Mathematisches Forschungsinstitut Oberwolfach

Report No. 05/2010

DOI: 10.4171/OWR/2010/05

Mini-Workshop: Semiparametric Modelling of Multivariate Economic Time Series With Changing Dynamics

Organised by
Luc Bauwens (Louvain-la-Neuve)
Qiwei Yao (London)
Rainer von Sachs (Louvain-la-Neuve)

January 17th – January 23rd, 2010

Abstract. Modelling multivariate time series of possibly high dimension calls for appropriate dimension-reduction, e.g. by some factor modelling, additive modelling, or some simplified parametric structure for the dynamics (i.e. the serial dependence) of the time series. This workshop aimed to bring together experts in this field in order to discuss recent methodology for multivariate time series dynamics which are changing over time: by an abrupt switch between two (or more) different regimes or rather smoothly evolving over time. The emphasis has been on mathematical methods for semiparametric modelling and estimation, where "semiparametric" is to be understood in a rather broad sense: parametric models where the parameters are themselves nonparametric functions (of time), regime-switching nonparametric models with a parametric specification of the transition mechanism, and alike. An ultimate goal of these models to be applied to economic and financial time series is prediction. Another emphasis has been on comparing Bayesian with frequentist approaches, and to cover both theoretical aspects of estimation, such as consistency and efficiency, and computational aspects.

Mathematics Subject Classification (2000): 62M10, 62M20, 62H12, 62P20 - IMU 10.

Introduction by the Organisers

Over the past 20 years statisticians have contributed by developing methodology to address the curse of dimensionality when modelling high-dimensional multivariate economic and financial data such that subsequent estimation remains possible. This ranges from parsimonious parametric models, over additive non-parametric

modelling to the quite general factor model approach based e.g. on principal components regression. Common to all of these approaches is the goal to describe the common structure of a panel of several tenths of time series by a "simple" model, either in a low-dimensional parametric space or a low-dimensional space of "common components" which up to some idiosynchratic behaviour specific to each time series dimension describe the co-movement over time of the whole panel. This latter factor approach has found lots of applications, not only in the context of macro-economic or financial data. As soon as the data become serially correlated, the afore-mentioned approaches need to be dynamic: (vector-) autoregressive models, either in the (conditional) mean or in the (conditional) variance structure of the data (leading to so called MGARCH models), either parametric or non-parametric (i.e. non-linear AR-models), dynamic (instead of static) factor models, conditional correlation models, and alike. Treating these dynamic models from the point of view of deriving theoretical properties of the accompanying estimation methods, such as consistency, asymptotic normality, efficiency, and to construct reliable prediction methods (intervals) for the future evolution of these time series, calls for more refined mathematical-statistical skills, and it is primarily towards the community of researchers with expertise in this field that the scope of our workshop is addressed. Based on recent empirical evidence, most of these multivariate time series cannot necessarily be considered to have a homogeneous dynamical structure over time: the impact of political and financial crises, changing monetary policies of central banks, and alike, suggests to refine the afore-mentioned models to allow for inhomogeneity, i.e. changing dynamics. Subsequently, the proposed estimation methods need to be refined, and quite naturally, constructing appropriate predictors becomes more challenging. The goal of this workshop attended by 15 researchers from 10 different countries has been to compare a variety of different approaches to model and treat the kind of inhomogeneity described above. These approaches to inhomogeneous time series modelling could be called "semi-parametric", e.g. via a parametrisation locally in time as a stationary processes. Examples for that are (vector-) autoregressive or MGARCH-type models where the parameters become now non-parametric functions of time. Those can be of low regularity (to include abrupt changes over time) or of higher smoothness over time, in order to model a slow evolution of the dynamics. Another instance of semiparametric modelling arises for a different approach to inhomogeneity: in regime-switching models, the underlying stochastic process is modelled as changing from one regime of stationarity to another via a random mechanism (e.g. an underlying Markov chain), i.e. the existence of latent (unobservable) state variables that control in which regime the mean and the variance-covariance structure (conditional or unconditional) are to be found. Modelling the transition of these (nonparametrically modelled) states by Bayesian or frequentist approaches, e.g. via a parametric matrix of transition probabilities, was one emphasis of our discussions, addressing among others questions of statistical inference and subsequent prediction, but also computational issues such as MCMC (Markov Chain Monte Carlo) or EM-algorithms. Summarizing the workshop gave us the opportunity

to enjoy lively discussions on the question how models for high-dimensional time series with changing dynamics can be made on the one hand sufficiently parsimonious (in "parameters" through either latent variables or functional parameters whose number is considerably smaller than the number of modelled time series) but on the other hand sufficiently flexible to capture the inhomogeneity of the time series.

Mini-Workshop: Semiparametric Modelling of Multivariate Economic Time Series With Changing Dynamics

Table of Contents

Abstracts

Dynamic Factor Models for Forecasting and Structural Identification

Matteo Barigozzi
(joint work with Lucia Alessi, Marco Capasso)

We consider new empirical applications of factor models, based on recent methodological advances in forecasting and structural analysis. The main idea underlying factor analysis is that a large set of variables can be explained by a small number of latent variables, the *factors*, which are responsible for all the relevant dynamics. Given a sequence (indexed by n) of nested vector stochastic processes $\{\mathbf{y}_{nt}, t \in \mathbb{Z}\}$ the dynamic factor represenation is

$$\mathbf{y}_{nt} = \mathbf{x}_{nt} + \mathbf{z}_{nt} = \mathbf{B}_n(L)\mathbf{u}_t + \mathbf{z}_{nt}, \quad t \in \mathbb{Z}. \tag{1}$$

The process $\mathbf{y}_{nt}$ is decomposed into two components: a *common* one $\mathbf{x}_{nt}$ that contains the information carried by the factors $\mathbf{u}_t$, and an *idiosyncratic* one $\mathbf{z}_{nt}$ which is simply the residual of the decomposition. Furthermore, the relation between the common part of the observable series and the factors is assumed to be linear. This decomposition is accomplished by analysing the spectral density matrix of the observable variables. If these are all driven by a common set of say q factors with $q < n$, then the q largest eigenvalues of the spectral density matrix of $\mathbf{y}_{nt}$ contain all the essential information and diverge as n diverges, while the remaining $n - q$ eigenvalues remain bounded. Factor analysis is therefore a technique of dimension reduction that takes the information contained in a large dataset and summarizes it by means of few unobservable common variables.

Starting from the most general case of Dynamic Factor Model presented in [7], and [6], we consider the restriced case presented in [5]. Indeed under the assumption of onesided finite lag loadings $\mathbf{B}_n(L)$ we can write (1) in state space form

$$\mathbf{y}_{nt} = \mathbf{A}_n\mathbf{F}_t + \mathbf{z}_{nt}, \quad (\mathbf{I} - \mathbf{C}(L))F_t = \mathbf{H}\mathbf{u}_t, \quad t \in \mathbb{Z}. \tag{2}$$

This is very similar to the models by [9], and [3].

We propose a modification of (2) for conditionally heteroskedastic series. In factor analysis nothing is usually said about the conditional covariance matrix of the data. We assume that the observed conditional heteroskedasticity observed for certain datasets is generated by conditionally heteroskedastic dynamic factors $\mathbf{u}_t$ that evolve according to a Multivariate GARCH. We call this new model Dynamic Factor-GARCH (DF-GARCH). Together with the assumptions of this model, we propose a three-steps estimator of the conditional covariance matrix which is consistent as both the cross-section dimension n and the sample size T diverge. We provide results from Monte Carlo simulations. We show forecast performance of the model when applied to volatilities and covolatilities of asset returns of the

London Stock Exchange. We then apply the model to a macroeconomic context, when forecasting US inflation and its conditional variance. Moreover, we are also able to get forecasts of conditional covariances between macroeconomic variables which may result valuable in monetary policy issues (see [2]).

