$

*where* $\bar{V}_2 = \int_0^h e^{rV}A_2'\Sigma^*A_2e^{rV'}dr.$

This last term can be calculated as $\bar{V}_2 = \bar{Z}_{22}'\bar{Z}_{12}$, where $\bar{Z}_{22}$ and $\bar{Z}_{12}$ are elements of the matrix exponential

$$\bar{Z} = \begin{pmatrix} \bar{Z}_{11} & \bar{Z}_{12} \\ 0 & \bar{Z}_{22} \end{pmatrix} = \exp\begin{pmatrix} -Vh & A_2'\Sigma^*A_2h \\ 0 & V'h \end{pmatrix}.$$

(see Jewitt and McCrorie, 2005).

Once the exact discrete analogue has been derived, standard estimation methods can be applied to obtain estimates of the continuous time parameters. Next section describes the Gaussian methods proposed by Bergstrom (1983, 1985, 1986), which have been commonly employed for this purpose.

## 3.4 Estimation

Assuming that the continuous time disturbance vector (i.e. $u(\tau)$ in the first-order system and $n(\tau)$ in the mixed-order system) is multivariate Gaussian, which implies a discrete time disturbance vector being Gaussian too, exact maximum likelihood estimates can be obtained by minimizing

$$\ln L = \sum_{t=1}^{n} \left( \varpi'_{th} \varpi_{th} + 2\ln|M_{tt}| \right), \tag{3.54}$$

where $\ln L$ denotes minus twice the logarithm of the likelihood function (ignoring the constant term), $\varpi_h, ..., \varpi_{nh}$ are functions of the observations and the parameters, which can be computed recursively as follows

$$\varpi_h = M_{11}^{-1} \xi_h, \tag{3.55}$$

$$\varpi_{2h} = M_{22}^{-1} (\xi_{2h} - M_{21} \varpi_h), \tag{3.56}$$

$$\varpi_{th} = M_{tt}^{-1} (\xi_{th} - M_{t,t-1} \varpi_{th-h} - M_{t,t-2} \varpi_{th-2h}), \ t = 3, ..., n, \tag{3.57}$$

and the coefficient matrices $M_{11}$, $M_{21}$, $M_{22}$, $M_{t,t-2}$, $M_{t,t-1}$ and $M_{tt}$ $(t = 3, ..., n)$ are obtained recursively from the equations

$$M_{11} M'_{11} = \Omega_{11}, \tag{3.58}$$

$$M_{21} = \Omega_{21} [M'_{11}]^{-1}, \tag{3.59}$$

$$M_{22} M'_{22} = \Omega_{22} - M_{21} M'_{21}, \tag{3.60}$$

$$M_{31} = \Omega_2 [M'_{11}]^{-1}, \tag{3.61}$$

$$M_{32} = [\Omega_{32} - M_{31} M'_{21}][M'_{22}]^{-1}, \tag{3.62}$$

$$M_{33} M'_{33} = \Omega_0 - M_{31} M'_{31} - M_{32} M'_{32}, \tag{3.63}$$

$$M_{t,t-2} = \Omega_2 [M'_{t-2,t-2}]^{-1}, \ t = 4, ..., n, \tag{3.64}$$

$$M_{t,t-1} = [\Omega_1 - M_{t,t-2} M'_{t-1,t-2}][M'_{t-1,t-1}]^{-1}, \ t = 4, ..., n, \tag{3.65}$$

$$M_{tt} M'_{tt} = \Omega_0 - M_{t,t-2} M'_{t,t-2} - M_{t,t-1} M'_{t,t-1}, \ t = 4, ..., n. \tag{3.66}$$

It should be noted that the coefficient matrices $M_{t,t-2}$, $M_{t,t-1}$ and $M_{tt}$ are time-dependent, although they converge to constant matrices as $t$ goes to infinity (See Bergstrom, 1990, Ch.7). Also, $M_{t,t-2} = 0$ $(t = 3, ..., n)$ in the first-order system since $\xi_{th}$ is a MA(1) process.

## 3.5 Monte Carlo simulations

This section explores the finite sample properties of the Gaussian estimator using Monte Carlo simulations. We consider two bivariate continuous time models, these being a first-order system and a mixed-order system. In both cases, it is assumed that the first variable of interest is a stock, while the second one is a flow, which is motivated by the theory of the Fisher effect.[10] Let $y(\tau) = (y^s(\tau), y^f(\tau))'$ and $b(\tau) = (b^s(\tau), b^f(\tau))'$, where $y^j(\tau)$ $(j = s, f)$ is a scalar continuous time process, $b^j(\tau)$ $(j = s, f)$ is a scalar Brownian motion process, and the superscripts $s$ and $f$ denote a stock or a flow, respectively. The covariance matrix of the increment of the Brownian motion $db(\tau) = [db^s(\tau), db^f(\tau)]'$, is $\Sigma d\tau$, where $\Sigma$ is a $2 \times 2$ symmetric positive definite matrix, taken to be

$$\Sigma = \begin{pmatrix} \sigma_{ss} & \sigma_{sf} \\ \sigma_{sf} & \sigma_{ff} \end{pmatrix} = \begin{pmatrix} 1 & \sigma_{sf} \\ \sigma_{sf} & \sigma_{ff} \end{pmatrix}.$$

To ensure that $\Sigma$ is positive definite during the estimation step, we use a Cholesky decomposition such that $\Sigma = PP'$, where $P$ is a lower triangular matrix, and estimate the elements of $P$ instead of the elements of $\Sigma$. Estimates of the original parameters can be recovered by using the following equations, which relate the elements of both matrices: $\sigma_{ss} = p_{11}^2$, $\sigma_{sf} = p_{11}p_{21}$, and $\sigma_{ff} = p_{21}^2 + p_{22}^2$.

In order to assess the effects of increasing the data frequency and the data span on the sampling properties of the Gaussian estimator, we consider five combinations of $T$ and $h$, these being $(T = 30, h = 1/4)$, $(T = 40, h = 1/4)$, $(T = 50, h = 1/4)$, $(T = 50, h = 1/12)$, and $(T = 50, h = 1/52)$, which correspond to 30, 40 and 50 years of quarterly data, 50 years of monthly data, and 50 years of weekly data, respectively. The experiment is replicated 10000 times.

### 3.5.1 First-order system

The first model considered is given by

$$dy^s(\tau) = [a^s + C(y^s(\tau) - \Pi y^f(\tau))]d\tau + db^s(\tau), \ \tau > 0, \tag{3.67}$$

$$dy^f(\tau) = a^f d\tau + db^f(\tau), \ \tau > 0. \tag{3.68}$$

Data are generated from the exact discrete representation, which is

$$y_h - y(0) = GJB'y(0) + \beta_h + \xi_h, \tag{3.69}$$

$$\Delta y_{th} = GJB'y_{th-h} + \beta + \xi_{th}, \ t = 2, ..., n, \tag{3.70}$$

[10] This theory introduced by Fisher (1930) postulates that the nominal interest rate is equal to the sum of the real interest rate and the expected rate of inflation. Fisher (1930) claimed a one-to-one relationship between the nominal interest rate (a stock variable) and the expected rate of inflation (a flow variable). The relationship has been often examined empirically by testing a long-run unitary coefficient on the expected inflation rate.

where $J = e^{hC} - 1$, and the vectors $\beta_h$ and $\beta$ are

$$\beta_h = S_1 H_0 a + S_2 \left( \frac{1}{2} I_2 + H_2 B' \right) ah,$$
$$\beta = H_0 a + \frac{1}{2} S_5 ah,$$

with $a = (a^s, a^f)'$, $H_0$ and $H_2$ defined in the Appendix, and the selection matrices $S_j$ $(j = 1, 2, 5)$ being

$$S_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad S_2 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \quad S_5 = \begin{pmatrix} 0 & -J\Pi \\ 0 & 0 \end{pmatrix}.$$

The disturbance vector $\xi_{th}$ is given by

$$\xi_h = S_1 \eta_h + S_2 \tilde{\upsilon}_h, \tag{3.71}$$
$$\xi_{2h} = S_1 \eta_{2h} + S_5 \eta_h + S_2 \upsilon_{2h} - S_5 \tilde{\upsilon}_h, \tag{3.72}$$
$$\xi_{th} = S_1 \eta_{th} + S_2 \upsilon_{th} + S_5 \epsilon_{th-h}, \ t = 3, ..., n, \tag{3.73}$$

and satisfies the following properties

$$\begin{aligned}
E(\xi_h \xi_h') &= \Omega_{11} = \int_0^h J_2(s) \Sigma J_2(s)' ds, \\
E(\xi_{th} \xi_{th}') &= \Omega_0 = \int_0^h J_0(s) \Sigma J_0(s)' ds + \int_0^h J_1(s) \Sigma J_1(s)' ds, \ t = 2, ..., n, \\
E(\xi_{th} \xi_{th-h}') &= \Omega_1 = \int_0^h J_1(s) \Sigma J_0(s)' ds, \ t = 2, ..., n, \\
E(\xi_{th} \xi_{th-jh}') &= 0, \ j > 1,
\end{aligned}$$

where $\int_0^h J_i(s) \Sigma J_k(s)' ds$ (see the Appendix for the definition of this term, and also for the definition of $\eta_{th}$, $\tilde{\upsilon}_h$, $\upsilon_{th}$, and $\epsilon_{th-h}$) is a function of the matrices $J_{ik}$ $(i = 0, 1; k = 0, 1, 2)$ given in the following table

| | $k$ | | |
|---|---|---|---|
| $i$ | 0 | 1 | 2 |
| 0 | $S_1(I_p - GB')$ | $\frac{1}{h} S_2$ | $S_1 G$ |
| 1 | $S_2 + S_5$ | $-\frac{1}{h}[S_2 + S_5]$ | 0 |

Table 3.1: Definition of the $J_{ik}$ matrices in a bivariate system.

To generate artificial data, we first compute the discrete time disturbance vector $\xi_{th}$

as follows

$$\xi_h = M_{11}\varrho_h,$$
$$\xi_{th} = M_{tt}\varrho_{th} + M_{t,t-1}\varrho_{th-h},\ t = 2, ..., n,$$

where $\varrho_{th} = (\varrho_{th}^s, \varrho_{th}^f)'$, $\varrho_{th}^j$ $(j = s, f)$ are random draws from a standard normal distribution, and the coefficient matrices $M_{11}$, $M_{t,t-1}$ and $M_{tt}$ $(t = 2, ..., n)$ are obtained recursively from the equations[11]

$$M_{11}M_{11}' = \Omega_{11},$$
$$M_{t,t-1} = \Omega_1[M_{t-1,t-1}']^{-1}, \quad t = 2, ..., n,$$
$$M_{tt}M_{tt}' = \Omega_0 - M_{t,t-1}M_{t,t-1}', \quad t = 2, ..., n.$$

Assuming that the initial state vector is $y(0) = 0$, we then compute simulated data for the vector $y$ by using equations (3.69)-(3.70). The following values for the parameters are employed: $a^s = a^f = 0$, $C = \{-0.3, -0.5, -1, -2\}$, $\Pi = 1$, $\sigma_{sf} = 0$ and $\sigma_{ff} = 1$. We only consider small values of the adjustment parameter $C$ because of the empirical evidence suggesting a slow reaction of the system variables to long-run disequilibrium (see for example, Rogoff, 1996; Johansen and Juselius, 1990). The initial value of the flow variable $y_1^f(0)$ is unobservable and must be estimated in addition to the continuous time parameters. For simplicity, we assume that $\xi_h = 0$ and compute the conditional maximum likelihood estimates, thus avoiding the computation of $\Omega_{11}$ and the estimation of $y_1^f(0)$.[12] Table 3.2 reports the bias and variance of the parameter estimates in the continuous time system with unknown mean.

We observe that the bias of $\hat{C}$ is negative and more pronounced as $C$ moves to zero. For example, for 30 years of quarterly data the percentage bias is 7% when $C = -2$, while it is 13%, 28%, and more than 50% when the adjustment coefficient is -1, -0.5 and -0.3, respectively. Increasing the sample size reduces the bias and variance of $\hat{C}$, these reductions being much larger when the span increases than when the sampling interval decreases. We also see from the table that the long-run equilibrium coefficient is downward biased, its bias and variance depend on the size of $C$, these being smaller as $C$ moves away from zero. Increasing the span reduces the bias and variance of $\hat{\Pi}$, but increasing the sampling frequency does not produce substantial changes in the figures. The variance of $\hat{\Pi}$ decreases faster than the variance of the other coefficients when the data span increases.

