Modeling Multivariate Time Series with Fractional Integration in Macroeconomics and Finance

Dissertation zur Erlangung des Grades eines Doktors der Wirtschaftswissenschaft

eingereicht an der

Fakultät für Wirtschaftswissenschaften
der Universität Regensburg

vorgelegt von Roland Weigand

Berichterstatter:

Prof. Dr. Rolf Tschernig, Universität Regensburg
Prof. Dr. Enzo Weber, Universität Regensburg

Tag der Disputation: 17. Juli 2014

Contents

List of Figures

List of Tables

1 Introduction and Overview

This thesis develops new approaches for modeling multivariate time series. It covers four essays on different setups from macroeconomics and finance. In the introduction, first the general framework of multivariate time series models is introduced with a mention of long-memory processes and factor models. Then, the empirical setups of the essays are outlined, while a third section states the contribution of this dissertation relative to the existing literature and gives an overview over the essays.

1.1 Methodological Framework

In this dissertation we consider collections of k economic variables $y_t = (y_{1t}, \ldots, y_{kt})'$ which are measured regularly at time periods $t = 1, \ldots, T$. The aim is a model-based statistical characterization of the dependencies between elements of y_t over time and among each other, which may serve different purposes. In macroeconomics, multivariate time series models have been used for analysing the effects of structural shocks which we denote by a vector ε_t. Elements of this process are associated with interpretable economic sources of fluctuations whose dynamic impacts on the observable series in y_t are studied. Moreover, such models are applied for forecasting. Here, the dependence of the economic variables over time is used to draw conclusions on the probabilistic behaviour of y_{T+h} in the future, given currently available information y_1, ..., y_T.

A very influential approach which has fostered numerous applications since its introduction to macroeconomics by Sims (1980) is the vector autoregressive (VAR) model

$$y_t = A_1 y_{t-1} + \ldots + A_p y_{t-p} + u_t, \qquad t = 1, \ldots, T; \tag{1.1}$$

see Lütkepohl (2005) for a textbook treatment. Here, u_t is an independent and identically distributed disturbance term with zero mean and covariance matrix Σ, denoted $u_t \sim IID(0, \Sigma)$, while correlation between observations over time is introduced by lagged values y_{t-l}, $l = 1, \ldots, p$, which enter the model equation through the $(k \times k)$ coefficient

matrices A_l. Nondiagonal A_l and contemporaneous correlation between the noise components u_t allow for dynamic linkages between the individual time series under consideration.

A large literature has considered the long-run behaviour of systems such as (1.1). In general, unit roots in the characteristic equation $|I - A_1 z - \ldots - A_p z^p| = 0$ induce nonstationary time series with stochastic trends, most prominently processes which are called integrated of order 1 or $I(1)$, but also $I(d)$ processes with $d = 2, 3, \ldots$ are possible. These can be rendered stationary autoregressive moving average (ARMA) processes, classified as $I(0)$, by d-th order differencing. Such difference-based modelling has been brought forward in applied statistics by Box and Jenkins (1970). The notion of long-run equilibria between $I(1)$ processes, so-called cointegration relations, has propelled the emphasis on the long-term properties of economic time series in recent decades, beginning with Granger (1983), Granger and Weiss (1983) and Engle and Granger (1987).

The VAR model is very popular among applied researchers but has shortcomings in several fields of empirical work. A major limitation of the VAR setup is the dichotomy between stationary $I(0)$ processes and the unit root case. This amounts to a sharp distinction between the rather extreme cases of short and perfect memory, respectively, and poses a considerable challenge to distinguish between these setups, especially when structural analyses and forecasts are sensitive with respect to the specification of the long-run properties. The unit root testing methodology often fails to provide a clear answer, and hence, different treatments of certain variables co-exist in the literature.

Additionally, processes of interest are often relatively high-dimensional which causes problems for VAR analysis. Given a k-dimensional time series y_t, the number of free parameters in the unrestricted VAR (1.1) is $k^2 p + 0.5k(k+1)$ and grows quadratically in k. Such a parameter affluence hinders precise estimation for larger k and limits the scope of VAR models to applications with a few variables only. Methodological progress has been spurred by these drawbacks of VARs for modeling the correlation structures, but also analyses from a new perspective have extended the scope of multivariate time series modeling; for example to time varying conditional covariance matrices; see Bollerslev, Engle, and Wooldridge (1988).

In this thesis, the mentioned shortcomings are tackled in different empirical setups, where the integration properties of the multivariate systems will be a primary focus. With this respect, fractionally integrated time series models are considered which avoid

the $I(0)/I(1)$ dichotomy by use of the fractional difference operator

$$\Delta^d = (1-L)^d = \sum_{j=0}^{\infty} \pi_j(d) L^j, \quad \pi_0(d) = 1, \quad \pi_j(d) = \frac{j-1-d}{j} \pi_{j-1}(d), \quad j = 1,2,\ldots,$$

where L denotes the lag operator ($Ly_t = y_{t-1}$) and d is a possibly non-integer scalar; see Baillie (1996) for a review. As a straightforward multivariate approach one may consider modeling the *fractionally differenced* process as a stationary VAR,

$$(I - A_1 L - \ldots - A_p L^p)(\Delta^{d_1} y_{1t}, \ldots, \Delta^{d_k} y_{kt})' = u_t, \qquad t = 1,\ldots,T; \tag{1.2}$$

see Sowell (1989) for an early treatment of fractionally integrated vector ARMA models and Nielsen (2004) who considers the fractionally integrated VAR (1.2) as a special case. The individual series y_{it} are integrated of fractional orders d_i which straightforwardly extends the classification into $I(d)$ variables for integer d. Analogously to the unit root literature, models with (fractional) cointegration have been developed and applied since the seminal work of Cheung and Lai (1993); see, e.g., Robinson and Hualde (2003) and Johansen and Nielsen (2012).

To cope with high dimensionality (large k) in time series analysis, which is also a key topic in this thesis, factor models have been a particularly successful field of research. Geweke (1977) extended classical cross-sectional factor analysis to a dynamic setup, while Quah and Sargent (1993) provided evidence on its applicability to the high-dimensional case. An elementary factor model is given by

$$y_t = \Lambda f_t + \varepsilon_t, \qquad t = 1,\ldots,T, \tag{1.3}$$

where f_t is an r-dimensional unobserved process which may be modelled as a VAR like (1.1), and where $r < k$. The assumptions on ε_t are crucial for a possible parsimonious parametrization for large k. The econometric literature has considered high-dimensional approximate factor models where, loosely speaking, f_t accounts for the bulk of cross-sectional correlation between elements of y_t (see Bai and Ng, 2008, for a survey). In contrast, the terminus of a statistical factor model has been used for setups where ε_t is serially uncorrelated and therefore, f_t accounts for the autocorrelation and hence for dimension-reduced dynamics in y_t (Pan and Yao, 2008; Lam, Yao, and Bathia, 2011; Lam and Yao, 2012).

1.2 Empirical Setups and Motivation

From an empirical point of view, this thesis is concerned with two distinct setups. The framework in chapters 2 and 3 is a bivariate time series with a focus on structural analysis, while the applications of chapters 4 and 5 aim on forecasting higher dimensional processes of realized covariance matrices.

In the first setup, structural shocks ε_t enter a system similar to (1.1) as $u_t = B\varepsilon_t$, and identification of the elements of B is achieved by imposing that a certain shock (ε_{2t} in our notation) has no long-term effect on a specified variable (y_{1t}). This is the well-known long-run restriction of Blanchard and Quah (1989) in the $I(1)$ case. Chapter 3 considers a system of output and prices. Here, the identification scheme is motivated by a possible association of restricted shocks (ε_{2t}) with aggregate demand shocks and unrestricted shocks (ε_{1t}) with shocks on the aggregate supply side of the economy. Other important empirical work has been done in similar frameworks. Most notably, applications to output and unemployment (Blanchard and Quah, 1989) as well as to productivity and hours worked (Gali, 1999) have aroused widespread interest.

In a large part of the related literature, the specification of long-run properties have been crucial for the outcomes. In the productivity and hours worked setup, a large debate has emerged over the effect of a technology shock (ε_{1t}) on hours worked (y_{2t}). Modeling hours worked as $I(1)$ and hence in first differences, a negative effect is found (Gali, 1999), while a specification in levels yields a contradictory conclusion with a positive effect; see Christiano, Eichenbaum, and Vigfusson (2003) and Christiano, Eichenbaum, and Vigfusson (2007). Similarly, for the output and price system considered in this thesis, Bayoumi and Eichengreen (1994) use an $I(1)$ specification for prices (y_{2t}) and find a negative effect of ε_{1t} (supply shocks) on prices, while in Quah and Vahey (1995), the effect of the non-core inflation shock (as they interpret ε_{1t}) on prices is positive but not significant for their $I(2)$ specification. Such ambiguities and their possible resolution by fractional integration techniques constitute the agenda for the first part of this thesis.

The second empirical setup of this thesis is concerned with modeling and forecasting the volatility of multiple financial assets. A relatively new literature has studied the dynamics of realized covariance matrices. Here, intra-day data on transaction prices are used to compute variances and covariances of asset returns within each trading day. Forecasts of the latter have been found valuable, e.g., for the purpose of portfolio selection (Liu, 2009).

From the perspective of dynamic modeling, standard approaches such as (1.1) exert the aforementioned problems. The strong persistence in the series has been tackled

by fractional integration techniques in the literature (Chiriac and Voev, 2011), while the typically high dimensionality of the series has led researchers to consider different forms of factor models; see Bauer and Vorkink (2011), Golosnoy and Herwartz (2012) and Gribisch (2013).

Another crucial distinction between the approaches considered in the literature is through the different transforms applied to the realized covariance matrices before fitting dynamic models. Chiriac and Voev (2011) use the elements of a triangular matrix square-root, while the matrix logarithm has been considered among others by Bauer and Vorkink (2011). Likewise, approaches that separate variance and correlation dynamics have been used; see Golosnoy and Herwartz (2012) and Halbleib and Voev (2011). In sum, the joint findings of long memory and factor structures poses a challenge to empirical researchers as does the coexistence of several transformations. The second part of this thesis is devoted to these problems.

1.3 Overview and Contribution

Despite the different setups, we employ similar strategies in the following chapters of this thesis. Implicitly or explicitly motivated by an empirical application, we identify key features of observed time series along with the problems which are most severe when standard methods are applied. In response, new modeling frameworks are developed which are well-suited to the empirical setups under consideration but also relevant in a wide range of other applications. For each of the models, econometric estimation poses further difficulties which are also tackled in this thesis.

The universe of existing approaches related to each of our empirical setups is characterized by certain dichotomies which turn out to be very influential for the outcomes of empirical work. Such discrete modeling choices may be harmful since they exclude possibly favorable in-between situations from the consideration set. Additionally, when deciding between the alternatives, there is typically no way to quantify an often substantial uncertainty regarding this decision in further steps of the analysis.

The methods we introduce are designed to overcome such dichotomies in the model specification process, most notably the distinction between integer integration orders. The key concept with this respect is the use of fractional integration techniques and their suitable adaptation to the characteristics of our empirical setups. Transformations of the variables are another important instance of such modeling decisions. We avoid the latter using a continuous framework in the spirit of Box and Cox (1964) which we

propose for the dynamic modeling of realized covariance matrices. To be more explicit on the contribution of this thesis we consider each of the essays in turn.

Long- versus Medium-Run Identification in Fractionally Integrated VAR Models. The following two chapters of this thesis extend structural VAR models with long-run restrictions to the case of fractional integration and hence overcome the dichotomy of integer orders which is typical for the related empirical literature. Chapter 2 is concerned with the interpretation of long-run restrictions to identify B when integration orders are non-integer. Whenever y_{1t} is integrated of an order less than one, these restrictions lose their original meaning from the $I(1)$ setup. In this case, the structural shocks do not exert a non-vanishing influence on y_{1t} regardless of the identification constraints. This case is empirically very relevant, since key macroeconomic variables such as output have been found mean-reverting $I(d)$ with $d < 1$; see, e.g., Diebold and Rudebusch (1989).

To obtain an economically meaningful restriction for this case, we consider a medium-run approach that constrains the variance contributions of ε_{2t} to $y_{1,t+h}$ over finite horizons h. For different such identification schemes, we investigate the case where relatively long horizons are appropriate from economic theory. Formally, we show that letting the horizon tend to infinity is equivalent to imposing the restriction of Blanchard and Quah (1989) introduced for the unit-root case. This finding justifies the use of a computationally straightforward approach in practice, while it retains interpretability of the resulting shocks and thus helps to overcome the dichotomy of integer integration orders in a range of empirically relevant situations.

Long-run Identification in a Fractionally Integrated System. In this paper, a model is proposed which has increased flexibility for structural analysis as compared to existing fractional processes. We derive the model's Granger representation and investigate the effects of long-run restrictions. In this way, we show that the impulse responses of y_{1t} to the restricted structural shock ε_{2t} are undesirably constraint for fractionally integrated VARs like (1.2), while our proposed FIVAR$_b$ model allows for very general patterns of decay.

Both in simulations and in empirical work, we find that enforcing integer integration orders can have severe consequences for impulse responses and hence, that it is indeed crucial to overcome this restrictive assumption in the current setup. Additionally, for the case of deterministic trends, a two-step estimation approach is proposed which outperforms the maximum likelihood estimator in a Monte Carlo study by Tschernig,

Weber, and Weigand (2013a).

In a system of U.S. real output and aggregate prices, shocks that are typically interpreted as demand disturbances have a very brief influence on gross domestic product if prices are modeled as $I(2)$ and exert a long-living effect if prices are taken to be $I(1)$. The fractional specification points to an in-between scenario, both in terms of the estimated integration orders and in the characterization of restricted impulse responses which are relatively short-living and hence closer to the $I(2)$ specification.

State Space Modeling of Fractional Cointegration. Chapter 4 of this thesis also considers multivariate fractionally integrated time series models, albeit with a different scope. A model setup is proposed which allows for fractional cointegration relations between the variables and is thus more general than the models of the previous chapters. In contrast to the autoregressive nature of (1.2), the model is formulated in terms of latent fractional and additive short memory components. This approach allows for a treatment of possibly nonstationary time series of different fractional integration orders. It features cointegration relations of different strengths and is therefore very flexible as compared to currently applied parametric models such as Robinson and Hualde (2003), Johansen (2008) or Avarucci and Velasco (2009). A further advantage is the clear interpretation of the cointegration properties in our representation.

The empirical setup of realized covariance matrices motivates the use of parsimonious models which are applicable to processes consisting of a large number of variance and covariance processes or transformations thereof. With a factor structure as in (1.3), our unobserved components formulation benefits the modeling of such high-dimensional series. We propose an according parametrization of the fractional components setup which is based on dimension reduction along the lines of statistical factor models and dynamic orthogonal components (Matteson and Tsay, 2011).

Estimation of our model is based on a state space representation where finite order ARMA approximations of the fractional processes are applied. This procedure outperforms the standard autoregressive or moving average truncation approach by providing a substantial reduction in state dimension for a desired approximation quality and is hence computationally convenient. Monte Carlo simulations document the successfulness of our approximation and show a reasonable performance of the proposed methods for cointegration modeling.

The methods are applied to realized covariance matrices using the dataset of Chiriac and Voev (2011). The sample consists of six U.S. stocks corresponding to a 21-dimensional time series of log variances and z-transformed correlations. It exhibits long memory

characteristics and a pronounced co-movement in the series' low-frequency dynamics. We find that common mutually orthogonal short- and long-memory components with two different fractional integration orders provide a reasonable fit to the data. An out-of-sample study shows that the fractional components model provides a superior forecasting accuracy compared to several competitor methods.

Matrix Box-Cox Models for Multivariate Realized Volatility. In the same framework of modeling realized covariance matrices, chapter 5 is concerned with data transforms and presents a flexible setup generalizing the Box-Cox approach (Box and Cox, 1964) to the matrix case. By proposing two specific models we face the otherwise discrete transformation decisions inherent to the modeling of realized covariance matrices. The matrix Box-Cox model of realized covariances (MBC-RCov) is based on transformations of the covariance matrix eigenvalues, while for the Box-Cox dynamic correlation (BC-DC) specification the variances are transformed individually and modeled jointly with the z-transformed correlations.

A key part of this paper is concerned with parameter estimation. A multivariate semiparametric estimator is proposed for the transformation parameters and feasible confidence intervals are derived. The estimator allows for a convenient two-step modeling strategy, first determining the transform, while specifying a dynamic model in a second step.

Since an emphasis in this empirical framework is on forecasting, we also provide a discussion of bias-corrected point forecasts for re-transformed covariance matrices and of density forecasts for daily returns. A simulation-based approach is proposed which is applicable also for other models from the realized covariance literature and which we find very valuable in an out-of-sample evaluation.

Using the same dataset as in chapter 4, our estimates suggest negative Box-Cox transformation parameters close to zero for both the MBC-RCov and the BC-DC model. The same values are supported by an out-of-sample forecast comparison. Here, the BC-DC model outperforms a wide range of competitor methods such as Cholesky-based and conditional Wishart models. In sum, modeling of log variances along with z-transformed correlations appears as a practically reasonable strategy, which also justifies the use of this transform in chapter 4.

Since this dissertation consists of four autonomous papers, the notation used in the following chapters differ and will be introduced for each chapter in turn. In the next two chapters, boldface symbols are used for vectors and matrices which reflects the practice in the published articles.

2 Long- versus Medium-run Identification in Fractionally Integrated VAR Models

This paper is joint work with Rolf Tschernig (University of Regensburg) and Enzo Weber (University of Regensburg, Institute for Employment Research (IAB) and Institute for East and Southeast European Studies). It is published as TSCHERNIG, R., E. WEBER, AND R. WEIGAND (2014): "Long- versus Medium-run Identification in Fractionally Integrated VAR Models," *Economics Letters*, 122(2), 299–302.

3 Long-run Identification in a Fractionally Integrated System

This paper is joint work with Rolf Tschernig (University of Regensburg) and Enzo Weber (University of Regensburg, Institute for Employment Research (IAB) and Institute for East and Southeast European Studies). It is published as TSCHERNIG, R., E. WEBER, AND R. WEIGAND (2013c): "Long-Run Identification in a Fractionally Integrated System," *Journal of Business & Economic Statistics*, 31(4), 438–450.

4 State Space Modeling of Fractional Cointegration Subspaces

Abstract. We investigate a setup for fractionally cointegrated time series which is formulated in terms of latent integrated and short-memory components. It accommodates nonstationary processes with different fractional orders and cointegration of different strengths and is applicable in high-dimensional settings. A convenient parametric treatment is achieved by finite-order ARMA approximations in the state space representation. Monte Carlo simulations reveal good estimation properties for processes of different dimensions. In an application to realized covariance matrices, we find that orthogonal short- and long-memory components provide a reasonable fit and outstanding out-of-sample performance compared to several competitor methods.

Keywords. Long memory, fractional cointegration, state space, unobserved components, factor model, realized covariance matrix.

JEL-Classification. C32, C51, C53, C58.

4.1 Introduction

Multivariate fractional integration and cointegration models have proven valuable in a wide range of empirical applications from macroeconomics and finance. They generalize the standard concept of cointegration by allowing for non-integer orders of integration both for the observations and for equilibrium errors; see Gil-Alana and Hualde (2008) for a literature review. In the field of macroeconomics, such models have turned out to be relevant in analyses of purchasing power parity beginning with Cheung and Lai (1993), of the relation between unemployment and input prices (Caporale and Gil-Alana, 2002) and of broader models for economic fluctuations (Morana, 2006). The empirical finance literature has considered fractional cointegration, e.g., for analysing international bond

returns (Dueker and Startz, 1998), for modeling co-movements of stock return volatilities (Beltratti and Morana, 2006), for assessing the link between realized and implied volatility (Nielsen, 2007) and for quantifying risk in strategic asset allocation problems (Schotman, Tschernig, and Budek, 2008). From a methodological point of view, semiparametric techniques for inference on the cointegration rank, the cointegration space and memory parameters have been very popular among empirical researchers, although the development of optimal parametric inferential methods for models with triangular or fractional vector error correction representations has recently made considerable progress (see, e.g., Robinson and Hualde, 2003; Avarucci and Velasco, 2009; Łasak, 2010; Johansen and Nielsen, 2012).

Despite their flexibility and their computationally simple treatment, semiparametric models are limited in scope since they aim to describe low-frequency properties only and are hence not appropriate for impulse response analysis and forecasting. While semiparametric techniques have been developed to cope with multivariate processes of different integration orders and multiple fractional cointegration relations of different strengths (Chen and Hurvich, 2006; Hualde and Robinson, 2010; Hualde, 2009), there seems to be a lack of parametric models of such generality. Furthermore, the usual error correction and triangular models with their typically abundant parametrization are not deemed appropriate for time series of dimension, say, larger than five.

In this paper, we investigate models for multivariate fractionally integrated and cointegrated time series which are formulated in terms of latent purely fractional and additive short-memory components. With a "type II" definition of fractional integration (Robinson, 2005), this approach allows for a flexible modeling of possibly nonstationary time series of different fractional integration orders. It permits cointegration relations of different strengths as well as polynomial cointegration (multicointegration in the terminology of Granger and Lee, 1989), i.e., cointegration between the levels of some time series and their (fractional) differences, and guarantees a clear representation of the long-run characteristics. The unobserved components formulation benefits the modeling of relatively high-dimensional time series. For this situation we propose a parsimonious parametrization based on dimension reduction and dynamic orthogonal components in the spirit of Pan and Yao (2008) and Matteson and Tsay (2011). We analyse the models in state space form which allows for missing values and a seamless treatment of additive seasonal, noise, break and cycle components familiar from structural time series models (Harvey, 1991).

Several authors have proposed fractional integration modeling by state space methods. Classical treatments of univariate stationary long memory include Chan and

Palma (1998), who study autoregressive fractionally integrated moving average (ARFIMA) processes and Grassi and de Magistris (2012), who consider ARFIMA models with noise, structural breaks and missing values. Bayesian simulation-based techniques have been proposed by Hsu and Breidt (2003) for noise-perturbed ARFIMA models and Brockwell (2007) for a so-called generalized long-memory model, where the conditional distribution of the process nonlinearly depends on a latent ARFIMA process.

Multivariate treatments of models based on latent fractional components have mostly been studied by semiparametric approaches. Ray and Tsay (2000) use semiparametric memory estimators and canonical correlations to infer the existence of common fractional components, Morana (2004) proposes a frequency domain principal component estimator, Morana (2007) estimate components of a single fractional integration order by univariate permanent-transitory (or persistent-transitory) decompositions followed by a principal component analysis of the permanent (or persistent) components and Luciani and Veredas (2012) estimate their fractional factor model by fitting long-memory models to the principal components of a large panel of time series. In a setup closest to ours, Chen and Hurvich (2006) suggest a semiparametric frequency domain methodology to identify and estimate cointegration subspaces which annihilate fractional components of different memory.

Parametric, likelihood-based methods for such models have so far been computationally demanding. Hsu, Ray, and Breidt (1998) discuss a Bayesian sampling algorithm for a bivariate process sharing one stationary long-memory component. More recently, Mesters, Koopman, and Ooms (2011) consider maximum likelihood estimation of stationary generalized long-memory models with one or more latent ARFIMA components. They propose an importance sampling scheme to obtain exact maximum likelihood estimators, but the methods become numerically challenging for more than two latent long-memory factors.

We consider a computationally straightforward classical treatment of our linear model in state space form. An approximation of potentially nonstationary fractional integration using finite-order ARMA structures is adapted. This procedure outperforms the standard truncation approach and provides a substantial reduction of the state dimension for a desired approximation quality, hence reducing the computational burden. Parameter estimation by means of the EM algorithm and analytical expressions for the likelihood score make the approach feasible even in high dimensions. In Monte Carlo simulations we study the performance of the proposed methods and quantify the accuracy of our state space approximation. For fractionally integrated and cointegrated processes of different dimensions we find favorable finite-sample estimation properties

also in light of alternative techniques.

The methods are applied to modeling and forecasting daily realized covariance matrices, where the strengths of our approach become apparent. In this setup, typically high-dimensional processes with strong persistence and a pronounced co-movement in the low-frequency dynamics are considered. In time series of log variances and z-transformed correlations for six US stocks, we find that common orthogonal short- and long-memory components with two different fractional integration orders provide a reasonable fit. A pseudo out-of-sample study shows that the fractional components model provides a superior forecasting accuracy compared to several competitor methods.

The paper is organized as follows. Section 4.2 introduces the general setup and a specific parsimonious model, section 4.3 discusses its state space form and maximum likelihood estimation, while in section 4.4 the estimation properties are investigated by means of Monte Carlo experiments. The empirical application to realized covariance matrices and a pseudo out-of-sample assessment are contained in section 4.5 before section 4.6 concludes.

4.2 Fractional Components Models

In this section, we introduce a general modeling setup and clarify its integration and cointegration properties. Furthermore, its relation to existing setups for multivariate integrated time series is discussed. A specific model appropriate for relatively high-dimensional processes is considered which will be the workhorse specification in the empirical application of section 4.5.

4.2.1 The General Setup

We consider a linear model for a p-dimensional observed time series y_t, which we label a *fractional components* (FC) setup,

$$y_t = \Lambda x_t + u_t, \qquad t = 1, \ldots, n. \tag{4.1}$$

The model is formulated in terms of the latent processes x_t and u_t where Λ will always be assumed to have full column rank and the components of the s-dimensional x_t are fractionally integrated noise according to

$$\Delta^{d_j} x_{jt} = \xi_{jt}, \qquad j = 1, \ldots, s. \tag{4.2}$$

For a generic scalar d, the fractional difference operator is defined by

$$\Delta^d = (1-L)^d = \sum_{j=0}^{\infty} \pi_j(d) L^j, \quad \pi_0(d) = 1, \quad \pi_j(d) = \frac{j-1-d}{j} \pi_{j-1}(d),\ j \geq 1, \tag{4.3}$$

where L denotes the lag or backshift operator, $Lx_t = x_{t-1}$. We adapt a nonstationary type II solution of these processes (Robinson, 2005) and hence treat $d_j \geq 0.5$ alongside the asymptotically stationary case $d_j < 0.5$ in a continuous setup, while setting starting values to zero, $x_{jt} = 0$ for $t \leq 0$. Nonzero initial values have been considered for observed fractional processes by Johansen and Nielsen (2012), but are not straightforwardly handled for our unobserved processes. The solution is based on the truncated operator $\Delta_+^{-d_j}$ (Johansen, 2008) and given by

$$x_{jt} = \Delta_+^{-d_j} \xi_{jt} = \sum_{i=0}^{t-1} \psi_i(d) \xi_{j,t-i}, \qquad j = 1, \ldots, s.$$

Without loss of generality let the components be arranged such that $d_1 \geq \ldots \geq d_s$.

We assume $d_j > 0$ for all j in what follows, so that x_t governs the long-term characteristics of the observations y_t. These are complemented by additive short-run dynamics which we describe by stationary vector ARMA specifications for u_t in the general case. The process is given by

$$\Phi(L) u_t = \Theta(L) e_t, \qquad t = 1, \ldots, n, \tag{4.4}$$

where $\Phi(L)$ and $\Theta(L)$ are a stable vector autoregressive polynomial and an invertible moving average polynomial, respectively. The disturbances ξ_t and e_t jointly follow a Gaussian white noise (*NID*) sequence such that

$$\xi_t \sim NID(0, \Sigma_\xi), \quad e_t \sim NID(0, \Sigma_e) \quad \text{and} \quad \mathrm{E}(\xi_t e_t') = \Sigma_{\xi e}, \tag{4.5}$$

where at this stage, before turning to identified and empirically relevant model specifications below, we do not consider restrictions on the joint covariance matrix, but only require Σ_ξ to have strictly positive entries on the main diagonal.

Some remarks regarding the general FC setup are in order. The model as given in (4.1) is not identified without further restrictions on the loading matrix Λ, on the vector ARMA coefficients and on the noise covariance matrix. While restrictions on Σ_ξ and Λ may be based on results in dynamic factor analysis as will be seen below, choosing specific parametrizations for u_t will depend on characteristics of the data and on the purpose of the empirical analysis. Identified vector ARMA structures like the echelon

form (see Lütkepohl, 2005, chapter 12) can be used for a rich parametrization, while a multivariate structural time series approach as described in Harvey (1991) integrates nicely with the unobserved components framework considered in this paper. Below, we introduce a parsimonious model well-suited to relatively high dimensions which is conceptually based on dimension reduction and orthogonal components.

For a characterization of the integration and cointegration properties of our model, we adapt the definitions of these concepts from Hualde and Robinson (2010), which prove useful here. Hence, a generic scalar process ρ_t is called integrated of order δ or $I(\delta)$ if it can be written as $\rho_t = \sum_{i=1}^{l} \Delta_+^{-\delta_i} \nu_{it}$, where $\delta = \max_{i=1,\dots,l}\{\delta_i\}$ and $\nu_t = (\nu_{1t}, \dots, \nu_{lt})'$ is a finite-dimensional covariance stationary process with spectral density matrix which is continuous and nonsingular at all frequencies. A vector process τ_t is called $I(\delta)$ if δ is the maximum integration order of its components. We call the process τ_t cointegrated if there exists a vector β such that $\beta'\tau_t$ is $I(\gamma)$ where $\delta - \gamma > 0$ will be referred to as the strength of the cointegration relation. The number of linearly independent cointegration relations with possibly differing γ is called cointegration rank of τ_t.

By these definitions, x_{jt} is clearly $I(d_j)$ while both x_t and y_t are integrated of order d_1. We observe at least two different integration orders in the individual series of y_t whenever $\Lambda_{i1} = 0$ for some i and $d_1 > d_2$. More generally, $y_{it} \sim I(d_j)$, if $\Lambda_{i1} = \ldots = \Lambda_{i,j-1} = 0$ but $\Lambda_{ij} \neq 0$.

To state the cointegration properties of the FC setup (4.1), we assume that $s \leq p$, so that all fractional components are reflected by the integration and cointegration structure of y_t and that Σ_ξ is nonsingular. It is useful to identify all q groups of x_{jt} with identical integration orders and denote their respective sizes by $s_1, \dots, s_q$, such that $d_{s_1+\ldots+s_{j-1}+1} = \ldots = d_{s_1+\ldots+s_j}$ and $s = \sum_{j=1}^{q} s_j$. Of course, if $q = s$, then $s_1 = \ldots = s_q = 1$ and all components of x_t have mutually different integration orders, while for $q = 1$ it holds that $s = s_1$ and we observe $d_1 = \ldots = d_s$.

To keep notation simple, for a generic matrix A for which a specific grouping of rows and columns is clear from the context, we denote by $A^{(i,j)}$ the block from intersecting the i-th group of rows with the j-th group of columns. A stacking of several groups of rows $i,\dots,j$ and columns $k,\dots,l$ is indicated by $A^{(i:j,k:l)}$. For a grouping in only one dimension we write $A^{(i)}$ or $A^{(i:j)}$, where it shall be clear from the context whether a grouping of rows or columns is considered. Furthermore, we denote the column space of a generic $k \times l$ matrix A by $sp(A) \subseteq \mathbb{R}^k$ and its orthogonal complement by $sp^\perp(A)$. Further, for $k > l$, the $k \times (k-l)$ orthogonal complement of A will be denoted by $A_\perp$, which spans the $(k-l)$-dimensional space $sp^\perp(A)$.

According to the grouping of equal individual integration orders in x_t, we may there-

fore rewrite the FC process (4.1) as

$$y_t = \Lambda^{(1)} x_t^{(1)} + \ldots + \Lambda^{(q)} x_t^{(q)} + u_t.$$

Here, $\Lambda^{(j)}$ is a $p \times s_j$ submatrix of Λ consisting of columns $\Lambda_{\cdot i}$ for which $s_1 + \ldots + s_{j-1} < i \leq s_1 + \ldots + s_j$, and $x_t^{(j)}$ is a s_j-dimensional subprocess of x_t corresponding to components with memory parameter $d^{(j)} := d_{s_1+\ldots+s_{j-1}+1} = \ldots = d_{s_1+\ldots+s_j}$. Whenever $s_1 < p$, there exist $p - s_1$ linearly independent linear combinations $\beta_i' y_t \sim I(\gamma_i)$ and $\gamma_i < d_1$, so that fractional cointegration occurs. Due to our definition of cointegration, this may be a trivial case where a single component y_{it} with integration order smaller than d_1 is selected. Since

$$\Lambda_\perp^{(1)'} y_t = \Lambda_\perp^{(1)'} \Lambda^{(2)} x_t^{(2)} + \ldots + \Lambda_\perp^{(1)'} \Lambda^{(q)} x_t^{(q)} + \Lambda_\perp^{(1)'} u_t$$

is integrated of order $d^{(2)}$, the columns of $\Lambda_\perp^{(1)}$ qualify as cointegration vectors and $\mathscr{S}^{(1)} := sp^\perp(\Lambda^{(1)})$ is the $(p - s_1)$-dimensional cointegration space of y_t.

Whenever $s_1 + s_2 < p$, there are subspaces of $\mathscr{S}^{(1)}$ forcing a stronger reduction in integration orders. More generally, it holds that $\Lambda_\perp^{(1:j)'} y_t \sim I(d^{(j+1)})$ whenever $\sum_{i=1}^j s_i < p$ and where we set $d^{(j+1)} = 0$ for $j > s$. Analogously to Hualde and Robinson (2010), for $s = p$ and $j = 1, \ldots, q-1$, we call $\mathscr{S}^{(j)} := sp^\perp(\Lambda^{(1:j)})$ the j-th cointegration subspace of y_t, for which $\mathscr{S}^{(q-1)} \subset \ldots \subset \mathscr{S}^{(1)}$. For $p > s$, $\mathscr{S}^{(q)} \subset \mathscr{S}^{(q-1)}$ is a further such subspace. Cointegration vectors in $\mathscr{S}^{(q)}$ cancel all fractional components and hence reduce the integration order from d_1 to zero, the strongest reduction possible in our setup.

Besides this general pattern of cointegration relations, our model features an interesting special case with so-called polynomial cointegration, that is, cointegration relations where lagged observations nontrivially enter a cointegration relation. To see this possibility, consider a bivariate example similar to Granger and Lee (1989), where $q = p = 2$ and $\xi_{1t} = \xi_{2t}$, so that Σ_ξ is singular and $x_{2t} = \Delta^{d_1 - d_2} x_{1t}$. Augmenting the variables by a fractional difference as $\tilde{y}_t := (y_{1t}, y_{2t}, \Delta^{d_1 - d_2} y_{2t})'$, we obtain a three-dimensional system where levels of y_t enter a nontrivial cointegration relation with a fractional difference to achieve a reduction in integration order from d_1 to $\max\{2d_2 - d_1, 0\} < d_2$; see equation (4.26) below for the cointegration space. Hence, our setup complements the model of Johansen (2008, section 4), which was the first to handle polynomial cointegration in a fractional setup.

4.2.2 Relations to Other Cointegration Models

In this section, we clarify the relation of the fractional components model (4.1) to popular existing representations for cointegrated processes and show how our model can be represented in alternative ways brought forward in the literature. While our model is among the most general setups with respect to its integration and cointegration properties, the additive modeling of short-run dynamics is new to the literature and gives rise to distinct parametrizations not possible within other representations in a similarly convenient way.

Error correction models. The most popular representation of cointegrated systems in the $I(1)$ setting is the vector error correction form. Since an early mention by Granger (1986), in the fractionally integrated case, e.g., Avarucci and Velasco (2009), Łasak (2010) and Johansen and Nielsen (2012) have recently considered such models. In terms of the integration and cointegration properties, the fractional error correction setups are typically restricted to the special case with $q = 2$ and $s = p$, such that the observed variables are integrated of order $d^{(1)}$ and there exist $p - s_1$ cointegration relations with errors of order $d^{(2)}$.

Defining the fractional lag operator $L_b := 1 - \Delta^b$ (Johansen, 2008), we are able to derive the error correction representation for this special case of our model; see appendix 4.A. It is given by

$$\Delta^{d^{(1)}} y_t = \alpha\beta' L_{d^{(1)}-d^{(2)}} \Delta^{d^{(2)}} y_t + \kappa_t, \tag{4.6}$$

where we find $\alpha\beta' = -\Lambda^{(2)}(\Lambda_\perp^{(1)\prime}\Lambda^{(2)})^{-1}\Lambda_\perp^{(1)\prime}$ to precede the error correction term, while

$$\kappa_t := M(\Lambda^{(1)}\xi_t^{(1)} + \Delta^{d^{(1)}} u_t) - \alpha\beta'(\Lambda^{(2)}\xi_t^{(2)} + \Delta^{d^{(2)}} u_t)$$

is integrated of order zero and M is defined in (4.29).

The model differs both from the models of Avarucci and Velasco (2009) and from the representation of Johansen (2008) in the way short-run dynamics are modeled. While the literature has considered (fractional) lags of differenced variables and possibly of error correction terms in the VECM representation, our setup generates autocorrelated κ_t through the introduction of a latent u_t. In practice, approximating our model by an autoregressive structure in $\Delta^{d^{(1)}} y_t$ may lead to an abundance of parameters whenever u_t is reasonably parametrized and vice versa.

As we have discussed above, Johansen (2008) proposes a polynomially cointegrated generalization of his model which allows terms integrated of orders d, $d-b$ and $d-2b$ in

the Granger representation (Johansen, 2008, theorem 9). Even compared to that specification, our model allows for more general patterns of integration orders and cointegration strengths. More in line with the generality envisaged in this paper, Tschernig, Weber, and Weigand (2013c, appendix A) present a model with error correction term and different integration orders, while Łasak and Velasco (2014) sequentially fit error correction models to test for cointegration relations of possibly different strengths.

Vector ARFIMA. An interesting special case of (4.1) occurs for $s = p$ and $\Lambda = I$, where each series in y_{it} is driven by a single fractional component and $y_{it} \sim I(d_i)$. This resembles standard vector ARFIMA models with possibly different integration orders; see, e.g., Lobato (1997) who labels the more popular vector ARFIMA class considered here as "model A". A frequently used submodel is the fractionally integrated vector autoregressive model discussed by Nielsen (2004). The main difference to these approaches is our additive modeling of short-run dynamics, whereas in the vector ARFIMA setup weakly dependent vector ARMA instead of noise processes are passed through the fractional integration filters.