Model (2) can be used also as a competitor of Structural VARs (SVARs) for economic analysis. Usually the literature faces the probelm of identification of the structural shocks hitting the economy in the context of SVAR. We show that typical small size reduced form VARs may suffer of the problem of nonfundamentalness, i.e. the structural shocks of the structural MA representation of the economy belong also the space spanned by future observations. This fact typically happens when the economic agents' information space is larger that the econometricians' one. In all these cases, the agent forms his future expectations by using the information supplied by additional variables, while SVARs make use only of a limited amount of information and clearly have information only about the past. Neglecting this problem may lead to wrong identification of structural shocks. This may be the case, for example, when we use SVARs for validating Dynamic Stochastic General Equilibrium models which are often used by Central Banks for policy making. We show how dynamic factor models in the state space representation (2) do not suffer of this problem and are therefore suitable for identification. Following [5], we compare the results obtained with a dynamic factor approach with the ones obtained from a New Keynesian model by [8]. Our results turn out to be robust with respect to the transformation used for hours per worker as the labour input variable. This is a very controversial issue as explained in [4]. As in [8], our method seems to hint towards a negative correlation between hours per worker and productivity growth rates, when conditioning on a technology shock while in [4] a positive conditional correlation is reported. This discrepancy between SVAR and factor models results, is imputable to thee presence of nonfundamental shocks in the bivariate system made of labour productivity and hours worked, which SVARs cannot identify (see [1]).

References

[1] L. Alessi, M. Barigozzi, and M. Capasso, *Nonfundamentalness in structural econometric models: A reiview*, Manuscript, available at http://www.barigozzi.eu/research.html, (2010).

[2] L. Alessi, M. Barigozzi, and M. Capasso, *Estimation and forecasting in large datasets with conditionally heteroskedastic dynamic common factors*, Manuscript, European Central Bank working paper 1115, (2009).

[3] J. Bai, *Inferential theory for factor models of large dimensions*, Econometrica **71** (2003), 135-171.

[4] L. J. Christiano, M. Eichenbaum, and R. Vigfusson, *The response of hours to a technology shock: evidence based on direct measures of technology*, Journal of the European Economic Association **2** (2004), 381-395.

[5] M. Forni, D. Giannone, M. Lippi, and L. Reichlin, *Opening the black box: structural factor models with large cross-sections*, Econometric Theory **25** (2009), 1319-1347.

[6] M. Forni, M. Hallin, M. Lippi, and L. Reichlin, , (2000), –.

[7] M. Forni and M. Lippi, *The generalized dynamic factor model: Representation theory*, Econometric Theory **17** (2001), 1113-1141.

[8] J. Gal, *Technology employment and the business cycle: Do technology shocks explain aggregate fluctuations?*, American Economic Review **89** (1999), 249-271.
[9] J. H. Stock and M. W. Watson, *Forecasting using principal components from a large number of predictors*, Journal of the Americal Statistical Association **97** (2002), 1167–1179.

Generalized Linear Dynamic Factor Models - An approach via Singular Autoregressions

Manfred Deistler
(joint work with Brian D.O. Anderson, Alexander Filler, Christiane Zinner, Weitian Chen)

Modeling and forecasting of high dimensional time series is plagued by the so called "curse of dimensionality". For instance in the unrestricted autoregressive case, for fixed maximum lag, the dimension of the parameter space is proportional to N^2 where N is the number of time series considered, whereas the number of data points, for a fixed sample size T, is linear in N. One way to overcome this "curse of dimensionality" is to use Generalized Dynamic Factor Models (GDFMs) which have been introduced in [2], [3] and in a slighty different form in [4] and [5]. The price to be paid for overcoming this curse of dimensionality is to require a certain kind of similarity or co-movement between the single time series.

The basic idea of GDFMs is that the N-dimensional observation at time t, y_t^N say, can be represented as

$$y_t^N = \widehat{y}_t^N + u_t^N, \tag{1}$$

where $(\widehat{y}_t^N)$ is the process of latent variables, which are strongly dependent in the cross-sectional dimension, and where (u_t^N) is the wide sense idiosyncratic noise, i.e. (u_t^N) is weakly dependent in the cross-sectional dimension. Under our assumptions the spectral densities corresponding to (1) are given by

$$f_y^N(\lambda) = f_{\widehat{y}}^N(\lambda) + f_u^N(\lambda). \tag{2}$$

We assume that N is tending to infinity, that $f_{\widehat{y}}^N$ is of rank q and that the latent variables can be represented as a state space system of minimal dimension n where both q and n are independent of N.

The core of the paper [1] is concerned with structure theory, in the sense that we want to obtain a model for the latent variables from the population spectral density of these variables rather than from data. Of course this is an idealized setting, however as we show, the results obtained for this idealized setting, are useful for estimation. Our structure analysis consists of the following steps:

- Factorization of the rational singular spectral density f_y^N. The corresponding stable and mini-phase factors are "tall" transfer functions and correspond to Wold decomposition. The corresponding white noise are dynamic factors.

- Extraction of static factors from the latent variables. Static factors are obtained by a static linear transformation of the latent variables and show the same dynamics as the latent variables. They are of smaller dimension than the latent variables and their dimension remains constant when N is going to infinity. For this reason we consider the modeling of static factors.
- Realization of the tall spectral factors of the spectral density of the static factors as state space or ARMA systems. Since the tall spectral factors can be shown to be generically (in a certain setting) zeroless, they can be shown to be realizable as an autoregressive system. This is a big advantage, since the estimation of autoregressive systems is much easier than estimation of state space or ARMA systems.
- Determination of the AR parameters by Yule-Walker equations. Since the AR systems here are typically singular (i.e. the innvoation variance is singular), the Yule-Walker equations may have multiple solutions. It is shown that the minimal norm solution always corresponds to a stable system.

If we commence from data, the process of static factors can be estimated by using a (static) PCA on the sample covariance matrix of the observations.

References

[1] B. D. O. Anderson and M. Deistler, *Generalized dynamic factor models-an approach via singular autoregressions*, European Journal of Control, (2010), forthcoming.
[2] M. Forni, M. Hallin, M. Lippi, and L. Reichlin, *The generalized dynamic factor model: identification and estimation*, The Review of Economics and Statistics **65** (2000), 453-473.
[3] M. Forni and M. Lippi, *The generalized dynamic factor model: representation theory*, Econometric Theory **17** (2001), 1113-1141.
[4] J. H. Stock and M. W. Watson, *Forecasting using principal components from a large number of predictors*, Journal of the Americal Statistical Association **97** (2002), 1167–1179.
[5] J. H. Stock and M. W. Watson, *Macroeconomic forecasting using diffusion indexes*, Journal of Business and Economic Statistics **20** (2002), 147-162.

Flexible Low Dimensional Dynamic Factor Models with Applications in Weather and Neuroeconomics

Song Song
(joint work with Wolfgang Haerdle, Yaacov Ritov)

Modeling for high-dimensional data is a challenging task in statistics especially when the data comes in a dynamic context and is observed at changing locations with different sample sizes. Such modeling challenges appear in many different fields. In agricultural economics and financial engineering, it is common to analyze the dynamics of temperatures and implied volatility surface for risk management. For functional magnetic resonance imaging data (fMRI), one may be interested in analyzing the brain's response over time to certain stimuli as well as identifying its activation area.

