On the asymptotic properties of the OLS estimator in regression models with fractionally integrated regressors and errors

Dissertation

zur Erlangung des akademischen Grades
des Doktors der Naturwissenschaften (Dr. rer. nat.)
an der
Universität Konstanz
Mathematisch-Naturwissenschaftliche Sektion
Fachbereich Mathematik und Statistik

vorgelegt von

Toni Clemens Stocker

Tag der mündlichen Prüfung: 20.05.2008

Referent: Prof. Dr. Siegfried Heiler

Referent: Prof. Dr. Jan Beran

Acknowledgements

I would like to thank my advisor Prof. Dr. Siegfried Heiler for his guidance during my time in Konstanz and supervision of my dissertation. I thank Prof. Dr. Jan Beran for refereeing my dissertation. I am grateful to Prof. Dr. Enno Mammen (University of Mannheim) for being a member of the examining committee. I am indebted to Prof. James Davidson for highly appreciated comments. I thank very much Martin Schützner and Dr. Kyusang Yu (University of Mannheim) for their comments and discussion. I cordially thank Milton Rueda-Varon and Hoda Elattar for their moral support and their kindness all the time.

Contents

Zusammenfassung

Die Kleinste-Quadrate-Methode (KQ-Methode) ist die in der Regressionsanalyse am häufigsten verwendete Schätzmethode. Ziel dieser Dissertation ist es, die asymptotischen Eigenschaften des KQ-Schätzers in stochastischen Regressionsmodellen mit fraktionell integrierten Regressoren und Störtermen herzuleiten. Die Analyse wird im diskreten Zeitbereich durchgeführt und beschränkt sich auf sogenannte Vektor autoregressive fraktionell integrierte Moving average Prozesse (VARFIMA). Diese parametrische Prozess-klasse ist flexibel genug, vielfältige Arten von Zeitreihen in der Praxis zu modellieren. Der Schwerpunkt liegt auf der Herleitung asymptotischer Verteilungen und ihrer korrespondierender Konvergenzraten. Dazu werden jüngste Entwicklungen in der Forschung bezüglich der Verteilung von Stichprobenkovarianzen mit nichtstationären Prozessen berücksichtigt. Unter anderem werden hinreichende Bedingungen für die asymptotische Normalität der KQ-Schätzung in einem nichtstationären fraktionell integrierten dynamischen Regressionsmodell aufgestellt. Ein früheres Resultat für ganzzahligen Integrationsgrad wird damit verallgemeinert.

Die vorliegende Dissertation stellt eine der derzeit umfassendsten Arbeiten zur KQ-Asymptotik in fraktionell integrierten Regressionsmodellen dar.

Summary

Ordinary least squares (OLS) is the most common method of estimation in regression analysis. This dissertation aims to derive the asymptotic properties of OLS in stochastic regression models with fractionally integrated regressors and error terms. The analysis is carried out in the discrete time domain and restricted to so-called vector autoregressive fractionally integrated moving average processes (VARFIMA). This parametric class of processes is flexible enough to model many different kinds of time series in practice. The main focus is on deriving the asymptotic distributions and its corresponding convergence rates. For this purpose latest developments in research concerning the distribution of sample covariances involving nonstationary processes are accounted for. Among other things, sufficient conditions for the asymptotic normality of OLS in a nonstationary fractionally integrated dynamic regression model are stated. This extends a result earlier derived for integer-valued integration order.

This dissertation is one of the most comprehensive papers at the moment dealing with OLS asymptotics in fractionally integrated regression models.

1 Introduction

In this dissertation, the asymptotic properties of ordinary least squares (OLS) are derived for the univariate regression models

(M1) $y_t = \beta x_t + e_t$,

(M2) $y_t = \beta_1 x_{t1} + \beta_2 x_{t2} + e_t$,

(M3) $y_t = \alpha y_{t-1} + \beta z_t + e_t$, $|\alpha| < 1$,

where the regressors and the errors are assumed to be jointly driven by vector autoregressive fractionally integrated moving average processes (VARFIMA), possibly nonstationary. VARFIMA processes constitute a parametric subclass of fractionally integrated processes, frequently denoted as I(d) processes, where the parameter d, which may be any real number, is the integration order of the process. The main focus is on deriving the asymptotic distributions of OLS and its corresponding convergence rates, but we also make statements about under which conditions OLS yields consistent or inconsistent estimates. Among other things, it is shown that the OLS estimator for $(\alpha, \beta)^T$ in (M3) is asymptotically normal if z_t is a nonstationary I(d_z) regressor with $0.75 < d_z < 1.25$ and e_t a stationary I(d_e) error with $-0.5 < d_e < 0.25$. This result can be regarded as new and it generalizes a theorem earlier derived by Maekawa et. al. (1996) which states asymptotic normality for the integer-valued case with I(1) regressor and I(0) error.

The outline of the dissertation is as follows. Chapter 2 deals with the notion of 'I(d) process'. Prevailing concepts in literature are shortly introduced and contrasted to each other. In fact, there is no single and unique I(d) definition. Therefore, we need to clarify how our framework fits into theses different concepts. Chapter 3 reviews to what extent asymptotic properties of OLS in regression have already been treated for I(d) processes. The objective is also to clarify to what extent this dissertation provides new results under model settings which have not been examined yet. Chapter 4 formally introduces VARFIMA processes with its most important properties needed in subsequent parts. Chapter 5 deals with the asymptotics of sample covariances involving I(d) processes. Sample covariances are the essential ingredients of OLS regression formulas. Chapter 6 contains the main results of this paper. The asymptotic properties of OLS are examined for models (M1)-(M3). Chapter 7 specializes these results to the frequently examined integer-valued case with I(1) regressors and I(0) errors. Chapter 8 provides suggestions for further research.

2 Fractionally integrated processes

The general notion of 'I(d) process' lacks a uniform definition in literature, and we also have to distinguish between time domain and frequency domain devoted definitions.

In the time domain, a common approach is to combine a more or less rigorous I(0) definition with the application of the fractional difference operator. For any real number d, the fractional difference operator $\Delta^d(L) = (1-L)^d$ is defined by means of the binomial expansion

$$\Delta^d(L) := (1-L)^d = \sum_{j=0}^{\infty} \pi_j L^j ,$$

where L represents the usual lag operator with $Lx_t = x_{t-1}$,

$$\pi_j = \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)} = \prod_{k=1}^{j} \frac{k-1-d}{k}, \; j = 0,1,2,\ldots,$$

and $\Gamma(\cdot)$ is the gamma function with

$$\Gamma(x) := \begin{cases} \int_0^{\infty} t^{x-1} e^{-t} dt, & x > 0, \\ \infty, & x = 0, \\ x^{-1}\Gamma(1+x), & x < 0. \end{cases}$$

A process z_t is said to be integrated of order *d*, or shortly I(d), if and only if

$$(2.1) \qquad \Delta^d(L) z_t = \sum_{j=0}^{\infty} \pi_j z_{t-j} = w_t ,$$

where the process w_t matches the I(0) definition. Now the problem is deferred to the definition of I(0) processes. Davidson (2007) provides a survey of common I(0) definitions in literature and shows that some of these definitions are by no means equivalent. It is even not quite clear how the concepts these definitions rely on (e.g., stationarity, short memory and finite variance) might be connected with one another. One more rigorous definition is given by Engle and Granger (1987, p. 252):

"Definition: A series with no deterministic component which has a stationary, invertible ARMA representation after differencing d times is said be integrated of order d ..."

In fact, this parametric approach implicitly defines 'autoregressive fractionally integrated moving average' processes (ARFIMA) and it can be easily extended to the vector case, i.e., VARFIMA processes. Another widely used approach (e.g., Park and Phillips [1988], [1989]) is to say, a zero mean time series w_t is I(0), if the partial sum process W_n defined on the unit interval by

$$W_n(s) := \frac{1}{\sqrt{n}} \sum_{t=1}^{[ns]} w_t \,, \text{ for } 0 \le s \le 1 \,,$$

where $[x]$ denotes the largest integer not exceeding x, converges weakly to ordinary Brownian motion B_0 with variance $\varpi^2 = \lim_{n\to\infty} Var(W_n(1))$ as $n \to \infty$. Hence we have

(2.2) $\quad W_n \Rightarrow B_0 \,,$

where '$\Rightarrow$' denotes weak convergence in the space of measures on $D[0,1]$, the space of cadlag functions of the unit interval equipped with the Skorohod-Topology. This approach of specifying conditions under which some asymptotic theory is valid is a pragmatic one. The first functional central limit theorem (FCLT) according to (2.2) was given by Donsker (1951) for the case when the w_t's are independent and identically distributed (i.i.d.) with zero mean and finite positive variance. This result was extended by various authors to more general processes and also to the multivariate case. Probably the most general conditions for (2.2) to hold are provided by Davidson and de Jong (2000a), and these include ARMA processes just as special cases.

Depending on the parameter *d*, I(d) processes may be stationary or nonstationary. Let ε_t be an i.i.d. sequence with zero mean and finite variance $\sigma^2 > 0$. When setting $w_t = \varepsilon_t$ in (2.1), the process z_t is stationary if $d < 0.5$ and otherwise nonstationary. For $|\, d \,| < 0.5$, the process z_t is frequently called 'fractionally integrated noise' and its autocovariance function $\gamma_z(h)$ is given by

(2.3) $$\gamma_z(h) = \sigma^2 \frac{\Gamma(h+d)\Gamma(1-2d)}{\Gamma(h+1-d)\Gamma(d)\Gamma(1-d)} \,,$$

where for h large enough,

(2.4) $$\gamma_z(h) \sim \sigma^2 h^{2d-1} \frac{\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)} .$$

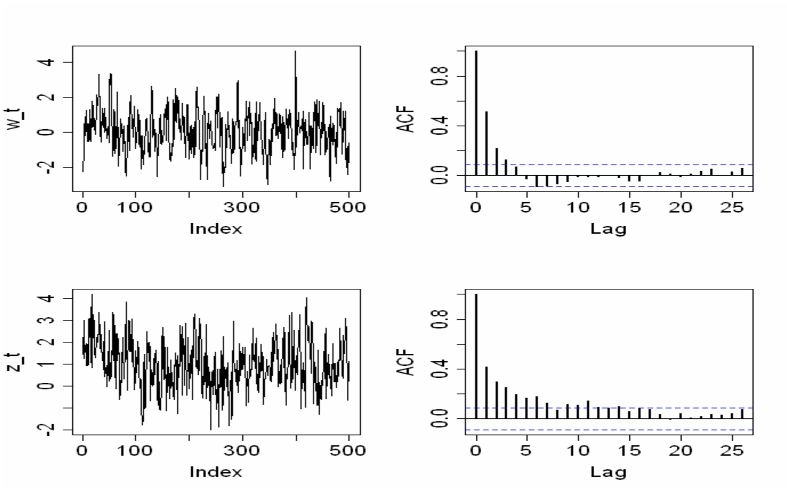
The notation '$\sim$' means, that the ratio of the left and right side tends to 1.

If we consider a stationary AR(1) process $w_t = \alpha\, w_{t-1} + \varepsilon_t$ with $|\, \alpha \,| < 1$ instead, we have

(2.5) $\quad \gamma_w(h) = \alpha^h \sigma^2 / 1 - \alpha^2 \,.$

Figure 1 (p. 4) shows sample paths and associated sample autocorrelation functions (ACF) of two processes. The first one is fractionally integrated noise with integration order $d = 0.3$, the second is a stationary AR(1) process with autocorrelation coefficient $\alpha = 0.5$. Despite the fact that both processes are stationary and approximately have the same variance, their remaining properties are quite different. According to their theoretical counterparts, the autocorrelations decline geometrically fast to zero in the AR(1) case, whereas in the fractionally integrated case the autocorrelations decrease at a slower rate. These properties are sometimes referred to as ‘short memory’ and ‘long memory’, respectively. These properties provide just another way to define I(0) and I(d) processes, respectively. Surveys on the theory and applications of long memory and related fields are provided by Beran (1994), Embrechts and Maejima (2002), Doukhan, Oppenheim and Taqqu (2003), Robinson (2003), and Samorodnitsky (2006).

Figure 1: Fractionally integrated noise and AR(1) process



Defining autoregressive and moving average lag polynomials

$$\alpha(L) := 1 - \alpha_1 L - \alpha_2 L^2 - \ldots - \alpha_p L^p ,$$

$$\beta(L) := 1 + \beta_1 L + \beta_2 L^2 + \ldots + \beta_q L^q$$

and corresponding characteristic polynomials when replacing *L* by a complex-valued variable *z*, the process w_t is a stationary and invertible ARMA(p, q) process, if

(i) $|\alpha(z)| \neq 0$, for $|z| \leq 1$,

(ii) $|\beta(z)| \neq 0$, for $|z| \leq 1$,

(iii) $\alpha(L) w_t = \beta(L) \varepsilon_t$.

Combining a stationary and invertible ARMA(p, q) process w_t with the fractional difference operator in (2.1) yields a stationary ARFIMA(p, d, q) process z_t, if $|d| < 0.5$. This parametric type of long memory model was simultaneously introduced by Hosking (1981) and Granger and Joyeux (1980). Now, defining the partial sum process Z_n by

(2.6) $$Z_n(s) := \frac{1}{n^{1/2+d}} \sum_{t=1}^{[ns]} z_t \text{ , for } 0 \le s \le 1 ,$$

it is known that

(2.7) $$Z_n \Rightarrow B_d ,$$

where B_d is fractional Brownian Motion (fBM) defined as

(2.8) $$B_d(s) = \frac{1}{\Gamma(d+1)} \int_0^s (s-t)^d \, dB_0(t) + \frac{1}{\Gamma(d+1)} \int_{-\infty}^0 \left[(s-t)^d - (-t)^d \right] dB_0(t), \text{ for } 0 \le s \le 1 .$$

The integrator process B_0 implied by (2.2) is ordinary Brownian motion extended to the real line. The process in (2.8) was introduced by Mandelbrot and Van Ness (1968) and, strictly speaking, this is fractional Brownian motion of type I, in the following denoted as type I fBM. Defining

(2.9) $$w_t^* = \mathrm{I}_{\{t>0\}} w_t$$

and z_t^* as the case corresponding to z_t in (2.1) when w_t^* replaces w_t , and Z_n^* like (2.6) with z_t^* replacing z_t , it is known that

(2.10) $$Z_n^* \Rightarrow B_d^* ,$$

where B_d^* is fractional Brownian Motion of type II (type II fBM), which is defined as

(2.11) $$B_d^*(s) = \frac{1}{\Gamma(d+1)} \int_0^s (s-t)^d \, dB_0(t), \text{ for } 0 \le s \le 1 .$$

Hence, in the type II case the second term in (2.8) is omitted. The two types of models correspond to the different cases in which pres-sample shocks are either included in the lag structure in (2.1), or suppressed. Type II fBM was introduced by Levy (1953), Mandelbrot and Van Ness (1968). The differences between type I and type II fBM are examined by Marinucci and Robinson (1999), Davidson and Hashimzade (2007a). Since the two summands in (2.8) are Gaussian and independent of each other, the variance of type I fBM exceeds that of type II fBM. Marinucci and Robinson (1999) show that the increments of type I fBM are stationary, whereas those of type II fBM are not. While type I fBM is well-defined only for $|d| < 0.5$, type II fBM is defined for all $d > -1$. Hence, only type II fBM can describe the limit behaviour of partial sums of nonstationary I(d) processes. Davydov (1970) established the first FCLT for the type I model when w_t is an i.i.d. process. This result was extended by various authors, e.g., Taqqu (1975), Gorodetskii (1977), Chan and Terrin (1995), Csőrgo and Mielniczuk (1995) to more general processes and also to the multivariate case.

Probably the most general conditions, where w_t is permitted to be near-epoch dependent on a mixing process, are provided by Davidson and de Jong (2000b), and they include stationary ARFIMA processes just as special cases. Weak convergence results corresponding to type II fBM, including nonstationary processes, were derived by Akonom and Gourieroux (1987), Silveira (1991), Marinucci and Robinson (2000). The latter paper also extends to the multivariate case.

Functional central limit theorems, both of type I and type II, combined with the continuous mapping theorem provide an important tool when deriving limiting distributions of various statistics involving nonstationary I(d) processes. Thereby, the type I framework requires to keep cumulation (integer integration) and stationary long memory in conceptually separate compartments. First, a stationary I(d) process z_t with $|d|<0.5$, for example, a stationary ARFIMA, is defined by (2.1). Then, a nonstationary I(d+1) process $\tilde{z}_t$ is defined by cumulating the process z_t, i.e.,

(2.12) $\quad \tilde{z}_t = \sum_{s=1}^{t} z_s$, or, equivalently,

(2.13) $\quad \Delta \tilde{z}_t = z_t$, for $t>0$ and $\tilde{z}_0 = 0$.

In this approach, a cumulation process $\tilde{z}_t$ must be assigned a finite start date, but its stationary increments z_t may be dependent on the remote past. Combining the FCLT for the type I case now with the continuous mapping theorem yields asymptotic results such as

(2.14) $\quad n^{-2-2d} \sum_{t=1}^{n} \tilde{z}_t^2 \xrightarrow{d} \int_0^1 B_d^2(s)\,ds$,

where '$\xrightarrow{d}$' denotes convergence in distribution. Such results are specifically useful in the context of regressions with nonstationary processes.

In the type II framework a model cannot be allowed to have an infinitely remote starting date, but must be conceived as a cumulation of increments initiated at date t = 1, with an initial condition that must be generated by a different mechanism. On the other hand, nonstationary integrated models are neatly incorporated into a general framework. As noted by Davidson (2007), there can be substantial differences between asymptotic distributions and functionals derived from them, when using either the type I or the type II approach.

Frequency domain devoted definitions of stationary I(d) processes usually refer to the spectral density of the process at frequency zero. Defining the spectral density matrix $\mathbf{f}(\lambda)$ of the p-dimensional process $\mathbf{z}(t)$ to satisfy

(2.15) $\boldsymbol{\Gamma}(m) = E\left[(\mathbf{z}_t - \boldsymbol{\mu})(\mathbf{z}_{t+m} - \boldsymbol{\mu})^T\right] = \int_{-\pi}^{\pi} \mathbf{f}(\lambda) e^{im\lambda} d\lambda$, where $\boldsymbol{\mu} = E(\mathbf{z}_t)$,

the $p \times p$ diagonal matrix

(2.16) $\boldsymbol{\Lambda}(\lambda) = Diag\left\{e^{i\pi d_1/2}\lambda^{-d_1}, \ldots, e^{i\pi d_p/2}\lambda^{-d_p}\right\}$

and its complex conjugate $\overline{\boldsymbol{\Lambda}}(\lambda)$, $\mathbf{z}(t)$ may be called $\mathrm{I}(d_1, \ldots, d_p)$ process if and only if

(2.17) $\mathbf{f}(\lambda) \sim \boldsymbol{\Lambda}(\lambda) G \overline{\boldsymbol{\Lambda}}(\lambda)$ as $\lambda \to 0+$,

where the $p \times p$ matrix G is positive definite and all its diagonal elements are nonzero. The notation '$\sim$' is here taken elementwise, to mean that the ratio of real parts and of imaginary parts of left and right sides tends to 1. For $p = 1$ this definition specializes to

(2.18) $f(\lambda) \sim G\lambda^{-2d}$ as $\lambda \to 0+$, where $G > 0$.

From (2.18) we see that $f(0)$ is finite if and only if $d \le 0$. The foregoing definition is taken from Robinson and Yajima (2002). In fact, these authors show, that by allowing the matrix G to be not of full rank, the process $\mathbf{z}(t)$ is 'fractionally cointegrated' in the sense of Engle and Granger (1987). Hence, there exists a linear combination of component processes of $\mathbf{z}(t)$ exhibiting the same integration order, say d', which yields a process with integration order $d'' < d'$. Our paper does not deal with fractional cointegration. Nevertheless, the phenomenon of fractional cointegration will arise in the context of Model (M3).

The spectral analysis of stationary stochastic processes has solid mathematical foundations, whereas this appears not to be the case for nonstationary processes. At least to our knowledge, literature lacks a rigorous definition for nonstationary I(d) processes in the frequency domain. Bujosa et. al. (2001) extend classical spectral analysis to the nonstationary case by defining pseudo-covariance generating functions and pseudo-spectra for the special case of nonstationary ARMA processes. Their approach could possibly extended to nonstationary ARFIMA processes or, more generally, to nonstationary I(d) processes.

In this paper, the analysis will be carried out solely in the discrete time domain and restricted to VARFIMA processes. Nonstationary VARFIMA processes will be defined in the type I sense. Hence, if we consider the univariate case for the moment, taking a stationary ARFIMA(p, d, q) process z_t in (2.13) yields a nonstationary ARFIMA(p, d+1, q) process $\tilde{z}_t$.

It should be noted that when taking VARFIMA processes all common I(d) definitions are met, or at least not contradicted, for both the time domain and the frequency domain. Moreover, this parametric class of processes is flexible enough to model many different kinds of time series encountered in practice.

3 Review of literature

The asymptotic properties of OLS in regression models have been analyzed for a long time under many different settings and assumptions by numerous authors. This chapter provides a rough review of the most important landmarks in literature. The objective is also to clarify to what extent I(d) processes already have been treated in literature and how our paper fits into it as a new contribution.

Probably the most important distinction has to be made between the two cases whether the regressors are assumed to be stochastic or nonstochastic.

3.1 Regression models with nonstochastic regressors

In many regression models the explanatory variables have been assumed to be nonstochastic. Under this setting, Grenander and Rosenblatt (1957) established the asymptotic estimation theory when the errors are short memory stationary processes. Their analysis is carried out in the frequency domain, i.e., the statistical properties of the error term are described via the spectral density. In this respect their paper can be regarded as a pioneer work. Their emphasis is on showing asymptotic efficiency of OLS under various settings which include, for example, trigonometric and polynomial regressors.

If z_t is a stationary ARFIMA(p, d, q) process as introduced in Chapter 2, a moving average representation of z_t can be set up which is given by

(3.1) $$z_t = \delta(L)\varepsilon_t = \sum_{j=0}^{\infty} \delta_j \varepsilon_{t-j} \,,$$

where $\delta(L) = \Delta^{-d}(L)\alpha^{-1}(L)\beta(L)$. Further, it can be shown that

(3.2) $$\sum_{j=0}^{\infty} \delta_j^2 < \infty \,.$$

Moving average processes of the form (3.1) are frequently termed either 'generalized linear processes' or 'linear processes' according to whether (3.2) holds or the stronger assumption

(3.3) $$\sum_{j=0}^{\infty} \left| \delta_j \right| < \infty \,.$$

This terminology traces back to Eicker (1965). In fact, linear regression models with nonstochastic regressors and error terms which are generalized linear processes were covered by Eicker (1967), whereas in all previous publications, e.g. Hannan (1961), the stronger assumption (3.3) was made. Thus, Eicker's paper can be regarded as an early contribution to the asymptotic theory of I(d) processes in the context of regression, though the terms ‘fractional integration’ and ‘long memory’ were not well-defined and parametric models such as ARFIMA were not known at that time. Eicker (1967) provides necessary and sufficient conditions for asymptotic normality and consistency of OLS, which include polynomial and trigonometric regressors as special cases.

Particularly strong consistency of OLS in regression models with nonstochastic regressors has been shown under various conditions on the error term by Anderson and Taylor (1976), Drygas (1976), Lai, Robbins and Wei (1979) and Solo (1981). One of the most general papers in this context, which unifies many of the foregoing results is given by Chen, Lai and Wei (1981). In fact, their conditions also cover generalized linear processes in the sense of (3.2) and hence certain types of I(d) processes. But again, neither ‘fractional integration’ nor ‘long memory’ are mentioned, let alone models such as ARFIMA.

Yajima (1988, 1991) studied the regression model with long memory errors, discussing both limit distribution theory and efficiency of OLS. He derived necessary and sufficient conditions for the least squares estimator to be asymptotically efficient relative to the best linear unbiased estimator (BLUE). This extended the work of Grenander (1954) and Grenander and Rosenblatt (1957) for the short memory case. Yajima also proved asymptotic normality of OLS for certain regressors, under conditions on the cumulants of the white noise process of the errors. One surprising result is that in the case of polynomial regression OLS is no longer asymptotically efficient, as it is in the short memory case. This parallels the results of Beran and Künsch (1985) who have pointed out that the arithmetic mean is no longer an efficient estimator for the location of a long memory process which it is for a short memory process. These results were extended to the case of weighted least squares estimation by Dahlhaus (1995) and Deo (1997). In fact, Deo (1997) proved asymptotic normality and efficiency for a weighted least squares estimator in the case of polynomial regressors.

Other related studies include research by Koul (1992), Koul and Mukherjee (1993) and Giraitis, Koul and Surgailis (1996), who consider a number of robust estimators. Giraitis, Koul, and Surgailis (1996) provide sufficient conditions for asymptotic normality of certain classes of M- and R-estimators in linear regresson models with long memory moving average errors. Their conditions also cover ARFIMA processes as special cases. It should be noted that the OLS estimator falls within the class of M-estimators.

Usami and Huzii (1995) and Usami (2002) investigate the asymptotic properties of OLS in a polynomial regression when the error is heteroskedastic and its variance is a polynomial function in time, i.e. changes deterministically with time varying. They show weak consistency and asymptotic efficiency under various assumptions on the covariance structure of the errors.

3.2 Regression models with stochastic regressors

Specifically in time series regression, regressors are usually assumed to be stochastic, and this can significantly affect the asymptotic theory. Such models are frequently called ‘stochastic regression models’, for which, as far as asymptotic properties are concerned, the most important distinction has to be made between stationary and nonstationary models.

Asymptotic results for estimating the coefficients of stationary models were first derived by Mann and Wald (1943) and Anderson (1959) for linear autoregressive models of the ARMA type. In fact, they proved consistency and asymptotic normality of OLS under various conditions. Generally, in such parametric settings some stationarity restrictions have to be put on the coefficients of the specified model.

An alternative approach is to assume that ‘classical properties’ such as stationarity, ergodicity or mixing directly hold for regressors and errors, possibly in conjunction with some moment restrictions. In such nonparametric settings OLS usually yields consistent estimates and asymptotic normality. A modern review in this context is provided by White (2001). It should be noted that long memory processes are not strong mixing as shown by Helson and Sarason (1967), for which reason I(d) processes seem not to be covered by the mixing concept.

The asymptotic distribution of OLS in stationary stochastic regression models including long memory processes was first examined by Robinson and Hidalgo (1997). They considered the model

(M4a) $y_t = \alpha + \beta_1 x_{t1} + \ldots + \beta_p x_{tp} + e_t$,

where the regressors $x_{t1}, \ldots, x_{tp}$ and the error term e_t are permitted to be long memory processes with

(3.4) $E(e_t) = 0$ and $E(x_{ti}) = 0$ for $i = 1, \ldots, p$.

In a unified manner they derived the asymptotic distribution of a generalized least squares estimator for the slope coefficient vector $\beta = (\beta_1, \ldots, \beta_p)^T$ which includes the OLS estimator and the best linear unbiased estimator as special cases. Specifically, they provide sufficient conditions for asymptotic normality. At this point it should be mentioned that their results do not apply to our results for Model (M3) by just setting $x_{t1} \equiv 1$. In fact, for Model (M3) we will later assume that one of the two regressors is nonstationary, and this yields quite different results.

Choy and Taniguchi (2001) discuss various estimates for β including OLS in a stationary model of the form (M1) under different combinations of long memory in x_t and e_t. Besides stationarity they assume

(3.5) $E(e_t) = 0$ and $E(x_t) \neq 0$.

Concretely, they examine the four different cases

(I) x_t and e_t are short memory,

(II) x_t is long memory and e_t is short memory,

(III) x_t is short memory and e_t is long memory,

(IV) x_t and e_t are long memory.

In all four cases they obtain asymptotic normality for OLS, a result that is quite different from ours as will be seen later. Specifically in case (IV), depending on the degree of long memory, or more precisely, on the degree of integration, OLS may yield non-normal distributions. This fact was pointed out by Robinson (1994) and earlier by Rosenblatt (1961) in the context of estimating the mean of a long memory process. The limiting distribution was later described first by Taqqu (1975) as the 'Rosenblatt distribution'. In fact, Choy and Taniguchi (2001) obtain normality even for case (IV), since they assume $E(x_t) \neq 0$, a seemingly harmless but critical assumption. We will assume

(3.6) $E(e_t) = 0$ and $E(x_t) = 0$

and similarly work out different cases for the integration orders of x_t and e_t. Our results in this context mostly rely on the work of Chung (2002), who analyzes the asymptotic distribution of OLS in (M4a) including the theory of Rosenblatt distributions. For that purpose, he first derives new asymptotic results for multivariate long memory processes, that generalize particularly Hosking's (1996) univariate results on sample autocovariances and Fox and Taqqu's (1987) theory on the product of two univariate long memory processes. In contrast to these earlier results, neither any stringent Gaussian assumption is made nor is it assumed that x_t and e_t are independent as in Robinson and Hidalgo (1997). It should be noted that our results for (M1) cannot be deducted from Chung´s results for (M4a). Instead we will make use of his asymptotic results on multivariate long memory processes.

Strong consistency of OLS in stochastic regression models has been examined by Anderson and Taylor (1979), Lai and Wei (1982a, 1982b) and Wei (1985). The common approach of these authors was to consider the multiple stochastic regression model

(M4b) $y_t = \beta_1 x_{t1} + \ldots + \beta_p x_{tp} + e_t$,

where e_t was assumed to be a martingale difference sequence. Defining

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & & x_{2p} \\ \vdots & \vdots & \ddots & \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix},$$

they provide sufficient conditions on the eigenvalues of the matrix

$$A = X^T X,$$

which assure strong consistency of OLS. In our paper, the question whether these results do also apply to I(d) regressors is not examined, since it is difficult to connect the asymptotic behaviour of the eigenvalues of A to the asymptotic behaviour of I(d) regressors. But there seems to be evidence that these results, at least partly, also apply to I(d) processes. In fact, Lai and Wei (1982b) provide explicit examples, where their results are applied to special autoregressive models containing unit roots, hence models with integer-valued integrated processes.