Our model belongs to the class of vector ARFIMA processes for integer $d_j \in \{1,2,\ldots\}$, but not for general fractional integration orders. For the case of integer d_j, note that $(x_t', u_t')'$ is a finite-order vector ARMA process, and hence y_t as a linear combination is itself in the ARMA class; see Lütkepohl (1984). For general vector ARFIMA processes, a similar conclusion does not hold. It is sufficient to consider a stylized univariate case of our model with $p = s = 1$, where $\Delta^d x_t = \xi_t$ and $(1 - \phi L)u_t = e_t$. First note that $(\Delta^d x_t, u_t)'$ has a vector ARMA structure, and hence $(x_t, u_t)'$ is a vector ARFIMA process. Expanding $(1 - \phi L)\Delta^d x_t = (1 - \phi L)\xi_t$ and $(1 - \phi L)\Delta^d u_t = \Delta^d e_t$, we can write the sum, belonging to the fractional components model class, as

$$(1 - \phi L)\Delta^d y_t = (1 - \phi L)\xi_t + \Delta^d e_t. \tag{4.7}$$

The right hand side of this expression is not a finite-order MA process in general, as it has nonzero autocorrelations for all lags, and hence, the process does not belong to the ARFIMA class.

Triangular representations. The models discussed so far have restricted integration or cointegration properties as compared to our model. In contrast, Hualde (2009) and Hualde and Robinson (2010) have proposed a very flexible model which adapts the triangular form of Phillips (1991) and its generalization to processes with multiple unit

roots (Stock and Watson, 1993) to the fractional cointegration setup.

To derive the triangular representation for our model, we assume that the variables in y_t are ordered in a way that $\Lambda^{(1:j,1:j)}$ is nonsingular for $j = 1,\dots,q$ and restrict attention to the case $s = p$. The variables are partitioned according to the groups of different integration orders in x_t as $y_t^{(j)} := (y_{s_1+\dots+s_{j-1}+1},\dots,y_{s_1+\dots+s_j})'$, $j = 1,\dots,q$. The first block in the triangular system is

$$\Delta^{d_1} y_t^{(1)} = \Lambda^{(1,1)}\xi_t^{(1)} + \Lambda^{(1,2)}\Delta^{d_1-d_2}\xi_t^{(2)} + \dots + \Lambda^{(1,q)}\Delta^{d_1-d_q}\xi_t^{(q)} + \Delta^{d_1}u_t \quad (:= \omega_t^{(1)}), \tag{4.8}$$

where $\omega_t^{(1)}$ is integrated of order zero. The general expression for the j-th block of the triangular system is derived in appendix 4.A for $j = 2,\dots,q$, and given by

$$\begin{aligned}\Delta^{d^{(j)}} y_t^{(j)} &= \Lambda^{(j,1:(j-1))}(\Lambda^{(1:(j-1),1:(j-1))})^{-1}\Delta^{d^{(j)}} y_t^{(1:(j-1))} + \omega_t^{(j)} \\ &= -B^{(j,1)}\Delta^{d^{(j)}} y_t^{(1)} - \dots - B^{(j,j-1)}\Delta^{d^{(j)}} y_t^{(j-1)} + \omega_t^{(j)},\end{aligned} \tag{4.9}$$

where also $\omega_t^{(j)}$ is integrated of order zero for $j = 2,\dots,q$. By inverting the fractional difference operators we obtain

$$By_t = (\Delta_+^{-d_1}\omega_t^{(1)'},\dots,\Delta_+^{-d_q}\omega_t^{(q)'})', \tag{4.10}$$

where B has a block triangular structure such that $B^{(i,i)} = I$ and $B^{(i,j)} = 0$ for $i < j$. A re-ordering of the variables in y_t yields the representation of Hualde and Robinson (2010).

This representation allows for a semiparametric cointegration analysis of our model using the methods of Hualde (2009) and Hualde and Robinson (2010). However, our model differs significantly from straightforward parametrizations of the triangular system, e.g., from assuming a vector ARMA process for ω_t, since in our setup ω_t as stated in (4.32) generally contains fractional differences that cannot be represented within the ARMA framework.

State space approaches. Bauer and Wagner (2012) have presented a state space canonical form for multiple frequency unit root processes of different (integer-valued) integration orders. Their discussion is based on unit root vector ARMA models which are separated in pure unit root structures and short-term dynamics. Although the analogy to our model is striking, there are notable differences between their unit root and our fractional setup. Firstly, as discussed in the paragraph on vector ARFIMA mod-

els (see (4.7)), the fractional components setup (4.1) is not nested within a general class comparable to the vector ARMA models, which form the basis of the discussion in Bauer and Wagner (2012). Secondly, in their setting, the introduction of different integration orders is achieved by repeated summation of lower order integrated processes which themselves enter the observations to achieve polynomial cointegration. This is in contrast to the continuous treatment of integration orders in our (type II) fractional setup.

However, fractional components models could be constructed to straightforwardly extend the setup of Bauer and Wagner (2012). Using the fractional lag operator $L_b = 1-\Delta^b$ instead of L in the short-run dynamic specification (4.4), a stable vector ARMA$_b$ process can be defined by $\tilde{\Phi}(L_b)\tilde{u}_t = \tilde{\Theta}(L_b)e_t$ under suitable stability conditions (Johansen, 2008, corollary 6). Then, replacing u_t by $\tilde{u}_t$ in the model setup (4.1) with d_j restricted to some multiple of b ($d_j = i_j b,\ i_j \in \{1,2,\ldots\}$), the process y_t is in the class of vector ARMA$_b$ models itself, while unit roots in the vector autoregressive polynomial generate the fractional $I(d_j)$ processes. Such a framework could be treated analogously to Bauer and Wagner (2012), but the restriction that all integration orders are multiples of b makes such a framework somewhat less flexible than ours.

4.2.3 A Dimension-reduced Orthogonal Components Specification

So far, we have considered a general modeling setup and discussed its integration and cointegration properties as well as its relation to existing approaches in the literature. We now turn to the discussion of a specific model from this class which bears potential for parsimonious modeling of long- and short-run dynamics in relatively high-dimensional applications. Besides its general interest, this will be the workhorse specification for the empirical application to realized covariance modeling in section 4.5.

To introduce the model and emphasize its restrictions as compared to (4.1), we decompose the short-term dependent process u_t into an autocorrelated component, Γz_t, where z_t is a vector of s_0 mutually uncorrelated components with $s+s_0 \leq p$, and a white noise component ε_t, respectively. We label the result the *dynamic orthogonal fractional components* (DOFC) model,

$$y_t = \Lambda^{(1)} x_t^{(1)} + \ldots + \Lambda^{(q)} x_t^{(q)} + \Gamma z_t + \varepsilon_t, \tag{4.11}$$

where x_t is generated by a purely fractional process (4.2) as above, while

$$(1 - \phi_{j1}L - \ldots - \phi_{jk}L^k) z_{jt} = \zeta_{jt}, \qquad j = 1,\ldots,s_0,$$

are s_0 univariate stationary autoregressive processes of order k. Regarding the mutually uncorrelated noise processes ξ_t, ζ_t and ε_t, we assume

$$\xi_t \sim NID(0,I), \qquad \zeta_t \sim NID(0,I) \qquad \text{and} \quad \varepsilon_t \sim NID(0,H),$$

where H is diagonal with entries $h_i > 0$, $i = 1,\ldots,p$.

The model as specified in (4.11) and below is not identifiable without further information. Considering $\tilde{y}_t := \Delta^{d^{(1)}} y_t$ instead of y_t to meet the assumptions of Heaton and Solo (2004), their theorem 4 suggests that groups of common components $\Delta^{d^{(1)}} x_t^{(1)}$, …, $\Delta^{d^{(1)}} x_t^{(q)}$, $\Delta^{d^{(1)}} z_t$ can be disentangled (up to rotations within these groups) through their different shapes in spectral densities whenever $d^{(1)} > \ldots > d^{(q)} > 0$. Still, there exist observationally equivalent structures with $\tilde{\Lambda}^{(j)} = \Lambda^{(j)} M^{-1}$ and $\tilde{x}_t^{(j)} = M x_t^{(j)}$ which satisfy the model restrictions for orthonormal M. Hence, we impose further restrictions on the loading matrices. As is standard practice in dynamic factor analysis, we set the upper triangular elements to zero such that $\Lambda_{rl}^{(j)} = 0$ for $r < l$, $j = 1,\ldots,q$, and $\Gamma_{rl} = 0$ for $r < l$. Certain observables are thus assumed not to be influenced by certain factors.

The model (4.11) is very parsimonious considering that it includes both a rich fractional structure as well as short-run dynamics with co-dependence. This is possible by comprising three components of parsimony which have been brought forward in the statistical time series literature. Firstly, there are $p - s - s_0 \geq 0$ white noise linear combinations of y_t. A strict inequality implies a reduced dimension in the dynamics of y_t which is characteristic for so-called statistical factor models; see Pan and Yao (2008), Lam, Yao, and Bathia (2011) and Lam and Yao (2012). In contrast, the model (4.1) does not belong to this class in general. Secondly, all cross-sectional correlation stems from the common components which is a familiar feature from classical factor analysis (Anderson and Rubin, 1956). Thirdly, both the fractional and the nonfractional components are mutually orthogonal for all leads and lags.

Combined with semiparametric techniques of fractional integration and cointegration analysis, existing methods for statistical factor and dynamic orthogonal components analysis (Matteson and Tsay, 2011) can be used to justify the model assumptions and may be useful in the course of model specification as we will illustrate in the empirical application below.

4.3 State Space Form and Estimation

Parametric treatments of unobserved components models are conveniently conducted in a state space setup. The next section considers the state space representation of fractional components models along with an approximation enabling computational feasibility. Classical likelihood-based inference is described subsequently. The discussion is based on the DOFC model (4.11), while other specifications from the general FC setup can be treated analogously with minor modifications.

4.3.1 Approximating Nonstationary Fractional Integration

Unlike the stationary long-memory processes considered in the literature, e.g., by Chan and Palma (1998), Hsu, Ray, and Breidt (1998), Hsu and Breidt (2003), Brockwell (2007), Mesters, Koopman, and Ooms (2011) as well as Grassi and de Magistris (2012), our nonstationary type II specification of fractional integration with zero initial values is straightforwardly represented in state space form. Since x_t has an autoregressive structure with coefficient matrices $\Pi_j^d = \text{diag}(\pi_j(d_1),\ldots,\pi_j(d_s))$, $j = 1,\ldots,n$, see (4.2), a Markovian state vector embodying x_t has to include $n-1$ lags of x_t and is initialized deterministically with $x_{-n+1} = \ldots = x_0 = 0$.

In principle, this exact state space form can be used to compute the Kalman filter, evaluate the likelihood and estimate the unknown model parameters by nonlinear optimization routines. Since the state vector is of dimension $sn + s_0 k$, this becomes computationally very cumbersome, particularly in large samples and for a large number s of fractional components. This makes a treatment of the system in its exact state space representation practically infeasible for a wide range of relevant applications from macroeconomics and finance.

The literature on stationary long-memory processes has considered approximations based on truncating the autoregressive representation, considering only $m-1$ lags of x_t for $m < n$ in the transition equation and thus effectively set Π_j^d to zero for $j > m$. Alternatively, the moving average representation has been truncated to arrive at a feasible state space model; see Palma (2007), sections 4.2 and 4.3.

We will apply ARMA approximations to the fractional state vectors which provide a better approximation quality than the autoregressive or moving average truncation. An ARMA approximation of long-memory processes has been considered in the importance sampling frameworks of Hsu and Breidt (2003) and Mesters, Koopman, and Ooms (2011), but did not find wider usage in applied research so far. In our setup, where

fractional integration appears in the form of purely fractional components rather than ARFIMA processes, this approach is particularly convenient.

We introduce an approximation method which is reasonable in our possibly nonstationary framework. For a univariate process $x_t = \Delta_+^{-d}\xi_t = \sum_{j=0}^{n-1}\psi_j(d)\xi_{t-j}$, for finite v and w we consider as a (nonstationary) approximation the process

$$\tilde{x}_t = \left[\frac{(1+m_1L+\ldots+m_wL^w)}{(1-a_1L-\ldots-a_vL^v)}\right]_+ \xi_t = \sum_{j=0}^{n-1}\tilde{\psi}_j(\varphi)\xi_{t-j}, \tag{4.12}$$

where $\varphi := (a_1,\ldots,a_v,m_1,\ldots,m_w)'$ and all a_i and m_j will be made functionally dependent on d to approximate x_t by $\tilde{x}_t$.

In order to determine the parameters ϕ, we minimize the distance between x_t and $\tilde{x}_t$, using the mean squared error (MSE) over $t = 1,\ldots,n$ as distance measure. For given t, d and φ, we observe

$$\tilde{x}_t - x_t = \sum_{j=0}^{t-1}\tilde{\psi}_j(\varphi)\xi_{t-j} - \sum_{j=0}^{t-1}\psi_j(d)\xi_{t-j} = \sum_{j=0}^{t-1}(\tilde{\psi}_j(\varphi)-\psi_j(d))\xi_{t-j}.$$

Hence, the MSE for period t is given by

$$E[(\tilde{x}_t - x_t)^2] = Var(\xi_t)\sum_{j=0}^{t-1}(\tilde{\psi}_j(\varphi)-\psi_j(d))^2,$$

while averaging over all periods for a given sample size n and ignoring the constant variance term yields the objective function for a given d,

$$\mathrm{MSE}_n^d(\varphi) = \frac{1}{n}\sum_{t=1}^{n}\sum_{j=0}^{t-1}(\tilde{\psi}_j(\varphi)-\psi_j(d))^2 = \frac{1}{n}\sum_{j=1}^{n}(n-j+1)(\tilde{\psi}_j(\varphi)-\psi_j(d))^2. \tag{4.13}$$

The approximating ARMA coefficients are thus given by

$$\hat{\varphi}_n(d) = \arg\min_{\varphi}\mathrm{MSE}_n^d(\varphi). \tag{4.14}$$

Technically, to obtain the approximating ARMA coefficients we impose stability of the autoregressive polynomial for $d < 1$ and appropriate unit roots for $d \geq 1$. In order to achieve numerically well-behaved optimizations, we work with transformed parameters. More precisely, the optimization is conducted in the space of partial autocorrelations of the stable autoregressive and moving average parts, respectively; see Barndorff-Nielsen and Schou (1973) and Veenstra (2012). Additionally, we apply Fisher's z-transform

$z = 0.5[\log(1+x) - \log(1-x)]$ to obtain unconstraint optimization problems. For a given sample size n, we carry out an optimization for each value on a grid for d. We smooth the values using splines before the result is re-transformed to the space of ARMA coefficients, and hence obtain a continuous and differentiable function $\hat{\varphi}_n(d)$. Whenever discontinuities occur in the space of transformed parameters (as for $d = 1$), we obtain a smooth transition between segments of $\hat{\varphi}_n(d)$ using the sine function. All computations in this paper are conducted using R (R Core Team, 2013).

To illustrate the results we plot the approximating ARMA(2,2) parameters as a function of d for $n = 500$; see figure 4.1. A closer look at the coefficients reveal that for $d > 0$ typically both the autoregressive and the moving average polynomial have roots close to unity which nearly cancel out. For example, to approximate a process with $d = 0.75$ we have $(1 - 1.932L + 0.932L^2)\tilde{x}_t = (1 - 1.285L + 0.306L^2)\xi_t$, which can be factorized as $(1 - 0.999L)(1 - 0.933L)\tilde{x}_t = (1 - 0.970L)(1 - 0.316L)\xi_t$.

To compare the ARMA(v,w) approximations with $v = w \in \{1,2,3,4\}$ to a truncated AR(m) process, we contrast the approximating impulse response function $\tilde{\psi}_j$ to the true one, $\psi_j(d)$, for a given d. The autoregressive truncation lag $m = 50$ is used for our comparison, since this is among the largest values which we consider as feasible in multivariate applications. The result of this comparison is shown in figure 4.2 for $n = 500$ and $d = 0.75$. The autoregressive truncation approach gives the exact impulse responses for horizons $j \leq 50$, but then tapers off too fast. The ARMA approximations improves significantly over the autoregressive truncation whenever $v = w \geq 2$. For orders 3 or 4, the approximation error is even hardly visible. We do not show the moving average truncation, where the impulse responses equal zero for horizons exceeding the truncation lag.

To perform the comparison for different d, we plot the square root of the MSE (4.13) as a function of d for different approximation methods. For negative integration orders, see figure 4.3, the moving average approach clearly outperforms the autoregression, while the ARMA method with orders $v = w > 2$ are better. The moving average approximation becomes inaccurate, however, for the case $d > 0$, and worse even than the autoregressive method as can be seen in figure 4.4. In contrast, the ARMA(3,3) and ARMA(4,4) approximations are well-suited to mimic fractional processes over the whole range of d. Further evidence in favor of the ARMA approximation will be presented in the Monte Carlo simulation of section 4.4.1. The following discussion on parameter estimation and the empirical application in section 4.5 will therefore be based on ARMA approximations.

4.3.2 The State Space Representations

Based on these methods we introduce the state space form of the multivariate model (4.11), where each x_{jt} is approximated by the ARMA approach. The general state space model is

$$y_t = Z\alpha_t + \varepsilon_t, \qquad \alpha_{t+1} = T\alpha_t + R\eta_t, \tag{4.15}$$

where in our case the states may be partitioned into $\alpha_t' = (\alpha_t^{(1)'}, \alpha_t^{(2)'})$, the states related to the fractional and the autoregressive components, respectively.

We define $A_j^d := \text{diag}(\hat{a}_j(d_1), \ldots, \hat{a}_j(d_s))$ and $M_j^d := \text{diag}(\hat{m}_j(d_1), \ldots, \hat{m}_j(d_s))$ containing the approximating AR and MA coefficients, respectively, while $A_j^d = 0$ for $j > v$ and $M_j^d = 0$ for $j > w$. For $u = \max(v, w+1)$, the first part of the state vector is a (us)-dimensional process $\alpha_t^{(1)'} = (\mu_t', \ldots, \mu_{t-u+1}')$, with $\mu_t = (I - A_1^d L - \ldots - A_v L^v)_+^{-1} \xi_t$, and $\alpha_t^{(2)'} = (z_t', \ldots, z_{t-k+1}')$ has dimension ks_0. The observation equation is therefore specified as

$$Z = \left(Z^{(1)}, Z^{(2)}\right), \qquad Z^{(1)} = \left(\Lambda \quad \Lambda M_1^d \quad \ldots \quad \Lambda M_{u-1}^d\right), \qquad Z^{(2)} = \left(\Gamma \quad 0\right),$$

and the transition matrix is given by

$$T = \begin{pmatrix} T^{(1,1)} & 0 \\ 0 & T^{(2,2)} \end{pmatrix}, \quad T^{(1,1)} = \begin{pmatrix} A_1^d & A_2^d & \ldots & A_u^d \\ I & & & 0 \\ \vdots & \ddots & & \vdots \\ 0 & \ldots & I & 0 \end{pmatrix}, \quad T^{(2,2)} = \begin{pmatrix} \Phi_1 & \Phi_2 & \ldots & \Phi_k \\ I & & & 0 \\ \vdots & \ddots & & \vdots \\ 0 & \ldots & I & 0 \end{pmatrix}.$$

Furthermore, from the definition of our model (4.11) we have $\Phi_j := \text{diag}(\phi_{1j}, \ldots, \phi_{s_0 j})$, $\eta_t := (\xi_{t+1}', \zeta_{t+1}')'$ and $Q := \text{Var}(\eta_t) = I$, while

$$R = \begin{pmatrix} R^{(1,1)} & 0 \\ 0 & R^{(2,2)} \end{pmatrix}, \qquad R^{(1,1)} = \begin{pmatrix} I \\ 0 \end{pmatrix}, \qquad R^{(2,2)} = \begin{pmatrix} I \\ 0 \end{pmatrix}.$$

The dynamics are complemented by the initial conditions for the states. From the definition of our type II fractional process we set $\alpha_0^{(1)} = 0$, while $\alpha_t^{(2)}$ is initialized by its stationary distribution.

The fractional components x_t do not explicitly appear as states in this representation. However, filtered and smoothed states can be constructed using the relation $x_t = \mu_t + \sum_{j=1}^{w} M_j^d \mu_{t-j}$. To obtain conditional covariance matrices for x_t, it is more convenient to use an alternative state space form of the ARMA process, where the MA coefficients

appear in $R^{(1,1)}$ rather than in $Z^{(1)}$; see Durbin and Koopman (2012), section 3.4. The current setup is appropriate for estimating the parameters via the EM algorithm which is discussed in the next section.

4.3.3 Maximum Likelihood Estimation

The EM algorithm has been proposed for maximum likelihood estimation of state space models by Shumway and Stoffer (1982) and Watson and Engle (1983). In the context of dynamic factor models, this method has been found very useful in locating maxima of high-dimensional likelihood functions for models of possibly more than hundred dependent variables; see, e.g., Quah and Sargent (1993), Doz, Giannone, and Reichlin (2012) and Jungbacker and Koopman (2014). Since after rapidly locating an approximate optimum the final steps until convergence are typically slow for the EM algorithm, it has been suggested to switch to gradient-based methods with analytical expressions for the likelihood score at a certain step.

We will present these algorithms for our fractional model, and thereby extend existing treatments in the literature. For the model represented by (4.15), the matrices T and Z both nonlinearly depend on d and other unknown parameters, so that there are nonlinear cross-equation restrictions linking the transition and the observation equation of the system.

The EM algorithm in general consists of two steps, which are repeated until convergence. In the E-step the expected complete data likelihood is computed, where the expectation is evaluated for a given set of parameters $\theta_{\{j\}}$, while the M-step maximizes this function to arrive at the parameters used in the next E-step, $\theta_{\{j+1\}}$. Thus, we define $Q(\theta,\tilde{\theta}) := \mathrm{E}_{\tilde{\theta}}[l(\theta)]$, where in this section all expectation operators are understood as conditional on the data $y_1,\dots,y_n$. In the course of the EM algorithm, after choosing suitable startup values $\theta_{\{1\}}$, the optimization $\theta_{\{j+1\}} = \arg\max_\theta Q(\theta,\theta_{\{j\}})$ is iterated for $j = 1,2,\dots$ until convergence.

To state the algorithm for the model (4.11), following Wu, Pai, and Hosking (1996), we obtain the expected complete data likelihood as

$$\begin{aligned} Q(\theta;\theta_{\{j\}}) = &-\frac{n}{2}\log|Q| - \frac{1}{2}\operatorname{tr}\left[RQ^{-1}R'(A_{\{j\}} - TB'_{\{j\}} - B_{\{j\}}T' + TC_{\{j\}}T')\right] \\ &-\frac{n}{2}\log|H| - \frac{1}{2}\operatorname{tr}\left[H^{-1}(D_{\{j\}} - ZE'_{\{j\}} - E_{\{j\}}Z' + ZF_{\{j\}}Z')\right], \end{aligned} \tag{4.16}$$

where in our case $Q = I$, while T, Z and H are functions of θ and a possible dependence of the initial conditions for α_0 on θ has been discarded for simplicity. The conditional

moment matrices $A_{\{j\}}$, $B_{\{j\}}$, ..., are given in appendix 4.B and can be computed by a single run of a state smoothing algorithm (Durbin and Koopman, 2012, section 4.4) based on the system determined by $\theta_{\{j\}}$.

Rather than carrying out the full maximization of $Q(\theta, \theta_{\{j\}})$ at each step, we obtain a computationally simpler modified algorithm. To this end, we partition the vector of unknown parameters as $\theta' = (\theta^{(1)'}, \theta^{(2)'})$ where $\theta^{(1)'} = (d', \phi', \lambda', \gamma')$ and ϕ, λ, γ contain the unknown elements in Φ_j, Λ and Γ, respectively, while the variance parameters are collected in $\theta^{(2)'} = (h_1, \ldots, h_p)$. First, the ECM algorithm described by Meng and Rubin (1993) in our setup amounts to a conditional optimization over $\theta^{(1)}$ for given variance parameters $\theta^{(2)}_{\{j\}}$ and optimization over $\theta^{(2)}$ for given $\theta^{(1)}_{\{j\}}$. Second, as suggested by Watson and Engle (1983), the optimization over $\theta^{(1)}$ is not finalized for each j, but rather a single Newton step is implemented for each iteration of the procedure. Neither of these departures from the basic EM algorithm hinders reasonable convergence properties.

A Newton step in the estimation of $\theta^{(1)}$ for given $\theta^{(2)}_{\{j\}}$ yields the estimate in the $(j+1)$-th step

$$\theta^{(1)}_{\{j+1\}} = (\Xi'_{\{j\}} G_{\{j\}} \Xi_{\{j\}})^{-1} \Xi'_{\{j\}} (g_{\{j\}} - G_{\{j\}} \xi_{\{j\}}). \tag{4.17}$$

The derivation of (4.17) and expressions for $\Xi_{\{j\}}$, $\xi_{\{j\}}$, $g_{\{j\}}$ and $G_{\{j\}}$ can be found in appendix 4.B. Finally, the free variance parameters h_i, collected in $\theta^{(2)}$, are estimated using the derivative of $Q(\theta, \theta_{\{j\}})$ with respect to H; see (4.36). The estimate is given by the diagonal elements of

$$\frac{1}{n} L_{\{j\}} := \frac{1}{n} \mathrm{E}_{\theta_{\{j\}}} \sum_{t=1}^{n} \varepsilon_t \varepsilon_t' = \frac{1}{n} (D_{\{j\}} - Z E'_{\{j\}} - E_{\{j\}} Z' + Z F_{\{j\}} Z').$$

For using gradient-based methods in later steps of the maximization, the likelihood score can be obtained with only one run of a state smoothing algorithm. This has been shown by Koopman and Shephard (1992), who draw on the result

$$\left. \frac{\partial Q(\theta, \theta_{\{j\}})}{\partial \theta} \right|_{\theta_{\{j\}}} = \left. \frac{\partial l(\theta)}{\partial \theta} \right|_{\theta_{\{j\}}},$$

where $l(\theta)$ denotes the Gaussian log-likelihood of the model. Evaluation of the score for our model can therefore be based on (4.34) and (4.36).

An estimate of the covariance matrix can be computed using an analytical expression for the information matrix. Denoting by v_t and F_t the model residuals and forecast error variances obtained from the Kalman filter, the i-th element of the gradient vector

for observation t is given by

$$\frac{\partial l_t(\theta)}{\partial \theta_i} = -\frac{1}{2}\operatorname{tr}\left[\left(F_t^{-1}\frac{\partial F_t}{\partial \theta_i}\right)(I - F_t^{-1}v_t v_t')\right] + \frac{\partial v_t'}{\partial \theta_i}F_t^{-1}v_t, \tag{4.18}$$

while the ij-th element of the information matrix $\mathscr{I}(\theta)$ is

$$\mathscr{I}_{ij}(\theta) = \frac{1}{2}\sum_{t=1}^{n}\operatorname{tr}\left[F_t^{-1}\frac{\partial F_t}{\partial \theta_i}F_t^{-1}\frac{\partial F_t}{\partial \theta_j}\right] + \mathrm{E}_\theta\left[\sum_{t=1}^{n}\frac{\partial v_t'}{\partial \theta_i}F_t^{-1}\frac{\partial v_t}{\partial \theta_j}\right]; \tag{4.19}$$

see Harvey (1991, section 3.4.5). To obtain a feasible estimator $\hat{\mathscr{I}}(\hat{\theta})$, either the expectation term in (4.19) is omitted, as suggested by Harvey (1991) and as we do in the empirical application in section 4.5, or the techniques of Cavanaugh and Shumway (1996) may be used to compute the exact Fisher information. An estimate of the covariance matrix of the estimator is then given by

$$\widehat{\mathrm{Var}}_{\mathrm{info}}(\hat{\theta}) = \hat{\mathscr{I}}(\hat{\theta})^{-1}, \tag{4.20}$$

or by the sandwich form

$$\widehat{\mathrm{Var}}_{\mathrm{sand}}(\hat{\theta}) = \hat{\mathscr{I}}(\hat{\theta})^{-1}\left[\sum_{t=1}^{n}\left.\frac{\partial l_t(\theta)}{\partial \theta}\right|_{\hat{\theta}}\left.\frac{\partial l_t(\theta)}{\partial \theta'}\right|_{\hat{\theta}}\right]\hat{\mathscr{I}}(\hat{\theta})^{-1}, \tag{4.21}$$

which is robust to certain violations of the model assumptions; see White (1982). In section 4.5.3, we will contrast these methods also to the results obtained by the bootstrap.

We note that an asymptotic theory for maximum likelihood estimation in the fractionally cointegrated state space setup is not available. Certain functions of the parameter estimates are expected to exert nonstandard asymptotic behavior, especially in the nonstationary case $d_j > 0.5$ for some j. However, normal and mixed normal asymptotics have been established and conventional methods of inference justified in different parametric fractional cointegration settings as well as in state space models with common unit root components (Chang, Miller, and Park, 2009; Chang, Jiang, and Park, 2012). We thus use standard parameter inference in the empirical application below, bearing the preceding caveats in mind.

Our estimation approach can be straightforwardly generalized to some very relevant practical situations. To include a treatment of further components causing nonstationarity such as deterministic trends or exogenous regressors, one can use diffuse initialization of one or more of the states which may be based on Koopman (1997). Since we

have discussed maximum likelihood estimation under a setting where all data in y_t are available, our algorithms can be generalized for arbitrary patterns of missing data using the approach of Banbura and Modugno (2012). The computational refinements of Jungbacker and Koopman (2014) may be used for very high-dimensional datasets.

4.4 A Monte Carlo Study

We study the performance of the described methods for several stylized processes which are nested in the general setup (4.1). The simulation study is designed to answer several questions. Firstly, we assess whether the finite-order ARMA approximation of the state space system performs well as compared to other approaches. Secondly, we investigate joint estimation of memory parameters and cointegration vectors in bivariate fractional systems with and without polynomial cointegration. To be able to assess the performance and potential gains of our parametric approach, we also consider popular semiparametric approaches as benchmarks. Thirdly, the precision of quantifying cointegration relations is also studied in its relation to the dimension of the observed time series.

For all processes, we simulate $R = 1000$ iterations for each specification and estimate the models using the true parameters as starting values for maximum likelihood estimation. The precision of the estimators is assessed by the root mean squared error (RMSE) criterion or the bias or median errors of the parameter estimators. We vary over different sample sizes $n \in \{250, 500, 1000\}$ which covers relevant situations in macroeconomics and finance.

4.4.1 Finite State Approximations in a Univariate Setup

As the simplest stylized setup of our model, we first assess the fractional integration plus noise case, which has been studied in a stationary setup, e.g., by Grassi and de Magistris (2012). For mutually independent ξ_t and ε_s, the data generating process is given by

$$\begin{aligned} y_t &= x_t + \varepsilon_t, \qquad t = 1, \dots, n, \\ \Delta^d x_t &= \xi_t, \quad \xi_t \sim NID(0, q), \quad \varepsilon_t \sim NID(0, 1). \end{aligned} \tag{4.22}$$

For the signal to noise ratio we consider $q \in \{0.5, 1, 2\}$, while the memory parameters $d \in \{0.25, 0.5, 0.75\}$ cover cases of asymptotically stationary and nonstationary fractional

integration. We estimate the free parameters d, q and the noise variance h by maximum likelihood using the state space approach.

We apply different approximations to avoid an otherwise n-dimensional state process. Firstly, the ARMA(v,w) approximation given by (4.12) and (4.14) is considered, setting $v = w \in \{2,3,4\}$. The corresponding estimators are denoted as $\hat{d}_{v,w}$ in the result tables. Additionally, we assess truncations of the autoregressive representation of the fractional process at $m = 20$ and $m = 50$ lags, and label these estimators $\hat{d}_{AR20}$ and $\hat{d}_{AR50}$, respectively. Furthermore, moving average representations are used, also with a truncation at $m = 20$ and $m = 50$ lags ($\hat{d}_{MA20}$ and $\hat{d}_{MA50}$). Finally, we employ the exact local Whittle ($\hat{d}_{EW}$) estimator of Shimotsu and Phillips (2005) as well as the univariate exact local Whittle approach ($\hat{d}_{UEW}$) as defined by (Sun and Phillips, 2004), which accounts for additive $I(0)$ perturbations. For both semiparametric estimators of the fractional integration order, we use $m = \lfloor n^{0.65} \rfloor$ Fourier frequencies which outperforms other choices for our data generating processes.

The root mean squared errors for estimates of d for this setup are shown in table 4.1. Not surprisingly, for this stylized process with only three free parameters, the parametric approaches clearly outperform the semiparametric Whittle estimators. For the EW approach, the performance gets worse for more volatile noise processes (lower q), which is not the case for the UEW estimator. The bias of the EW estimator is negative due to the additive noise; see table 4.2 and also Sun and Phillips (2004). In contrast, the UEW estimator is positively biased, independently of q. Overall, the latter has inferior estimation properties, so that we do not show the UEW results for the other data generating processes.

Focusing on the state space approximations, we find that the ARMA approach for $v,w \geq 3$ is always among the best competitors. Overall, the ARMA(3,3) and ARMA(4,4) approximations exert a very similar performance, and their superiority does not seem to depend on the specification of d and q. The truncation methods, in contrast, show mixed results. The moving average approximation tends to dominate the autoregressive one for smaller $d < 0.5$, which mirrors the conclusion from Grassi and de Magistris (2012) in their stationary setting. However, we find that the autoregression is better whenever nonstationary $d \geq 0.5$ or higher signal to noise ratios are considered.

Directing attention to table 4.2, we find that the bias for the ARMA approach for $v,w \geq 3$ does not add significantly to the estimation errors. Often, it does not appear until the third decimal place. The bias is generally small also for the truncation approaches, but there exist some situations where it is noticeable, mostly for larger d. There, larger sample sizes even tend to increase the bias, while higher truncation lags do not always

Table 4.1: Root mean squared error (RMSE) for memory parameters in DGP1 (4.22). The columns show maximum likelihood estimators under ARMA(v,w) approximations of the fractional process with $v = w \in \{2,3,4\}$ ($\hat{d}_{v,w}$). Additionally, the truncated AR(m) representation ($\hat{d}_{ARm}$), and truncated MA(m) representations ($\hat{d}_{MAm}$) are given. Finally, we show the exact local Whittle ($\hat{d}_{EW}$) and the univariate exact local Whittle estimator ($\hat{d}_{UEW}$), each with $\lfloor n^{0.65} \rfloor$ Fourier frequencies.

q	d	n	$\hat{d}_{2,2}$	$\hat{d}_{3,3}$	$\hat{d}_{4,4}$	$\hat{d}_{AR20}$	$\hat{d}_{AR50}$	$\hat{d}_{MA20}$	$\hat{d}_{MA50}$	$\hat{d}_{EW}$	$\hat{d}_{UEW}$
.5	.25	250	.113	.112	.115	.122	.116	.117	.112	.225	.408
		500	.070	.069	.069	.073	.070	.073	.070	.176	.350
		1000	.048	.047	.047	.053	.049	.052	.049	.143	.286
	.50	250	.090	.091	.091	.119	.091	.105	.094	.223	.340
		500	.063	.063	.063	.080	.065	.082	.067	.173	.281
		1000	.043	.043	.043	.077	.052	.078	.055	.130	.220
	.75	250	.093	.090	.093	.101	.090	.193	.110	.208	.287
		500	.067	.065	.065	.080	.065	.198	.087	.160	.236
		1000	.047	.044	.043	.086	.053	.252	.072	.120	.178
1.0	.25	250	.078	.078	.078	.081	.079	.080	.078	.204	.408
		500	.053	.053	.053	.055	.053	.055	.053	.160	.354
		1000	.038	.037	.037	.041	.038	.040	.038	.126	.279
	.50	250	.073	.073	.073	.080	.074	.082	.073	.201	.333
		500	.052	.051	.051	.063	.052	.064	.054	.158	.283
		1000	.037	.035	.035	.061	.042	.061	.045	.120	.213
	.75	250	.076	.074	.074	.079	.074	.126	.091	.194	.284
		500	.056	.053	.053	.065	.053	.122	.071	.153	.237
		1000	.043	.036	.036	.071	.044	.161	.062	.116	.178
2.0	.25	250	.065	.065	.065	.067	.065	.067	.065	.192	.411
		500	.045	.044	.044	.046	.045	.046	.045	.154	.361
		1000	.032	.031	.031	.035	.033	.034	.032	.118	.288
	.50	250	.063	.063	.063	.068	.064	.070	.063	.193	.336
		500	.046	.044	.044	.054	.045	.055	.047	.153	.285
		1000	.033	.031	.031	.052	.036	.051	.039	.116	.209
	.75	250	.066	.064	.064	.068	.064	.095	.083	.188	.286
		500	.051	.045	.045	.056	.046	.096	.070	.150	.235
		1000	.055	.032	.031	.062	.038	.130	.057	.114	.177

lessen the problem.

In sum, we find superior performance of the ARMA approximations. The ARMA(3,3) approach appears sufficient in typical empirical applications. This finding is very appreciable in light of the great reduction in computational effort: a fractional component is represented by 3 states, rather than by 50 in a truncation setup with inferior performance. Overall, the differences between the approximations account for a small fraction of the overall estimation uncertainty, even for poor approximations in relative terms and even in this stylized setting with high overall estimation precision. Together with the finding of accurate ARMA-approximations in section (4.3.1), this suggests that the need of approximations might not be a serious obstacle to the state space modeling of fractional cointegration.

4.4.2 A Basic Fractional Cointegration Setup

The performance of the state space approach in estimating fractionally cointegrated systems is studied in a bivariate process with short-run dynamics,

$$\begin{aligned} y_{1t} = x_t + cz_{1t}, \qquad y_{2t} = x_t + (ce)z_{1t} + cz_{2t} \\ \Delta^d x_t = \xi_t, \quad \xi_t \sim NID(0,1), \\ (1-0.5L)z_{it} = \zeta_{it}, \quad \zeta_{it} \sim NID(0,1), \quad i = 1,2, \quad t = 1,\dots,n, \end{aligned} \tag{4.23}$$

where again the innovation processes are mutually independent. We vary over values of the fractional integration order $d \in \{0.25, 0.5, 0.75\}$ and the perturbation parameter $c \in \{0.5, 1, 2\}$ and introduce short-memory correlation between the processes, which will be governed by different values of $e \in \{0, 0.5, 1\}$.