A successful modeling approach utilizes factor type models, which allow low-dimensional representation of the data. In an orthogonal L-factor model, a J-dimensional random vector $Y_t = (Y_{t,1}, \ldots, Y_{t,J})^\top$ can be represented as

$$Y_{t,j} = Z_{t,1} m_{1,j} + \cdots + Z_{t,L} m_{L,j} + \varepsilon_{t,j}, \tag{1}$$

where $Z_{t,l}$ are common factors, $\varepsilon_{t,j}$ are errors and the coefficients $m_{l,j}$ are factor loadings. In most applications, the index $t = 1, \ldots, T$ reflects the time evolution of the whole system, and Y_t can be considered as a multidimensional not necessarily stationary time series. The study of high-dimensional Y_t is then simplified to the modeling of $Z_t = (Z_{t,1}, \ldots, Z_{t,L})^\top$, which is a more feasible task when $L \ll J$. In a variety of applications, one has explanatory variable $X_{t,j} \in \mathbb{R}^d$ at hand that may influence the factor loadings m_l. An important refinement of the model (1) is to incorporate the existence of observable covariates $X_{t,j}$. The factor loadings are now generalized to functions of $X_{t,j}$, so that the model (1) is generalized to

$$\begin{aligned} Y_{t,j} &= \sum_{l=1}^{L} Z_{t,l}\ m_l(X_{t,j}) + \varepsilon_{t,j},\ \ 1 \le j \le J_t,\ 1 \le t \le T. \\ &\overset{\text{def}}{=} Z_t^\top m(X_{t,j}) + \varepsilon_{t,j} \end{aligned}$$

where $Z_t = (Z_{t,1}, \ldots, Z_{t,L})^\top$ is an unobservable L-dimensional process and m is an L-tuple $(m_1, \ldots, m_L)$ of unknown real-valued functions m_l defined on a subset of $\mathbb{R}^d$. The variables $X_{1,1}, \ldots, X_{T,J_T}$, $\varepsilon_{1,1}, \ldots, \varepsilon_{T,J_T}$ are independent. The errors $\varepsilon_{t,j}$ have zero means and finite second moments. For simplicity of notation, we will assume that the covariates $X_{t,j}$ have support $[0,1]^d$, and also that $J_t = J$ do not depend on t.

[1] consider an efficient nonparametric method of fitting the model (2) and call it dynamic semiparametric factor model (DSFM).

Although it displays very interesting performance in applications, it suffers several main disadvantages in practice. Firstly, for the estimation of m, they use a series estimator, e.g. take tensor B-spline basis as $\{\psi_k : [0,1]^d \to \mathbb{R}, 1 \le k \le K\}$ and approximate $m(x)$ by $A\psi(x)$, where $A = (\alpha_{l,k})$ is a $L \times K$ matrix and $\psi = (\psi_1, \ldots, \psi_K)^\top$. However, in practice, when $d > 3$, this usually leads to quite large K. Secondly, when periodic dynamics are present, which are very common in weather, fMRI and implied volatility modelings, VAR modeling the low-dimensional time series Z_t usually requires a high order process. Thirdly, the Newton-Raphson algorithm for the quadratic minimization they proposed actually actually can not be implemented in practice since they need invert a large matrix with dimension $L(T + K)$. Fourthly, a two step estimation procedure, i.e. estimating Z and m at first, and then using VAR process to fit $\hat{Z}$ leads to additional errors.

We will consider a model in which patterns in time and space are incorporated:

$$Y_t(x) = \sum_{l=1}^{L} Z_{t,l} m_l(x) + \varepsilon_t(x), \tag{2}$$

One of the main motivations of the second point comes from the special structure of the temperature data. They are the moving average (of 365 nearby days) temperatures of Germany from Sep 1st, 1957 to Jul 24th, 2006. It shows the special "large period" structure of the Germany temperatures besides the normal seasonal effect. On may find another application of the model (2) in the analysis of functional magnetic resonance imaging (fMRI) data. The fMRI is a noninvasive technique of recording brain's signals on spatial area in every particular time period (2.5 sec for our data set). One obtains a series of three-dimensional images of the blood-oxygen-level-dependent (BOLD) fMRI signals, when an exercised person is subject to certain stimuli (periodically). The main aims of the statistical methods in this field are identification of the brain's activation areas for risky decisions and analysis of its response over time. For this purpose the model (2) can be applied.

We observe $(X_{t,j}, Y_{t,j})$ for $j = 1, \ldots, J_t$ and $t = 1, \ldots, T$ such that

$$Y_{t,j} = \sum_{l=1}^{L}\sum_{r=1}^{R} u_r(t)\gamma_{rl}\sum_{k=1}^{K} a_{lk}\psi_k(X_{t,j}) + \varepsilon_{tj} \tag{3}$$

$$Y_t^\top = \underbrace{U_t^\top\Gamma}_{Z_t^\top}\underbrace{A\Psi_t}_{m} + \varepsilon_t \stackrel{\text{def}}{=} U_t^\top\beta^\top\Psi_t + \varepsilon_t. \tag{4}$$

Here $U_t^\top = (u_1(t), \ldots, u_R(t))$ is a $1\times R$ matrix with $u_r(t)$ as the pre-specified initial basis function in time, which we introduce to capture the global trend and periodic variations. $\Psi_t = (\psi_1(X_t), \ldots, \psi_K(X_t))^\top$ is a $K \times J$ matrix with ψ_k a basis of space functions. Γ, A and β are $R\times L$, $L\times K$ and $R\times K$ matrixes consisting of γ_{rl}, a_{lk} and β_{rk} respectively. For every β matrix, we introduce $\beta^r = (\beta_{kr}, 1 \le k \le K)^\top$, that is, the row vector formed by the coefficients corresponding to the r-th basis in time. Additionally we define $\|\beta\|_{2,1} = \sum_r \sqrt{\sum_k \beta_{rk}^2}$. Finally we set $R(\beta) = \{r : \beta^r \neq 0\}$ and $M(\beta) = |R(\beta)|$ where $|R|$ denotes the cardinality of set R.

The variables $X_{1,1}, \ldots, X_{T,J_T}$, $\varepsilon_{1,1}, \ldots, \varepsilon_{T,J_T}$ are independent. Throughout the paper we assume that the $X_{t,j}$ are deterministic. The errors $\varepsilon_{t,j}$ are *i.i.d.*, have zero means and finite second moments. For simplicity of notation, we will assume that the covariates $X_{t,j}$ have support $[0,1]^d$, and also that $J_t \equiv J$ do not depend on t unless otherwise specified.

Someone may criticize that there is no random effect in the time part. If we insert an unobservable L-dimensional random process $Z_{0,t}$ with $\mathsf{E}(Z_{0,t}|X_t) = 0$ into the time part, similar to equation (4) we get

$$Y_t^\top = (Z_{0,t}^\top + U_t^\top\Gamma)A\Psi_t + \varepsilon_t = U_t^\top\beta^\top\Psi_t + Z_{0,t}^\top A\Psi_t + \varepsilon_t \stackrel{\text{def}}{=} U_t^\top\beta^\top\Psi_t + \varepsilon_t^*.$$

with $\mathsf{E}(\varepsilon_t^*|X_t) = 0$, which indicates that this is indifferent from the original model (4). In essence, we do not assume that the behavior in time is stationary or Markovian in . We believe that the temporal and spacial behavior should be treated by similar smoothing tools.

Since not all initially included basis (in time) are significant, basis selection are necessary. A popular variable selection method is the l_1-norm penalized least squares estimation, which is originally from [2] and commonly referred to as the

lasso method. A natural extension of the lasso method, the so-called group Lasso is given in [3], in which the penalty term is a mixed $(2,1)$-norm of the coefficient matrix.

Under the additional Gaussian error assumption, we first show that this group Lasso-type estimator enjoys nice sparsity oracle inequalities and variable selection properties. In particular, we are able to essentially remove the effect of the number of initial basis functions in time in the bound. Finally, we show how our results can be extended to more general noise distributions, of which we only require the variance to be finite.

Our methods produce estimates of the true unobservable Z_t, say $\hat{Z}_t$, as well as estimates of the unknown basis in space m_l. In practice, one operates on $\hat{Z}_t$ for further inference of the data. The main question that arises from these application is whether the inference based on $\hat{Z}_t$ is equivalent to the one based on Z_t. Attempting to give an answer to this questions forms Theorem 3.3.

References

[1] Park, Byeong U., Mammen, Enno, Wolfgang, Härdle, and Borak, Szymon, *Time Series Modelling With Semiparametric Factor Dynamics*, Journal of the American Statistical Association **104** (2009), 284–298.