[11]Note that these are the same equations given in (3.58)-(3.66) with $M_{t,t-2} = 0$.

[12]Alternatively, we could obtain exact Gaussian estimates (see the iterative estimation procedure proposed by Bergstrom (1990, pg. 113-114). The conditional maximum likelihood estimator should yield a good approximation to the exact Gaussian estimator when the sample is large.

| | True value | Quarterly T=30 | Quarterly T=40 | Quarterly T=50 | Monthly T=50 | Weekly T=50 |
|---|---|---|---|---|---|---|
| | | | Bias | | | |
| $C$ | -0.3 | -0.1532 | -0.1065 | -0.0797 | -0.0782 | -0.0753 |
| $\Pi$ | 1 | -0.0624 | -0.0429 | -0.0317 | -0.0319 | -0.0310 |
| $\sigma_{ss}$ | 1 | 0.0163 | 0.0110 | 0.0083 | 0.0023 | 0.0006 |
| $\sigma_{sf}$ | 0 | -0.0189 | -0.0148 | -0.0119 | -0.0043 | -0.0014 |
| $\sigma_{ff}$ | 1 | -0.0017 | -0.0012 | -0.0004 | 0.0001 | 0.0006 |
| $a^s$ | 0 | 0.0040 | 0.0066 | 0.0038 | 0.0014 | 0.0006 |
| $a^f$ | 0 | -0.0035 | -0.0035 | -0.0023 | -0.0031 | -0.0025 |
| $C$ | -0.5 | -0.1393 | -0.0987 | -0.0745 | -0.0719 | -0.0686 |
| $\Pi$ | 1 | -0.0297 | -0.0192 | -0.0139 | -0.0142 | -0.0135 |
| $\sigma_{ss}$ | 1 | 0.0148 | 0.0101 | 0.0077 | 0.0020 | 0.0005 |
| $\sigma_{sf}$ | 0 | -0.0198 | -0.0154 | -0.0121 | -0.0044 | -0.0014 |
| $\sigma_{ff}$ | 1 | -0.0017 | -0.0012 | -0.0005 | 0.0001 | 0.0006 |
| $a^s$ | 0 | 0.0032 | 0.0065 | 0.0033 | 0.0019 | 0.0001 |
| $a^f$ | 0 | -0.0033 | -0.0034 | -0.0023 | -0.0031 | -0.0025 |
| $C$ | -1 | -0.1326 | -0.0958 | -0.0732 | -0.0677 | -0.0629 |
| $\Pi$ | 1 | -0.0093 | -0.0058 | -0.0042 | -0.0044 | -0.0040 |
| $\sigma_{ss}$ | 1 | 0.0143 | 0.0101 | 0.0079 | 0.0019 | 0.0005 |
| $\sigma_{sf}$ | 0 | -0.0206 | -0.0158 | -0.0124 | -0.0045 | -0.0014 |
| $\sigma_{ff}$ | 1 | -0.0016 | -0.0012 | -0.0005 | 0.0001 | 0.0006 |
| $a^s$ | 0 | 0.0032 | 0.0061 | 0.0031 | 0.0026 | -0.0003 |
| $a^f$ | 0 | -0.0030 | -0.0031 | -0.0020 | -0.0030 | -0.0025 |
| $C$ | -2 | -0.1415 | -0.1036 | -0.0800 | -0.0671 | -0.0603 |
| $\Pi$ | 1 | -0.0025 | -0.0017 | -0.0013 | -0.0014 | -0.0012 |
| $\sigma_{ss}$ | 1 | 0.0162 | 0.0118 | 0.0094 | 0.0020 | 0.0005 |
| $\sigma_{sf}$ | 0 | -0.0211 | -0.0161 | -0.0126 | -0.0045 | -0.0014 |
| $\sigma_{ff}$ | 1 | -0.0014 | -0.0012 | -0.0006 | 0.0000 | 0.0005 |
| $a^s$ | 0 | 0.0037 | 0.0065 | 0.0032 | 0.0032 | -0.0003 |
| $a^f$ | 0 | -0.0022 | -0.0025 | -0.0015 | -0.0029 | -0.0025 |
| | | | Variance | | | |
| $C$ | -0.3 | 0.0378 | 0.0216 | 0.0142 | 0.0132 | 0.0125 |
| $\Pi$ | 1 | 0.1575 | 0.0905 | 0.0545 | 0.0549 | 0.0547 |
| $\sigma_{ss}$ | 1 | 0.0195 | 0.0141 | 0.0113 | 0.0035 | 0.0008 |
| $\sigma_{sf}$ | 0 | 0.0144 | 0.0105 | 0.0085 | 0.0027 | 0.0006 |
| $\sigma_{ff}$ | 1 | 0.0174 | 0.0129 | 0.0104 | 0.0034 | 0.0008 |
| $a^s$ | 0 | 0.2090 | 0.1406 | 0.0969 | 0.0955 | 0.0949 |
| $a^f$ | 0 | 0.0334 | 0.0247 | 0.0197 | 0.0203 | 0.0200 |
| $C$ | -0.5 | 0.0438 | 0.0270 | 0.0189 | 0.0168 | 0.0157 |
| $\Pi$ | 1 | 0.0567 | 0.0315 | 0.0194 | 0.0193 | 0.0194 |
| $\sigma_{ss}$ | 1 | 0.0199 | 0.0144 | 0.0115 | 0.0035 | 0.0008 |
| $\sigma_{sf}$ | 0 | 0.0146 | 0.0107 | 0.0086 | 0.0028 | 0.0006 |
| $\sigma_{ff}$ | 1 | 0.0174 | 0.0128 | 0.0104 | 0.0034 | 0.0008 |
| $a^s$ | 0 | 0.1684 | 0.1164 | 0.0824 | 0.0809 | 0.0813 |
| $a^f$ | 0 | 0.0334 | 0.0246 | 0.0197 | 0.0203 | 0.0200 |
| $C$ | -1 | 0.0697 | 0.0464 | 0.0341 | 0.0279 | 0.0252 |
| $\Pi$ | 1 | 0.0138 | 0.0077 | 0.0047 | 0.0046 | 0.0047 |
| $\sigma_{ss}$ | 1 | 0.0211 | 0.0153 | 0.0122 | 0.0036 | 0.0008 |
| $\sigma_{sf}$ | 0 | 0.0154 | 0.0112 | 0.0091 | 0.0028 | 0.0006 |
| $\sigma_{ff}$ | 1 | 0.0174 | 0.0128 | 0.0104 | 0.0034 | 0.0008 |
| $a^s$ | 0 | 0.1387 | 0.0997 | 0.0717 | 0.0703 | 0.0712 |
| $a^f$ | 0 | 0.0334 | 0.0246 | 0.0197 | 0.0203 | 0.0200 |
| $C$ | -2 | 0.1509 | 0.1048 | 0.0799 | 0.0544 | 0.0457 |
| $\Pi$ | 1 | 0.0034 | 0.0019 | 0.0012 | 0.0011 | 0.0011 |
| $\sigma_{ss}$ | 1 | 0.0243 | 0.0176 | 0.0140 | 0.0038 | 0.0008 |
| $\sigma_{sf}$ | 0 | 0.0171 | 0.0125 | 0.0101 | 0.0029 | 0.0006 |
| $\sigma_{ff}$ | 1 | 0.0174 | 0.0128 | 0.0103 | 0.0033 | 0.0008 |
| $a^s$ | 0 | 0.1285 | 0.0918 | 0.0672 | 0.0655 | 0.0664 |
| $a^f$ | 0 | 0.0334 | 0.0245 | 0.0197 | 0.0203 | 0.0200 |

Table 3.2: Bias and variance of parameter estimates in system (3.67)-(3.68) with unknown mean.

The magnitude of $C$ has a smaller effect or no effect on the bias and variance of the other parameter estimates. Estimation of the covariance parameters enjoys small bias and variance, both statistics decreasing as the sample size becomes larger, although not

| | True value | Quarterly T=30 | Quarterly T=40 | Quarterly T=50 | Monthly T=50 | Weekly T=50 |
|---|---|---|---|---|---|---|
| | | | Bias | | | |
| $C$ | -0.3 | -0.0909 | -0.0634 | -0.0490 | -0.0484 | -0.0463 |
| $\Pi$ | 1 | -0.0152 | -0.0114 | -0.0092 | -0.0097 | -0.0107 |
| $\sigma_{ss}$ | 1 | 0.0132 | 0.0091 | 0.0070 | 0.0019 | 0.0005 |
| $\sigma_{sf}$ | 0 | -0.0124 | -0.0096 | -0.0078 | -0.0030 | -0.0010 |
| $\sigma_{ff}$ | 1 | 0.0067 | 0.0051 | 0.0045 | 0.0019 | 0.0009 |
| $C$ | -0.5 | -0.0857 | -0.0605 | -0.0469 | -0.0456 | -0.0430 |
| $\Pi$ | 1 | -0.0076 | -0.0048 | -0.0038 | -0.0042 | -0.0047 |
| $\sigma_{ss}$ | 1 | 0.0125 | 0.0087 | 0.0068 | 0.0018 | 0.0005 |
| $\sigma_{sf}$ | 0 | -0.0131 | -0.0100 | -0.0080 | -0.0031 | -0.0011 |
| $\sigma_{ff}$ | 1 | 0.0067 | 0.0050 | 0.0045 | 0.0018 | 0.0009 |
| $C$ | -1 | -0.0839 | -0.0604 | -0.0468 | -0.0439 | -0.0400 |
| $\Pi$ | 1 | -0.0020 | -0.0011 | -0.0010 | -0.0013 | -0.0014 |
| $\sigma_{ss}$ | 1 | 0.0126 | 0.0090 | 0.0071 | 0.0018 | 0.0004 |
| $\sigma_{sf}$ | 0 | -0.0137 | -0.0104 | -0.0082 | -0.0032 | -0.0011 |
| $\sigma_{ff}$ | 1 | 0.0068 | 0.0050 | 0.0044 | 0.0018 | 0.0009 |
| $C$ | -2 | -0.0904 | -0.0666 | -0.0517 | -0.0437 | -0.0386 |
| $\Pi$ | 1 | -0.0002 | -0.0002 | -0.0003 | -0.0004 | -0.0005 |
| $\sigma_{ss}$ | 1 | 0.0144 | 0.0106 | 0.0085 | 0.0019 | 0.0004 |
| $\sigma_{sf}$ | 0 | -0.0141 | -0.0108 | -0.0085 | -0.0032 | -0.0011 |
| $\sigma_{ff}$ | 1 | 0.0070 | 0.0049 | 0.0044 | 0.0017 | 0.0009 |
| | | | Variance | | | |
| $C$ | -0.3 | 0.0278 | 0.0165 | 0.0117 | 0.0110 | 0.0103 |
| $\Pi$ | 1 | 0.0935 | 0.0477 | 0.0282 | 0.0294 | 0.0293 |
| $\sigma_{ss}$ | 1 | 0.0190 | 0.0138 | 0.0111 | 0.0034 | 0.0008 |
| $\sigma_{sf}$ | 0 | 0.0143 | 0.0104 | 0.0084 | 0.0027 | 0.0006 |
| $\sigma_{ff}$ | 1 | 0.0175 | 0.0130 | 0.0105 | 0.0034 | 0.0008 |
| $C$ | -0.5 | 0.0356 | 0.0226 | 0.0166 | 0.0150 | 0.0139 |
| $\Pi$ | 1 | 0.0305 | 0.0161 | 0.0097 | 0.0100 | 0.0100 |
| $\sigma_{ss}$ | 1 | 0.0194 | 0.0141 | 0.0114 | 0.0035 | 0.0008 |
| $\sigma_{sf}$ | 0 | 0.0146 | 0.0106 | 0.0086 | 0.0028 | 0.0006 |
| $\sigma_{ff}$ | 1 | 0.0175 | 0.0130 | 0.0105 | 0.0034 | 0.0008 |
| $C$ | -1 | 0.0621 | 0.0421 | 0.0319 | 0.0265 | 0.0238 |
| $\Pi$ | 1 | 0.0070 | 0.0039 | 0.0024 | 0.0023 | 0.0024 |
| $\sigma_{ss}$ | 1 | 0.0207 | 0.0150 | 0.0121 | 0.0036 | 0.0008 |
| $\sigma_{sf}$ | 0 | 0.0153 | 0.0112 | 0.0091 | 0.0028 | 0.0006 |
| $\sigma_{ff}$ | 1 | 0.0176 | 0.0129 | 0.0105 | 0.0034 | 0.0008 |
| $C$ | -2 | 0.1420 | 0.0993 | 0.0771 | 0.0531 | 0.0444 |
| $\Pi$ | 1 | 0.0017 | 0.0010 | 0.0006 | 0.0006 | 0.0006 |
| $\sigma_{ss}$ | 1 | 0.0238 | 0.0173 | 0.0139 | 0.0038 | 0.0008 |
| $\sigma_{sf}$ | 0 | 0.0170 | 0.0125 | 0.0101 | 0.0029 | 0.0006 |
| $\sigma_{ff}$ | 1 | 0.0175 | 0.0129 | 0.0104 | 0.0034 | 0.0008 |

Table 3.3: Bias and variance of parameter estimates in system (3.67)-(3.68) with known mean.

uniformly for the the bias of $\sigma_{ff}$ which increases for weekly data. Results indicate that increasing the span reduces the variance of the intercepts, but does not produce a clear effect on the bias of these coefficients. The bias of $a^s$ also reduces as the sampling interval goes to zero.