The asymptotic distribution of OLS in nonstationary stochastic regression models was first treated by White (1958, 1959) and Anderson (1959) for the AR(1) model

(M5) $y_t = \alpha\, y_{t-1} + e_t$,

where e_t was assumed to be an independent normally distributed sequence or a white noise sequence. These two authors considered the limiting distribution of OLS for the three different cases $|\alpha| < 1$, $|\alpha| = 1$ and $|\alpha| > 1$ under various conditions on the starting value y_0. In the context of unit root testing the limiting distribution has been of interest again. Investigations by Dickey (1976), Dickey and Fuller (1979, 1981), Evans and Savin (1981, 1984) have been at the forefront. But the most general and comprehensive contributions were the ones given by Phillips. First, the conditions on e_t in Model (M5) were considerably relaxed in Phillips (1987), allowing for dependent and heterogeneously distributed errors. The limiting distributions of OLS were found using functional central limit theory in conjunction with the continuous mapping theorem, a groundbreaking approach in this context. In the same way, the asymptotics for much more general nonstationary regression models were examined.

This comprises the papers by Phillips and Durlauf (1986), Phillips (1988), and Park and Phillips (1988). Probably the most general paper is that of Park and Phillips (1989), which includes all foregoing results as special cases. Related papers are that by Stock (1987), Chan and Wei (1988) and Sims, Stock and Watson (1990), which emphasize on very special models or different aspects. Specifically, Sims, Stock and Watson (1990) provide sufficient conditions for asymptotic normality of OLS in nonstationary regression models. But all of these mentioned papers are restricted to integer-valued integrated models, where the regressors are assumed to be driven by I(0), I(1) or I(2) processes, and the error term is assumed to follow an I(0) process. Andrews (1987) analyzed the asymptotics of OLS in a multiple regression with integrated regressors allowing for nonstationary errors. Maekawa et. al. (1996) showed asymptotic normality of OLS for Model (M3), when z_t is an I(1) and e_t an I(0) process. This result can be regarded as a special case of regression with cointegrated regressors, which was earlier treated by Park and Phillips (1989) in a much more general setting. In our paper this result will be generalized to the fractional case.

Sowell (1990) derived the asymptotic distribution of OLS in (M5) for the nonstationary case when $\alpha = 1$ and e_t is an I(d) process. This generalizes the results given by Phillips (1987) for the integer-valued integrated case. Using the type I approach the 'unit root distribution' is extended to a so-called 'fractional unit root distribution'. However, Sowell wrongly uses type II fBM instead of type I fBM when considering the limiting distribution of OLS, as noted by Marinucci and Robinson (1999). More generally, Chan and Terrin (1995) consider an unstable AR(p) model with I(d) errors, i.e. Model (M5) extended to p lags. In contrast to Sowell´s work, their analysis is carried out in the frequency domain and the asymptotic distribution of OLS is described via the spectral representations of various processes.

Robinson and Marinucci (2001) study the asymptotic behaviour of averaged periodograms and cross-periodograms of nonstationary I(d) processes. Their purpose is to examine the performance of the so-called 'narrow-band least squares estimator' (NBLSE) which estimates the cointegration relation between two fractionally cointegrated processes. In fact, the OLS estimator in (M1) turns out to be just a special NBLSE, and the asymptotic distributions for various cases with a nonstationary regressor and a stationary error term are derived (Propositions 6.1-6.5). Employing functional limit theory developed in Marinucci and Robinson (2000) the asymptotic distributions are nonstandard and partly functionals of type II fBM, since nonstationary processes are assumed to be generated by the type II approach. Marinucci and Robinson (2001) similarly deal with the asymptotics of OLS in the more general Model (M4a), but merely provide convergence rates. It should be noted that Theorem

6.2.2 in our paper, which applies to Model (M1), is closest to Propositions 6.1-6.5 in Robinson and Marinuncci (2001). However, in our setting the type I approach is employed to generate nonstationary processes. As a consequence our limiting distributions are different and even slight differences appear with respect to convergence rates. Further, the special case of an I(1) regressor and an I(0) error has to be examined separately in Robinson and Marinucci (2001). In contrast, the I(1)/I(0) case needs no special treatment in our framework, but fits smoothly in as a special case of fractional integration.
Dolado and Marmol (2004) derive the asymptotic distribution of OLS in reduced-form autoregressive distributive lag models whose variables include nonstationary I(d) processes of type II. They provide sufficient conditions for asymptotic normality of OLS and thereby generalize the well-known results of Sims, Stock and Watson (1990) for the integer unit root case. Though their analysis is carried out in the time domain, their results are difficult to compare to ours due to the specific form of their model.

The main objective of the subsequent chapters is to derive the asymptotic distributions of OLS in (M1)-(M3) for various integration orders of regressors and errors. The analysis will be carried out in the time domain and will use the type I approach to generate nonstationary series. To our knowledge, there is no such comprehensive work for OLS asymptotics under this setting and for these types of models at the moment. An obstacle has until recently been the absence of an asymptotic theory to describe the limiting distributions of various terms arising in the OLS estimator, specifically, sample covariances of nonstationary (regressors) and stationary (errors) processes.
For example, in Model (M1)

(3.7) $$\hat{\beta} - \beta = \frac{\sum \tilde{x}_t e_t}{\sum \tilde{x}_t^2},$$

where $\hat{\beta}$ denotes the ordinary OLS estimator, the nominator may be given by the sample covariance of a nonstationary $\mathrm{I}(\tilde{\mathrm{d}})$ regressor $\tilde{x}_t$ and a stationary I(d) error e_t. Using relations (2.12)-(2.14) when replacing $\tilde{z}_t$ by $\tilde{x}_t$, a common approach in a I(1)/I(0) environment is to split up the nominator of (3.7) according to

(3.8) $$\sum \tilde{x}_t e_t = \sum \tilde{x}_{t-1} e_t + \sum x_t e_t .$$

The asymptotics of the two right-hand side terms of (3.8) are then examined separately (see for example Park and Phillips [1988, 1989]). Thereby, the first term is more difficult to handle, since we have a product of a nonstationary and a stationary process. Working with

I(d) processes complicates matters. Recent research in this area has focused on semiparametric approaches to estimation. The analyses by Chan and Terrin (1995), Robinson and Marinucci (2001) are carried out in the frequency domain using weak convergence results of I(d) processes by applying the Wiener-Itô calculus of their spectral representations. Davidson and Hashimzade (2007b) complement the above-mentioned work by adopting the time domain setting. They consider the asymptotic distribution of the covariance of a nonstationary I(d) process with the stationary increments of another such process. In other words, a new asymptotic theory for the first right-hand side term of (3.8) is developed. We will make use of these results when extending the asymptotics of (3.8) to the fractional case.

4 Properties of VARFIMA processes

In this chapter, vector autoregressive fractionally integrated moving average processes (VARFIMA) are formally introduced along with its most important properties needed in subsequent parts. Specifically, VARFIMA processes are noted as strictly stationary and covariance ergodic processes. Moreover, crucial properties of the corresponding spectral density matrix will be stated, too.

Definition 4.1: White Noise

We call the p-dimensional process $\boldsymbol{\varepsilon}_t = (\varepsilon_{t1}, \ldots, \varepsilon_{tp})^T$ *white noise,*

if the following conditions are satisfied:

(i) $E(\boldsymbol{\varepsilon}_t) = \mathbf{0}$,

(ii) $\boldsymbol{\varepsilon}_t$ *is an i.i.d. process with*

$E|\varepsilon_{ti}\varepsilon_{tj}\varepsilon_{tk}\varepsilon_{tl}| < \infty$, *for all* $i, j, k, l = 1, \ldots, p$,

(iii) $\Sigma := Cov(\boldsymbol{\varepsilon}_t) = E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^T)$ *is nonsingular.*

We shortly denote this by $\boldsymbol{\varepsilon}_t \sim WN(\mathbf{0}, \Sigma)$.

Hence, we assume a white noise sequence to be independent and not just uncorrelated. In addition, we assume finite fourth moments. Such a sequence will be the innovation process for linear and generalized linear processes defined next.

Definition 4.2: Linear and Generalized Linear Processes

Let $\mathbf{z}_t$ *be a vector process with*

$$\mathbf{z_t} = \sum_{j=-\infty}^{\infty} D_j \, \boldsymbol{\varepsilon}_{t-j}, \text{ where } \boldsymbol{\varepsilon}_t \sim WN(\mathbf{0}, \Sigma).$$

We call $\mathbf{z}_t$

(i) a linear process, if $\sum_{j=-\infty}^{\infty} \| D_j \| < \infty$,

(ii) a generalized linear process, if $\sum_{j=-\infty}^{\infty} \| D_j \|^2 < \infty$,

where $\|\mathbf{A}\|$ *is the norm of the matrix A as given by Definition A-4.1 in the appendix.*

Obviously, any linear process is a generalized linear process but not vice versa (Lemma A-4.2 in the appendix).

Next, the notation for dealing with stationary processes is introduced. For the different concepts of stationarity, i.e., 'strict stationarity', 'mean stationarity' and 'covariance stationarity', see Definition A-4.3 in the appendix. Throughout this paper the terms 'stationary' and 'covariance stationary' will be used synonymously.

Let $\mathbf{z}_t = (z_{t1}, \ldots, z_{tp})^T$ be a p-dimensional stationary process with mean $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)^T$.
Then the autocovariance matrix function will be denoted by

$$\boldsymbol{\Gamma}(m) := E\left[(\mathbf{z}_t - \boldsymbol{\mu})(\mathbf{z}_{t+m} - \boldsymbol{\mu})^T\right], \text{ for integers } m,$$

whereas the components of $\boldsymbol{\Gamma}(m)$ are denoted by $\gamma_{ij}(m)$ with

$$\gamma_{ij}(m) := E\left[(z_{ti} - \mu_i)(z_{(t+m)j} - \mu_j)\right], \; i, j = 1, \ldots, p.$$

Note that $\boldsymbol{\Gamma}(m) = \boldsymbol{\Gamma}(-m)$, due to stationarity.
For the special case $m = 0$, we may shortly write γ_{ij} instead of $\gamma_{ij}(0)$.
The sample mean vector is given by

$$\overline{\mathbf{z}} := (\overline{z}_1, \ldots, \overline{z}_p)^T = \left(\frac{1}{n}\sum_{t=1}^{n} z_{t1}, \ldots, \frac{1}{n}\sum_{t=1}^{n} z_{t1}\right)^T$$

and the sample autocovariance matrix function by

$$\widetilde{\boldsymbol{\Gamma}}(m) := \frac{1}{n-m}\sum_{t=1}^{n-m}(\mathbf{z}_t - \boldsymbol{\mu})(\mathbf{z}_{t+m} - \boldsymbol{\mu})^T.$$

The components of $\widetilde{\Gamma}(m)$ are denoted by $\widetilde{\gamma}_{ij}(m)$, where

$$\widetilde{\gamma}_{ij}(m) := \frac{1}{n-m}\sum_{t=1}^{n-m}(z_{ti} - \mu_i)(z_{(t+m)j} - \mu_j).$$

We also define the matrix $\hat{\Gamma}(m)$ and its components $\hat{\gamma}_{ij}(m)$ by

$$\hat{\boldsymbol{\Gamma}}(m) := \frac{1}{n-m}\sum_{t=1}^{n-m}(\mathbf{z}_t - \overline{\mathbf{z}})(\mathbf{z}_{t+m} - \overline{\mathbf{z}})^T \quad \text{and} \quad \hat{\gamma}_{ij}(m) := \frac{1}{n-m}\sum_{t=1}^{n-m}(z_{ti} - \overline{z}_i)(z_{(t+m)j} - \overline{z}_j),$$

respectively.
Note that we have the approximation

$$\hat{\boldsymbol{\Gamma}}(m) \approx \frac{1}{n-m}\sum_{t=1}^{n-m}(\mathbf{z}_t - \boldsymbol{\mu})(\mathbf{z}_{t+m} - \boldsymbol{\mu})^T - (\overline{\mathbf{z}} - \boldsymbol{\mu})(\overline{\mathbf{z}} - \boldsymbol{\mu})^T = \widetilde{\boldsymbol{\Gamma}}(m) - (\overline{\mathbf{z}} - \boldsymbol{\mu})(\overline{\mathbf{z}} - \boldsymbol{\mu})^T.$$

Proposition 4.1

If $\mathbf{z}_t$ is a generalized linear process, then $\mathbf{z}_t$ is

(i) strictly stationary with finite fourth moments,

(ii) covariance stationary.

Further, we have

(iii) $\overline{\mathbf{z}} \xrightarrow{a.s.} \mathbf{0}$,

(iv) $\widetilde{\mathbf{\Gamma}}(m) \xrightarrow{a.s.} \mathbf{\Gamma}(m)$,

(v) $\hat{\mathbf{\Gamma}}(m) \xrightarrow{a.s.} \mathbf{\Gamma}(m)$.

If $\mathbf{z}_t$ is a linear process, then also

(vi) $\widetilde{\mathbf{\Gamma}}(m) \xrightarrow{m.s.} \mathbf{\Gamma}(m)$.

Proof:

For (i) see Theorem 3 of Part II in Hannan (1970). Note that any generalized linear process is generated by an i.i.d. white noise sequence with finite fourth moments. It can be easily shown then that the process itself also has finite fourth moments. For the rest, see Theorems 2 and 6 of Part II in Hannan (1970). Of course (ii) is implied by (i). Further, (v) follows from (iii) and (iv). This can be seen as follows. For ease of exposition we consider only the univariate case. On the one hand,

$$
\begin{aligned}
\hat{\gamma}(m) &= \frac{1}{n-m}\sum_{t=1}^{n-m}(z_t-\bar{z})(z_{t+m}-\bar{z}) \\
&= \frac{1}{n-m}\sum_{t=1}^{n-m} z_t z_{t+m} - \frac{1}{n-m}\sum_{t=1}^{n-m} z_t \bar{z} - \frac{1}{n-m}\sum_{t=1}^{n-m} z_{t+m}\bar{z} + \bar{z}^2 \\
&= \frac{1}{n-m}\sum_{t=1}^{n-m} z_t z_{t+m} - \frac{\bar{z}}{n-m}\left(\sum_{t=1}^{n} z_t + \sum_{t=m+1}^{n-m} z_t\right) + \bar{z}^2 \\
&= \frac{1}{n-m}\sum_{t=1}^{n-m} z_t z_{t+m} - \frac{n}{n-m}\bar{z}^2 - \frac{n-2m}{n-m}\bar{z}_{(n-2m)}\,\bar{z} + \bar{z}^2,
\end{aligned}
$$

where

$$
\bar{z}_{(n-2m)} = \frac{1}{n-2m}\sum_{t=m+1}^{n-m} z_t
$$

is the sample mean based on the central $n-2m$ observations. On the other hand,

$$
\widetilde{\gamma}(m) = \frac{1}{n-m}\sum_{t=1}^{n-m}(z_t-\mu)(z_{t+m}-\mu) = \frac{1}{n-m}\sum_{t=1}^{n-m} z_t z_{t+m},
$$

since $\mu = 0$.

Together we have

$$\hat{\gamma}(m) - \tilde{\gamma}(m) = \bar{z}^2\left(1 - \frac{n}{n-m}\right) - \frac{n-2m}{n-m}\bar{z}_{(n-2m)}\,\bar{z}\,,$$

from which follows $\hat{\gamma}(m) - \tilde{\gamma}(m) \xrightarrow{a.s.} 0$, since $\bar{z} \xrightarrow{a.s.} 0$ and $\bar{z}_{(n-2m)} \xrightarrow{a.s.} 0$.

□

Results (i)-(v) of Proposition 4.1 also hold if we assume only finite second moments for the innovation process. We just have to replace 'finite fourth moments' by 'finite second moments' in (i). Finite fourth moments are only needed for (vi) to hold.

For showing $\hat{\Gamma}(m) \xrightarrow{m.s.} \Gamma(m)$ we may not argue as simply as in the 'almost sure' case. The argumentation is more complicated and involves some knowledge about the autocovariance function of $\mathbf{z}_t$. We do not investigate further this property, since it will be sufficient for us to assume 'covariance ergodicity' defined in the following way.

Definition 4.3: Covariance Ergodicity

The vector process $\mathbf{z}_t$ is covariance ergodic, if

(i) $\mathbf{z}_t$ *is covariance stationary,*

(ii) $\bar{\mathbf{z}} \xrightarrow{p} \boldsymbol{\mu}$,

(iii) $\tilde{\mathbf{\Gamma}}(m) \xrightarrow{p} \mathbf{\Gamma}(m)$ *and* $\hat{\mathbf{\Gamma}}(m) \xrightarrow{p} \mathbf{\Gamma}(m)$.

Using this definition, the following corollary to Proposition 4.1 can be stated immediately without any further proof.

Corollary 4.1

Any generalized linear (and hence any linear) process is covariance ergodic.

Definition 4.4: VARMA process

First define vector autoregressive moving average (VARMA) filters

$A(L) := I_p - A_1L - A_2L^2 - \ldots - A_pL^a$ *and*

$B(L) := I_p + B_1L + B_2L^2 + \ldots + B_bL^b$,

which are $p \times p$ matrix polynomials in L, where L denotes the usual lag operator. Replacing L by a complex-valued variable z yields the corresponding characteristic *polynomials. We call the p-dimensional process* $\mathbf{w}_t$ *vector autoregressive moving average process of order* (a,b)*, or shortly* $VARMA(a,b)$ *process, if the following conditions are satisfied:*

(i) $A(L)\mathbf{w}_t = B(L)\boldsymbol{\varepsilon}_t$,

(ii) $\boldsymbol{\varepsilon}_t \sim WN(\mathbf{0},\Sigma)$,

(iii) $\det[A(z)] \neq 0$, *for* $|z| \leq 1$,

(iv) $\det[B(z)] \neq 0$, *for* $|z| \leq 1$.

The process $\mathbf{w}_t$ *may alternatively be represented in moving average form by*

$$\mathbf{w_t} = C(L)\boldsymbol{\varepsilon}_t, \text{ where } C(L) = A^{-1}(L)B(L).$$

Condition (iii) is called 'stability condition', and we therefore also say that $\mathbf{w}_t$ *is a stable* $VARMA(a,b)$ *process.*

For the exact meaning of $C(L) = A^{-1}(L)B(L)$, see the proof of the following proposition.

Proposition 4.2

Any $VARMA(a,b)$ *process* $\mathbf{w}_t$ *according to Definition 4.4 with*

$$A(L)\mathbf{w}_t = B(L)\boldsymbol{\varepsilon}_t$$

is a linear (and hence generalized linear) process with absolutely summable coefficient matrices, which converge geometrically fast to zero, i.e.,

$$\mathbf{w}_t = \sum_{j=0}^{\infty} C_j \boldsymbol{\varepsilon}_{t-j} \text{ with } \sum_{j=0}^{\infty} \| C_j \| < \infty \text{ and } \| C_j \| \leq K\alpha^j,$$

for some constant $K \geq 0$ *and* $\alpha \in [0,1)$. *It holds,* $C(L) = A^{-1}(L)B(L)$.

Proof: See appendix. □

For defining VARFIMA processes the fractional difference operator concept first needs to be extended to the multivariate case. A p-dimensional fractional difference operator $\boldsymbol{\Delta}^{\mathbf{d}}(L)$ with $\mathbf{d} = (d_1,\ldots,d_p)$ is a $p \times p$ diagonal matrix with $\Delta^{d_i}(L)$ on the diagonal, i.e.,

$$\boldsymbol{\Delta}^{\mathbf{d}}(L) = Diag(\Delta^{d_i}(L)),$$

where $\Delta^{d_i}(L)$ are univariate operators as defined in Chapter 2, for $i = 1,\ldots,p$.

Definition 4.5: VARFIMA process

We call the p-dimensional process $\mathbf{z}_t$ *vector autoregressive fractionally integrated moving average process of order* (a,b) *with memory parameter vector* $\mathbf{d} = (d_1,\ldots,d_p)$*, or shortly* $VARFIMA(a,\mathbf{d},b)$ *process, if the following conditions are satisfied:*

(i) $\mathbf{\Delta}^{\mathbf{d}}(L)\,\mathbf{z}_t = \mathbf{w}_t$,

(ii) $|\,d_i\,| < 0.5$*, for* $i = 1,\ldots,p$,

(iii) $\mathbf{w}_t$ *is a* $VARMA(a,b)$ *process.*

The process $\mathbf{z}_t$ *may alternatively be represented by*

$$A(L)\,\mathbf{\Delta}^{\mathbf{d}}(L)\,\mathbf{z}_\mathbf{t} = B(L)\boldsymbol{\varepsilon}_t,$$

or in moving average form by

$$\mathbf{z}_t = \mathbf{\Delta}^{-\mathbf{d}}(L)C(L)\boldsymbol{\varepsilon}_t,$$

where $C(L) = A^{-1}(L)B(L)$ *and* $\boldsymbol{\varepsilon}_t \sim WN(\mathbf{0},\Sigma)$.

Denoting the coefficients of the moving average representation by D_j *we have*

$$\mathbf{z}_\mathbf{t} = \sum_{j=0}^{\infty} D_j\ \boldsymbol{\varepsilon}_t,$$

where $D(L) = \mathbf{\Delta}^{-\mathbf{d}}(L)C(L)$.

This type of process was introduced independently by Granger and Joyeux (1980) and Hosking (1981). For the univariate ARFIMA(0, d, 0) process (fractionally integrated noise) Brockwell and Davis (1991) prove, that z_t is mean square convergent and thus well-defined. As noted by these authors, the technique of their proof can be easily extended to general ARFIMA(a, d, b) processes. It should be noted that each component process of any VARFIMA process has an ARFIMA representation (Proposition A-4.6 in the appendix), and hence $\mathbf{z}_t$ in Definition 4.5 is also well-defined in a mean square sense. A way of calculating the moving average coefficient matrices D_j, occasionally referred to as impulse responses, is suggested by Chung (2001).

Since $\Delta^0(L) = 1$, any VARMA process is a VARFIMA with $\mathbf{d} = \mathbf{0}$, at least according to our definition. Though the VARMA class is within the VARFIMA class, the two cases $d = 0$ and $|\,d\,| \neq 0$ need to be strictly distinguished when stating general properties of VARFIMA processes.

Theorem 4.3

Any VARFIMA process is a generalized linear process with hyperbolically fast declining coefficients. More precisely, let

$$\mathbf{z}_t = D(L)\boldsymbol{\varepsilon}_t = \sum_{j=0}^{\infty} D_j \boldsymbol{\varepsilon}_{t-j}$$

be the moving average representation of a VARFIMA process with

$$D(L) = \boldsymbol{\Delta}^{-\mathbf{d}}(L)C(L) \text{ and } C(L) = A^{-1}(L)B(L),$$

then

$$D_j \sim Diag\left(\Gamma(d_i)^{-1} j^{d_i - 1}\right)C(1) \text{ as } j \to \infty, \text{ and hence } \sum_{j=0}^{\infty} \left\| D_j \right\|^2 < \infty .$$

Proof: See Corollary 2 of Chung (2001). □

Hence, the moving average matrices of VARMA and VARFIMA processes behave substantially different. In the former case, these decline geometrically fast to zero, whereas in the latter case they converge to zero at a hyperbolic rate, which implies only square summability. Summarizing the foregoing results we can state the following corollary.

Corollary 4.3

Any VARFIMA process according to Definition 4.5 (and hence any VARMA process) is a generalized linear process and hence strictly stationary and covariance ergodic.

Proof:

This immediately follows from Proposition 4.1, Corollary 4.1 and Theorem 4.3. □

Theorem 4.4: Spectral density matrix for generalized linear processes

If $\mathbf{z}_t$ *is a generalized linear process, i.e.,*

$$\mathbf{z}_t = \sum_{j=-\infty}^{\infty} D_j \, \boldsymbol{\varepsilon}_{t-j}, \text{ where}$$

$$\boldsymbol{\varepsilon}_t \sim WN(\mathbf{0}, \Sigma) \text{ and } \sum_{j=-\infty}^{\infty} \left\| D_j \right\|^2 < \infty ,$$

the spectral density matrix of the process is given by

$$\mathbf{f}(\lambda) = \frac{1}{2\pi}\left(\sum_{j=-\infty}^{\infty} D_j e^{ij\lambda} \right) \Sigma \left(\sum_{j=-\infty}^{\infty} D_j^T e^{-ij\lambda} \right).$$

Proof: See Hannan (1970), p. 61.

□

In conjunction with Proposition 4.3 the following corollary can be stated immediately.

Corollary 4.4: Spectral density matrix of a VARFIMA process

If $\mathbf{z}_t$ *is a* $VARFIMA(a,\mathbf{d},b)$ *process, the spectral density matrix is given by*

$$\mathbf{f}(\lambda)=\frac{1}{2\pi}D(e^{i\lambda})\,\Sigma\,D^T(e^{-i\lambda})$$
$$=\frac{1}{2\pi}\mathbf{\Delta}^{-\mathbf{d}}(e^{i\lambda})A^{-1}(e^{i\lambda})B(e^{i\lambda})\,\Sigma\,B(e^{-i\lambda})^T A^{-1}(e^{-i\lambda})^T\mathbf{\Delta}^{-\mathbf{d}}(e^{-i\lambda}).$$

The next proposition states crucial properties of the spectral density matrix, which will be relevant to the asymptotic distribution of sample autocovariances in the next chapter.

Proposition 4.5

Let $\mathbf{f}(\lambda)$ *be the spectral density matrix of a p-dimensional VARFIMA(a,* $\mathbf{d}$*, b) process with* $\mathbf{d}=(d_1,\ldots,d_p)$*. The j-th diagonal element of* $\mathbf{f}(\lambda)$*, denoted by* $\mathbf{f}_{jj}(\lambda)$*,*

(i) is integrable, if $d_j < 0.5$*,*

(ii) is square integrable, if $d_j < 0.25$*,*

for j = 1, ..., p.

Proof:

The *j*-th diagonal element of $\mathbf{f}(\lambda)$ is given by

$$\mathbf{f}_{jj}(\lambda)=\left[\frac{1}{2\pi}D(e^{i\lambda})\Sigma\,D(e^{-i\lambda})^T\right]_{jj}$$
$$=\frac{1}{2\pi}(1-e^{i\lambda})^{-d_j}(1-e^{-i\lambda})^{-d_j}\left[A^{-1}(e^{i\lambda})B(e^{i\lambda})\,\Sigma\,B(e^{-i\lambda})^T A^{-1}(e^{-i\lambda})^T\right]_{jj}$$
$$=\frac{1}{2\pi}\left|1-e^{-i\lambda}\right|^{-2d_j}\left[A^{-1}(e^{i\lambda})B(e^{i\lambda})\,\Sigma\,B(e^{-i\lambda})^T A^{-1}(e^{-i\lambda})^T\right]_{jj},$$

where $[\,\cdot\,]_{jj}$ denotes the *j*-th diagonal element of the matrix inside the brackets.

The terms inside the brackets are representing the VARMA part of the model. From the moving average representation of this part we have

$$C(e^{i\lambda})=A^{-1}(e^{i\lambda})B(e^{i\lambda}),$$

where

$$C\left(e^{i\lambda}\right) = \sum_{j=0}^{\infty} C_j e^{i\lambda} = \sum_{j=0}^{\infty} C_j \left(\cos\lambda + i\sin\lambda\right).$$

Now we have

$$\| C\left(e^{i\lambda}\right) \| = \left\| \sum_{j=0}^{\infty} C_j \left(\cos(\lambda) + i\sin(\lambda)\right) \right\| \le \sum_{j=0}^{\infty} \| C_j \left(\cos(\lambda) + i\sin(\lambda)\right) \| \le \sum_{j=0}^{\infty} \| C_j \| < \infty$$

and analogously, $\| C\left(e^{-i\lambda}\right) \| < \infty$.

By setting

$$\widetilde{m}_j := \left[A^{-1}\left(e^{i\lambda}\right) B\left(e^{i\lambda}\right) \Sigma\, B\left(e^{-i\lambda}\right)^T A^{-1}\left(e^{-i\lambda}\right)^T\right]_{jj} = \left[C\left(e^{i\lambda}\right) \Sigma\, C^{-1}\left(e^{-i\lambda}\right)^T\right]_{jj},$$

we can conclude $\left| \widetilde{m}_j \right| < \infty$.

Note that

$$\left| 1 - e^{i\lambda} \right| = \sqrt{\left(1 - \cos\lambda\right)^2 + \left(\sin\lambda\right)^2} = 2\sqrt{\frac{1 - \cos\lambda}{2}} = \left| 2\sin\left(\frac{\lambda}{2}\right) \right|.$$

Hence we have

$$f_{jj}(\lambda) = \left|1 - e^{i\lambda}\right|^{-2d_j} \frac{\widetilde{m}_j}{2\pi}$$

with

$$f_{jj}(\lambda) \sim \lambda^{-2d_j} \frac{\widetilde{m}_j}{2\pi} \text{ as } \lambda \to 0 \text{, since } \sin\lambda \sim \lambda \text{ as } \lambda \to 0.$$

The behaviour of $f_{jj}(\lambda)$ as $\lambda \to 0$ implies integrability for $d_j < 0.5$ and square integrability for $d_j < 0.25$.

□

Appendix 4

Definition A-4.1: Matrix Norm

Let A be a matrix of possibly complex numbers. Then the norm of A we choose is defined as

$$\|A\| = \left[\lambda_{\max}\left(A^* A\right)\right]^{1/2},$$

i.e., the positive square root of the greatest eigenvalue of $\|A^* A\|$, where the star indicates transposition combined with conjugation.