[2] Tibshirani, Robert, *Regression Shrinkage and Selection via the Lasso*, Journal of the Royal Statistical Society, Series B **58** (1996), 267–288.

[3] Ming Yuan and Yi Lin, *Model selection and estimation in regression with grouped variables*, Journal of the Royal Statistical Society, Series B **68** (2006), 49–67.

Forecasting Inflation Using Dynamic Model Averaging

Gary Koop

(joint work with Dimitris Korobilis)

Forecasting inflation is one of the more important, but difficult, exercises in macroeconomics. Many different approaches have been suggested. Perhaps the most popular are those based on extensions of the Phillips curve. This literature is voluminous, with a few representative and influential papers include [1], [2], [11], [21] and [20]. The details of these papers differ, but the general framework involves a dependent variable such as inflation (or the change in inflation) and explanatory variables including lags of inflation, the unemployment rate and other predictors. Recursive, regression-based methods, have had some success. However, three issues arise when using such methods.

First, the coefficients on the predictors can change over time. For instance, it is commonly thought that the slope of the Phillips curve has changed over time. If so, the coefficients on the predictors that determine this slope will be changing. More broadly, there is a large literature in macroeconomics which documents structural breaks and other sorts of parameter change in many time series variables (see, among many others, [22]). Recursive methods are poorly designed to capture such parameter change. It is better to build models designed to capture it.

Second, the number of potential predictors can be large. For instance, [11] consider ten predictors. Researchers working with factor models such as [21] typically have many more than this. The existence of so many predictors can result in a huge number of models. For instance, if the set of models is defined by whether each of m potential predictors is included or excluded, then the researcher has 2^m models. This raises substantive statistical problems for model selection strategies. In light of this, many authors have turned to Bayesian methods, either to do Bayesian model averaging (BMA) or to automate the model selection process. Examples in macroeconomics and finance include [3], [7] and [14]. Furthermore, computational demands can become daunting when the research is facing 2^m models.

Third, the model relevant for forecasting can potentially change over time. For instance, the set of predictors for inflation may have been different in the 1970s than now. Or some variables may predict well in recessions but not in expansions. Furthermore, papers such as [20] find that Phillips curve forecasts work well in some periods, but at other periods simpler univariate forecasting strategies work better. Such arguments suggest that the forecasting model is changing over time. This kind of issue further complicates an already difficult econometric exercise. That is, if the researcher has 2^m models and, at each point in time, a different forecasting model may apply, then the number of combinations of models which must be estimated in order to forecast at time τ is $2^{m\tau}$. Even in relatively simple forecasting exercises, it can be computationally infeasible to forecast by simply going through all of these $2^{m\tau}$ combinations. For this reason, to our knowledge, there is no literature on forecasting inflation with many predictors where the coefficients on those predictors may change over time and where a different forecasting model might hold at each point in time. A purpose of this paper is to fill this gap.

In this paper, we consider a strategy developed by [18] which they refer to as dynamic model averaging or DMA. Their approach can also be used for dynamic model selection or DMS where a single (potentially different) model can be used as the forecasting model at each point in time. DMA or DMS seem ideally suited for the problem of forecasting inflation since they allow for the forecasting model to change over time while, at the same time, allowing for coefficients in each model to evolve over time. They involve only standard econometric methods for state space models such as the Kalman filter but (via some empirically-sensible approximations) achieve vast gains in computational efficiency so as to allow DMA and DMS to be done in real time despite the computational problem described in the preceding paragraph.

We use these methods in the context of a forecasting exercise with quarterly US data from 1959Q1 through 2008Q2. We use two measures of inflation and fifteen predictors and compare the forecasting performance of DMA and DMS to a wide variety of alternative forecasting procedures. DMA and DMS indicate that the set of good predictors for inflation changes substantially over time. Due to this, we find DMA and DMS to forecast very well (in terms of forecasting metrics such

as log predictive likelihoods, MSFEs and MAFEs), in most cases leading to large improvements in forecast performance relative to alternative approaches.

References

[1] A. Ang, G. Bekaert and M. Wei, *Do macro variables, asset markets, or surveys forecast inflation better?*, Journal of Monetary Economics, **54** (2007), 1163–1212.

[2] A. Atkeson and L. Ohanian, *Are Phillips curves useful for forecasting inflation?*, Federal Reserve Bank of Minneapolis Quarterly Review, **25** (2001), 2–11.

[3] D. Avramov, *Stock return predictability and model uncertainty*, Journal of Financial Economics, **64** (2002), 423–458.

[4] R. Chen and J. Liu, *Mixture Kalman filters*, Journal of the Royal Statistical Society, Series B, **62** (2000), 493–508.

[5] T. Cogley and T. Sargent, *Drifts and volatilities: Monetary policies and outcomes in the post WWII U.S.*, Review of Economic Dynamics, **8** (2005), 262–302.

[6] T. Cogley, S. Morozov, and T. Sargent, *Bayesian fan charts for U.K. inflation: Forecasting and sources of uncertainty in an evolving monetary system*, Journal of Economic Dynamics and Control **29** (2005), 1893–1925.

[7] K. Cremers, *Stock return predictability: A Bayesian model selection perspective*, Review of Financial Studies **15** (2002), 1223–1249.

[8] S. Fagin, *Recursive linear regression theory, optimal filter theory, and error analyses of optimal systems*, IEEE International Convention Record Part 1 (1964), 216–240.

[9] J. Geweke and G. Amisano, *Hierarchical Markov normal mixture models with applications to financial asset returns*, Manuscript available at http://www.biz.uiowa.edu/faculty/jgeweke/papers/paperA/paper.pdf, (2007).

[10] P. Green, *Reversible jump Markov Chain Monte Carlo computation and Bayesian model determination*, Biometrika **82** (1995), 711–732.

[11] J. Groen, R. Paap, and F. Ravazzolo, *Real-time inflation forecasting in a changing world*, Manuscript, Erasmus University, (2008).

[12] A. Jazwinsky, *Stochastic processes and filtering theory*, New York: Academic Press, (1970).

[13] G. Koop, R. Leon-Gonzalez, and R. Strachan, *On the evolution of the monetary policy transmission mechanism*, Journal of Economic Dynamics and Control **33** (2009), 997–1017.

[14] G. Koop and S. Potter, *Forecasting in dynamic factor models using Bayesian model averaging*, The Econometrics Journal **7** (2004), 550–565.

[15] D. Korobilis, *Assessing the transmission of monetary policy shocks using dynamic factor models*, Discussion Paper 9-14, University of Strathclyde, (2009).

[16] D. Madigan and J. York, *Bayesian graphical models for discrete data*, International Statistical Review **63** (1995), 215–232.

[17] G. Primiceri, *Time varying structural vector autoregressions and monetary policy*, Review of Economic Studies **72** (2005), 821–852.

[18] A. Raftery, M. Karny, J. Andrysek, and P. Ettler, *Online prediction under model uncertainty via dynamic model averaging: Application to a cold rolling mill*, Technical report 525, Department of Statistics, University of Washington, (2007).

[19] J. Smith and J. Miller, *A non-Gaussian state space model and application to prediction records*, Journal of the Royal Statistical Society, Series B **48** (1986), 79–88.

[20] J. Stock and M. Watson, *Phillips curve inflation forecasts*, NBER Working Paper No. 14322, (2008).

[21] J. Stock and M. Watson, *Forecasting inflation*, Journal of Monetary Economics **44** (1999), 293–335.

[22] J. Stock and M. Watson, *Evidence on structural instability in macroeconomic time series relations*, Journal of Business and Economic Statistics **14** (1996), 11–30.

Time Series, Breaks and Economic Forecasting

Jeroen V.K. Rombouts

The general objective of this project is to improve economic forecasts with evolutionary econometric models, i.e. models that adapt to structural changes in the economic environment. They are also called change-point models. The need for adaptive modelling is obvious in the light of the current economic downturn.