Table 3.3 reports the bias and variance of the parameter estimates for the first-order system with known mean. A similar picture emerges from these results. Comparing Tables 3.2 and 3.3, we also observe that the bias of $\hat{C}$ and $\hat{\Pi}$ increases considerably when intercepts are estimated. So far we have only considered cases when the Brownian motions driving $y^s(\tau)$ and $y^f(\tau)$ have equal variances (i.e. the ratio $\sigma_{ff}/\sigma_{ss} = 1$) and are uncorrelated (i.e. $\sigma_{sf} = 0$). Tables 3.4 and 3.5 report the corresponding results for three values of $\sigma_{ff} = \{0.5, 1, 2\}$ and $\sigma_{sf} = \{-0.5, 0, 0.5\}$, respectively. For both cases we only consider the system with unknown mean and one value for the adjustment coefficient,

| | True value | Quarterly T=30 | Quarterly T=40 | Quarterly T=50 | Monthly T=50 | Weekly T=50 |
|---|---|---|---|---|---|---|
| | | | Bias | | | |
| $C$ | -1 | -0.1732 | -0.1260 | -0.0967 | -0.0885 | -0.0825 |
| $\Pi$ | 1 | -0.0116 | -0.0075 | -0.0054 | -0.0057 | -0.0052 |
| $\sigma_{ss}$ | 1 | 0.0230 | 0.0166 | 0.0129 | 0.0035 | 0.0008 |
| $\sigma_{sf}$ | 0 | -0.0134 | -0.0104 | -0.0082 | -0.0030 | -0.0009 |
| $\sigma_{ff}$ | 0.5 | -0.0009 | -0.0006 | -0.0002 | 0.0001 | 0.0003 |
| $a^s$ | 0 | 0.0026 | 0.0060 | 0.0031 | 0.0029 | -0.0003 |
| $a^f$ | 0 | -0.0023 | -0.0023 | -0.0015 | -0.0022 | -0.0018 |
| $C$ | -1 | -0.1326 | -0.0958 | -0.0732 | -0.0677 | -0.0629 |
| $\Pi$ | 1 | -0.0093 | -0.0058 | -0.0042 | -0.0044 | -0.0040 |
| $\sigma_{ss}$ | 1 | 0.0143 | 0.0101 | 0.0079 | 0.0019 | 0.0005 |
| $\sigma_{sf}$ | 0 | -0.0206 | -0.0158 | -0.0124 | -0.0045 | -0.0014 |
| $\sigma_{ff}$ | 1 | -0.0016 | -0.0012 | -0.0005 | 0.0001 | 0.0006 |
| $a^s$ | 0 | 0.0032 | 0.0061 | 0.0031 | 0.0026 | -0.0003 |
| $a^f$ | 0 | -0.0030 | -0.0031 | -0.0020 | -0.0030 | -0.0025 |
| $C$ | -1 | -0.0901 | -0.0648 | -0.0493 | -0.0460 | -0.0426 |
| $\Pi$ | 1 | -0.0065 | -0.0040 | -0.0029 | -0.0031 | -0.0028 |
| $\sigma_{ss}$ | 1 | 0.0051 | 0.0034 | 0.0027 | 0.0003 | 0.0001 |
| $\sigma_{sf}$ | 0 | -0.0279 | -0.0213 | -0.0166 | -0.0061 | -0.0019 |
| $\sigma_{ff}$ | 2 | -0.0030 | -0.0025 | -0.0011 | 0.0002 | 0.0011 |
| $a^s$ | 0 | 0.0035 | 0.0063 | 0.0031 | 0.0025 | -0.0002 |
| $a^f$ | 0 | -0.0037 | -0.0041 | -0.0026 | -0.0042 | -0.0035 |
| | | | Variance | | | |
| $C$ | -1 | 0.0941 | 0.0629 | 0.0461 | 0.0371 | 0.0336 |
| $\Pi$ | 1 | 0.0261 | 0.0147 | 0.0091 | 0.0090 | 0.0091 |
| $\sigma_{ss}$ | 1 | 0.0223 | 0.0161 | 0.0128 | 0.0036 | 0.0008 |
| $\sigma_{sf}$ | 0 | 0.0075 | 0.0055 | 0.0044 | 0.0014 | 0.0003 |
| $\sigma_{ff}$ | 0.5 | 0.0043 | 0.0032 | 0.0026 | 0.0008 | 0.0002 |
| $a^s$ | 0 | 0.1466 | 0.1052 | 0.0747 | 0.0725 | 0.0733 |
| $a^f$ | 0 | 0.0167 | 0.0123 | 0.0098 | 0.0102 | 0.0100 |
| $C$ | -1 | 0.0697 | 0.0464 | 0.0341 | 0.0279 | 0.0252 |
| $\Pi$ | 1 | 0.0138 | 0.0077 | 0.0047 | 0.0046 | 0.0047 |
| $\sigma_{ss}$ | 1 | 0.0211 | 0.0153 | 0.0122 | 0.0036 | 0.0008 |
| $\sigma_{sf}$ | 0 | 0.0154 | 0.0112 | 0.0091 | 0.0028 | 0.0006 |
| $\sigma_{ff}$ | 1 | 0.0174 | 0.0128 | 0.0104 | 0.0034 | 0.0008 |
| $a^s$ | 0 | 0.1387 | 0.0997 | 0.0717 | 0.0703 | 0.0712 |
| $a^f$ | 0 | 0.0334 | 0.0246 | 0.0197 | 0.0203 | 0.0200 |
| $C$ | -1 | 0.0459 | 0.0303 | 0.0225 | 0.0186 | 0.0169 |
| $\Pi$ | 1 | 0.0074 | 0.0041 | 0.0025 | 0.0024 | 0.0024 |
| $\sigma_{ss}$ | 1 | 0.0198 | 0.0144 | 0.0116 | 0.0035 | 0.0008 |
| $\sigma_{sf}$ | 0 | 0.0319 | 0.0232 | 0.0188 | 0.0057 | 0.0013 |
| $\sigma_{ff}$ | 2 | 0.0696 | 0.0513 | 0.0415 | 0.0134 | 0.0031 |
| $a^s$ | 0 | 0.1312 | 0.0940 | 0.0686 | 0.0680 | 0.0691 |
| $a^f$ | 0 | 0.0668 | 0.0491 | 0.0394 | 0.0406 | 0.0400 |

Table 3.4: Bias and variance of parameter estimates in system (3.67)-(3.68) with unknown mean for different values of $\sigma_{ff}$.

this being $C = -1$.

From Table 3.4 we observe that the bias and variance of $\hat{C}$, $\hat{\Pi}$ and $\hat{\sigma}_{ss}$ and the variance of $\hat{a}^s$ decrease as the ratio $\sigma_{ff}/\sigma_{ss}$ increases,[13] as opposed to the bias and variance of $\hat{\sigma}_{sf}$, $\hat{\sigma}_{ff}$ and $\hat{a}^f$, which increase when the ratio $\sigma_{ff}/\sigma_{ss}$ becomes larger. There is a tendency for the bias of $\hat{a}^s$ to increase for quarterly data, and to decrease for higher frequency data. Table 3.5 suggests that the covariance parameter affects the bias and variance of some parameter estimates, although not in the same way. For example, the bias and variance of the adjustment coefficient are larger when $\sigma_{sf}$ goes from -0.5 to 0.5, while those of the long-run coefficient are higher when $\sigma_{sf} = 0$. The effect of increasing the sample size is

[13]The result for the bias of the long-run coefficient is consistent with the findings in Gonzalo (1994) for discrete time cointegrated models.

| | True value | Quarterly T=30 | Quarterly T=40 | Quarterly T=50 | Monthly T=50 | Weekly T=50 |
|---|---|---|---|---|---|---|
| | | | Bias | | | |
| $C$ | -1 | -0.0703 | -0.0509 | -0.0381 | -0.0350 | -0.0329 |
| $\Pi$ | 1 | -0.0065 | -0.0045 | -0.0030 | -0.0032 | -0.0033 |
| $\sigma_{ss}$ | 1 | 0.0125 | 0.0089 | 0.0068 | 0.0017 | 0.0004 |
| $\sigma_{sf}$ | -0.5 | -0.0192 | -0.0143 | -0.0113 | -0.0039 | -0.0013 |
| $\sigma_{ff}$ | 1 | -0.0006 | -0.0003 | 0.0002 | 0.0005 | 0.0008 |
| $a^s$ | 0 | 0.0013 | 0.0049 | 0.0028 | 0.0026 | 0.0014 |
| $a^f$ | 0 | -0.0045 | -0.0043 | -0.0030 | -0.0034 | -0.0023 |
| $C$ | -1 | -0.1326 | -0.0958 | -0.0732 | -0.0677 | -0.0629 |
| $\Pi$ | 1 | -0.0093 | -0.0058 | -0.0042 | -0.0044 | -0.0040 |
| $\sigma_{ss}$ | 1 | 0.0143 | 0.0101 | 0.0079 | 0.0019 | 0.0005 |
| $\sigma_{sf}$ | 0 | -0.0206 | -0.0158 | -0.0124 | -0.0045 | -0.0014 |
| $\sigma_{ff}$ | 1 | -0.0016 | -0.0012 | -0.0005 | 0.0001 | 0.0006 |
| $a^s$ | 0 | 0.0032 | 0.0061 | 0.0031 | 0.0026 | -0.0003 |
| $a^f$ | 0 | -0.0030 | -0.0031 | -0.0020 | -0.0030 | -0.0025 |
| $C$ | -1 | -0.2049 | -0.1486 | -0.1151 | -0.1002 | -0.0923 |
| $\Pi$ | 1 | -0.0071 | -0.0046 | -0.0033 | -0.0031 | -0.0023 |
| $\sigma_{ss}$ | 1 | 0.0133 | 0.0093 | 0.0074 | 0.0015 | 0.0004 |
| $\sigma_{sf}$ | 0.5 | -0.0187 | -0.0145 | -0.0111 | -0.0040 | -0.0011 |
| $\sigma_{ff}$ | 1 | -0.0019 | -0.0015 | -0.0007 | -0.0003 | 0.0002 |
| $a^s$ | 0 | 0.0066 | 0.0066 | 0.0026 | 0.0018 | 0.0020 |
| $a^f$ | 0 | -0.0008 | -0.0013 | -0.0005 | -0.0019 | -0.0020 |
| | | | Variance | | | |
| $C$ | -1 | 0.0333 | 0.0227 | 0.0165 | 0.0137 | 0.0127 |
| $\Pi$ | 1 | 0.0119 | 0.0064 | 0.0038 | 0.0037 | 0.0037 |
| $\sigma_{ss}$ | 1 | 0.0209 | 0.0152 | 0.0123 | 0.0036 | 0.0008 |
| $\sigma_{sf}$ | -0.5 | 0.0149 | 0.0109 | 0.0089 | 0.0028 | 0.0006 |
| $\sigma_{ff}$ | 1 | 0.0170 | 0.0127 | 0.0104 | 0.0034 | 0.0008 |
| $a^s$ | 0 | 0.1036 | 0.0770 | 0.0567 | 0.0543 | 0.0552 |
| $a^f$ | 0 | 0.0333 | 0.0249 | 0.0197 | 0.0202 | 0.0198 |
| $C$ | -1 | 0.0697 | 0.0464 | 0.0341 | 0.0279 | 0.0252 |
| $\Pi$ | 1 | 0.0138 | 0.0077 | 0.0047 | 0.0046 | 0.0047 |
| $\sigma_{ss}$ | 1 | 0.0211 | 0.0153 | 0.0122 | 0.0036 | 0.0008 |
| $\sigma_{sf}$ | 0 | 0.0154 | 0.0112 | 0.0091 | 0.0028 | 0.0006 |
| $\sigma_{ff}$ | 1 | 0.0174 | 0.0128 | 0.0104 | 0.0034 | 0.0008 |
| $a^s$ | 0 | 0.1387 | 0.0997 | 0.0717 | 0.0703 | 0.0712 |
| $a^f$ | 0 | 0.0334 | 0.0246 | 0.0197 | 0.0203 | 0.0200 |
| $C$ | -1 | 0.1225 | 0.0785 | 0.0584 | 0.0443 | 0.0388 |
| $\Pi$ | 1 | 0.0102 | 0.0055 | 0.0034 | 0.0032 | 0.0033 |
| $\sigma_{ss}$ | 1 | 0.0197 | 0.0142 | 0.0113 | 0.0035 | 0.0008 |
| $\sigma_{sf}$ | 0.5 | 0.0158 | 0.0115 | 0.0093 | 0.0028 | 0.0006 |
| $\sigma_{ff}$ | 1 | 0.0172 | 0.0128 | 0.0102 | 0.0033 | 0.0008 |
| $a^s$ | 0 | 0.1307 | 0.0850 | 0.0629 | 0.0593 | 0.0606 |
| $a^f$ | 0 | 0.0334 | 0.0245 | 0.0197 | 0.0200 | 0.0204 |