Lemma A-4.1

Let $A = (a_{mn})$ be a $p \times p$ matrix of real numbers. Then

$$\|A\| \leq \left(\sum_{m=1}^{p} \sum_{n=1}^{p} a_{mn}^2 \right)^{1/2} \leq p^{1/2} \cdot \|A\|.$$

Proof:

We first note that

$$\|A\| = \left[\lambda_{\max}\left(A^T A\right)\right]^{1/2}.$$

Since the matrix $A^T A$ is nonnegative definite, all eigenvalues of $A^T A$ are real nonnegative numbers. We also have

$$\sum_{i=1}^{p} \lambda_i = tr\left(A^T A\right) = \sum_{m=1}^{p} \sum_{n=1}^{p} a_{mn}^2 ,$$

where $\lambda_1, \ldots \lambda_p$ are the eigenvalues of $A^T A$, and $tr\left(A^T A\right)$ is the trace of $A^T A$.

Hence we have

$$p^{1/2} \cdot \left[\lambda_{\max}\left(A^T A\right)\right]^{1/2} \geq \left(\sum_{m=1}^{p} \sum_{n=1}^{p} a_{mn}^2 \right)^{1/2}$$

and also

$$\left[\lambda_{\max}\left(A^T A\right)\right]^{1/2} \leq \left(\sum_{m=1}^{p} \sum_{n=1}^{p} a_{mn}^2 \right)^{1/2}.$$

□

Definition A-4.2: Absolute summability and square summability

(i) A doubly infinite sequence of real numbers $\{ a_i \}$, $i = 0, \pm 1, \pm 2, \ldots$, *is*

absolutely summable, if $\lim_{n\to\infty} \sum_{i=-n}^{n} |a_i|$ *exists and is finite, and is*

square summable, if $\lim_{n\to\infty} \sum_{i=-n}^{n} |a_i|^2$ *exists and is finite.*

The limits are denoted by $\sum_{i=-\infty}^{\infty} | a_i |$ *and* $\sum_{i=-\infty}^{\infty} | a_i |^2$, *respectively.*

To state conditions for absolute or square summability we use the notation

$\sum_{i=-\infty}^{\infty} | a_i | < \infty$ *and* $\sum_{i=-\infty}^{\infty} | a_i |^2 < \infty$, *respectively.*

(ii) A sequence of $p \times p$ *matrices* $\{ A_i = (a_{mn,i}) \}$, $i = 0, \pm 1, \pm 2, \ldots$, *is absolutely summable, if each component sequence* $\{ a_{mn,i} \}$, $m, n = 1, \ldots, p$, $i = 0, \pm 1, \pm 2, \ldots$, *is absolutely summable, or equivalently,*

(ii)' if the sequence $\{ \|A_i\| \}$ *is absolutely summable, i.e.,*

$$\sum_{i=-\infty}^{\infty} \|A_i\| < \infty .$$

(iii) A sequence of $p \times p$ *matrices* $\{ A_i = (a_{mn,i}) \}$, $i = 0, \pm 1, \pm 2, \ldots$, *is square summable, if each component sequence* $\{ a_{mn,i} \}$, $m, n = 1, \ldots, p$, $i = 0, \pm 1, \pm 2, \ldots$, *is square summable or equivalently,*

(iii)' if the sequence $\{ \|A_i\| \}$ *is square summable, i.e.,*

$$\sum_{i=-\infty}^{\infty} \|A_i\|^2 < \infty .$$

To see the equivalence of (ii) and (ii)', note that

$$|a_{mn,i}| \le \left(\sum_m \sum_n a_{mn,i}^2 \right)^{0.5} \le p^{1/2} \|A_i\|$$

according to Lemma A-4.1. Hence,

$$\sum_{i=-\infty}^{\infty} | a_{mn,i} | < \infty \text{ if } \sum_{i=-\infty}^{\infty} \|A_i\| < \infty .$$

In turn, we also have

$$\|A_i\| \le \left(\sum_m \sum_n a_{mn,i}^2 \right)^{0.5} \le \sum_m \sum_n \left| a_{mn,i} \right|.$$

Hence, if

$$\sum_{i=-\infty}^{\infty} \left| a_{mn,i} \right| < \infty \text{, for all } m,n = 1,\ldots,p \text{, then}$$

$$\sum_{i=-\infty}^{\infty} \sum_m \sum_n \left| a_{mn,i} \right| < \infty \text{ and hence, } \sum_{i=-\infty}^{\infty} \|A_i\| < \infty .$$

To see the equivalence of (iii) and (iii)', similarly note that

$$a_{mn,i}^2 \le \sum_m \sum_n a_{mn,i}^2 \le p \cdot \|A_i\|^2 .$$

Hence,

$$\sum_{i=-\infty}^{\infty} a_{mn,i}^2 < \infty \text{ if } \sum_{i=-\infty}^{\infty} \|A_i\|^2 < \infty .$$

Again, we also have

$$\|A_i\|^2 \le \sum_m \sum_n a_{mn,i}^2 \le a_{mn,i}^2 .$$

Hence, if

$$\sum_{i=-\infty}^{\infty} a_{mn,i}^2 < \infty \text{, for all } m,n = 1,\ldots,p \text{, then}$$

$$\sum_{i=-\infty}^{\infty} \sum_m \sum_n a_{mn,i}^2 < \infty \text{ and hence, } \sum_{i=-\infty}^{\infty} \|A_i\|^2 < \infty .$$

We further note that square summability is implied by absolute summability, since

$$\|A_i\|^2 \le \left(\sum_m \sum_n a_{mn,i}^2 \right) \le \left(\sum_m \sum_n \left| a_{mn,i} \right| \right)^2 .$$

This is formally stated in the next lemma.

Lemma A-4.2

Let $\{ A_i = (a_{mn,i}) \}$, $i = 0, \pm 1, \pm 2, \ldots,$ *be a sequence of* $p \times p$ *matrices.*

If $\sum_{i=-\infty}^{\infty} \| A_i \| < \infty$, *then* $\sum_{i=-\infty}^{\infty} \| A_i \|^2 < \infty$.

Definition A-4.3: Stationarity

Let $\mathbf{z}_t = (z_{t1}, \ldots, z_{tp})^T$ *be a p-dimensional process.*

a) We call $\mathbf{z}_t$ *strictly stationary, if the joint distributions of*

$$\mathbf{z}_{t_1}, \ldots, \mathbf{z}_{t_k} \text{ and } \mathbf{z}_{t_{1+m}}, \ldots, \mathbf{z}_{t_{k+m}}$$

are the same for all $t_1, \ldots, t_k$*, all integers* m *and all positive integers* k.

b) Assume that the second moments of $\mathbf{z}_t$ *are finite, and let the mean function of* $\mathbf{z}_t$ *be defined by*

$$\boldsymbol{\mu}(t) := [\, E(z_{t1}), \ldots, E(z_{tp}) \,]^T$$

and the autocovariance matrix function by

$$\breve{\boldsymbol{\Gamma}}(t,s) := E\left[(\mathbf{z}_t - \boldsymbol{\mu}(t))(\mathbf{z}_s - \boldsymbol{\mu}(s))^T \right].$$

Then we call $\mathbf{z}_t$

(i) mean stationary, if $\boldsymbol{\mu}(t) = \boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)^T$*, for all* t*, and*

(ii) covariance stationary, if $\mathbf{z}_t$ *is mean stationary and* $\breve{\boldsymbol{\Gamma}}(t,s)$ *is a function only of* $|\, t - s|$.

Definition A-4.4: VAR(1) process

We call the p-dimensional process $\mathbf{w}_t$ *vector autoregressive process of order 1, or shortly* $VAR(1)$ *process, if the following conditions are satisfied:*

(i) $\mathbf{w}_t = A\mathbf{w}_{t-1} + \boldsymbol{\varepsilon}_t$,

(ii) $\boldsymbol{\varepsilon}_t \sim WN(\mathbf{0}, \Sigma)$,

(iii) $\det[I_p - Az] \neq 0$ *for* $|z| \leq 1$,

where I_p *denotes the p-dimensional identity matrix.*

Condition (iii) is called 'stability condition' and therefore we also say that $\mathbf{w}_t$ *is a stable VAR(1) process.*

Note that condition (iii) is equivalent to the condition that all eigenvalues of A have modulus less than 1.

Proposition A-4.3

Let $\mathbf{w}_t = A\mathbf{w}_{t-1} + \boldsymbol{\varepsilon}_t$ *be a VAR(1) process according to Definition A-4.4.*

The infinite sum

$$\sum_{j=1}^{\infty} A^j \boldsymbol{\varepsilon}_{t-j}$$

exists in mean square, and $\mathbf{w}_t$ *is the well-defined linear process*

$$\mathbf{w}_t = \sum_{j=0}^{\infty} A^j \boldsymbol{\varepsilon}_{t-j} \ .$$

We call the last expression the moving average representation of $\mathbf{w}_t$.

It further holds, that

$$\sum_{j=0}^{\infty} \left\| A^j \right\| < \infty, \ \sum_{j=0}^{\infty} \left\| A^j \right\| = \left(I_p - A \right)^{-1} \text{ and } \left\| A^j \right\| \le K\alpha^j$$

for some constant $K \ge 0$ *and* $\alpha \in [0,1)$*, i.e , the sequence* $\{A^j\}$ *is absolutely summable and converges geometrically fast to zero.*

Proof:

See Lütkepohl (2005), p. 14.

Further, note that for any $p \times p$ matrix A with n distinct eigenvalues $\lambda_1, \ldots, \lambda_n$ (possibly complex-valued) there exists a nonsingular matrix P (possibly complex-valued) such that

$$P^{-1}AP = \begin{bmatrix} \Lambda_1 & & 0 \\ & \ddots & \\ 0 & & \Lambda_n \end{bmatrix} =: \Lambda \quad \text{or} \quad A = P\Lambda P^{-1} \text{, where}$$

$$\Lambda_i = \begin{bmatrix} \lambda_i & 1 & 0 & \ldots & 0 \\ 0 & \lambda_i & 1 & & 0 \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & 0 & & \ddots & 1 \\ 0 & 0 & \ldots & \ldots & \lambda_i \end{bmatrix} \text{ is a } r_i \times r_i \text{ matrix , } i = 1, \ldots, n \, .$$

Thereby, r_i denotes the multiplicity of λ_i, $i = 1, \ldots, n$. This decomposition of A is called the ‘Jordan canonical form’. The Jordan canonical form implies that

(4.1) $\quad A^j = \left(P^{-1} \Lambda P \right)^j = P \Lambda^j P^{-1},$

and it can be shown that

$$\Lambda_i = \begin{bmatrix} \lambda_i^j & \binom{j}{1}\lambda_i^{j-1} & \dots & \binom{j}{r_i-1}\lambda_i^{j-r_i+1} \\ 0 & \lambda_i^j & \dots & \binom{j}{r_i-2}\lambda_i^{j-r_i+2} \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_i^j \end{bmatrix},$$

where

$$\binom{p}{q} = \frac{p!}{(p-q)!q!}$$

denotes a binomial coefficient. Hence, we can easily deduct from (4.1) that A^j converges geometrically fast to zero, since all eigenvalues of A have modulus less than 1

□

Proof of Proposition 4.2

Any $\mathrm{VARMA}(a,b)$ process according to Definition 4.4 has a VAR(1) representation according to Definition A-4.4. To see this, let $A(L)\mathbf{w}_t = B(L)\boldsymbol{\varepsilon}_t$ be a $\mathrm{VARMA}(a,b)$ process. Define

$$W_t := \begin{pmatrix} \mathbf{w}_t \\ \vdots \\ \mathbf{w}_{t-a+1} \\ \boldsymbol{\varepsilon}_t \\ \vdots \\ \boldsymbol{\varepsilon}_{t-b+1} \end{pmatrix}, \qquad E_t := \begin{pmatrix} \boldsymbol{\varepsilon}_t \\ 0 \\ \vdots \\ 0 \\ \boldsymbol{\varepsilon}_t \\ 0 \\ \vdots \\ 0 \end{pmatrix} \begin{matrix} \left.\vphantom{\begin{matrix}a\\a\\a\\a\end{matrix}}\right\} pa \times 1 \\ \left.\vphantom{\begin{matrix}a\\a\\a\\a\end{matrix}}\right\} pb \times 1 \end{matrix},$$

$$p(a+b)\times 1$$

and

$$\mathbf{A} := \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \quad p(a+b)\times p(a+b), \text{ where}$$

$$\mathbf{A}_{11} := \begin{bmatrix} A_1 & \dots & A_{a-1} & A_a \\ I_a & & 0 & 0 \\ & \ddots & & \\ 0 & \dots & I_a & 0 \end{bmatrix}, \qquad \mathbf{A}_{12} := \begin{bmatrix} M_1 & \dots & M_{b-1} & M_b \\ 0 & \dots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & 0 & 0 \end{bmatrix},$$

$$pa \times pa \qquad\qquad pa \times pb$$

$$\mathbf{A}_{21} := 0 \quad \text{and} \quad \mathbf{A}_{22} := \begin{bmatrix} 0 & \dots & 0 & 0 \\ I_p & & 0 & 0 \\ & \ddots & & \vdots \\ 0 & \dots & I_p & 0 \end{bmatrix} .$$

$pb \times pa$ $\qquad\qquad$ $pb \times pb$

With this notation, the VAR(1) representation is given by

(4.2) $\quad W_t = \mathbf{A} W_{t-1} + E_t$.

If $a = 0$, we choose $a = 1$ and set $A_1 = 0$ in this representation.

It should be mentioned that the process W_t by construction is not a VAR(1) process according to Definition A-4.4 in a strict sense, since the covariance matrix of E_t in (4.2) is singular. However, the following considerations remain true in spite of this fact. The next lemma states the relationship between W_t and $\mathbf{w}_t$ concerning the stability of the process.

Lemma A-4.4

The p(a+b)-dimensional VAR(1) process W_t, which corresponds to the VAR(1) representation (4.2) of a p-dimensional VARMA(a,b) process $\mathbf{w}_t$, is stable if and only if $\mathbf{w}_t$ is stable. In other words,

$$\det\left[I_{p(a+b)} - \mathbf{A}\, z\right] \neq 0 \text{ for } |z| \leq 1 \quad \Leftrightarrow \quad \det[A(z)] = \det\left[I_p - A_1 z - \ldots - A_a z^a\right] \neq 0 \text{ for } |z| \leq 1.$$

Proof:

This follows by simple rules for the determinant of a partitioned matrix.

□

We know from Proposition A-4.3, if $\mathbf{w}_t$ and hence W_t is stable, the latter process has an moving average representation

(4.3) $\quad W_t = \sum_{j=0}^{\infty} \mathbf{A}^j E_{t-j}$.

Premultiplying (4.3) by the $p \times p(a+b)$ matrix $J := \left[I_p : 0 : \cdots : 0\right]$ yields

$$\mathbf{w}_t = \sum_{j=0}^{\infty} J\mathbf{A}^j E_{t-j} = \sum_{j=0}^{\infty} J\mathbf{A}^j HJE_{t-j} = \sum_{j=0}^{\infty} J\mathbf{A}^j H\boldsymbol{\varepsilon}_{t-j} = \sum_{j=0}^{\infty} C_j \boldsymbol{\varepsilon}_{t-j} ,$$

where

$$H:=\begin{pmatrix} I_p \\ 0 \\ \vdots \\ 0 \\ I_p \\ 0 \\ \vdots \\ 0 \end{pmatrix} \begin{matrix} \left.\vphantom{\begin{matrix}I_p\\0\\\vdots\\0\end{matrix}}\right\} pa\times p \\ \left.\vphantom{\begin{matrix}I_p\\0\\\vdots\\0\end{matrix}}\right\} pb\times p \end{matrix}$$

and

(4.4) $\quad C_j = J\mathbf{A}^j H$.

Hence, from (4.3), (4.4) and Proposition A-4.3 it follows that VARMA processes are linear and hence generalized linear processes with absolutely summable coefficient matrices, which converge geometrically fast to zero.

The stability condition (iii) in Definition 4.4 implies that the inverse filter of $A(L)$ denoted by $A^{-1}(L)$ exists. It can be shown that computing the moving average coefficient matrices by (4.4) is just another way of computing the coefficient matrices of the power series

$$\sum_{j=0}^{\infty} C_j L^j = \left(I_p - A_1 L - \ldots - A_a L^a\right)^{-1}\left(I_p + M_1 L + \ldots + M_b L^b\right).$$

Hence, $\mathbf{w}_t$ is a stable process which alternatively can be expressed in lag operator notation by

$$\mathbf{w}_t = C(L)\boldsymbol{\varepsilon}_t = A^{-1}(L)B(L)\boldsymbol{\varepsilon}_t .$$

This completes the proof of Proposition 4.2.

□

Note that, due to condition (iv) of Definition 4.4, $\mathbf{w}_t$ is also an invertible process, which alternatively can be expressed in pure autoregressive form by

$$B^{-1}(L)A(L)\mathbf{w}_t = \boldsymbol{\varepsilon}_t .$$

The way of computing the coefficients of the power series $B^{-1}(L)A(L)$ is quite similar to the moving average case (see Lütkepohl (2005), p. 423-429 for further details).

Proposition A-4.5: Linear transformations of VARMA processes

Let $\mathbf{w}_t$ be a p-dimensional VARMA(a,b) process and let F be an ($m\times p$) matrix of rank m. Then the process $\tilde{\mathbf{w}}_t = F\,\mathbf{w}_t$ has a VARMA$(\tilde{a},\tilde{b})$ representation with

$\tilde{a} \le pa$ and $\tilde{b} \le (p-1)a + b$.

Proof: See Lütkepohl (2005), p. 436.

□

Hence, the VARMA class is closed with respect to linear transformations and hence, each individual component processes of $\mathbf{w}_t$ has a finite order ARMA representation. This is stated in the next corollary.

Corollary A-4.5

Let $\mathbf{w}_t = (w_{t1},\ldots,w_{tp})^T$ be a p-dimensional VARMA(a,b) process, then the i-th component process w_{ti} has an ARMA$(\widetilde{a}_i,\widetilde{b}_i)$ representation with $\widetilde{a}_i \le pa$ and $\widetilde{b}_i \le (p-1)a+b$, for all $i=1,\ldots,p$.

Proposition A-4.6

Let $\mathbf{z}_t = (z_{t1},\ldots,z_{tp})^T$ be a p-dimensional VARFIMA$(a,\mathbf{d},b)$ process according to Definition 4.5, then the i-th component process z_{ti} has an ARFIMA$(\widetilde{a}_i,d_i,\widetilde{b}_i)$ representation, i.e., $\Delta^{d_i}(L)\,z_{ti} = w_{ti}$, where w_{ti} is an ARMA$(\widetilde{a}_i,\widetilde{b}_i)$ process with $\widetilde{a}_i \le pa$ and $\widetilde{b}_i \le (p-1)a+b$, for $i=1,\ldots,p$.

Proof:

This immediately follows from Definition 4.5 and Corollary A-4.5.

□

5 Asymptotics for fractionally integrated processes

This chapter deals with the asymptotics of sample covariances of I(d) processes. For this purpose a couple of theorems are provided and specialized to VARFIMA type processes. Sample covariances constitute the ingredient parts in OLS regression formulas. The limiting distributions are not necessarily normal but may be expressed by stochastic integrals. Therefore, the notion of fractional Brownian motion has to be introduced. Using notation from Chapter 4 the first theorem stated traces back to Hannan (1976).

Theorem 5.1

Let $\mathbf{z}_t$ *be a p-dimensional generalized linear process according to Definition 4.2 and let* $\mathbf{f}_{ij}(\lambda)$, *for i, j = 1, ..., p, be the indicated elements of the spectral density matrix* $\mathbf{f}(\lambda)$ *given in Theorem 4.4. Now define*

$$\hat{\tau}_{ij}(m) := \sqrt{n}\left(\hat{\gamma}_{ij}(m) - \gamma_{ij}(m)\right), \text{ where}$$

$$\gamma_{ij}(m) = E\left(z_{ti} z_{(t+m)j}\right) \text{ and } \hat{\gamma}_{ij}(m) = \frac{1}{n-m}\sum_{t=1}^{n-m}\left(z_{ti} - \bar{z}_i\right)\left(z_{(t+m)j} - \bar{z}_j\right).$$

A necessary and sufficient condition that any finite set of the $\hat{\tau}_{ij}(m)$ *be jointly asymptotically normal is that the* $\mathbf{f}_{ii}(\lambda)$ *are all square integrable, for* $i = 1, \ldots, p$.

The covariance between $\hat{\tau}_{ij}(m_1)$ *and* $\hat{\tau}_{kl}(m_2)$ *in their limiting distribution is given by*

$$2\pi \int_{-\pi}^{\pi} \left\{ f_{ik}(\lambda)\overline{f_{jl}(\lambda)} e^{-i^*(m_2 - m_1)\lambda} + f_{il}(\lambda)\overline{f_{jk}(\lambda)} e^{i^*(m_2 + m_1)\lambda} \right\} d\lambda$$
$$+ \sum_{a=1}^{p}\sum_{b=1}^{p}\sum_{c=1}^{p}\sum_{d=1}^{p} \kappa_{abcd} \frac{1}{2\pi} \int_{-\pi}^{\pi} \left\{ \left[D\left(e^{i^*\lambda}\right)\right]_{ia} \left[D\left(e^{-i^*\lambda}\right)\right]_{jb} e^{i^* m_1 \lambda} + \left[D\left(e^{-i^*\lambda}\right)\right]_{kc} \left[D\left(e^{i^*\lambda}\right)\right]_{ld} e^{-i^* m_2 \lambda} \right\} d\lambda,$$

where $D\left(e^z\right) = \sum_{j=0}^{\infty} D_j e^{zj}$,

$i^* = \sqrt{-1}$ *and* κ_{abcd} *denotes the fourth cumulant between* ε_{ta}, ε_{tb}, ε_{tc}, ε_{td}.

Proof: See Hannan (1976).

□

According to Proposition 4.5, the diagonal elements of the spectral density matrix of VARFIMA processes are square integrable, if the integration order of each component process is less than 0.25. Hence, we get the following result.

Corollary 5.1

Let $\mathbf{z}_t$ *be a p-dimensional VARFIMA* $(a,\mathbf{d},b)$ *process according to Definition 4.5.*

Then any finite set of the $\hat{\tau}_{kl}(m)$*, for k, l = 1, ..., p, is jointly asymptotically normal, if* $d_i < 0.25$*, for i = 1, ..., p.*

The same holds if $\hat{\tau}_{kl}(m)$ *is replaced by* $\tilde{\tau}_{kl}(m) := \sqrt{n}(\tilde{\gamma}_{kl}(m) - \gamma_{kl}(m))$.

For the last result note that $\hat{\tau}_{kl}(m) - \tilde{\tau}_{kl}(m) = o_p(1)$ (see proof of Proposition 4.1).

Next, we define fractional Brownian motion (fBM). As noted in Chapter 2, there are two different definitions in literature being known respectively as type I and type II fBM. The difference between these processes is substantial, and we emphasize that all our results are merely based on the type I definition. This is due to the fact that we will base our analysis partly on results proved by Davidson and Hashimzade (2007b) for nonstationary I(d) processes which are based on the type I approach.

Definition 5.1: Fractional Brownian Motion (Type I fBM)

The p-dimensional fractional Brownian motion $\mathbf{B_d}(s)$ *with memory parameter vector* $\mathbf{d} = (d_1, \ldots, d_p)$ *and covariance matrix* $\mathbf{\Psi}$ *is defined as*

$$\mathbf{B_d}(s) = \left[B_{d_1}(s), \ldots, B_{d_p}(s)\right]^T.$$

Thereby, each $B_{d_i}(s)$ *is a one-dimensional fractional Brownian motion with memory parameter* d_i *defined as*

$$B_{d_i}(s) = \frac{1}{\Gamma(d_i+1)}\left\{\int_0^s (s-t)^{d_i}\, dB_0^{(i)}(t) + \int_{-\infty}^0 \left[(s-t)^{d_i} - (-t)^{d_i}\right] dB_0^{(i)}(t)\right\}, \ for\ 0 \le s \le 1,$$

where $B_0^{(i)}(t)$ *is the i-th element of a p-dimensional Brownian motion with covariance matrix* $\mathbf{\Omega}$ *(Definition A-5.2 in the appendix).*

Letting ϖ_{ij} *denote the* (i,j)*th element of* $\mathbf{\Omega}$*, the elements of the matrix* $\mathbf{\Psi}$ *are defined by*

$$\psi_{ij} = \frac{\varpi_{ij}}{\Gamma(d_i+1)\ \Gamma(d_j+1)}\left\{\frac{1}{1+d_i+d_j} + \int_0^\infty \left[(1+\tau)^{d_i} - \tau^{d_i}\right]\left[(1+\tau)^{d_j} - \tau^{d_j}\right] d\tau\right\}$$

for $i, j = 1, \ldots, p$.

The univariate fractional Brownian motion $B_{d_i}(s)$ was first proposed by Mandelbrot and Van Ness (1968), from whose work it can be derived that the expectation of $B_{d_i}(s)$ is zero.

The covariance between $B_{d_i}(s)$ and $B_{d_j}(s)$ is ψ_{ij}, i.e.,

$$E\left(\mathbf{B_d}(1)\mathbf{B_d}(1)^T\right) = \mathbf{\Psi}.$$

Hence, the variance of the *i*-th element of $\mathbf{B_d}(s)$ is given by

$$\psi_{ii} = \frac{\varpi_{ii}}{\Gamma(d_i+1)^2}\left\{\frac{1}{1+2d_i} + \int_0^\infty \left((1+\tau)^{d_i} - \tau^{d_i}\right)^2 d\tau\right\}.$$

A proof of the closed-form representation

$$\psi_{ii} = \frac{\varpi_{ii}\,\Gamma(1-2d_i)}{(2d_i+1)\Gamma(1+d_i)\Gamma(1-d_i)}$$

is given in Davidson and Hashimzade (2008).

The next theorem parallels Corollary 5.1 for the case when the integration order of the underlying process is strictly between 0.25 and 0.5. We note that in this case the corresponding spectral densities are not square integrable any more.

Theorem 5.2 (Chung, 2002)

Let $\mathbf{z}_t$ *be a p-dimensional VARFIMA* $(a,\mathbf{d},b)$ *process according to Definition 4.5*

We additionally assume $E\,|\,\varepsilon_{ti}\,|^{4+\eta} < \infty$, *for i = 1, ..., p, and some* $\eta > 0$.

If $0.25 < d_i < 0.5$, *for all* $i = 1,\dots,p$, *then for any finite integer m we have*

$$(5.1)\qquad Diag\left(n^{-d_i}\right)\left\{\sum_{t=1}^{[ns]-m}\left[\mathbf{z}_t\mathbf{z}_{t+m}^T - \mathbf{\Gamma}(m)\right]\right\}Diag\left(n^{-d_i}\right) \Rightarrow \mathbf{R}(s),\ \ for\ 0 \le s \le 1,$$

where $\mathbf{\Gamma}(m)$ *are population autocovariances and* $\mathbf{R}(s)$ *is a* $p\times p$ *matrix of processes of which the* (k,l)*th element is defined by*

$$R^{(kl)}(s) = \frac{1}{\Gamma(d_k)\Gamma(d_l)}\int_{-\infty}^{s}\int_{-\infty}^{s}\left[\int_0^s (\tau-r_k)_+^{d_k-1}(\tau-r_l)_+^{d_l-1}\,d\tau\right]dB_0^{(k)}(r_k)\,dB_0^{(l)}(r_l)$$

Here $(z)_+^\delta = 0$ *if* $z \le 0$, *and* $(z)_+^\delta \equiv z^\delta$ *if* $z > 0$, *and* $B_0^{(k)}(r)$ *is the k-th element of a p-dimensional Brownian motion with the variance-covariance matrix* $\mathbf{\Omega} = C(1)\Sigma\, C(1)^T$.

Proof:

See Chung (2002), Theorem 2. □

Remarkably, the limit does not depend on m. Unlike fractional Brownian motion, $R^{(kl)}(s)$ is not Gaussian. Fox and Taqqu (1987) and Hosking (1996) have derived cumulants for the univariate version of $R^{(kl)}(s)$ with $k \neq l$ and with $k = l$, respectively. In particular, their results imply that $E[R^{(kl)}(s)] = 0$, for all k and l, from which it can be concluded that the mean of the limiting process $\mathbf{R}(s)$ in Theorem 5.2 is also zero. In addition, Tudor (2007) analyzes the univariate Rosenblatt process from a stochastic calculus point of view.

From Corollary 5.1 and Theorem 5.2 we conclude that the convergence rate of the sample variance grows toward the standard $n^{-1/2}$ rate as the values of d_i decline from 0.5 to 0.25. When the values of d_i pass 0.25, then the convergence rate of the sample variance stops changing and settles at the $n^{-1/2}$ rate, whereas the limiting distribution changes from a nonnormal distribution to a normal one. The case where $d_i = 0.25$ is not available under the present assumptions. Hosking (1996) has derived a result for the univariate case under the Gaussian innovations assumption that $\sqrt{n/\ln n}\,[\hat{\gamma}_{kk}(m) - \gamma_{kk}(m)]$ will have a Gaussian limiting distribution given $d_i = 0.25$. Note that Corollary 5.1 and Theorem 5.2 do not cover any case. For example, if we assume the integration order of one process to be less than 0.25 and the integration order of another process to be greater than 0.25, then there seems to be no distributional result in literature yet.

The next theorems will enable us to deal with the asymptotics of nonstationary I(d) processes. Common tools in this context are functional central limit theorems (FCLTs) and the continuous mapping theorem. Together they provide a way to represent limiting distributions as functionals of fractional Brownian motions.

First, an FCLT proved by Davidson and de Jong (2000b) is stated, which is then specialized to VARFIMA type processes. Then, we apply the continuous mapping theorem to derive the limiting distribution of the sample variance when appropriately standardized.

Theorem 5.3 (FCLT for I(d) processes)

Assumption 1:

The sequence w_t

(i) has zero mean,

(ii) is uniformly L_r*-bounded for* $r > 2$,

(iii) is L_2*-NED of size -0.5 on* ε_t *with* $d_t = 1$, *where* ε_t *is either an* α*-mixing*

sequence of size $-r/(r-2)$, or a ϕ-mixing sequence of size $-r/(2(r-1))$,

(iv) is covariance stationary with

$$0<\sigma_w^2=\lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\sum_{s=1}^{n}E(w_t w_s)<\infty .$$

Assumption 2:

Each component of the p-dimensional process $\mathbf{w}_t=(w_{t1},\ldots,w_{tp})^T$ satisfies Assumption 1, and

$$\mathbf{\Omega}=\lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\sum_{s=1}^{n}E(\mathbf{w}_t\mathbf{w}_s^T)$$

is finite and positive definite.