Economic forecasting is essential for decision making with respect to fiscal and monetary policy, public spending, and investment. For example, the monetary transmission mechanism has long and uncertain lags. Therefore, monetary policy should be forward-looking. To effectively ensure price stability, the central bank needs to make forecast about the evolution of prices and output among others. Forecasting member states deficits is also needed to monitor the implementation of the stability pact of the EU.

Forecasting future observations of economic time series is done conditional on past information since there is a wide consensus, based on experience, that the future depends on the past. However, when there are structural changes we have to be extremely careful. Structural changes (also called breaks), may change the dynamics and even the trend of a time series and its volatility. Ignoring these breaks, that is assuming constant parameters in the econometric model, typically leads to forecasts that are far from realisations. As a consequence, an important econometric challenge is to detect as soon as possible a structural break. Once we know there is a break we can produce forecasts by either only using the information since that break, or using all the past data but acknowledging there was a break in the past. The latter option is more promising especially if we allow that breaks can also occur in the forecasting period and if we believe that past breaks are informative for future breaks. In fact, if breaks have happened in the past, we can expect them to happen also in the future, especially for large forecast horizons.

This project seeks to develop a forecasting methodology based on econometric models that account automatically for structural breaks if they occur and to apply them to economic and financial variables. Modelling and forecasting this kind of variables is a major field of empirical research in economics, for example understanding and predicting the temporal dependence in economic growth and inflation rates, interest rates, exchange rates and the volatilities and co-volatilities of financial returns is important for several theoretical and applied issues. Economic models involve potentially several relations and many unknown parameters the values of which may change given certain economic conditions. For example, it is likely the case that the parameters of an economic growth model are different in booms and recessions and are furthermore likely to evolve over time.

It is well known that economic time series are not stationary, especially when observed for long periods. One source of non-stationarity are the technological and institutional changes of the economic environment, which may occur more or less abruptly. When changes are gradual, they do not translate immediately into observed data as there may be threshold or ratchet effects. Econometric models with a fixed structural form or constant parameters are thus potentially misspecified.

Markov-switching models allow capturing regime changes in econometric models, by driving the parameter changes through a discrete hidden Markov chain. At each date, the parameters of the model are in a given state determined by a value of a discrete latent variable. At the next date, they can stay in the same state or change to another state, among a few possible values. At a further date, they may switch back to a previous state, i.e. states can be recurrent. In the recent literature, Markov-switching models with non-recurrent states (also called change-point models or structural break models) have emerged as promising alternatives to recurrent state models. It may indeed be argued that booms and recessions in the economy are always different and therefore have to be modeled by an evolutionary model such as the change-point model. Another reason for these models is that with recurrent states, the model is more difficult to handle because it involves more parameters that are ultimately difficult to identify. To date, there are only a few recent papers (see the context section below) that develop and use change-point models.

Bayesian Inference for Markov-switching models or change-point models is particularly convenient for several reasons. Compared to the classical statistical paradigm, Bayesian inference considers the unknown parameters of the models as random variables. This opens the door for computer-intensive techniques that allow us to learn about the unknown parameters and simultaneously about the structural break dates. The same techniques can be directly used for calculating predictive densities. Unlike prediction in the classical framework, predictive densities take into account parameter uncertainty by construction. Bayesian inference can also deal easily with model uncertainty. For example, it may happen that of two models, model 1 is the best for the period 1980-1987, then model 2 until 1993, then again model 1 etc. Finally, another crucial feature of Bayesian inference is that we can take into account potential future structural breaks when simulating the predictive densities. Bayesian inference is also very useful for combining forecasts from several models through Bayesian model averaging.

The literature on change-point models in economics is recent (though, an influential paper from the statistics literature on mean and variance shifts in autoregressive time series is [6]) and many questions are still unresolved. The following papers draw our particular attention.

First, [7] provide a new approach to forecasting time series that are subject to discrete structural breaks. Using Bayesian inference, they propose a prediction procedure that allows for the possibility of new breaks occurring over the forecast horizon, taking account of the size and duration of past breaks (if any) by means of a hierarchical hidden Markov chain model. Predictions are formed by integrating over the parameters from the meta-distribution that characterizes the stochastic break-point process. In an application to U.S. Treasury bill rates, they find that the method leads to better out-of-sample forecasts than a range of alternative methods.

Second, [5] develop a new approach to change-point modelling that allows the number of change-points in the observed sample to be unknown. Their model assumes that regime durations have a Poisson distribution. It approximately nests the two most common approaches: the time-varying parameter model with a change-point every period and the change-point model with a small number of regimes. A Markov chain Monte Carlo posterior sampler is constructed to estimate a version of their model, which allows for change in conditional means and variances. They show how real-time forecasting can be done in an efficient manner using sequential importance sampling. Their small empirical exercise involves U.S. GDP growth and inflation.

Third, [4] use a Bayesian approach to investigate the evidence for structural breaks in reduced form time-series models of realized volatility. Using Monte Carlo simulations they demonstrate that their estimation approach is effective in identifying and dating structural breaks. Applied to daily SP 500 data from 1993-2004, they find strong evidence of a structural break in early 1997. The main effect of the break is a reduction in the variance of log-volatility.

The three papers above rely extensively on [2] who provides a new approach for models with multiple change points. He formulates the change-point model in terms of a latent discrete state variable that indicates the regime from which a particular observation has been drawn. This state variable is specified to evolve according to a discrete-time discrete-state Markov process with the transition probabilities constrained so that the state variable can either stay at the current value or jump to the next higher value associated to the next regime. This parameterization exactly corresponds to the change point model.

An alternative approach to breaks in time series is proposed by [3]. They model the break process in a state-space representation through mixture distributions for the state innovations. Similarly to [5], this allows for a random number of breaks.

The use of change-point models in forecasting has been shown to be promising when applied to a few US time series. It is now time to study the relevance of this model class to a large variety of macroeconomic and financial time series.

References

[1] S. Chib, *Marginal Likelihood from the Gibbs output*, Journal of the American Statistical Association, **90** (1995), 1313–1321.

[2] S. Chib, *Estimation and comparison of multiple change-point models*, Journal of Econometrics, **86** (1998), 221–241

[3] P. Giordani and R. Kohn, *Efficient bayesian inference for multiple change-point and mixture innovation models*, Journal of Business and Economics Statistics, **26** (2008), 66–77.

[4] C. Liu and J. Maheu, *Are there structural breaks in realized volatility?*, Journal of Financial Econometrics, **6** (2008), 326–360.

[5] G. Koop and S. M. Potter, *Estimation and Forecasting with Multiple Breaks*, Review of Economic Studies, **74** (2007), 763–789.

[6] R. E. McCulloch and R. S. Tsay, *Bayesian inference and prediction for mean and variance shifts in autoregressive time series*, Journal of the American Statistical Association, **88** (1993), 968–978.

[7] M. H. Pesaran, D. Pettenuzzo and A. Timmermann, *Forecasting time series subject to multiple structural breaks*, Review of Economic Studies, **73** (2006), 1057–1084.

On the Estimation of Dynamic Conditional Correlation Models

Christian M. Hafner
(joint work with Olga Reznikova)

Over the last 20 years, modelling of time-varying covariances of financial asset returns has become an integral part of financial econometrics. Estimation of time-varying volatilities and co-movements between financial series has proved to be a useful tool in financial management. Correlations between returns are relevant for problems such as time-varying beta coefficients in CAPM-type models, estimation of hedge ratios, and Value-at-Risk (VaR) of a portfolio.

Alternative multivariate GARCH (generalized autoregressive conditional heteroscedasticity) models, which are aiming to give simple solutions to the problems described above, have become the subject of wide discussions. The most well known of them are the VEC model of [4], the BEKK model of [9], and the DCC model of [7]. Recent proposals include the Flexible Multivariate GARCH model of [16], the asymmetric DCC model of [5], Generalized Autoregressive Conditional Correlation (GARCC) model of [20], and the Dynamic Equicorrelation (DE) model of [8]. For recent reviews of multivariate GARCH models, see [2] and [22].