Here and henceforth, we apply the ARMA(3,3) approximation for maximum likelihood estimation of the unknown model parameters. In the current setup, the latter consist of the eight entries in $\theta' = (d, \phi_1, \phi_2, \Lambda_{11}, \Lambda_{21}, \Gamma_{11}, \Gamma_{21}, \Gamma_{22})$, while the variance parameters are normalized to achieve identification. To contrast the properties to standard semiparametric approaches again, we apply the EW estimator componentwise to the univariate processes and investigate the mean of the univariate estimates. For the cointegration relation we apply the narrow-band least squares estimator which has been studied by Robinson and Marinucci (2001) in the nonstationary single equation case and by Hualde (2009) in a setup with cointegration subspaces. We follow the literature which suggests a small number of frequencies and use $\lfloor n^{0.3} \rfloor$, amounting to 5, 6 and 7 frequencies for our sample sizes.

Table 4.2: Bias for memory parameters in DGP1 (4.22). The columns show maximum likelihood estimators under ARMA(v,w) approximations of the fractional process with $v = w \in \{2,3,4\}$ ($\hat{d}_{v,w}$). Additionally, the truncated AR(m) representation ($\hat{d}_{ARm}$), and truncated MA(m) representations ($\hat{d}_{MAm}$) are given. Finally, we show the exact local Whittle ($\hat{d}_{EW}$) and the univariate exact local Whittle estimator ($\hat{d}_{UEW}$), each with $\lfloor n^{0.65} \rfloor$ Fourier frequencies.

q	d	n	$\hat{d}_{2,2}$	$\hat{d}_{3,3}$	$\hat{d}_{4,4}$	$\hat{d}_{AR20}$	$\hat{d}_{AR50}$	$\hat{d}_{MA20}$	$\hat{d}_{MA50}$	$\hat{d}_{EW}$	$\hat{d}_{UEW}$
.5	.25	250	-.019	-.019	-.019	-.009	-.016	-.011	-.018	-.125	.016
		500	-.009	-.010	-.010	.003	-.006	-.001	-.007	-.103	.034
		1000	-.004	-.005	-.005	.010	.001	.007	-.000	-.091	.044
	.50	250	-.012	-.010	-.010	.017	-.005	.010	-.012	-.125	.060
		500	-.011	-.007	-.007	.027	.003	.023	-.002	-.092	.059
		1000	-.011	-.004	-.004	.045	.013	.042	.011	-.069	.051
	.75	250	-.015	-.009	-.008	.013	-.006	.034	-.045	-.099	.053
		500	-.016	-.007	-.006	.031	.003	.070	-.031	-.065	.051
		1000	-.011	-.007	-.004	.058	.015	.138	-.009	-.043	.040
1.0	.25	250	-.011	-.011	-.011	-.004	-.009	-.007	-.010	-.088	.072
		500	-.008	-.007	-.008	.002	-.005	-.002	-.006	-.070	.075
		1000	-.003	-.003	-.003	.008	.001	.005	.000	-.061	.071
	.50	250	-.011	-.008	-.008	.010	-.004	.001	-.012	-.083	.095
		500	-.011	-.007	-.006	.019	.001	.012	-.005	-.059	.086
		1000	-.010	-.004	-.003	.034	.010	.029	.007	-.045	.063
	.75	250	-.011	-.007	-.007	.010	-.005	-.014	-.054	-.065	.077
		500	-.005	-.008	-.006	.024	.001	.012	-.043	-.043	.067
		1000	.013	-.006	-.003	.047	.012	.058	-.022	-.030	.049
2.0	.25	250	-.008	-.008	-.008	-.002	-.006	-.005	-.007	-.061	.107
		500	-.007	-.006	-.006	.001	-.004	-.002	-.005	-.047	.106
		1000	-.003	-.003	-.003	.007	.001	.004	.000	-.041	.098
	.50	250	-.008	-.006	-.006	.009	-.003	-.004	-.012	-.056	.117
		500	-.007	-.006	-.006	.016	.000	.005	-.006	-.040	.100
		1000	-.005	-.003	-.003	.029	.008	.020	.004	-.031	.071
	.75	250	-.002	-.006	-.005	.009	-.003	-.039	-.059	-.045	.093
		500	.010	-.007	-.006	.020	.000	-.015	-.050	-.031	.075
		1000	.040	-.004	-.003	.039	.010	.025	-.032	-.023	.053

Since the cointegration vectors are not identified without further restrictions, we investigate the angle ϑ between true and estimated cointegration spaces. Nielsen (2010, equation (22)) provides an expression for the sine of this angle, which is given for the j-th cointegration (sub)space in our framework by

$$\sin(\vartheta_j) = \frac{\operatorname{tr}(\Lambda^{(1:j)'}\hat{B}_j)}{\|\Lambda^{(1:j)}\|\|\hat{B}_j\|}, \tag{4.24}$$

where $\hat{B}_j$ is an estimated $p \times (p - s_j)$ cointegration matrix. In the current bivariate setup with $q = 1$ and $p - s_1 = 1$ cointegration relation, we have $\hat{B} = \hat{\Lambda}_\perp$ for the maximum likelihood estimator and $\hat{B}_{NB} = (1, -\hat{\beta}_{NB})'$ for the narrow-band least squares estimator $\hat{\beta}_{NB}$ applied to $y_{1t} = \beta y_{2t} + \text{error}$. Values of $\sin(\vartheta_j)$ close to zero indicate precise estimates and thus we compute the corresponding root mean squared error criterion as the square root of $\frac{1}{R}\sum_{i=1}^{R}\sin(\vartheta_j^i)^2$ in what follows.

In table 4.3 we show root mean squared errors for memory parameters ($\hat{d}^{ML}$ and $\hat{d}^{EW}$) and evaluate estimated cointegration spaces (by ϑ^{ML} and ϑ^{NB}) applying either the maximum likelihood or the semiparametric technique, respectively. Regarding the memory estimators, we find relatively large errors for this data generating process, with root mean squared errors frequently around 0.2 or larger, most prominently when the variances of the short-memory processes are large ($c = 2$). The Whittle estimator often performs better than maximum likelihood, especially for smaller c and d and in smaller samples.

For estimating the cointegration space, however, the state space approach appears worthwhile and always outperforms narrow band least squares for this process. Not surprisingly, strong cointegration ($d = 0.75$) is precisely estimated, as is cointegration with small short-memory disturbances ($c = 0.5$). While the relative superiority of maximum likelihood is not changed for different cointegration strengths, we find that strong perturbations foster the favorability of the state space estimators. For $c = 2$, the RMSE of the semiparametric approach often outnumbers the parametric RMSE by a factor of three.

Short memory correlation as introduced through $e > 0$ overall decreases the precision of the memory estimators. Interestingly, however, the performance of the cointegration estimators improves when $e > 0$ is considered. This is the case for both the maximum likelihood and the narrow band approach. To gain some insights into this finding, we assess the typical signed errors of the cointegration estimates. To this end, we consider a normalization of the cointegration vectors $(1, -\beta)$, and assess estimated β for both

Table 4.3: RMSE for parameters in DGP2 (4.23) for different specifications. The estimators arranged in columns are the ML estimator for d ($\hat{d}^{ML}$), the exact local Whittle estimator for d ($\hat{d}^{EW}$), the ML estimator for the cointegration space (ϑ^{ML}) and narrow band least squares for the cointegration space (ϑ^{NB}). The RMSE for cointegration spaces is based on the sine of the angle ϑ between the true and the estimated space (4.24).

			$e=0$				$e=0.5$				$e=1$			
c	d	n	$\hat{d}^{ML}$	$\hat{d}^{EW}$	ϑ^{ML}	ϑ^{NB}	$\hat{d}^{ML}$	$\hat{d}^{EW}$	ϑ^{ML}	ϑ^{NB}	$\hat{d}^{ML}$	$\hat{d}^{EW}$	ϑ^{ML}	ϑ^{NB}
.5	.25	250	.197	.024	.098	.178	.222	.026	.098	.131	.233	.031	.093	.127
		500	.108	.037	.059	.132	.136	.040	.055	.097	.177	.048	.057	.096
		1000	.053	.042	.042	.105	.081	.046	.037	.077	.106	.057	.035	.078
	.50	250	.189	.037	.061	.070	.203	.042	.058	.054	.213	.057	.050	.051
		500	.107	.043	.032	.044	.120	.048	.025	.034	.139	.064	.023	.031
		1000	.074	.042	.019	.029	.086	.047	.015	.022	.095	.062	.013	.021
	.75	250	.162	.040	.023	.026	.175	.045	.033	.021	.189	.058	.020	.020
		500	.111	.040	.010	.015	.115	.044	.008	.012	.129	.056	.007	.011
		1000	.092	.035	.005	.008	.090	.039	.004	.007	.093	.049	.004	.006
1.0	.25	250	.215	.053	.147	.383	.265	.055	.140	.256	.288	.060	.142	.232
		500	.135	.083	.106	.317	.175	.086	.101	.210	.215	.094	.113	.194
		1000	.061	.097	.071	.263	.109	.103	.066	.176	.144	.113	.063	.170
	.50	250	.211	.121	.095	.166	.257	.130	.083	.122	.280	.151	.074	.118
		500	.129	.131	.054	.102	.156	.140	.042	.076	.185	.163	.039	.075
		1000	.076	.125	.031	.066	.090	.135	.025	.050	.103	.158	.021	.049
	.75	250	.158	.119	.032	.059	.185	.128	.035	.045	.224	.153	.027	.043
		500	.111	.111	.016	.032	.125	.119	.012	.025	.145	.143	.011	.022
		1000	.086	.097	.009	.017	.087	.105	.007	.014	.094	.125	.006	.013
2.0	.25	250	.301	.079	.174	.576	.324	.080	.156	.368	.350	.081	.143	.306
		500	.221	.124	.147	.545	.265	.126	.136	.337	.277	.129	.126	.280
		1000	.153	.150	.117	.499	.221	.153	.112	.312	.238	.158	.106	.269
	.50	250	.293	.231	.125	.363	.348	.238	.110	.243	.365	.253	.095	.222
		500	.202	.249	.075	.250	.254	.258	.073	.171	.276	.275	.063	.163
		1000	.123	.247	.047	.167	.153	.257	.041	.118	.178	.278	.036	.117
	.75	250	.200	.264	.047	.141	.231	.278	.040	.103	.274	.311	.035	.100
		500	.137	.248	.021	.071	.162	.263	.015	.054	.195	.298	.014	.052
		1000	.091	.220	.011	.038	.105	.235	.008	.029	.122	.268	.007	.028

approaches. The median errors $(\text{median}_i(\hat{\beta}_j^i) - \beta_j)$ for this data generating process are shown in table 4.4.[1]

The typical deviations for the narrow band estimates exert a negative median bias of the estimates. A positive correlation between the short-memory components appears to work in the opposite direction so that the negative bias is reduced. In contrast, we find that the maximum likelihood estimators are essentially median-unbiased. Here, correlation between the short-memory components may improve the distinction between short and long-memory components and hence reduce variability.

4.4.3 Correlated Fractional Shocks and Polynomial Cointegration

A further simulation setup is concerned with correlation between the fractional components and possible polynomial cointegration. We consider

$$\begin{aligned} y_{1t} &= x_{1t} + a x_{2t} + \varepsilon_{1t}, \qquad y_{2t} = x_{1t} - a x_{2t} + \varepsilon_{2t} \\ \Delta^{d_i} x_{it} &= \xi_{it}, \quad \xi_{it} \sim NID(0,1), \quad \text{Corr}(\xi_{1t}, \xi_{2t}) = r \\ \varepsilon_{it} &\sim NID(0,1), \quad i = 1,2, \quad t = 1,\ldots,n. \end{aligned} \tag{4.25}$$

Here, correlation between the innovations to the fractional processes is introduced through the parameter r. Besides the standard setting $r = 0$, we refrain from the assumption of independent components for $r = 0.5$, while $r = 1$ amounts to $\xi_{1t} = \xi_{2t}$ which is the case of polynomial cointegration mentioned in the end of section 4.2.1. Combinations of $d_1 \in \{0.2, 0.4\}$ and $d_2 \in \{0.6, 0.8\}$ contrast relatively weak and strong cases of cointegration, while the importance of the component x_{2t} varies with $a \in \{0.5, 1, 2\}$. We treat $\theta = (d_1, d_2, \Lambda_{11}, \Lambda_{21}, \Lambda_{12}, \Lambda_{22}, r, h_{11}, h_{22})'$ as free parameters, but also investigate estimates imposing the singularity $r = 1$ when it is appropriate.

Consider the results for $r = 0.5$ first. The root mean squared errors, shown in table 4.5, include estimators of cointegration spaces as above (evaluated by ϑ_1^{ML} and ϑ_1^{NB} in the table). Now, there are two memory parameters to be estimated either by maximum likelihood ($\hat{d}_1^{ML}$ and $\hat{d}_2^{ML}$) or by the Whittle approach ($\hat{d}_1^{EW}$ and $\hat{d}_2^{EW}$). Semiparametric estimates of d_2 are obtained from the narrow band least squares residuals. The table also contains the maximum likelihood estimate of the correlation parameter r ($\hat{r}^{ML}$).

For most parameter settings, we observe that the parametric memory estimators perform satisfactorily and in most cases outperform the semiparametric approach, most

[1] For the state space approach an estimate for β is given by $\hat{\beta}_{ml} = \hat{\Lambda}_{21}/\hat{\Lambda}_{11}$ and produces large outliers for $\hat{\Lambda}_{11} \approx 0$. It is hence informative to compute an outlier-robust measure of the typical signed deviation.

Table 4.4: Median errors for parameters in DGP2 (4.23) for different specifications. The estimators arranged in columns are the ML estimator for d ($\hat{d}^{ML}$), the exact local Whittle estimator for d ($\hat{d}^{EW}$), the ML estimator for the cointegration coefficient (β^{ML}) and narrow band least squares for the cointegration coefficient (β^{NB}).

			$e=0$				$e=0.5$				$e=1$			
c	d	n	$\hat{d}^{ML}$	$\hat{d}^{EW}$	β^{ML}	β^{NB}	$\hat{d}^{ML}$	$\hat{d}^{EW}$	β^{ML}	β^{NB}	$\hat{d}^{ML}$	$\hat{d}^{EW}$	β^{ML}	β^{NB}
.5	.25	250	-.016	-.024	-.002	-.192	-.012	-.026	-.007	-.134	-.014	-.030	-.003	-.165
		500	-.015	-.037	-.008	-.143	-.006	-.040	-.006	-.106	.003	-.047	-.000	-.127
		1000	-.007	-.042	-.000	-.119	.000	-.046	-.007	-.086	.014	-.055	-.000	-.107
	.50	250	.027	-.037	.000	-.047	.025	-.041	-.003	-.037	.013	-.054	-.002	-.049
		500	.025	-.043	.001	-.024	.026	-.048	-.001	-.019	.031	-.062	.001	-.028
		1000	.025	-.042	.000	-.017	.029	-.047	-.000	-.012	.048	-.059	-.001	-.015
	.75	250	.060	-.040	-.000	-.010	.039	-.045	-.001	-.008	.019	-.056	-.000	-.010
		500	.062	-.040	.000	-.003	.054	-.044	-.001	-.002	.046	-.055	.000	-.004
		1000	.069	-.035	-.000	-.002	.064	-.038	-.000	-.001	.058	-.048	-.000	-.001
1.0	.25	250	-.020	-.053	-.003	-.489	-.011	-.055	-.018	-.326	-.013	-.060	-.022	-.327
		500	-.012	-.083	-.007	-.416	-.004	-.086	-.022	-.269	.002	-.093	-.007	-.282
		1000	-.004	-.097	-.001	-.345	-.002	-.103	-.007	-.238	-.000	-.112	.005	-.259
	.50	250	.018	-.121	.004	-.180	.017	-.130	.001	-.121	.019	-.149	-.002	-.146
		500	.021	-.131	.002	-.098	.024	-.140	-.003	-.072	.025	-.160	.001	-.091
		1000	.018	-.125	.003	-.062	.022	-.134	-.000	-.047	.030	-.155	-.001	-.057
	.75	250	.044	-.119	.002	-.035	.030	-.128	-.002	-.029	.025	-.150	.001	-.036
		500	.045	-.111	.002	-.013	.032	-.119	-.001	-.008	.030	-.140	-.001	-.015
		1000	.054	-.097	.001	-.007	.049	-.104	-.000	-.004	.044	-.123	-.000	-.006
2.0	.25	250	-.031	-.079	-.049	-.786	-.039	-.080	-.105	-.492	-.027	-.081	-.131	-.443
		500	-.016	-.124	-.031	-.753	-.024	-.126	-.073	-.457	-.034	-.129	-.068	-.413
		1000	-.007	-.150	-.000	-.675	-.015	-.153	-.018	-.428	-.020	-.158	-.023	-.405
	.50	250	.013	-.231	-.002	-.454	.001	-.238	-.010	-.304	-.017	-.252	-.024	-.313
		500	.018	-.249	-.002	-.305	.015	-.258	-.002	-.207	.010	-.274	-.011	-.230
		1000	.017	-.247	.002	-.201	.018	-.257	-.002	-.144	.012	-.276	-.004	-.169
	.75	250	.002	-.264	.004	-.128	-.011	-.278	-.002	-.095	-.014	-.308	-.002	-.115
		500	.012	-.248	.004	-.055	.007	-.263	-.001	-.038	.006	-.294	-.000	-.053
		1000	.028	-.220	.002	-.025	.018	-.234	.000	-.018	.011	-.265	-.000	-.024

Table 4.5: RMSE for parameters in DGP3 (4.25) with $r = 0.5$. The estimators arranged in columns are the ML estimators for d_1 and d_2 ($\hat{d}_1^{ML}$ and $\hat{d}_2^{ML}$), the EW estimator for d_1 and d_2 ($\hat{d}_1^{EW}$ and $\hat{d}_2^{EW}$), the ML and NBLS estimators for the cointegration space $\mathscr{S}^{(1)}$ (ϑ_1^{ML} and ϑ_1^{NB}), as well as ML for r ($\hat{r}^{ML}$). The RMSE for cointegration spaces is based on the sine of the angle ϑ_j between the true and the estimated space (4.24).

a	d2	d1	n	$\hat{d}_1^{ML}$	$\hat{d}_1^{EW}$	$\hat{d}_2^{ML}$	$\hat{d}_2^{EW}$	ϑ_1^{ML}	ϑ_1^{NB}	$\hat{r}^{ML}$
.5	.2	.6	250	.134	.142	.198	.172	.175	.070	.120
			500	.100	.112	.153	.143	.112	.051	.094
			1000	.062	.081	.104	.126	.049	.038	.060
		.8	250	.118	.133	.192	.172	.075	.035	.053
			500	.090	.106	.153	.140	.041	.023	.032
			1000	.057	.079	.094	.123	.013	.014	.015
	.4	.6	250	.140	.137	.184	.214	.277	.122	.186
			500	.089	.108	.121	.178	.231	.102	.167
			1000	.053	.079	.087	.150	.136	.085	.147
		.8	250	.118	.131	.179	.214	.133	.064	.114
			500	.080	.105	.123	.176	.072	.048	.102
			1000	.053	.079	.090	.147	.032	.036	.072
2.0	.2	.6	250	.120	.247	.069	.119	.218	.361	.129
			500	.083	.216	.045	.089	.151	.222	.111
			1000	.059	.182	.030	.066	.093	.124	.066
		.8	250	.105	.248	.064	.113	.097	.137	.061
			500	.071	.200	.041	.086	.059	.074	.029
			1000	.052	.160	.028	.066	.033	.045	.016
	.4	.6	250	.124	.180	.067	.127	.429	.756	.221
			500	.085	.164	.047	.107	.366	.675	.206
			1000	.061	.147	.034	.085	.285	.551	.189
		.8	250	.109	.226	.068	.124	.215	.331	.164
			500	.076	.193	.045	.092	.154	.205	.148
			1000	.055	.164	.031	.069	.095	.129	.124

pronouncedly for a strong influence of the x_{2t} components ($a = 2$) and in larger samples. There is a distinction between specifications where state space estimates of cointegration spaces are superior and those where narrow band estimates dominate. Also there, higher values of a favor the parametric method. The correlation parameter is estimated with increasing precision in larger samples, while also the strength of the cointegration relation is relevant for estimation of this parameter. For $d_1 = d_2$, the correlation parameter (and also certain elements of Λ) would not be identifiable, and hence setups with small $d_1 - d_2$ are problematic.

For $r = 1$, we additionally consider the properties of estimators for the polynomial cointegration relation. To evaluate estimators of the polynomial cointegration spaces, note that the cointegration subspace for $(y_{1t}, y_{2t}, \Delta^{d_1-d_2} y_{2t})'$ is

$$\mathscr{S}^{(2)} = sp^{\perp} \begin{pmatrix} \Lambda^{(1)} & \Lambda^{(2)} \\ 0 & \Lambda_{12} \end{pmatrix}. \tag{4.26}$$

This is estimated replacing Λ_{ij} by its maximum likelihood estimators, where $r = 1$ is imposed. For the narrow band least squares estimator, this space is determined by $sp((1, -\hat{\beta}_1, -\hat{\beta}_2)')$, where the coefficients are narrow band least square estimates from $y_{1t} = \beta_1 y_{2t} + \beta_2 \Delta^{d_1-d_2} y_{2t} + \text{error}$ with d_1 and d_2 replaced by local Whittle estimates.

In table 4.6, the corresponding root mean squared errors are given. The elementary cointegration space is estimated with a very similar precision by the unrestricted estimator (see ϑ_1^{ML}) and the restricted estimator (see ϑ_1^{RML}, imposing $r = 1$). This is in accordance with the notably precise estimation of r in this case. The parametric estimators of the cointegration spaces are again better than semiparametric approaches (1) in large samples and (2) when a strong second fractional component is present. Overall, the results suggest that polynomial fractional cointegration analysis is feasible in our setup, while the maximum likelihood approach has reasonable estimation properties at least for larger sample sizes.

4.4.4 Cointegration Subspaces in Higher Dimensions

Until now, we have considered one- or two-dimensional processes in our simulations which limits the empirical relevance of our findings. We claim that modeling high-dimensional time series constitutes a strength of our approach, at least if suitably sparse parametrizations like the dimension-reduced process (4.11) are empirically rea-

Table 4.6: RMSE for parameters in DGP3 (4.25) with $r = 1$. The estimators arranged in columns are the ML estimator for d_1 and d_2 ($\hat{d}_1^{ML}$ and $\hat{d}_2^{ML}$), the restricted ML (setting $r = 1$), the ML and NBLS estimator for the cointegration space $\mathscr{S}^{(1)}$ (ϑ_1^{RML}, ϑ_1^{ML} and ϑ_1^{NB}), the restricted ML (setting $r = 1$) and NBLS estimator for the cointegration subspace $\mathscr{S}^{(2)}$ (ϑ_2^{RML} and ϑ_2^{NB}), as well as ML for r ($\hat{r}^{ML}$). The RMSE for cointegration spaces is based on the sine of the angle ϑ_j between the true and the estimated space (4.24).

a	d2	d1	n	$\hat{d}_1^{ML}$	$\hat{d}_2^{ML}$	ϑ_1^{ML}	ϑ_1^{RML}	ϑ_1^{NB}	ϑ_2^{RML}	ϑ_2^{NB}	$\hat{r}^{ML}$
.5	.2	.6	250	.149	.142	.154	.161	.095	.099	.093	.099
			500	.116	.102	.085	.087	.080	.056	.057	.056
			1000	.086	.070	.043	.043	.065	.029	.034	.029
		.8	250	.137	.120	.049	.049	.039	.035	.038	.035
			500	.088	.078	.020	.020	.028	.013	.018	.013
			1000	.063	.052	.008	.008	.019	.005	.009	.005
	.4	.6	250	.212	.185	.322	.337	.209	.145	.133	.145
			500	.146	.124	.241	.245	.196	.107	.099	.107
			1000	.107	.084	.137	.137	.180	.070	.072	.070
		.8	250	.155	.126	.105	.103	.092	.066	.049	.066
			500	.112	.090	.051	.050	.074	.034	.025	.034
			1000	.081	.058	.024	.023	.059	.015	.015	.015
2.0	.2	.6	250	.126	.076	.211	.210	.203	.066	.120	.066
			500	.090	.052	.125	.122	.206	.033	.073	.033
			1000	.062	.033	.063	.061	.196	.013	.029	.013
		.8	250	.095	.060	.058	.055	.094	.013	.040	.013
			500	.060	.042	.021	.021	.077	.004	.014	.004
			1000	.041	.030	.010	.010	.060	.002	.005	.002
	.4	.6	250	.125	.082	.376	.371	.820	.093	.276	.093
			500	.089	.055	.294	.290	.542	.069	.247	.069
			1000	.072	.040	.225	.212	.362	.043	.189	.043
		.8	250	.119	.067	.163	.162	.210	.050	.094	.050
			500	.076	.047	.077	.074	.204	.018	.040	.018
			1000	.049	.034	.031	.029	.183	.005	.014	.005

sonable. To assess the performance in higher dimensions, consider the process

$$\begin{aligned} y_{it} &= a x_{1t} + a(-1)^{i+1} x_{2t} + \varepsilon_{it}, \\ \Delta^{d_j} x_{jt} &= \xi_{jt}, \quad \xi_{jt} \sim NID(0,1), \\ \varepsilon_{it} &\sim NID(0,1), \quad j = 1,2, \quad i = 1,\dots,p, \quad t = 1,\dots,n, \end{aligned} \tag{4.27}$$

with mutually independent noise sequences. We now vary over the dimension $p \in \{3,10,50\}$, while again combinations of $d_1 \in \{0.2,0.4\}$ and $d_2 \in \{0.6,0.8\}$ are considered. The parameter $a \in \{0.5,1,2\}$ gives the relative importance of the fractional components and hence plays the role of a signal to noise ratio. We estimate d_j, Λ_{ij}, h_i for $j = 1,2$ and $i = 1,\dots,p$ as free parameters.

Along with the memory estimates, we show results for estimating the cointegration subspaces $\mathscr{S}^{(1)}$ and $\mathscr{S}^{(2)}$, which are straightforwardly obtained for the maximum likelihood approach and evaluated by ϑ_1^{ML} and ϑ_2^{ML}. The narrow-band least squares method (evaluated by ϑ_1^{NB} and ϑ_2^{NB}) estimates cointegration matrices under specific normalizations as above. Estimating $\mathscr{S}^{(1)}$, we construct $\hat{B}_1$ to have free entries $-\hat{\beta}_2$, …, $-\hat{\beta}_p$ in the first row and a $p-1$ identity matrix below, such that β_j is obtained from $y_{jt} = \beta_j y_{1t} + \text{error}$ for $j = 2,\dots,p$. In the estimation of $\mathscr{S}^{(2)}$, we have two free rows in $\hat{B}_2$ which are given by $(-\hat{\beta}_{13}, \dots, -\hat{\beta}_{1p})$, and $(-\hat{\beta}_{23}, \dots, -\hat{\beta}_{2p})$, respectively, and can be estimated from $y_{jt} = \beta_{1j} y_{1t} + \beta_{2j} y_{2t} + \text{error}$ for $j = 3,\dots,p$.

In table 4.7, results are shown for $a = 0.5$ while the other specifications yield qualitatively similar outcomes. The process allows for a precise estimation of both d_1 and d_2 by maximum likelihood. The estimates are clearly better than the Whittle estimates, which are again obtained by averaging univariate estimates for d_1 and using narrow band residuals to estimate d_2. An increasing dimension p leads to a better estimation by maximum likelihood which is not the case for the Whittle technique.

Also regarding the estimation of the cointegration spaces, maximum likelihood is superior. Both parametric and semiparametric estimators have smaller errors for higher dimension, whereas this "blessing of dimensionality" is more pronounced for the state space approach. Generally, the fraction between the maximum likelihood RMSE and the semiparametric RMSE decrease for larger p.

Not surprisingly, the case with strongest basic cointegration ($d_1 - d_2$ large) is the one with highest precision in estimating $\mathscr{S}^{(1)}$. For estimating $\mathscr{S}^{(2)}$, a slightly different logic applies, with a larger d_2 supporting the estimation. E.g., in the case $d_1 = 0.6$ and $d_2 = 0.4$ higher precision is achieved than for $d_1 = 0.6$ and $d_2 = 0.2$. Overall, we find that our approach profits from a suitable dimension-reduced structure which is not the

Table 4.7: RMSE for parameters in DGP4 (4.27) with $a = 0.5$. The estimators arranged in columns are the ML estimator for d_1 and d_2 ($\hat{d}_1^{ML}$ and $\hat{d}_2^{ML}$), the EW estimator for d_1 and d_2 ($\hat{d}_1^{EW}$ and $\hat{d}_2^{EW}$), the ML and NBLS estimator for the cointegration space $\mathscr{S}^{(1)}$ (ϑ_1^{ML} and ϑ_1^{NB}), and the ML and NBLS estimator for the cointegration subspace $\mathscr{S}^{(2)}$ (ϑ_2^{ML} and ϑ_2^{NB}). The RMSE for cointegration spaces is based on the sine of the angle ϑ_j between the true and the estimated space (4.24).

d2	d1	p	n	$\hat{d}_1^{ML}$	$\hat{d}_1^{EW}$	$\hat{d}_2^{ML}$	$\hat{d}_2^{EW}$	ϑ_1^{ML}	ϑ_1^{NB}	ϑ_2^{ML}	ϑ_2^{NB}
.2	.6	3	250	.127	.263	.109	.158	.074	.107	.036	.057
			500	.087	.224	.078	.131	.046	.075	.020	.035
			1000	.064	.187	.055	.115	.032	.048	.013	.023
		10	250	.063	.268	.069	.157	.013	.029	.013	.036
			500	.044	.226	.046	.129	.008	.019	.009	.033
			1000	.029	.191	.030	.115	.006	.013	.006	.027
		50	250	.054	.271	.059	.150	.002	.006	.002	.007
			500	.037	.228	.039	.128	.001	.004	.001	.006
			1000	.028	.189	.026	.113	.001	.002	.001	.005
	.8	3	250	.107	.274	.107	.167	.033	.047	.018	.028
			500	.077	.219	.076	.138	.020	.028	.009	.016
			1000	.059	.171	.053	.119	.013	.017	.005	.009
		10	250	.065	.281	.070	.164	.007	.013	.012	.036
			500	.045	.222	.047	.135	.004	.007	.009	.033
			1000	.030	.175	.030	.120	.002	.004	.006	.027
		50	250	.055	.285	.060	.158	.001	.002	.002	.007
			500	.037	.224	.039	.136	.001	.001	.001	.006
			1000	.029	.173	.027	.118	.000	.001	.001	.005
.4	.6	3	250	.122	.237	.121	.190	.163	.192	.034	.059
			500	.079	.204	.081	.150	.124	.159	.020	.035
			1000	.055	.173	.054	.125	.091	.129	.014	.024
		10	250	.062	.241	.065	.188	.028	.051	.011	.024
			500	.044	.204	.045	.149	.017	.043	.008	.017
			1000	.029	.175	.030	.124	.013	.035	.005	.011
		50	250	.054	.243	.059	.184	.004	.010	.002	.004
			500	.036	.207	.038	.148	.003	.008	.001	.003
			1000	.028	.174	.027	.123	.002	.006	.001	.002

case for the benchmark methods applied in this comparison.

4.5 An Application to Realized Covariance Modeling

We apply the fractional components approach to the modeling and forecasting of multivariate realized stock market volatility which has recently received considerable interest in the financial econometrics literature.

4.5.1 Data and Recent Approaches

We use the dataset of Chiriac and Voev (2011) which comprises realized variances and covariances from six US stocks, namely (1) American Express Inc., (2) Citigroup, (3) General Electric, (4) Home Depot Inc., (5) International Business Machines and (6) JP-Morgan Chase & Co for the period from 2000-01-01 to 2008-07-30 ($n = 2156$). The data are available from `http://qed.econ.queensu.ca/jae/2011-v26.6/chiriac-voev`.

Different transformations of the realized covariance matrices have been applied to fit dynamic models to data of this kind. Weigand (2014) discusses these transforms and considers a general framework nesting several previously applied approaches. His results suggest that applying linear models to a multivariate time series of log realized variances along with z-transformed realized correlations is a reasonable choice in practice. We follow this approach and base our empirical study on the 21-dimensional time series

$$y_t = (\log(X_{11,t}), \ldots, \log(X_{66,t}), Z_{21,t}, Z_{31,t}, \ldots, Z_{65,t})', \tag{4.28}$$

where X_t is the 6×6 realized covariance matrix at period t, and

$$Z_{ij,t} = 0.5[\log(1 + R_{ij,t}) - \log(1 - R_{ij,t})], \quad R_{ij,t} = \frac{X_{ij,t}}{\sqrt{X_{ii,t} X_{jj,t}}}.$$

All time series (grey) of log variances and their maxima and minima for a given day t (black) are depicted in figure 4.5, while z-transformed correlations are shown in figure 4.6.

Recent approaches to modeling realized covariance matrices have successfully used long-memory specifications (Chiriac and Voev, 2011), or found co-movements between the processes well-represented by dynamic factor structures; see Bauer and Vorkink (2011) and Gribisch (2013). Our specific transformation (4.28) also makes recent results on forecasting univariate realized variances applicable, where also factor models with long-memory dynamics have been proposed. While Beltratti and Morana (2006) use frequency-domain principal components techniques to assess the low-frequency co-movements, Luciani and Veredas (2012) apply time-domain principal components to

their high-dimensional series and apply fractional integration techniques to both estimated factors and idiosyncratic components. Very recently, Asai and McAleer (2014) have considered long-memory factor dynamics also for the modeling of realized covariance matrices, where again a semiparametric factor approach precedes a long-memory analysis in their two-step approach.

By applying our fractional components model DOFC (4.11) to the time series (4.28), we contribute to the literature in several ways. Our methods offer new insights in the integration and cointegration properties of stock market volatilities, for which fractional components structures of different integration orders have not been investigated so far. Fractional cointegration between variances and correlations is of particular interest for the understanding of longer-term portfolio hedging and systemic risk assessment, but has not found attention in the existing literature. Our state space approach for variances and correlations also features other relevant aspects of volatility modeling. It offers a separation into short-term and long-term components in the spirit of Engle and Lee (1999), directly accounts for measurement noise, and is applicable in datasets of higher dimensions. Practicability in case of missing values is achieved, while our parameter-driven model may be straightforwardly carried over to stochastic volatility frameworks for daily data in the spirit of Harvey, Ruiz, and Shephard (1994).

4.5.2 Preliminary Analysis and Model Specification

We investigate whether the constraints imposed in the DOFC model (4.11) are reasonable for the dataset under investigation. Semiparametric methods are used to assess these restrictions and to obtain reasonable starting values for the parametric estimation of our model.

The model (4.11) implies that there are $s+s_0$ components which govern the dynamics of y_t, and hence, for $p > s+s_0$, there is a dimension reduction in terms of the autocorrelation characteristics. Pan and Yao (2008) study time series with such properties and propose a sequential test to infer the dynamic dimension of the process, allowing for nonstationarity of the autocorrelated components. The algorithm sequentially finds the least serially correlated linear combinations of y_t, subsequently testing the null of no autocorrelation of this series.

Applying this approach to our dataset and thereby evaluating autocorrelations of possible white noise linear combinations using 3 lags, we do not reject the null for eight linear combinations which can hence be treated as white noise. For the ninth such combination, the p-value for the multivariate Ljung-Box test drops from 0.1935 to 0.0002,

so that the white noise hypothesis is rejected for reasonable significance levels. We conclude that there are $s + s_0 = 21 - 8 = 13$ components which account for the dynamic properties of the process. Pan and Yao (2008) also propose an estimator for the space of dynamic components $(x_t', z_t')'$, from which an identified rotation is selected as the full set of principal components and called the factors in what follows.

Our model implies that $(x_t', z_t')'$ and hence a suitable rotation of the factors can be modelled as $s + s_0$ univariate time series which are mutually orthogonal at all leads and lags. This corresponds to the notion of dynamic orthogonal components as introduced by Matteson and Tsay (2011) who provide methods to test for the presence of such a structure and to estimate the appropriate rotation. Using first differences of the factors to achieve stationarity as required by Matteson and Tsay (2011) for a suitable range of d_j, and using 3 lags in the cross-covariances, we find highly significant cross-correlations of the factors (the test statistic takes the value 4198.94 for a level 0.01 critical value of 625.80) while a dynamic orthogonal structure is not rejected for the rotated series, with a test statistic of 445.55 and a corresponding p-value close to one. The test result is robust to a specification in levels. In what follows, the dynamic orthogonal components are computed from the factors in levels which slightly outperforms the difference-approach in simulations with fractional processes which are available from the author.

Due to their dynamic orthogonality, the rotation of Matteson and Tsay (2011) identifies the single processes in $(x_t', z_t')'$ up to scale, sign and order. A preliminary analysis of the integration orders of x_t can hence be undergone by a univariate treatment of these series. We investigate these integration orders by the exact local Whittle estimator allowing for an unknown mean (Shimotsu, 2010).

A possible grouping of components with equal integration orders is assessed by the methods proposed by Robinson and Yajima (2002), with the modifications for possibly nonstationary integration orders by Nielsen and Shimotsu (2007). The specific-to-general approach of Robinson and Yajima (2002) sequentially tests for existence of $j = 1, 2, \ldots$ groups of equal integration orders. The sequence is terminated if for some j^* there is a grouping for which within-group equality is not rejected, and for $j^* > 1$ the grouping with highest p-value is selected. In our application, we restrict attention to possible groupings where for $\hat{d}_{i_1} > \hat{d}_{i_2} > \hat{d}_{i_3}$, there is no group including both $\hat{d}_{i_1}$ and $\hat{d}_{i_3}$ but not $\hat{d}_{i2}$. For the tests of equal integration orders within the sequential approach, we consider the Wald test proposed by Nielsen and Shimotsu (2007), jointly testing all hypothesized equalities for a given grouping. We choose $m = \lfloor n^{0.5} \rfloor = 46$ as bandwidth and set the trimming parameter h to zero, since the dynamic orthogonal components structure does not permit fractional cointegration.

The estimated integration orders for the dynamic orthogonal components range from 0.0087 to 0.7328 and indicate that some of the components may have short memory while others behave like stationary or nonstationary fractionally integrated processes. We clearly reject equality of all integration orders, while also each of the groupings in two groups can be rejected on a 0.01 significance level. For three groups, we do not reject the hypothesis of equal integration orders within groups. The sequential test for groups with equal memory yields $j^* = 3$ with a p-value of 0.3181, where groups of three ($\hat{d}^{(1)} = 0.6717$), seven ($\hat{d}^{(2)} = 0.3448$) and three ($\hat{d}^{(3)} = 0.0523$) components are identified, respectively. The hypothesis that $d^{(3)} = 0$ is not rejected. We may therefore treat the members of the third group as short-range dependent and belonging to z_t. Thus, $s_1 = 3$, $s_2 = 7$ and $s_0 = 3$ appear as a reasonable specification for model (4.11) due to the preliminary analysis.