Define the process

$\mathbf{z}_t=\mathbf{\Delta}^{-\mathbf{d}}(L)\mathbf{w}_t$, *where* $|d_i|<0.5$, *for* $i=1,\ldots,p$, *and*

$$Z_n(s):=\sum_{t=1}^{[ns]}Z_{nt}\text{, where } Z_{nt}:=\hat{D}_n^{-1}\mathbf{z}_t \text{ and } \hat{D}_n:=Diag(n^{d_i+0.5}).$$

Then,

$$Z_n\Rightarrow \mathbf{B_d},$$

where $\mathbf{B_d}$ is a p-dimensional fractional Brownian motion as in Definition 5.1.

Proof: See Davidson and de Jong (2000b) Theorem 3.2.

□

For the concept of near-epoch dependence see Definition A-5.4 in the appendix.
We next specialize Theorem 5.3 to VARFIMA processes.

Corollary 5.3: FCLT for VARFIMA processes

Let $\mathbf{z}_t$ be a p-dimensional VARFIMA process according to Definition 4.5 with

(5.2) $\mathbf{\Delta}^{\mathbf{d}}(L)\mathbf{z}_t=C(L)\boldsymbol{\varepsilon}_t,$

then

(5.3) $Diag(n^{-0.5-d_i})\sum_{t=1}^{[ns]}\mathbf{z}_t \Rightarrow \mathbf{B_d}(s)$, *for* $0\le s\le 1$,

where $\mathbf{B_d}(s)$ is a p-dimensional fractional Brownian process as in Definition 5.1 with $\mathbf{\Omega}=C(1)\Sigma C(1)^T$.

Proof:

All assumptions of Theorem 5.3 are fulfilled. This follows from Proposition A-5.4 in the appendix when choosing $\mathbf{w}_t = C(L)\varepsilon_t$.

Note that the process $\mathbf{w}_t$ is a stationary and invertible VARMA process with finite fourth moments, which is driven by an i.i.d. and hence mixing white noise sequence $\boldsymbol{\varepsilon}_t$. Further, each component process of $\mathbf{z}_t$ has a finite ARFIMA representation (Proposition A-4.6 in the appendix), and hence the conditions in a) hold componentwise. Since the covariance matrix Σ of $\boldsymbol{\varepsilon}_t$ is assumed to be finite and positive definite ,

$$\boldsymbol{\Omega} = C(1)\Sigma C(1)^T$$

is also finite and positive definite.

□

A useful result for the sample variance of a nonstationary ARFIMA process is the following.

Proposition 5.4

Let z_t be a one-dimensional VARFIMA process. Then,

$$n^{-2-2d}\sum_{t=1}^{n} x_t^2 \xrightarrow{d} \int_0^1 B_d^2(s)\,ds\,,$$

where $B_d^2(s)$ is as in (5.3) in Corollary 5.3, $\Delta x_t = z_t$, for $t > 0$ and $x_0 = 0$.

Proof: See appendix. □

Next, new asymptotic results derived by Davidson and Hashimzade (2007b) for the covariances of a nonstationary I(d) process with the stationary increments of another such process will be stated. We will then specialize these results to VARFIMA processes. The specific assumptions to be adopted are as follows.

Assumption 1

$\boldsymbol{\varepsilon}_t = (\varepsilon_{t1}, \varepsilon_{t2})^T$ is i.i.d. with zero mean and covariance matrix

$$E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^T) = \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} \text{ and } E(\varepsilon_{t1}^2\varepsilon_{t2}^2) < \infty.$$

$\varepsilon_{t1} = \varepsilon_{t2}$ is an admissible case.

Assumption 2

$z_{t1} = \sum_{j=0}^{\infty} D_{1j}\varepsilon_{(t-j)1}$ *and* $z_{t2} = \sum_{j=0}^{\infty} D_{2j}\varepsilon_{(t-j)2}$, *where the sequences* $\{D_{1j}\}$ *and* $\{D_{2j}\}$ *depend on parameters* $d_1 \in (-0.5, 0.5)$ *and* $d_2 \in (-0.5, 0.5)$, *respectively, and sequences* $\{L_1(j)\}$ *and* $\{L_2(j)\}$ *that are slowly varying at infinity. These sequences satisfy one of the following conditions, stated for* $\{D_{1j}\}$ *as representative case:*

(a) If $0 < d_1 < 0.5$, *then* $D_{1j} = \Gamma(d_1)^{-1}(j+1)^{d_1-1}L_1(j)$.

(b) If $d_1 = 0$, *then* $0 < \left|\sum_{j=0}^{\infty} D_{1j}\right| < \infty$, *and* $D_{1j} = O\left(j^{-1-\eta}\right)$ *for* $\eta > 0$.

(c) If $-0.5 < d_1 < 0$, *then* $D_{10} = \delta_0$ *and* $D_{1j} = \delta_j - \delta_{j-1}$ *for* $j > 0$, *where*

$$\delta_j = \Gamma(1+d_1)^{-1}(j+1)^{d_1}L_1(j).$$

Note, a sequence $L(j)$ is said to be slowly varying at infinity, if it satisfies

$L(\alpha\, j)/L(j) \to 1$ as $j \to \infty$, for any $\alpha > 0$.

Now define

(5.4) $\quad G_n = \frac{1}{K(n)}\sum_{t=1}^{n-1}\sum_{s=1}^{t} z_{s2}z_{(t+1)1}$, where $K(n) = n^{1+d_1+d_2}L_1(n)L_2(n)$.

Expressions of the form G_n correspond to the first right-hand side term of (3.8). The asymptotic distribution of G_n is well-known for the case when z_{t1} and z_{t2} are I(0) processes with z_{t1} having the interpretation of a stationary error term and z_{t2} the differences of a nonstationary I(1) regressor in the type I sense. The next theorem states the asymptotics of G_n for I(d) processes with $d \neq 0$.

Theorem 5.5

Let Assumptions 1 and 2 hold.

(i) If $d_2 + d_1 > 0$, *then* $G_n \xrightarrow{d} \Xi_{21} + \lambda_{21}$,

where $\lambda_{21} = \lim_{n\to\infty} E(G_n) = \frac{\sigma_{21}\Gamma(1-d_2-d_1)}{\pi(1+d_2+d_1)(d_2+d_1)}\sin(\pi\, d_1)$.

(ii) If $d_2 + d_1 = 0$, *then* $G_n \xrightarrow{d} \Xi_{21} + \lambda_{21}^{*}$, *where* $\lambda_{21}^{*} = \lim_{n\to\infty} E\left(\frac{K(n)}{n}G_n\right)$.

(iii) If $d_2 + d_1 < 0$ *and* $\lambda_{21}^* \neq 0$, *then* $\frac{K(n)}{n} G_n \xrightarrow{p} \lambda_{21}^*$.

(iv) If $d_2 + d_1 < 0$ *and* $\lambda_{21}^* = 0$, *then* $G_n \xrightarrow{d} \Xi_{21}$.

The random variable Ξ_{21} *can be represented as*

$$\Xi_{21} = \Xi_{1,21} + \Xi_{2,21},$$

where

$$\Xi_{1,21} = \int_{-\infty}^{1} Q(r)\, dB^{(1)}(r), \; \Xi_{2,21} = \int_{-\infty}^{1} H(p)\, dB^{(2)}(p),$$

$$Q(r) = \int_{-\infty}^{r} A_{21}(r,p)\, dB^{(2)}(p), \; H(p) = \int_{-\infty}^{p} E_{21}(p,r)\, dB^{(1)}(r).$$

Thereby,

$$A_{21}(r,p) = \frac{(r-p)^{d_2}(1-r)^{d_1} F\left(-d_2, d_1, 1+d_1;\, -\frac{1-r}{r-p}\right)}{\Gamma(1+d_2)\Gamma(1+d_1)}$$
$$- \mathrm{I}_{\{r<0\}} \frac{(r-p)^{d_2}(-r)^{d_1} F\left(-d_2, d_1, 1+d_1;\, \frac{r}{r-p}\right)}{\Gamma(1+d_2)\Gamma(1+d_1)}$$
$$- \mathrm{I}_{\{p<0\}} \frac{(-p)^{d_2}\left[(1-r)^{d_1} - \mathrm{I}_{\{r<0\}}(-r)^{d_1}\right]}{\Gamma(1+d_2)\Gamma(1+d_1)},$$

for $-\infty < p \leq r \leq 1$, *and*

$$E_{21}(p,r) = \frac{1}{\Gamma(d_1+1)\Gamma(d_2+1)}$$
$$\times\left[(1-p)^{d_2}\left((1-r)^{d_1} - (p-r)^{d_1} F\left(-d_1, d_2, 1+d_2; -\frac{1-p}{p-r}\right)\right)\right.$$
$$\left. - \mathrm{I}_{\{p<0\}}(-p)^{d_2}\left((1-r)^{d_1} - (p-r)^{d_1} F\left(-d_1, d_2, 1+d_2; \frac{-p}{p-r}\right)\right)\right],$$

for $-\infty < r \leq p \leq 1$, *where*

$$F(a,b,c;z) = \frac{\Gamma(c)}{\Gamma(a)\Gamma(b)} \sum_{j=0}^{\infty} \frac{\Gamma(a+j)\Gamma(b+j)}{\Gamma(c+j)\, j!}(-z)^j$$

represents the hypergeometric function.

Finally, $B^{(1)}(r)$ *and* $B^{(2)}(r)$ *are the components of a 2-dimensional Brownian motion with covariance matrix* Σ.

Proof: Davidson and Hashimzade (2007b), Proposition 4.1.

□

Note, by setting $\boldsymbol{\varepsilon}_t = (\varepsilon_{t1}, \varepsilon_{t2})^T$, we have

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]}\boldsymbol{\varepsilon}_t \xrightarrow{d} \mathbf{B}(r) = \left(B^{(1)}(r), B^{(2)}(r)\right)^T \text{, for real-valued r,}$$

when extending Corollary 5.3 to the whole real line, where $\mathbf{B}(r)$ is a 2-dimensional Brownian motion with covariance matrix

$$\boldsymbol{\Omega} = Cov\left(\frac{1}{\sqrt{n}}\sum \boldsymbol{\varepsilon}_t\right) = \Sigma = Cov(\boldsymbol{\varepsilon}_t).$$

Davidson and Hashimzade (2007b) note that both $\Xi_{1,21}$ and $\Xi_{2,21}$ are stochastic integrals of Gaussian integrand processes with respect to Brownian motion. Therefore, these integrals are of Itô type. Further, $Q(r) = B_{d_2}(r)$ for $r \geq 0$ and 0 for $r < 0$ when $d_1 = 0$, i.e., $\Xi_{1,21}$ reduces to a regular Itô integral of a fractional Brownian integrand. The term $\Xi_{2,21}$ arises only in the case of fractional integrator functions, since $E_{21}(p, r) = 0$ for all p and r when $d_1 = 0$.

Theorem 5.5 is not applicable to any arbitrary VARFIMA process , since we have to assume some diagonal structure for the lag coefficients. As can be seen from Assumption 2, the innovations of each process affect the other one only by contemporaneous correlation expressed via Σ. Assuming this restriction, we can easily extend the results of Theorem 5.5 to *p*-dimensional processes.

Corollary 5.5

Let $\mathbf{z}_t$ *be a* p *-dimensional VARFIMA(a,* **d***, b) process according to Definition 4.5 exhibiting diagonal structure, i.e.,*

$$\boldsymbol{\Delta}^{\mathbf{d}}(L)\mathbf{z}_t = \mathbf{w}_t, \text{ where}$$

$$A(L)\mathbf{w}_t = B(L)\boldsymbol{\varepsilon}_t,$$

and we additionally assume

$$A(L) = Diag(a_i(L)),$$

$$B(L) = Diag(b_i(L)),$$

$$C(L) = Diag(c_i(L)), \text{ where}$$

$$c_i(L) = a_i(L)^{-1} b_i(L), \text{ for } i = 1, ..., p.$$

Let $\mathbf{z}_t = \left(z_{t1}, \ldots, z_{tp}\right)^T$, $\widetilde{\mathbf{z}}_t = \left(\widetilde{z}_{t1}, \ldots, \widetilde{z}_{tp}\right)^T$ *and* $\widetilde{\mathbf{d}} = \left(\widetilde{d}_1, \ldots, \widetilde{d}_p\right)^T$, *where*

$\Delta\widetilde{z}_{ti} = z_{ti}$, *for t > 0,* $\widetilde{z}_{0i} = 0$ *and* $\widetilde{d}_i = d_i + 1$.

Define

$$G_n^{(ij)} = \frac{1}{K_{ij}(n)} \sum_{t=1}^{n-1} \sum_{s=1}^{t} z_{si} z_{(t+1)j} = \frac{1}{K_{ij}(n)} \sum_{t=1}^{n} \widetilde{z}_{(t-1)i} z_j \text{, where } K_{ij}(n) = n^{1+d_i+d_j} = n^{\widetilde{d}_i + d_j}.$$

Under these assumptions we have the following results for i, j = 1, ..., p:

(i) *If* $\widetilde{d}_i + d_j > 1$, *then* $G_n^{(ij)} \xrightarrow{d} c_i(1) c_j(1) \left(\Xi_{ij} + \lambda_{ij}\right)$,

where $\lambda_{ij} = c_i(1)^{-1}\, c_j(1)^{-1} \lim_{n\to\infty} E\left(G_n^{(ij)}\right) = \dfrac{\sigma_{ij}\Gamma\left(1 - d_i - d_j\right)}{\pi\left(1 + d_i + d_j\right)\left(d_i + d_j\right)} \sin\left(\pi\, d_j\right)$.

(ii) *If* $\widetilde{d}_i + d_j = 1$, *then* $G_n^{(ij)} \xrightarrow{d} c_i(1) c_j(1) \left(\Xi_{ij} + \lambda_{ij}^*\right)$, *where*

$$\lambda_{ij}^* = \lim_{n\to\infty} E\left(\frac{1}{n} \sum_{t=1}^{n} \widetilde{z}_{(t-1)i} z_{tj}\right).$$

(iii) *If* $\widetilde{d}_i + d_j < 1$ *and* $\lambda_{ij}^* \neq 0$, *then* $\dfrac{1}{n} \sum_{t=1}^{n} \widetilde{z}_{(t-1)i} z_{tj} \xrightarrow{p} \lambda_{ij}^*$.

(iv) *If* $\widetilde{d}_i + d_j < 1$ *and* $\lambda_{ij}^* = 0$, *then* $G_n^{(ij)} \xrightarrow{d} c_i(1)\, c_j(1)\, \Xi_{ij}$.

The random variables Ξ_{ij} *and the terms* λ_{ij} *and* λ_{ij}^* *are defined in analogy to the results of theorem 5.5.*

Proof:

Since we assume diagonal structure, it is obviously sufficient to show that Assumptions 1 and 2 of Theorem 5.5 hold. According to the assumptions put on the white noise sequence $\boldsymbol{\varepsilon}_t$, Assumption 1 is fulfilled. Now let z_t be an ARFIMA process (univariate) according to Definition 4.5 with moving average representation

$$z_t = D(L)\varepsilon_t = \sum_{j=0}^{\infty} D_j \varepsilon_{t-j} \text{, where } D(L) = \boldsymbol{\Delta}^{-d}(L) c(L) \text{ and } c(L) = a^{-1}(L) b(L).$$

From Theorem 4.3 we have

$$D_j \sim \Gamma(d)^{-1} j^{d-1} c(1) \quad as \quad j \to \infty.$$

Hence, Assumption (2a) holds when choosing the sequence $L(j)$ to be implicitly defined by

$$\mathrm{L}(j) = \frac{D_j}{\Gamma(d) j^{d-1}}.$$

Obviously, $L(j) \to c(1)$, and the sequence is hence slowly varying at infinity.

If $d = 0$, the process z_t is an ARMA process with absolutely summable coefficients, which decline geometrically fast to zero (Proposition 4.2). Hence Assumption (2b) is fulfilled.

For the more elaborate case when $d \in (-0.5, 0)$ we rewrite

$$z_t = \sum_{j=0}^{\infty} D_j \varepsilon_{t-j} \quad \text{as} \quad \Delta^d(L) z_t = c(L) \varepsilon_t .$$

Differencing both sides of the latter representation yields

$$\Delta\Delta^d(L) z_t = \Delta\, c(L) \varepsilon_t \quad \text{and hence} \quad \Delta^{1+d} z_t = c(L) \Delta\, \varepsilon_t .$$

The moving average representation of z_t is now given by $z_t = \Delta^{-d-1}(L) c(L) \Delta\, \varepsilon_t$.

Note that $\Delta^{-d-1}(L) = (1-L)^{-d-1}$ is defined as

$$(1-L)^{-d-1} = \sum_{j=0}^{\infty} \frac{\Gamma(j+d+1) L^j}{\Gamma(j+1)\,\Gamma(d+1)} = \sum_{j=0}^{\infty} \theta_j L^j ,$$

where

$$\theta_j = \frac{\Gamma(j+d+1)}{\Gamma(j+1)\Gamma(d+1)} .$$

Convolution of the filters $(1-L)^{-d-1}$ and $c(L)$ yields

$$\Delta^{-d-1}(L) c(L) = \sum_{i=0}^{\infty} \theta_i L^i \sum_{i=0}^{\infty} c_j L^j = \sum_{j=0}^{\infty} \delta_j L^j ,$$

where

(5.5) $$\delta_k = \sum_{i=0}^{\infty} \theta_i c_{k-i}$$

(see Gourieroux and Montfort [1997], p. 140).

Hence, z_t can alternatively be represented by

(5.6) $$z_t = \sum_{j=0}^{\infty} \delta_j \Delta\, \varepsilon_{t-j} .$$

Writing (5.6) out as

$$\begin{aligned} z_t &= \delta_0(\varepsilon_t - \varepsilon_{t-1}) + \delta_1(\varepsilon_{t-1} - \varepsilon_{t-2}) + \delta_2(\varepsilon_{t-2} - \varepsilon_{t-3}) + \ldots \\ &= \delta_0 \varepsilon_t + (\delta_1 - \delta_0)\varepsilon_{t-1} + (\delta_2 - \delta_1)\varepsilon_{t-2} + (\delta_3 - \delta_2)\varepsilon_{t-3} + \ldots \end{aligned}$$

we have $D_0 = \delta_0$ and $D_j = \delta_j - \delta_{j-1}$, for $j \geq 1$.

On using Stirling's approximation for large j that

$$\frac{\Gamma(j+a)}{\Gamma(j+b)} \sim j^{a-b} ,$$

it can be established that

$$\theta_j = \frac{\Gamma(j+d+1)}{\Gamma(j+1)\Gamma(d+1)} \sim \frac{j^d}{\Gamma(d+1)},$$

and hence,

$$\delta_j \sim j^d \; \Gamma(d+1)^{-1} c(1).$$

The last result can be easily seen when writing the coefficients δ_k in (5.5) out as

$$\delta_0 = 1,$$

$$\delta_1 = c_1 + \theta_1,$$

$$\delta_2 = c_2 + \theta_1 c_1 + \theta_2,$$

$$\delta_3 = c_3 + \theta_1 c_2 + \theta_2 c_1 + \theta_3,$$

$$\delta_4 = c_4 + \theta_1 c_3 + \theta_2 c_2 + \theta_3 c_1 + \theta_4,$$

$$\ldots$$

and noting that the coefficients c_j decline geometrically fast to zero.

Hence, Assumption (2c) holds when choosing the sequence $L(j)$ to be implicitly defined by

$$\mathrm{L}(j) = \frac{\delta_j}{\Gamma(d+1) j^d}$$

with $L(j) \to c(1)$.

□

Appendix 5

Theorem A-5.1:

If $\mathbf{w}_t$ *is a stationary p-dimensional process with zero mean and absolutely continuous spectrum* $\mathbf{f}(\lambda)$*, continuous at the origin, then*

$$\lim_{n\to\infty} n^{-1}\sum_{t=1}^{n}\sum_{s=1}^{n} E\left(\mathbf{w}_t\mathbf{w}_s^T\right) = 2\pi\,\mathbf{f}(0).$$

Proof: See Hannan (1970), p. 208.

□

Corollary A-5.1:

If $\mathbf{w}_t$ *is a VARMA*(a,b) *process then*

$$\lim_{n\to\infty} n^{-1}\sum_{t=1}^{n}\sum_{s=1}^{n} E\left(\mathbf{w}_t\mathbf{w}_s^T\right) = C(1)\,\Sigma\, C(1)^T =: \mathbf{\Omega},$$

where $\mathbf{\Omega}$ *is positive definite.*

Proof:

This follows from Theorem A-5.1 by application of Corollary 4.4 to the VARMA case.

□

Note that the expression

$$\lim_{n\to\infty} n^{-1}\sum_{t=1}^{n}\sum_{s=1}^{n} E\left(\mathbf{w}_t\mathbf{w}_s^T\right)$$

may be interpreted as the asymptotic covariance matrix of $n^{-1/2}\sum_{t=1}^{n}\mathbf{w}_t$.

Definition A-5.1: Standard Brownian Motion

a) Standard Brownian motion

Standard Brownian Motion is a stochastic process $\{W(t), t\ge 0\}$ *satisfying the conditions*

(i) $W(0)=0$*, a.s.,*

(ii) $W(t_2)-W(t_1), W(t_3)-W(t_2),\ldots,W(t_n)-W(t_{n-1})$*, are independent for every* $n\in\{3,4,\ldots\}$ *and every* $\mathbf{t}=(t_1,\ldots,t_n)^T$ *such that* $0\le t_1<t_2<\ldots<t_n$*,*

(iii) $W(t)-W(s)$ *is distributed as* $N(0,t-s)$ *for* $t\ge s$.

b) Multivariate standard Brownian motion

A stochastic process $\mathbf{W} = (W_1, \ldots, W_p)^T$ *is a multivariate (p-dimensional) standard Brownian motion, if* $W_1, \ldots, W_p$ *are independent (univariate) standard Brownian motions as in a).*

Definition A-5.2: Correlated Brownian Motion

The p-dimensional Brownian Motion $\mathbf{B}$ *with covariance matrix* $\mathbf{\Omega}$*, is defined as*

$$\mathbf{B} = \mathbf{\Omega}^{1/2}\mathbf{W},$$

where $\mathbf{W}$ *is a p-dimensional standard Brownian motion and* $\mathbf{\Omega}$ *is a symmetric positive definite matrix.*

Definition A-5.3 (L_p-bounded)

Let X be a random variable on a probability space (Ω, F, P). If $E|X|^p < 0$, for any real $p > 0$, X is said to be L_p-bounded. The L_p-norm of X is defined as

$$\|X\|_p = \left(E|X|^p\right)^{1/p}.$$

A stochastic sequence $\{X_t\}_{-\infty}^{\infty}$ on (Ω, F, P) is said to be uniformly L_p-bounded for $p > 0$, if

$$\sup_t \|X_t\|_p \leq B < \infty .$$

Lemma A-5.2 (Minkowski inequality)

For $r \geq 1$ and two random variables X and Y,

$$\|X + Y\|_r \leq \|X\|_r + \|Y\|_r .$$

Proof:

Davidson (1994), p. 139.

□

Lemma A-5.3 (Conditional modulus inequality)

Let Y be an integrable random variable on a probability space (Ω, F, P) and G a σ-field contained in F. Then,

$$| E(Y \mid G)| \leq E(\,| Y | \mid G) \text{ a.s.}$$

Proof:

Davidson (1994), p. 151.

□

Definition A-5.4: Near-epoch dependence (Davidson [1994], p. 261)

For a stochastic sequence $\{ \varepsilon_t \}_{-\infty}^{\infty}$ on a probability space (Ω, F, P), let

$$F_{t-m}^{t+m} = \sigma(\varepsilon_{t-m}, \ldots, \varepsilon_{t+m})$$

the σ-field generated by $\varepsilon_{t-m}, \ldots, \varepsilon_{t+m}$, such that $\{ F_{t-m}^{t+m} \}_{m=0}^{\infty}$ is an increasing sequence of σ-fields. If, for $p > 0$, a sequence of integrable random variables $\{w_t\}_{-\infty}^{\infty}$ satisfies

$$\| w_t - E(w_t \mid F_{t-m}^{t+m}) \|_p \le d_t \nu_m,$$

where $\nu_m \to 0$, and $\{ d_t \}_{-\infty}^{\infty}$ is a sequence of positive constants, w_t will be said to be near-epoch dependent in L_p-norm (L_p-NED) on $\{\varepsilon_t\}_{-\infty}^{\infty}$. We will say that the sequence $\{w_t\}_{-\infty}^{\infty}$ is L_p-NED of size $-\varphi_0$ when $\nu_m = O(m^{-\varphi})$ for $\varphi > \varphi_0$.

Proposition A-5.4

Let w_t be an ARMA(a, b) process with moving average representation

$$w_t = \sum_{j=0}^{\infty} c_j \varepsilon_{t-j} .$$

Then, w_t is L_2-NED of size -0.5 on ε_t with $d_t = 1$.

Proof:

According to the Minkowski and conditional modulus inequalities (Lemmas A-5.2 and A-5.3)

$$\begin{aligned}
\| w_t - E(w_t \mid F_{t-m}^{t+m}) \|_2 &= \left\| \sum_{j=0}^{\infty} c_j \left(\varepsilon_{t-j} - E(\varepsilon_{t-j} \mid F_{t-m}^{t+m})\right) \right\|_2 = \left\| \sum_{j=m+1}^{\infty} c_j \left(\varepsilon_{t-j} - E(\varepsilon_{t-j} \mid F_{t-m}^{t+m})\right) \right\|_2 \\
&\le \left\| \sum_{j=m+1}^{\infty} c_j \varepsilon_{t-j} \right\|_2 + \left\| \sum_{j=m+1}^{\infty} c_j E(\varepsilon_{t-j} \mid F_{t-m}^{t+m}) \right\|_2 \\
&\le \sum_{j=m+1}^{\infty} | c_j | \left(\|\varepsilon_{t-j}\|_2 + \| E(\varepsilon_{t-j} \mid F_{t-m}^{t+m}) \|_2 \right) \\
&\le \sum_{j=m+1}^{\infty} | c_j | \left(\|\varepsilon_{t-j}\|_2 + \left\{ E\left[E(| \varepsilon_{t-j} |^2 | F_{t-m}^{t+m})\right] \right\}^{1/2} \right) \\
&= \sum_{j=m+1}^{\infty} | c_j | \left(\|\varepsilon_{t-j}\|_2 + \|\varepsilon_{t-j}\|_2 \right) \\
&= \sum_{j=m+1}^{\infty} | c_j | 2\sigma \\
&= d_t \nu_m ,
\end{aligned}$$

where $\nu_m = 2\sigma \sum_{j=m+1}^{\infty} | c_j |$ and $d_t = 1$.

According to Proposition 4.2 the coefficients c_j are absolutely summable and decline geometrically fast. Hence, we conclude $c_j = O\left(j^{-1-\varphi}\right)$ for every $\varphi > 0$ and hence, for $\varphi_0 = -0.5$. Hence, $\nu_m \to 0$ with $\nu_m = O\left(m^{-\varphi}\right)$ for $\varphi > -0.5$.

□

Theorem A-5.5 (Continuous mapping theorem)

Let *h(x)* be a continuous function defined on D.

If $X_n \Rightarrow X$, then $h(X_n) \Rightarrow h(X)$.

Proof:

Billingsley (1968), p. 29.

□

Proof of Proposition A-5.4:

First define

$$Z_n(s) := n^{-1/2-d_i} \sum_{t=1}^{[ns]} z_t \ .$$

By Corollary 5.3 we have $Z_n(s) \Rightarrow B_d(s)$. Next, rewrite $n^{-2-2d} \sum_{t=1}^{n} x_t^2$ in terms of

$$Z_n(a_t) = n^{-1/2-d} x_t = n^{-1/2-d} \sum_{s=1}^{t} z_s \ ,$$

where $a_t = t/n$, so that

$$n^{-2-2d} \sum_{t=1}^{n} x_t^2 = n^{-1} \sum_{t=1}^{n} Z_n(a_t)^2 \ .$$

Because $Z_n(s)$ is a constant for $t/n \le s \le (t+1)/n$, we have

$$n^{-1} \sum_{t=1}^{n} Z_n(a_t)^2 = \sum_{t=1}^{n} \int_{t/n}^{(t+1)/n} Z_n(s)^2 \, ds = \int_0^1 Z_n(s)^2 \, ds \ .$$

The continuous mapping theorem (Theorem A-5.5) applies to

$$h(Z_n) = \int_0^1 Z_n(s)^2 \, ds \ .$$

It follows $h(Z_n) \Rightarrow h(B_d)$, so that

$$n^{-2-2d} \sum_{t=1}^{n} x_t \xrightarrow{\ d\ } \int_0^1 B_d^2(s) \, ds \text{ , as claimed.}$$

□

6 Asymptotics for fractionally integrated regressors

In this chapter, the asymptotic distributions of OLS are derived for Models (M1)-(M3). Therefore, most results from the foregoing chapters are summarized in Section 6.1 in a uniform manner with its proper notation. The one-regressor model (M1) is treated in Section 6.2 for stationary and nonstationary settings, whereas the two-regressor model (M2) is treated in Section 6.3. The dynamic regresssion model (M3) is examined in Section 6.4 for the nonstationary case.

6.1 Preliminary results

The following lemma brings together all preliminary results needed for dealing with OLS asymptotics. This essentially requires asymptotic results on sample covariances.