The Dynamic Conditional Correlation (DCC) model of [7] is one of the most cited works related to the parametric modelling of time-varying correlations for multivariate portfolios. It is a generalization of the Constant Conditional Correlation (CCC) model of [3], where volatilities are time-varying but conditional correlations are assumed to be constant. The CCC model is, however, too restrictive as it does not take into account the time variation in co-movements of the assets during the periods of economic stability, growth or crises. [7] and [23] extended the CCC model by allowing the correlation to change over time. Both models are quite similar but, while [23] model the correlation process directly, [7] specifies the model using a nonlinear transformation of a GARCH-type process to ensure that the resulting process is a sequence of correlation matrices. [11] provide some theoretical properties for the DCC model, but due to the complexity of the model, a rigorous treatment of the theory is not yet available.

Although the DCC model is easy to implement and is widely used, it is now well accepted that it does not perform well for the case of large dimensions. To a minor extent, the reason is the assumption of the same parameters driving all correlations. [12] extend the DCC model by allowing the parameters to vary across the assets. More importantly, it turned out that in high dimensions the parameter estimates encounter severe negative biases, resulting in increasingly smooth correlation trajectories, which eventually become virtually flat and constant. [11] recognize the presence of downward bias in the estimated parameters but do not propose solutions.

[1] reveals a weak point in the DCC model of [7]. His theoretical computations and results of the simulation study show that the DCC model possesses a significant asymptotic bias in the estimator of the sample covariance matrix which is a constituent of the correlation evolution process. [1] proposes a consistent DCC

(cDCC) model. He modifies the form of the correlation driving process of the DCC model in such a way that it has martingale difference innovations.

[10] suggest a composite likelihood estimator based on summing up the quasi-likelihood functions of subsets of assets and thus avoids working with high dimensional matrices. They work with the specification of [1] and suppose that the bias problem of the standard DCC model stems from the bias in the covariance targeting parameter. However, we show in this paper that the problem of biased parameter estimates prevails in the specification proposed by [1] in high dimensions. This suggests that the problem is genuine to the dimensionality issue.

We claim that the main problem in estimating either the standard DCC model or the modified version of [1] in high dimensions is an ill-conditioned estimator of the sample covariance matrix, which is used for covariance targeting. We suggest using the shrinkage technique of [13], which is a solution to obtain a well-conditioned and asymptotically accurate estimator of the covariance matrix. We show that this approach considerably improves the downward bias of the DCC model as well as the cDCC model. Moreover, our estimator is asymptotically efficient, because the shrinkage intensity goes to zero when keeping the dimension fixed and letting the sample size go to infinity. This contrasts the composite likelihood estimator of [10], which will be less efficient due to the information loss by ignoring the joint likelihood.

References

[1] G. P. Aielli, *Consistent estimation of large scale dynamic conditional correlations*, Working paper n. 47, University of Messina, Department of Economics, Statistics, Mathematics and Sociology, (2008)

[2] L. Bauwens, S. Laurent, and J. Rombouts, *Multivariate GARCH models: a survey*, Journal of Applied Econometrics **21** (2006), 79–109.

[3] T. Bollerslev, *Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH approach*, Review of Economic and Statistics **72** (1990), 498–505.

[4] T. Bollerslev, R. F. Engle, and J. Wooldridge, *A capital asset pricing mode1 with time varying covariances*, Journal of Political Economy **96** (1988), 116–131.

[5] L. Cappiello, R. F. Engle, and K. Sheppard, *Asymmetric dynamics in the correlations of global equity and bond returns*, Journal of Financial Econometrics **4** (2006), 537–572.

[6] R. F. Engle, *Anticipating Correlations: A New Paradigm for Risk Management*, Princeton University Press, (2009).

[7] R. F. Engle, *Dynamic conditional correlation: a simple class of multivariate generalized autoregressive conditional heteroskedasticity models*, Journal of Business and Economic Statistics **20** (2002), 339–350.

[8] R. F. Engle and B. Kelly, *Dynamic equicorrelation*, Manuscript, Stern Business School, NYU, (2008).

[9] R. F. Engle and K. F. Kroner, *Multivariate simultaneous GARCH*, Econometric Theory **11** (1995), 122–150.

[10] R. F. Engle, N. Shephard, and K. Sheppard, *Fitting and testing vast dimensional time-varying covariance models*, Working Paper n. 30, Oxford Financial Research Centre, Financial Economics, (2008).

[11] R. F. Engle and K. Sheppard, *Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH*, Working Paper 2001-15, University of California at San Diego, (2001).

[12] C. M. Hafner and P. H. Franses, *A generalized dynamic conditional correlation model: Simulation and application to many assets*, Econometric Reviews **28** (2009), 612-631.
[13] O. Ledoit and M. Wolf, *A well-conditioned estimator for large-dimensional covariance matrices*, Journal of Multivariate Analysis **88** (2004a), 365-411.
[14] O. Ledoit and M. Wolf, *Honey, I shrunk the sample covariance matrix*, Journal of Portfolio Management **30** (2004b), 110–119.
[15] O. Ledoit and M. Wolf, *Improved estimation of the covariance matrix of stock returns with an application to portfolio selection*, Journal of Empirical Finance **10** (2003), 603-621.
[16] O. Ledoit, P. Santa Clara, and M. Wolf, *Flexible multivariate GARCH modeling with an application to international stock markets*, Review of Economics and Statistics **85** (2003), 735-747.
[17] X. Long, L. Su, and A. Ullah, *Estimation of dynamic conditional covariance: A semiparametric multivariate model*. Journal of Business Economics and Statistics (2009), forthcoming.
[18] F. Longin and B. Solnik, *Is the correlation in international equity returns constant: 1960-1990?*, Journal of International Money and Finance **14** (1995), 3–26.
[19] H. Lütkepohl, *Introduction to multiple time series analysis*, 2nd ed. Berlin, New York, Springer, (1993).
[20] M. McAleer, F. Chan, S. Hoti, and O. Lieberman, *Generalized autoregressive conditional correlation*, Econometric Theory **24** (2008), 1554-1583.
[21] W. F. Sharpe, *A simplified model for portfolio analysis*, Management Science **9** (1963), 277–293.
[22] A. Silvennoinen and T. Teräsvirta, *Multivariate GARCH models*, In T. G. Andersen, R. A. Davis, J. P. Kreiss, and T. Mikosch (Eds.), Handbook of Financial Time Series, Springer-Verlag, (2008), forthcoming.
[23] Y. K. Tse and A. K. C. Tsui, *A multivariate GARCH model with time-varying correlations*, Journal of Business and Economic Statistics **20** (2002), 351-362.

Semiparametric Estimation of Locally Stationary Diffusion Models

Oliver Linton
(joint work with Bonsoo Koo)

The theory of asset pricing has been one of the fastest growing fields of study over the past decades.The diffusion process lays at the heart of modeling the dynamics of economic variables, including the term structure of interest rates.

$$dX_t = \mu_t dt + \sigma_t dW_t, \tag{1}$$

where $\{W_t : t \geq 0\}$ is a standard Brownian motion defined on the filtered probability space $(\Omega, \mathcal{F}^W, (\mathbb{F}_t^W), P)$, $\mathcal{F}^W$ is a σ-algebra, $\mathbb{F}_t^W$ is a filtration. Here, μ_t and σ_t^2 are commonly referred to as the conditional drift or instantaneous return function, and the conditional diffusion or volatility function of the process respectively. They can depend on X_t and time.

In line with the remarkable progress in asset pricing theory, the estimation methodologies concerning continuous-time stochastic processes have improved immensely over the most recent decade. Because one of the main goals of financial econometrics is to investigate the expected returns and volatilities of the underlying dynamics of economic variables such as stocks, interest rates, exchange rates,

and their derivatives, the econometric treatment of estimating the above two functions of interest has advanced quite significantly and become more sophisticated.