Table 3.5: Bias and variance of parameter estimates in system (3.67)-(3.68) with unknown mean for different values of $\sigma_{sf}$.

similar to that found for the case of uncorrelated errors with equal variances.

According to the Monte Carlo study, it seems that the bias and the variance of $\hat{C}$ and $\hat{\Pi}$ as well as the variance of the intercepts are mainly determined by the data span, which is consistent with the results in Tang and Chen (2009) for stationary continuous time models. In contrast, the bias and variance of $\hat{\sigma}_{ij}$ appear to depend on the sample size, implying that these parameters will be estimated very accurately in both cases when the span gets larger or when the sampling interval gets smaller. This last result is also consistent with the Tang and Chen's findings for stationary continuous time models.

### 3.5.2 Mixed-order system

The second model considered is

$$d[Dy^s(\tau)] = \left[\mu^s + \Psi Dy^s(\tau) + \tilde{C}(y^s(\tau) - \Pi y^f(\tau))\right] d\tau + db^s(\tau),\ \tau > 0, \tag{3.74}$$

$$dy^f(\tau) = a^f d\tau + db^f(\tau),\ \tau > 0. \tag{3.75}$$

Similar to the first-order system, we simulate data from the exact discrete representation, which is given by

$$y_h - y(0) = GR_1B'y(0) + GR_2Dy^s(0) + \alpha_h + \xi_h, \tag{3.76}$$

$$\Delta y_{2h} = GQB'y_h - F_2\Delta y_h + \alpha_{2h} + \xi_{2h}, \tag{3.77}$$

$$\Delta y_{th} = GQB'y_{th-h} - F_2\Delta y_{th-h} + \alpha + \xi_{th},\ t = 3, ..., n, \tag{3.78}$$

where $R_1 = [e^{h\bar{A}}]_{11} - 1,\ R_2 = [e^{h\bar{A}}]_{13},\ Q = ([e^{h\bar{A}}]_{33} - 1)(1 - [e^{h\bar{A}}]_{11}) + [e^{h\bar{A}}]_{13}[e^{h\bar{A}}]_{31}$, the matrix $F_2$ is

$$\begin{aligned}
F_2 &= G(F_{2,11} \quad F_{2,12}),\\
F_{2,11} &= [e^{h\bar{A}}]_{13}[e^{h\bar{A}}]_{31} - [e^{h\bar{A}}]_{33}[e^{h\bar{A}}]_{11},\\
F_{2,12} &= -\left([e^{h\bar{A}}]_{13}[e^{h\bar{A}}]_{31} + [e^{h\bar{A}}]_{33}(1 - [e^{h\bar{A}}]_{11})\right)\Pi,
\end{aligned}$$

and the vectors $\alpha$'s are

$$\begin{aligned}
\alpha_h &= N_1\tilde{S}_1\Theta(h)\mu^* + \frac{1}{h}N_2\tilde{S}_1\Upsilon(h)\mu^*,\\
\alpha_{2h} &= (\tilde{S}_1 + \tilde{S}_2)\Theta(h)\mu^* + N_3\tilde{S}_1\left[\Theta(h) - \frac{1}{h}\Upsilon(h)\right]\mu^*,\\
\alpha &= (\tilde{S}_1 + \tilde{S}_2)\Theta(h)\mu^* + \frac{1}{2}(N_3 + N_4)\mu h,
\end{aligned}$$

with the matrices $S_j$ $(j = 1, 2)$ and $N_j$ $(j = 1, 2, 3, 4)$ being

$$\tilde{S}_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix},\quad \tilde{S}_2 = \begin{pmatrix} -[e^{h\bar{A}}]_{33} & 0 & [e^{h\bar{A}}]_{13} \\ 0 & 0 & 0 \end{pmatrix},$$

$$N_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},\quad N_2 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix},\quad N_3 = \begin{pmatrix} 0 & -Q\Pi - F_{2,12} \\ 0 & 0 \end{pmatrix},\quad N_4 = \begin{pmatrix} 0 & F_{2,12} \\ 0 & 0 \end{pmatrix}.$$

The disturbance vector $\xi_{th}$ is a vector MA(2) process with autocovariances given in Theorem 3.3.3.

We simulate the discrete time disturbances as follows

$$\xi_h = M_{11}\varrho_h,$$
$$\xi_{2h} = M_{22}\varrho_{2h} + M_{21}\varrho_h,$$
$$\xi_{th} = M_{tt}\varrho_{th} + M_{t,t-1}\varrho_{th-h} + M_{t,t-2}\varrho_{th-2h}, \quad t = 3, ..., n,$$

where $\varrho_{th} = (\varrho_{th}^s, \varrho_{th}^f)'$, $\varrho_{th}^j$ $(j = s, f)$ are random draws from a standard normal distribution, and the coefficient matrices $M_{11}$, $M_{21}$, $M_{22}$, $M_{t,t-2}$, $M_{t,t-1}$ and $M_{tt}$ $(t = 3, ..., n)$ are obtained recursively from equations (3.58)-(3.66). We then compute simulated data for $y$ by using (3.76)-(3.78), assuming that the initial state vector is $x(0) = (y(0), Dy^s(0)) = 0$. Three combinations of $\Psi$ and $\tilde{C}$ are considered, these being $(\Psi = -1.5, \tilde{C} = -0.5)$, $(\Psi = -2, \tilde{C} = -1)$ and $(\Psi = -3, \tilde{C} = -2)$. In all three cases we use $\mu^s = a^f = 0$, $\sigma_{ss} = \sigma_{ff} = 1$, and $\sigma_{sf} = 0$. For estimation it is assumed that $y^s(0)$ is observable and equal to zero. We also assume that $\xi_h = 0$, thus avoiding the computation of $\Omega_{11}$, $\Omega_{21}$, $R_1$ and $R_2$, and the estimation of $y_1^f(0)$ and $Dy^s(0)$. Table 3.6 reports the conditional maximum likelihood estimates.

Results are similar to those obtained for the first-order system. First, the biases of the short-run, long-run and adjustment coefficients are negative and more pronounced as $\Psi$ and $\tilde{C}$ get smaller (in absolute value). Second, estimation of the long-run equilibrium coefficient and covariance parameters enjoys small bias and variance. Third, it appears that the bias and variance of the estimators of $\Psi$, $\tilde{C}$ and $\Pi$ as well as the variance of the intercepts are mainly determined by the data span. Fourth, the bias and variance of the covariance coefficients seem to depend on the sample size.

## 3.6 Conclusions

This paper has derived the exact discrete representation corresponding to a mixed-order system of stochastic differential equations with a mixture of stock and flow variables. We have also studied the finite sample properties of the Gaussian estimator using a Monte Carlo experiment. Simulations based on a first-order system reveals that the magnitude of the bias of $\hat{C}$ and $\hat{\Pi}$ depends on the size of the adjustment coefficient, the covariance and the ratio of the variances of the two Brownian motions.

Results for both a first-order system and a mixed-order system are consistent with what has been found for stationary continuous time models. The bias and variance of the estimators of the short-run, long-run and adjustment coefficients as well as the variance of the intercepts are mainly determined by the data span, while the bias and variance of the covariance coefficients depend on the sample size. Estimation of the long-run coefficient and covariance parameters enjoys smaller bias and variance than the other parameters. It appears that despite $\hat{\Pi}$ being estimated very precisely, there may be some bias in finite samples, this being especially the case when the model includes an intercept and the reaction of the system variables to deviations from equilibrium is slow (i.e. $C$ is close to zero).