Lemma 6.1

Let $\mathbf{z}_t = \left(z_{t0}, \ldots, z_{tp}\right)^T$ *be a (p+1)-dimensional VARFIMA process with*

$$\mathbf{\Delta}^{\mathbf{d}}(L)\mathbf{z}_t = \mathbf{w}_t, \textit{ where } A(L)\mathbf{w}_t = B(L)\boldsymbol{\varepsilon}_t.$$

Further, let $\tilde{\mathbf{z}}_t = \left(\tilde{z}_{t0}, \ldots, \tilde{z}_{tp}\right)^T$ *and* $\tilde{\mathbf{d}} = \left(\tilde{d}_0, \ldots, \tilde{d}_p\right)^T$, *where* $\Delta \tilde{z}_{ti} = z_{ti}$, *for t > 0,* $\tilde{z}_{0i} = 0$ *and* $\tilde{d}_i = d_i + 1$, *for i = 0,1, ..., p. Denoting the covariance matrix of* $\boldsymbol{\varepsilon}_t$ *by*

$$Cov(\boldsymbol{\varepsilon}_t) = \Sigma = \left(\sigma_{ij}\right)_{i,j=0,\ldots p}$$

and the covariance matrix of $\mathbf{z}_t$ *by*

$$Cov(\mathbf{z}_t) = \Gamma_{\mathbf{z}} = \left(\gamma_{ij}\right)_{i,j=0,\ldots,p} ,$$

we have the following results:

(P1) $$\frac{1}{\sqrt{n}}\sum_{t=1}^{[ns]}\boldsymbol{\varepsilon}_t \Rightarrow \mathbf{B}_{(\boldsymbol{\varepsilon})}(s), \textit{ for } 0 \le s \le 1,$$

where $\mathbf{B}_{(\boldsymbol{\varepsilon})}(s)$ *is a (p+1)-dimensional Brownian motion with covariance matrix* Σ.

(P2) $$\frac{1}{\sqrt{n}}\sum_{t=1}^{[ns]}\mathbf{w}_t \Rightarrow \mathbf{B}_{(\mathbf{w})}(s), \textit{ for } 0 \le s \le 1,$$

where $\mathbf{B}_{(\mathbf{w})}(s)$ *is a (p+1)-dimensional Brownian motion with covariance matrix*

$$\mathbf{\Omega} = \left(\omega_{ij}\right)_{i,j=0,\ldots,p} = C(1)\Sigma C(1)^T .$$

(P3) $\frac{1}{\sqrt{n}}\sum_{t=1}^{[ns]}\mathbf{z}_t \Rightarrow \mathbf{B_d}(s)$, *for* $0 \le s \le 1$,

where $\mathbf{B_d}(s)$ *is a (p+1)-dimensional fractional Brownian motion with memory parameter vector* $\mathbf{d} = (d_0,\ldots,d_p)$ *and covariance matrix*

$$\mathbf{\Psi} = (\psi_{ij})_{i,j=0,\ldots,p},$$

where

$$\psi_{ij} = \frac{\varpi_{ij}}{\Gamma(d_i+1)\ \Gamma(d_j+1)}\left\{\frac{1}{1+d_i+d_j} + \int_0^\infty \left[(1+\tau)^{d_i} - \tau^{d_i}\right]\left[(1+\tau)^{d_j} - \tau^{d_j}\right]d\tau\right\}.$$

(P4) $\frac{1}{n}\sum z_{ti}z_{(t+m)j} \xrightarrow{p} \gamma_{ij}(m)$, *for* $m \in (\ldots,-2,-1,0,1,2,\ldots)$,

where $\gamma_{ij}(-m) = \gamma_{ij}(m)$ *and* $\gamma_{ij}(0) = \gamma_{ij}$, *for i, j = 0, ..., p.*

(P5) $n^{-2\tilde{d}_i}\sum_{t=1}^{n}\tilde{z}_{ti}^2 \xrightarrow{d} \int_0^1 B_{d_i}^2(s)\,ds$,

where $B_{d_i}(s)$ *is the (i+1)-th component in (P3), for i = 0, ..., p.*

(P6)

(i) Any finite set of $\sqrt{n}(\tilde{\gamma}_{kl}(m) - \gamma_{kl}(m))$, *for k, l = 0, ..., p, is jointly asymptotically normal, if* $d_i < 0.25$ *for i = 0, ..., p.*

(ii) If we additionally assume $E\,|\,\varepsilon_{ti}\,|^{4+\eta} < \infty$, *for i =0, ..., p, and some* $\eta > 0$,

then for any finite integer m we have

$$Diag\left(n^{-d_i}\right)\left\{\sum_{t=1}^{[ns]-m}\left[\mathbf{z}_t\mathbf{z}_{t+m}^T - \Gamma(m)\right]\right\}Diag\left(n^{-d_i}\right) \Rightarrow \mathbf{R}(s),$$

if $0.25 < d_i < 0.5$, *for all i =0, ..., p. Thereby* $\mathbf{R}(s)$ *is defined as in Theorem 5.2.*

(P7) *Define*

$$G_n^{(ij)} = \frac{1}{K_{ij}(n)}\sum_{t=1}^{n-1}\sum_{s=1}^{t} z_{si}z_{(t+1)j} = \frac{1}{K_{ij}(n)}\sum_{t=1}^{n}\tilde{z}_{(t-1)i}z_j,\ \text{where}\ \ K_{ij}(n) = n^{1+d_i+d_j} = n^{\tilde{d}_i+d_j}.$$

We additionally assume $\mathbf{z}_t$ *to exhibit diagonal structure, i.e.,*

$A(L) = Diag(a_i(L))$, $B(L) = Diag(b_i(L))$, $C(L) = Diag(c_i(L))$, *where*

$c_i(L) = a_i(L)^{-1}b_i(L)$, *for i = 0, ..., p.*

(i) If $\tilde{d}_i + d_j > 1$, *then* $G_n^{(ij)} \xrightarrow{d} c_i(1)c_j(1)(\Xi_{ij} + \lambda_{ij})$,

where $\lambda_{ij} = c_i(1)^{-1}\,c_j(1)^{-1}\lim_{n\to\infty}E\left(G_n^{(ij)}\right) = \frac{\sigma_{ij}\Gamma(1-d_i-d_j)}{\pi(1+d_i+d_j)\ (d_i+d_j)}\sin(\pi\, d_j).$

(ii) If $\widetilde{d}_i + d_j = 1$, *then* $G_n^{(ij)} \xrightarrow{d} c_i(1)c_j(1)(\Xi_{ij} + \lambda_{ij}^*)$, *where*

$$\lambda_{ij}^* = \lim_{n\to\infty} E\left(\frac{1}{n}\sum_{t=1}^{n} \widetilde{z}_{(t-1)i} z_{tj}\right).$$

(iii) If $\widetilde{d}_i + d_j < 1$ *and* $\lambda_{ij}^* \neq 0$, *then* $\frac{1}{n}\sum_{t=1}^{n} \widetilde{z}_{(t-1)i} z_{tj} \xrightarrow{p} \lambda_{ij}^*$.

(iv) If $\widetilde{d}_i + d_j < 1$ *and* $\lambda_{ij}^* = 0$, *then* $G_n^{(ij)} \xrightarrow{d} c_i(1)\, c_j(1)\, \Xi_{ij}$.

Thereby Ξ_{ij}, λ_{ij} *and* λ_{ij}^* *are defined as in Corollary 5.5 for i, j = 0, 1, ..., p.*

(P8) *We have joint weak convergence of (P3) and (P7).*

Proof:

Results (P1)-(P3) follow from Corollary 5.3. Result (P4) follows by the property of covariance ergodicity (Corollary 4.3). Result (P5) follows from Proposition 5.4. The results in (P6) are taken from Corollary 5.1 and Theorem 5.2, for $i = 0, 1, ..., p$. The results in (P7) are the results of Corollary 5.5, for $i, j = 0, 1, ..., p$. Result (P8) follows directly from Theorem 4.1 in Davidson and Hashimzade (2007b).

□

6.2 Regression model with one regressor

In this section, we consider the univariate regression model (M1), where a single regressor and the error term are generated by a process to be specified. The first theorem deals with the case, where both regressor and error are stationary, whereas the two subsequent theorems consider the cases in which either the regressor or the error is nonstationary.

Theorem 6.2.1 (Stationary regressor and stationary error)

Let in Model (M1) the process $\mathbf{z}_t = (e_t, x_t)^T$ *be driven by a VARFIMA process according to the assumptions of Lemma 6.1 with (P1)-(P5). We set* $z_{t0} = e_t$ *and* $z_{t1} = x_t$.

(i) If $d_0, d_1 \in (-0.5, 0.25)$ *and hence, we additionally assume (P6)(i), then*

$$n^{1/2}\left(\hat{\beta} - \beta - \frac{\gamma_{10}}{\gamma_{11}}\right) \xrightarrow{d} \gamma_{11}^{-1} N(0, V)$$

with $V > 0$ *whose definition can be found in the proof.*

For the remaining cases we additionally assume (P6)(ii) and define

$$R^{(10)}(1) = \frac{1}{\Gamma(d_1)\ \Gamma(d_0)} \int_{-\infty}^{1}\int_{-\infty}^{1} \left[\int_{0}^{1} (\tau - r_1)_+^{d_1 - 1} (\tau - r_0)_+^{d_0 - 1} d\tau \right] dB_{(\varepsilon)}^{(1)}(r_1)\, dB_{(\varepsilon)}^{(0)}(r_0),$$

$$R^{(11)}(1) = \frac{1}{\Gamma(d_1)^2} \int_{-\infty}^{1}\int_{-\infty}^{1} \left[\int_{0}^{1} (\tau - r_1)_+^{d_1 - 1} (\tau - r_0)_+^{d_1 - 1} d\tau \right] dB_{(\varepsilon)}^{(1)}(r_1)\, dB_{(\varepsilon)}^{(1)}(r_0).$$

Thereby $B_{(\varepsilon)}^{(i)}(r)$ *denotes the (i +1)th component of* $\mathbf{B}_{(\varepsilon)}(r)$.

(ii) If $d_0, d_1 \in (0.25, 0.5)$ *and* $\sigma_{10} = 0$, *then*

$$n^{1-d_1-d_0}\left(\hat{\beta} - \beta\right) \xrightarrow{d} \gamma_{11}^{-1} R^{(10)}(1).$$

(iii) If $d_0, d_1 \in (0.25, 0.5)$ *and* $d_1 = d_0$, *then*

$$n^{1-2d_1}\left(\hat{\beta} - \beta - \frac{\gamma_{10}}{\gamma_{11}}\right) \xrightarrow{d} \gamma_{11}^{-1} R^{(10)}(1) - \gamma_{10}\gamma_{11}^{-2} R^{(11)}(1).$$

(iv) If $d_0, d_1 \in (0.25, 0.5)$ *and* $d_1 < d_0$, *then*

$$n^{1-d_1-d_0}\left(\hat{\beta} - \beta - \frac{\gamma_{10}}{\gamma_{11}}\right) \xrightarrow{d} \gamma_{11}^{-1} R^{(10)}(1).$$

(v) If $d_0, d_1 \in (0.25, 0.5)$ *and* $d_1 > d_0$, *then*

$$n^{1-2d_1}\left(\hat{\beta} - \beta - \frac{\gamma_{10}}{\gamma_{11}}\right) \xrightarrow{d} \gamma_{10}\gamma_{11}^{-2} R^{(11)}(1).$$

According to our cases, we either have a normal distribution or a Rosenblatt distribution. The convergence rate in the latter case is strictly slower than $n^{1/2}$. It may also be a noteworthy and surprising fact that in this case the convergence rate decreases as d_1 increases. In general, the estimates are asymptotically biased if the regressor is correlated with the error term. A sufficient condition yielding consistent estimates is $\sigma_{10} = 0$. At least to our knowledge, there is no distributional result available at the moment for the case, when $d_0 \in (-0.5, 0.25]$ and $d_1 \in [0.25, 0.5]$ or vice versa. Note that Chung (2002) derived the asymptotic distribution in the multiple regression model (M4a), p. 11, and our results may not be deducted from his results.

Proof of Theorem 6.2.1:

For $\hat{\beta}$ we have the relation

(6.1) $$\hat{\beta} - \beta - \frac{\gamma_{10}}{\gamma_{11}} = \frac{\sum x_t e_t}{\sum x_t^2} - \frac{\gamma_{10}}{\gamma_{11}}.$$

Generally, we have

(6.2) $n^{-1}\sum x_t^2 \xrightarrow{p} \gamma_{11}$

because of covariance ergodicity. For (i) we multiply (6.1) by $n^{1/2}$, which yields

$$n^{1/2}\left(\frac{\sum x_t e_t}{\sum x_t^2}-\frac{\gamma_{10}}{\gamma_{11}}\right)=n^{1/2}\frac{\sum x_t e_t}{\sum x_t^2}-n^{1/2}\frac{\gamma_{10}}{n^{-1}\sum x_t^2}-n^{1/2}\frac{\gamma_{10}}{\gamma_{11}}+n^{1/2}\frac{\gamma_{10}}{n^{-1}\sum x_t^2}$$
$$=\frac{n^{-1/2}\sum\left(x_t e_t-\gamma_{10}\right)}{n^{-1}\sum x_t^2}-\frac{\gamma_{10}}{\gamma_{11}}\frac{n^{-1/2}\sum\left(x_t^2-\gamma_{11}\right)}{n^{-1}\sum x_t^2}.$$

According to (P6) (i) the terms $n^{-1/2}\sum\left(x_t e_t-\gamma_{10}\right)$ and $n^{-1/2}\sum\left(x_t^2-\gamma_{11}\right)$ are jointly normally distributed with variances v_{10} and v_{11}, respectively, and covariance v_{1011} whose definitions are implicitly given by the results of Theorem 5.1. Thus we have a normal distribution in the limit with variance

$$V=v_{10}+\gamma_{10}^2\gamma_{11}^{-2}v_{11}-2\gamma_{11}^{-1}\gamma_{10}v_{1011}.$$

For (ii) to (v) we multiply (6.1) by $n^{1-d_1-d_0}$, which similarly yields

$$n^{1-d_1-d_0}\left(\frac{\sum x_t e_t}{\sum x_t^2}-\frac{\gamma_{10}}{\gamma_{11}}\right)=\frac{n^{-d_1-d_0}\sum\left(x_t e_t-\gamma_{10}\right)}{n^{-1}\sum x_t^2}-\frac{\gamma_{10}}{\gamma_{11}}\frac{n^{-d_1-d_0}\sum\left(x_t^2-\gamma_{11}\right)}{n^{-1}\sum x_t^2}.$$

Further,

(6.3) $n^{-d_1-d_0}\sum\left(x_t e_t-\gamma_{10}\right) \xrightarrow{d} R^{(10)}(1)$, where

$$R^{(10)}(1)=\frac{1}{\Gamma(d_1)\,\Gamma(d_0)}\int_{-\infty}^{1}\int_{-\infty}^{1}\left[\int_0^1(\tau-r_1)_+^{d_1-1}(\tau-r_0)_+^{d_0-1}\,d\tau\right]dB_{(\varepsilon)}^{(1)}(r_1)\,dB_{(\varepsilon)}^{(0)}(r_0)$$

and

(6.4) $n^{-2d_1}\sum\left(x_t^2-\gamma_{11}\right) \xrightarrow{d} R^{(11)}(1)$, where

$$R^{(11)}(1)=\frac{1}{\Gamma(d_1)^2}\int_{-\infty}^{1}\int_{-\infty}^{1}\left[\int_0^1(\tau-r_1)_+^{d_1-1}(\tau-r_0)_+^{d_1-1}\,d\tau\right]dB_{(\varepsilon)}^{(1)}(r_1)\,dB_{(\varepsilon)}^{(1)}(r_0).$$

If $\sigma_{10}=0$, then $\gamma_{10}=0$ and (ii) can be easily deducted from (6.2) and (6.3).

If $d_1=d_0$, we have joint weak convergence of (6.3) and (6.4),

if $d_1<d_0$, then

$$n^{-d_1-d_0}\sum\left(x_t e_t-\gamma_{10}\right) \xrightarrow{d} R^{(10)}(1) \text{ and } n^{-d_1-d_0}\sum\left(x_t^2-\gamma_{11}\right) \xrightarrow{p} 0,$$

and if $d_1>d_0$, then

$$n^{-2d_1}\sum\left(x_t e_t-\gamma_{10}\right) \xrightarrow{p} 0 \text{ and } n^{-2d_1}\sum\left(x_t^2-\gamma_{11}\right) \xrightarrow{d} R^{(11)}(1).$$

□

Theorem 6.2.2 (Nonstationary regressor and stationary error)

Let in Model (M1) the process $\mathbf{z}_t = (e_t, \Delta x_t)^T$ *be driven by a VARFIMA process according to the assumptions of Lemma 6.1 with (P1)-(P5), (P7) and (P8). We set* $z_{t0} = e_t$ *and* $z_{t1} = \Delta x_t$.

(i) *If* $\tilde{d}_1 + d_0 > 1$, *then*

$$n^{\tilde{d}_1 - d_0}\left(\hat{\beta} - \beta\right) \xrightarrow{d} \frac{c_1(1)c_0(1)\left(\Xi_{10} + \lambda_{10}\right)}{\int_0^1 B_{\tilde{d}_1}^2(s)\,ds}.$$

(ii) *If* $\tilde{d}_1 + d_0 = 1$, *then*

$$n^{\tilde{d}_1 - d_0}\left(\hat{\beta} - \beta\right) \xrightarrow{d} \frac{c_1(1)c_0(1)\left(\Xi_{10} + \lambda_{10}^*\right) + \gamma_{10}}{\int_0^1 B_{\tilde{d}_1}^2(s)\,ds}.$$

(iii) *If* $\tilde{d}_1 + d_0 < 1$ *and* $\lambda_{10}^* \neq 0$, *then*

$$n^{2\tilde{d}_1 - 1}\left(\hat{\beta} - \beta\right) \xrightarrow{d} \frac{\lambda_{10}^* + \gamma_{10}}{\int_0^1 B_{\tilde{d}_1}^2(s)\,ds}.$$

(iv) *If* $\tilde{d}_1 + d_0 < 1$ *and* $\lambda_{10}^* = 0$, *then*

$$n^{2\tilde{d}_1 - 1}\left(\hat{\beta} - \beta\right) \xrightarrow{d} \frac{\gamma_{10}}{\int_0^1 B_{\tilde{d}_1}^2(s)\,ds}.$$

Generally, OLS yields consistent estimates. The convergence rate increases as d_1 increases and d_0 decreases, but is unaffected by d_0, if $\tilde{d}_1 + d_0 < 1$. Further, if $\sigma_{10} = 0$ in (iv), then $\gamma_{10} = 0$ and hence $\hat{\beta} - \beta = o_p\left(n^{1-2\tilde{d}_1}\right)$. Case (i) parallels Proposition 6.5 of Robinson and Marinucci (2001). Case (iii) and (iv) parallel Proposition 6.1 of the same paper. Since these authors use the type II approach for generating nonstationary sequences, their limiting distributions generally differ from ours. At the same time their convergence rates match ours for these cases. But it should be noted that they additionally have to assume $d_0 > 0$ throughout their work. Further, the case $d_0 = 0$ does not fit in smoothly but needs separate treatment in their work. Their convergence rate for the special case $\tilde{d}_1 + d_0 = 0$ is slightly slower than ours by the factor $1/\log n$. Hence, there seem to be some advantages of our type I based approach over the harmonic analysis taken by Robinson and Marinucci (2001). These specifically include the facts, that the special $I(1)/I(0)$ case needs no separate analysis but can be regarded just as a special case of fractional integration, and that we can treat the full range of long memory parameters for both regressor and error term.

Proof of Theorem 6.2.2:

We first note that

$$\sum x_t e_t = \sum x_{t-1} e_t + \sum \Delta x_t e_t = \sum x_{t-1} e_t + O_p(n), \text{ since}$$

$$n^{-1} \sum \Delta x_t e_t \xrightarrow{p} \gamma_{10}.$$

The rest follows directly from (P5), (P7) and (P8).

(i) If $\tilde{d}_1 + d_0 > 1$, then

$$n^{-\tilde{d}_1 - d_0} \sum x_t e_t = n^{-\tilde{d}_1 - d_0} \sum x_{t-1} e_t + o_p(1) \xrightarrow{d} c_1(1) c_0(1) (\Xi_{10} + \lambda_{10}) \text{ and}$$

$$n^{-2\tilde{d}_1} \sum_{t=1}^{n} x_t^2 \xrightarrow{d} \int_0^1 B_{d_1}^2(s)\, ds,$$

and by joint weak convergence,

$$n^{\tilde{d}_1 - d_0} \frac{\sum x_t e_t}{\sum x_t^2} = n^{\tilde{d}_1 - d_0} \left(\hat{\beta} - \beta\right) \xrightarrow{d} \frac{c_1(1) c_0(1) (\Xi_{10} + \lambda_{10})}{\int_0^1 B_{d_1}^2(s)\, ds}.$$

(ii) If $\tilde{d}_1 + d_0 = 1$, the argumentation is as in (i) with

$$\frac{1}{n} \sum x_t e_t = \frac{1}{n} \sum x_{t-1} e_t + \frac{1}{n} \sum \Delta x_t e_t \xrightarrow{d} c_1(0) c_0(0) \left(\Xi_{10} + \lambda_{10}^*\right) + \gamma_{10}.$$

(iii) If $\tilde{d}_1 + d_0 < 1$ and $\lambda_{10}^* \neq 0$, then

$$\frac{1}{n} \sum x_{t-1} e_t + \frac{1}{n} \sum \Delta x_t e_t \xrightarrow{p} \lambda_{10}^* + \gamma_{10}$$

and hence,

$$n^{2\tilde{d}_1 - 1} \frac{\sum x_t e_t}{\sum x_t^2} \xrightarrow{d} \frac{\lambda_{10}^* + \gamma_{10}}{\int_0^1 B_{d_1}^2(s)\, ds}.$$

(iv) If $\tilde{d}_1 + d_0 < 1$ and $\lambda_{10}^* = 0$, then

$$n^{-\tilde{d}_1 - d_0} \sum x_{t-1} e_t \xrightarrow{d} c_1(1) c_0(1)\, \Xi_{10}$$

and hence,

$$n^{\tilde{d}_1 - d_0} \frac{\sum x_{t-1} e_t}{\sum x_t^2} \xrightarrow{d} \frac{c_1(1) c_0(1) \Xi_{10}}{\int_0^1 B_{d_1}^2(s)\, ds}.$$

Obviously,

$$n^{2\tilde{d}_1 - 1} \frac{\sum x_{t-1} e_t}{\sum x_t^2} \xrightarrow{p} 0,$$

since $2\tilde{d}_1 - 1 < \tilde{d}_1 - d_0$, or equivalently, $\tilde{d}_1 + d_0 < 1$.

Hence,

$$n^{2\tilde{d}_1-1}\frac{\sum x_t e_{t1}}{\sum x_t^2} = n^{2\tilde{d}_1-1}\frac{\sum \Delta x_t e_t}{\sum x_t^2} + n^{2\tilde{d}_1-1}\frac{\sum x_{t-1} e_t}{\sum x_t^2} = n^{2\tilde{d}_1-1}\frac{\sum \Delta x_t e_t}{\sum x_t^2} + o_p(1)$$

and since

$$n^{2\tilde{d}_1-1}\frac{\sum \Delta x_t e_t}{\sum x_t^2} \xrightarrow{d} \frac{\gamma_{10}}{\int_0^1 B_{d_1}^2(s)\,ds},$$

we can conclude the result in (iv).

□

Theorem 6.2.3 (Stationary regressor and nonstationary error)

Let in Model (M1) the process $\mathbf{z}_t = (\Delta e_t, x_t)^T$ *be driven by a VARFIMA process according to the assumptions of Lemma 6.1 with (P1)-(P5), (P7) and (P8). We set* $z_{t0} = \Delta e_t$ *and* $z_{t1} = x_t$*.*

(i) If $d_1 + \tilde{d}_0 > 1$*, then*

$$n^{1-d_1-\tilde{d}_0}\left(\hat{\beta}-\beta\right) \xrightarrow{d} \frac{c_0(1)c_1(1)}{\gamma_{11}}\left(\Xi_{01} + \lambda_{01}\right).$$

(ii) If $d_1 + \tilde{d}_0 = 1$*, then*

$$\hat{\beta}-\beta \xrightarrow{d} \frac{c_0(1)c_1(1)\left(\Xi_{01} + \lambda_{01}^*\right) + \gamma_{10}}{\gamma_{11}}.$$

(iii) If $d_1 + \tilde{d}_0 < 1$ *and* $\lambda_{01}^* \neq 0$*, then*

$$\hat{\beta}-\beta \xrightarrow{p} \frac{\lambda_{01}^* + \gamma_{10}}{\gamma_{11}}.$$

(iv) If $d_1 + \tilde{d}_0 < 1$ *and* $\lambda_{01}^* = 0$*, then*

$$\hat{\beta}-\beta \xrightarrow{p} \frac{\gamma_{10}}{\gamma_{11}}.$$

(v) If $d_1 + \tilde{d}_0 \in (0.5, 1)$ *with* $d_0, d_1 < 0.25$ *and* $\sigma_{10} = 0$*, then*

$$n^{1-\tilde{d}_0-d_1}\left(\hat{\beta}-\beta\right) \xrightarrow{d} c_0(1)c_1(1)\frac{\Xi_{01}}{\gamma_{11}}.$$

(vi) If $d_1 + \tilde{d}_0 < 0.5$ *with* $d_0, d_1 < 0.25$ *and* $\sigma_{10} = 0$*, then*

$$n^{1/2}\left(\hat{\beta}-\beta\right) \xrightarrow{d} \gamma_{11}^{-1} N(0, V_{10})$$

with $V_{10} > 0$ *whose definition can be found in the proof.*

In (vi), $d_0, d_1 < 0.25$ in fact is implied by $d_1 + \tilde{d}_0 < 0.5$, since in our definition the integration order of stationary VARFIMA is restricted to the interval $(-0.5, 0.5)$. Hence, OLS may yield consistent and even asymptotically normally distributed estimates, if the degree of long memory in both regressor and error is not too large, and if regressor and error are independent. Though, the corresponding convergence rates imply no super-consistency, i.e., the rates are slower than the usual root-*n* rate. At least to our knowledge, there is no comparable result to Theorem 6.2.3 in literature for the fractional case at the moment.

Proof of Theorem 6.2.3:

The proof is analogous to the proof of Theorem 6.2.2 when noting that

$$\sum x_t e_t = \sum x_t e_{t-1} + \sum x_t \Delta e_t = \sum x_t e_{t-1} + O_p(n), \text{ since } n^{-1}\sum \Delta e_t x_t \xrightarrow{p} \gamma_{10},$$

and $n^{-1}\sum x_t^2 \xrightarrow{p} \gamma_{11}$.

In (v) and (vi) we have

$$(6.5) \quad n^{1-\tilde{d}_0-d_1} \frac{\sum x_t e_{t-1}}{\sum x_t^2} \xrightarrow{d} c_0(1)c_1(1)\frac{\Xi_{01}}{\gamma_{11}} \text{ and}$$

$$(6.6) \quad n^{1/2} \frac{\sum x_t \Delta e_t}{\sum x_t^2} \xrightarrow{d} \gamma_{11}^{-1} N(0, V_{10})$$

according to (P7) (iv) and (P6) (ii), respectively, since we assume $d_0, d_1 < 0.25$ and $\sigma_{10} = 0$ and hence $\lambda_{01}^* = 0$. For (v) we note, if we assume $\tilde{d}_0 + d_1 > 0.5$, then

$$n^{1-\tilde{d}_0-d_1} \frac{\sum x_t \Delta e_t}{\sum x_t^2} \xrightarrow{p} 0 \text{ and hence,}$$

$$n^{1-\tilde{d}_0-d_1} \frac{\sum x_t e_t}{\sum x_t^2} = n^{1-\tilde{d}_0-d_1} \frac{\sum x_t e_{t-1}}{\sum x_t^2} + o_p(1) \xrightarrow{d} c_0(1)c_1(1)\frac{\Xi_{01}}{\gamma_{11}}.$$

For (vi) we note, if we additionally assume $\tilde{d}_0 + d_1 < 0.5$, then

$$n^{1/2} \frac{\sum x_t e_{t-1}}{\sum x_t^2} \xrightarrow{p} 0 \text{ and hence,}$$

$$n^{1/2} \frac{\sum x_t e_t}{\sum x_t^2} = n^{1/2} \frac{\sum x_t \Delta e_t}{\sum x_t^2} + o_p(1) \xrightarrow{d} \gamma_{11}^{-1} N(0, V_{10}).$$

Note that $n^{-1/2}\sum x_t \Delta e_t$ is normally distributed with variance $V_{10} > 0$ whose definition is implicitly given by the results of Theorem 5.1.

□

Note, the case $\tilde{d}_0 + d_1 = 0.5$ is not available at the present, since for that we would need joint weak convergence of (6.5) and (6.6).