We obtain starting values for the parametric estimation from this procedure. Firstly, d and ϕ are estimated from the dynamic orthogonal components. Secondly, from regressing observed data on standardized estimated orthogonal components with unit innovation variance, we obtain starting values for h, Λ and Γ, while certain columns of the latter matrices are rotated to satisfy the zero restrictions.

In very high-dimensional cases, the approach of Pan and Yao (2008) is not applicable, but Lam, Yao, and Bathia (2011) and Lam and Yao (2012) provide feasible methods for stationary settings and comment on possible extensions to nonstationarity. In cases where the dynamic orthogonal components specification (4.11) is not appropriate, but the general setup (4.1) is, a specification search and preliminary estimates for the integration and cointegration parameters of the more general model could be based on the algorithm of Hualde (2009) which is capable of identifying and estimating cointegration subspaces by semiparametric methods.

4.5.3 A Parametric Fractional Components Analysis

We proceed with maximum likelihood estimation of the fractional components model using the techniques of section 4.3. First, we complement the results of the previous section by a parametric specification search. After diagnostic checking of the selected model, we will take a closer look at its parameter estimates and implied long-run characteristics. Forecasts are then carried out in the next section.

As justified in section 4.3 and by Monte Carlo simulations, we use an ARMA(3,3) approximation for the fractional components. Constant terms are included by a further column c in the observation matrix and estimated along with the free elements of Λ and

Γ. Setting the autoregressive order of z_t to one and using starting values as described above, we estimate models with $q \in \{1,2,3\}$ groups of equal integration orders $d^{(j)} > 0$ and additional autoregressive components. The Bayesian information criterion (BIC) is used to select sizes $s_0,\ldots,s_q$ and the value of q with appropriate in-sample fit.[2] We apply the BIC even if consistency is not established in this fractional setting. We expect that existing results hold for specification choices not involving the fractional components, while it is not clear to what extent the results of Chang, Jiang, and Park (2012) carry over to the fractional setup. There, consistency of the BIC is shown for the number of stochastic trends in a unit root state space model.

The best models for each q are shown in table 4.8, where estimated integration orders are given along with the log-likelihood (log-lik) and the BIC. Regarding the integration orders, we find that for $q > 1$ estimates of $d^{(1)}$ are always above 0.5 suggesting nonstationarity of at least s_1 series in y_t. Overall, the models with $q = 2$ are superior, in particular the grouping in $s_1 = 2$ and $s_2 = 9$ fractional and $s_0 = 2$ nonfractional components. This specification is similar to the one selected by the semiparametric approach and also suggests a dynamic dimension of $s + s_0 = 13$. Interestingly, the same specification with full noise covariance matrix H is inferior ($BIC = -16.626$) as is the model with a full vector autoregressive matrix Φ ($BIC = -17.150$). Furthermore, considering more lags in z_t does not sufficiently improve the fit ($BIC = -17.155$ for $k = 2$, $BIC = -17.046$ for $k = 3$ and $BIC = -17.139$ for $k = 4$).

We conduct several diagnostic tests on standardized model residuals $e_{it} = v_{it} / \sqrt{F_{ii,t}}$, where v_t and F_t are filtered residuals and forecast error covariance matrices, respectively. The residuals corresponding to log variances and z-transformed correlations for the first three assets are plotted in figure 4.7, while residual autocorrelations are depicted in figure 4.8, autocorrelations of squared residuals in figure 4.9 and histograms of the residuals along with the normal density in figure 4.10. The visual inspection shows some but no overwhelming evidence against the model assumptions. Autocorrelation both of residuals and squared residuals are generally below 0.1 in absolute value and mostly within the $\pm$ 2 standard error bands which are shown as horizontal lines. Some deviations from normality are visible, but not the sort of skewness and fat tails observed

[2] Instead of estimating all reasonable combinations of s_0, ..., s_q for each q, we begin by the optimal grouping for a given q obtained from the semiparametric methods of the previous section. From this specification, denoted as $s_j^{\{0\}}$, $j = 0,\ldots,q$, we estimate all models characterized by $s_j \in \{s_j^{\{0\}} - 1, s_j^{\{0\}}, s_j^{\{0\}} + 1\}$, $j = 0,\ldots,q$. The model with least value of the BIC is selected and its indices denoted as $s_j^{\{1\}}$, and again models with indices close to $s_j^{\{1\}}$ are estimated and compared. This process is iterated until $s_j^{\{i\}} = s_j^{\{i-1\}}$ holds for all $j = 0,\ldots,q$. As a result, also the number of white noise combinations may differ from 8, the result of the semiparametric analysis in the previous section.

Table 4.8: Estimation results for different specifications of the models estimated in section 4.5.3. We show the combinations of s_j, $j = 0,\ldots,q$ with best values of the BIC for $q = 1$ (above), $q = 2$ (middle) and $q = 3$ (below).

s_1	s_2	s_3	s_0	log-lik	$d^{(1)}$	$d^{(2)}$	$d^{(3)}$	BIC
12			3	-19572.2	0.368			-17.116
10			4	-19553.8	0.390			-17.106
11			4	-19590.6	0.390			-17.101
12			4	-19620.7	0.383			-17.094
10			5	-19601.8	0.411			-17.087
11			5	-19636.4	0.406			-17.080
2	9		2	-19573.8	0.631	0.338		-17.157
2	10		1	-19538.0	0.596	0.319		-17.156
1	10		2	-19530.9	0.653	0.370		-17.146
2	10		2	-19605.0	0.551	0.303		-17.143
3	9		1	-19546.7	0.619	0.315		-17.139
2	8		3	-19575.9	0.634	0.353		-17.134
2	2	7	2	-19600.7	0.639	0.422	0.304	-17.129
2	3	7	2	-19666.6	0.565	0.417	0.252	-17.122
2	3	6	2	-19607.5	0.634	0.407	0.288	-17.121
2	4	6	2	-19676.5	0.634	0.412	0.234	-17.121
3	3	5	2	-19617.4	0.629	0.398	0.272	-17.119
3	2	6	2	-19601.0	0.618	0.395	0.292	-17.115

for models of untransformed residual variances and covariances.

Table 4.9 presents the diagnostic tests on standardized residuals. The p-values are shown for the Ljung-Box test (LB) and the ARCH-LM test for conditional heteroscedasticity (CH) for different lag length 5, 10 and 22. Additionally, the Jarque-Bera test result (JB) is shown in the last column. The null of no autocorrelation is not rejected at the 0.01 level for all but two or three residuals, depending on lag length. Clear evidence of conditional heteroskedasticity is found for the residuals of the log variance series, that is $e_{2t},\ldots,e_{6t}$, where also the normality assumption is clearly rejected, but also for a few correlation series such as $e_{15,t}$ or $e_{19,t}$. A more flexible data transformation like the matrix Box-Cox approach of Weigand (2014) would typically ameliorate these findings, but we do not follow this approach further here.

Estimates of several of the model parameters are shown in table 4.10. Along with the maximum likelihood estimates, we also show the mean of the estimators from a

Table 4.9: P-values of diagnostic tests for the residuals from the DOFC model (4.11) estimated in section 4.5.3. We conducted Ljung-Box tests for residual correlation (LB), ARCH-LM tests for conditional heteroskedasticity (CH), each with different lags, and Jarque-Bera tests (JB) for deviations from normality.

	LB5	LB10	LB22	CH5	CH10	CH22	JB
$e_{1,t}$	0.944	0.848	0.101	0.083	0.359	0.373	0.000
$e_{2,t}$	0.022	0.008	0.032	0.000	0.001	0.001	0.000
$e_{3,t}$	0.191	0.110	0.253	0.008	0.018	0.006	0.000
$e_{4,t}$	0.474	0.459	0.109	0.043	0.038	0.152	0.000
$e_{5,t}$	0.000	0.004	0.002	0.000	0.004	0.000	0.000
$e_{6,t}$	0.035	0.197	0.063	0.000	0.000	0.000	0.000
$e_{7,t}$	0.091	0.054	0.178	0.741	0.142	0.382	0.002
$e_{8,t}$	0.071	0.075	0.103	0.587	0.569	0.509	0.000
$e_{9,t}$	0.208	0.365	0.295	0.109	0.280	0.212	0.219
$e_{10,t}$	0.108	0.117	0.459	0.861	0.915	0.717	0.001
$e_{11,t}$	0.326	0.090	0.092	0.207	0.436	0.877	0.000
$e_{12,t}$	0.468	0.477	0.442	0.538	0.033	0.037	0.175
$e_{13,t}$	0.080	0.158	0.800	0.571	0.318	0.060	0.000
$e_{14,t}$	0.235	0.162	0.026	0.080	0.167	0.079	0.000
$e_{15,t}$	0.242	0.328	0.072	0.001	0.011	0.011	0.102
$e_{16,t}$	0.354	0.541	0.589	0.272	0.180	0.367	0.000
$e_{17,t}$	0.000	0.000	0.000	0.057	0.039	0.003	0.369
$e_{18,t}$	0.158	0.376	0.480	0.245	0.326	0.349	0.000
$e_{19,t}$	0.557	0.514	0.849	0.003	0.011	0.019	0.001
$e_{20,t}$	0.685	0.882	0.942	0.412	0.216	0.790	0.000
$e_{21,t}$	0.122	0.014	0.055	0.600	0.446	0.256	0.000

bootstrap resampling exercise with 1000 iterations and generally find a low bias for the corresponding estimates. Again, this finding supports the use of the ARMA(3,3) approximation of the fractional processes. We also show standard errors, obtained in three ways, namely by the bootstrap (SE.boot), using the sandwich form (4.21), denoted by SE.sand, and by the information matrix approach (4.20), labelled SE.info in the table. The different methods of computing standard errors give similar results, except for the variance parameters h_i, where the sandwich estimates are large compared to the others. Overall, including the parameters not shown in the table, the median ratio between bootstrap and sandwich standard errors is 1.31, while a typical sandwich estimate is 1.20 times larger than the corresponding estimate from the information matrix.

Table 4.10: Estimated parameters along with bootstrap mean and standard errors from bootstrap (SE.boot), sandwich (SE.sand) and information matrix (SE.info) as described in section 4.3.3 for the DOFC model (4.11) estimated in section 4.5.3.

	Estimate	Mean	SE.boot	SE.sand	SE.info
d_1	0.6308	0.6361	0.0190	0.0217	0.0178
d_2	0.3382	0.3334	0.0094	0.0116	0.0086
ϕ_1	0.2468	0.2360	0.0345	0.0417	0.0348
ϕ_2	0.0768	0.0636	0.0370	0.0419	0.0402
h_1	0.2028	0.1844	0.1122	0.1106	0.0759
h_2	0.3858	0.3727	0.0522	0.0551	0.0321
h_3	0.3289	0.3309	0.0930	0.0957	0.0714
h_4	0.1758	0.1638	0.1222	0.1371	0.0861
h_5	0.7649	0.7618	0.0558	0.0676	0.0482
h_6	0.2459	0.2413	0.0772	0.0810	0.0588
h_7	0.0615	0.0611	0.0037	0.0079	0.0027
h_8	0.0746	0.0739	0.0032	0.0063	0.0026
h_9	0.0799	0.0793	0.0033	0.0060	0.0027
h_{10}	0.0778	0.0771	0.0034	0.0060	0.0028
h_{11}	0.0725	0.0718	0.0036	0.0072	0.0030
h_{12}	0.0563	0.0557	0.0036	0.0062	0.0025
h_{13}	0.0545	0.0543	0.0032	0.0063	0.0026
h_{14}	0.0509	0.0505	0.0031	0.0060	0.0025
h_{15}	0.0570	0.0564	0.0056	0.0077	0.0045
h_{16}	0.0739	0.0733	0.0033	0.0059	0.0029
h_{17}	0.0889	0.0880	0.0036	0.0053	0.0032
h_{18}	0.0441	0.0438	0.0040	0.0082	0.0030
h_{19}	0.0919	0.0910	0.0038	0.0059	0.0035
h_{20}	0.0621	0.0615	0.0037	0.0064	0.0031
h_{21}	0.0601	0.0595	0.0035	0.0060	0.0032

We hence use the bootstrap methods in order to avoid a possible underestimation of the variances and spurious inference.

The estimated memory parameters d_1 and d_2 exert a marked difference in the integration orders of fractional components. The two series in the first group are the cause of significant nonstationarity in our dataset. The second group of nine series introduces stationary long-memory persistence. In contrast, the nonfractional components in z_t are only mildly autocorrelated, with small but significant autoregression

Table 4.11: Bootstrap t-ratios for fractional components loadings ($\Lambda^{(1)}$ and $\Lambda^{(2)}$) and nonfractional loadings (Γ) from the DOFC model (4.11) estimated in section 4.5.3.

	$\Lambda^{(1)}$		$\Lambda^{(2)}$									Γ	
$y_{1,t}$	17.4		11.9									3.6	
$y_{2,t}$	22.7	1.6	1.8	-11.9								0.4	-0.7
$y_{3,t}$	13.1	-1.6	3.4	-6.1	-15.6							1.1	1.3
$y_{4,t}$	11.9	-2.9	2.1	-3.3	-5.2	16.2						0.6	-1.6
$y_{5,t}$	12.5	-11.1	4.4	-10.7	-1.1	1.1	-3.4					1.6	-1.7
$y_{6,t}$	18.9	1.2	2.0	-6.0	-2.3	3.1	-3.1	6.8				1.0	1.7
$y_{7,t}$	4.1	6.0	5.6	-8.0	-1.1	2.0	-4.7	-1.5	4.9			-0.5	-2.7
$y_{8,t}$	3.5	3.2	5.4	-10.2	-1.2	2.4	-1.6	0.5	4.4	7.4		0.9	0.2
$y_{9,t}$	4.7	4.2	2.8	-7.4	-1.0	2.2	-5.8	-2.2	-0.1	3.4	-3.0	3.9	-0.5
$y_{10,t}$	2.7	4.9	2.5	-9.0	-2.9	2.0	-3.3	-0.9	2.5	3.6	4.2	6.1	-4.0
$y_{11,t}$	4.0	5.1	3.5	-9.9	0.6	2.7	-2.8	-1.7	7.1	-1.8	-2.5	3.2	1.7
$y_{12,t}$	2.9	3.8	9.7	-9.8	-1.0	2.5	-1.3	0.2	0.5	3.5	3.5	-2.5	3.0
$y_{13,t}$	3.9	4.3	5.5	-7.0	-0.9	2.2	-5.7	-2.7	-3.1	0.6	-1.0	0.2	2.7
$y_{14,t}$	2.0	4.4	5.1	-9.8	-2.6	2.6	-2.3	-0.8	-1.8	-0.8	6.1	2.6	-3.4
$y_{15,t}$	3.4	6.0	6.2	-12.5	0.5	2.5	-1.6	0.6	1.4	-7.2	-0.4	-0.3	3.5
$y_{16,t}$	4.4	3.1	5.6	-10.6	-1.1	2.3	-1.7	-0.4	-2.9	5.4	-2.2	1.2	2.2
$y_{17,t}$	3.2	4.4	5.7	-12.5	-1.5	2.8	0.5	1.7	-0.8	3.9	4.2	3.8	-0.9
$y_{18,t}$	2.5	3.1	6.6	-12.6	0.0	2.8	0.4	0.3	2.3	2.2	1.2	1.5	10.2
$y_{19,t}$	3.7	5.2	3.6	-8.9	-1.9	1.7	-2.6	-0.9	-5.1	2.1	-0.7	6.2	-1.9
$y_{20,t}$	3.7	4.0	3.4	-8.1	0.1	2.6	-4.7	-3.2	-2.4	0.2	-2.2	3.9	6.4
$y_{21,t}$	1.6	4.0	3.1	-9.9	-1.4	3.0	-1.2	-0.5	-0.3	-2.1	4.5	8.0	1.4

parameters. Figure 4.11 gives a visual impression of the factor dynamics, showing full sample (smoothed) estimates of the two nonstationary components (above), of the first two stationary long-memory components (middle) and of the short-memory components (below). The ± 2 standard error bands suggest a relatively precise estimation of the components. The different persistence of the three groups is clearly visible.

We turn to a discussion of the cointegration properties of the estimated system. In our preferred specification with a cointegration rank of $p - s_1 = 19$, and an 11-dimensional cointegration subspace, the loadings of fractional components provide an easier interpretation than the corresponding cointegration vectors, while in other cases the latter can be easily obtained and suitably normalized.

With the abovementioned caveat that asymptotic results are not available for this

fractional cointegration setting, we show t-ratios for constants, for fractional loadings and for nonfractional loadings in table 4.11, where the bootstrap standard errors are used. The t-ratios for $\Lambda^{(1)}$ suggest that each of the series in y_t is influenced by the nonstationary components, and hence all components of y_t are nonstationary themselves. The first component loads very significantly on all variances with the same sign and can hence be interpreted as the main common risk factor. The second component represent joint common nonstationarity of the correlations, which is negatively associated with the IBM return variances. Except those corresponding to the first, the second and the forth stationary components with their equal signs, the columns of $\Lambda^{(2)}$ have a rather mixed pattern. Like the nonstationary factors, also the $I(d^{(2)})$ components affect variance and correlation dynamics at the same time and therefore induce fractional cointegration between log variances and z-transformed correlations. Having discussed model specification and estimation results using the full sample of realized covariance data, we turn to an evaluation of the forecasting precision in the next section.

4.5.4 An Out-of-sample Comparison

We assess the forecasting performance of our model by means of an out-of-sample comparison. To avoid reference of the forecasts on the out-of-sample periods, we conduct a semiparametric specification search along the lines of section 4.5.2 for the first estimation sample only, i.e. for y_t, $t = 1,\dots,1508$, while $t = 1509,\dots,2156$ is reserved for prediction and therefore not used for selecting the specification. In this way, the model for the forecasting comparison includes $s_1 = 2$, $s_2 = 7$ and $s_0 = 3$ components of different integration orders. Rather than conducting comprehensive comparisons of a wide range of available methods which is beyond the scope of this paper, we select straightforward and simple benchmark models which have performed excellently in previous studies.

We choose the same out-of-sample setup as in Weigand (2014). Thus, for each $T' \in [1508; 2156-h]$, different models are estimated for a rolling sample with $n = 1508$ observations, $y_{T'-1507},\dots,y_{T'}$. From these estimates, forecasts of $y_{T'+h}$, $h = 1,5,10,20$, are computed. Also in line with Weigand (2014), we compute bias-corrected forecasts of the realized covariance matrices $\hat{X}_{T'+h|T'}$ by the simulation-based technique discussed there. We evaluate the forecasting accuracy using the ex-post available data of the respective period.

The forecasting precision is assessed using different loss functions defined in appendix 4.C. We consider the Frobenius norm $LF_{T',h}$ (4.37), the Stein norm $LS_{T',h}$ (4.38) and the asymmetric loss $L3_{T',h}$ (4.39); see Laurent, Rombouts, and Violante (2011) and

Laurent, Rombouts, and Violante (2013). Additionally, the ex-ante minimum variance portfolio is computed from the forecast and its realized variance $LMV_{T',h}$ (5.32) used as a loss with obvious economic relevance. Furthermore, we assess density forecasts f_r of the daily returns using covariance matrices, which are evaluated at the daily returns $r_{T'+h}$ in a logarithmic scoring rule $LD_{T',h}$ (5.25).

As benchmarks, we consider two linear models for the log variance and z-transformed correlation series y_t, namely a diagonal vector ARMA(2,1) and a diagonal vector ARFIMA(1,d,1) model, which have been found to perform well by Weigand (2014). Additionally, the diagonal vector ARFIMA(1,d,1) model is applied to the Cholesky factors of the covariance matrices (Chiriac and Voev, 2011). Furthermore, we consider models with a conditional Wishart distribution, namely the conditional autoregressive Wishart (CAW) model of Golosnoy, Gribisch, and Liesenfeld (2012), a dynamic correlation specification (CAW-DCC) of Bauwens, Storti, and Violante (2012), and additive and multiplicative components Wishart models as proposed by Jin and Maheu (2013). For further details on the comparison models consult appendix 4.C.

For each loss function and horizon h, we compute the average losses (risks) for all models and obtain model confidence sets of Hansen, Lunde, and Nason (2011), bootstrapping the max-t statistic with a block lengths of $\max\{5,h\}$. In tables 4.12, 4.13, 4.14 and 4.15, we present the risks for $h = 1,5,10,20$. The best performing model (***) as well as members of the 80% model confidence set (**) and models contained in the 90% but not in the 80% set (*) are indicated.

The fractional components model is among the best competitors for all horizons and loss functions. It has lowest risks for almost all setups. Exceptions occur for $h \geq 10$ where the ARFIMA model for log variances and z-correlations performs best in some cases. Overall, the ARFIMA model on y_t appears as a second best in terms of forecasting precision.

The DOFC model is always contained in the 80% model confidence set whereas all other models are rejected at least in some cases. For the Stein loss and the minimum-variance loss, the DOFC model is significantly superior than most competitors for small horizons, while with the Frobenius and asymmetric loss, rejections of other models are achieved for $h = 10$ and $h = 20$.

The performance of the fractional components model in terms of density forecasting is noteworthy. In each case there, our model is either the single member or one of two models in the confidence set and hence significantly outperforms most of the competitors. Since daily returns are often more important than the realized measures themselves, this finding is particularly strong.

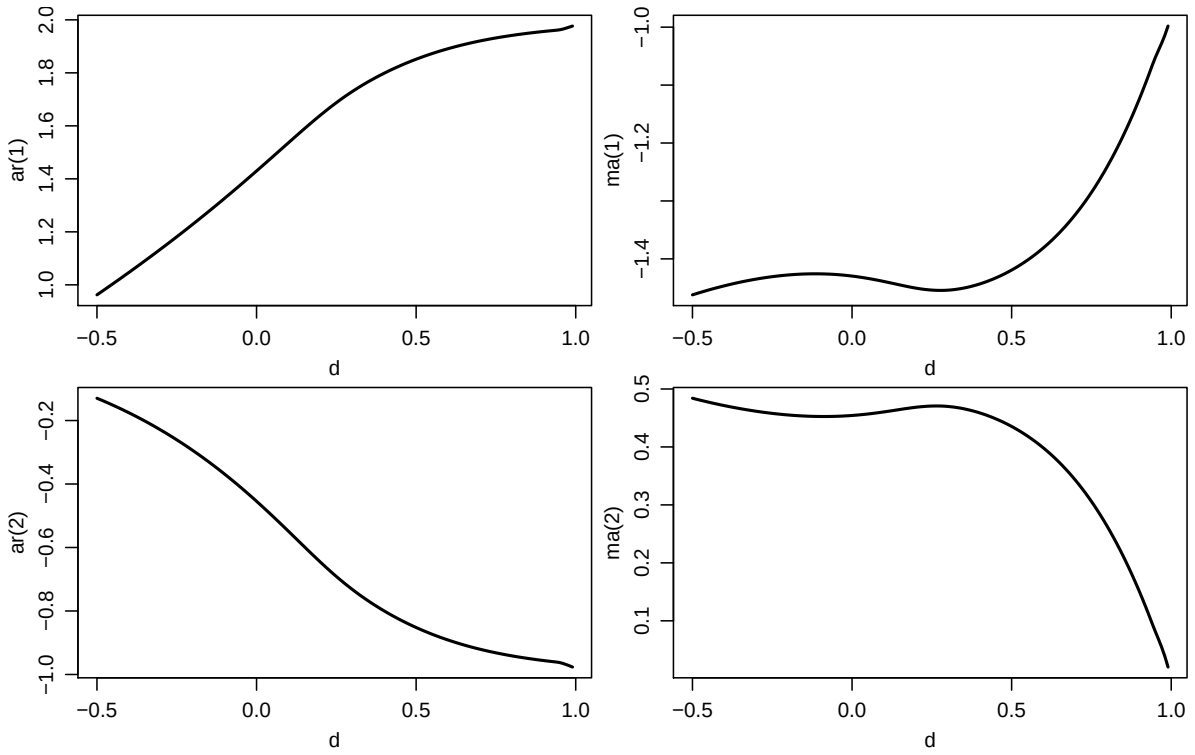

Figure 4.1: ARMA(2,2) coefficients (4.14) in the approximation of fractional processes for $d \in [-0.5; 1]$ and $n = 500$.

Table 4.12: Out-of-sample risks for $h = 1$ as described in section 4.5.4. In different rows, we consider the fractional components (FC) and several diagonal vector ARFIMA, ARMA and CAW models. Asterisks denote the best performing model (***), models in the 80% model confidence set (**) and additional models in the 90% model confidence set (*). As loss functions, we consider the Frobenius norm (LF), the Stein norm (LS), the predictive densities (LD), the minimum-variance portfolio variance (LMV) and the L3-Loss (L3).

$h = 1$	LF	LS	L3	LMV	LD
FC	84.28^{***}	0.9660^{***}	1807^{***}	0.7905^{***}	8.1319^{***}
ARMA	85.09^{**}	0.9950	1823^{**}	0.7916	8.1533
ARFIMA	86.22^{**}	0.9955	1829^{**}	0.7911^{**}	8.1586
ARFIMA.chol	87.82^{**}	1.0830	1860^{**}	0.7920^{*}	8.1723^{**}
CAW.diag	85.97^{**}	1.0254	1843^{**}	0.7930	8.1869
CAW.dcc	86.32^{**}	1.0037	1866^{**}	0.7928	8.3021
CAW.acomp	85.77^{**}	1.0268	1814^{**}	0.7932	8.2482
CAW.mcomp	90.72^{**}	1.0301	1904^{**}	0.7929	8.2478

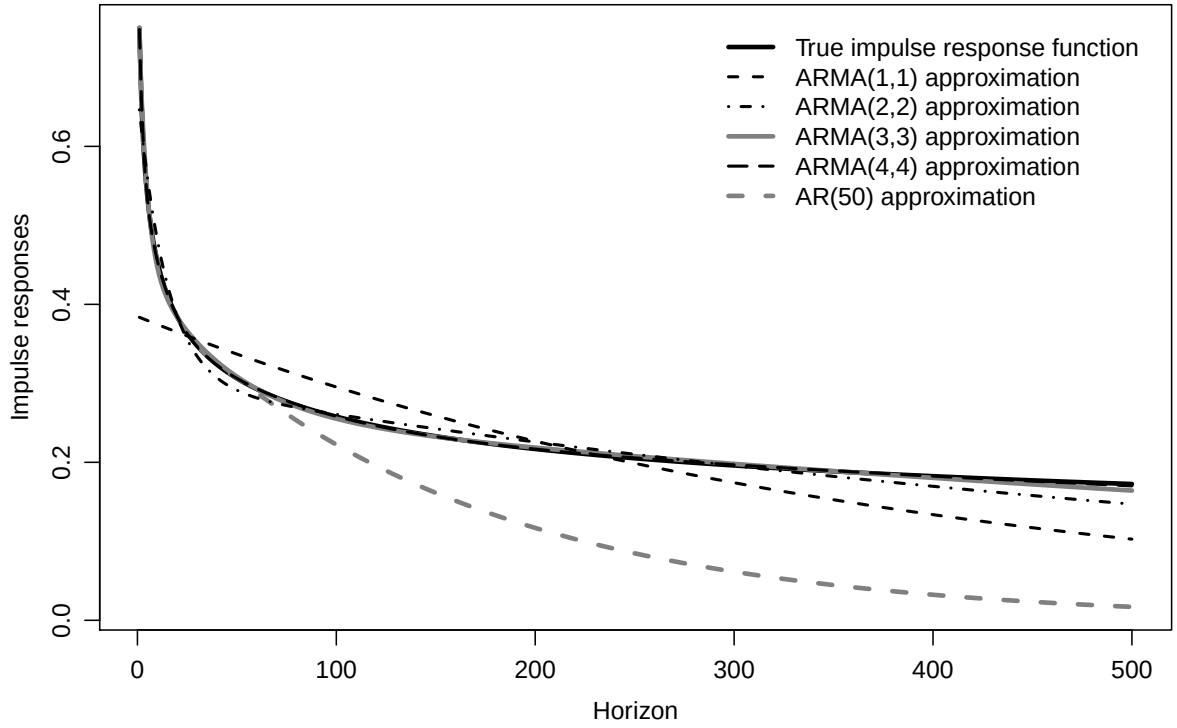

Figure 4.2: Impulse response functions $\tilde{\psi}_j$ (see (4.13)) for different approximating models for $d = 0.75$ and $n = 500$.

Table 4.13: Out-of-sample risks for $h = 5$ as described in section 4.5.4. In different rows, we consider the fractional components (FC) and several diagonal vector ARFIMA, ARMA and CAW models. Asterisks denote the best performing model (***), models in the 80% model confidence set (**) and additional models in the 90% model confidence set (*). As loss functions, we consider the Frobenius norm (LF), the Stein norm (LS), the predictive densities (LD), the minimum-variance portfolio variance (LMV) and the L3-Loss (L3).

$h = 5$	LF	LS	L3	LMV	LD
FC	135.28^{***}	1.3766^{***}	2463^{***}	0.8011^{***}	8.2490^{***}
ARMA	134.43^{**}	1.4046	2498^{**}	0.8025	8.2688
ARFIMA	135.12^{**}	1.3974	2492^{**}	0.8015^{**}	8.2664
ARFIMA.chol	140.43^{**}	1.5348	2557^{**}	0.8021^{*}	8.3113^{**}
CAW.diag	137.34^{**}	1.4356	2612^{**}	0.8038	8.3184
CAW.dcc	137.77^{**}	1.4094	2627^{**}	0.8039	8.4150
CAW.acomp	139.22^{**}	1.4443	2558^{**}	0.8028	8.3311
CAW.mcomp	142.26^{**}	1.4489	2590^{**}	0.8028	8.3399

Table 4.14: Out-of-sample risks for $h = 10$ as described in section 4.5.4. In different rows, we consider the fractional components (FC) and several diagonal vector ARFIMA, ARMA and CAW models. Asterisks denote the best performing model (***), models in the 80% model confidence set (**) and additional models in the 90% model confidence set (*). As loss functions, we consider the Frobenius norm (LF), the Stein norm (LS), the predictive densities (LD), the minimum-variance portfolio variance (LMV) and the L3-Loss (L3).

$h = 10$	LF	LS	L3	LMV	LD
FC	170.07^{**}	1.7033^{**}	2837^{**}	0.8102^{**}	8.3118^{***}
ARMA	172.03^{**}	1.6985^{**}	2890^{**}	0.8088^{*}	8.3519
ARFIMA	168.55^{***}	1.6716^{***}	2837^{***}	0.8076^{***}	8.3372
ARFIMA.chol	173.43^{**}	1.8455	2900^{**}	0.8103^{**}	8.3893
CAW.diag	176.50^{**}	1.7399^{**}	2980^{**}	0.8110^{**}	8.4105
CAW.dcc	178.11^{**}	1.7120^{**}	2986^{**}	0.8101^{**}	8.4973
CAW.acomp	177.37^{**}	1.7366^{**}	2947^{**}	0.8093^{**}	8.4146
CAW.mcomp	181.04^{**}	1.7265^{**}	3009^{**}	0.8096^{**}	8.4248

Table 4.15: Out-of-sample risks for $h = 20$ as described in section 4.5.4. In different rows, we consider the fractional components (FC) and several diagonal vector ARFIMA, ARMA and CAW models. Asterisks denote the best performing model (***), models in the 80% model confidence set (**) and additional models in the 90% model confidence set (*). As loss functions, we consider the Frobenius norm (LF), the Stein norm (LS), the predictive densities (LD), the minimum-variance portfolio variance (LMV) and the L3-Loss (L3).

$h = 20$	LF	LS	L3	LMV	LD
FC	199.42^{***}	2.0461^{**}	3144^{***}	0.8225^{**}	8.3778^{***}
ARMA	208.07^{**}	2.0980^{**}	3231	0.8224^{**}	8.4314
ARFIMA	200.33^{**}	2.0305^{***}	3162^{**}	0.8209^{***}	8.4049
ARFIMA.chol	203.22^{**}	2.1910^{**}	3201^{**}	0.8219^{**}	8.4738
CAW.diag	214.60^{*}	2.1580^{**}	3326^{*}	0.8241^{**}	8.5034
CAW.dcc	217.52	2.1698^{**}	3331	0.8231^{**}	8.6158
CAW.acomp	211.33	2.1165^{**}	3289^{*}	0.8214^{**}	8.5028
CAW.mcomp	209.78^{**}	2.0858^{**}	3282^{**}	0.8225^{**}	8.5158

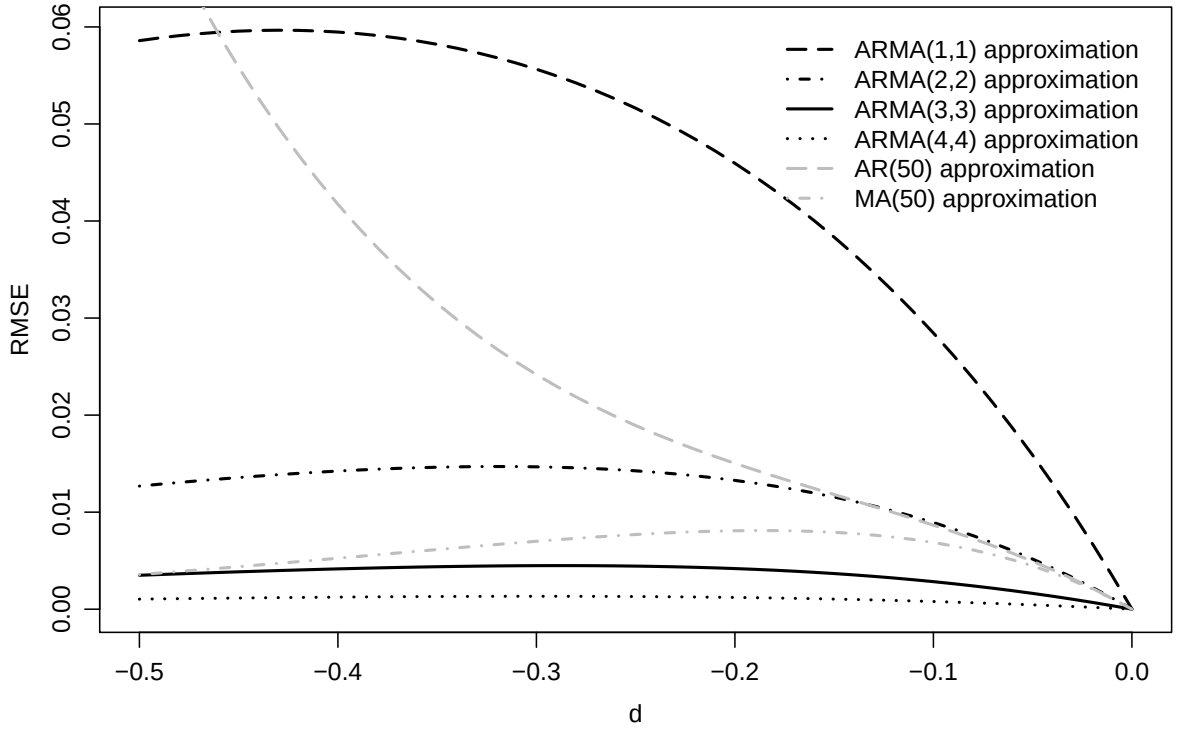

Figure 4.3: Root mean squared error (square root of (4.13)) for different approximating models, $d \in [-0.5; 0]$ and $n = 500$.

Overall, we find a very good forecast performance of the model proposed in this paper. Although for some criteria and horizons statistical significance is lacking, the model yields very precise forecasts in relation to different competitors for all considered horizons and for several ways to measure this precision.

4.6 Conclusion

We have suggested a general setup and a parsimonious model with very general fractional integration and cointegration properties. The model can be estimated by maximum likelihood using an approximate state space form which is computationally convenient and has good estimation properties which we found in simulations. We discussed the usefulness of our approach for multivariate realized volatility modeling. In our application it was shown to provide a reasonable in-sample fit and excellent out-of-sample forecasting accuracy.

Several questions remain for further research. An asymptotic theory of our maximum likelihood approach is missing. For other existing parametric fractional cointegration setups, asymptotic distributions have been derived and tests for the cointegration rank

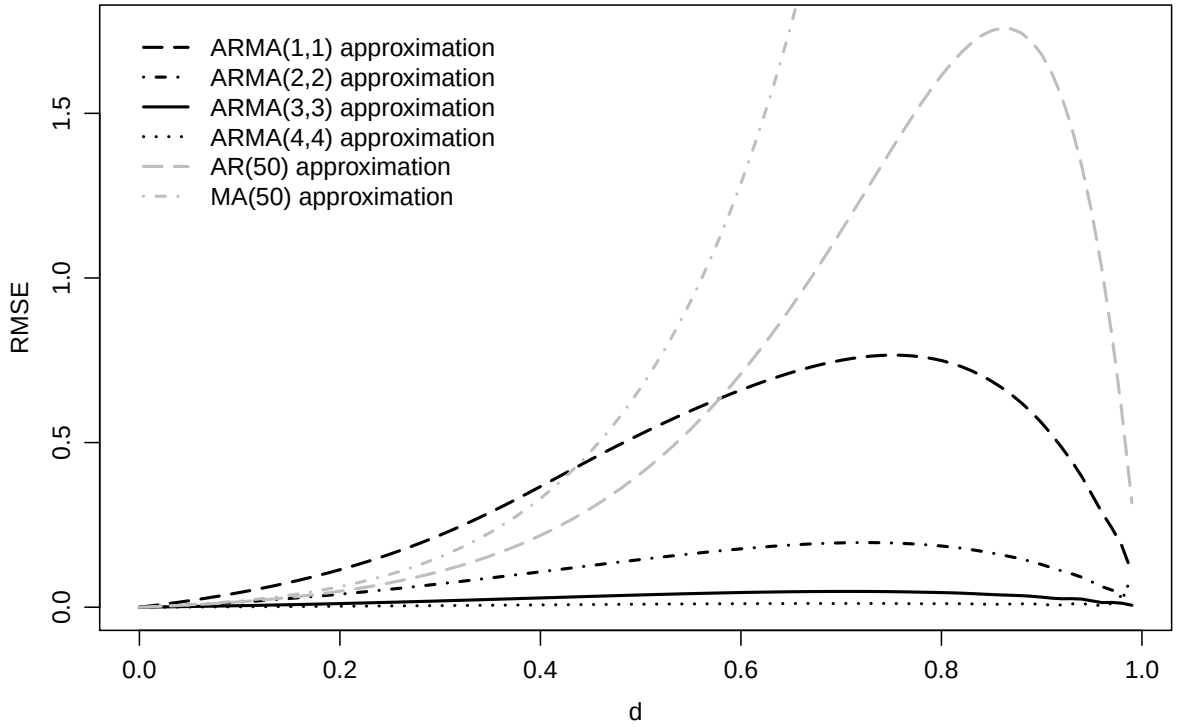

Figure 4.4: Root mean squared error (square root of (4.13)) for different approximating models, $d \in [0;1]$ and $n = 500$.

have been developed. It remains to be shown to what extent existing results carry over to the state space model proposed in this paper.