In spite of the progress on various fronts of econometric theory on continuous-time stochastic processes, quite a few challenges still remain to be addressed. For instance, identification, estimation and studies of the asymptotic properties of the continuous-time processes have turned out to be quite demanding, mainly because we only have discretely sampled observations drawn from processes whose dynamics are continuous in time. Moreover, while the diffusion process in (1) is quite extensively used, asset pricing theory doesn't narrow down the number of possible specifications for the drift and volatility terms, let alone pin down their exact forms. For example, an array of different specifications have been proposed for the term structure dynamics (see [1]).

As a result, many of stochastic models have been chosen simply due to mathematical manipulability and simplicity of statistical inferences. One salient example could be the assumption of stationarity, i.e., the assumption of the existence of a time invariant stationary distribution. Indeed, most of the financial econometrics theories depend on the assumption of stationarity of the observed process. Undoubtedly, this is because a stationarity assumption provides a powerful device for identification and estimation of the underlying continuous-time data generating process. More specifically, there are conspicuous benefits attributed to this assumption. Among them, most importantly, it enables us to avoid the serious identification issue known as the aliasing problem, since cross-restrictions can be imposed. It is worth mentioning that in general, there is no one-to-one correspondence (bijection) between the parameters or functionals of the continuous-time model and its corresponding discrete time model. Moreover, under suitable conditions, a strict stationarity assumption guarantees that the distribution of the diffusion process is completely characterized by two functionals of our concern. [1] and [3] used this property to substantiate their arguments. Another compelling reason for the stationarity assumption is because well-established asymptotic results are readily available. Therefore, it makes establishment of estimation and inference procedures much simpler.

Nevertheless, there are a plethora of cases in which an assumption of stationarity is unrealistic and unjustifiable. For example, irregularity of the trade at the beginning or the end of financial markets, volatility clustering, and ruptures arising from shocks or structural changes are among those. In truth, nonstationary properties of financial data are often found in many problems of interest in economics and finance. In addition, time series in economic and financial markets have inherently dynamic and time-varying nature. Therefore, the stationarity assumption is unlikely to hold in many cases. Consequently, it would be ideal, if not indispensable, for the sake of many applications in economics and finance that nonstationarity could be allowed for. Not only does this relaxation enable us to fit the data better, but may also allow us to augment the model built upon the assumption of stationarity to a more general model with nonstationary characteristics.

With these concerns as a motivation, this paper provides a semiparametric estimation procedure with respect to the time-inhomogeneous diffusion processes along the line of an important class of nonstationarity, *local stationarity*. More specifically, our approach is based on the density matching method of [1] along with the concept of local stationarity in [2]. We consider a method which is a generalization of [1]'s method for stationary diffusion processes.

This paper proposes appropriate estimators of the quantities of interest of the locally stationary diffusion processes in an attempt to establish the appropriate asymptotic theory in a non- and semi-parametric framework. This paper suggests our proposed estimators of the drift and volatility of processes are consistent and asymptotically normal.

In addition, we obtain the uniform rate of convergence of the estimator of the drift and the volatility functions. Finally, we present a simulation study and an application to weekly interest rate data to illustrate the finite sample properties of the proposed estimators of the drift and diffusion.

References

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

[2] R. Dahlhaus, *On the Kullback-Leiber information divergence of locally stationary processes*, Stochastic Processes and their Applications **62** (1996), 139–168.

[3] G. Jiang and J. Knight, *A nonparmetric approach to the estimation of diffusion processes, with an application to a short-term interest rate model*, Econometric Theory **13** (1997), 615–645.

Factor Modelling for Multiple Time Series: A Simple Approach with Simple Inference

Qiwei Yao

Let $\{\mathbf{Y}_t\}$ be a $d \times 1$ time series. When d is large, the number of parameters in a VARMA or even VAR model is large. In fact it may be too large to lead to a statistically effective and practically meaningful fitting. Under those circumstances, an attempt to reduce the dimensionality is pertinent. If we are interested in the linear dynamic structure of $\mathbf{Y}_t$ only, conceptually we may think that $\mathbf{Y}_t$ consists of two parts: a dynamic component driven by, hopefully, a low-dimensional process and a static part (i.e. a white noise). This leads to the decomposition:

$$\mathbf{Y}_t = \mathbf{A}\mathbf{X}_t + \boldsymbol{\varepsilon}_t, \tag{1}$$

where $\mathbf{X}_t$ is an $r \times 1$ latent process with (unknown) $r < d$, $\mathbf{A}$ is a $d \times r$ unknown constant matrix, and $\boldsymbol{\varepsilon}_t \sim \mathrm{WN}(\boldsymbol{\mu}_\varepsilon, \boldsymbol{\Sigma}_\varepsilon)$ is a vector white noise process. When r is much smaller than d, we achieve an effective dimension-reduction, as then the serial dependence of $\mathbf{Y}_t$ is driven by that of a much lower-dimensional process $\mathbf{X}_t$. We call $\mathbf{X}_t$ a factor process.

Since none of the elements on the RHS of (1) are observable, we have to characterize them further to make them identifiable. First we assume that no linear

combinations of $\mathbf{X}_t$ are white noise, as any such component can be absorbed into $\boldsymbol{\varepsilon}_t$. We also assume that the rank of $\mathbf{A}$ is r. (Otherwise (1) may be expressed equivalently in terms of a lower-dimensional factor.) Furthermore, since (1) is unchanged if we replace $(\mathbf{A}, \mathbf{X}_t)$ by $(\mathbf{AH}, \mathbf{H}^{-1}\mathbf{X}_t)$ for any invertible $r \times r$ matrix $\mathbf{H}$, we may assume that the columns of $\mathbf{A} = (\mathbf{a}_1, \cdots, \mathbf{a}_r)$ are orthonormal, i.e., $\mathbf{A}'\mathbf{A} = \mathbf{I}_r$, where $\mathbf{I}_r$ denotes the $r \times r$ identity matrix. Note that even with this constraint, $\mathbf{A}$ and $\mathbf{X}_t$ are not uniquely determined in (1), as the aforementioned replacement is still applicable for any orthogonal $\mathbf{H}$. However the factor loading space, i.e. the r-dimensional linear space spanned by the columns of $\mathbf{A}$, denoted by $\mathcal{M}(\mathbf{A})$, is uniquely defined.

We summarize into condition C1 all the assumptions introduced so far.

C1. In model (1), $\boldsymbol{\varepsilon}_t \sim \mathrm{WN}(\boldsymbol{\mu}_{\boldsymbol{\varepsilon}}, \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}})$, $\mathbf{c}'\mathbf{X}_t$ is not white noise for any constant $\mathbf{c} \in \mathbb{R}^d$. Furthermore $\mathbf{A}'\mathbf{A} = \mathbf{I}_r$.

The key in the statistical inference for model (1) is to estimate $\mathbf{A}$, or more precisely $\mathcal{M}(\mathbf{A})$. Once we have obtained an estimator, say, $\widehat{\mathbf{A}}$, a natural estimator for the factor process is

$$\widehat{\mathbf{X}}_t = \widehat{\mathbf{A}}'\mathbf{Y}_t, \tag{2}$$

and the resulting residuals are

$$\widehat{\boldsymbol{\varepsilon}}_t = (\mathbf{I}_d - \widehat{\mathbf{A}}\widehat{\mathbf{A}}')\mathbf{Y}_t. \tag{3}$$

The dynamic modelling for $\mathbf{Y}_t$ is achieved via $\widehat{\mathbf{Y}}_t = \widehat{\mathbf{A}}\widehat{\mathbf{X}}_t$, and such a modelling for $\widehat{\mathbf{X}}_t$. A parsimony fitting for $\widehat{\mathbf{X}}_t$ may be obtained by rotating $\widehat{\mathbf{X}}_t$ appropriately (see [23]). Such a rotation is equivalent to replace $\widehat{\mathbf{A}}$ by $\widehat{\mathbf{A}}\mathbf{H}$ for an $r \times r$ orthogonal matrix $\mathbf{H}$. Note that $\mathcal{M}(\widehat{\mathbf{A}}) = \mathcal{M}(\widehat{\mathbf{A}}\mathbf{H})$, and the residuals (3) are unchanged with such a replacement.