6.3 Regression model with two regressors

In the following we consider the univariate regression model with two regressors

(M2) $\quad y_t = \beta_1 x_{t1} + \beta_2 x_{t2} + e_t$, for $t = 1,\ldots,n$,

where $\mathbf{z}_t = (e_t, x_{t1}, \Delta x_{t2})^T$ is driven by a VARFIMA process according to the assumptions of Lemma 6.1. Hence, we assume x_{t2} to be a nonstationary $\mathrm{I}(\tilde{d}_2)$ process with $\tilde{d}_2 = d_2 + 1$, whereas the regressor x_{t1} is a stationary $\mathrm{I}(d_1)$ process and the error e_t a stationary $\mathrm{I}(d_0)$ process. Defining

$$X = \begin{pmatrix} x_{11} & x_{12} \\ \vdots & \vdots \\ x_{n1} & x_{n2} \end{pmatrix}, \; y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \text{ and } e = \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix},$$

the OLS estimator for $\beta = (\beta_1, \beta_2)^T$ is given by

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = (X^T X)^{-1} X^T y = (X^T X)^{-1} X^T (X\beta + e).$$

This yields the relation

$$\hat{\beta} - \beta = \begin{pmatrix} \sum x_{t1}^2 & \sum x_{t2}x_{t1} \\ \sum x_{t2}x_{t1} & \sum x_{t2}^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum x_{t1}e_t \\ \sum x_{t2}e_t \end{pmatrix}$$

and hence,

$$\text{(6.7)} \quad \hat{\beta}_1 - \beta_1 = \frac{\dfrac{\sum x_{t1}e_t}{\sum x_{t1}^2} - \dfrac{\sum x_{t2}x_{t1}}{\sum x_{t1}^2}\dfrac{\sum x_{t2}e_t}{\sum x_{t2}^2}}{1 - \dfrac{\sum x_{t2}x_{t1}}{\sum x_{t1}^2}\dfrac{\sum x_{t2}x_{t1}}{\sum x_{t2}^2}}$$

and

$$\text{(6.8)} \quad \hat{\beta}_2 - \beta_2 = \frac{\dfrac{\sum x_{t2}e_t}{\sum x_{t2}^2} - \dfrac{\sum x_{t2}x_{t1}}{\sum x_{t2}^2}\dfrac{\sum x_{t1}e_t}{\sum x_{t1}^2}}{1 - \dfrac{\sum x_{t2}x_{t1}}{\sum x_{t1}^2}\dfrac{\sum x_{t2}x_{t1}}{\sum x_{t2}^2}}$$

respectively. We note that (6.7) and (6.8) each consist of 4 different terms, which are

(t1) $\dfrac{\sum x_{t1}e_t}{\sum x_{t1}^2}$, (t2) $\dfrac{\sum x_{t2}x_{t1}}{\sum x_{t1}^2}$, (t3) $\dfrac{\sum x_{t2}x_{t1}}{\sum x_{t2}^2}$ and (t4) $\dfrac{\sum x_{t2}e_t}{\sum x_{t2}^2}$.

The asymptotic properties of OLS in (M2) are determined by the interdependences of these four terms. We will not consider all possible parameter combinations for d_0, d_1 and d_2. This would require a bulk of case differentations not really giving more insight into the material. Instead, we restrict ourselves to one specific case. In the next theorem, sufficient conditions are stated for $n^{1/2}\left(\hat{\beta}_1 - \beta_1 - \gamma_{10}\gamma_{11}^{-1}\right)$ to be normally distributed, and for $n^{1/2}\left(\hat{\beta}_2 - \beta_2\right)$ to be $o_p(1)$. The latter property is sometimes called 'super-consistency' in the literature and a well-known phenomenon when dealing with I(1) processes. It refers to the convergence rate which is faster than the 'classical' root-*n* rate.

Theorem 6.3.1

Let in Model (M2) the process $\mathbf{z}_t = \left(e_t, x_{t1}, \Delta x_{t2}\right)^T$ *be driven by a VARFIMA process according to the assumptions of Lemma 6.1 with (P1)-(P8).*

We set $z_{t0} = e_t$, $z_{t1} = x_{t1}$ *and* $z_{t2} = \Delta x_{t2}$. *If* $d_0, d_1 \in (-0.5, 0.25)$ *and* $\widetilde{d}_2 > 0.75$ *, then*

$$n^{1/2}\begin{pmatrix} \hat{\beta}_1 - \beta_1 - \dfrac{\gamma_{10}}{\gamma_{11}} \\ \hat{\beta}_2 - \beta_2 \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \gamma_{11}^{-1} N(0, V) \\ 0 \end{pmatrix},$$

with V > 0 whose definition can be found in the proof of Theorem 6.2.1.

Proof:

Defining

$$A := \frac{\sum x_{t2}x_{t1}}{\sum x_{t1}^2}\frac{\sum x_{t2}e_t}{\sum x_{t2}^2}, \quad B := \frac{\sum x_{t2}x_{t1}}{\sum x_{t1}^2}\frac{\sum x_{t2}x_{t1}}{\sum x_{t2}^2} \text{ and } C := \frac{\sum x_{t2}x_{t1}}{\sum x_{t2}^2}\frac{\sum x_{t1}e_t}{\sum x_{t1}^2},$$

we can rewrite (6.7) and (6.8) as

(6.9) $$\hat{\beta}_1 - \beta_1 - \frac{\gamma_{10}}{\gamma_{11}} = \frac{\sum x_{t1}e_t}{\sum x_{t1}^2(1-B)} - \frac{\gamma_{10}}{\gamma_{11}} - \frac{A}{1-B} \quad \text{and}$$

(6.10) $$\hat{\beta}_2 - \beta_2 = \frac{\sum x_{t2}e_t}{\sum x_{t2}^2(1-B)} - \frac{C}{1-B},$$

respectively. The rest follows by application of the following two Lemmas to (6.9) and (6.10).

Lemma 6.3.2

If the assumptions of Theorem 6.3.1 hold, then

(i) $n^{1/2}A \xrightarrow{p} 0$,

(ii) $n^{1/2}B \xrightarrow{p} 0$,

(iii) $n^{1/2}C \xrightarrow{p} 0$,

(iv) $n^{1/2}\frac{\sum x_{t2}e_t}{\sum x_{t2}^2} \xrightarrow{p} 0$.

Proof: See Appendix 6.

Lemma 6.3.3

If the assumptions of Theorem 6.3.1 hold, then

$$n^{1/2}\left(\frac{\sum x_{t1}e_t}{\sum x_{t1}^2(1-B)} - \frac{\gamma_{10}}{\gamma_{11}}\right) \xrightarrow{d} \gamma_{11}^{-1}N(0,\ V)$$

with $V > 0$ *whose definition is given in the proof of Theorem 6.2.1.*

Proof:

We have

$$n^{1/2}\left(\frac{\sum x_{t1}e_t}{\sum x_{t1}^2} - \frac{\gamma_{10}}{\gamma_{11}}\right) \xrightarrow{d} \gamma_{11}^{-1}N(0,V) \text{ with V} > 0,$$

if we put $\mathbf{z}_t = (e_t, x_{t1})^T$ in Theorem 6.2.1 (i). Now,

$$n^{1/2}\left(\frac{\sum x_{t1}e_t}{(1-B)\sum x_{t1}^2} - \frac{\gamma_{10}}{\gamma_{11}}\right) - n^{1/2}\left(\frac{\sum x_{t1}e_t}{\sum x_{t1}^2} - \frac{\gamma_{10}}{\gamma_{11}}\right) = n^{1/2}B\frac{\sum x_{t1}e_t}{(1-B)\sum x_t^2} = o_p(1),$$

since

$$\frac{\sum x_{t1}e_t}{\sum x_t^2} \xrightarrow{p} \frac{\gamma_{10}}{\gamma_{11}},$$

and $n^{1/2}B = o_p(1)$ according to Lemma 6.3.2 (ii).

□

Theorem 6.3.1 states that $\hat{\beta}_1$ generally is inconsistent, if x_{t1} is correlated with e_t, whereas $\hat{\beta}_2$ yields superconsistent estimates. Hence, we have a degenerate bivariate normal distribution. At least to our knowledge, there is no comparable result to Theorem 6.3.1 in literature for the fractional case at the moment.

6.4 Dynamic-Regression model

In the following we consider the univariate dynamic regression model

(M3) $y_t = \alpha\, y_{t-1} + \beta\, z_t + e_t$, for $t = 1,\dots,n$,

where $|\alpha| < 1$, $y_0 = 0$ and $\mathbf{z}_t^* = (e_t, \Delta z_t)^T$ is driven by a VARFIMA process according to the assumptions of Lemma 6.1 with $d_0 \in (-0.5, 0.25)$ and $d_1 \in (-0.25, 0.25)$. Hence, we assume z_t to be a nonstationary $\mathrm{I}(\tilde{d}_1)$ process with $\tilde{d}_1 \in (0.75, 1.25)$ and e_t to be a stationary $\mathrm{I}(d_0)$ process. If we define

$$A = \begin{pmatrix} y_0 & z_1 \\ \vdots & \vdots \\ y_{n-1} & z_n \end{pmatrix},\ y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix},\ e = \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix} \text{ and } \gamma = (\alpha, \beta)^T,$$

the model can be represented by

(6.11) $y = Z\gamma + e$.

The OLS estimator for γ in (6.11) is given by

$$\hat{\gamma} = \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} = (Z^T Z)^{-1} Z^T y = (Z^T Z)^{-1} Z^T (Z\gamma + e),$$

which yields the relation

$$\hat{\gamma} - \gamma = (Z^T Z)^{-1} Z^T e.$$

If we linearly transform (6.11) by a nonsingular 2 × 2 matrix *C*, we get

(6.12) $y = X\delta + e$,

where $X = ZC$ and $\delta = C^{-1}\gamma$.

The OLS estimator for δ in (6.12) is given by

$$\hat{\delta} = (X^T X)^{-1} X^T y = (X^T X)^{-1} X^T (X\delta + e),$$

which yields the relation

$$\tilde{\delta} - \delta = (X^T X)^{-1} X^T e.$$

Since

$$\begin{aligned} C\hat{\delta} &= C(X^T X)^{-1} X^T y = C(C^T Z^T Z C)^{-1} C^T Z^T (ZCC^{-1}\gamma + e) \\ &= CC^{-1}(C^T Z^T Z)^{-1} C^T Z^T (Z\gamma + e) \\ &= (Z^T Z)^{-1} (C^T)^{-1} C^T Z^T (Z\gamma + e) \\ &= \hat{\gamma}, \end{aligned}$$

we have the important relation

(6.13) $C^{-1}(\hat{\gamma} - \gamma) = \hat{\delta} - \delta$.

Equation (6.13) will be the basis for deriving the asymptotic distribution of OLS in (M3).

If we define $(x_{t1}, x_{t2}) = (y_{t-1}, z_t)C$ and $\delta = (\delta_1, \delta_2)^T$, Model (M3) can alternatively be represented in the form

$$(6.14) \quad y_t = (y_{t-1}, z_t)CC^{-1}\begin{pmatrix}\alpha \\ \beta\end{pmatrix} + e_t = (x_{t1}, x_{t2})\begin{pmatrix}\delta_1 \\ \delta_2\end{pmatrix} + e_t, \text{ for } t = 1, \ldots, n.$$

For the subsequent analysis we need to define the process $\mathbf{z}_t = (\bar{e}_t, \Delta\bar{z}_t)^T$, where

$$(6.15) \quad \bar{e}_t = \alpha(L)^{-1} e_t \text{ and } \Delta\bar{z}_t = \alpha(L)^{-1}\Delta z_t \text{ with } \alpha(L) = 1 - \alpha L.$$

Obviously, $\mathbf{z}_t$ is also driven by a VARFIMA process according to the assumptions of Lemma 6.1 with $d_0 \in (-0.5, 0.25)$ and $d_1 \in (-0.25, 0.25)$. The lag polynomial $\alpha(L)$ just extends the orders of the autoregressive lag polynomials of e_t and Δz_t. For example, if e_t is a 1-dimensional VARFIMA(a, d_0, b) process, then $\bar{e}_t$ is a VARFIMA(a + 1, d_0, b) process. We note that the product of lag polynomials is commutative in the scalar case.

Theorem 6.4.1

Let in Model (M3) the process $\mathbf{z}_t = (\bar{e}_t, \Delta\bar{z}_t)^T$ be driven by a VARFIMA process according to the assumptions of Lemma 6.1 with (P1)-(P8). We set $z_{t0} = \bar{e}_t$ and $z_{t1} = \Delta\bar{z}_t$. Define

$$\tau := \frac{\beta}{1-\alpha} \text{ and } b^* := \frac{b_1^*}{b_2^*}, \text{ where}$$

$$b_1^* := \gamma_{00}(1) - \alpha\,\gamma_{00} - \tau\,\gamma_{10} + \alpha\,\tau\,\gamma_{10}(1) \text{ and } b_2^* := \gamma_{00} - 2\,\tau\,\gamma_{10}(1) + \tau^2\gamma_{11}.$$

If $d_0 \in (-0.5, 0.25)$ and $\tilde{d}_1 \in (0.75, 1.25)$, then

$$n^{1/2}\begin{pmatrix}\hat{\delta}_1 - \delta_1 - b^* \\ \hat{\delta}_2 - \delta_2\end{pmatrix} \xrightarrow{d} \begin{pmatrix} Z \\ 0\end{pmatrix} \text{ and}$$

$$n^{1/2}\begin{pmatrix}\hat{\alpha} - \alpha - b^* \\ \hat{\beta} - \beta + \tau b^*\end{pmatrix} \xrightarrow{d} \begin{pmatrix} Z \\ -\tau Z\end{pmatrix},$$

where $Z \sim (b_2^*)^{-1} N(0, V^*)$ with $V^* > 0$ (see Lemma 6.4.3 for further details).

Proof:

The technique of the proof is similar to the proof of Theorem 6.3.1.

Defining

$$C = \begin{pmatrix} 1 & 0 \\ -\tau & 1 \end{pmatrix},$$

we have $x_{t1} = y_{t-1} - \tau\, z_t$ and $x_{t2} = z_t$.

Since

$$\begin{aligned} y_{t-1} - \tau\, z_t &= \alpha\left(y_{t-2} - \tau\, z_{t-1}\right) + \alpha\,\tau\, z_{t-1} - \tau\, z_t + \beta\, z_{t-1} + e_{t-1} \\ &= \alpha\left(y_{t-2} - \tau\, z_{t-1}\right) + e_{t-1} - \tau\,\Delta z_t\,, \end{aligned}$$

we have

$$\alpha(L)\, x_{t1} = e_{t-1} - \tau\,\Delta z_t \text{, or equivalently, } x_{t1} = \bar{e}_{t-1} - \tau\,\Delta\bar{z}_t\,.$$

Hence,

(6.16) $\quad x_{t1} = z_{(t-1)0} - \tau\, z_{t1}$ and $x_{t2} = z_t$.

Note, according to (6.15) we have

$$z_t = \bar{z}_t - \alpha\,\bar{z}_{t-1} \quad \text{and} \quad e_t = \bar{e}_t - \alpha\,\bar{e}_{t-1}$$

and hence,

(6.17) $\quad z_t = \tilde{z}_{t1} - \alpha\,\tilde{z}_{(t-1)1}$ and $e_t = z_{t0} - \alpha\, z_{(t-1)0}$

according to the notation of Lemma 6.1. Following (6.7), (6.8), (6.16) and (6.17) we get

(6.18)

$$\hat{\delta}_1 - \delta_1$$

$$= \frac{\dfrac{\sum\left(z_{(t-1)0} - \tau\, z_{t1}\right)\left(z_{t0} - \alpha\, z_{(t-1)0}\right)}{\sum\left(z_{(t-1)0} - \tau\, z_{t1}\right)^2} - \dfrac{\sum\left(\tilde{z}_{t1} - \alpha\,\tilde{z}_{(t-1)1}\right)\left(z_{(t-1)0} - \tau\, z_{t1}\right)}{\sum\left(z_{(t-1)0} - \tau\, z_{t1}\right)^2}\dfrac{\sum\left(\tilde{z}_{t1} - \alpha\,\tilde{z}_{(t-1)1}\right)\left(z_{t0} - \alpha\, z_{(t-1)0}\right)}{\sum\left(\tilde{z}_{t1} - \alpha\,\tilde{z}_{(t-1)1}\right)^2}}{1 - \dfrac{\sum\left(\tilde{z}_{t1} - \alpha\,\tilde{z}_{(t-1)1}\right)\left(z_{(t-1)0} - \tau\, z_{t1}\right)}{\sum\left(z_{(t-1)0} - \tau\, z_{t1}\right)^2}\dfrac{\sum\left(\tilde{z}_{t1} - \alpha\,\tilde{z}_{(t-1)1}\right)\left(z_{(t-1)0} - \tau\, z_{t1}\right)}{\sum\left(\tilde{z}_{t1} - \alpha\,\tilde{z}_{(t-1)1}\right)^2}}$$

and

(6.19)

$$\hat{\delta}_2 - \delta_2$$

$$= \frac{\dfrac{\sum\left(\tilde{z}_{t1} - \alpha\,\tilde{z}_{(t-1)1}\right)\left(z_{t0} - \alpha\, z_{(t-1)0}\right)}{\sum\left(\tilde{z}_{t1} - \alpha\,\tilde{z}_{(t-1)1}\right)^2} - \dfrac{\sum\left(\tilde{z}_{t1} - \alpha\,\tilde{z}_{(t-1)1}\right)\left(z_{(t-1)0} - \tau\, z_{t1}\right)}{\sum\left(\tilde{z}_{t1} - \alpha\,\tilde{z}_{(t-1)1}\right)^2}\dfrac{\sum\left(z_{(t-1)0} - \tau\, z_{t1}\right)\left(z_{t0} - \alpha\, z_{(t-1)0}\right)}{\sum\left(z_{(t-1)0} - \tau\, z_{t1}\right)^2}}{1 - \dfrac{\sum\left(\tilde{z}_{t1} - \alpha\,\tilde{z}_{(t-1)1}\right)\left(z_{(t-1)0} - \tau\, z_{t1}\right)}{\sum\left(z_{(t-1)0} - \tau\, z_{t1}\right)^2}\dfrac{\sum\left(\tilde{z}_{t1} - \alpha\,\tilde{z}_{(t-1)1}\right)\left(z_{(t-1)0} - \tau\, z_{t1}\right)}{\sum\left(\tilde{z}_{t1} - \alpha\,\tilde{z}_{(t-1)1}\right)^2}}\,.$$

Defining

$$A^* := \frac{\sum(\tilde{z}_{t1}-\alpha\,\tilde{z}_{(t-1)1})(z_{(t-1)0}-\tau\,z_{t1})}{\sum(z_{(t-1)0}-\tau\,z_{t1})^2}\frac{\sum(\tilde{z}_{t1}-\alpha\,\tilde{z}_{(t-1)1})(z_{t0}-\alpha\,z_{(t-1)0})}{\sum(\tilde{z}_{t1}-\alpha\,\tilde{z}_{(t-1)1})^2},$$

$$B^* := \frac{\sum(\tilde{z}_{t1}-\alpha\,\tilde{z}_{(t-1)1})(z_{(t-1)0}-\tau\,z_{t1})}{\sum(z_{(t-1)0}-\tau\,z_{t1})^2}\frac{\sum(\tilde{z}_{t1}-\alpha\,\tilde{z}_{(t-1)1})(z_{(t-1)0}-\tau\,z_{t1})}{\sum(\tilde{z}_{t1}-\alpha\,\tilde{z}_{(t-1)1})^2},$$

$$C^* := \frac{\sum(\tilde{z}_{t1}-\alpha\,\tilde{z}_{(t-1)1})(z_{(t-1)0}-\tau\,z_{t1})}{\sum(\tilde{z}_{t1}-\alpha\,\tilde{z}_{(t-1)1})^2}\frac{\sum(z_{(t-1)0}-\tau\,z_{t1})(z_{t0}-\alpha\,z_{(t-1)0})}{\sum(z_{(t-1)0}-\tau\,z_{t1})^2},$$

we can rewrite (6.18) and (6.20) as

$$\text{(6.20)}\qquad \hat{\delta}_1-\delta_1-b^* = \frac{\sum(z_{(t-1)0}-\tau\,z_{t1})(z_{t0}-\alpha\,z_{(t-1)0})}{(1-B^*)\sum(z_{(t-1)0}-\tau\,z_{t1})^2}-b^*-\frac{A^*}{1-B^*}$$

and

$$\text{(6.21)}\qquad \hat{\delta}_2-\delta_2 = \frac{\sum(\tilde{z}_{t1}-\alpha\,\tilde{z}_{(t-1)1})(z_{t0}-\alpha\,z_{(t-1)0})}{\sum(\tilde{z}_{t1}-\alpha\,\tilde{z}_{(t-1)1})^2(1-B^*)}-\frac{C^*}{1-B^*},$$

respectively.

The rest follows by application of the following two Lemmas to (6.20) and (6.21), and (6.13).

Lemma 6.4.2

If the assumptions of Theorem 6.4.1 hold, then

(i) $n^{1/2}A^* \xrightarrow{p} 0$,

(ii) $n^{1/2}B^* \xrightarrow{p} 0$,

(iii) $n^{1/2}C^* \xrightarrow{p} 0$,

(iv) $n^{1/2}\frac{\sum(\tilde{z}_{t1}-\alpha\,\tilde{z}_{(t-1)1})(z_{t0}-\alpha\,z_{(t-1)0})}{\sum(\tilde{z}_{t1}-\alpha\,\tilde{z}_{(t-1)1})^2} \xrightarrow{p} 0$.

Proof:

See Appendix 6.

Lemma 6.4.3

If the assumptions of Theorem 6.4.1 hold, then

$$n^{1/2}\left(\frac{\sum(z_{(t-1)0}-\tau\,z_{t1})(z_{t0}-\alpha\,z_{(t-1)0})}{\sum(z_{(t-1)0}-\tau\,z_{t1})^2(1-B^*)}-b^*\right) \xrightarrow{d} (b_2^*)^{-1}N(0,V^*)$$

with $V^*>0$ *whose definition is given in the proof.*

Proof:

Note,

$$n^{1/2}\left(\frac{\sum\left(z_{(t-1)0}-\tau\, z_{t1}\right)\left(z_{t0}-\alpha\, z_{(t-1)0}\right)}{\sum\left(z_{(t-1)0}-\tau\, z_{t1}\right)^2}-b^*\right)$$

$$=n^{1/2}\,\frac{\sum\left(z_{(t-1)0}-\tau\, z_{t1}\right)\left(z_{t0}-\alpha\, z_{(t-1)0}\right)}{\sum\left(z_{(t-1)0}-\tau\, z_{t1}\right)^2}-n^{1/2}\,\frac{b_1^*}{\sum\left(z_{(t-1)0}-\tau\, z_{t1}\right)^2}$$

$$-n^{1/2}b^*+n^{1/2}\,\frac{b_1^*}{\sum\left(z_{(t-1)0}-\tau\, z_{t1}\right)^2}$$

$$=\frac{n^{-1/2}\sum\left(z_{(t-1)0}z_{t0}-\alpha\, z_{(t-1)0}^2-\tau\, z_{t1}z_{t0}+\alpha\,\tau\, z_{t1}z_{(t-1)0}-b_1^*\right)}{n^{-1}\sum\left(z_{(t-1)0}-\tau\, z_{t1}\right)^2}$$

$$-b^*\,\frac{n^{-1/2}\sum\left(z_{(t-1)0}^2-2\tau\, z_{(t-1)0}z_{t1}+\tau^2z_{t1}^2-b_2^*\right)}{n^{-1}\sum\left(z_{(t-1)0}-\tau\, z_{t1}\right)^2}.$$

The terms

$$n^{-1/2}\sum\left(z_{(t-1)0}z_{t0}-\gamma_{00}(1)\right),\; n^{-1/2}\sum\left(z_{t1}z_{t0}-\gamma_{10}\right),$$

$$n^{-1/2}\sum\left(z_{(t-1)0}^2-\gamma_{00}\right),\; n^{-1/2}\sum\left(z_{(t-1)0}z_{t1}-\gamma_{10}(1)\right),$$

$$n^{-1/2}\sum\left(z_{t1}^2-\gamma_{11}\right)$$

are jointly normally distributed with variances and covariances whose definitions are implicitly given by the results of Theorem 5.1. Hence, we have a normal distribution in the limit with some variance $V^*>0$.

Now,

$$n^{1/2}\left(\frac{\sum\left(z_{(t-1)0}-\tau\, z_{t1}\right)\left(z_{t0}-\alpha\, z_{(t-1)0}\right)}{\left(1-B^*\right)\sum\left(z_{(t-1)0}-\tau\, z_{t1}\right)^2}-b^*\right)-n^{1/2}\left(\frac{\sum\left(z_{(t-1)0}-\tau\, z_{t1}\right)\left(z_{t0}-\alpha\, z_{(t-1)0}\right)}{\sum\left(z_{(t-1)0}-\tau\, z_{t1}\right)^2}-b^*\right)$$

$$=n^{1/2}B^*\,\frac{\sum\left(z_{(t-1)0}-\tau\, z_{t1}\right)\left(z_{t0}-\alpha\, z_{(t-1)0}\right)}{\left(1-B^*\right)\sum\left(z_{(t-1)0}-\tau\, z_{t1}\right)^2}=o_p(1),$$

since

$$\frac{\sum\left(z_{(t-1)0}-\tau\, z_{t1}\right)\left(z_{t0}-\alpha\, z_{(t-1)0}\right)}{\sum\left(z_{(t-1)0}-\tau\, z_{t1}\right)^2}\xrightarrow{\;p\;}b^*$$

and $n^{1/2}B^*=o_p(1)$ according to Lemma 6.4.2 (ii).

□

Though a nonstationary regressor z_t is involved in Model (M4), the asymptotic results are the same as in the standard asymptotics in classical dynamic regression models, i.e., if the error e_t is serially correlated, then the OLS estimator of α and β have root-*n* inconsistency.

The result in Theorem 6.4.1 can be regarded as new and it generalizes Theorem 1 of Maekawa et al. (1996) who showed inconsistency and asymptotic normality of OLS for the special case when z_t is a simple random walk, i.e., a special I(1) process, and e_t is a stationary AR(1) process, i.e., a special I(0) process. In fact, we will specialize Theorem 6.4.1 to the I(1)/I(0) case in Section 7.3 of Chapter 7. Our result for the fractional case may be interpreted as a special type of 'robustness', where normality is kept, if deviations from the integer values 1 (regressor) and 0 (error) are not too large.

Appendix 6

Proof of Lemma 6.3.2:

The asymptotic properties of the Terms (t1)-(t4) can be easily derived by the results in Section 6.2. Since we assume diagonal structure according to (P7) for $\mathbf{z}_t = (e_t, x_{t1}, \Delta x_{t2})^T$, limiting expressions of each term are unaffected by the third variable and hence, each term can be separately interpreted as an OLS estimate in a type (M1) model. This will be carried out in Lemmas A-6.1 to A-6.4 and summarized in Lemma A-6.5 Finally, it is shown that under the given assumptions the results of Lemma 6.3.2 hold for any admissible parameter combination of d_0, d_1 and d_2.

Lemma A-6.1 [Term (t1)]

If $d_0, d_1 \in (-0.5, 0.25)$*, then*

$$n^{1/2}\left(\frac{\sum x_{t1}e_t}{\sum x_{t1}^2} - \frac{\gamma_{10}}{\gamma_{11}}\right) \xrightarrow{d} \gamma_{11}^{-1} N(0, V)$$

with V > 0, whose definition can be found in the proof of Theorem 6.2.1.

Proof:

This follows by Theorem 6.2.1 (i), if we put $\mathbf{z}_t = (e_t, x_{t1})^T$.

□

Lemma A-6.2 [Term (t2)]

(i) If $\tilde{d}_2 + d_1 > 1$*, then*

$$n^{1-\tilde{d}_2-d_1}\frac{\sum x_{t2}x_{t1}}{\sum x_{t1}^2} \xrightarrow{d} \frac{c_1(1)c_2(1)}{\gamma_{11}}(\Xi_{21} + \lambda_{21}).$$

(ii) If $\tilde{d}_2 + d_1 = 1$*, then*

$$\frac{\sum x_{t2}x_{t1}}{\sum x_{t1}^2} \xrightarrow{d} \frac{c_1(1)c_2(1)(\Xi_{21} + \lambda_{21}^*) + \gamma_{21}}{\gamma_{11}}.$$

(iii) If $\tilde{d}_2 + d_1 < 1$ *and* $\lambda_{21}^* \neq 0$*, then*

$$\frac{\sum x_{t2}x_{t1}}{\sum x_{t1}^2} \xrightarrow{p} \frac{\lambda_{21}^* + \gamma_{21}}{\gamma_{11}}.$$

(iv) If $\widetilde{d}_2 + d_1 < 1$ *and* $\lambda_{21}^* = 0$*, then*

$$\frac{\sum x_{t2} x_{t1}}{\sum x_{t1}^2} \xrightarrow{p} \frac{\gamma_{21}}{\gamma_{11}} .$$

Proof:

This follows by Theorem 6.2.3 (i)-(iv), if we replace $\mathbf{z}_t = (\Delta e_t, x_t)^T$ by $\mathbf{z}_t = (\Delta x_{t2}, x_{t1})^T$.

□

Lemma A-6.3 [Term (t3)]

(i) If $\widetilde{d}_2 + d_1 \geq 1$*, then*

$$\frac{\sum x_{t2} x_{t1}}{\sum x_{t2}^2} = O_p\left(n^{-\widetilde{d}_2 + d_1}\right).$$

(ii) If $\widetilde{d}_2 + d_1 < 1$ *and* $\lambda_{21}^* \neq 0$*, then*

$$n^{2\widetilde{d}_2 - 1} \frac{\sum x_{t2} x_{t1}}{\sum x_{t2}^2} \xrightarrow{d} \frac{\lambda_{21}^* + \gamma_{21}}{\int_0^1 B_{d_2}^2(s)\,ds}.$$

(iii) If $\widetilde{d}_2 + d_1 < 1$ *and* $\lambda_{21}^* = 0$*, then*

$$n^{2\widetilde{d}_2 - 1} \frac{\sum x_{t2} x_{t1}}{\sum x_{t2}^2} \xrightarrow{d} \frac{\gamma_{21}}{\int_0^1 B_{d_2}^2(s)\,ds}.$$

Proof:

This follows by Theorem 6.2.2 (i)-(iv), if we replace $\mathbf{z}_t = (e_t, \Delta x_t)^T$ by $\mathbf{z}_t = (x_{t1}, \Delta x_{t2})^T$.

□

Lemma A-6.4 [Term (t4)]

(i) If $\widetilde{d}_2 + d_0 \geq 1$*, then*

$$\frac{\sum x_{t2} e_t}{\sum x_{t2}^2} = O_p\left(n^{-\widetilde{d}_2 + d_0}\right).$$

(ii) If $\widetilde{d}_2 + d_0 < 1$ *and* $\lambda_{20}^* \neq 0$*, then*

$$n^{2\widetilde{d}_2 - 1} \frac{\sum x_{t2} e_t}{\sum x_{t2}^2} \xrightarrow{d} \frac{\lambda_{20}^* + \gamma_{20}}{\int_0^1 B_{d_2}^2(s)\,ds}.$$

(iii) If $\widetilde{d}_2 + d_0 < 1$ *and* $\lambda_{20}^* = 0$*, then*

$$n^{2\widetilde{d}_2-1}\frac{\sum x_{t2}e_t}{\sum x_{t2}^2} \xrightarrow{d} \frac{\gamma_{20}}{\int_0^1 B_{d_2}^2(s)\,ds}.$$

Proof:

This follows by Theorem 6.2.2 (i)-(iv), if we replace $\mathbf{z}_t = (e_t, \Delta x_t)^T$ by $\mathbf{z}_t = (e_t, \Delta x_{t2})^T$.