From an empirical point of view, we have shown the relevance of a very restricted specification in financial econometrics, but the general setup we introduced has a broader scope. Fractional components models with rich short-run dynamics may be considered for models of smaller dimension. In several empirical setups, fractional integration and cointegration has been found relevant but very rarely dynamic modeling, forecasting, identification of structural shocks and impulse response analyses have been conducted in an according framework.

Appendix 4.A Details on Alternative Representations

In this appendix we provide more details on the derivation of the alternative representations of the fractional components model (4.1) which we discuss in section 4.2.2.

To derive the error correction representation (4.6), we start from the FC setup with $q = 2$ and $s = p$,

$$y_t = \Lambda^{(1)} x_t^{(1)} + \Lambda^{(2)} x_t^{(2)} + u_t,$$

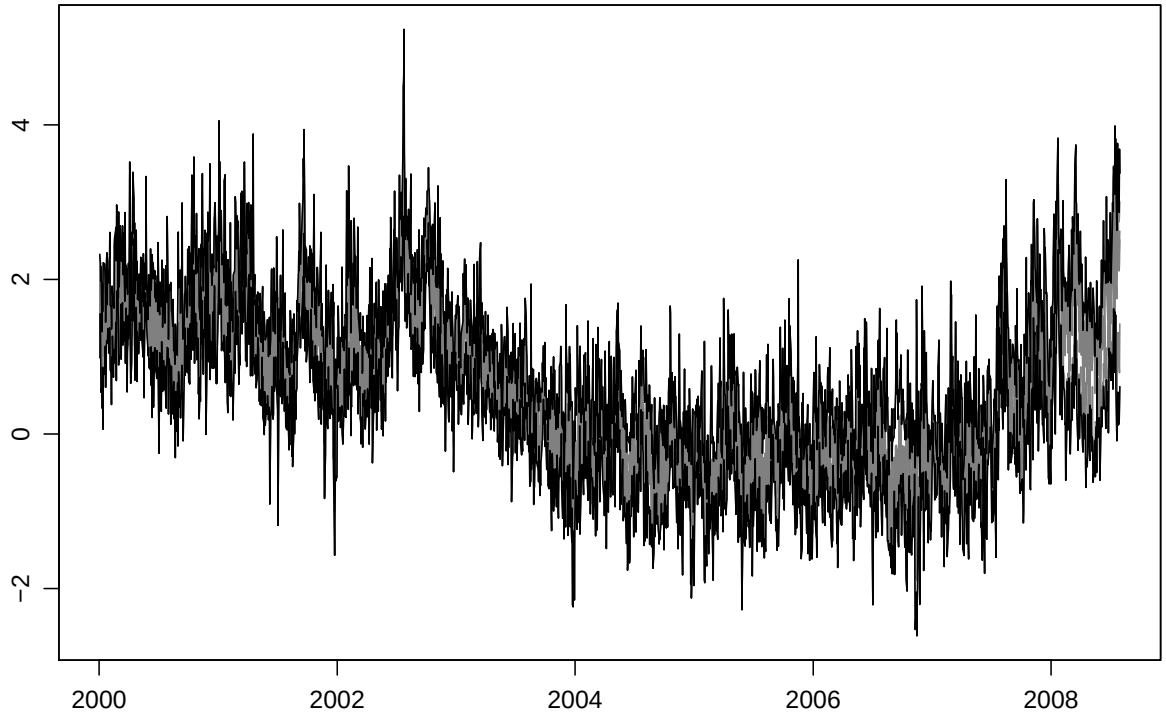

Figure 4.5: Time series plots of log realized variances for the dataset described in section 4.5 (grey) together with maximum and minimum for all periods (black).

from which we note that

$$\Lambda_\perp^{(1)'}\Delta^{d^{(2)}}y_t = \Lambda_\perp^{(1)'}\Lambda^{(2)}\xi_t^{(2)} + \Lambda_\perp^{(1)'}u_t \quad \text{and} \quad \Lambda_\perp^{(2)'}\Delta^{d^{(1)}}y_t = \Lambda_\perp^{(2)'}\Lambda^{(1)}\xi_t^{(1)} + \Lambda_\perp^{(2)'}u_t.$$

We define

$$N := \Lambda^{(2)}(\Lambda_\perp^{(1)'}\Lambda^{(2)})^{-1}\Lambda_\perp^{(1)'} \quad \text{and} \quad M := \Lambda^{(1)}(\Lambda_\perp^{(2)'}\Lambda^{(1)})^{-1}\Lambda_\perp^{(2)'}, \tag{4.29}$$

and make use of $I = N + M$ (Johansen, 2008), to obtain

$$\Delta^{d^{(1)}}y_t = M(\Lambda^{(1)}\xi_t^{(1)} + \Delta^{d^{(1)}}u_t) + \Delta^{d^{(1)}-d^{(2)}}\Delta^{d^{(2)}}Ny_t. \tag{4.30}$$

Adding and subtracting $\Delta^{d^{(2)}}Ny_t$ on the right side of (4.30) and the decomposition $N = -\alpha\beta'$ yields (4.6).

Next, we consider the triangular representation; see (4.8) and (4.9). The first block, (4.8), is easily obtained. Since $\Lambda^{(1,1)}$ is nonsingular and we also assumed a nonsingular covariance matrix of the white noise sequence ξ_t, we find that the first term on the right is $I(0)$ with positive definite spectral density while the other terms have integration orders lower than zero, leading to $\omega_t^{(1)} \sim I(0)$. To arrive at the j-th block of the system,

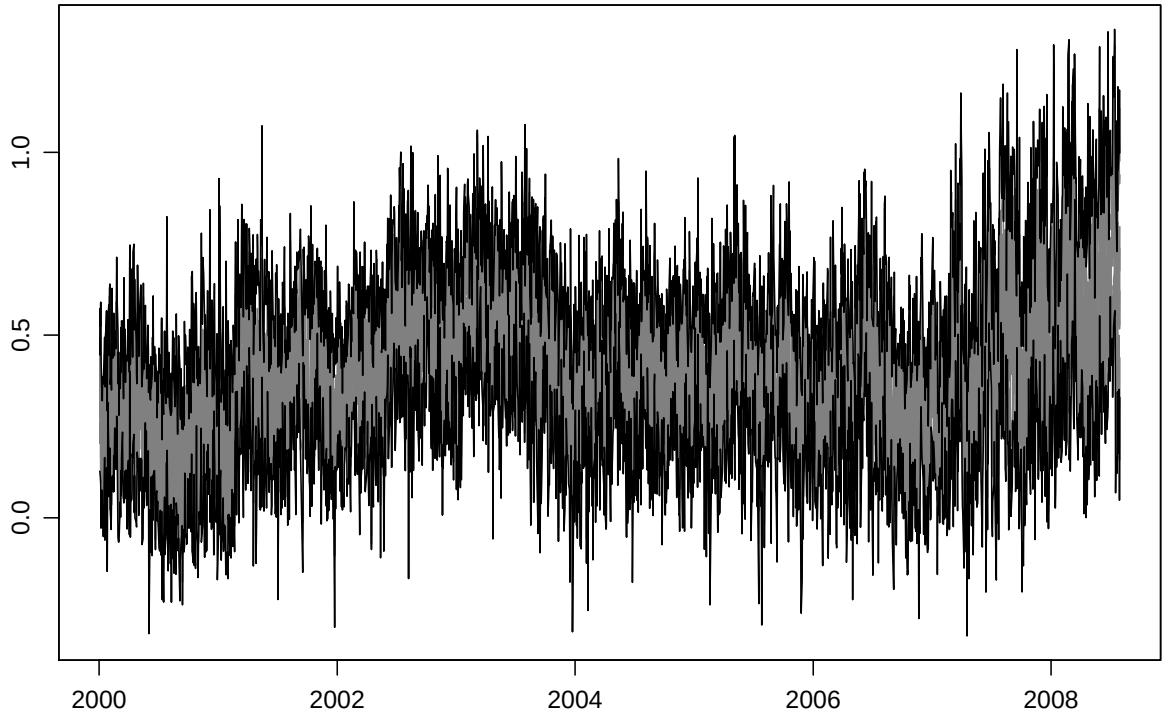

Figure 4.6: Time series plots of z-transformed realized correlations for the dataset described in section 4.5 (grey) together with maximum and minimum for all periods (black).

consider the expression for $y_t^{(j)}$,

$$\Delta^{d^{(j)}} y_t^{(j)} = \Lambda^{(j,1)} \Delta^{d^{(j)}} x_t^{(1)} + \ldots + \Lambda^{(j,q)} \Delta^{d^{(j)}} x_t^{(q)} + \Delta^{d^{(j)}} u_t^{(j)}.$$

Since $\Delta^{d^{(j)}} x_t^{(i)}$ is integrated of order zero or lower for $i \geq j$, we can write

$$\begin{aligned} \Delta^{d^{(j)}} y_t^{(j)} &= \Lambda^{(j,1)} \Delta^{d^{(j)}} x_t^{(1)} + \ldots + \Lambda^{(j,j-1)} \Delta^{d^{(j)}} x_t^{(j-1)} + \tilde{\omega}_t^j \\ &= \Lambda^{(j,1:(j-1))} \Delta^{d^{(j)}} x_t^{(1:(j-1))} + \tilde{\omega}_t^j, \end{aligned} \tag{4.31}$$

where $\tilde{\omega}_t^j \sim I(0)$. To substitute for the latent variables in this expression, consider

$$\Delta^{d^{(j)}} y_t^{(1:(j-1))} = \Lambda^{(1:(j-1),1:(j-1))} \Delta^{d^{(j)}} x_t^{(1:(j-1))} + \breve{\omega}_t^j,$$

with $\breve{\omega}_t^j \sim I(0)$ which we can solve for

$$\Delta^{d^{(j)}} x_t^{(1:(j-1))} = (\Lambda^{(1:(j-1),1:(j-1))})^{-1} \Delta^{d^{(j)}} y_t^{(1:(j-1))} - (\Lambda^{(1:(j-1),1:(j-1))})^{-1} \breve{\omega}_t^j.$$

Substituting this expression into (4.31) yields the general expression (4.9) for the j-th

block of the triangular system for $j = 2, \ldots, q$, where

$$\omega_t^{(j)} = \tilde{\omega}_t^j - \Lambda^{(j,1:(j-1))}(\Lambda^{(1:(j-1),1:(j-1))})^{-1}\breve{\omega}_t^j,$$

which can be stated in greater detail as

$$\begin{aligned}\omega_t^{(j)} = &-\Lambda^{(j,1:(j-1))}(\Lambda^{(1:(j-1),1:(j-1))})^{-1}\Lambda^{(1:(j-1),j:q)}\Delta^{d^{(j)}}x_t^{(j:q)} \\ &- \Lambda^{(j,1:(j-1))}(\Lambda^{(1:(j-1),1:(j-1))})^{-1}\Delta^{d^{(j)}}u_t^{(1:(j-1))} + \Lambda^{(j,j:q)}\Delta^{d^{(j)}}x_t^{(j:q)} + \Delta^{d^{(j)}}u_t^{(j)}. \quad (4.32)\end{aligned}$$

This process is the sum of several additive negatively integrated plus a white noise process

$$\left[\Lambda^{(j,j)} - \Lambda^{(j,1:(j-1))}(\Lambda^{(1:(j-1),1:(j-1))})^{-1}\Lambda^{(1:(j-1),j)}\right]\xi_t^{(j)},$$

so that we conclude that $\omega_t^{(j)}$ is $I(0)$ with positive definite spectral density at zero frequency.

We arrive at the representation (4.10) where B is partitioned into blocks according to

$$B = \begin{pmatrix} I & 0 & \ldots & 0 \\ B^{(1,1)} & I & & 0 \\ \vdots & \ddots & \ddots & \vdots \\ B^{(q,1)} & \ldots & B^{(q,q-1)} & I \end{pmatrix}.$$

In case $p > s$, we have

$$\begin{aligned}y_t^{(q+1)} &= \Lambda^{(q+1,1:q)}(\Lambda^{(1:q,1:q)})^{-1}y_t^{(1:q)} + u_t^{(q+1)} - \Lambda^{(q+1,1:q)}(\Lambda^{(1:j,1:j)})^{-1}u_t^{(1:q)} \\ &= B^{(q+1,1)}y_t^{(1)} + \ldots + B^{(q+1,q)}y_t^{(j-1)} + \omega_t^{(q+1)},\end{aligned}$$

and the representation (4.10) is changed to

$$By_t = (\Delta_+^{-d_1}\omega_t^{(1)}, \ldots, \Delta_+^{-d_q}\omega_t^{(q)}, \omega_t^{(q+1)})'$$

where B is extended by the $p - s$ rows $(B^{(q+1,1)}, \ldots, B^{(q+1,q)}, I)$.

Appendix 4.B Details on the EM Algorithm

In this appendix, all necessary expressions for the computation of the EM algorithm will be given. The log-likelihood where the unobserved state process α_t is assumed known is called the complete data log likelihood and given by

$$\begin{aligned} l(\theta;\{y_t,\alpha_t\}_{t=1}^n) = &-\frac{1}{n}\log|Q| - \frac{1}{2}\operatorname{tr}\left[RQ^{-1}R'\sum_{t=2}^n(\alpha_t - T\alpha_{t-1})(\alpha_t - T\alpha_{t-1})'\right] \\ &-\frac{n}{2}\log|H| - \frac{1}{2}\operatorname{tr}\left[H^{-1}\sum_{t=1}^n(y_t - Z\alpha_t)(y_t - Z\alpha_t)'\right]. \end{aligned}$$

The expectation of the complete data likelihood, with expectation evaluated at parameters $\theta_{\{j\}}$, is denoted by $Q(\theta,\theta_{\{j\}})$ and given by (4.16). The terms involving expectations of the (partially unobserved) data and its cross-moments are

$$\begin{aligned} A_{\{j\}} &:= \mathrm{E}_{\theta_{\{j\}}}\left[\sum_{t=2}^n \alpha_t\alpha_t'\right] = \sum_{t=2}^n \hat\alpha_t\hat\alpha_t' + \sum_{t=2}^n V_{t,t} \\ B_{\{j\}} &:= \mathrm{E}_{\theta_{\{j\}}}\left[\sum_{t=2}^n \alpha_t\alpha_{t-1}'\right] = \sum_{t=2}^n \hat\alpha_t\hat\alpha_{t-1}' + \sum_{t=1}^n V_{t,t-1} \\ C_{\{j\}} &:= \mathrm{E}_{\theta_{\{j\}}}\left[\sum_{t=2}^n \alpha_{t-1}\alpha_{t-1}'\right] = \sum_{t=2}^n \hat\alpha_{t-1}\hat\alpha_{t-1}' + \sum_{t=2}^n V_{t-1,t-1} \\ D_{\{j\}} &:= \mathrm{E}_{\theta_{\{j\}}}\left[\sum_{t=1}^n y_t y_t'\right] = \sum_{t=1}^n y_t y_t', \quad E_{\{j\}} := \mathrm{E}_{\theta_{\{j\}}}\left[\sum_{t=1}^n y_t\alpha_t'\right] = \sum_{t=1}^n y_t\hat\alpha_t', \\ F_{\{j\}} &:= \mathrm{E}_{\theta_{\{j\}}}\left[\sum_{t=1}^n \alpha_t\alpha_t'\right] = \sum_{t=1}^n \hat\alpha_t\hat\alpha_t' + \sum_{t=1}^n V_{t,t}. \end{aligned}$$

Here, $\hat\alpha_t = \mathrm{E}_{\theta_{\{j\}}}[\alpha_t]$ and $V_{t,s} = \mathrm{E}_{\theta_{\{j\}}}[(\alpha_t - \hat\alpha_t)(\alpha_s - \hat\alpha_s)']$ can be computed by state smoothing algorithms based on the state space representation for given $\theta_{\{j\}}$ (Durbin and Koopman, 2012, section 4.4).

We turn to the derivation of (4.17). For notational convenience we denote the objective function for optimization over $\theta^{(1)}$ by $Q^1_{\{j\}}(\theta^{(1)}) \equiv Q((\theta^{(1)'},\theta^{(2)'}_{\{j\}})';\theta_{\{j\}})$. To describe the Newton step in the optimization of $Q^1_{\{j\}}$ in detail, we explicitly state the nonlinear dependence of $\operatorname{vec}(T,Z)' = (\operatorname{vec}(T)',\operatorname{vec}(Z)')$ on $\theta^{(1)}$ by $\operatorname{vec}(T,Z) = f(\theta^{(1)})$ and consider the linearization at $\theta_{\{j\}}$,

$$\operatorname{vec}(T,Z) \approx \Xi_{\{j\}}\theta^{(1)} + \xi_{\{j\}}, \quad \text{where} \quad \Xi \equiv \frac{\partial f(\theta^{(1)})}{\partial\theta^{(1)'}}, \tag{4.33}$$

$\xi \equiv f(\theta^{(1)}) - \Xi\theta^{(1)}$, and the $\{j\}$ subscript indicates evaluation of a specific expression at

$\theta_{\{j\}}$. The optimization over $\theta^{(1)}$ jointly involves elements in T and Z, since d enters the expression of both system matrices and hence, Ξ is not diagonal.

A single iteration of the Newton optimization algorithm is carried out by expanding the gradient around $\theta^{(1)}_{\{j\}}$. The gradient is given by

$$\frac{\partial Q^1_{\{j\}}(\theta^{(1)})}{\partial \theta^{(1)}} = \frac{\partial(\mathrm{vec}(T)', \mathrm{vec}(Z)')}{\partial \theta^{(1)}} \frac{\partial Q^1_{\{j\}}}{\partial(\mathrm{vec}(T)', \mathrm{vec}(Z)')'} = \Xi' \,\mathrm{vec}\begin{pmatrix} \frac{\partial Q^1_{\{j\}}}{\partial T} & \frac{\partial Q^1_{\{j\}}}{\partial Z} \end{pmatrix}. \tag{4.34}$$

For the derivatives with respect to the system matrices we have

$$\frac{\partial Q^1_{\{j\}}}{\partial T} = (RQ^{-1}R')(B_{\{j\}} - TC_{\{j\}}) \qquad \text{and} \qquad \frac{\partial Q^1_{\{j\}}}{\partial Z} = H^{-1}(E_{\{j\}} - ZF_{\{j\}}),$$

so that

$$\mathrm{vec}\begin{pmatrix} \frac{\partial Q^1_{\{j\}}}{\partial T} & \frac{\partial Q^1_{\{j\}}}{\partial Z} \end{pmatrix} = \begin{pmatrix} \mathrm{vec}(RQ^{-1}R'B_{\{j\}}) \\ \mathrm{vec}(H^{-1}E_{\{j\}}) \end{pmatrix} - \begin{pmatrix} C'_{\{j\}} \otimes RQ^{-1}R' & 0 \\ 0 & F'_{\{j\}} \otimes H^{-1} \end{pmatrix} \begin{pmatrix} \mathrm{vec}(T) \\ \mathrm{vec}(Z). \end{pmatrix}$$

Hence, for $G_{\{j\}}$ and $g_{\{j\}}$ given by

$$g_{\{j\}} = \mathrm{vec}(RQ^{-1}R'B_{\{j\}}, H^{-1}E_{\{j\}}), \quad \text{and} \quad G_{\{j\}} = \mathrm{diag}(C'_{\{j\}} \otimes RQ^{-1}R', F'_{\{j\}} \otimes H^{-1}_{\{j\}}),$$

we obtain the linear expansion

$$\frac{\partial Q^1_{\{j\}}(\theta^{(1)})}{\partial \theta^{(1)}} \approx \Xi'_{\{j\}} g_{\{j\}} - \Xi'_{\{j\}} G_{\{j\}} (\Xi_{\{j\}} \theta^{(1)} + \xi_{\{j\}}).$$

Equating to zero and solving for $\theta^{(1)}$ yields (4.17). For the estimation of H, see (4.36), we define

$$L_{\{j\}} := \mathrm{E}_{\theta_{\{j\}}} \left[\sum_{t=1}^{n} \varepsilon_t \varepsilon_t' \right] = D_{\{j\}} - ZE'_{\{j\}} - E_{\{j\}} Z' + ZF_{\{j\}} Z'. \tag{4.35}$$

and use

$$\frac{\partial Q(\theta, \theta_{\{j\}})}{\partial H} = (H^{-1}L_{\{j\}} - nI)H^{-1} - 0.5\,\mathrm{diag}((H^{-1}L_{\{j\}} - nI)H^{-1}), \tag{4.36}$$

to derive the estimator of the variance parameters; see Jungbacker and Koopman (2014, appendix A.3).

Appendix 4.C Details on the Out-of-sample Comparison

In this section we give further details on the out-of-sample evaluation of section 4.5.4. We state the loss functions to evaluate the forecasts as well as the specifications of the benchmark models and their estimation.

For given forecasted realized covariance matrices $X_{T'+h|T'}$ and realizations $X_{T'+h}$, the loss functions considered in this paper are the Frobenius norm ($LF_{T',h}$), the Stein distance ($LS_{T',h}$), the asymmetric loss ($L3_{T',h}$), the realized variance of the ex-ante minimum variance portfolio ($LMV_{T',h}$), and the negative log-score of density forecasts f_r ($LD_{T',h}$), given by

$$LF_{T',h} = \sum_{i=1}^{k}\sum_{j=1}^{k}(X_{ij,T'+h} - X_{ij,T'+h|T'})^2, \tag{4.37}$$

$$LS_{T',h} = \operatorname{tr}\left[X_{T'+h|T'}^{-1}X_{T'+h}\right] - \log\left|X_{T'+h|T'}^{-1}X_{T'+h}\right| - k, \tag{4.38}$$

$$L3_{T',h} = \frac{1}{6}\operatorname{tr}\left[X_{T'+h|T'}^{3} - X_{T'+h}^{3}\right] - \frac{1}{2}\operatorname{tr}\left[X_{T'+h|T'}^{2}(X_{T'+h} - X_{T'+h|T'})\right], \tag{4.39}$$

$$LMV_{T',h} = w'X_{T'+h}w, \quad w = (\iota'X_{T'+h|T'}\iota)^{-1}X_{T'+h|T'}\iota, \quad \iota = (1,\dots,1)', \tag{4.40}$$

$$LD_{T',h} = -\log f_r(r_{T'+h}). \tag{4.41}$$

As comparison models we consider three linear models in transformed covariance matrices, namely the diagonal vector ARMA(2,1) model

$$(1 - \phi_{i1}L - \phi_{i2}L^2)(y_{it} - c_i) = (1 + \theta_{i1}L)v_{it}, \quad i = 1,\dots,21,$$

for the log variance and z-correlation series y_t, a diagonal vector ARFIMA(1,d,1) model

$$(1 - \phi_{i1}L)(1 - L)^{d_i}(y_{it} - c_i) = (1 + \theta_{i1}L)v_{it}, \quad i = 1,\dots,21, \tag{4.42}$$

for y_t and the same model (4.42) applied to Cholesky factors. The same model orders have been used by Chiriac and Voev (2011) and Weigand (2014) and were found to compete favorably with other choices. The dynamic parameters of these models are estimated by Gaussian quasi maximum likelihood equation by equation, with no cross-equation restrictions such as equality of memory parameters. A full covariance matrix of the error terms is estimated from the residuals.

The other four benchmark models are based on a conditional Wishart distribution,

$$X_t|\mathscr{I}_{t-1} \sim W_n(\nu, S_t/\nu),$$

where $\mathscr{I}_t$ is the information set consisting of X_s, $s \leq t$, W_n denotes the central Wishart density, ν is the scalar degrees of freedom parameter and S_t/ν is a (6×6) positive definite scale matrix, which is related to the conditional mean of X_t by $E[X_t|\mathscr{I}_{t-1}] = S_t$. The baseline CAW(p,q) model of Golosnoy, Gribisch, and Liesenfeld (2012) specifies the conditional mean as

$$S_t = CC' + \sum_{j=1}^{p} B_j S_{t-j} B_j' + \sum_{j=1}^{q} A_j X_{t-j} A_j',$$

C, B_j and A_j denoting (6×6) parameter matrices, while the CAW-DCC model of Bauwens, Storti, and Violante (2012) employs a decomposition $S_t = H_t P_t H_t'$ where H_t is diagonal and P_t is a well-defined correlation matrix. As a sparse and simple DCC benchmark we apply univariate realized GARCH(p_v,q_v) specifications for the realized variances

$$H_{ii,t}^2 = c_i + \sum_{j=1}^{p_v} b_{i,j}^v H_{ii,t-j}^2 + \sum_{j=1}^{q_v} a_{i,j}^v X_{ii,t-j},$$

along with the 'scalar Re-DCC' model (Bauwens, Storti, and Violante, 2012) for the realized correlation matrix R_t,

$$P_t = \bar{P} + \sum_{j=1}^{p_c} b_j^c P_{t-j} + \sum_{j=1}^{q_c} a_j^c R_{t-j}.$$

The diagonal CAW(p,q) and the CAW-DCC(p,q) specification with $p = p_v = p_c = 2$ and $q = q_v = q_c = 1$ are selected since they provide a reasonable in-sample fit among various order choices. They are estimated by maximum likelihood using variance and correlation targeting.

Acknowledgements

This research has partly been done at the Institute of Economics and Econometrics of the University of Regensburg. Very valuable comments by Rolf Tschernig, by Enzo Weber and by participants of the Interdisciplinary Workshop on Multivariate Time Series Modeling 2011 in Louvain La Neuve, at the Statistische Woche 2011 in Leipzig, and of research seminars at the Universities of Regensburg, Augsburg and Bielefeld are gratefully acknowledged. The author is also thankful to Niels Aka for providing R codes to estimate model confidence sets.

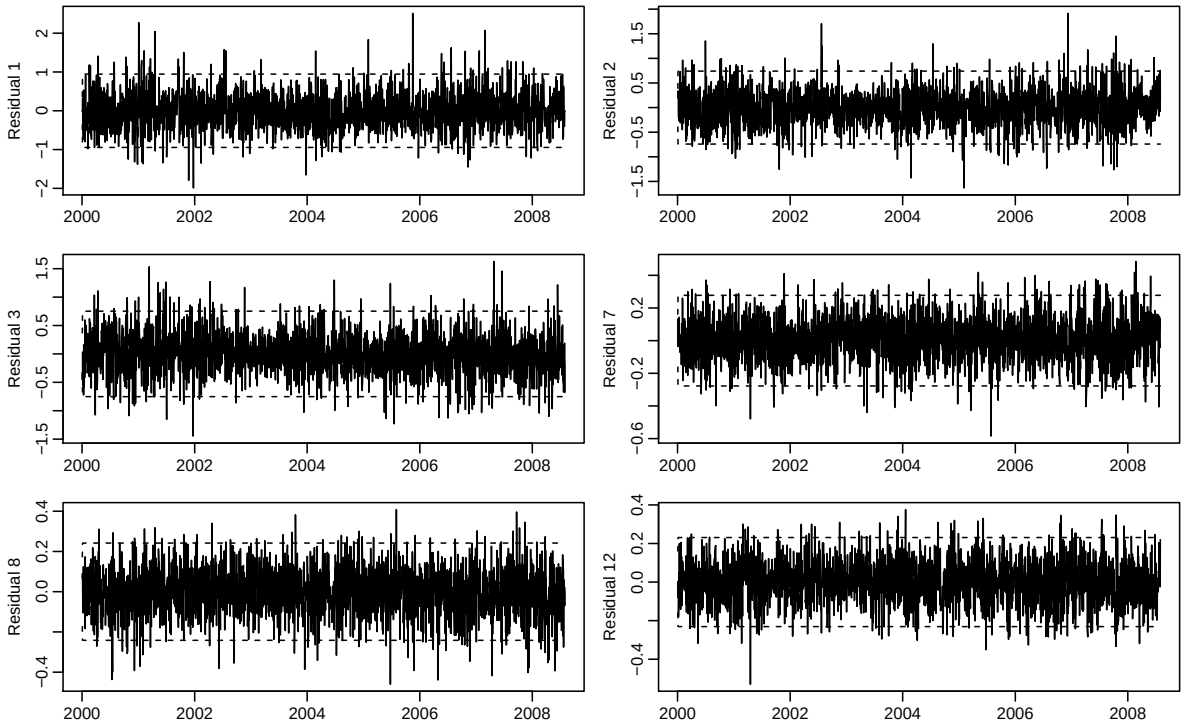

Figure 4.7: Residuals corresponding to log variances and z-transformed correlations for the first three assets for the fractional components model estimated in section 4.5.

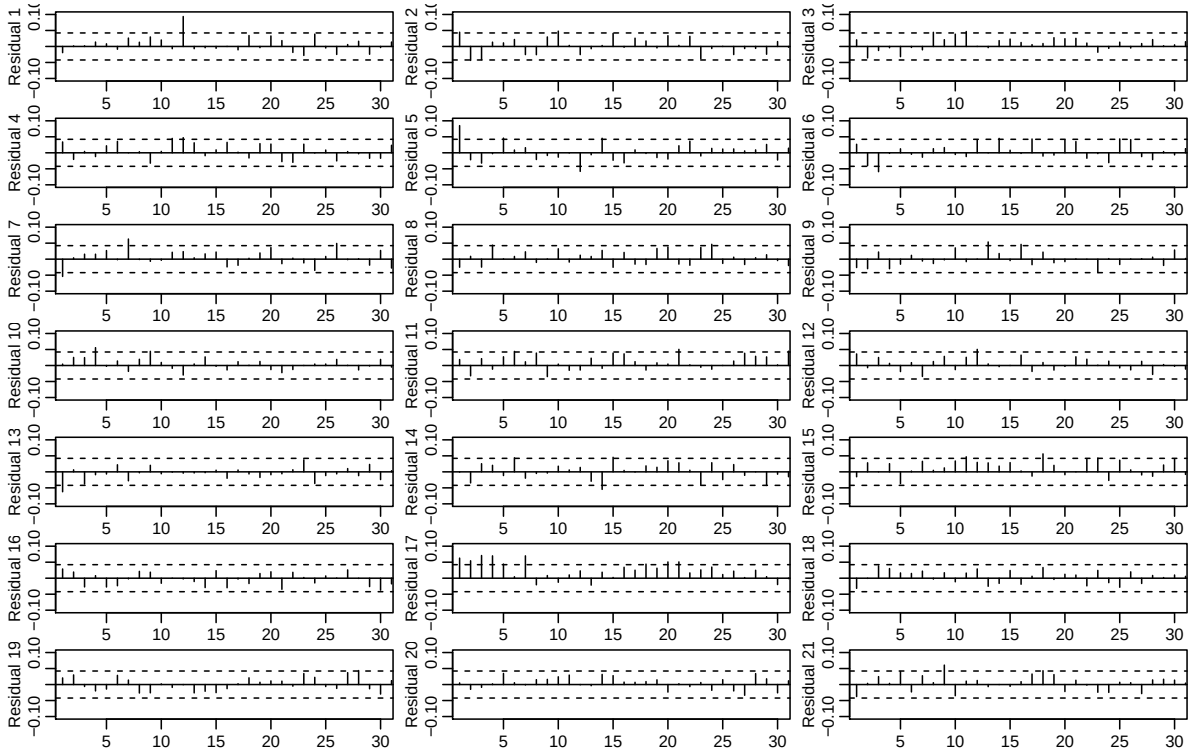

Figure 4.8: Residual autocorrelations for the fractional components model estimated in section 4.5.

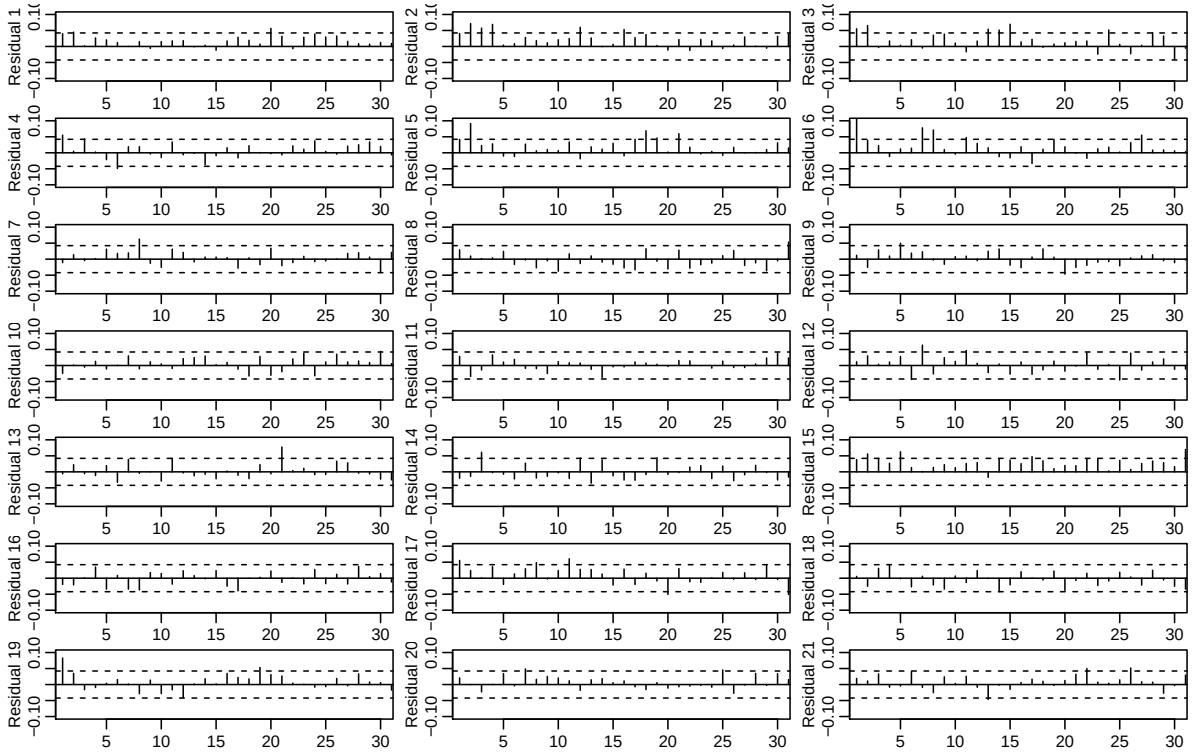

Figure 4.9: Autocorrelations of squared residuals for the fractional components model estimated in section 4.5.

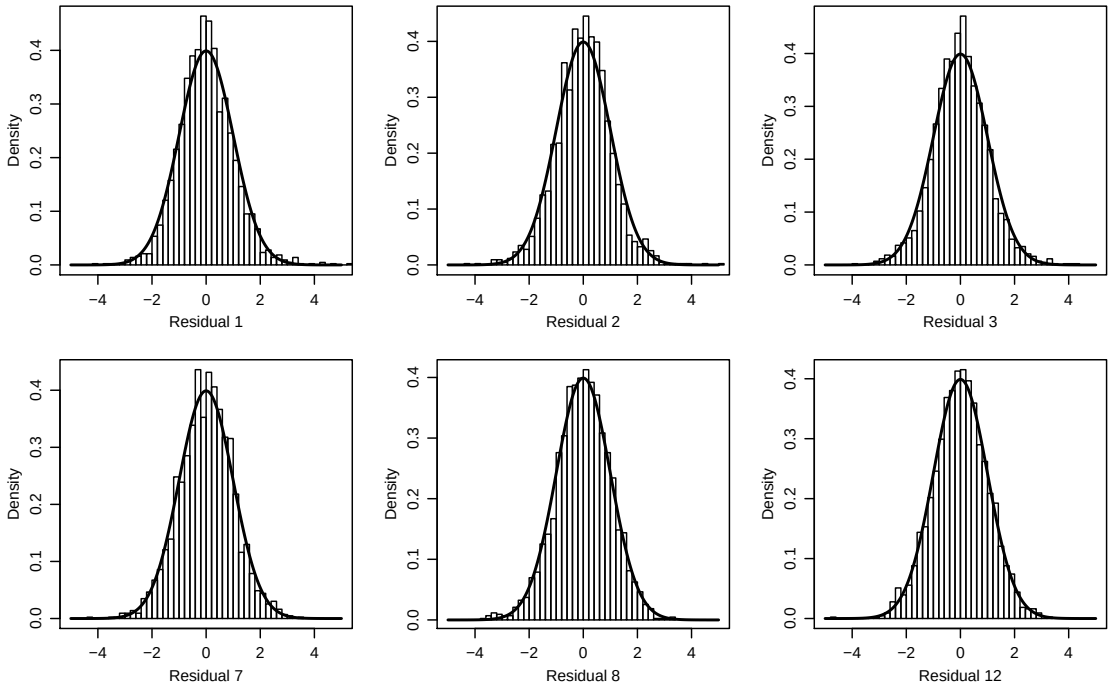

Figure 4.10: Histogram of residuals corresponding to log variances and z-transformed correlations for the first three assets for the fractional components model estimated in section 4.5 and normal density.