| | | Quarterly | | | Monthly | Weekly |
|---|---|---|---|---|---|---|
| | True value | T=30 | T=40 | T=50 | T=50 | T=50 |
| | | | Bias | | | |
| $\Psi$ | -1.5 | -0.1789 | -0.1308 | -0.1026 | -0.0981 | -0.0939 |
| $\tilde{C}$ | -0.5 | -0.1565 | -0.1100 | -0.0830 | -0.0813 | -0.0779 |
| $\Pi$ | 1 | -0.0167 | -0.0131 | -0.0109 | -0.0114 | -0.0100 |
| $\sigma_{ss}$ | 1 | 0.0084 | 0.0061 | 0.0050 | 0.0008 | 0.0002 |
| $\sigma_{sf}$ | 0 | -0.0143 | -0.0114 | -0.0091 | -0.0033 | -0.0009 |
| $\sigma_{ff}$ | 1 | -0.0102 | -0.0075 | -0.0055 | -0.0015 | 0.0002 |
| $\mu^s$ | 0 | 0.0048 | 0.0060 | 0.0030 | 0.0020 | 0.0004 |
| $a^f$ | 0 | -0.0034 | -0.0036 | -0.0024 | -0.0031 | -0.0025 |
| $\Psi$ | -2 | -0.1795 | -0.1305 | -0.1031 | -0.0949 | -0.0903 |
| $\tilde{C}$ | -1 | -0.1516 | -0.1092 | -0.0841 | -0.0794 | -0.0745 |
| $\Pi$ | 1 | -0.0056 | -0.0039 | -0.0034 | -0.0034 | -0.0027 |
| $\sigma_{ss}$ | 1 | 0.0098 | 0.0068 | 0.0056 | 0.0008 | 0.0002 |
| $\sigma_{sf}$ | 0 | -0.0159 | -0.0124 | -0.0099 | -0.0035 | -0.0010 |
| $\sigma_{ff}$ | 1 | -0.0101 | -0.0075 | -0.0055 | -0.0015 | 0.0002 |
| $\mu^s$ | 0 | 0.0042 | 0.0064 | 0.0029 | 0.0023 | -0.0003 |
| $a^f$ | 0 | -0.0032 | -0.0034 | -0.0022 | -0.0031 | -0.0027 |
| $\Psi$ | -3 | -0.1893 | -0.1370 | -0.1085 | -0.0923 | -0.0878 |
| $\tilde{C}$ | -2 | -0.1648 | -0.1202 | -0.0942 | -0.0814 | -0.0747 |
| $\Pi$ | 1 | -0.0015 | -0.0011 | -0.0011 | -0.0011 | -0.0007 |
| $\sigma_{ss}$ | 1 | 0.0121 | 0.0083 | 0.0069 | 0.0006 | 0.0002 |
| $\sigma_{sf}$ | 0 | -0.0182 | -0.0140 | -0.0112 | -0.0037 | -0.0010 |
| $\sigma_{ff}$ | 1 | -0.0100 | -0.0075 | -0.0056 | -0.0015 | 0.0001 |
| $\mu^s$ | 0 | 0.0038 | 0.0072 | 0.0033 | 0.0026 | -0.0003 |
| $a^f$ | 0 | -0.0028 | -0.0030 | -0.0019 | -0.0030 | -0.0024 |
| | | | Variance | | | |
| $\Psi$ | -1.5 | 0.1780 | 0.1220 | 0.0909 | 0.0690 | 0.0622 |
| $\tilde{C}$ | -0.5 | 0.0507 | 0.0294 | 0.0199 | 0.0185 | 0.0175 |
| $\Pi$ | 1 | 0.0850 | 0.0423 | 0.0239 | 0.0229 | 0.0240 |
| $\sigma_{ss}$ | 1 | 0.0260 | 0.0190 | 0.0149 | 0.0038 | 0.0008 |
| $\sigma_{sf}$ | 0 | 0.0091 | 0.0068 | 0.0054 | 0.0017 | 0.0004 |
| $\sigma_{ff}$ | 1 | 0.0171 | 0.0127 | 0.0103 | 0.0034 | 0.0008 |
| $\mu^s$ | 0 | 0.1830 | 0.1263 | 0.0878 | 0.0871 | 0.0881 |
| $a^f$ | 0 | 0.0334 | 0.0247 | 0.0197 | 0.0203 | 0.0200 |
| $\Psi$ | -2 | 0.2150 | 0.1485 | 0.1132 | 0.0811 | 0.0709 |
| $\tilde{C}$ | -1 | 0.0730 | 0.0474 | 0.0345 | 0.0291 | 0.0268 |
| $\Pi$ | 1 | 0.0185 | 0.0097 | 0.0057 | 0.0055 | 0.0057 |
| $\sigma_{ss}$ | 1 | 0.0279 | 0.0203 | 0.0160 | 0.0039 | 0.0008 |
| $\sigma_{sf}$ | 0 | 0.0096 | 0.0071 | 0.0056 | 0.0017 | 0.0004 |
| $\sigma_{ff}$ | 1 | 0.0171 | 0.0127 | 0.0103 | 0.0034 | 0.0008 |
| $\mu^s$ | 0 | 0.1482 | 0.1051 | 0.0758 | 0.0748 | 0.0763 |
| $a^f$ | 0 | 0.0334 | 0.0246 | 0.0197 | 0.0203 | 0.0200 |
| $\Psi$ | -3 | 0.3214 | 0.2239 | 0.1744 | 0.1098 | 0.0904 |
| $\tilde{C}$ | -2 | 0.1485 | 0.1023 | 0.0784 | 0.0546 | 0.0467 |
| $\Pi$ | 1 | 0.0045 | 0.0023 | 0.0014 | 0.0013 | 0.0014 |
| $\sigma_{ss}$ | 1 | 0.0326 | 0.0235 | 0.0186 | 0.0041 | 0.0008 |
| $\sigma_{sf}$ | 0 | 0.0108 | 0.0079 | 0.0063 | 0.0018 | 0.0004 |
| $\sigma_{ff}$ | 1 | 0.0171 | 0.0127 | 0.0103 | 0.0033 | 0.0008 |
| $\mu^s$ | 0 | 0.1329 | 0.0944 | 0.0694 | 0.0688 | 0.0704 |
| $a^f$ | 0 | 0.0334 | 0.0246 | 0.0197 | 0.0203 | 0.0200 |

Table 3.6: Bias and variance of the parameter estimates in system (3.74)-(3.75).

## 3.7 Appendix

The matrices Φ's and the vectors $\beta_{th}$ $(t = 1, ..., n)$ are given by[14]

$$\Phi_{10} = \begin{pmatrix} J_s \\ K_f \\ 0 \\ 0 \end{pmatrix}, \quad \Phi_{21} = \begin{pmatrix} J_s \\ J_f \\ 0 \\ 0 \end{pmatrix}, \quad \Phi_{20} = \begin{pmatrix} J_{sf}(J_f - K_f) \\ J_{fs}(K_s - J_s) \\ 0 \\ 0 \end{pmatrix},$$

$$\Phi_0 = \begin{pmatrix} P_0 \\ 0 \end{pmatrix}, \quad P_0 = \begin{pmatrix} P_0^{ss} & P_0^{sf} \\ 0 & P_0^{ff} \end{pmatrix}, \quad \Phi_1 = \begin{pmatrix} P_1 & 0 \\ 0 & 0 \end{pmatrix}, \quad P_1 = \begin{pmatrix} P_1^{ss} & 0 \\ P_1^{fs} & 0 \end{pmatrix},$$

$$\beta_h = \begin{pmatrix} c_{1,h}^s \\ \tilde{\psi}_{1,h}^f \\ c_{2,h}^s \\ \tilde{\psi}_{2,h}^f \end{pmatrix}, \quad \beta_{2h} = \begin{pmatrix} \beta_{1,2h}^s \\ \beta_{1,2h}^f \\ c_{2,2h}^s \\ \psi_{2,2h}^f \end{pmatrix}, \quad \beta_{th} = \begin{pmatrix} \beta_{1,th}^s \\ \beta_{1,th}^f \\ c_{2,th}^s \\ \psi_{2,th}^f \end{pmatrix},$$

where

$$J = \begin{pmatrix} J_s \\ J_f \end{pmatrix} = \begin{pmatrix} J_{ss} & J_{sf} \\ J_{fs} & J_{ff} \end{pmatrix} = e^{hC} - I_{p_1},$$

$$K = \begin{pmatrix} K_s \\ K_f \end{pmatrix} = \begin{pmatrix} K_{ss} & K_{sf} \\ K_{fs} & K_{ff} \end{pmatrix} = \frac{1}{h} C^{-1} J - I_{p_1},$$

and the remaining terms are defined in Table 3.7.

The disturbance vectors $\xi_{th}$ are

$$\begin{aligned} \xi_h &= S_6 \eta_h + S_7 \tilde{\upsilon}_h, \\ \xi_{2h} &= S_6 \eta_{2h} + S_8 \eta_h + S_7 \upsilon_{2h} - S_8 \tilde{\upsilon}_h, \\ \xi_{th} &= S_1 \eta_{th} + S_2 \upsilon_{th} + S_3 e_{th} + S_4 e_{th-h} + S_5 \epsilon_{th-h}, \quad t = 3, ..., n, \end{aligned}$$

where $\eta_{th} = \int_{th-h}^{th} e^{(th-s)AB'} u(s) ds$, $e_{th} = \int_{th-h}^{th} u(r) dr$, $\epsilon_{th} = \frac{1}{h} \int_{th-h}^{th} [h - (th - s)] u(s) ds$, $\upsilon_{th} = \frac{1}{h} \int_{th-h}^{th} \int_{r-h}^{r} e^{(r-s)AB'} u(s) ds dr$, $\tilde{\upsilon}_h = \frac{1}{h} \int_0^h \int_0^r e^{(r-s)AB'} u(s) ds dr$, and the selection

[14]The matrices and vectors are partitioned conformably with $y_1$, $y_2$, as well as with the stock and flow components. Their elements are ordered according to the vector of observations $y_{th}$, which takes the form

$$y_{th} = \begin{pmatrix} y_{1,th}^s \\ y_{1,th}^f \\ y_{2,th}^s \\ y_{2,th}^f \end{pmatrix} = \begin{pmatrix} y_1^s(th) \\ \frac{1}{h} \int_{th-h}^{th} y_1^f(r) dr \\ y_2^s(th) \\ \frac{1}{h} \int_{th-h}^{th} y_2^f(r) dr \end{pmatrix}.$$

| Submatrices | | | | |
|---|---|---|---|---|
| | $[\cdot]^{ss}$ | $[\cdot]^{sf}$ | $[\cdot]^{fs}$ | $[\cdot]^{ff}$ |
| $P_0$ | $J_{ss} + J_{sf}\Gamma_2$ | $J_{sf}\Gamma_3$ | $0$ | $J_{ff} - J_{fs}C_{ss}^{-1}C_{sf}$ |
| $P_1$ | $J_{sf}\Gamma_1$ | $0$ | $J_{fs}C_{ss}^{-1}h^{-1}$ | $0$ |
| $M$ | — | $J_{ss}\Pi_{sf} + J_{sf}\Pi_{ff}$ | $J_{fs}\Pi_{ss} + J_{ff}\Pi_{fs}$ | — |
| | | | | |
| $\Gamma_1$ | $J_{ff}^{-1}\left(C_{fs}\left(J_{ss} + I_{p_1^s}\right) + C_{ff}J_{fs}\right)C_{ss}^{-1}$ | | | |
| $\Gamma_2$ | $-J_{ff}^{-1}J_{fs}$ | | | |
| $\Gamma_3$ | $\left[J_{ff}^{-1}\left(C_{fs}J_{sf} + C_{ff}\left(J_{ff} + I_{p_1^f}\right)\right) - \Gamma_1 C_{sf}\right]h$ | | | |
| $\Gamma_4$ | $\left[C_{fs}B_s' + C_{ff}B_f'\right]h$ | | | |

| Subvectors | | |
|---|---|---|
| | $[\cdot]^s$ | $[\cdot]^f$ |
| $\beta_{1,2h}$ | $c_{1,2h}^s - M_{sf}\left(c_{2,h}^f - \tilde{\psi}_{2,h}^f\right) + J_{sf}\left(c_{1,h}^f - \tilde{\psi}_{1,h}^f\right)$ | $\psi_{1,2h}^f - M_{fs}\left(\tilde{\psi}_{2,h}^s - c_{2,h}^s\right) + J_{fs}\left(\tilde{\psi}_{1,h}^s - c_{1,h}^s\right)$ |
| $\beta_{1,th}$ | $c_{1,th}^s + J_{sf}g_{th}^f - P_0^{ss}\Pi_{sf}\phi_{2,th-h}^f + P_0^{sf}\Pi_{fs}\phi_{2,th-h}^s$ | $\psi_{1,th}^f - P_1^{fs}h_{1,th-h}^s + P_0^{ff}\Pi_{fs}\phi_{2,th-h}^s$ |
| $g_{th}$ | — | $J_{ff}^{-1}\left(h_{1,th}^f - c_{1,th}^f + \Gamma_4\psi_{th}\right) - \Gamma_1 h_{1,th-h}^s$ |
| | | |
| $h_{1,th}$ | $\left[a_1 + b_1\left(th - \frac{1}{2}h\right)\right]h$ | |
| $\phi_{2,th}$ | $\left[\left(\frac{1}{2}a_2 - \frac{1}{6}b_2 h\right) + \frac{1}{2}b_2 th\right]h$ | |

| Matrices | |
|---|---|
| $G$ | $\left(I_{p_1}, 0_{p_1\times p_2}\right)'$ |
| $H_0$ | $\left(I_p + GKB'\right)h$ |
| $H_1$ | $G\left(K_1 h^{-2} - \frac{1}{2}I_p\right)$ |
| $H_2$ | $G\left(K_2 h^{-2} - \frac{1}{2}I_{p_1}\right)$ |
| $H_3$ | $G\left(C^{-1}K_1 h^{-3} - \frac{1}{2h}C^{-1} - \frac{1}{3}I_{p_1}\right)$ |
| $H_4$ | $G\left(C^{-1}K_1 h^{-3} - C^{-2}Kh^{-2} - \frac{1}{6}I_{p_1}\right)$ |
| $K_1$ | $C^{-1}e^{hC}h - C^{-2}J$ |
| $K_2$ | $C^{-2}J - C^{-1}$ |

| Vectors | |
|---|---|
| $c_{th}$ | $H_0 a - \left(\frac{1}{2}I_p + H_1 B'\right)bh^2 + H_0 bth$ |
| $\psi_{th}$ | $H_0 a - \left(\frac{1}{2}(I_p + h^{-1}H_0) + H_1 B'\right)bh^2 + H_0 bth$ |
| $\tilde{\psi}_{th}$ | $\left(\frac{1}{2}I_p + H_2 B'\right)ah + \left(\frac{1}{6}I_p + (H_3 - H_4)B'\right)bh^2$ |

Table 3.7: Definition of additional matrices and vectors.