□

Next, the results from Lemmas A-6.2 to A-6.4 are summarized in such a way that we can analyze all admissible parameter combinations of d_0, d_1 and d_2 in a convenient way. We distinct between the two cases $\widetilde{d}_2 + d_0 \geq 1$ and $\widetilde{d}_2 + d_0 < 1$, henceforth Case 0a and Case 0b, respectively, and between the two cases $\widetilde{d}_2 + d_1 \geq 1$ and $\widetilde{d}_2 + d_1 < 1$, henceforth Case 1a and Case 1b, respectively.

Lemma A-6.5

Case 0a: *If* $\widetilde{d}_2 + d_0 \geq 1$*, then*

$$\frac{\sum x_{t2}e_t}{\sum x_{t2}^2} = O_p\left(n^{-\widetilde{d}_2+d_0}\right).$$

Case 0b: *If* $\widetilde{d}_2 + d_0 < 1$*, then*

$$\frac{\sum x_{t2}e_t}{\sum x_{t2}^2} = O_p\left(n^{1-2\widetilde{d}_2}\right).$$

Case 1a: *If* $\widetilde{d}_2 + d_1 \geq 1$*, then*

$$\frac{\sum x_{t2}x_{t1}}{\sum x_{t1}^2} = O_p\left(n^{\widetilde{d}_2+d_1-1}\right) \text{ and}$$

$$\frac{\sum x_{t2}x_{t1}}{\sum x_{t2}^2} = O_p\left(n^{-\widetilde{d}_2+d_1}\right).$$

Case 1b: *If* $\widetilde{d}_2 + d_1 < 1$*, then*

$$\frac{\sum x_{t2}x_{t1}}{\sum x_{t1}^2} = O_p(1) \text{ and}$$

$$\frac{\sum x_{t2}x_{t1}}{\sum x_{t2}^2} = O_p\left(n^{1-2\widetilde{d}_2}\right).$$

Lemma A-6.1 and Lemma A-6.5 are now the basis for studying the asymptotics of A, B and C. Under the assumptions of Theorem 6.3.1 there are 4 different parameter constellations,

(i) Case 0a combined with Case 1a, henceforth Case 0a-1a,

(ii) Case 0b combined with Case 1b, henceforth Case 0b-1b,

(ii) Case 0b combined with Case 1a, henceforth Case 0b-1a,

(iv) Case 0a combined with Case 1b, henceforth Case 0a-1b.

Case 0a-1a: $\widetilde{d}_2 + d_0 \geq 1$ and $\widetilde{d}_2 + d_1 \geq 1$

For Term A we have

$$A = \frac{\sum x_{t2}x_{t1}}{\sum x_{t1}^2} \frac{\sum x_{t2}e_t}{\sum x_{t2}^2} = O_p\left(n^{\widetilde{d}_2+d_1-1}\right) O_p\left(n^{-\widetilde{d}_2+d_0}\right)$$

and hence,

$$n^{1/2}A = O_p\left(n^{\widetilde{d}_2+d_1-1/2}\right) O_p\left(n^{\widetilde{d}_2+d_0}\right) = O_p\left(n^{d_1+d_0-1/2}\right),$$

which is $o_p(1)$, since $d_0 + d_1 < 0.5$.

For Term B we have

$$B = \frac{\sum x_{t2}x_{t1}}{\sum x_{t1}^2} \frac{\sum x_{t2}x_{t1}}{\sum x_{t2}^2} = O_p\left(n^{\widetilde{d}_2+d_1-1}\right) O_p\left(n^{-\widetilde{d}_2+d_1}\right)$$

and hence,

$$n^{1/2}B = O_p\left(n^{\widetilde{d}_2+d_1-1/2}\right) O_p\left(n^{-\widetilde{d}_2+d_1}\right) = o_p\left(n^{2d_1-1/2}\right),$$

which is $o_p(1)$, since $2d_1 < 0.5$.

For Term C we have

$$C = \frac{\sum x_{t2}x_{t1}}{\sum x_{t2}^2} \frac{\sum x_{t1}e_t}{\sum x_{t1}^2} = \frac{\sum x_{t2}x_{t1}}{\sum x_{t2}^2}\left(\frac{\sum x_{t1}e_t}{\sum x_{t1}^2} - \frac{\gamma_{10}}{\gamma_{11}}\right) + \frac{\gamma_{10}}{\gamma_{11}} \frac{\sum x_{t2}x_{t1}}{\sum x_{t2}^2}$$
$$= O_p\left(n^{-\widetilde{d}_2+d_1}\right) O_p\left(n^{-1/2}\right) + O_p\left(n^{-\widetilde{d}_2+d_1}\right)$$

and hence,

$$n^{1/2}C = O_p\left(n^{-\widetilde{d}_2+d_1}\right) O_p(1) + O_p\left(n^{-\widetilde{d}_2+d_1+1/2}\right)$$
$$= O_p\left(n^{-\widetilde{d}_2+d_1}\right) + O_p\left(n^{-\widetilde{d}_2+d_1+1/2}\right) = O_p\left(n^{-\widetilde{d}_2+d_1+1/2}\right),$$

which is $o_p(1)$, since $\widetilde{d}_2 - d_1 > 1/2$.

Finally, we note

$$n^{1/2}\frac{\sum x_{t2}e_t}{\sum x_{t2}^2}=n^{1/2}O_p\left(n^{-\widetilde{d}_2+d_0}\right)=O_p\left(n^{-\widetilde{d}_2+d_0+1/2}\right),$$

which is $o_p(1)$, since $\widetilde{d}_2-d_0>1/2$.

Case 0b-1b: $\widetilde{d}_2+d_0<1$ and $\widetilde{d}_2+d_1<1$

For Term A we have

$$A=\frac{\sum x_{t2}x_{t1}}{\sum x_{t1}^2}\frac{\sum x_{t2}e_t}{\sum x_{t2}^2}=O_p(1)\ O_p\left(n^{1-2\widetilde{d}_2}\right)=O_p\left(n^{1-2\widetilde{d}_2}\right)$$

and hence,

$$n^{1/2}A=O_p\left(n^{3/2-2\widetilde{d}_2}\right),$$

which is $o_p(1)$, since $2\widetilde{d}_2>3/2$.

For Terms B and C we similarly have

$$B=\frac{\sum x_{t2}x_{t1}}{\sum x_{t1}^2}\frac{\sum x_{t2}x_{t1}}{\sum x_{t2}^2}=O_p(1)\ O_p\left(n^{1-2\widetilde{d}_2}\right)=O_p\left(n^{1-2\widetilde{d}_2}\right)=o_p\left(n^{-1/2}\right)$$

and

$$C=\frac{\sum x_{t2}x_{t1}}{\sum x_{t2}^2}\frac{\sum x_{t1}e_t}{\sum x_{t1}^2}=O_p\left(n^{1-2\widetilde{d}_2}\right)O_p(1)=O_p\left(n^{1-2\widetilde{d}_2}\right)=o_p\left(n^{-1/2}\right),$$

respectively.

Finally,

$$\frac{\sum x_{t2}e_t}{\sum x_{t2}^2}=O_p\left(n^{1-2\widetilde{d}_2}\right)=o_p\left(n^{-1/2}\right).$$

Case 0b-1a: $\widetilde{d}_2+d_0<1$ and $\widetilde{d}_2+d_1\geq 1$.

For Term A we have

$$A=\frac{\sum x_{t2}x_{t1}}{\sum x_{t1}^2}\frac{\sum x_{t2}e_t}{\sum x_{t2}^2}=O_p\left(n^{\widetilde{d}_2+d_1-1}\right)\ O_p\left(n^{1-2\widetilde{d}_2}\right)$$

and hence,

$$n^{1/2}A=O_p\left(n^{\widetilde{d}_2+d_1-1/2}\right)\ O_p\left(n^{1-2\widetilde{d}_2}\right)=O_p\left(n^{-\widetilde{d}_2+d_1+1/2}\right),$$

which is $o_p(1)$, since $\widetilde{d}_2-d_1>1/2$.

For Terms B and C we similarly have

$$B = \frac{\sum x_{t2}x_{t1}}{\sum x_{t1}^2}\frac{\sum x_{t2}x_{t1}}{\sum x_{t2}^2} = O_p\left(n^{\tilde{d}_2+d_1-1}\right) O_p\left(n^{-\tilde{d}_2+d_1}\right)$$

and hence,

$$n^{1/2}B = O_p\left(n^{\tilde{d}_2+d_1-1/2}\right) O_p\left(n^{-\tilde{d}_2+d_1}\right) = O_p\left(n^{2d_1-1/2}\right),$$

which is $o_p(1)$ and

$$C = \frac{\sum x_{t2}x_{t1}}{\sum x_{t2}^2}\frac{\sum x_{t1}e_t}{\sum x_{t1}^2} = O_p\left(n^{-\tilde{d}_2+d_1}\right)O_p(1) = O_p\left(n^{-\tilde{d}_2+d_1}\right) = o_p\left(n^{-1/2}\right).$$

Finally,

$$n^{1/2}\frac{\sum x_{t2}e_t}{\sum x_{t2}^2} = n^{1/2}O_p\left(n^{1-2\tilde{d}_2}\right) = O_p\left(n^{3/2-2\tilde{d}_2}\right) = o_p(1).$$

Case 0a-1b: $\tilde{d}_2 + d_0 \geq 1$ and $\tilde{d}_2 + d_1 < 1$

For Terms A, B and C we have

$$A = \frac{\sum x_{t2}x_{t1}}{\sum x_{t1}^2}\frac{\sum x_{t2}e_t}{\sum x_{t2}^2} = O_p(1)\ O_p\left(n^{-\tilde{d}_2+d_0}\right) = o_p\left(n^{-1/2}\right),$$

$$B = \frac{\sum x_{t2}x_{t1}}{\sum x_{t1}^2}\frac{\sum x_{t2}x_{t1}}{\sum x_{t2}^2} = O_p(1)\ O_p\left(n^{1-2\tilde{d}_2}\right) = O_p\left(n^{1-2\tilde{d}_2}\right) = o_p\left(n^{-1/2}\right)$$

and

$$C = \frac{\sum x_{t2}x_{t1}}{\sum x_{t2}^2}\frac{\sum x_{t1}e_t}{\sum x_{t1}^2} = O_p\left(n^{1-2\tilde{d}_2}\right)O_p(1) = O_p\left(n^{1-2\tilde{d}_2}\right) = o_p\left(n^{-1/2}\right).$$

Finally,

$$\frac{\sum x_{t2}e_t}{\sum x_{t2}^2} = O_p\left(n^{-\tilde{d}_2+d_0}\right) = o_p\left(n^{-1/2}\right).$$

This completes the proof of Lemma 6.3.2.

□

Proof of Lemma 6.4.2:

The results of Lemma 6.4.2 can be largely derived by the results of Lemma 6.3.2, since most parts of Terms A^*, B^* and C^* have the same structure with respect to convergence rates as Terms A, B and C.

For Term A^* we have

$$\begin{aligned} A^* &= \frac{\sum\left(\tilde{z}_{t1} - \alpha\, \tilde{z}_{(t-1)1}\right)\left(z_{(t-1)0} - \tau\, z_{t1}\right)}{\sum\left(z_{(t-1)0} - \tau\, z_{t1}\right)^2} \frac{\sum\left(\tilde{z}_{t1} - \alpha\, \tilde{z}_{(t-1)1}\right)\left(z_{t0} - \alpha\, z_{(t-1)0}\right)}{\sum\left(\tilde{z}_{t1} - \alpha\, \tilde{z}_{(t-1)1}\right)^2} \\ &= \frac{\sum\left(\tilde{z}_{t1} - \alpha\, \tilde{z}_{(t-1)1}\right) z_{(t-1)0}}{\sum\left(z_{(t-1)0} - \tau\, z_{t1}\right)^2} \frac{\sum\left(\tilde{z}_{t1} - \alpha\, \tilde{z}_{(t-1)1}\right)\left(z_{t0} - \alpha\, z_{(t-1)0}\right)}{\sum\left(\tilde{z}_{t1} - \alpha\, \tilde{z}_{(t-1)1}\right)^2} \\ &\quad - \tau \frac{\sum\left(\tilde{z}_{t1} - \alpha\, \tilde{z}_{(t-1)1}\right) z_{t1}}{\sum\left(z_{(t-1)0} - \tau\, z_{t1}\right)^2} \frac{\sum\left(\tilde{z}_{t1} - \alpha\, \tilde{z}_{(t-1)1}\right)\left(z_{t0} - \alpha\, z_{(t-1)0}\right)}{\sum\left(\tilde{z}_{t1} - \alpha\, \tilde{z}_{(t-1)1}\right)^2} \end{aligned}$$

and hence,

(6.22) $A^* = A_1^* - \tau\, A_2^*$,

where

$$\begin{aligned} A_1^* &= \frac{\sum\left(\tilde{z}_{t1} - \alpha\, \tilde{z}_{(t-1)1}\right) z_{(t-1)0}}{\sum\left(z_{(t-1)0} - \tau\, z_{t1}\right)^2} \frac{\sum\left(\tilde{z}_{t1} - \alpha\, \tilde{z}_{(t-1)1}\right)\left(z_{t0} - \alpha\, z_{(t-1)0}\right)}{\sum\left(\tilde{z}_{t1} - \alpha\, \tilde{z}_{(t-1)1}\right)^2} \\ &= \frac{\sum\left(\tilde{z}_{t1} - \alpha\, \tilde{z}_{(t-1)1}\right) z_{(t-1)0}}{\sum\left(z_{(t-1)0} - \tau\, z_{t1}\right)^2} \frac{\sum\left(\tilde{z}_{t1} - \alpha\, \tilde{z}_{(t-1)1}\right)\left(z_{t0} - \alpha\, z_{(t-1)0}\right)}{\sum\left(\tilde{z}_{t1} - \alpha\, \tilde{z}_{(t-1)1}\right)^2} \\ &\quad + \frac{\sum\left(z_{t1} - \alpha\, z_{(t-1)1}\right) z_{(t-1)0}}{\sum\left(z_{(t-1)0} - \tau\, z_{t1}\right)^2} \frac{\sum\left(\tilde{z}_{t1} - \alpha\, \tilde{z}_{(t-1)1}\right)\left(z_{t0} - \alpha\, z_{(t-1)0}\right)}{\sum\left(\tilde{z}_{t1} - \alpha\, \tilde{z}_{(t-1)1}\right)^2} \end{aligned}$$

and

$$A_2^* = \frac{\sum\left(\tilde{z}_{t1} - \alpha\, \tilde{z}_{(t-1)1}\right) z_{t1}}{\sum\left(z_{(t-1)0} - \tau\, z_{t1}\right)^2} \frac{\sum\left(\tilde{z}_{t1} - \alpha\, \tilde{z}_{(t-1)1}\right)\left(z_{t0} - \alpha\, z_{(t-1)0}\right)}{\sum\left(\tilde{z}_{t1} - \alpha\, \tilde{z}_{(t-1)1}\right)^2}.$$

First we note

(6.23) $A_2^* = o_p\left(n^{-1/2}\right)$,

since we could treat A_2^* similarly as Term A in Lemma 6.3.2. Further, if we define

(6.24) $A_1^* = A_{11}^* + A_{12}^*$,

where

$$A_{11}^* = \frac{\sum\left(\tilde{z}_{t1} - \alpha\, \tilde{z}_{(t-1)1}\right) z_{(t-1)0}}{\sum\left(z_{(t-1)0} - \tau\, z_{t1}\right)^2} \frac{\sum\left(\tilde{z}_{t1} - \alpha\, \tilde{z}_{(t-1)1}\right)\left(z_{t0} - \alpha\, z_{(t-1)0}\right)}{\sum\left(\tilde{z}_{t1} - \alpha\, \tilde{z}_{(t-1)1}\right)^2}$$

and

$$A_{12}^{*} = \frac{\sum\left(z_{t1} - \alpha\, z_{(t-1)1}\right) z_{(t-1)0}}{\sum\left(z_{(t-1)0} - \tau\, z_{t1}\right)^{2}} \frac{\sum\left(\tilde{z}_{t1} - \alpha\, \tilde{z}_{(t-1)1}\right)\left(z_{t0} - \alpha\, z_{(t-1)0}\right)}{\sum\left(\tilde{z}_{t1} - \alpha\, \tilde{z}_{(t-1)1}\right)^{2}},$$

we have

(6.25) $A_{1}^{*} = o_{p}\left(n^{-1/2}\right) + o_{p}\left(n^{-1/2}\right) = o_{p}\left(n^{-1/2}\right),$

since we could treat A_{11}^{*} similarly as Term B and A_{12}^{*} as Term C of Lemma 6.3.2.

Collecting the results from (6.22)-(6.25) yields $A^{*} = o_{p}\left(n^{-1/2}\right)$.

For Term B^{*} we have

$$\begin{aligned}
B^{*} &= \frac{\sum\left(\tilde{z}_{t1} - \alpha\, \tilde{z}_{(t-1)1}\right)\left(z_{(t-1)0} - \tau\, z_{t1}\right)}{\sum\left(z_{(t-1)0} - \tau\, z_{t1}\right)^{2}} \frac{\sum\left(\tilde{z}_{t1} - \alpha\, \tilde{z}_{(t-1)1}\right)\left(z_{(t-1)0} - \tau\, z_{t1}\right)}{\sum\left(\tilde{z}_{t1} - \alpha\, \tilde{z}_{(t-1)1}\right)^{2}} \\
&= \frac{\sum\left(\tilde{z}_{t1} - \alpha\, \tilde{z}_{(t-1)1}\right) z_{(t-1)0} - \tau \sum\left(\tilde{z}_{t1} - \alpha\, \tilde{z}_{(t-1)1}\right) z_{t\,1}}{\sum\left(z_{(t-1)0} - \tau\, z_{t1}\right)^{2}} \\
&\quad \times \frac{\sum\left(\tilde{z}_{t1} - \alpha\, \tilde{z}_{(t-1)1}\right) z_{(t-1)0} - \tau \sum\left(\tilde{z}_{t1} - \alpha\, \tilde{z}_{(t-1)1}\right) z_{t\,1}}{\sum\left(\tilde{z}_{t1} - \alpha\, \tilde{z}_{(t-1)1}\right)^{2}} \\
&= \frac{\sum\left(\tilde{z}_{t1} - \alpha\, \tilde{z}_{(t-1)1}\right) z_{(t-1)0}}{\sum\left(z_{(t-1)0} - \tau\, z_{t1}\right)^{2}} \frac{\sum\left(\tilde{z}_{t1} - \alpha\, \tilde{z}_{(t-1)1}\right) z_{(t-1)0}}{\sum\left(\tilde{z}_{t1} - \alpha\, \tilde{z}_{(t-1)1}\right)^{2}} \\
&\quad - 2\,\tau\, \frac{\sum\left(\tilde{z}_{t1} - \alpha\, \tilde{z}_{(t-1)1}\right) z_{(t-1)0}}{\sum\left(z_{(t-1)0} - \tau\, z_{t1}\right)^{2}} \frac{\sum\left(\tilde{z}_{t1} - \alpha\, \tilde{z}_{(t-1)1}\right) z_{t\,1}}{\sum\left(\tilde{z}_{t1} - \alpha\, \tilde{z}_{(t-1)1}\right)^{2}} \\
&\quad + \tau^{2}\, \frac{\sum\left(\tilde{z}_{t1} - \alpha\, \tilde{z}_{(t-1)1}\right) z_{t\,1}}{\sum\left(z_{(t-1)0} - \tau\, z_{t1}\right)^{2}} \frac{\sum\left(\tilde{z}_{t1} - \alpha\, \tilde{z}_{(t-1)1}\right) z_{t\,1}}{\sum\left(\tilde{z}_{t1} - \alpha\, \tilde{z}_{(t-1)1}\right)^{2}}
\end{aligned}$$

and hence,

(6.26) $B^{*} = B_{1}^{*} - 2\tau\, B_{2}^{*} + \tau^{2} B_{3}^{*},$

where

$$\begin{aligned}
B_{1}^{*} &= \frac{\sum\left(\tilde{z}_{t1} - \alpha\, \tilde{z}_{(t-1)1}\right) z_{(t-1)0}}{\sum\left(z_{(t-1)0} - \tau\, z_{t1}\right)^{2}} \frac{\sum\left(\tilde{z}_{t1} - \alpha\, \tilde{z}_{(t-1)1}\right) z_{(t-1)0}}{\sum\left(\tilde{z}_{t1} - \alpha\, \tilde{z}_{(t-1)1}\right)^{2}} \\
&= \left(\frac{\sum\left(\tilde{z}_{t1} - \alpha\, \tilde{z}_{(t-1)1}\right) z_{(t-1)0}}{\sum\left(z_{(t-1)0} - \tau\, z_{t1}\right)^{2}} + \frac{\sum\left(z_{t1} - \alpha\, z_{(t-1)1}\right) z_{(t-1)0}}{\sum\left(z_{(t-1)0} - \tau\, z_{t1}\right)^{2}} \right) \\
&\quad \times \left(\frac{\sum\left(\tilde{z}_{t1} - \alpha\, \tilde{z}_{(t-1)1}\right) z_{(t-1)0}}{\sum\left(\tilde{z}_{t1} - \alpha\, \tilde{z}_{(t-1)1}\right)^{2}} + \frac{\sum\left(z_{t1} - \alpha\, z_{(t-1)1}\right) z_{(t-1)0}}{\sum\left(\tilde{z}_{t1} - \alpha\, \tilde{z}_{(t-1)1}\right)^{2}} \right)
\end{aligned}$$

$$= \frac{\sum\left(\tilde{z}_{t1} - \alpha\,\tilde{z}_{(t-1)1}\right) z_{(t-1)0}}{\sum\left(z_{(t-1)0} - \tau\, z_{t1}\right)^2} \frac{\sum\left(\tilde{z}_{t1} - \alpha\,\tilde{z}_{(t-1)1}\right) z_{(t-1)0}}{\sum\left(\tilde{z}_{t1} - \alpha\,\tilde{z}_{(t-1)1}\right)^2}$$
$$+ 2 \frac{\sum\left(\tilde{z}_{t1} - \alpha\,\tilde{z}_{(t-1)1}\right) z_{(t-1)0}}{\sum\left(\tilde{z}_{t1} - \alpha\,\tilde{z}_{(t-1)1}\right)^2} \frac{\sum\left(z_{t1} - \alpha\, z_{(t-1)1}\right) z_{(t-1)0}}{\sum\left(z_{(t-1)0} - \tau\, z_{t1}\right)^2}$$
$$+ \frac{\sum\left(z_{t1} - \alpha\, z_{(t-1)1}\right) z_{(t-1)0}}{\sum\left(z_{(t-1)0} - \tau\, z_{t1}\right)^2} \frac{\sum\left(z_{t1} - \alpha\, z_{(t-1)1}\right) z_{(t-1)0}}{\sum\left(\tilde{z}_{t1} - \alpha\,\tilde{z}_{(t-1)1}\right)^2},$$

$$B_2^* = \frac{\sum\left(\tilde{z}_{t1} - \alpha\,\tilde{z}_{(t-1)1}\right) z_{(t-1)0}}{\sum\left(z_{(t-1)0} - \tau\, z_{t1}\right)^2} \frac{\sum\left(\tilde{z}_{t1} - \alpha\,\tilde{z}_{(t-1)1}\right) z_{t1}}{\sum\left(\tilde{z}_{t1} - \alpha\,\tilde{z}_{(t-1)1}\right)^2}$$

and

$$B_3^* = \frac{\sum\left(\tilde{z}_{t1} - \alpha\,\tilde{z}_{(t-1)1}\right) z_{t1}}{\sum\left(z_{(t-1)0} - \tau\, z_{t1}\right)^2} \frac{\sum\left(\tilde{z}_{t1} - \alpha\,\tilde{z}_{(t-1)1}\right) z_{t1}}{\sum\left(\tilde{z}_{t1} - \alpha\,\tilde{z}_{(t-1)1}\right)^2}.$$

First we note

(6.27) $B_2^* = o_p\left(n^{-1/2}\right)$ and $B_3^* = o_p\left(n^{-1/2}\right)$,

since B_2^* has the same structure as A_1^* and B_3^* has the same structure as A_2^*.

Further, if we define

(6.28) $B_1^* = B_{11}^* + B_{12}^* + B_{13}^*$,

where

$$B_{11}^* = \frac{\sum\left(\tilde{z}_{t1} - \alpha\,\tilde{z}_{(t-1)1}\right) z_{(t-1)0}}{\sum\left(z_{(t-1)0} - \tau\, z_{t1}\right)^2} \frac{\sum\left(\tilde{z}_{t1} - \alpha\,\tilde{z}_{(t-1)1}\right) z_{(t-1)0}}{\sum\left(\tilde{z}_{t1} - \alpha\,\tilde{z}_{(t-1)1}\right)^2},$$

$$B_{12}^* = \frac{\sum\left(\tilde{z}_{t1} - \alpha\,\tilde{z}_{(t-1)1}\right) z_{(t-1)0}}{\sum\left(\tilde{z}_{t1} - \alpha\,\tilde{z}_{(t-1)1}\right)^2} \frac{\sum\left(z_{t1} - \alpha\, z_{(t-1)1}\right) z_{(t-1)0}}{\sum\left(z_{(t-1)0} - \tau\, z_{t1}\right)^2},$$

and

$$B_{13}^* = \frac{\sum\left(z_{t1} - \alpha\, z_{(t-1)1}\right) z_{(t-1)0}}{\sum\left(z_{(t-1)0} - \tau\, z_{t1}\right)^2} \frac{\sum\left(z_{t1} - \alpha\, z_{(t-1)1}\right) z_{(t-1)0}}{\sum\left(\tilde{z}_{t1} - \alpha\,\tilde{z}_{(t-1)1}\right)^2},$$

we have

(6.29) $B_1^* = o_p\left(n^{-1/2}\right) + o_p\left(n^{-1/2}\right) + o_p\left(n^{-1/2}\right) = o_p\left(n^{-1/2}\right)$,

since B_{11}^* could be treated as Term B and B_{12}^* as Term C of Lemma 6.3.2. Since we need only the same rate of convergence, it does not matter that the indices in the nominator and denominator differ.

For B_{13}^* we note that

$$\frac{\sum\left(z_{t1}-\alpha\, z_{(t-1)1}\right)z_{(t-1)0}}{\sum\left(z_{(t-1)0}-\tau\, z_{t1}\right)^2}=O_p(1)$$

because of covariance ergodicity and

$$\frac{\sum\left(z_{t1}-\alpha\, z_{(t-1)1}\right)z_{(t-1)0}}{\sum\left(\widetilde{z}_{t1}-\alpha\, \widetilde{z}_{(t-1)1}\right)^2}=o_p\left(n^{2\widetilde{d}_1-1}\right)=o_p\left(n^{-1/2}\right)$$

according to (P4) and (P5) of Lemma 6.1. The last equality holds, since according to our assumptions, $\widetilde{d}_1>0.75$. Hence, $B_{13}^*=o_p\left(n^{-1/2}\right)$.

Collecting the results from (6.26) to (6.29) yields $B^*=o_p\left(n^{-1/2}\right)$.

For Term C^* we have

$$\begin{aligned}
C^* &= \frac{\sum\left(\widetilde{z}_{t1}-\alpha\, \widetilde{z}_{(t-1)1}\right)\left(z_{(t-1)0}-\tau\, z_{t1}\right)}{\sum\left(\widetilde{z}_{t1}-\alpha\, \widetilde{z}_{(t-1)1}\right)^2}\frac{\sum\left(z_{(t-1)0}-\tau\, z_{t1}\right)\left(z_{t0}-\alpha\, z_{(t-1)0}\right)}{\sum\left(z_{(t-1)0}-\tau\, z_{t1}\right)^2}\\
&= \frac{\sum\left(\widetilde{z}_{t1}-\alpha\, \widetilde{z}_{(t-1)1}\right)z_{(t-1)0}}{\sum\left(\widetilde{z}_{t1}-\alpha\, \widetilde{z}_{(t-1)1}\right)^2}\frac{\sum\left(z_{(t-1)0}-\tau\, z_{t1}\right)\left(z_{t0}-\alpha\, z_{(t-1)0}\right)}{\sum\left(z_{(t-1)0}-\tau\, z_{t1}\right)^2}\\
&\quad -\tau\, \frac{\sum\left(\widetilde{z}_{t1}-\alpha\, \widetilde{z}_{(t-1)1}\right)z_{t1}}{\sum\left(\widetilde{z}_{t1}-\alpha\, \widetilde{z}_{(t-1)1}\right)^2}\frac{\sum\left(z_{(t-1)0}-\tau\, z_{t1}\right)\left(z_{t0}-\alpha\, z_{(t-1)0}\right)}{\sum\left(z_{(t-1)0}-\tau\, z_{t1}\right)^2}
\end{aligned}$$

and hence,

(6.30) $C^*=C_1^*-\tau\, C_2^*$,

where

$$\begin{aligned}
C_1^* &= \frac{\sum\left(\widetilde{z}_{t1}-\alpha\, \widetilde{z}_{(t-1)1}\right)z_{(t-1)0}}{\sum\left(\widetilde{z}_{t1}-\alpha\, \widetilde{z}_{(t-1)1}\right)^2}\frac{\sum\left(z_{(t-1)0}-\tau\, z_{t1}\right)\left(z_{t0}-\alpha\, z_{(t-1)0}\right)}{\sum\left(z_{(t-1)0}-\tau\, z_{t1}\right)^2}\\
&= \frac{\sum\left(\widetilde{z}_{t1}-\alpha\, \widetilde{z}_{(t-1)1}\right)z_{(t-1)0}}{\sum\left(\widetilde{z}_{t1}-\alpha\, \widetilde{z}_{(t-1)1}\right)^2}\frac{\sum\left(z_{(t-1)0}-\tau\, z_{t1}\right)\left(z_{t0}-\alpha\, z_{(t-1)0}\right)}{\sum\left(z_{(t-1)0}-\tau\, z_{t1}\right)^2}\\
&\quad +\frac{\sum\left(z_{t1}-\alpha\, z_{(t-1)1}\right)z_{(t-1)0}}{\sum\left(\widetilde{z}_{t1}-\alpha\, \widetilde{z}_{(t-1)1}\right)^2}\frac{\sum\left(z_{(t-1)0}-\tau\, z_{t1}\right)\left(z_{t0}-\alpha\, z_{(t-1)0}\right)}{\sum\left(z_{(t-1)0}-\tau\, z_{t1}\right)^2}
\end{aligned}$$

and

$$C_2^*=\frac{\sum\left(\widetilde{z}_{t1}-\alpha\, \widetilde{z}_{(t-1)1}\right)z_{t1}}{\sum\left(\widetilde{z}_{t1}-\alpha\, \widetilde{z}_{(t-1)1}\right)^2}\frac{\sum\left(z_{(t-1)0}-\tau\, z_{t1}\right)\left(z_{t0}-\alpha\, z_{(t-1)0}\right)}{\sum\left(z_{(t-1)0}-\tau\, z_{t1}\right)^2}.$$

First we note

(6.31) $C_2^*=o_p\left(n^{-1/2}\right)$, since C_2^* could be treated as Term C of Lemma 4.3.2.