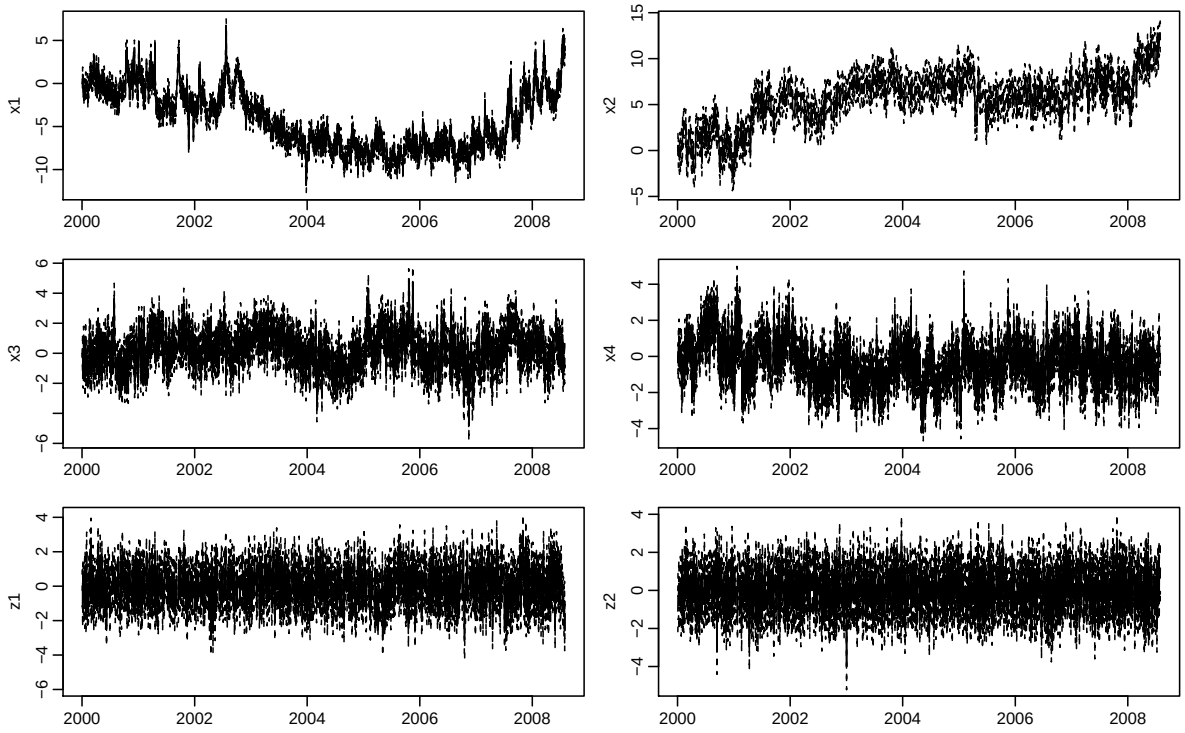

Figure 4.11: Selected smoothed fractional and nonfractional components (solid) ± 2 standard deviations (dashed) for the fractional components model estimated in section 4.5. Both nonstationary components (above), the first two stationary long-memory components (middle) and the short-memory components (below) are given.

5 Matrix Box-Cox Models for Multivariate Realized Volatility

This paper is also available as WEIGAND, R. (2014): "Matrix Box-Cox Models for Multivariate Realized Volatility," University of Regensburg Working Papers in Business, Economics and Management Information Systems 478.

Abstract. We propose flexible models for multivariate realized volatility dynamics which involve generalizations of the Box-Cox transform to the matrix case. The matrix Box-Cox model of realized covariances (MBC-RCov) is based on transformations of the covariance matrix eigenvalues, while for the Box-Cox dynamic correlation (BC-DC) specification the variances are transformed individually and modeled jointly with the correlations. We estimate transformation parameters by a new multivariate semiparametric estimator and discuss bias-corrected point and density forecasting by simulation. The methods are applied to stock market data where excellent in-sample and out-of-sample performance is found.

JEL-Classification. C14, C32, C51, C53, C58.

Keywords. Realized covariance matrix, dynamic correlation, semiparametric estimation, density forecasting.

5.1 Introduction

Dynamic modeling of multivariate financial volatility has recently gained significant interest. On the one hand, it constitutes an essential part of portfolio decisions, in empirical asset pricing models and for derivative analysis. On the other hand, recent financial crises have accentuated the importance of quantifying systemic risk. The latter also requires multivariate rather than univariate models. Such models require a precise measure of the otherwise latent asset variance and covariance processes and a

framework for modeling the dynamics. Precise measures are available due to recent and significant achievements on multivariate realized financial volatility modelling; see, e.g., Andersen, Bollerslev, Diebold, and Labys (2003) and Barndorff-Nielsen and Shephard (2004).

There are several approaches for modeling covariance dynamics. A prominent model class is based on conditionally Wishart distributed processes (see, e.g., Golosnoy, Gribisch, and Liesenfeld, 2012). Alternatively, linear vector time series models are applied to specific transformations of realized covariance matrices. The latter approach has the advantage of simplicity; model estimation, checking and inference is implemented in econometric software packages, while suitable ways of handling high-dimensional panels of time series are well-established. Various transformations have recently been suggested: Chiriac and Voev (2011), for instance, use the elements of a triangular matrix square-root transform, while the matrix logarithm has been considered by Bauer and Vorkink (2011) as well as Gribisch (2013). For these models, fitted covariance matrices and out-of-sample forecasts are automatically positive definite through the corresponding retransformations.

Likewise, approaches that separate variance and correlation dynamics, so called dynamic correlation (DC) models have been a fruitful direction of research. With appropriate factor or panel structure assumptions, Golosnoy and Herwartz (2012) model the z-transformed realized correlations (cf. (5.6) below). Correlation eigenvalues along with locally constant eigenvectors, sampled at different frequencies, are used by Hautsch, Kyj, and Malec (2014). Separate realized variance and correlation dynamics in mixed frequency models are also investigated by Halbleib and Voev (2011).

In the univariate time series literature, where transformation-based methods have a long tradition, the model of Box and Cox (1964) has become popular to find a suitable transform prior to ARIMA analysis (Box and Jenkins, 1970). Similar approaches were used for univariate volatility modeling, e.g., by Higgins and Bera (1992), Yu, Yang, and Zhang (2006), Zhang and King (2008) and Goncalves and Meddahi (2011).

We propose two flexible models in the spirit of Box and Cox (1964) for the dynamic multivariate realized volatility setup. Both generalize the univariate Box-Cox transform to the matrix case and contain several well-known transforms as special cases. The matrix Box-Cox model of realized covariances (MBC-RCov) is based on transformations of the covariance matrix eigenvalues. On the other hand, for the Box-Cox dynamic correlation (BC-DC) model, the variances are transformed individually and modeled together with the z-transformed correlations.

We introduce a semiparametric estimator of the transformation parameters in the

multivariate setup by generalizing the univariate approach of Proietti and Lütkepohl (2013). It does not require the specification of a dynamic model and makes a computationally simple two-step approach feasible. A simulation-based forecasting procedure is presented to reduce the bias of the naïve re-transform forecasts. Simulated paths of the realized volatilities may also be used to obtain density forecasts of the daily returns which will often be the aim of studying covariance matrix dynamics.

We apply these methods to the data set of Chiriac and Voev (2011) and find that a sparse vector autoregressive vector moving average (VARMA) specification provides a reasonable fit to the transformed series. A pseudo out-of-sample forecast comparison is conducted, where the BC-DC specification either with estimated transformation parameters or restricted to the logarithmic case emerges as favorable in practice. Bias correction provides significant improvements over the naïve forecasts. Notably, also the conditional Wishart models as popular benchmarks are outperformed by our transformation-based approach. These results are robust to different dynamic specifications and remain qualitatively intact for most of the loss functions recently used for evaluations of this kind.

The paper is organized as follows: In section 2 the new models are introduced. Parameter estimation and forecasting is described in sections 3 and 4, respectively. Section 5 presents the estimation results, while section 6 contains the out-of-sample forecast evaluation. Section 7 concludes.

5.2 Multivariate Box-Cox Volatility Models

In univariate regression and time series models, the Box-Cox transformation (Box and Cox, 1964) has been applied to obtain a linear, homoscedastic specification for the transformed dependent variable. For a scalar $x > 0$, it is parameterized by δ and given by

$$h(x;\delta) = \begin{cases} \frac{x^\delta - 1}{\delta} & \text{for } \delta \neq 0, \\ \log(x) & \text{for } \delta = 0. \end{cases} \tag{5.1}$$

For specific choices of δ, the transform corresponds to a linear mapping of the raw series ($\delta = 1$) or of various popular transforms such as the square root ($\delta = 0.5$), the logarithm ($\delta = 0$) and the inverse ($\delta = -1$). The reverse transform is given by

$$h^{-1}(y;\delta) = \begin{cases} (\delta y + 1)^{\frac{1}{\delta}} & \text{for } \delta \neq 0, \\ \exp(y) & \text{for } \delta = 0, \end{cases} \tag{5.2}$$

which is defined for $y > -\frac{1}{\delta}$ if $\delta > 0$ and for $y < -\frac{1}{\delta}$ if $\delta < 0$, and gives strictly positive values.

5.2.1 The Matrix Box-Cox Model of Realized Covariances

To generalize the Box-Cox method for modeling covariance matrices we define a matrix version of the latter, the matrix Box-Cox (MBC) transform. For a positive definite $(k \times k)$ covariance matrix X_t, and $t = 1, 2, \ldots$ denoting time periods, we suggest to apply Box-Cox transformations to the eigenvalues of X_t, each with a distinct transformation parameter collected in $\delta = (\delta_1, \ldots, \delta_k)'$,

$$Y_t(\delta) := H(X_t; \delta) = V_t \begin{pmatrix} h(\lambda_{1t}; \delta_1) & 0 & \ldots & 0 \\ 0 & h(\lambda_{2t}; \delta_2) & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & h(\lambda_{kt}; \delta_k) \end{pmatrix} V_t'. \tag{5.3}$$

Here $\lambda_{1t} \geq \ldots \geq \lambda_{kt} \geq 0$ are the eigenvalues of X_t, $h(\lambda_{it}; \delta_i)$, $i = 1, \ldots, k$, are their univariate Box-Cox transforms and V_t denotes the matrix of eigenvectors of X_t.

To understand the consequences of the MBC approach for modelling covariance matrices, it is useful to consider the inverse transformation, applied to a symmetric $(k \times k)$ matrix Y_t,

$$H^{-1}(Y_t; \delta) = V_t \begin{pmatrix} h^{-1}(\lambda_{1t}^y; \delta_1) & 0 & \ldots & 0 \\ 0 & h^{-1}(\lambda_{2t}^y; \delta_2) & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & h^{-1}(\lambda_{kt}^y; \delta_k) \end{pmatrix} V_t'. \tag{5.4}$$

Here, by λ_{jt}^y, $j = 1, \ldots, k$, we denote the eigenvalues of Y_t, while V_t contains the eigenvectors of both Y_t and $H^{-1}(Y_t; \delta)$, which remain unaffected by the transform. Notably, when the inverse MBC transform is well defined and applied to a symmetric matrix, the re-transformed fitted or forecasted matrices are always positive definite.

The reverse Box-Cox transform is not always well-defined, however. As mentioned below (5.2), existence of h^{-1} and hence of H^{-1} requires that the eigenvalues satisfy certain restrictions, namely $\lambda_{jt}^y(\delta_j) > -\frac{1}{\delta_j}$ for $\delta_j > 0$ and $\lambda_{jt}^y(\delta_j) < -\frac{1}{\delta_j}$ for $\delta_j < 0$. This requirement limits the set of feasible values of δ for a given sequence of matrices (e.g., forecasts) to which the inverse transform has to be applied. Our empirical results for

stock market data suggest that this potential drawback may be irrelevant as long as the applied transformation parameters are not chosen grossly at odds with estimates from the data (i.e. for $\delta_j > -0.25$ in our application).

As in the univariate setup, the matrix transform contains as special cases linear combinations of the raw matrix entries ($\delta_1 = \delta_2 = \ldots = \delta_k = 1$), of the (symmetric) matrix square root $\delta_1 = \delta_2 = \ldots = \delta_k = 0.5$), of the matrix logarithm ($\delta_1 = \delta_2 = \ldots = \delta_k = 0$) and of the inverse ($\delta_1 = \delta_2 = \ldots = \delta_k = -1$). It thus incorporates several empirically relevant approaches to covariance modeling within a common framework. We call this approach for modeling and forecasting multivariate realized volatility the matrix Box-Cox model of realized covariances (MBC-RCov).

For all periods $t = 1, \ldots, T$, the MBC-transform is applied to the realized covariance matrices X_t for an appropriate vector of parameters δ. In this way we obtain a sequence of symmetric matrices $Y_t(\delta)$ from which only the lower triangular elements (including the main diagonal) need to be modeled. A time series model is thus fitted only to the $k(k+1)/2$-dimensional vector process $y_t(\delta) := \text{vech}(Y_t(\delta))$[1]. For generality, we assume a linear process

$$y_t(\delta) = \sum_{j=0}^{\infty} \Psi_j(\theta) u_{t-j}, \qquad u_t \sim IID(0, \Sigma_u), \tag{5.5}$$

with $\Psi_0 = I$. We let θ as well as Σ_u consist of unknown parameters. Specific models will be considered in the empirical application in section 5.5. Here, we apply diagonal vector autoregressive moving average (VARMA) models as well as fractionally integrated VARMA (VARFIMA) and multivariate heterogeneous autoregressive (HAR) models.

5.2.2 The Box-Cox Dynamic Correlation Model

As an alternative to the matrix version of the Box-Cox transform, we consider a decomposition of variances and correlations. Applying the Box-Cox transform to the individual asset variances we introduce the Box-Cox dynamic correlation (BC-DC) model. In the spirit of dynamic conditional correlation models (Engle, 2002), we write $X_t = D_t R_t D_t$, where $D_t = \text{diag}(\sqrt{X_{11,t}}, \ldots, \sqrt{X_{kk,t}})$ is a diagonal matrix containing the univariate realized standard deviations while R_t is the sequence of realized correlation matrices.

[1] A different strategy would be to fit a model directly to the transformed eigenvalues and free elements of the eigenvectors, analogously to the approach of Hautsch, Kyj, and Malec (2014). They find that the eigenvectors are rather noisy and unstable at daily frequency which is not the case for our vech-transformation.

Applying Fisher's z-transformation

$$\tilde{R}_{ij,t} := \frac{1}{2}\log\frac{1+R_{ij,t}}{1-R_{ij,t}} \tag{5.6}$$

to the correlations has several advantages as compared to using the raw correlations (see Golosnoy and Herwartz, 2012), so that we propose modelling the vector time series

$$z_t(\delta) := g(X_t;\delta) := (h(X_{11,t};\delta_1),\ldots,h(X_{kk,t};\delta_k),\tilde{R}_{21,t},\tilde{R}_{31,t},\ldots,\tilde{R}_{k\,k-1,t})', \tag{5.7}$$

as a linear process analogous to (5.5).

The inverse BC-DC transform g^{-1}, when applied to forecasted z_{T+h}, yields positive variances due to the inverse Box-Cox and correlations in the range $(-1;1)$ due to the inverse Fisher transform. In contrast to the matrix Box-Cox approach, positive definiteness is not guaranteed for $k > 2$, however.[2] Whenever positive definiteness fails, it has to be enforced and a well-conditioned matrix must be obtained by some sort of eigenvalue trimming or shrinkage procedure. Positive definiteness, however, is not problematic empirically even in high-dimensional stock market applications for z-transformed correlation matrices as the results of Golosnoy and Herwartz (2012) suggest. Compared to the MBC-RCov approach, the estimated dynamics of the linear model (5.5) fitted to $z_t(\delta)$ are easily interpreted. The matrix Σ_u, for example, provides guidance about the extent of instantaneous co-movement within groups of variances or correlations but also between correlations and variances. Dynamic spill-overs may be modeled by non-diagonal specifications for $\Psi_j(\theta)$.

In addition to enabling a linear homoskedastic specification, the Box-Cox transform has originally been introduced to reduce the deviation from normality of the involved variables or model residuals. However, for the univariate transform (5.1) it holds that $h(x;\delta) > -\frac{1}{\delta}$ for $\delta > 0$ and $h(x;\delta) < -\frac{1}{\delta}$ for $\delta < 0$. Due to its bounded support, hence, the BC-transformed variable cannot literally be Gaussian whenever $\delta \neq 0$; see, e.g., Amemiya and Powell (1981). Merits of the transform even in cases where Gaussianity fails have been pointed out by Draper and Cox (1969). Although in the matrix case the MBC-transformed series are not individually bounded, the same logic implies that the MBC- and BD-DC-transformed series cannot be exactly multivariate normal. We do not need the Gaussianity assumption at this stage but empirically assess whether the

[2] As a counterexample where the unrestricted forecasts do not yield a valid correlation matrix, consider $k = 3$ and suppose that the inverse Fisher transform gives $R_{12,t} = R_{13,t} = 0.8$ along with $R_{23,t} = -0.8$. The quadratic form $\gamma' R_t \gamma$ is negative, e.g., for $\gamma = (1, -1, -1)'$.

transformed data are at least *approximately* Gaussian later on.

5.3 Semiparametric Estimation of the Transformation Parameter

In this section we discuss semiparametric estimation of the vector of transformation parameters δ. Among others, Han (1987) has proposed a semiparametric approach to estimate the transformation parameter of a single variable. Likewise, the recently developed estimator of Proietti and Lütkepohl (2013) for time series data does not involve specifying a parametric dynamic model. It is computed by minimizing a frequency-domain estimate of the prediction error variance of the transformed series. In the following, we generalize their approach to multivariate BC-DC and MBC-RCov setups. For the BC-DC model, our multivariate method provides a potentially more efficient estimator than applying the univariate estimator to all k variance series individually and allows to impose cross-equation restrictions. Moreover, in the MBC-RCov context, estimation is inherently multivariate and hence the existing semiparametric approaches would not be applicable without modifications.

In multivariate (vector) Box-Cox regression models, where each of the k nonnegative endogenous variables, say realized variances $(X_{11,t},\dots,X_{kk,t})'$, are transformed individually, the standard estimation strategy has been maximum likelihood under the auxiliary assumption of Gaussian transformed variables; see, e.g., Velilla (1993). Maximum likelihood estimation can be straightforwardly extended to the MBC-RCov model, as we outline in Appendix 5.A. In case of dynamic models, the likelihood is simultaneously maximized with respect to both, the dynamic and the transformation parameters. In contrast, our approach allows the researcher to proceed in two steps: After the estimation of the transformation parameters, which involves a k-dimensional optimization for both the BC-DC and MBC-RCov approach, the dynamic model specification and estimation is carried out for the transformed series as if δ was known.

To sketch our semiparametric approach for a generic k-dimensional vector process x_t with strictly positive elements, we consider the Jacobian of the vector Box-Cox transform

$$J_t(\delta) := \left.\frac{\partial(h(x_1;\delta_1),\dots,h(x_k;\delta_k))'}{\partial x'}\right|_{x=x_t},$$

such that a normalized transform with unit Jacobi determinant is given by

$$\xi_t(\delta) := \left| \prod_{s=1}^{T} J_s(\delta) \right|^{-\frac{1}{kT}} (h(x_{1,t};\delta_1), \ldots, h(x_{k,t};\delta_k))'. \tag{5.8}$$

The transformed values $h(x_{j,t};\delta_j)$ are corrected for the change in scale induced by the Box-Cox function. The Jacobian is diagonal with elements given by $J_{jj,s}(\delta) = x_{j,s}^{\delta_j - 1}$, so that $\xi_t(\delta)$ is easily computed from $\{x_t\}_{t=1}^{T}$ and δ. Alternatively, the well-known normalization $\check{\xi}_{j,t}(\delta_j) := (\prod_{s=1}^{T} x_{j,s})^{\frac{\delta_j - 1}{T}} h(x_{j,t};\delta_j)$ can be applied to the individual time series. It also succeeds in obtaining scale-invariance and gives numerically identical results for the estimation procedure described in this section.

Without referring to a specific dynamic process, denote the one-step ahead prediction error as $\eta_t(\delta) := \xi_t(\delta) - Proj(\xi_t(\delta)|\mathscr{I}_{t-1})$, where $Proj(\cdot|\mathscr{I}_{t-1})$ is the best linear predictor of a time series given an information set $\mathscr{I}_{t-1}$ which consists of the series' own past in this case and let $\Sigma_\eta(\delta) := Var(\eta_t(\delta))$. Under the assumption that there exists a vector δ^* for which $E(\xi_t(\delta^*)|\mathscr{I}_{t-1})$ is linear in $\xi_{t-j}(\delta^*)$, $j \geq 1$, we characterize this true value δ^* as minimizing the determinant of the prediction error covariance matrix $|\Sigma_\eta(\delta)|$, the so-called generalized variance of $\eta_t(\delta)$.

A *least generalized variance* estimator for δ becomes feasible by utilizing the nonparametric methods proposed by Jones (1976) and further developed by Mohanty and Pourahmadi (1996) to obtain nonparametric estimates of $|\Sigma_\eta(\delta)|$, for a given δ. This generalized prediction error variance is related to the $(k \times k)$ spectral density matrix $g_\xi(\omega)$ of ξ_t by a multivariate extension of the Szegö-Kolmogoroff-Formula (cf. Priestley, 1982, p. 761),

$$\log|\Sigma_\eta| = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log 2\pi |g_\xi(\omega)| d\omega.$$

In practice, the integral may be approximated by the mean over a finite number M of frequencies, $\omega_j = (\pi j)/(M+1)$ for $j = 0, 1, \ldots, M-1$,

$$\log|\Sigma_\eta^M| = \frac{1}{M} \sum_{j=0}^{M-1} \log 2\pi |g_\xi(\omega_j)|,$$

while the unknown spectral density can be estimated by smoothing the $(k \times k)$ periodogram matrix $I_\xi(\omega;\delta)$ of $\xi_t(\delta)$ over frequencies in the neighborhood of ω,

$$\hat{g}_\xi(\omega;m) = \sum_{|l|<m} W_m(l) I_\xi\left(\omega + \frac{2\pi l}{T};\delta\right).$$

To this end, a bandwidth m and a kernel $W_m(l)$ are applied for which $m \to \infty$, $m/T \to 0$ as $T \to \infty$ and $\sum_{|l|<m} W_m(l) = 1$ hold, and which satisfy also further regularity conditions of Mohanty and Pourahmadi (1996). Taken together, a straightforward estimator for the innovation generalized variance satisfies

$$\log|\widehat{\Sigma}_{\eta}^{M}(\delta;m)| = \frac{1}{M}\sum_{j=0}^{M-1}\log\left|2\pi\sum_{|l|<m} W_m(l) I_{\xi}\left(\omega_j + \frac{2\pi l}{T};\delta\right)\right|. \tag{5.9}$$

A possible bias correction term for estimating $|\widehat{\Sigma}_{\eta}^{M}(\delta;m)|$ is not considered here since it does not change the optimization problem for the resulting estimator,

$$\hat{\delta} = \arg\min_{\delta} \log|\hat{\Sigma}_{\eta}^{M}(\delta;m)|. \tag{5.10}$$

The univariate minimum prediction error variance approach of Proietti and Lütkepohl (2013) results as a special case for $k = 1$ by choosing the uniform kernel $W_m(l) = 1/(2m-1)$ and averaging the smoothed log periodogram over $M = [(T-1)/(2m)]$ frequencies.

In the matrix Box-Cox model, semiparametric estimation of the transformation parameter is more demanding since a scale-preserving normalized transform as in (5.8) is not available in closed form. In this context, define $\tilde{x}_t := \text{vech}(X_t)$, denote the MBC transformation in vech-space as $\varphi : \tilde{x}_t \mapsto y_t(\delta)$ and the corresponding Jacobi-matrix as

$$\tilde{J}_t(\delta) := \left.\frac{\partial\varphi(\tilde{x};\delta)}{\partial\tilde{x}'}\right|_{\tilde{x}=\tilde{x}_t} \tag{5.11}$$

for a given observation. A normalized transform is obtained by

$$\tilde{\xi}_t(\delta) := \left|\prod_{s=1}^{T}\tilde{J}_s(\delta)\right|^{-\frac{1}{0.5k(k+1)T}} y_t(\delta).$$

For computational reasons, it is often preferable to work with log determinants by substituting the latter expression into the log of the innovation generalized variance,

$$\log|\widehat{\Sigma}_{\eta}^{M}(\delta;m)| = \log|\widehat{\Sigma}_{u}^{M}(\delta;m)| - \frac{2}{T}\sum_{t=1}^{T}\log\left|\tilde{J}_t(\delta)\right|, \tag{5.12}$$

where $|\widehat{\Sigma}_{u}^{M}(\delta;m)|$ is the estimated generalized innovation variance of the non-normalized transform $y_t(\delta)$; see (5.5). When used as an objective for minimization with respect to δ, the Jacobi determinant has to be evaluated numerically.

As a first check if a transformation is relevant for a specific problem at all, it is useful to construct interval estimates for the transformation parameters. If the intervals include unity, then an untransformed approach may be used. Alternatively, matrix logarithmic or square-root models may be a reasonable approximation if the corresponding δ (0 or 0.5, respectively) is contained in the confidence region. Such regions for BC-DC and MBC-RCov transformation parameters can be based on the pivot method, see Casella and Berger (2002, Sec. 9.2.2).

To see how this can be achieved in the current setup, note that for a given δ, the asymptotic distribution of the log innovation generalized variance estimate does not depend on unknown parameters. Mohanty and Pourahmadi (1996, Theorem 3.1(c)) show that under reasonable conditions, for M fixed and $T \to \infty$,

$$\frac{\sqrt{M}}{\sqrt{k \sum_{|j|<m} W_m(j)^2}} \left(\log|\widehat{\Sigma}_\eta^M(\delta; m)| - \log|\Sigma_\eta^M(\delta)| \right) \xrightarrow{d} N(0;1),$$

which is an asymptotically pivotal statistic. A feasible confidence interval for $\log|\Sigma_\eta^M(\delta^*)|$ is given by

$$\log|\widehat{\Sigma}_\eta^M(\hat{\delta}; m)| \pm \frac{\sqrt{M}}{\sqrt{k \sum_{|j|<m} W_m(j)^2}} z_{1-\alpha/2}, \tag{5.13}$$

where $z_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of the standard normal distribution. Following Proietti and Lütkepohl (2013), the confidence region for δ consists of all values δ for which $\log|\widehat{\Sigma}_\eta^M(\delta; m)|$ is contained in (5.13).

As a practical issue, for both point and interval estimation, the bandwidth parameter m has to be selected. For the univariate approach, a value of $m = 3$ has been found to provide a good balance between bias and variance in Monte Carlo simulations by Proietti and Lütkepohl (2013). In the multivariate case, $m > k$ is required to have positive definite spectral density estimates and hence a nonzero determinant of $\hat{g}(\omega; \delta)$. We try different choices of m in the empirical application below to assess the robustness with respect to the bandwidth choice. Furthermore, we follow Mohanty and Pourahmadi (1996) and set M to the integer part of $0.5T \sum_{|j|<m} W_m(j)^2$, while a uniform kernel is used throughout.

Using this procedure for the realized covariance models introduced in section 5.2, the transformation parameters can be estimated in a first step. While the MBC-RCov model calls for the multivariate approach, for the BC-DC model either the individual asset variances $X_{ii,t}$ may be used to determine δ_i, $i = 1, \ldots, k$ in turn, or the minimization (5.10) is carried out for the full vector of realized variances. Leaving dynamic model

specification and estimation to a second step makes the analysis computationally convenient: Estimates of the dynamic parameters θ and innovation covariance matrix Σ_u are determined from $y_t(\hat{\delta})$ or $z_t(\hat{\delta})$, respectively. Depending on the dynamic specification, e.g., least squares or Gaussian quasi maximum likelihood methods may be considered.

5.4 Forecasting and Bias Correction

5.4.1 Realized Covariance Forecasting

Once the parameters of the MBC-RCov model have been appropriately determined, it can be used for forecasting

$$y_{T+h|T}(\delta) := E[y_{T+h}(\delta)|\mathscr{I}_T],$$

where $\mathscr{I}_T$ consists of both returns and realized covariances up to period T. To obtain forecasts of the realized covariance matrices it is necessary to re-transform these predictions into positive definite matrices. Reconstructing a symmetric $(k \times k)$ matrix $Y_{T+h|T}(\delta) = \text{vech}^{-1}(y_{T+h|T}(\delta))$ and applying the inverse of the MBC transform

$$\widehat{X}_{T+h|T} = H^{-1}(Y_{T+h|T}; \hat{\delta}) \tag{5.14}$$

may be used as a naïve point forecasts of the realized covariance matrix X_{T+h}.

Due to the nonlinearity of the MBC-transform and its inverse, point forecasts obtained in this way may be severely biased for the conditional mean of X_{T+h}. We therefore propose a simple simulation-based bias correction. Given estimated or pre-specified parameters δ, Σ_u, θ and assuming normally distributed disturbances, we simulate realizations of $y^{(i)}_{T+h}(\delta)$ given $\mathscr{I}_T$ from the model (5.5) using simulated errors

$$u^{(i)}_t \sim N(0, \Sigma_u), \quad t = T+1, \ldots, T+h, \quad i = 1, \ldots, R.$$

The reverse MBC-transform yields positive definite $X^{(i)}_{T+h}$, $i = 1, \ldots, R$. Averaging over these simulated covariance matrices provides an approximately unbiased point forecast

$$\widehat{X}_{T+h|T} = \frac{1}{R} \sum_{i=1}^{R} X^{(i)}_{T+h}, \tag{5.15}$$

provided that the normality assumption gives a good description of the actual data generating process. A re-sampling of the model residuals to draw paths of $y^{(i)}_{T+j}(\delta)$ may

lead to a procedure which is more robust to deviations from normality. The same procedure can be straightforwardly applied to the BC-DC model and other approaches to transformation-based forecasting as well.

As has been pointed out in section 5.2, the normality assumption for transformed variables cannot be satisfied whenever $\delta \neq 0$ due to their bounded support. Correspondingly, the re-transformed values of simulated trajectories may not always exist. We circumvent this shortcoming by using draws from a truncated distribution as follows: We first draw paths $u_t^{(i)}$, $t = T, \ldots, T+h$ for $i = 1, \ldots R$ as described above. Whenever a simulated value $y_{T+j}^{(i)}(\delta)$ cannot be re-transformed, we discard the whole trajectory and average over the remaining ones in our bias-correction.

5.4.2 Forecasting the Return Distribution

In addition to point forecasts of realized covariance matrices, we consider density prediction of daily returns r_{T+h} conditional on information in period T, denoted $f_r(r_{T+h}|\mathscr{I}_T)$, as this is a key input, e.g., to portfolio decisions and value-at-risk assessment. Joint models of realized covariance and return dynamics have been found beneficial to obtain suitable density forecasts, see Noureldin, Shephard, and Sheppard (2012) or Jin and Maheu (2013). In a Bayesian framework of conditional Wishart models the latter propose computation of such predictive densities that also involves the parameter uncertainty. Our frequentist setup naturally differs from their approach by treating the parameters as fixed and ignoring the estimation error in the computation of density forecasts.

Depending on the intra-day dynamics of returns, the method for computing X_t and the time-span from which daily returns are computed (open-to-close versus close-to-close returns), the unconditional mean of X_t may differ from the unconditional covariance matrix of daily returns, and a re-scaling of the realized measure will be needed. We follow Jin and Maheu (2013) and assume for the daily returns

$$r_t | X_t \sim N(0, X_t^{\frac{1}{2}} \Lambda X_t^{\frac{1}{2}}), \tag{5.16}$$

where the parameters of the symmetric $(k \times k)$ scaling matrix Λ are estimated by maximum likelihood using daily returns r_t, $t = 1, \ldots, T$. Suppressing the conditioning on $\mathscr{I}_T$ for notational convenience, we obtain draws from $f_X(X_{T+h})$ by simulating $X_{T+h}^{(i)}$ as described in section 5.4.1 and hence approximate the predictive density at a point r_{T+h}

according to

$$f_r(r_{T+h}) = \int f_{r|X}(r_{T+h}|X_{T+h}) f_X(X_{T+h}) \mathrm{d}X_{T+h} \approx \frac{1}{R} \sum_{i=1}^{R} f_{r|X}(r_{T+h}|X_{T+h}^{(i)}), \tag{5.17}$$

where $f_{r|X}$ is the multivariate normal density of (5.16).

5.5 Estimation Results

We apply the proposed models to US stock market return and volatility data to assess their usefulness in practice. The data set of Chiriac and Voev (2011) is used which is based on tick-by-tick bid and ask quotes from the NYSE Trade and Quotations (TAQ) database for the period from 2000-01-01 to 2008-07-30 ($T = 2156$). Six liquid stocks are considered, namely (1) American Express Inc., (2) Citigroup, (3) General Electric, (4) Home Depot Inc., (5) International Business Machines and (6) JPMorgan Chase & Co. Chiriac and Voev (2011) describe the computation of the realized covariance matrices from intraday data. They are available from `http://qed.econ.queensu.ca/jae/2011-v26.6/chiriac-voev`.

Estimates of the transformation parameters for the MBC-RCov model are shown in the upper panel of Table 5.1. They are computed by minimizing (5.12) for different bandwidths m, using the uniform kernel and setting M to the integer part of $0.5T\sum_{|j|<m} W_m(j)^2$ in (5.9). Both the unrestricted estimates and the restricted ones under $\delta_1 = \ldots = \delta_k$ are presented. The table reveals that the estimated δ_j are negative but close to zero. The values are all in the range from -0.05 to 0 and are insensitive with regard to the choice of the bandwidth parameters. The confidence intervals which are shown below the respective unrestricted point estimates and at the right of the restricted ones provide statistical evidence against untransformed, matrix square-root or inverse models. In contrast, the matrix logarithmic model is supported, since all confidence intervals include zero.

To assess the robustness with respect to the estimation method, we also provide the maximum likelihood results for a simple baseline dynamic specification. Details on the estimation approach are given in Appendix 5.A. We assume that the vector of MBC-transformed series follows a VARMA process without dynamic spill-overs, so that each

Table 5.1: Estimates of the transformation parameters $\delta_1, \ldots, \delta_k$ for the covariance eigenvalues in the MBC-RCov model based on (5.3) and (5.5). The semiparametric estimator outlined in section 5.3 as well as the Maximum Likelihood estimator for the VARMA(2,1) specification (5.18) (see Appendix 5.A) are applied to the realized covariance matrices. Estimates under the constraint $\delta_1 = \ldots = \delta_k$ ("restricted") and without this constraint ("unrestricted") are presented.

m	δ_1	δ_2	δ_3	δ_4	δ_5	δ_6
			Semiparametric unrestricted			
21	-0.0306	-0.0397	-0.0180	-0.0144	-0.0176	-0.0266
	[-0.11; 0.06]	[-0.11; 0.03]	[-0.09; 0.05]	[-0.09; 0.06]	[-0.09; 0.06]	[-0.12; 0.07]
42	-0.0265	-0.0315	-0.0272	-0.0271	-0.0143	-0.0231
	[-0.11; 0.06]	[-0.10; 0.03]	[-0.09; 0.04]	[-0.09; 0.06]	[-0.10; 0.06]	[-0.12; 0.07]
63	-0.0230	-0.0274	-0.0206	-0.0157	-0.0162	-0.0253
	[-0.10; 0.06]	[-0.11; 0.03]	[-0.08; 0.06]	[-0.09; 0.06]	[-0.09; 0.06]	[-0.12; 0.07]
			Semiparametric restricted $\delta_1 = \ldots = \delta_k$			
21			-0.0328	[-0.12; 0.05]		
42			-0.0270	[-0.12; 0.05]		
63			-0.0252	[-0.12; 0.05]		
			Maximum Likelihood			
unrestr.	-0.0238	-0.0279	-0.0225	-0.0180	-0.0213	-0.0316
restr.			-0.0239			

series is represented by an ARMA(p,q)

$$(1 - \alpha_{i1}L - \ldots - \alpha_{ip}L^p)(y_{it}(\delta) - \mu_i) = (1 + \phi_{i1}L + \ldots + \phi_{iq}L^q)u_{it}(\delta),$$
$$i = 1, \ldots, k(k+1)/2. \quad (5.18)$$

If ARMA models are estimated for $y_{it}(\hat{\delta})$, $i = 1, \ldots, k$, with $\hat{\delta}$ determined by the semiparametric method, the BIC favors orders (p, q) of (1,1), (2,1) or, less frequently, (2,2) for all but 2 of the 21 series. We choose the diagonal VARMA(2,1) model as it reconciles these outcomes with a quest for parsimony. With three dynamic parameters per series, this model is similar in complexity to recent successful approaches to daily (co-)variance modeling; see, e.g., Chiriac and Voev (2011). As in their model, Granger-causality relations between the series are excluded. The maximum likelihood estimates, given in the lower panel of Table 5.1, are close to the semiparametric ones.

Estimation results for the transformation parameters of univariate asset variances, which are utilized in the BC-DC approach, are provided in Table 5.2. The Box-Cox parameters are again negative with small magnitude. In contrast to the MBC-RCov case, the univariate estimates are spread over the range $[-0.1;0]$; the multivariate estimates are closer to zero again. By including zero, the confidence intervals suggest that a linear time series model may simply be applied to the log realized variances and z-transformed correlations. This conclusion does not change when the maximum likelihood estimator for the diagonal VARMA(2,1) model is considered. Again, this is the specification which is individually favored by the BIC for most series.

Figure 5.1 shows two time series plots of each, the raw series X_t, the MBC-transformed $y_t(\hat{\delta})$ (with semiparametric estimates $\hat{\delta}_j$ and $m = 42$) and the BC-DC transformed $z_t(\hat{\delta})$ (with univariate semiparametric estimates $\hat{\delta}_j$ and $m = 3$) in the left two panels. While large positive outliers occur and volatility strongly co-moves with the level of the untransformed variances and covariances, the distribution of the transformed series is more stable and symmetric around their persistent movements. Interestingly, the dynamics of the first diagonal entry of the MBC series, $y_{1,t}(\hat{\delta})$, are very similar to the univariate transformed variance $z_{1,t}(\hat{\delta})$, while the nondiagonal MBC entry $y_{2,t}(\hat{\delta})$ evolves in close parallel to the corresponding z-transformed correlation $z_{7,t}(\hat{\delta})$. This reflects the finding of Gribisch (2013, p.4) who found and discussed the closeness of the matrix logarithm's diagonal and off-diagonal elements to log variances and correlations, respectively.

We assess the in-sample success of various nonlinear transforms for our dataset. The appropriateness of a transform for modeling purposes can be seen by a parsimonious ARMA representation for the transformed variable. Additionally, the stabilization of conditional variances, i.e. the conditional homoscedasticity of the residuals, is an important goal of the transformation. Further, the approximate normality of transformed variable or model residuals are frequently stated as the motivation for a transformation-based approach. To evaluate these goals, diagnostic residual tests are carried out for our benchmark VARMA(2,1) specification which is applied to different transforms. As straightforward and familiar choices, we use univariate Ljung-Box tests both on raw residuals and on squared residuals to check serial correlation and conditional heteroscedasticity, respectively, while Jarque-Bera tests are used to detect deviations from normality. In our comparison, we consider the raw realized variances and covariances (vech), the nonzero terms of the Cholesky factors (chol) as well as the unique terms of the symmetric matrix square-root (sym-root), of the matrix logarithm (mlog) and of the inverse covariance matrices (inverse).