matrices $S_j$ $(j = 1, ..., 8)$ are

$$S_1 = \begin{pmatrix} I_{p_1^s} & -J_{sf}J_{ff}^{-1} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & I_{p_2^s} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad S_2 = \begin{pmatrix} & J_{sf}J_{ff}^{-1}\Gamma_4 & & \\ 0 & I_{p_1^f} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I_{p_2^f} \end{pmatrix},$$

$$S_3 = \begin{pmatrix} 0 & J_{sf}J_{ff}^{-1} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad S_4 = \begin{pmatrix} -J_{sf}\Gamma_1 & 0 & 0 & 0 \\ -P_1^{fs} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

$$S_5 = \begin{pmatrix} 0 & 0 & P_0^{sf}\Pi_{fs} & -P_0^{ss}\Pi_{sf} \\ 0 & 0 & P_0^{ff}\Pi_{fs} & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad S_6 = \begin{pmatrix} I_{p_1^s} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & I_{p_2^s} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

$$S_7 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & I_{p_1^f} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I_{p_2^f} \end{pmatrix}, \quad S_8 = \begin{pmatrix} 0 & J_{sf} & 0 & -M_{sf} \\ -J_{fs} & 0 & M_{fs} & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

The disturbance $\xi_{th}$ is a vector MA(1) satisfying the following properties

$$\begin{aligned} E(\xi_h\xi_h') &= \Omega_{11} = \int_0^h J_2(s)\Sigma J_2(s)'ds, \\ E(\xi_{2h}\xi_{2h}') &= \Omega_{22} = \int_0^h J_2(s)\Sigma J_2(s)'ds + \int_0^h J_3(s)\Sigma J_3(s)'ds, \\ E(\xi_{th}\xi_{th}') &= \Omega_0 = \int_0^h J_0(s)\Sigma J_0(s)'ds + \int_0^h J_1(s)\Sigma J_1(s)'ds, \ t = 3, ..., n, \\ E(\xi_{2h}\xi_h') &= \Omega_{21} = \int_0^h J_3(s)\Sigma J_2(s)'ds, \\ E(\xi_{3h}\xi_{2h}') &= \Omega_{32} = \int_0^h J_1(s)\Sigma J_2(s)'ds, \\ E(\xi_{th}\xi_{th-h}') &= \Omega_1 = \int_0^h J_1(s)\Sigma J_0(s)'ds, \ t = 4, ..., n, \\ E(\xi_{th}\xi_{th-jh}') &= 0, \quad j > 1, \end{aligned}$$

where

$$\int_0^h J_i(s)\Sigma J_k(s)'ds = J_{i0}\Sigma J_{k0}'h + \frac{1}{2}J_{i0}\Sigma J_{k1}'h^2 + J_{i0}\Sigma B(K' + I_{p_1})J_{k2}'h + \frac{1}{2}J_{i1}\Sigma J_{k0}'h^2$$
$$+ \frac{1}{3}J_{i1}\Sigma J_{k1}'h^3 + J_{i1}\Sigma BK_1'J_{k2}' + J_{i2}(K + I_{p_1})B'\Sigma J_{k0}'h$$
$$+ J_{i2}K_1B'\Sigma J_{k1}' + J_{i2}K_3J_{k2}',$$

$K_3 = \int_0^h e^{sC}B'\Sigma Be^{sC'}ds$ and the matrices $J_{ik}$ $(i = 0, 1, 2, 3; k = 0, 1, 2)$ are defined in the following table.

| | $k$ | | |
|---|---|---|---|
| $i$ | 0 | 1 | 2 |
| 0 | $S_1(I_p - GB') - \frac{1}{h}S_2GC^{-1}B' + S_3$ | $\frac{1}{h}S_2(I_p - GB')$ | $S_1G + \frac{1}{h}S_2GC^{-1}$ |
| 1 | $S_2(I_p - GB') + \frac{1}{h}S_2GC^{-1}e^{hC}B' + S_5 + S_4$ | $-\frac{1}{h}S_2(I_p - GB') - \frac{1}{h}S_5$ | $-\frac{1}{h}S_2GC^{-1}$ |
| 2 | $S_6(I_p - GB') - \frac{1}{h}S_7GC^{-1}B'$ | $\frac{1}{h}S_7(I_p - GB')$ | $S_6G + \frac{1}{h}S_7GC^{-1}$ |
| 3 | $-J_{31}h + \frac{1}{h}S_7GC^{-1}e^{hC}B' + \frac{1}{h}S_8GC^{-1}B'$ | $-\frac{1}{h}(S_7 + S_8)(I_p - GB')$ | $S_8G - \frac{1}{h}(S_7 + S_8)GC^{-1}$ |

Table 3.8: Definition of the $J_{ik}$ matrices.

The term $K_3$ can be obtained from the matrix exponential

$$\bar{M} = \exp\begin{pmatrix} -Ch & B'\Sigma Bh \\ 0 & C'h \end{pmatrix} = \begin{pmatrix} \bar{M}_{11} & \bar{M}_{12} \\ 0 & \bar{M}_{22} \end{pmatrix},$$

as $K_3 = \bar{M}_{22}'\bar{M}_{12}$ (see Jewitt and McCrorie, 2005).

## 3.8 Appendix: Proofs

*The matrix exponential.* The matrix $\bar{A}$ can be decomposed into a product of two matrices as follows

$$\bar{A} = A_1A_2', \tag{3.79}$$

where $A_1$ and $A_2'$ are matrices of dimension $(p+p_1)\times 2p_1$ and $2p_1\times(p+p_1)$, respectively, given by

$$A_1 = \begin{pmatrix} I_{p_1} & 0 \\ 0 & 0 \\ 0 & I_{p_1} \end{pmatrix}, \quad A_2' = \begin{pmatrix} 0 & 0 & I_{p_1} \\ \tilde{C} & -\tilde{C}\Pi & \Psi \end{pmatrix}.$$

It can be shown that $\bar{A}^j = A_1V^{j-1}A_2'$ $(j = 1, ..., \infty)$, where $V = A_2'A_1$, which enables the matrix exponential $e^{h\bar{A}}$ to be written as

$$e^{h\bar{A}} = I_{p+p_1} + A_1 V^{-1} \sum_{j=1}^{\infty} \frac{(hV)^j}{j!} A_2'. \tag{3.80}$$

Adding the term $A_1V^{-1}A_2' - A_1V^{-1}A_2'$ to the right-hand side of (3.80) yields

$$\begin{aligned} e^{h\bar{A}} &= I_{p+p_1} + A_1 V^{-1} \left[ \sum_{j=0}^{\infty} \frac{(hV)^j}{j!} - I_{2p_1} \right] A_2', \\ &= I_{p+p_1} + A_1 \bar{V} A_2', \end{aligned} \tag{3.81}$$

where $\bar{V} = V^{-1}(e^{hV} - I_{2p_1})$.

Partitioning $\bar{V}$, pre-multiplying by $A_1$, post-multiplying by $A_2'$, and adding the identity matrix $I_{p+p_1}$, we obtain

$$e^{h\bar{A}} = \begin{pmatrix} [e^{h\bar{A}}]_{11} & (I_{p_1} - [e^{h\bar{A}}]_{11})\Pi & [e^{h\bar{A}}]_{13} \\ 0 & I_{p_2} & 0 \\ [e^{h\bar{A}}]_{31} & -[e^{h\bar{A}}]_{31}\Pi & [e^{h\bar{A}}]_{33} \end{pmatrix},$$

where

$$\begin{aligned} [e^{h\bar{A}}]_{11} &= I_{p_1} + [\bar{V}]_{12}\tilde{C}, \\ [e^{h\bar{A}}]_{13} &= [\bar{V}]_{11} + [\bar{V}]_{12}\Psi, \\ [e^{h\bar{A}}]_{31} &= [\bar{V}]_{22}\tilde{C}, \\ [e^{h\bar{A}}]_{33} &= [\bar{V}]_{21} + [\bar{V}]_{22}\Psi + I_{p_1}. \end{aligned}$$

□

*Derivation of equations (3.18).* From (3.14) we obtain

$$x(th) = e^{h\bar{A}}x(th-h) + \int_{th-h}^{th} e^{(th-s)\bar{A}}(\mu^* + \gamma^* s)ds + \int_{th-h}^{th} e^{(th-s)\bar{A}} n^*(s)ds. \tag{3.82}$$

Setting $t = 1$ in (3.82), subtracting $x(0)$ on both sides of the equation and pre-multiplying by $\tilde{S}_1$ gives equation (3.18). □

*Derivation of equations (3.20) and (3.21).* Integrating equation (3.14) over the interval $(0, h)$ gives

$$\int_0^h x(r)dr = \int_0^h e^{r\bar{A}}dr x(0) + \int_0^h \int_0^r e^{(r-s)\bar{A}}(\mu^* + \gamma^* s)dsdr + \int_0^h \int_0^r e^{(r-s)\bar{A}} n^*(s)dsdr. \tag{3.83}$$

Equation (3.20) is obtained by subtracting $hx(0)$ on both sides of (3.83), and then pre-multiplying by $h^{-1}\tilde{S}_1$.

Integrating (3.82) from $h$ to $2h$, subtracting $\int_0^h x(r)dr$ on both sides of the equation and pre-multiplying by $h^{-1}\tilde{S}_1$ yields

$$\begin{aligned}\Delta\bar{y}(2h) = G&\left[([e^{h\bar{A}}]_{11} - I_{p_1})B'\bar{y}(h) + \frac{1}{h}[e^{h\bar{A}}]_{13}\int_0^h Dy_1(r)dr\right] \\ &+ \frac{1}{h}\tilde{S}_1\int_h^{2h}\int_{r-h}^r e^{(r-s)\bar{A}}(\mu^* + \gamma^* s)dsdr + \frac{1}{h}\tilde{S}_1\int_h^{2h}\int_{r-h}^r e^{(r-s)\bar{A}}n^*(s)dsdr. \quad (3.84)\end{aligned}$$

Note that (3.84) contains the unobservable $\int_0^h Dy_1(r)dr$, which can be eliminated by using the last $p_1$ equations of the system (3.83). To extract the last $p_1$ components from $\int_0^h x(r)dr$, we pre-multiply (3.83) by $\tilde{S} = (0, 0, I_{p_1})$, which results in

$$\begin{aligned}\int_0^h Dy_1(r)dr = &\int_0^h [e^{r\bar{A}}]_{31}drB'y(0) + \int_0^h [e^{r\bar{A}}]_{33}drDy_1(0) \\ &+ \tilde{S}\int_0^h\int_0^r e^{(r-s)\bar{A}}(\mu^* + \gamma^* s)dsdr + \tilde{S}\int_0^h\int_0^r e^{(r-s)\bar{A}}n^*(s)dsdr. \quad (3.85)\end{aligned}$$

Substituting (3.85) into (3.84) gives equation (3.21).

□

*Proof of Theorem (3.3.1)* . In order to derive the exact discrete model with mixed data it is convenient to partition equations (3.17)-(3.21) conformably with $y_1$, $y_2^s$ and $y_2^f$ as following.