Further, if we define

(6.32) $C_1^* = C_{11}^* + C_{12}^*$,

where

$$C_{11}^* = \frac{\sum\left(\tilde{z}_{t1} - \alpha\, \tilde{z}_{(t-1)1}\right) z_{(t-1)0}}{\sum\left(\tilde{z}_{t1} - \alpha\, \tilde{z}_{(t-1)1}\right)^2} \frac{\sum\left(z_{(t-1)0} - \tau\, z_{t1}\right)\left(z_{t0} - \alpha\, z_{(t-1)0}\right)}{\sum\left(z_{(t-1)0} - \tau\, z_{t1}\right)^2}$$

and

$$C_{12}^* = \frac{\sum\left(z_{t1} - \alpha\, z_{(t-1)1}\right) z_{(t-1)0}}{\sum\left(\tilde{z}_{t1} - \alpha\, \tilde{z}_{(t-1)1}\right)^2} \frac{\sum\left(z_{(t-1)0} - \tau\, z_{t1}\right)\left(z_{t0} - \alpha\, z_{(t-1)0}\right)}{\sum\left(z_{(t-1)0} - \tau\, z_{t1}\right)^2},$$

we have

(6.33) $C_1^* = o_p\left(n^{-1/2}\right) + o_p\left(n^{-1/2}\right) = o_p\left(n^{-1/2}\right)$,

since C_{11}^* has the same structure as B_{12}^* and C_{12}^* has the same structure as B_{13}^*.

Collecting the results from (6.30)-(6.33) yields $C^* = o_p\left(n^{-1/2}\right)$.

Finally, we have

$$n^{1/2} \frac{\sum\left(\tilde{z}_{t1} - \alpha\, \tilde{z}_{(t-1)1}\right)\left(z_{t0} - \alpha\, z_{(t-1)0}\right)}{\sum\left(\tilde{z}_{t1} - \alpha\, \tilde{z}_{(t-1)1}\right)^2} \xrightarrow{\;p\;} 0,$$

since we have the same structure as in Lemma 6.3.2 (iv).

This completes the proof of Lemma 6.4.2.

□

7 The special I(1)/I(0) case

In this chapter, we check how asymptotic results earlier derived for the standard case with I(1) regressors and I(0) errors fit into our work. The natural question arises whether these results can be deducted from our theorems or not. We will see, for the most part we can say so. In the following, Models (M1)-(M3) will be reconsidered, and the results from Chapter 6 are specialized to the I(1)/I(0) case. For ease of exposition Lemma 6.1 will be slightly modified. Specifically, we set $C(1)=1$ in (P2) with no loss of generality, since the covariance matrix Σ is unrestricted. Hence, we have $\mathbf{\Omega}=\Sigma$ and $c_i(1)=1$, for i = 0, ..., p, in (P7).

7.1 Regression model with one regressor

The asymptotics in Model (M1) for the I(1)-I(0) case are given by Theorem 6.2.2 (ii). We have

$$(7.1)\qquad n\left(\hat{\beta}-\beta\right)\xrightarrow{d}\frac{\Xi_{10}+\lambda_{10}^{*}+\gamma_{10}}{\int_0^1 B_{d_1}^2(s)\,ds}.$$

Now according to Theorem 5.5,

$$\Xi_{10}=\Xi_{1,10}+\Xi_{2,10},$$

where

$$\Xi_{1,10}=\int_{-\infty}^{1}Q(r)\,dB_{(\varepsilon)}^{(0)}(r),\ \Xi_{2,10}=\int_{-\infty}^{1}H(p)\,dB_{(\varepsilon)}^{(1)}(p),$$

$$Q(r)=\int_{-\infty}^{r}A_{10}(r,p)\,dB_{(\varepsilon)}^{(1)}(p),\ H(p)=\int_{-\infty}^{p}E_{10}(p,r)\,dB_{(\varepsilon)}^{(0)}(r).$$

If we assume $d_1=d_0=0$, then

$$A_{10}(r,p)=1-\mathrm{I}_{\{r<0\}}-\mathrm{I}_{\{p<0\}}+\mathrm{I}_{\{p<0\}}\,\mathrm{I}_{\{r<0\}},\ for\ -\infty<p\le r\le 1,$$

$$E_{21}(p,r)=0,\ for\ -\infty<r\le p\le 1.$$

and hence, $\Xi_{10}=\Xi_{1,10}$, where

$$\Xi_{1,10}=\int_0^1 B_{(\varepsilon)}^{(1)}(r)\,dB_{(\varepsilon)}^{(0)}(r),\ \text{since}\ Q(r)=\int_0^r dB_{(\varepsilon)}^{(1)}(p)=B_{(\varepsilon)}^{(1)}(r),\ \text{for}\ 0\le r\le 1.$$

Altogether, (7.1) simplifies to

$$(7.2)\qquad n\left(\hat{\beta}-\beta\right)\xrightarrow{d}\frac{\int_0^1 B_{(\varepsilon)}^{(1)}(r)\,dB_{(\varepsilon)}^{(0)}(r)}{\int_0^1\left[B_{(\varepsilon)}^{(1)}\right]^2(r)\,dr}.$$

Result (7.2) was first derived in Theorem 3.1 of Park and Phillips (1988) under quite similar assumptions. In fact, these authors assumed the process $\mathbf{z}_t = (e_t, \Delta x_t)^T$ to be driven by a process which satisfies a multivariate invariance principle, i.e.,

$$(7.3) \qquad \frac{1}{\sqrt{n}} \sum_{t=1}^{[ns]} \mathbf{z}_t \Rightarrow \mathbf{B}(s), \text{ for } 0 \le s \le 1,$$

where $\mathbf{B}(s)$ is ordinary Brownian motion on the unit interval with covariance matrix $\mathbf{\Omega}$. Insofar they were less precise in their model assumptions but allowed for more general processes than VARMA processes. Further, they did not assume any diagonal structure of the model process $\mathbf{z}_t$. Hence, in their paper $\mathbf{\Omega}$ cannot be set equal to Σ without any loss of generality.

7.2 Regression model with two regressors

The asymptotics in Model (M2) for the I(1)/I(0) case are given partly by Theorem 6.3.1, if we assume $d_2 = d_1 = d_0 = 0$. According to Lemma 6.1 including the foregoing modifications, we have

$$(7.4) \qquad \frac{1}{n}\sum x_{t2}e_t = \frac{1}{n}\sum x_{(t-1)2}e_t + \frac{1}{n}\sum \Delta x_{t2}e_t \xrightarrow{d} \Xi_{20} + \lambda_{20}^* + \gamma_{20},$$

$$(7.5) \qquad \frac{1}{n}\sum x_{t2}x_{t1} = \frac{1}{n}\sum x_{(t-1)2}x_{t1} + \frac{1}{n}\sum \Delta x_{t2}x_{t1} \xrightarrow{d} \Xi_{21} + \lambda_{21}^* + \gamma_{21}$$

and

$$(7.6) \qquad n^{-2}\sum_{t=1}^{n} x_{t2}^2 \xrightarrow{d} \int_0^1 B_{d_2}^2(s)\,ds.$$

For the moment we assume joint weak convergence of (7.4) through (7.7), which has not been formally proved yet and is not covered by Theorem 4.1 of Davidson and Hashimzade (2007b). Since

$$(7.7) \qquad \hat{\beta}_2 - \beta_2 = \frac{\dfrac{\sum x_{t2}e_t}{\sum x_{t2}^2} - \dfrac{\sum x_{t2}x_{t1}}{\sum x_{t2}^2}\dfrac{\sum x_{t1}e_t}{\sum x_{t1}^2}}{1 - \dfrac{\sum x_{t2}x_{t1}}{\sum x_{t1}^2}\dfrac{\sum x_{t2}x_{t1}}{\sum x_{t2}^2}},$$

we get

$$(7.8) \qquad n\left(\hat{\beta}_2 - \beta_2\right) \xrightarrow{d} \frac{\Xi_{20} - \gamma_{10}\gamma_{11}^{-1}\Xi_{21} + \lambda_{20}^* + \gamma_{20} - \gamma_{10}\gamma_{11}^{-1}\left(\lambda_{21}^* + \gamma_{21}\right)}{\int_0^1 B_{d_2}^2(s)\,ds}.$$

Note that the denominator of (7.7) converges in probability to 1.

Following the same lines as in Section 7.1, (7.8) can be simplified to

(7.9)

$$n\left(\hat{\beta}_2-\beta_2\right) \xrightarrow{d} \frac{\int_0^1 B_{(\varepsilon)}^{(2)}(s)\,dB_{(\varepsilon)}^{(0)}(s)-\gamma_{10}\gamma_{11}^{-1}\int_0^1 B_{(\varepsilon)}^{(2)}(s)\,dB_{(\varepsilon)}^{(1)}(s)+\lambda_{20}^*+\gamma_{20}-\gamma_{10}\gamma_{11}^{-1}\left(\lambda_{21}^*+\gamma_{21}\right)}{\int_0^1\left[B_{(\varepsilon)}^{(2)}\right]^2(s)\,ds}.$$

Further, according to Theorem 6.3.1 we have

(7.10) $$n^{1/2}\left(\hat{\beta}_1-\beta_1-\frac{\gamma_{10}}{\gamma_{11}}\right) \xrightarrow{d} \gamma_{11}^{-1}N(0,V).$$

The results in (7.9) and (7.10) were first derived in Theorems 3.1 and 3.2, respectively, of Park and Phillips (1989) under quite similar assumptions. They assumed $\mathbf{z}_t=\left(e_t, x_{t1}, \Delta x_{t2}\right)^T$ to be driven by a process which satisfies a multivariate invariance principle according to (7.3). As before, they were less precise in their model assumptions but allowed for more general processes than VARMA processes. Further, they did not assume any diagonal structure of the model process $\mathbf{z}_t$.

7.3 Dynamic regression model

The asymptotics in Model (M3) for the I(1)-I(0) case are given by Theorem 6.4.1, if we assume $d_1=d_0=0$, i.e.,

(7.11) $$n^{1/2}\begin{pmatrix}\hat{\alpha}-\alpha-b^*\\ \hat{\beta}-\beta+\tau\, b^*\end{pmatrix} \xrightarrow{d} \begin{pmatrix}Z\\ -\tau\, Z\end{pmatrix},$$

where $Z \sim \left(b_2^*\right)^{-1}N\left(0,V^*\right)$ with $V^*>0$ and

$$b^*=\frac{\gamma_{00}(1)-\alpha\,\gamma_{00}-\tau\,\gamma_{10}+\alpha\,\tau\,\gamma_{10}(1)}{\gamma_{00}-2\,\tau\,\gamma_{10}(1)+\tau^2\gamma_{11}} \quad \text{with} \quad \tau=\frac{\beta}{1-\alpha}.$$

If we additionally assume z_t and e_t to be independent and hence, $\sigma_{10}=0$, we get

$$b^{**}=\frac{\gamma_{00}(1)-\alpha\,\gamma_{00}}{\gamma_{00}+\tau^2\gamma_{11}}.$$

Result (7.11) for the case $\sigma_{10}=0$ was earlier proved in Theorem 1 of Maekawa et al. (1996). In contrast, these authors provide only the marginal limit distributions, and it is not evident that we asymptotically obtain a singular normal distribution. Furthermore, they assume z_t to be driven by a simple random walk and e_t by an independent stationary AR(1) process. Hence, our theorem is both more precise and more general.

8 Suggestions for further research

So far, we have not dealt with inference yet. Within the I(1)/I(0) framework Park and Phillips (1989) also derive limiting distributions of test statistics for linear hypotheses about the regression coefficients. In a similar way, Maekawa et. al. (1996) examine the asymptotic properties of t-ratios and other types of test statistics in the context of Model (M3). The problem of hypotheses testing in a fractional, possibly nonstationary, framework has to be deferred to future research for lack of important prerequisite results such as the estimation of the covariance matrix of OLS. Further, making inference would require knowledge about the integration orders of the model processes. In practice, these are typically regarded as unknown and need to be estimated in a prior step. Methods of estimating the integration order in parametric settings such as ARFIMA models are suggested by Beran (1995).

Theorem 6.4.1 extends Theorem 1 of Maekawa et. al. (1996) to the fractional case. The aforementioned authors show that their model can be reduced to a regression with cointegrated regressors for which a general asymptotic theory was given in Section 5.2 of Park and Phillips (1989). It seems to be possible to connect Theorem 6.4.1 with the concept of 'fractional cointegration' in a similar way. Obviously, y_{t-1} and z_t in (M3) exhibit the same integration order $\tilde{d}_1$, whereas the error is integrated of order $d_0 < \tilde{d}_1$. Hence, upon an appropriate definition of 'fractional cointegration', it should be possible to interpret the setting of Theorem 6.4.1 as a regression with fractionally cointegrated regressors. One objective in this context could be to provide sufficient conditions for asymptotic normality when regressors are fractionally cointegrated.

With exception of Theorem 6.2.1, all results in Chapter 6 assume the model driving VARFIMA process to exhibit diagonal structure according to (P7) in Lemma 6.1. Hence, the innovations of each process affect the other one only via contemporaneous correlations. This might be a quite restrictive assumption to be removed in future. In fact, Davidson and Hashimzade (2007b) note, that a multivariate analysis would typically invoke a vector Wold representation, such as

$$\begin{bmatrix} z_{t1} \\ z_{t2} \end{bmatrix} = \begin{bmatrix} \Delta^{-d_1}(L) & 0 \\ 0 & \Delta^{-d_2}(L) \end{bmatrix} \begin{bmatrix} \theta_{11}(L) & \theta_{12}(L) \\ \theta_{21}(L) & \theta_{22}(L) \end{bmatrix} \begin{bmatrix} w_{t1} \\ w_{t2} \end{bmatrix}$$

in the bivariate case. Extending the results to non-diagonal structures is an application of the continuous mapping theorem to the limit distributions of (P7). In such a setting the term G_n

in (5.4) would become a sum of four terms involving respectively the driving pairs

$\{w_{t1}, w_{t1}\}$, $\{w_{t1}, w_{t2}\}$, $\{w_{t2}, w_{t1}\}$ and $\{w_{t2}, w_{t2}\}$.

Our analysis could be applied to each of these cases then. Hence, allowing for non-diagonal structures should be rather an essential complication of notation than a theoretical problem. Finally, our asymptotic results apply to infinitely large samples. In practice, we have to deal with finite samples. Literature seems to be much less extensive with respect to finite sample studies on nonstationary OLS asymptotics, specifically, for the fractional case. This might be partly due to the fact that simulation studies involving nonstationary processes and stochastic integrals are not a simple task. Considering I(d) processes complicates matters. Nevertheless, asymptotic results are of minor use, if they lack any knowledge about their validity in finite samples.

References

Akonom, J. and Gourieroux, C. (1987):
A functional central limit theorem for fractional processes.
Technical Report #8801, CEPREMAP, Paris.
Anderson, T.W. (1959):
On asymptotic distributions of estimates of parameters of stochastic difference equations.
The Annals of Mathematical Statistics, 30, 676-687.
Anderson, T. W. and Taylor, John B. (1979):
Strong consistency of least squares estimates in dynamic models.
The Annals of Statistics, 7, 484-489.
Anderson, T.W., and Taylor, John B. (1976):
Strong consistency of least squares estimates in normal linear regression.
The Annals of Statistics, 4, 788-790.
Beran, J. (1994):
Statistics for long-memory processes.
New York: Chapman & Hall.
Beran, J. (1995):
Maximum likelihood estimation of the differencing parameter for invertible short- and long-memory ARIMA models. Journal of the the Royal Statistical Society, Series B, 57, 659-672.
Beran, J. and Künsch, H. (1985):
Location estimators for processes with long range dependence.
Technical Report 40, Seminar für Statistik, ETH Zürich.
Billingsley, P. (1968):
Convergence of probability measures.
New York: Wiley.
Bujosa, A., Bjuosa, M. and García-Ferrer, A. (2001):
A note on the pseudo-spectra and the pseudo-covariance generating functions of ARMA processes. Universidad Complutense de Madrid, Facultad de Ciencias Económicas y Empresariales. Working Paper.
Brockwell, P.J. and Davis, R.A. (1991):
Time Series: Theory and Methods, 2nd edn.
New York: Springer-Verlag.
Chan, N.H. and Wei, C.Z. (1988):
Limiting distributions of least squares estimates of unstable autoregressive processes.
Annals of Statistics, 16, 367-401.
Chan and Terrin (1995):
Inference for unstable long-memory processes with applications to fractional unit root autoregressions. The Annals of Statistics, 23, 1662-1683.
Chen, G.J.,Lai, T.L., and Wei,C.Z. (1981):
Convergence Systems and strong consistency of least squares estimates in regression models.
Journal of Multivariate Analysis, 11, 319-33.
Choy and Taniguchi (2001):
Stochastic regression model with dependent disturbances.
Journal of Time Series Analysis, 22, 175-196.
Chung, Ching-Fan (2001):
Calculating and analyzing impulse responses for the vector ARFIMA model.
Economics Letters, 71, 17-25.

Chung, Ching-Fan (2002):
Sample means, sample autocovariances, and linear regression of stationary multivariate long memory processes. Econometric Theory, 18 (1), 51-78.
Csőrgo, S. and Mielniczuk, J. (1995):
Distant long-range dependent sums and regression estimation.
Stochastic Processes and their Applications, 59, 143-155.
Dahlhaus (1995):
Efficient location and regression estimation for long range dependent regression models.
The Annals of Statistics, 23, 1029-1047.
Davidson, J.E.H. (1994):
Stochastic Limit Theory.
Oxford: Oxford University Press.
Davidson, J. (2007):
When is a time series I(0)?
School of Business and Economics, University of Exeter. Discussion paper.
At *http://www.people.ex.ac.uk/jehd201/research*. Accessed 01/20/08.
Davidson, J. and de Jong, Robert M. (2000a):
The functional central limit theorem and weak convergence to stochastic integrals I:
weakly dependent processes. Econometric Theory, 16, 621-642.
Davidson, J. and de Jong, Robert M. (2000b):
The functional central limit theorem and weak convergence to stochastic integrals II:
fractionally integrated processes. Econometric Theory, 16, 643-666.
Davidson, J. and Hashimzade, N. (2007a):
Type I and type II fractional Brownian motions: a reconsideration.
School of Business and Economics, University of Exeter. Discussion Paper.
At *http://www.people.ex.ac.uk/jehd201/research*. Accessed 01/20/08.
Davidson, J. and Hashimzade, N. (2007b):
Representation and weak convergence of stochastic integrals with fractional integrator processes. University of Exeter. Discussion Paper.
At *http://www.people.ex.ac.uk/jehd201/research*. Accessed 01/20/08.
Davidson, J. and Hashimzade, N. (2008):
Alternative frequency and time domain versions of fractional Brownian motion.
Econometric Theory, 24, 256-293.
Davydov, Y. (1970):
The invariance principle for stationary processes.
Theory of Probability and Applications, 15, 487-498.
Deo (1997):
Asymptotic theory for certain regression models with long memory errors.
Journal of Time Series Analysis, 18, 385-393.
Dickey, D.A. (1976):
Estimation and hypothesis testing in nonstationary time series.
Ph.D. thesis, Iowa State University.
Dickey, D.A. and Fuller, W.A. (1979):
Distribution of the estimators for autoregressive time series with a unit root.
Journal of the American Statistical Association, 74, 427-431.
Dickey, D.A. and Fuller, W.A. (1981):
Likelihood ratio statistics for autoregressive time series with a unit root.
Econometrica, 49, 1057-1072.
Dolado, J. J. and Marmol, F. (2004):
Asymptotic inference results for multivariate long-memory processes.
Econometrics Journal, 2004, 168-190.

Donsker, M.D. (1951):
An invariance principle for certain probability limit theorems.
Memoirs of the American Mathematical Society, 79, 355-367.
Doukhan, P., Oppenheim, G., and Taqqu, M. (2003):
Theory and Applications of Long-Range Dependence.
Boston: Birkhäuser Press.
Drygas, Hilmar (1976):
Weak and strong consistency of the least squares estimators in regression models.
Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, 34, 119-127.
Eicker (1965):
Ein zentraler Grenzwertsatz für Summen von Variablen einer verallgemeinerten linearen Zufallsfolge. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, 3, 317-327.
Eicker, F. (1967):
Limit theorems for regressions with unequal and dependent errors.
In: *Proceedings of the Fifth Berkeley Symposium in Mathematical Statistics and Probability*, Vol. 1, University of California Press, 59-82.
Embrechts, P., and Maejima, M (2002):
Selfsimilar Processes.
Princeton, N.J.: Princeton University Press.
Engle, R.F. and Granger, C.W.J. (1987):
Cointegration and error correction: representation, estimation, and testing.
Econometrica 55, 251-267.
Evans, G., and Savin, N. (1981):
Testing for unit roots: I.
Econometrica, 49, 753-779.
Evans, G., and Savin, N. (1984):
Testing for unit roots: II.
Econometrica, 52, 1241-1269.
Fox and Taqqu (1987):
Multiple stochastic integrals with dependent integrators.
Journal of multivariate analysis, 21, 105-127.
Giraitis, L., Koul, H. L. and Surgailis, D. (1996):
Asymptotic normality of regression estimators with long memory errors.
Statistics & Probability Letters, 29, 317-335.
Gorodetskii, V.V. (1977):
On convergence to semi-stable Gaussian processes.
Theory of Probability and Applications, 22, 498-508.
Gourieroux, C. and Monfort, A.M. (1997):
Time series and dynamic models.
Cambridge University Press, Cambridge, UK.
Granger, C.W.J. and Joyeux, R. (1980):
An introduction to long memory time series models and fractional differencing.
Journal of Time Series Analysis, 1, 15-29.
Grenander, Ulf (1954):
On the estimation of regression coefficients in the case of an autocorrelated disturbance.
The Annals of Mathematicals Statistics, 25, 252-272.
Grenander Ulf and Rosenblatt Murray (1957):
Statistical analysis of stationary time series.
Stockholm: Almquist and Wiksell.

Hannan, E.J. (1961):
A central limit theorem for systems of regressions.
Proceedings of the Cambridge Philosophical Society, 57, 583-588.
Hannan, E.J. (1970):
Multiple Time Series.
New York: Wiley.
Hannan, E.J. (1976):
The asymptotic distribution of serial covariances.
The Annals of Statistics, 4, 396-399.
Helson, H., and Sarason, D. (1967):
Past and future.
Mathematica Scandinavia, 21, 5-16.
Hosking, J.R.M (1981):
Fractional differencing.
Biometrika, 68, 165-176.
Hosking, J.R.M (1996):
Asymptotic distributions of the sample mean, autocovariances, and autocorrelations of long memory time series. Journal of Econometrics, 73, 261-284.
Koul, Hira L. (1992):
M-estimators in linear models with long range dependent errors.
Statistics & Probability Letters, 14, 153-164.
Koul, H. L. and Mukherjee, K. (1993):
Asymptotics of R-, MD- and LAD-estimators in linear regression models with long range dependent errors. Proaibility Theory and Related Fields, 95, 535-553.
Lai, T.L., Robbins, Herbert, and Wei, C.Z. (1979):
Strong consistency of least squares estimates in multiple regression, II.
Journal of Multivariate Analysis, 9, 343-361.
Lai, T.L., and Wei, C.Z. (1982a):
Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems. The Annals of Statistics, 10, 154-166.
Lai, T.L., and Wei, C.Z. (1982b):
Asymptotic properties of projections with applications to stochastic regression problems.
Journal of Multivariate Analysis, 12, 346-370.
Levy, P. (1953):
Random functions: general theory with special reference in Laplacian random functions.
Univ. California Publ. Statist., 1, 331-390.
Lütkepohl, Helmut (2005):
New Introduction to Multiple Time Series Analysis.
Berlin: Springer.
Maekawa, K., Yamamoto, T., Takeuchi, Y. and Hatanaka, M (1996):
Estimation in dynamic regression with an integrated process.
Journal of Statistical Planning and Inference, 49, 279-303.
Mandelbrot, M.M. and Van Ness, J.W. (1968):
Fractional brownian motions, fractional noises and applications.
SIAM Reviews, 10, 422-437.
Mann, H.B. and Wald, A. (1943):
On the statistical treatment of linear stochastic difference equations.
Econometrica, 11, 172-200.
Marinucci and Robinson (1999):
Alternative forms of fractional Brownian motion.
Journal of Statistical Planning and Inference, 80, 111-122.

Marinucci and Robinson (2000):
Weak convergence of multivariate fractional processes.
Stochastic Processes and their Applications, 86, 103-120.
Marinucci and Robinson (2001):
Semiparametric fractional cointegration analysis.
Journal of Econometrics, 105, 225-247.
Phillips, P.C.B. and Durlauf, S. (1986):
Multiple time series regression with integrated processes.
Review of Economic Studies, 53, 473-495.
Phillips, P.C.B. (1987):
Time series regression with a unit root.
Econometica, 55, 277-302.
Phillips, P.C.B. (1988):
Weak convergence to the matrix stochastic integral $\int_0^1 B\, dB$.
Journal of Multivariate Analysis, 24, 252-264.
Park, J. and Phillips, P.C.B. (1988):
Statistical inference in regressions with integrated processes: part I.
Econometric Theory, 4, 468-497.
Park, J. and Phillips, P.C.B. (1989):
Statistical inference in regressions with integrated processes: part II.
Econometric Theory, 5, 95-131
Robinson, P.M. (2003):
„Long Memory Time Series“ in Time Series with Long Memory“ (P.M. Robinson, ed.),
Oxford University Press, Oxford (2003), 1-48.
Robinson, P.M. (1994):
Time series with strong dependence. In C.A. Sims (ed.), *Advances in Econometrics: Sixth World Congress*. Cambridge University Press.
Robinson, P.M. and Hidalgo, J.F. (1997):
Time series regression with long-range dependence.
The Annals of Statistics, 25, 77-104.
Robinson, P.M., and Marinucci, D. (2001):
Narrow-band analysis of nonstationary processes.
The Annals of Statistics, 29, 947-986.
Robinson, Peter M., and Yajima, Yoshihiro (2002):
Determination of cointegrating rank in fractional systems.
Journal of Econometrics, 106, 217-241.
Rosenblatt, M. (1961):
Independence and dependence. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, 2, 431-443. Berkeley: University of California Press.
Samorodnitsky, G. (2006):
Long range dependence.
Foundations and Trends in Stochastic Systems, 1 (3), 163-257.
Silveira, G. (1991):
Contributions to strong approximations in time series with applications in nonparametric statistics and functional central limit theorems. Ph.D. thesis, University of London.
Sims, C.A., Stock, J.H. and Watson, M.W. (1990):
Inference in linear time series models with some unit roots.
Econometrica, 58, 113-144.

Solo, V. (1981):
Strong consistency of least squares estimators in regression with correlated disturbances.
The Annals of Statistics, 9, 689-693.
Sowell (1990):
The fractional unit root distribution.
Econometrica, 58, 495-505.
Stock, J.H. (1987):
Asymptotic properties of least squares estimators of cointegrating vectors.
Econometrica, 55, 1035-1056.
Taqqu, M.S. (1975):
Weak convergence to fractional Brownian motion and to the Rosenblatt process.
Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, 31, 287-302.
Tudor, C.A. (2007):
Analysis of the Rosenblatt process. Working paper, SAMOS/MATISSE,
Centre d'Economie de La Sorbonne, Université de Panthéon-Sorbonne Paris 1.
Usami, Yoshihiro, and Huzii, Mituaki (1995):
Estimation of coefficients of time series regression with a nonstationary error process.
Journal of Time Series Analysis, 16, 105-118.
Usami, Yoshihiro (2002):
Some asymptotic properties of the least squares estimators of a polynomial regression with heteroskedastic error. Communications in Statistics - Theory and Methods, 31, 625-637.
Wei, C. Z. (1985):
Asymptotic properties of least-squares estimates in stochastic regression models.
The Annals of Statistics, 13, 1498-1508.
White (1958):
The limiting distribution of the serial correlation coefficient in the explosive case.
Annals of Mathematical Statistics, 29, 1188-1197.
White, John S. (1959):
The limiting distribution of the serial correlation coefficient in the explosive case, II.
Annals of Mathematical Statistics, 30, 831-834.
White, Halbert (2001):
Asymptotic theory for econometricians.
London: Academic Press.
Yajima (1988):
On estimation of a regression model with long-memory stationary errros.
The Annals of Statistics, 16, 791-807.
Yajima (1991):
Asymptotic properties of the LSE in a regression model with long-memory stationary errors.
The Annals of Statistics, 19, 158-177.