Table 5.2: Estimates of the transformation parameter $\delta_1, \ldots, \delta_k$ for the realized variances in the BC-DC model based on (5.7) and (5.5). The semiparametric estimator outlined in section 5.3 as well as the Maximum Likelihood estimator for the VARMA(2,1) specification (5.18) (see Appendix 5.A). The variances are ordered as (1) American Express Inc., (2) Citigroup, (3) General Electric, (4) Home Depot Inc., (5) International Business Machines and (6) JPMorgan Chase & Co. Estimates under the constraint $\delta_1 = \ldots = \delta_k$ ("restricted") and without this constraint ("unrestricted") are presented, while this restriction is not possible for the univariate estimators.

m	δ_1	δ_2	δ_3	δ_4	δ_5	δ_6
			Semiparametric univariate			
1	-0.0616	-0.0934	-0.0715	-0.0346	-0.0762	-0.0569
	[-0.20; 0.07]	[-0.23; 0.05]	[-0.23; 0.08]	[-0.22; 0.14]	[-0.25; 0.09]	[-0.19; 0.08]
3	-0.0619	-0.0888	-0.0645	-0.0382	-0.0758	-0.0525
	[-0.20; 0.07]	[-0.22; 0.04]	[-0.22; 0.09]	[-0.22; 0.14]	[-0.24; 0.08]	[-0.19; 0.08]
10	-0.0613	-0.0832	-0.0592	-0.0353	-0.0708	-0.0533
	[-0.19; 0.07]	[-0.22; 0.05]	[-0.21; 0.09]	[-0.22; 0.13]	[-0.24; 0.09]	[-0.19; 0.07]
			Semiparametric multivariate unrestricted			
6	-0.0093	-0.0354	-0.0433	-0.0165	-0.0501	-0.0127
	[-0.2; 0.17]	[-0.2; 0.12]	[-0.25; 0.15]	[-0.29; 0.24]	[-0.3; 0.18]	[-0.18; 0.14]
12	-0.0189	-0.0365	-0.0346	-0.0114	-0.0435	-0.0095
	[-0.2; 0.16]	[-0.2; 0.12]	[-0.24; 0.16]	[-0.27; 0.23]	[-0.29; 0.18]	[-0.18; 0.15]
36	-0.0205	-0.0376	-0.0311	-0.0101	-0.0443	-0.0158
	[-0.21; 0.16]	[-0.21; 0.12]	[-0.24; 0.17]	[-0.28; 0.24]	[-0.29; 0.18]	[-0.19; 0.14]
			Semiparametric multivariate restricted			
6			-0.0268 [-0.13; 0.08]			
12			-0.0259 [-0.12; 0.07]			
36			-0.0267 [-0.12; 0.07]			
			Maximum Likelihood			
univ.	-0.0569	-0.0795	-0.0596	-0.0367	-0.0676	-0.0604
unrestr.	-0.0121	-0.0123	-0.0121	-0.0121	-0.0121	-0.0114
restr.			-0.0122			

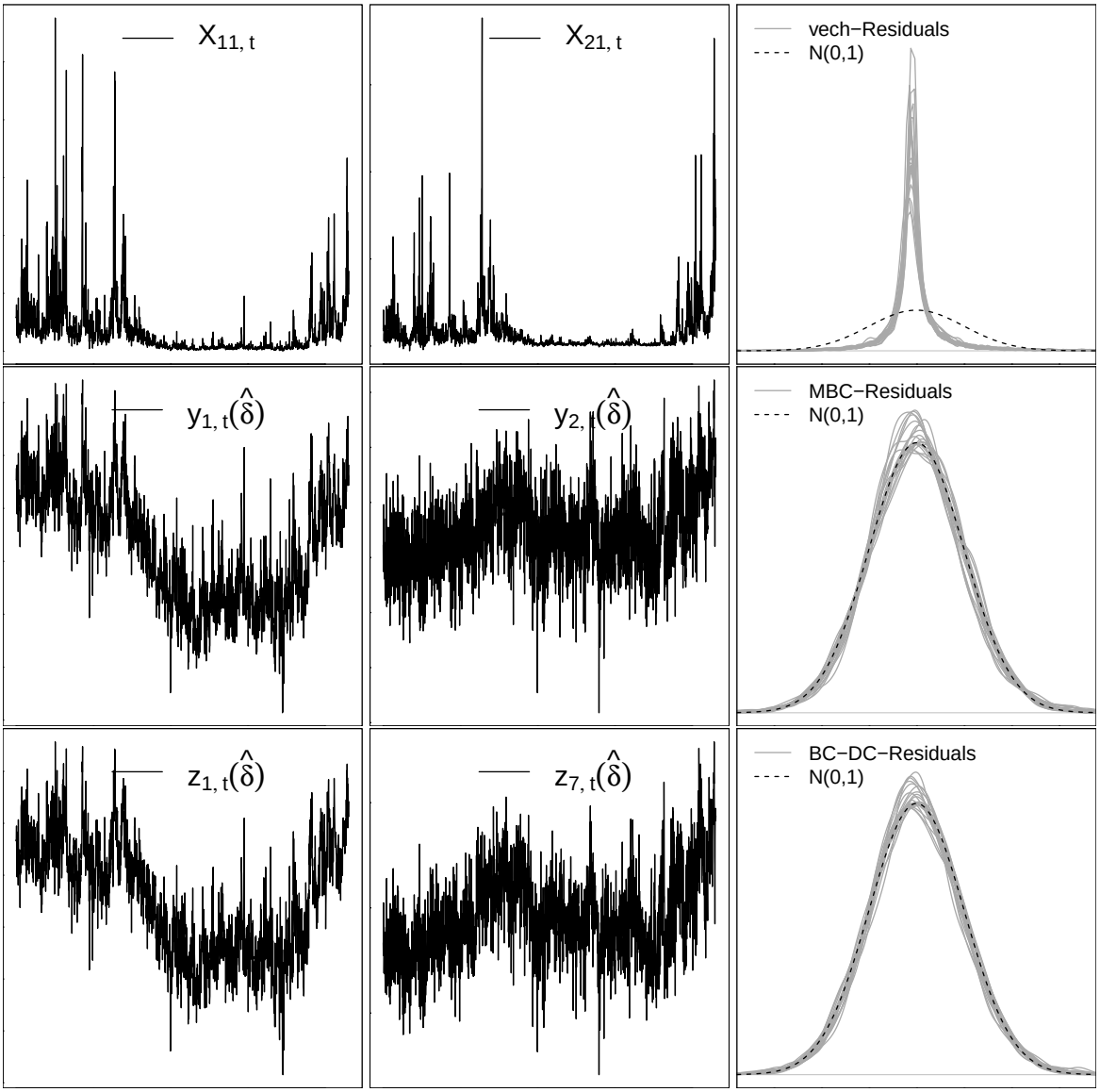

Figure 5.1: Left two panels: Time series plots of raw variance and covariance series (above), of the MBC-transformed data with δ estimated by the unrestricted semiparametric estimator (middle), as well as the BC-transformed variance $z_{1t}(\hat{\delta})$ and z-transformed correlation $z_{7t}(\hat{\delta}) = \tilde{R}_{12,t}$. Right panel: Kernel density estimates of standardized VARMA(2,1) residuals (grey) of the vech, MBC-RCov and BC-DC series. The standard normal density is shown as the black dashed line.

The results of these diagnostic tests, more precisely the number of rejections across series for different significance levels, is given in Table 5.3. Transformations of the realized covariance matrices which are not appropriate may be detectable by the failure of linear time series models to produce serially uncorrelated residuals. Therefore, consider the test for residual autocorrelation, given in the upper panel. Both the MBC-RCov and the BC-DC models have a non-negligible fraction of rejections. There remains significant autocorrelation even at the 0.1% level for one of the 21 series. Autocorrelation is modest compared to some of the other transformations, e.g., the Cholesky factorization, however. For the latter, the p-values are smaller than 1% in 9 cases and the majority (14 out of 21) of the residual series exert significant autocorrelation at the 5% level. The matrix logarithm transform is the hardest competitor but still exceeds the BC-DC model with respect to the number of rejections (4 versus 2 at the 1% level). A better fit of the models could be obtained by using larger VARMA model orders globally, or by specifying the ARMA models individually for each series. We have consistently chosen an intuitive and sparse specification to enhance comparability with other model classes in light of the out-of-sample forecasting study carried out below.

Compared to the other transforms and special cases, the BC-DC and even more the MBC-RCov model (2 and 1 rejection, respectively, at the 1% level) succeed in stabilizing the residual variance. Except for the mlog-transform, all other models suffer from extreme conditional heteroscedasticity with very high rejection rates when autocorrelation of the squared residuals is tested.

Normality of the model residuals is frequently rejected for the Box-Cox specifications, namely 10 times out of 21 at the 1% level. It thus performs worse than the matrix logarithm. Given that the matrix logarithm is contained in the family of MBC-transforms it turns out that when choosing the transformation parameters for our dataset, we face a tradeoff between obtaining linear time series models, stabilizing variance and yielding normally distributed variables. The estimated $\delta_j < 0$ thus prioritize the former goals but fail with respect to the latter. Still, the normal approximation is strongly improved compared to the raw variance and covariance series. Kernel density plots of the model residuals in the right panel of Figure 5.1 show the approximate normality of the transformed series as compared to the untransformed ones.

Table 5.3: Number of rejections for univariate diagnostic residuals tests of the VARMA(2,1) model (5.18) based on different transformations: the raw variance and covariance processes (vech), the Cholesky decomposition (chol), the symmetric matrix square root (sym-root), the matrix logarithm (mlog), the inverse as well as MBC and BC-DC given by (5.3) and (5.7), respectively, with estimated transformation parameters. Upper panel: Ljung-Box test with 10 lags for no autocorrelation in residuals. Middle panel: Ljung-Box tests with 10 lags for no autocorrelation in *squared* residuals. Lower panel: Jarque-Bera test for normality of the residuals.

P-Value	vech	chol	sym-root	mlog	inverse	MBC	BC-DC
Residual autocorrelation							
0.05	21	14	14	8	15	7	6
0.01	21	9	11	4	11	4	2
0.001	17	6	10	1	7	1	1
Conditional heteroskedasticity							
0.05	21	21	21	4	21	3	4
0.01	20	21	21	2	21	1	2
0.001	19	21	21	1	21	1	1
Non-normality							
0.05	21	21	21	11	21	13	12
0.01	21	21	21	9	21	10	10
0.001	21	21	21	8	21	8	8

5.6 Forecast Comparison

We assess the forecast performance of the MBC-RCov and the BC-DC models, also in light of popular competitor methods. To this end, we use the data set introduced in section 5.5 and conduct a quasi out-of-sample forecast exercise, recursively using a pre-specified window of data for parameter estimation and forecasting, and then, subsequently, evaluating the forecasts against realized data outside that range.

We address the following questions in the forecasting exercise. To take a closer look at the methods introduced in section 5.4.1, we first assess whether bias-correction leads to a significant improvement of the forecasts. Further, for both the MBC-RCov and the BC-DC model, we evaluate which value of the transformation parameter δ dominates in terms of out-of-sample precision and whether the estimates presented in section 5.5 are also superior out of sample. We are also interested in whether the dynamic correlation specification outperforms the matrix transform or vice versa.

Another main objective of the study is to assess the value of the transformation-based approach as compared to other methods. As a recently suggested and popular competitor we consider the class of models that assume conditionally Wishart distributed realized covariance matrices. More specifically, two models are included in our baseline comparison, the Conditional Autoregressive Wishart (CAW) model of Golosnoy, Gribisch, and Liesenfeld (2012) and the Conditional Autoregressive Wishart Dynamic Conditional Correlation (CAW-DCC) specification of Bauwens, Storti, and Violante (2012).

5.6.1 Models and Setup

To tackle the questions above, we apply the baseline diagonal VARMA(2,1) specification for the MBC-RCov and BC-DC model and apply a grid of fixed values for the transformation parameter which seem relevant for a specific comparison, e.g., $\delta_1 = \ldots = \delta_k \in \{-0.1, -0.05, 0, 0.5, 1\}$. The other model parameters are re-estimated for each estimation window.

For the Wishart models used as benchmarks, the distributional assumption is

$$X_t|\mathscr{I}_{t-1} \sim W_n(\nu, S_t/\nu), \tag{5.19}$$

where W_n denotes the central Wishart density, ν is the scalar degrees of freedom parameter and S_t/ν is a $(k \times k)$ positive definite scale matrix, which is related to the conditional mean of X_t by $E[X_t|\mathscr{I}_{t-1}] = S_t$. The baseline CAW(p,q) model of Golosnoy, Gribisch, and Liesenfeld (2012) specifies the conditional mean as

$$S_t = CC' + \sum_{j=1}^{p} B_j S_{t-j} B_j' + \sum_{j=1}^{q} A_j X_{t-j} A_j', \tag{5.20}$$

C, B_j and A_j denoting $(k \times k)$ parameter matrices, while the CAW-DCC model of Bauwens, Storti, and Violante (2012) employs a decomposition

$$S_t = H_t P_t H_t', \tag{5.21}$$

where H_t is diagonal and P_t is a well-defined correlation matrix. As a sparse and simple DCC benchmark we apply univariate realized GARCH(p_v,q_v) specifications for the realized variances

$$H_{ii,t}^2 = c_i + \sum_{j=1}^{p_v} b_{i,j}^v H_{ii,t-j}^2 + \sum_{j=1}^{q_v} a_{i,j}^v X_{ii,t-j}, \tag{5.22}$$

along with the 'scalar Re-DCC' model (Bauwens, Storti, and Violante, 2012) for the

realized correlation matrix R_t,

$$P_t = \bar{P} + \sum_{j=1}^{p_c} b_j^c P_{t-j} + \sum_{j=1}^{q_c} a_j^c R_{t-j}. \tag{5.23}$$

The diagonal CAW(p,q) and the CAW-DCC(p,q) specification with $p = p_v = p_c = 2$ and $q = q_v = q_c = 1$ are selected since they are similar in complexity to the diagonal VARMA(2,1) model and provide a reasonable in-sample fit among various order choices.

For a given forecasting method, the evaluation is carried out as follows: We split the available data in a sample $X_1,\ldots,X_{1508}$ which is used only for estimation and an evaluation sample $X_{1509},\ldots,X_{2156}$. For each $T' \in [1508; 2156-h]$, the model is estimated using a rolling sample $X_{T'-1507},\ldots,X_{T'}$ of 1508 observations and forecasts of $X_{T'+h}$, $h = 1,5,10,20$, are computed. For the transformation-based forecasts, we consider both the naïve forecast $\widetilde{X}_{T'+h|T'}$, based on (5.14), and the bias-corrected forecast $\widehat{X}_{T'+h|T'}$, see (5.15), using $R = 1000$ simulated realizations. In addition, density forecasts of the returns $r_{T'+h}$ given $\mathscr{I}_{T'}$ are computed using the same simulated trajectories as for the bias-corrected covariance forecasts.

As outlined in section 5.4.1, in the simulations we discard trajectories where the transformation from $y_t^{(i)}$ to $X_t^{(i)}$ is not well-defined. Additionally, to attenuate the effect of extreme outliers in the simulated paths, we replace an element of the bias-corrected covariance matrix forecast by the uncorrected forecast if the fraction between the two exceeds 5 in absolute value. Such a procedure reflects a practically feasible plausibility check. Both modifications of the forecasts are needed only for small values of the transformation parameters $\delta \leq -0.25$ in our study which are anyway inconsistent with the empirical interval estimates presented above.

We compare the forecasting models by presenting average losses, i.e. risks, over the evaluation period. To gain insights about statistical significance of the differences, model confidence sets (MCS) are constructed following Hansen, Lunde, and Nason (2011) using the MulCom package (Hansen and Lunde, 2010) in Ox Console Version 6.21. The Max-t statistic is bootstrapped with a block lengths of $d = \max\{5,h\}$ and 10000 iterations. A confidence level of 90% is used throughout.

5.6.2 Baseline Results

We begin with an evaluation of the forecasting performance using a simple *squared prediction error* loss function, evaluated using the true realized covariance matrices. For $h = 1,5,10,20$ and $T' = 1508,\ldots,2156-h$, we compute the period loss as the Frobenius

Table 5.4: Fraction of mean Frobenius loss (5.24) between bias-corrected (5.15) vs. naive (5.14) forecasts. The forecasts are from the diagonal VARMA(2,1) model (5.18) based on the MBC transform (5.3) and the BC-DC transform (5.7).

	$h=1$		$h=5$		$h=10$		$h=20$	
δ	MBC	BC-DC	MBC	BC-DC	MBC	BC-DC	MBC	BC-DC
-0.5	1.8150	4.0653	2.2496	2.1426	1.6510	1.7354	1.6164	1.3337
-0.1	0.8758	0.9068	0.8750	0.9083	0.8882	0.9120	0.8922	0.9043
-0.05	0.8847	0.9177	0.8920	0.9192	0.9031	0.9216	0.9052	0.9133
0	0.9037	0.9326	0.9098	0.9304	0.9172	0.9306	0.9166	0.9212
0.05	0.9286	0.9513	0.9317	0.9464	0.9357	0.9445	0.9336	0.9345
0.1	0.9375	0.9576	0.9393	0.9518	0.9414	0.9491	0.9378	0.9393
0.5	0.9956	0.9925	0.9893	0.9866	0.9853	0.9826	0.9792	0.9765
1	1.0011	1.3840	1.0021	1.3087	1.0026	1.2046	1.0029	1.1479

norm of the forecast error

$$LF^{\{s\}}_{T',h} = \sum_{i=1}^{k}\sum_{j=1}^{k}\left(X_{ij,T'+h} - X^{\{s\}}_{ij,T'+h|T'}\right)^2, \tag{5.24}$$

where $X^{\{s\}}_{T'+h|T'}$ is a covariance matrix forecast obtained from one of the different methods.

To assess the value of the bias-correction we compare the risk

$$\overline{LF}^{\{s\}}_{h} = \frac{1}{2156-h-1507}\sum_{T'=1508}^{2156-h} LF^{\{s\}}_{T',h}$$

of the corrected forecasts to the naïve ones by calculating the fraction of the two for the different models and transformation parameters. The results are given in Table 5.4. Despite the adjustment of miss-behaved bias-corrected forecasts outlined in section 5.6.1, the simulation-based forecasts are worse than the uncorrected ones for $\delta \leq -0.25$. This is most pronounced for the MBC-RCov model and for short horizons. In such cases, the normality assumption provides a poor description of the transformed variables and hence the simulated $y^{(i)}_t$ series do not produce well-behaved re-transformed forecasts.

The valuation of bias correction changes fundamentally when $\delta \geq -0.1$ is considered. This is the empirically relevant span as the estimates of section 5.5 suggest. The simulation-based procedure leads to marked improvements of the forecasts. The reduction in risks for the MBC-RCov model is as high as 12% for $h = 1$ and $\delta = -0.1$;

Table 5.5: Risks from VARMA(2,1) forecasts based on the MBC transform (5.3) and the BC-DC transform (5.7), as well as CAW(2,1) and CAW-DCC(2,1) benchmarks (models (5.20) and (5.21)-(5.23), respectively). Left: Frobenius loss function (5.24) from bias-corrected forecasts. Right: Negative logarithmic score (5.25) of density forecasts. Asterisks denote the 90% model confidence set for a given loss function and horizon h. The best-performing forecast is in boldface. Models for which at least one of the forecasts is not positive definite have missing predictive densities (—).

	Frobenius Loss				Predictive Density			
	$h=1$	$h=5$	$h=10$	$h=20$	$h=1$	$h=5$	$h=10$	$h=20$
MBC(-0.1)	92.45	138.98	176.74	211.33*	8.155	8.276*	8.356*	8.438*
MBC(-0.05)	92.53	139.75	177.42	212.45*	8.152	8.274*	8.351*	8.431*
MBC(0)	93.17	140.34	177.68	212.98*	**8.146***	**8.268***	**8.348***	**8.428***
MBC(0.5)	86.58*	141.55	179.39	217.35*	8.166	8.313	8.414	8.514
MBC(1)	89.50*	169.19	212.93	246.54*	—	—	—	—
BC-DC(-0.1)	84.86*	**133.85***	**170.98***	**206.42***	8.160	8.276*	8.355*	8.432*
BC-DC(-0.05)	**84.73***	134.01*	171.47*	207.29*	8.158	8.277*	8.359*	8.439*
BC-DC(0)	85.09*	134.43*	172.03*	208.07*	8.153	8.269*	8.352*	8.431*
BC-DC(0.5)	85.43*	142.84	181.56	218.82*	8.200	8.350	8.450	8.585
BC-DC(1)	122.39	218.40	258.76	284.48	—	—	—	—
Cholesky	86.20*	141.42	178.03	215.75*	8.179	8.318	8.415	8.519
CAW(2,1)	85.97*	137.34*	176.50	214.60*	8.187	8.318	8.410	8.503
CAW-DCC(2,1)	86.32*	137.77*	178.11	217.52*	8.302	8.415	8.497	8.616

it gradually reduces as δ approaches one. There, the MBC-transform corresponds to the raw covariance series and bias-correction is not needed. This broad picture is reflected also by the BC-DC transformed series, where $\delta = 0.5$, corresponding to a model of realized standard deviations, is minimally prone to bias. For $\delta = 1$, when untransformed realized variances are approximated by a Gaussian process in the simulations, the latter fail to reduce bias, and even devastate the forecast accuracy.

Having shown the usefulness of bias correction, we now turn to a comparison of bias-corrected forecasts of the proposed models along with the CAW and CAW-DCC specification for which a correction is not needed. The corresponding Frobenius risks are shown in the left part of Table 5.5. The boldface numbers, which indicate the best-performing model for each horizon, show that the BC-DC specification with $\delta \in \{-0.1, -0.05\}$ emerges as favorable for all horizons.

The asterisks indicate models contained in the 90% model confidence set for a given horizon h. The MBC-RCov forecasts are contained only for a few specific values of δ.

For one-step forecasts, only the matrix square root transformed ($\delta = 0.5$) and untransformed ($\delta = 1$) forecasts cannot be rejected. For some other horizons, no MBC-RCov specification ($h = 5$) or only those with negative δ ($h = 20$) resist a rejection. Neither the matrix logarithm ($\delta = 0$) nor the semiparametric estimates of δ provide a reasonable performance which is robust with respect to the chosen horizon. The BC-DC model forecasts are elements of the MCS for a wide range of transformation parameters including the estimates from section 5.5 as well as $\delta = 0$. This holds for all considered forecast horizons. A comparison to the Wishart models reveals that despite their larger risk, both models cannot be rejected by the MCS approach except for the two-weeks horizon $h = 10$.

These conclusions about superiority change when the density forecasts are considered. We evaluate the forecasts by the logarithmic scoring rule; see, e.g., Gneiting, Stanberry, Grimit, Held, and Johnson (2008),

$$LD_{T',h}^{\{s\}} = -\log f_r^{\{s\}}(r_{T'+h}|\mathscr{I}_{T'}), \tag{5.25}$$

computing negative logarithms of the density forecast $f_r^{\{s\}}$, evaluated at the realized daily returns h periods ahead. The logarithmic rule is a strictly proper scoring rule, rewarding careful and honest assessments (Gneiting and Raftery, 2007). It is local in the sense that no point of $f_r^{\{s\}}$ other than the realized return is evaluated, which is also an intuitively and computationally appealing property.

The results are given in the right part of Table 5.5. Here, the MBC-RCov with $\delta \approx 0$ outperforms the BC-DC model; significantly for $h = 1$ as the MCS consists of only one specification there. For larger horizons, the differences are less pronounced and both models with transformation parameters close to zero seem appropriate.

To understand this outcome, note that the whole density of $X_{T'+h}$ given $X_{T'}$ is involved in computing the conditional return density forecasts. The matrix logarithmic model stands out from its competitors in yielding relatively close-to-Gaussian residuals, as has been seen in section 5.3. Since we use the truncated normal distribution in our simulation of $X_{T'+h}^{(i)}$, the different model rankings with respect to covariance forecasts and return density forecasts can be understood in light of this finding.

Also the Wishart models are rejected using the log density metric. We regard this as evidence that the Gaussianity assumption for transformed realized covariances provides a better approximation than the Wishart specification — at least when it comes to forecasting the return distribution.

To conclude the baseline results, the BC-DC model with small negative δ outper-

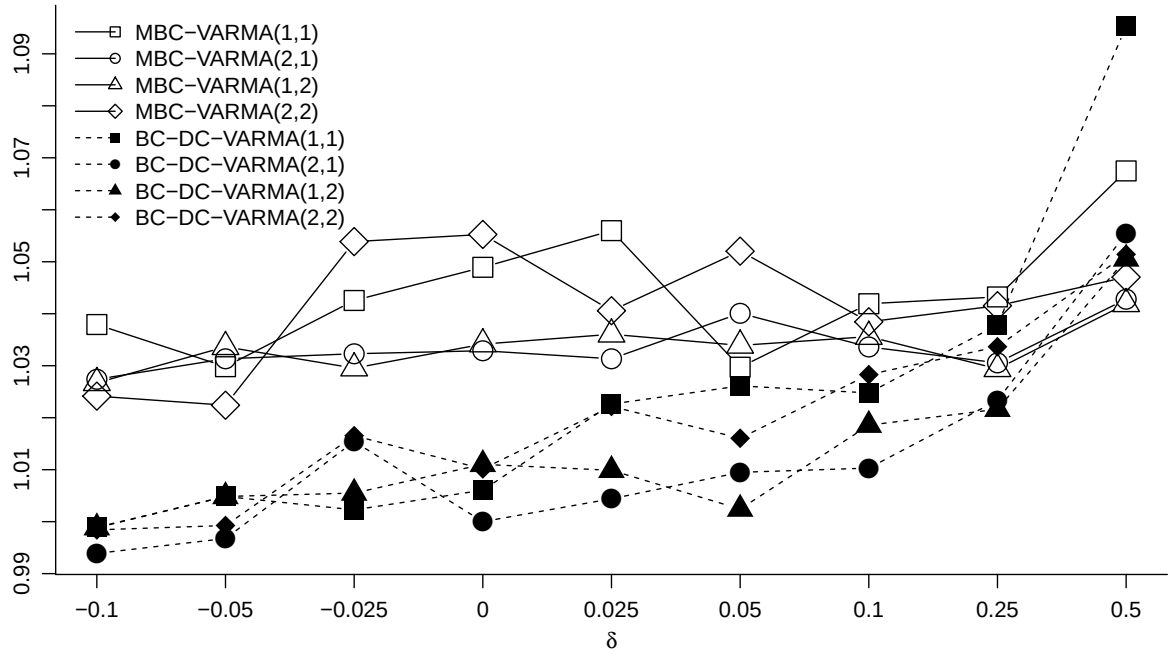

Figure 5.2: Robustness of out-of-sample results with respect to the order specification of the VARMA model (5.18). For $h = 10$, the Frobenius loss (5.24) is plotted as a fraction of the loss for the BC-DC-VARMA(2,1) specification with $\delta = 0$.

forms in terms of covariance matrix forecasting, while the MBC-RCov model with $\delta \approx 0$ emerges when the aim is forecasting the return density. The Wishart models are outperformed significantly in the latter case.

5.6.3 Robustness Regarding Model Specification

Up to now, the results for MBC-RCov and BC-DC forecasts are based on a simple diagonal VARMA(2,1) specification. We check whether our conclusions with regard to the data transformations remain intact for other models which have been used for volatility dynamics.

We first assess whether the choice of VARMA order matters for our conclusion. To this end, we compute forecasts and risks also for other specifications. The result for the Frobenius loss and $h = 10$ is exemplarily shown in Figure 5.2 and mirrors the result for other horizons. To make the figures comparable, here and henceforth, the risks are plotted as a fraction of a common benchmark, the BC-DC model with VARMA(2,1) dynamics and $\delta = 0$. It turns out that the VARMA(2,1) specification is among the best choices for most of the different transformation parameters. Importantly, the conclusion about favorable transforms does not interfere with the choice of model orders.

Daily financial volatility is often associated with a long memory behaviour, so we also include such models to our robustness checks. As a first alternative, we consider the heterogeneous autoregressive model of Corsi (2009). Lags of $y_{it}(\delta)$, averaged over 1, 5 and 20 trading days in the process

$$y_{it}(\delta) = c_i + \alpha_{i1} y_{i,t-1}(\delta) + \alpha_{i5}\Big(\frac{1}{5}\sum_{j=1}^{5} y_{i,t-j}(\delta)\Big) + \alpha_{i20}\Big(\frac{1}{20}\sum_{j=1}^{20} y_{i,t-j}(\delta)\Big) + u_{it} \tag{5.26}$$

introduce long-memory-like persistence. The parameters are estimated by least squares. In contrast, Chiriac and Voev (2011) use a flexible fractionally integrated vector ARMA (VARFIMA) specification with "real" long memory behavior. We also follow their approach but do not restrict the memory parameter to be the same across series, and hence estimate series-specific parameters $\theta_i = (d_i, \alpha_{i1}, \ldots, \alpha_{ip}, \phi_{i1}, \ldots, \phi_{iq}, \mu_i)$ of

$$(1 - \alpha_{i1}L - \ldots - \alpha_{ip}L^p)(1-L)^{d_i}(y_{it}(\delta) - \mu_i) = (1 + \phi_{i1}L + \ldots + \phi_{iq}L^q)u_{it}. \tag{5.27}$$

Again, correlation between the series is introduced only through the noise covariance matrix Σ_u. Like Chiriac and Voev (2011) we set $p = q = 1$ which gives the same model complexity as in our benchmark VARMA(2,1).

Overall, the VARFIMA setup provides smaller forecast errors than the VARMA benchmark, while the HAR is outperformed by both competitors; see the results in Figure 5.3. The excellent results for the ARFIMA model are in line with the results of Chiriac and Voev (2011). Further gains may be attainable by considering more sophisticated models, e.g., taking possible dynamic spillovers, factor structures and structural breaks into account. While a comprehensive comparison of different dynamic specifications is beyond the scope of this paper, we direct attention to the relative benefits of the various transformations for a given model. The relative rankings remain remarkably unchanged, independently of the dynamic specification. Again, the BC-DC model with small negative δ stands out.

Lastly, we compare our transformation-based approach to other models of the conditional Wishart family. To this end, we conduct a comparison of several diagonal CAW(p,q) models and CAW-DCC(p,q) models with different orders p and q. Additionally, the component models proposed by Jin and Maheu (2013) are considered. Regarding the latter, we estimate a Wishart Additive Component (CAW-ACOMP) model. The distribu-

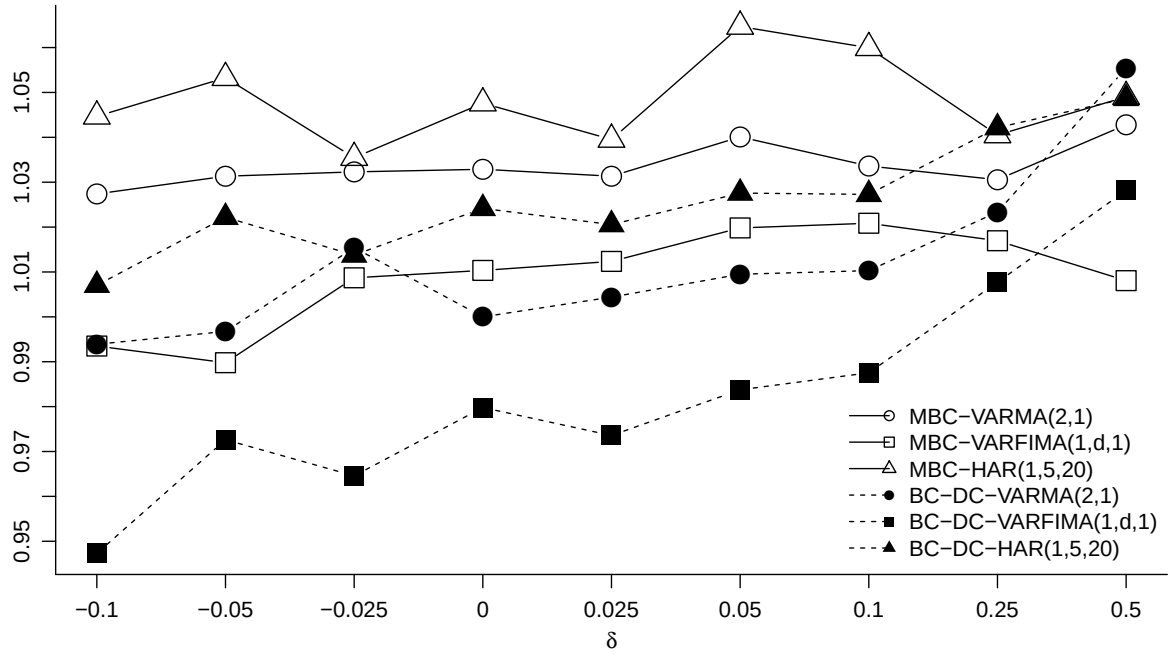

Figure 5.3: Robustness of out-of-sample results with respect to specification of dynamic persistence. The baseline VARMA(2,1) model (5.18) is compared to the VARFIMA model (5.27) and the HAR model (5.26). For $h = 10$, the Frobenius loss (5.24) is plotted as a fraction of the loss for the BC-DC VARMA(2,1) model with $\delta = 0$.

tional assumption (5.19) is complemented by

$$S_t = CC' + \sum_{j=1}^{K} B_j \odot \Gamma_{t,l_j}, \qquad B_j = b_j b_j', \qquad \Gamma_{t,l} = \frac{1}{l} \sum_{i=0}^{l-1} X_{t-i}, \tag{5.28}$$

where $\odot$ denotes the elementwise (Hadamard) product and, analogously to the HAR model, past averages of the covariances enter the conditional mean equation in a linear manner. Similarly, such lower frequency components are also involved in the Wishart Multiplicative Component (CAW-MCOMP) model

$$S_t = \left[\prod_{j=K}^{1} \Gamma_{t,l_j}^{\frac{\gamma_j}{2}} \right] CC' \left[\prod_{j=1}^{K} \Gamma_{t,l_j}^{\frac{\gamma_j}{2}} \right], \tag{5.29}$$

which we also assess in our study. As for the HAR model, we set $K = 3$ and average over $l_1 = 1$, $l_2 = 5$ and $l_3 = 20$ past observations. In accordance with all other Wishart models considered so far, the parameters are estimated by maximum likelihood for all rolling samples.

The results which are shown in Figure 5.4 for the Frobenius loss are clear-cut. None of

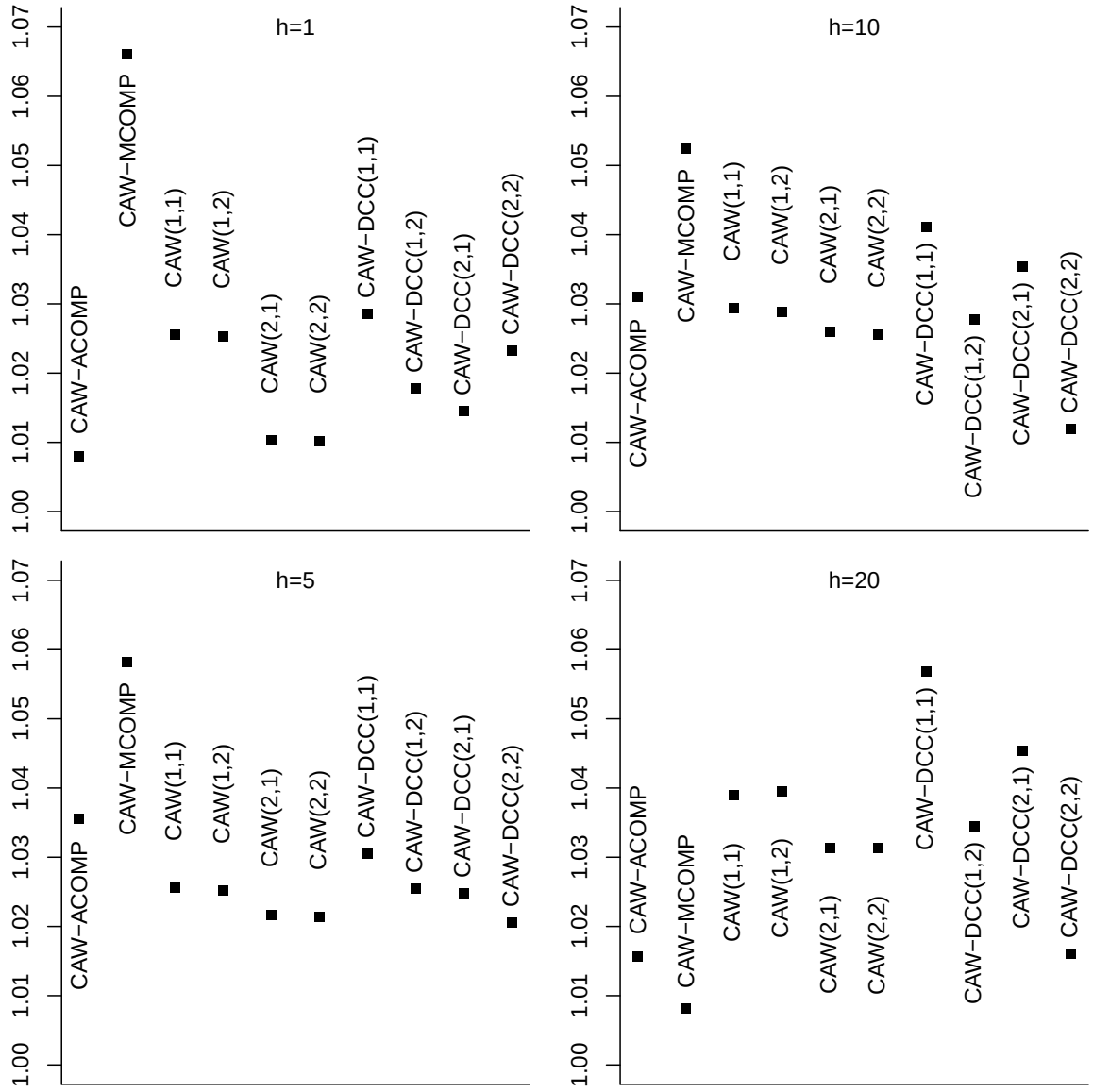

Figure 5.4: Robustness of out-of-sample results with respect to order specification of the CAW model (5.20) and the CAW-DCC model (5.21)-(5.23) and to the CAW-ACOMP model (5.28) and the CAW-MCOMP model (5.29). For each horizon, the Frobenius loss (5.24) is plotted as a fraction of the loss for the BC-DC-VARMA(2,1) model with $\delta = 0$.

the models outperforms the BC-DC benchmark, irrespective of the forecasting horizon. This is indicated by the relative risks that are above one for all models. In contrast, the ranking among the Wishart models varies with the horizon. At least for the smaller horizons, the chosen benchmark orders $p = 2$ and $q = 1$ correspond to well-performing models. The component models show an ambiguous figure. The additive model does well for most horizons, but the multiplicative approach is worthwhile only for rather long-term forecasts ($h = 20$).