Equation (3.17) (Discrete model if the sample was comprised entirely of stocks, $t = 2, ..., n$)

$$\begin{aligned}\Delta y_{1,th} = &Q\left[y_{1,th-h} - \Pi^s y^s_{2,th-h} - \Pi^f w^f_{2,th-h}\right] - F_{2,11}\Delta y_{1,th-h} - F^s_{2,12}\Delta y^s_{2,th-h} \\ &- F^f_{2,12}\Delta w^f_{2,th-h} + \kappa_{1,th} + \vartheta_{1,th} \quad (3.86)\end{aligned}$$

$$\Delta y^s_{2,th} = \kappa^s_{2,th} + \vartheta^s_{2,th} \quad (3.87)$$

$$\Delta w^f_{2,th} = \kappa^f_{2,th} + \vartheta^f_{2,th} \quad (3.88)$$

Equation (3.18) (Discrete model if the sample was comprised entirely of stocks, $t = 1$)

$$y_{1,h} - y_1(0) = R_1 B'y(0) + R_2 Dy_1(0) + \tilde{\kappa}_{1,h} + \tilde{\vartheta}_{1,h} \quad (3.89)$$

$$y^s_{2,h} - y^s_2(0) = \tilde{\kappa}^s_{2,h} + \tilde{\vartheta}^s_{2,h} \quad (3.90)$$

$$w^f_{2,h} - y^f_2(0) = \tilde{\kappa}^f_{2,h} + \tilde{\vartheta}^f_{2,h} \quad (3.91)$$

Equation (3.19) (Discrete model if the sample was comprised entirely of flows, $t = 3, ..., n$)

$$\Delta w_{1,th} = Q\left[w_{1,th-h} - \Pi^s w^s_{2,th-h} - \Pi^f y^f_{2,th-h}\right] - F_{2,11}\Delta w_{1,th-h} - F^s_{2,12}\Delta w^s_{2,th-h}$$
$$- F^f_{2,12}\Delta y^f_{2,th-h} + \iota_{1,th} + \varsigma_{1,th} \tag{3.92}$$
$$\Delta w^s_{2,th} = \iota^s_{2,th} + \varsigma^s_{2,th} \tag{3.93}$$
$$\Delta y^f_{2,th} = \iota^f_{2,th} + \varsigma^f_{2,th} \tag{3.94}$$

Equation (3.20) (Discrete model if the sample was comprised entirely of flows, $t = 1$)

$$w_{1,h} - y_1(0) = U_1 B' y(0) + U_2 D y_1(0) + \tilde{\iota}_{1,h} + \tilde{\varsigma}_{1,h} \tag{3.95}$$
$$w^s_{2,h} - y^s_2(0) = \tilde{\iota}^s_{2,h} + \tilde{\varsigma}^s_{2,h} \tag{3.96}$$
$$y^f_{2,h} - y^f_2(0) = \tilde{\iota}^f_{2,h} + \tilde{\varsigma}^f_{2,h} \tag{3.97}$$

Equation (3.21) (Discrete model if the sample was comprised entirely of flows, $t = 2$)

$$\Delta w_{1,2h} = R_1\left[w_{1,h} - \Pi^s w^s_{2,h} - \Pi^f y^f_{2,h}\right] + R_2 U_3 B' y(0) + R_2 U_4 D y_1(0) + \tilde{\iota}_{1,2h}$$
$$+ \tilde{\varsigma}_{1,2h} \tag{3.98}$$
$$\Delta w^s_{2,2h} = \tilde{\iota}^s_{2,2h} + \tilde{\varsigma}^s_{2,2h} \tag{3.99}$$
$$\Delta y^f_{2,2h} = \tilde{\iota}^f_{2,2h} + \tilde{\varsigma}^f_{2,2h} \tag{3.100}$$

Equation (3.22) is obtained by combining the observable stock components of (3.18) (i.e. equations (3.89) and (3.90)) and the observable flow components of (3.20) (i.e. equation (3.97)). To derive equation (3.23) we use (3.86), (3.87) and (3.100), after setting $t = 2$ in the first two equations. Note that (3.86) contains the unobservable component $w^f_{2,h}$, and its first difference $\Delta w^f_{2,h}$, which must be eliminated and replaced with the observables $y^f_{2,h}$ and $\Delta y^f_{2,h}$, respectively. The unobservable flows can be eliminated by taking the difference of (3.91) and (3.97), which gives

$$\Delta w^f_{2,h} - \Delta y^f_{2,h} = \tilde{\kappa}^f_{2,h} - \tilde{\iota}^f_{2,h} + \tilde{\vartheta}^f_{2,h} - \tilde{\varsigma}^f_{2,h}. \tag{3.101}$$

It follows from the above equations that $\Delta w^f_{2,h} - \Delta y^f_{2,h} = w^f_{2,h} - y^f_{2,h}$. Substituting the last expressions into (3.86) results in

$$\Delta y_{1,2h} = QB' y_h - F_{2,1}\Delta y_h + \kappa_{1,2h} - (Q\Pi^f + F^f_{2,12})(\tilde{\kappa}^f_{2,h} - \tilde{\iota}^f_{2,h})$$
$$+ \vartheta_{1,2h} - (Q\Pi^f + F^f_{2,12})(\tilde{\vartheta}^f_{2,h} - \tilde{\varsigma}^f_{2,h}), \tag{3.102}$$

where $F_{2,1} = (F_{2,11}, F^s_{2,12}, F^f_{2,12})$. Equation (3.23) is obtained by combining (3.102), (3.87) and (3.100).

To derive equation (3.24) we use (3.86), (3.87) and (3.94). As before, the unobservable components contained in (3.86) (i.e. $w^f_{2,th-h}$ and $\Delta w^f_{2,th-h}$) must be eliminated and replaced with observable terms (i.e. $y^f_{2,th-h}$ and $\Delta y^f_{2,th-h}$). To do this, we use the result given in Lemma 1 in Chambers (2009)

$$y_2(th) - \bar{y}_2(th) = \phi_{2,th} + \epsilon_{2,th}, \quad t = 1, ..., T, \tag{3.103}$$

where

$$\phi_{2,th} = h\left[\left(\frac{1}{2}a_2 - \frac{1}{6}b_2 h\right) + \frac{1}{2}b_2 th\right], \tag{3.104}$$

$$\epsilon_{2,th} = \frac{1}{h}\int_{th-h}^{th} [h - (th - s)]u_2(s)ds. \tag{3.105}$$

Applying the difference operator to (3.103), we have

$$\Delta y_2(th) - \Delta\bar{y}_2(th) = \Delta\phi_{2,th} + \Delta\epsilon_{2,th}, \quad t = 2, ..., T. \tag{3.106}$$

Partitioning (3.103) into stocks and flows yields

$$y^s_{2,th} - w^s_{2,th} = \phi^s_{2,th} + \epsilon^s_{2,th} \tag{3.107}$$

$$w^f_{2,th} - y^f_{2,th} = \phi^f_{2,th} + \epsilon^f_{2,th}. \tag{3.108}$$

Similarly, partitioning (3.106) into stocks and flows gives

$$\Delta y^s_{2,th} - \Delta w^s_{2,th} = \Delta\phi^s_{2,th} + \Delta\epsilon^s_{2,th} \tag{3.109}$$

$$\Delta w^f_{2,th} - \Delta y^f_{2,th} = \Delta\phi^f_{2,th} + \Delta\epsilon^f_{2,th}. \tag{3.110}$$

The variable $\Delta y_{1,th}$ is expressed in terms of observables by using (3.108) and (3.110) to eliminate $w^f_{2,th-h}$ and $\Delta w^f_{2,th-h}$, which gives

$$\begin{aligned}\Delta y_{1,th} = {} & QB'y_h - F_{2,1}\Delta y_h + \kappa_{1,th} - (Q\Pi^f + F^f_{2,12})\phi^f_{2,th-h} + F^f_{2,12}\phi^f_{2,th-2h} \\ & + \vartheta_{1,th} - (Q\Pi^f + F^f_{2,12})\epsilon^f_{2,th-h} + F^f_{2,12}\epsilon^f_{2,th-2h}.\end{aligned} \tag{3.111}$$

Equation (3.24) is obtained by combining (3.111), (3.87) and (3.94).

□

*Proof of Theorem (3.3.2)* . The proof is very similar to that of Theorem (3.3.1) with the difference that the first element of equations (3.17)-(3.21) (i.e. the equation corresponding to $y_1$) are now given by

$$\begin{aligned}
\Delta w_{1,th} &= Q\left[w_{1,th-h} - \Pi^s y^s_{2,th-h} - \Pi^f w^f_{2,th-h}\right] - F_{2,11}\Delta w_{1,th-h} - F^s_{2,12}\Delta y^s_{2,th-h} \\
&\quad - F^f_{2,12}\Delta w^f_{2,th-h} + \kappa_{1,th} + \vartheta_{1,th} && (3.112)\\
w_{1,h} - y_1(0) &= R_1 B' y(0) + R_2 D y_1(0) + \tilde{\kappa}_{1,h} + \tilde{\vartheta}_{1,h} && (3.113)\\
\Delta y_{1,th} &= Q\left[y_{1,th-h} - \Pi^s w^s_{2,th-h} - \Pi^f y^f_{2,th-h}\right] - F_{2,11}\Delta y_{1,th-h} - F^s_{2,12}\Delta w^s_{2,th-h} \\
&\quad - F^f_{2,12}\Delta y^f_{2,th-h} + \iota_{1,th} + \varsigma_{1,th} && (3.114)\\
y_{1,h} - y_1(0) &= U_1 B' y(0) + U_2 D y_1(0) + \tilde{\iota}_{1,h} + \tilde{\varsigma}_{1,h} && (3.115)\\
\Delta y_{1,2h} &= R_1\left[y_{1,h} - \Pi^s w^s_{2,h} - \Pi^f y^f_{2,h}\right] + R_2 U_3 B' y(0) + R_2 U_4 D y_1(0) + \tilde{\iota}_{1,2h} + \tilde{\varsigma}_{1,2h} \\
& && (3.116)
\end{aligned}$$

To derive equations (3.25)-(3.27) we use (3.114)-(3.116) as well as the observable vectors corresponding to $y_2$, and follow the steps described in the proof of Theorem (3.3.1) to eliminate the unobservables $w^s_{2,th-h}$ and $\Delta w^s_{2,th-h}$.

□

*Proof of Theorem (3.3.3)* . The autocovariance matrices of $\xi_{th}$ are computed from the moving average representation given in (3.37)-(3.39), noting that $\zeta_{th}$ is a vector white noise process with covariance matrix $\Omega_\zeta$; this last matrix being obtained by considering the covariances between the elements of $\zeta_{th}$. As an illustration, the entry (1,1) of $\Omega_\zeta$ is

$$\begin{aligned}
E(\zeta_{1,th}\zeta'_{1,th}) &= E\left[\int_{th-h}^{th} F(th-s)n^*(s)ds\right]\left[\int_{th-h}^{th} F(th-s)n^*(s)ds\right]', \\
&= \int_{th-h}^{th} F(th-s)\Sigma^* F(th-s)' ds, \\
&= \int_0^h F(s)\Sigma^* F(s)' ds. && (3.117)
\end{aligned}$$

□

*Proof of Theorem (3.3.4)* . First note that using the alternative expression for the matrix exponential, we can write $F(s) = e^{s\bar{A}}$ as

$$F(s) = J_1 + J_2 V e^{sV} A_2', \tag{3.118}$$

from which the following expressions are obtained

$$\Theta(s) = \int_0^s \digamma(r)dr = J_0 + J_1 s + J_2 e^{sV} A_2', \tag{3.119}$$

$$\Upsilon(s) = \int_0^s \Theta(r)dr = J_0 s + \frac{1}{2} J_1 s^2 + J_2 \bar{V} A_2', \tag{3.120}$$

$$\beth(s) = \int_0^s r\digamma(r)dr = \frac{1}{2} J_1 s^2 + J_2 V \bar{V}_1 A_2', \tag{3.121}$$

$$\aleph(s) = \int_0^s r\Theta(r)dr = \frac{1}{2} J_0 s^2 + \frac{1}{3} J_1 s^3 + J_2 \bar{V}_1 A_2', \tag{3.122}$$

with $\bar{V}$ and $\bar{V}_1$ being

$$\bar{V} = \int_0^h e^{rV} dr = V^{-1}(e^{hV} - I_{2p_1}), \tag{3.123}$$

$$\bar{V}_1 = \int_0^h r e^{rV} dr = V^{-1}(h e^{hV} - \bar{V}). \tag{3.124}$$

Now, to derive (3.40) and (3.41) we use a change of variable in the integration, which results in

$$\tilde{\kappa}_h = \tilde{S}_1 \int_0^h \digamma(r)[\mu^* + \gamma^*(h - r)]dr, \tag{3.125}$$

$$\kappa_{th} = \tilde{S}_1 \int_0^h \digamma(r)[\mu^* + \gamma^*(th - r)]dr + \tilde{S}_2 \int_0^h \digamma(r)[\mu^* + \gamma^*(th - h - r)]dr. \tag{3.126}$$

Evaluating the integrals gives the expressions for $\tilde{\kappa}_h$ and $\kappa_{th}$ in the Theorem.

The equation for $\iota_{th}$ is obtained by integrating $\kappa_{th}$ again from $th - h$ to $th$, after dividing by $h$ and evaluating the integrals.

To obtain (3.43), first note that $\tilde{\iota}_{2h}$ can be written as

$$\begin{aligned} \tilde{\iota}_{2h} &= \frac{1}{h}\left\{ \tilde{S}_1 \int_h^{2h} \Theta(2h - s)(\mu^* + \gamma^* s)ds + \tilde{S}_1 \int_0^h [\Theta(h) - \Theta(h - s)](\mu^* + \gamma^* s)ds \right. \\ &\quad \left. + \tilde{S}_3 \int_0^h \Theta(h - s)(\mu^* + \gamma^* s)ds \right\}, \\ &= \frac{1}{h}\left\{ \tilde{S}_1 \int_0^h \Theta(s)(\mu^* + \gamma^*(2h - s))ds + \tilde{S}_1 \int_0^h [\Theta(h) - \Theta(s)](\mu^* + \gamma^*(h - s))ds \right. \\ &\quad \left. + \tilde{S}_3 \int_0^h \Theta(s)(\mu^* + \gamma^*(h - s))ds \right\}, \end{aligned} \tag{3.127}$$

where the first line follows from an interchange of the orders of integration and the last line from a change of variable in the integration. Evaluating the integrals gives the expression for $\tilde{\iota}_{2h}$ in the Theorem.

Similarly, equation (3.42) is obtained from

$$\tilde{\iota}_h = \frac{1}{h}\tilde{S}_1 \int_0^h \Theta(s)[\mu^* + \gamma^*(h-s)]ds. \tag{3.128}$$

The integrals determining $\Omega_\zeta$ are obtained by using the expressions given in (3.118)-(3.122) and noting that $\Xi(s) = \Theta(h) - \Theta(s)$. For example,

$$\begin{aligned}
\int_0^h F(r)\Sigma^* F(r)' dr &= \int_0^h [J_1 + J_2 V e^{sV} A_2'] \Sigma^* [J_1 + J_2 V e^{sV} A_2']' dr \\
&= \int_0^h \Big[ J_1 \Sigma^* J_1' + J_1 \Sigma^* A_2 e^{sV'} V' J_2' \\
&\quad + J_2 V e^{sV} A_2' \Sigma^* J_1' + J_2 V e^{sV} A_2' \Sigma^* A_2 e^{sV'} V' J_2' \Big] dr \\
&= h J_1 \Sigma^* J_1' + J_1 \Sigma^* A_2 \bar{V}' V' J_2' + J_2 V \bar{V} A_2' \Sigma^* J_1' + J_2 V \bar{V}_2 V' J_2'. \qquad (3.129)
\end{aligned}$$

□

# Conclusions

In the first chapter, we obtain analytical expressions to approximate the bias of the maximum likelihood estimator in a univariate model with a known mean. Explicit expressions for the asymptotic variance and the asymptotic bias that arises when the correlation of the disturbances induced by temporal aggregation is not taken into account are also provided. A Monte Carlo study suggests that the performance of the formulae is reasonably good. Findings also suggest that the bias is larger when the initial condition is fixed than when it is random and smaller for flows than for stocks.

The second chapter provides a comparison of four bias corrected estimators. A Monte Carlo experiment shows that all approaches deliver substantial bias reductions when the original estimator is consistent. Results are robust to non-normality. In contrast, heteroskedasticity has a significant impact on all estimators. The indirect inference method of bias reduction is found to work particularly well when the original estimator is inconsistent, a case occurring when the serial correlation in the discrete time disturbances is ignored.

The third chapter has derived the exact discrete representation corresponding to a mixed-order system of stochastic differential equations with a mixture of stock and flow variables. This chapter also explores the finite sample behaviour of the Gaussian estimator in cointegrated continuous time systems. Results are consistent with what has been found for stationary continuous time models. In particular, the Monte Carlo simulations suggest that the bias and variance of the estimators of the short-run, long-run and adjustment coefficients, as well as the variance of the intercepts are mainly determined by the data span, while the bias and variance of the covariance coefficients seem to depend on the sample size. This finding explains why estimation of the covariance parameters enjoys smaller bias and variance than the other parameters.

The approach used to derive bias expressions as well as the bias reduction methods employed in this dissertation are expected to work well under much more general models, for example higher order differential equations with mixed stock and flow data. Their implementation in empirical applications provides an exciting agenda for future research.

# Bibliography

Agbeyegbe, T. D. (1984). The exact discrete analog to a closed linear mixed-order system. *Journal of Economic Dynamics and Control*, 7:363–375.

Andrews, D. W. K. (1993). Exactly median-unbiased estimation of first order autoregressive/unit root models. *Econometrica*, 61:139–165.

Andrews, D. W. K. and Chen, H.-Y. (1994). Approximately median-unbiased estimation of autoregressive models. *Journal of Business & Economic Statistics*, 12:187–204.

Bao, Y. and Ullah, A. (2007). The second-order bias and mean squared error of estimators in time-series models. *Journal of Econometrics*, 140:650–669.

Bao, Y., Ullah, A., Wang, Y., and Yu, J. (2015). Bias in the estimation of mean reversion in continuous-time Lévy processes. *Economics Letters*, 134:16–19.

Bergstrom, A. R. (1983). Gaussian estimation of structural parameters in higher order continuous time dynamic models. *Econometrica*, 51:117–152.

Bergstrom, A. R. (1985). The estimation of parameters in nonstationary higher-order continuous-time dynamic models. *Econometric Theory*, 1:369–385.

Bergstrom, A. R. (1986). The estimation of open higher-order continuous time dynamic models with mixed stock and flow data. *Econometric Theory*, 2:350–373.

Bergstrom, A. R. (1990). *Continuous Time Econometric Modelling.* Oxford University Press.

Bergstrom, A. R. (1996). Survey of continuous time econometrics. In Barnett, W. A., Gandolfo, G., and Hillinger, C., editors, *Dynamic Disequilibrium Modeling*, pages 3–25. Cambridge University Press.

Bergstrom, A. R. (1997). Gaussian estimation of mixed-order continuous-time dynamic models with unobservable stochastic trends from mixed stock and flow data. *Econometric Theory*, 13:467–505.

Bergstrom, A. R. (2009). The effects of differencing on the gaussian likelihood of models with unobservable stochastic trends: a simple example. *Econometric Theory*, 25:903–913.

Bergstrom, A. R. and Nowman, K. B. (2007). *A Continuous Time Econometric Model of the United Kingdom with Stochastic Trends*. Cambridge University Press.

Bergstrom, A. R. and Wymer, C. R. (1976). A model of disequilibrium neoclassical growth and its application to the United Kingdom. In *Statistical Inference in Continuous Time Economic Models*, pages 267–327. North-Holland.

Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods*. Springer.

Chambers, M. J. (1999). Discrete time representation of stationary and non-stationary continuous time systems. *Journal of Economic Dynamics and Control*, 23:619–639.

Chambers, M. J. (2009). Discrete time representations of cointegrated continuous time models with mixed sample data. *Econometric Theory*, 25:1030–1049.

Chambers, M. J. (2011). Cointegration and sampling frequency. *The Econometrics Journal*, 14:156–185.

Chambers, M. J. (2013). Jackknife estimation of stationary autoregressive models. *Journal of Econometrics*, 172:142–157.

Chambers, M. J. and Kyriacou, M. (2013). Jackknife estimation with a unit root. *Statistics & Probability Letters*, 83:1677–1682.

Chen, Y. and Yu, J. (2015). Optimal jackknife for unit root models. *Statistics and Probability Letters*, 99:135–142.

Cordeiro, G. M. and Klein, R. (1994). Bias correction in ARMA models. *Statistics & Probability Letters*, 19:169–176.

Cox, D. R. and Snell, J. (1968). A general definition of residuals. *Journal of the Royal Statistical Society B*, 30:248–275.

Cox, J. C., Ingersoll, J. E., and Ross, S. A. (1985). A theory of the term structure of interest rates. *Econometrica*, 53:385–407.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *The Annals of Statistics*, 7:1–26.

Efron, B. (1982). *The Jackknife, the Bootstrap, and Other Resampling Plans*. Society for Industrial and Applied Mathematics, Philadelphia.

Fisher, I. (1930). *The Theory of Interest as Determined by Impatience to Spend Income and Opportunity to Invest it*. Macmillan.

Gallant, A. R. and Tauchen, G. (1996). Which moments to match? *Econometric Theory*, 12:657–681.

Gonçalves, S. and Kilian, L. (2004). Bootstrapping autoregressions with conditional heteroskedasticity of unknown form. *Journal of Econometrics*, 123:89–120.

Gonzalo, J. (1994). Five alternative methods of estimating long-run equilibrium relationships. *Journal of Econometrics*, 60:203–233.

Gourieroux, C. and Monfort, A. (1997). *Simulation-Based Econometric Methods*. Oxford University Press.

Gourieroux, C., Monfort, A., and Renault, E. (1993). Indirect inference. *Journal of Applied Econometrics*, 8:S85–S118.

Gouriéroux, C., Renault, E., and Touzi, N. (2000). Calibration by simulation for small sample bias correction. In *Simulation-Based Inference in Econometrics: Methods and Applications*, pages 328–358. Cambridge University Press.

Harvey, A. C. and Stock, J. H. (1985). The estimation of higher-order continuous time autoregressive models. *Econometric Theory*, 1:97–117.

Harvey, A. C. and Stock, J. H. (1988). Continuous time autoregressive models with common stochastic trends. *Journal of Economic Dynamics and Control*, 12:365–384.

Jewitt, G. and McCrorie, J. R. (2005). Computing estimates of continuous time macroeconometric models on the basis of discrete data. *Computational Statistics & Data Analysis*, 49:397–416.

Johansen, S. and Juselius, K. (1990). Maximum likelihood estimation and inference on cointegration - with applications to the demand for money. *Oxford Bulletin of Economics and Statistics*, 52:169–210.

Klau, M. and Fung, S. S. (2006). The new BIS effective exchange rate indices. *BIS Quarterly Review, March*.

MacKinnon, J. G. and Smith, A. A. (1998). Approximate bias correction in econometrics. *Journal of Econometrics*, 85:205–230.

Maeshiro, A. (1999). A lagged dependent variable, autocorrelated disturbances, and unit root tests-peculiar OLS bias properties - A pedagogical note. *Applied Economics*, 31:381–396.

Marriott, F. and Pope, J. (1954). Bias in the estimation of autocorrelations. *Biometrika*, 41:390–402.

McCrorie, J. R. (2000). Deriving the exact discrete analog of a continuous time system. *Econometric Theory*, 16:998–1015.

McCrorie, J. R. and Chambers, M. J. (2006). Granger causality and the sampling of economic processes. *Journal of Econometrics*, 132:311–336.

Phillips, P. C. and Wickens, M. R. (1978). *Exercises in Econometrics*, volume 2. P. Allan.

Phillips, P. C. and Yu, J. (2005). Jackknifing bond option prices. *Review of Financial Studies*, 18:707–742.

Phillips, P. C. B. (1991a). Error correction and long-run equilibrium in continuous time. *Econometrica*, 59:967–980.

Phillips, P. C. B. (1991b). Optimal inference in cointegrated systems. *Econometrica*, 59:283–306.

Phillips, P. C. B. and Yu, J. (2009). Maximum likelihood and gaussian estimation of continuous time models in finance. In *Handbook of Financial Time Series*, pages 497–530. Springer.

Quenouille, M. H. (1949). Approximate tests of correlation in time-series. *Journal of the Royal Statistical Society B*, 11:68–84.

Robinson, P. (1993). Continuous-time models in econometrics: closed and open systems, stocks and flows. In Phillips, P. C. B., editor, *Models, Methods, and Applications of Econometrics: Essays in Honor of A.R. Bergstrom*, pages 71–90. Blackwell.

Rogoff, K. (1996). The purchasing power parity puzzle. *Journal of Economic Literature*, 34:647–668.

Rudebusch, G. D. (1992). Trends and random walks in macroeconomic time series: a re-examination. *International Economic Review*, 33:661–680.

Smith, A. A. (1993). Estimating nonlinear time-series models using simulated vector autoregressions. *Journal of Applied Econometrics*, 8:S63–S84.

Smith, A. A. J., Sowell, F., and Zin, S. E. (1997). Fractional integration with drift: estimation in small samples. *Empirical Economics*, 22:103–116.

Tang, C. Y. and Chen, S. X. (2009). Parameter estimation and bias correction for diffusion processes. *Journal of Econometrics*, 149:65–81.

Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics*, 5:177–188.

Wang, X., Phillips, P. C. B., and Yu, J. (2011). Bias in estimating multivariate and univariate diffusions. *Journal of Econometrics*, 161:228–245.

Working, H. (1960). Note on the correlation of first differences of averages in a random chain. *Econometrica*, 28:916–918.

Yu, J. (2012). Bias in the estimation of the mean reversion parameter in continuous time models. *Journal of Econometrics*, 169:114–122.

Zadrozny, P. (1988). Gaussian likelihood of continuous-time ARMAX models when data are stocks and flows at different frequencies. *Econometric Theory*, 4:108–124.