Overall, the robustness checks find that the results of the baseline setup remain qualitatively unchanged also when other dynamic models, both for the Box-Cox models and for the Wishart family, are taken into account. Among the considered alternatives, long memory BC-DC models are the most relevant direction for improvements.

5.6.4 Robustness Regarding Loss Function

So far we have focussed on the Frobenius norm of the forecast error when evaluating the covariance matrix forecasts. In the matrix case, there are several other loss functions which may be appropriate for different practical forecasting situations; see, e.g., Laurent, Rombouts, and Violante (2013) for a discussion. In our out-of-sample study, we additionally consider the Stein distance

$$LS_{T',h} = \operatorname{tr}\left[X_{T'+h|T'}^{-1} X_{T'+h}\right] - \log\left|X_{T'+h|T'}^{-1} X_{T'+h}\right| - k, \tag{5.30}$$

and the asymmetric loss

$$L3_{T',h} = \frac{1}{6}\operatorname{tr}\left[X_{T'+h|T'}^{3} - X_{T'+h}^{3}\right] - \frac{1}{2}\operatorname{tr}\left[X_{T'+h|T'}^{2}(X_{T'+h} - X_{T'+h|T'})\right], \tag{5.31}$$

which is used by Laurent, Rombouts, and Violante (2011). Forecast comparisons based on LF, LS and $L3$ may differ because LS penalizes underpredictions more heavily than LF while overpredictions are more influential with the $L3$ loss, see Laurent, Rombouts, and Violante (2011), section 2.3. Additionally, the loss functions differ in their relative importance of high versus low volatility periods since only the Stein distance is homogeneous of order 0 and hence scale invariant.

The results of the evaluation with the Frobenius norm is replicated using both the LS and the $L3$ norms. Again, as Table 5.6 reveals, the BC-DC models perform better than their MBC-RCov counterparts and their forecasting superiority is not rejected with zero or small negative transformation parameter. The CAW models are statistically rejected in some cases, even if the power of the MCS procedure appears small for the L3 loss.

Table 5.6: Risks from bias-corrected VARMA(2,1) forecasts based on the MBC transform (5.3) and the BC-DC transform (5.7), as well as CAW(2,1) and CAW-DCC(2,1) benchmarks (models (5.20) and (5.21)-(5.23), respectively). Left: Stein loss function (5.30). Right: L3 loss function (5.31). Asterisks denote the 90% model confidence set for a given loss function and horizon h. The best-performing forecast is in boldface. The case $\delta = 1$ is missing since there at least one of the forecasts is not positive-definite.

	Stein Loss				L3 Loss			
	$h=1$	$h=5$	$h=10$	$h=20$	$h=1$	$h=5$	$h=10$	$h=20$
MBC(-0.1)	0.997*	1.416*	1.720*	2.118*	201.0*	254.9*	293.9*	325.0*
MBC(-0.05)	0.997*	1.420	1.731*	2.142*	197.2*	254.4*	293.4*	324.3*
MBC(0)	0.996*	1.420	1.733*	2.149*	197.1*	254.9*	293.4*	324.4*
MBC(0.5)	1.076	1.573	1.936	2.430	183.0*	256.5*	294.5*	328.1*
BC-DC(-0.1)	1.000*	1.410*	1.697*	**2.074***	182.5*	249.6*	**288.3***	322.8*
BC-DC(-0.05)	0.996*	**1.404***	**1.696***	2.090*	181.9*	**249.5***	288.7*	**322.7***
BC-DC(0)	**0.995***	1.405*	1.698*	2.098*	182.3*	249.8*	289.0*	323.1*
BC-DC(0.5)	1.052	1.527	1.891	2.414	**181.7***	258.6	296.8*	327.9*
Cholesky	1.099	1.581	1.945	2.440	181.7*	256.6*	293.3*	327.3*
CAW(2,1)	1.025	1.436	1.740*	2.158*	184.3*	261.2	298.0*	332.6*
CAW-DCC(2,1)	1.004*	1.409*	1.712*	2.170*	186.6*	262.7	298.6	333.1*

A reasonable loss function may also be chosen to involve the economic cost of prediction errors. A risk-averse investor may be interested in the variance of an ex-ante minimum-variance portfolio (MVP) which is computed from the covariance matrix forecast. Using the realized variance of the MVP as the ex-post loss, we therefore consider

$$LMV_{T',h} = w' X_{T'+h} w, \qquad \text{where} \qquad w = (\iota' X_{T'+h|T'} \iota)^{-1} X_{T'+h|T'} \iota, \tag{5.32}$$

where $\iota = (1,\dots,1)'$. Alternatively, the squared daily return $w' r_{T'+h} r'_{T'+h} w$ of the ex-ante MVP is used instead of the realized variance.

Table 5.7 shows rather inconclusive results. The discriminating power is weak for these two losses, so that many models are included in the model confidence sets. Notably, however, BC-DC with $\delta = 0$ outperforms in three out of eight horizon-loss combinations and is always included in the MCS.

To summarize, the forecasting results are unchanged if other dynamic models are considered and reveal relatively little ambiguity also with alternative loss functions. Overall, the Box-Cox dynamic correlation specification with log variances ($\delta = 0$) seems to be a good and robust choice in practice. It is close to the best performing model

Table 5.7: Risks from bias-corrected VARMA(2,1) forecasts based on the MBC transform (5.3) and the BC-DC transform (5.7), as well as CAW(2,1) and CAW-DCC(2,1) benchmarks (models (5.20) and (5.21)-(5.23), respectively). Left: Realized variance of minimum variance portfolio (5.32). Right: Squared daily return of minimum variance portfolio as defined below (5.32). Asterisks denote the 90% model confidence set for a given loss function and horizon h. The best-performing forecast is in boldface. The case $\delta = 1$ is missing since there at least one of the forecasts is not positive-definite.

	Realized variance of MVP				Squared daily return of MVP			
	$h = 1$	$h = 5$	$h = 10$	$h = 20$	$h = 1$	$h = 5$	$h = 10$	$h = 20$
MBC(-0.1)	0.7924	0.8036*	0.8107*	0.8253*	0.6414*	0.6550*	0.6703*	0.6906
MBC(-0.05)	0.7915*	0.8027*	0.8099*	0.8245*	0.6409*	0.6542*	0.6694*	0.6887
MBC(0)	**0.7912***	**0.8022***	0.8092*	0.8233*	0.6409*	0.6538*	0.6685*	0.6870*
MBC(0.5)	0.7920*	0.8030*	0.8119	0.8267*	0.6425*	0.6543*	0.6692*	0.6905
BC-DC(-0.1)	0.7915*	0.8025*	0.8089*	0.8230*	0.6396*	0.6520*	0.6663*	0.6842*
BC-DC(-0.05)	0.7919*	0.8030*	0.8098*	0.8243*	0.6399*	0.6520*	0.6667*	0.6856*
BC-DC(0)	0.7916*	0.8025*	**0.8088***	**0.8224***	0.6397*	0.6514*	0.6656*	**0.6840***
BC-DC(0.5)	0.7928	0.8034*	0.8121	0.8272*	0.6434*	0.6534*	0.6702*	0.6913
Cholesky	0.7921*	0.8026*	0.8109*	0.8249*	0.6409*	0.6535*	0.6665*	0.6896
CAW(2,1)	0.7930	0.8038*	0.8110	0.8241*	**0.6385***	0.6514*	**0.6630***	0.6891
CAW-DCC(2,1)	0.7928	0.8039*	0.8101*	0.8231*	0.6443*	**0.6496***	0.6632*	0.6869*

for most horizons and with regard to many of the evaluation criteria. The MBC-RCov model, however, has a superior forecasting performance for specific criteria such as predictive densities. Further research appears fruitful to further clarify these facts, e.g., in light of datasets for different asset classes.

5.7 Conclusion

We have proposed two new approaches to multivariate realized volatility modeling and applied them to US stock market data. The empirical results, including an out-of-sample forecasting comparison, seem promising, also in comparison to the main competitors, the conditional autoregressive Wishart model of Golosnoy, Gribisch, and Liesenfeld (2012) and several variants thereof. Our assessment of various transformation parameters supports a convenient special case of our Box-Cox approaches: the use of standard linear time series models to a multivariate time series of log realized variances and z-transformed correlations. Its appropriateness can be easily checked for a specific dataset using the inferential methods introduced in this paper.

The present study leaves significant questions for further research. With a focus on forecasting, investigating more advanced dynamic specifications appears worthwhile, possibly including dynamic spillovers and structural changes along with the long memory dynamics briefly considered in this paper. Additionally, in applications to data sets of higher dimensions, our approach allows the assessment of cross-sectional properties such as factor structures in a methodologically and computationally straightforward setup. In addition to the realized volatility setup with utilization of intraday data, our models are also relevant for the study of multivariate stochastic volatility based on a latent covariance matrix specification.

Appendix 5.A Maximum Likelihood Estimation

This appendix describes maximum likelihood estimation of the MBC-RCov model as conducted in section 5.5 and shown in the lower panel of Table 5.1. Although the transformed series cannot be exactly Gaussian due to the bounded support, we use

$$u_t \sim NID(0, \Sigma_u)$$

as an approximating auxiliary assumption for parameter estimation, alongside a dynamic model specification (5.5), the VARMA model (5.18) in our case. Under this assumption, the conditional distribution of $y_t(\delta)$ is also Gaussian $N(\mu_t, \Sigma_u)$ with conditional mean $\mu_t(\theta)$ determined by the time series model. Denoting, as in section 5.3, the vector of untransformed variances and covariances by $\tilde{x}_t := \text{vech}(X_t)$, the MBC transformation in vech-space as $y_t(\delta) = \varphi(\tilde{x}_t; \delta)$ and the Jacobi matrix as $\tilde{J}_t(\delta)$ (see (5.11)), the joint density of $\tilde{x}_t$ is given by

$$f_x(\tilde{x}_t; \delta, \Sigma_u, \theta) = |\tilde{J}_t(\delta)||2\pi\Sigma_u|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\varphi(\tilde{x}_t;\delta) - \mu_t(\theta))'\Sigma_u(\varphi(\tilde{x}_t;\delta) - \mu_t(\theta))\right\},$$

see, e.g., Härdle and Simar (2007, section 3.7).

The log-likelihood, conditional on pre-sample values $x_0, x_{-1}, \ldots$, is hence

$$l(\delta, \Sigma_u, \theta) = \sum_{t=1}^{T} \log|\tilde{J}_t(\delta)| - \frac{T}{2}\log|\Sigma_u| - \frac{1}{2}\sum_{t=1}^{T}(\varphi(\tilde{x}_t;\delta) - \mu_t(\theta))'\Sigma_u^{-1}(\varphi(\tilde{x}_t;\delta) - \mu_t(\theta)). \quad (5.33)$$

Given δ and θ, the unrestricted maximum likelihood estimator of Σ_u is computed by $\Sigma_u(\theta, \delta) = \frac{1}{T}\sum_{t=1}^{T}(\varphi(\tilde{x}_t;\delta) - \mu_t(\theta))(\varphi(\tilde{x}_t;\delta) - \mu_t(\theta))'$ which can be plugged into (5.33) to ob-

tain the concentrated likelihood

$$l^c(\delta,\theta) = \sum_{t=1}^{T} \log|\tilde{J}_t(\delta)| - \frac{T}{2}\log\left|\frac{1}{T}\sum_{t=1}^{T}(\varphi(\tilde{x}_t;\delta)-\mu_t(\theta))(\varphi(\tilde{x}_t;\delta)-\mu_t(\theta))'\right|. \tag{5.34}$$

In practice, a further concentration step seems worthwhile. Compute, for a given parameter vector δ, the maximum likelihood estimator for the dynamic parameters θ. The term $\log|\tilde{J}_t(\delta)|$ does not affect this optimization so that computation of the Jacobian can be suppressed. The concentrated likelihood (5.34) is then maximized with respect to δ only, with the Jacobian numerically computed at each likelihood evaluation.

Acknowledgments

This research has partly been done at the Institute of Economics and Econometrics of the University of Regensburg. The author is very grateful for helpful comments by Rolf Tschernig, Enzo Weber and participants of the Humboldt–Copenhagen Conference on Financial Econometrics 2013. Support by BayEFG is gratefully acknowledged.

Bibliography

AMEMIYA, T., AND J. L. POWELL (1981): "A Comparison of the Box-Cox Maximum Likelihood Estimator and the Non-linear Two-stage Least Squares Estimator," *Journal of Econometrics*, 17(3), 351–381.

ANDERSEN, T. G., T. BOLLERSLEV, X. DIEBOLD, FRANCIS, AND P. LABYS (2003): "Modelling and Forecasting Realized Volatility," *Econometrica*, 71(2), 579–625.

ANDERSON, T. W., AND H. RUBIN (1956): "Statistical Inference in Factor Analysis," in *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Vol. V*, ed. by J. Neyman. University of California Press.

ASAI, M., AND M. MCALEER (2014): "Forecasting Co-Volatilities via Factor Models with Asymmetry and Long Memory in Realized Covariance," Tinbergen Institute Discussion Paper 2014-037.

AVARUCCI, M., AND C. VELASCO (2009): "A Wald Test for the Cointegration Rank in Nonstationary Fractional Systems," *Journal of Econometrics*, 151(2), 178 – 189.

BAI, J., AND S. NG (2008): "Large Dimensional Factor Analysis," *Foundations and Trends in Econometrics*, 3(2), 89–163.

BAILLIE, R. T. (1996): "Long Memory Processes and Fractional Integration in Econometrics," *Journal of Econometrics*, 73(1), 5–59.

BANBURA, M., AND M. MODUGNO (2012): "Maximum Likelihood Estimation of Factor Models on Datasets with Arbitrary Pattern of Missing Data," *Journal of Applied Econometrics*, 29(1), 133–160.

BARNDORFF-NIELSEN, O. E., AND G. SCHOU (1973): "On the Parametrization of Autoregressive Models by Partial Autocorrelations," *Journal of Multivariate Analysis*, 3(4), 408–419.

BARNDORFF-NIELSEN, O. E., AND N. SHEPHARD (2004): "Econometric Analysis of Realized Covariation: High Frequency Based Covariance, Regression, and Correlation in Financial Economics," *Econometrica*, 72(3), 885–925.

BAUER, D., AND M. WAGNER (2012): "A State Space Canonical Form for Unit Root Processes," *Econometric Theory*, 28(6), 1313–1349.

BAUER, G. H., AND K. VORKINK (2011): "Forecasting Multivariate Realized Stock Market Volatility," *Journal of Econometrics*, 160(1), 93–101.

BAUWENS, L., G. STORTI, AND F. VIOLANTE (2012): "Dynamic Conditional Correlation Models for Realized Covariance Matrices," CORE Discussion Paper 2012-60.

BAYOUMI, T., AND B. EICHENGREEN (1994): "One Money or Many? Analyzing the Prospects for Monetary Unification in Various Parts of the World," *Princeton Studies in International Finance*, 76, 1–44.

BELTRATTI, A., AND C. MORANA (2006): "Breaks and Persistency: Macroeconomic Causes of Stock Market Volatility," *Journal of Econometrics*, 131(1-2), 151–177.

BLANCHARD, O., AND D. QUAH (1989): "The Dynamic Effects of Aggregate Demand and Supply Disturbances," *The American Economic Review*, 79(4), 655–673.

BOLLERSLEV, T., R. F. ENGLE, AND J. M. WOOLDRIDGE (1988): "A Capital Asset Pricing Model with Time-Varying Covariances," *Journal of Political Economy*, 96(1), 116–131.

BOX, G. E. P., AND D. R. COX (1964): "An Analysis of Transformations," *Journal of the Royal Statistical Society. Series B*, 26(2), 211–252.

BOX, G. E. P., AND G. M. JENKINS (1970): *Time Series Analysis: Forecasting and Control*. Holden Day.

BROCKWELL, A. E. (2007): "Likelihood-based Analysis of a Class of Generalized Long-Memory Time Series Models," *Journal of Time Series Analysis*, 28(3), 386–407.

CAPORALE, G., AND L. GIL-ALANA (2013): "Long Memory in US Real Output per Capita," *Empirical Economics*, 44(2), 591–611.

CAPORALE, G. M., AND L. A. GIL-ALANA (2002): "Unemployment and Input Prices: A Fractional Cointegration Approach," *Applied Economics Letters*, 9(6), 347–351.

CAPORALE, G. M., AND L. A. GIL-ALANA (2011): "Fractional Integration and Impulse Responses: A Bivariate Application to Real Output in the USA and Four Scandinavian Countries," *Journal of Applied Statistics*, 38(1), 71–85.

CASELLA, G., AND R. L. BERGER (2002): *Statistical Inference*. Thomson Learning.

CAVANAUGH, J. E., AND R. H. SHUMWAY (1996): "On Computing the Expected Fisher Information Matrix for State-space Model Parameters," *Statistics & Probability Letters*, 26(4), 347–355.

CHAN, N. H., AND W. PALMA (1998): "State Space Modeling of Long-Memory Processes," *The Annals of Statistics*, 26(2), 719–740.

CHANG, Y., B. JIANG, AND J. Y. PARK (2012): "Using Kalman Filter to Extract and Test for Common Stochastic Trends," unpublished manuscript.

CHANG, Y., J. I. MILLER, AND J. Y. PARK (2009): "Extracting a Common Stochastic Trend: Theory with Some Applications," *Journal of Econometrics*, 150(2), 231 – 247.

CHEN, W. W., AND C. M. HURVICH (2006): "Semiparametric Estimation of Fractional Cointegrating Subspaces," *The Annals of Statistics*, 34(6), 2939–2979.

CHEUNG, Y.-W., AND K. S. LAI (1993): "A Fractional Cointegration Analysis of Purchasing Power Parity," *Journal of Business & Economic Statistics*, 11(1), 103–112.

CHIRIAC, R., AND V. VOEV (2011): "Modelling and Forecasting Multivariate Realized Volatility," *Journal of Applied Econometrics*, 26(6), 922–947.

CHRISTIANO, L. J., M. EICHENBAUM, AND R. VIGFUSSON (2003): "What Happens After a Technology Shock?," NBER Working Paper 9819.

——— (2007): "Assessing Structural VARs," *NBER Macroeconomics Annual 2006*, pp. 1–72.

CHUNG, C. (2001): "Calculating and Analyzing Impulse Responses for the Vector ARFIMA Model.," *Economics Letters*, 71(1), 17 – 25.

CORSI, F. (2009): "A Simple Approximate Long Memory Model of Realized Volatility," *Journal of Financial Econometrics*, 7(2), 174–195.

DIEBOLD, F. X., AND G. D. RUDEBUSCH (1989): "Long Memory and Persistence in Aggregate Output," *Journal of Monetary Economics*, 24(2), 189–209.

DO, H. X., R. D. BROOKS, AND S. TREEPONGKARUNA (2013): "Generalized Impulse Response Analysis in a Fractionally Integrated Vector Autoregressive Model," *Economics Letters*, 118, 462–465.

DOZ, C., D. GIANNONE, AND L. REICHLIN (2012): "A Quasi–Maximum Likelihood Approach for Large, Approximate Dynamic Factor Models," *The Review of Economics and Statistics*, 94(4), 1014–1024.

DRAPER, N. R., AND D. R. COX (1969): "On Distributions and Their Transformation to Normality," *Journal of the Royal Statistical Society. Series B*, 31(3), 472–476.

DUEKER, M., AND R. STARTZ (1998): "Maximum-Likelihood Estimation of Fractional Cointegration with an Application to U.S. and Canadian Bond Rates," *The Review of Economics and Statistics*, 80(3), 420–426.

DURBIN, J., AND S. J. KOOPMAN (2012): *Time Series Analysis by State Space Methods: Second Edition*. Oxford Statistical Science Series.

ENGLE, R. (2002): "Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models," *Journal of Business & Economic Statistics*, 20(3), 339–350.

ENGLE, R., AND G. LEE (1999): "A Permanent and Transitory Component Model of Stock Return Volatility," in *Cointegration, Causality, and Forecasting: A Festschrift in Honor of Clive W.J. Granger*, ed. by R. Engle, and H. White. Oxford University Press.

ENGLE, R. F., AND C. W. J. GRANGER (1987): "Co-Integration and Error Correction: Representation, Estimation, and Testing," *Econometrica*, 55(2), 251–276.

FRANCIS, N., M. T. OWYANG, J. E. ROUSH, AND R. DICECIO (2014): "A Flexible Finite-Horizon Alternative to Long-run Restrictions with an Application to Technology Shocks," *The Review of Economics and Statistics*, forthcoming.

GALI, J. (1999): "Technology, Employment, and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations?," *American Economic Review*, 89(1), 249–271.

GEWEKE, J. (1977): "The Dynamic Factor Analysis of Economic Time Series," in *Latent Variables in Socio-Economic Models*, ed. by D. J. Aigner, and A. S. Goldberger. North-Holland.

GIL-ALANA, L. (2011): "Inflation in South Africa. A Long Memory Approach," *Economics Letters*, 111(3), 207–209.

GIL-ALANA, L., AND J. HUALDE (2008): "Fractional Integration and Cointegration: An Overview and an Empirical Application," in *Palgrave Handbook of Econometrics, Vol. II*, ed. by K. Patterson, and T. Mills. Palgrave, MacMillan.

GIL-ALANA, L., AND A. MORENO (2009): "Technology Shocks and Hours Worked: A Fractional Integration Perspective.," *Macroeconomic Dynamics*, 13(5), 580 – 604.

GNEITING, T., AND A. E. RAFTERY (2007): "Strictly Proper Scoring Rules, Prediction, and Estimation," *Journal of the American Statistical Association*, 102, 359–378.

GNEITING, T., L. I. STANBERRY, E. P. GRIMIT, L. HELD, AND N. A. JOHNSON (2008): "Assessing Probabilistic Forecasts of Multivariate Quantities, with an Application to Ensemble Predictions of Surface Winds," *TEST*, 17(2), 211–235.

GOLOSNOY, V., B. GRIBISCH, AND R. LIESENFELD (2012): "The Conditional Autoregressive Wishart Model for Multivariate Stock Market Volatility," *Journal of Econometrics*, 167(1), 211–223.

GOLOSNOY, V., AND H. HERWARTZ (2012): "Dynamic Modeling of High-Dimensional Correlation Matrices in Finance," *International Journal of Theoretical and Applied Finance*, 15(5), 1–22.

GONCALVES, S., AND N. MEDDAHI (2011): "Box-Cox Transforms for Realized Volatility," *Journal of Econometrics*, 160(1), 129–144.

GOSPODINOV, N. (2010): "Inference in Nearly Nonstationary SVAR Models with Long-Run Identifying Restrictions," *Journal of Business & Economic Statistics*, 28(1), 1–12.

GOSPODINOV, N., A. MAYNARD, AND H. PESAVENTO (2011): "Sensitivity of Impulse Responses to Small Low-Frequency Comovements: Reconciling the Evidence on the Effects of Technology Shocks," *Journal of Business & Economic Statistics*, 29(4), 455–467.

GRANGER, C. (1983): "Co-Integrated Variables and Error-Correcting Models," UCSD Discussion Paper 83-13.

GRANGER, C. W., AND A. A. WEISS (1983): "Time Series Analysis of Error-correction Models," in *Studies in Econometrics, Time Series, and Multivariate Statistics*. Academic Press New York.

GRANGER, C. W. J. (1986): "Developments in the Study of Cointegrated Economic Variables," *Oxford Bulletin of Economics and Statistics*, 48(3), 213–228.

GRANGER, C. W. J., AND T. H. LEE (1989): "Multicointegration," in *Advances in Econometrics: Cointegration, Spurious Regressions, and Unit Roots*, ed. by G. F. Rhodes, and T. B. Fomby. JAI Press.

GRASSI, S., AND P. S. DE MAGISTRIS (2012): "When Long Memory Meets the Kalman Filter: A Comparative Study," *Computational Statistics & Data Analysis*, in press.

GRIBISCH, B. (2013): "A Latent Dynamic Factor Approach to Forecasting Multivariate Stock Market Volatility," in *Beiträge zur Jahrestagung des Vereins für Socialpolitik 2013: Wettbewerbspolitik und Regulierung in einer globalen Wirtschaftsordnung - Session: Volatility, F01-V2*.

HALBLEIB, R., AND V. VOEV (2011): "Forecasting Covariance Matrices: A Mixed Frequency Approach," Working Papers ECARES 2010-002.

HAN, A. K. (1987): "A Non-parametric Analysis of Transformations," *Journal of Econometrics*, 35, 191–209.

HANSEN, P. R., AND A. LUNDE (2010): *MulCom 2.00 - Econometric Toolkit for Multiple Comparisons*.

HANSEN, P. R., A. LUNDE, AND J. NASON (2011): "The Model Confidence Set," *Econometrica*, 79(2), 453–497.

HÄRDLE, W., AND L. SIMAR (2007): *Applied Multivariate Statistical Analysis*. Springer.

HARVEY, A., E. RUIZ, AND N. SHEPHARD (1994): "Multivariate Stochastic Variance Models," *The Review of Economic Studies*, 61(2), 247–264.

HARVEY, A. C. (1991): *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge Books. Cambridge University Press.

HAUTSCH, N., L. M. KYJ, AND P. MALEC (2014): "Do High-Frequency Data Improve High-Dimensional Portfolio Allocations?," *Journal of Applied Econometrics*, forthcoming.

HEATON, C., AND V. SOLO (2004): "Identification of Causal Factor Models of Stationary Time Series," *Econometrics Journal*, 7(2), 618–627.

HIGGINS, M. L., AND A. BERA (1992): "A Class of Nonlinear ARCH Models," *International Economic Review*, 33, 137–158.

HSU, N.-J., AND F. J. BREIDT (2003): "Bayesian Analysis of Fractionally Integrated ARMA with Additive Noise," *Journal of Forecasting*, 22(6-7), 491–514.

HSU, N.-J., B. K. RAY, AND F. J. BREIDT (1998): "Bayesian Estimation of Common Long-range Dependent Models," in *Probability Theory and Mathematical Statistics: Proceedings of the Seventh Vilnius Conference*, ed. by B. Grigelionis, J. Kubilius, V. Paulauskas, V. Statulevicius, and H. Pragarauskas. VSP.

HUALDE, J. (2009): "Consistent Estimation of Cointegrating Subspaces," Universidad Pública de Navarra. Preprint.

HUALDE, J., AND P. ROBINSON (2010): "Semiparametric Inference in Multivariate Fractionally Cointegrated Systems," *Journal of Econometrics*, 157(2), 492–511.

JIN, X., AND J. M. MAHEU (2013): "Modelling Realized Covariances and Returns," *Journal of Financial Econometrics*, 11(2), 335–369.

JOHANSEN, S. (2008): "A Representation Theory for a Class of Vector Autoregressive Models for Fractional Processes," *Econometric Theory*, 24(3), 651–676.

JOHANSEN, S., AND M. Ø. NIELSEN (2012): "Likelihood Inference for a Fractionally Cointegrated Vector Autoregressive Model," *Econometrica*, 80(6), 2667–2732.

JONES, R. H. (1976): "Estimation of the Innovation Generalized Variance of a Multivariate Stationary Time Series," *Journal of the American Statistical Association*, 71(354), 386–388.

JUNGBACKER, B., AND S. J. KOOPMAN (2014): "Likelihood-based Dynamic Factor Analysis for Measurement and Forecasting," *The Econometrics Journal*, forthcoming.

KOKOSZKA, P. S., AND M. S. TAQQU (1995): "Fractional ARIMA with Stable Innovations," *Stochastic Processes and their Applications*, 60, 19–47.

KOOPMAN, S. J. (1997): "Exact Initial Kalman Filtering and Smoothing for Nonstationary Time Series Models," *Journal of the American Statistical Association*, 92(440), 1630–1638.

KOOPMAN, S. J., AND N. SHEPHARD (1992): "Exact Score for Time Series Models in State Space Form," *Biometrika*, 79(4), 823–826.

LAM, C., AND Q. YAO (2012): "Factor Modeling for High-Dimensional Time Series: Inference for the Number of Factors," *The Annals of Statistics*, 40(2), 694–726.

LAM, C., Q. YAO, AND N. BATHIA (2011): "Estimation of Latent Factors for High-dimensional Time Series," *Biometrika*, 98(4), 901–918.

ŁASAK, K. (2010): "Likelihood Based Testing for No Fractional Cointegration," *Journal of Econometrics*, 158(1), 67–77.

ŁASAK, K., AND C. VELASCO (2014): "Fractional Cointegration Rank Estimation," Tinbergen Institute Discussion Paper 2014-021.

LAURENT, S., J. V. K. ROMBOUTS, AND F. VIOLANTE (2011): "On the Forecasting Accuracy of Multivariate GARCH Models," *Journal of Applied Econometrics*, 27(6), 934–955.

——— (2013): "On Loss Functions and Ranking Forecasting Performances of Multivariate Volatility Models," *Journal of Econometrics*, 173(1), 1–10.

LIU, Q. (2009): "On Portfolio Optimization: How and When Do We Benefit from High-frequency Data?," *Journal of Applied Econometrics*, 24(4), 560–582.

LOBATO, I. N. (1997): "Consistency of the Averaged Cross-periodogram in Long Memory Series," *Journal of Time Series Analysis*, 18(2), 137–155.

LOVCHA, J. (2009): "Hours Worked - Productivity Puzzle: A Seasonal Fractional Integration Approach," unpublished manuscript.

LÜTKEPOHL, H. (1984): "Linear Transformations of Vector ARMA Processes," *Journal of Econometrics*, 26(3), 283 – 293.

LÜTKEPOHL, H. (1996): *Handbook of Matrices*. Wiley & Sons.

LÜTKEPOHL, H. (2005): *New Introduction to Multiple Time Series Analysis*. Springer.

LUCIANI, M., AND D. VEREDAS (2012): "A Model for Vast Panels of Volatilities," Documentos de Trabajo 1230, Banco de Espana.

MATTESON, D. S., AND R. S. TSAY (2011): "Dynamic Orthogonal Components for Multivariate Time Series," *Journal of the American Statistical Association*, 106(496), 1450–1463.

MENG, X.-L., AND D. B. RUBIN (1993): "Maximum Likelihood Estimation via the ECM Algorithm: A General Framework," *Biometrika*, 80(2), 267–278.

MESTERS, G., S. KOOPMAN, AND M. OOMS (2011): "Monte Carlo Maximum Likelihood Estimation for Generalized Long-Memory Time Series Models," Tinbergen Institute Discussion Paper 2011-090.

MOHANTY, R., AND M. POURAHMADI (1996): "Estimation of the Generalized Prediction Error Variance of a Multiple Time Series," *Journal of the American Statistical Association*, 91(433), 294–299.

MORANA, C. (2004): "Frequency Domain Principal Components Estimation of Fractionally Cointegrated Processes," *Applied Economics Letters*, 11(13), 837–842.

——— (2006): "A Small Scale Macroeconometric Model for the Euro-12 Area," *Economic Modelling*, 23(3), 391–426.

——— (2007): "Multivariate Modelling of Long Memory Processes with Common Components," *Computational Statistics & Data Analysis*, 52(2), 919 – 934.

NIELSEN, M. Ø. (2004): "Efficient Inference in Multivariate Fractionally Integrated Time Series Models," *Econometrics Journal*, 7(1), 63–97.

——— (2007): "Local Whittle Analysis of Stationary Fractional Cointegration and the Implied–Realized Volatility Relation," *Journal of Business & Economic Statistics*, 25(4), 427–446.

——— (2010): "Nonparametric Cointegration Analysis of Fractional Systems with Unknown Integration Orders," *Journal of Econometrics*, 155(2), 170–187.

NIELSEN, M. Ø., AND K. SHIMOTSU (2007): "Determining the Cointegrating Rank in Nonstationary Fractional Systems by the Exact Local Whittle Approach," *Journal of Econometrics*, 141(2), 574 – 596.

NOURELDIN, D., N. SHEPHARD, AND K. SHEPPARD (2012): "Multivariate High-frequency-based Volatility (HEAVY) Models," *Journal of Applied Econometrics*, 27(6), 907–933.

PALMA, W. (2007): *Long-Memory Time Series: Theory and Methods*. Wiley.

PAN, J., AND Q. YAO (2008): "Modelling Multiple Time Series via Common Factors," *Biometrika*, 95(2), 365–379.

PHILLIPS, P. C. B. (1991): "Optimal Inference in Cointegrated Systems," *Econometrica*, 59(2), 283–306.

PRIESTLEY, M. B. (1982): *Spectral Analysis and Time Series*. Academic Press.

PROIETTI, T., AND H. LÜTKEPOHL (2013): "Does the Box–Cox Transformation Help in Forecasting Macroeconomic Time Series?," *International Journal of Forecasting*, 29(1), 88–99.

QUAH, D., AND T. J. SARGENT (1993): "A Dynamic Index Model for Large Cross Sections," in *Business Cycles, Indicators and Forecasting*, ed. by J. H. Stock, and M. W. Watson. University of Chicago Press.

QUAH, D., AND S. VAHEY (1995): "Measuring Core Inflation," *The Economic Journal*, 105(432), 1130–1144.

R CORE TEAM (2013): *R: A Language and Environment for Statistical Computing*.

RAY, B. K., AND R. S. TSAY (2000): "Long-range Dependence in Daily Stock Volatilities," *Journal of Business & Economic Statistics*, 18(2), 254–262.

ROBINSON, P. (2005): "The Distance Between Rival Nonstationary Fractional Processes," *Journal of Econometrics*, 128(2), 283–300.

ROBINSON, P., AND D. MARINUCCI (2001): "Narrow-band Analysis of Nonstationary Processes," *Annals of Statistics*, 29(4), 947–986.

ROBINSON, P. M., AND J. HUALDE (2003): "Cointegration in Fractional Systems with Unknown Integration Orders," *Econometrica*, 71(6), 1727–1766.

ROBINSON, P. M., AND Y. YAJIMA (2002): "Determination of Cointegrating Rank in Fractional Systems," *Journal of Econometrics*, 106(2), 217 – 241.

SCHOTMAN, P. C., R. TSCHERNIG, AND J. BUDEK (2008): "Long Memory and the Term Structure of Risk," *Journal of Financial Econometrics*, 2(1), 1–37.

SHIMOTSU, K. (2010): "Exact Local Whittle Estimation of Fractional Integration with Unknown Mean and Time Trend," *Econometric Theory*, 26(2), 501–540.

SHIMOTSU, K., AND P. C. B. PHILLIPS (2005): "Exact Local Whittle Estimation of Fractional Integration," *The Annals of Statistics*, 33(4), 1890–1933.

SHUMWAY, R. H., AND D. S. STOFFER (1982): "An Approach to Time Series Smoothing and Forecasting Using the EM Algorithm," *Journal of Time Series Analysis*, 3(4), 253–264.

SIMS, C. A. (1980): "Macroeconomics and Reality," *Econometrica*, 48(1), 1–48.

SOWELL, F. (1989): "Maximum Likelihood Estimation of Fractionally Integrated Time Series Models," Discussion paper, Carnegie Mellon University.

STOCK, J. H., AND M. W. WATSON (1993): "A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems," *Econometrica*, 61(4), 783–820.

SUN, Y., AND P. C. B. PHILLIPS (2004): "Understanding the Fisher Equation," *Journal of Applied Econometrics*, 19(7), 869–886.

TSCHERNIG, R., E. WEBER, AND R. WEIGAND (2013a): "Fractionally Integrated VAR Models with a Fractional Lag Operator and Deterministic Trends: Finite Sample Identification and Two-step Estimation," University of Regensburg Working Papers in Business, Economics and Management Information System 471.

——— (2013b): "Long- versus Medium-run Identification in Fractionally Integrated VAR Models," University of Regensburg Working Papers in Business, Economics and Management Information Systems 476.

——— (2013c): "Long-Run Identification in a Fractionally Integrated System," *Journal of Business & Economic Statistics*, 31(4), 438–450.

——— (2014): "Long- versus Medium-run Identification in Fractionally Integrated VAR Models," *Economics Letters*, 122(2), 299–302.

UHLIG, H. (2004): "Do Technology Shocks Lead to a Fall in Total Hours Worked?," *Journal of the European Economic Association*, 2(2-3), 361–371.

VEENSTRA, J. Q. (2012): "Persistence and Anti-persistence: Theory and Software," Ph.D. thesis, Western University.

VELILLA, S. (1993): "A Note on the Multivariate Box-Cox Transformation to Normality," *Statistics and Probability Letters*, 17, 259–263.

WATSON, M. W., AND R. F. ENGLE (1983): "Alternative Algorithms for the Estimation of Dynamic Factor, MIMIC and Varying Coefficient Regression Models," *Journal of Econometrics*, 23(3), 385–400.

WEIGAND, R. (2014): "Matrix Box-Cox Models for Multivariate Realized Volatility," University of Regensburg Working Papers in Business, Economics and Management Information Systems 478.

WHITE, H. (1982): "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50(1), 1–26.

WU, L. S.-Y., J. S. PAI, AND J. HOSKING (1996): "An Algorithm for Estimating Parameters of State-space Models," *Statistics & Probability Letters*, 28(2), 99 – 106.

YU, J., Z. YANG, AND X. ZHANG (2006): "A Class of Nonlinear Stochastic Volatility Models and Its Implications for Pricing Currency Options," *Computational Statistics and Data Analysis*, 51, 2218–2231.

ZHANG, X., AND M. L. KING (2008): "Box-Cox Stochastic Volatility Models with Heavy-tails and Correlated Errors," *Journal of Empirical Finance*, 15(3), 549–566.