Below we outline two methods for estimating $\mathbf{A}$ (and also r). The method via expanding the white noise space is more general. It can handle nonstationary factors. Unfortunately it involves solving some nonlinear optimization problems with upto $(d-1)$ variables, therefore is only applicable with moderately large d. If we are prepared to entertain the stationarity condition (see C2 in section 3 below), the problem boils down to finding eigenvalues and eigenvectors for a $d \times d$ non-negative definite matrix. Furthermore r is the number of non-zero eigenvalues of this matrix. Hence the method can be applied to the cases with d in the order of a few thousands.

Our goal is to estimate $\mathcal{M}(\mathbf{A})$, or its orthogonal complement $\mathcal{M}(\mathbf{B})$, where $\mathbf{B} = (\mathbf{b}_1, \cdots, \mathbf{b}_{d-r})$ is a $d \times (d-r)$ matrix for which $(\mathbf{A}, \mathbf{B})$ forms a $d \times d$ orthogonal matrix, i.e. $\mathbf{B}'\mathbf{A} = 0$ and $\mathbf{B}'\mathbf{B} = \mathbf{I}_{d-r}$ (see also C1). We call $\mathcal{M}(\mathbf{B})$ the white noise space.

It follows from (1) that

$$\mathbf{B}'\mathbf{Y}_t = \mathbf{B}'\boldsymbol{\varepsilon}_t. \tag{4}$$

Hence for any $1 \le j \le d-r$, $\{\mathbf{b}_j'\mathbf{Y}_t,\ t = 0, \pm 1, \cdots\}$ is a white noise process. Hence, we may search for mutually orthogonal directions $\mathbf{b}_1, \mathbf{b}_2, \cdots$ one by one

such that the projection of $\mathbf{Y}_t$ on each of those directions is a white noise. We stop the search when such a direction is no longer available, and take $d-k$ as the estimated value of r, where k is the number of directions obtained in the search. For further details on the implementation of this method, we refer to [17].

In principle, this method works under condition C1 which is rather weak. For example, we do not require condition C2 below, as we only make use of the property that the project of $\mathbf{Y}_t$ in $\mathcal{M}(\mathbf{B}) = \mathcal{M}(\mathbf{A})^{\perp}$ is white noise. The theoretical exploration of the method in [17] requires more regularity conditions. Intuitively some of those conditions are not essential.

A much simpler method is available with the stationarity condition on the factor process as follows.

C2. $\mathbf{X}_t$ is weakly stationary, and $\mathrm{Cov}(\mathbf{X}_t, \boldsymbol{\varepsilon}_{t+k}) = 0$ for any $k > 0$.

Put

$$\boldsymbol{\Sigma}_y(k) = \mathrm{Cov}(\mathbf{Y}_{t+k}, \mathbf{Y}_t), \quad \boldsymbol{\Sigma}_x(k) = \mathrm{Cov}(\mathbf{X}_{t+k}, \mathbf{X}_t), \quad \boldsymbol{\Sigma}_{x\varepsilon}(k) = \mathrm{Cov}(\mathbf{X}_{t+k}, \boldsymbol{\varepsilon}_t).$$

It follows from (1) and C2 that

$$\boldsymbol{\Sigma}_y(k) = \mathbf{A}\boldsymbol{\Sigma}_x(k)\mathbf{A}' + \mathbf{A}\boldsymbol{\Sigma}_{x\varepsilon}(k), \qquad k \geq 1. \tag{5}$$

For a prescribed integer $k_0 \geq 1$, define

$$\mathbf{M} = \sum_{k=1}^{k_0} \boldsymbol{\Sigma}_y(k)\boldsymbol{\Sigma}_y(k)'.$$

Then $\mathbf{M}$ is a $d \times d$ non-negative matrix. It follows from (5) that $\mathbf{MB} = 0$, i.e. the columns of $\mathbf{B}$ are the eigenvectors of $\mathbf{M}$ corresponding to zero-eigenvalues. Hence under conditions C1 and C2 we may conclude:

> *The factor loading space $\mathcal{M}(\mathbf{A})$ are spanned by the eigenvectors of $\mathbf{M}$ corresponding to its non-zero eigenvalues, and the number of the non-zero eigenvalues is r.*

To estimate $\mathcal{M}(\mathbf{A})$, we only need to perform the eigenanalysis on

$$\widehat{\mathbf{M}} = \sum_{k=1}^{k_0} \widehat{\boldsymbol{\Sigma}}_y(k)\widehat{\boldsymbol{\Sigma}}_y(k)',$$

where $\widehat{\boldsymbol{\Sigma}}_y(k)$ denotes the sample covariance matrix of $\mathbf{Y}_t$ at lag k.

When d is fixed, the estimators for the zero-eigenvalues of $\mathbf{M}$ converge at the fast rate n while the estimators for non-zero eigenvalues converge at the standard rate $n^{1/2}$, as the sample size $n \to \infty$; see, e.g. Theorem 1 of [5]. The fast rate was due to the quadratic form in the definition of $\mathbf{M}$ and $\widehat{\mathbf{M}}$.

This estimation method has been taken further in modelling curve time series by [5], in large factor modelling (i.e. $d \to \infty$) by [14], and [15], in modelling multivariate volatility processes by [16] and [24].

The approach outlined above is statistical, aiming to reduce the dimensionality in the modelling. The setting (1) may be traced back at least to [18]; see also its further development in dealing with cointegrated factors in [19].

Factor modelling for time series has been more predominately featured in econometrics literature. Motivated by modelling economic and financial behaviours, [20] and [12] proposed dynamic-factor models, [7] and [6] proposed static (approximate) factor models. Combining the above two approaches, [9], [10], [11] proposed and further developed the so-called generalized dynamic-factor model. See also, among others, [21], [22], [2], [3], [4], [13], [8], and [1].

Similar but also radically different from statistical model (1), econometric factor models also decompose $\mathbf{Y}_t$ into two parts: a common factor and the so-called idiosyncratic 'noise'. The common factor drives the dynamics of *most* components of $\mathbf{Y}_t$ while each idiosyncratic component only affects the dynamics of a few components of $\mathbf{Y}_t$. Obviously it is extremely appealing and also important to isolate those common factor components from many idiosyncratic components in analysing economic and financial phenomena. Since idiosyncratic 'noise' is not white noise, the identification and the inference for those models are inevitably more challenging. For example, the two parts (i.e. the common factor and the idiosyncratic 'noise') are only asymptotically identifiable when d, the number of components $\mathbf{Y}_t$, tends to ∞; see [7] and [9].

References

[1] L. Alessi, M. Barigozzi, and M. Capasso, *A robust certerion for determining the number of static factors in approximate factor models*, European Central Bank Working Paper 903 (2009).

[2] J. Bai, *Inferential theory for factor models of large dimensions*, Econometrica **71** (2003), 135–171.

[3] J. Bai and S. Ng, *Determining the number of factors in approximate factor models*, Econometrica **70** (2002), 191–221.

[4] J. Bai and S. Ng, *Determining the number of primitive shocks in factor models*, Journal of Business and Economic Statistics **25** (2007), 52–60.

[5] N. Bathia, Q. Yao, and F. Zieglemann, *Identifying the finite dimensionality of curve time series*, Manuscript available at http://stats.lse.ac.uk/q.yao/qyao.links/paper/byz.pdf, (2010).

[6] G. Chamberlain, *Funds, factors, and diversification in arbitrage pricing models*, Econometrica **51** (1983), 1305–1323.

[7] G. Chamberlain and M. Rothschild, *Arbitrage, factor structure, and mean-variance analysis on large asset markets*, Econometrica **51** (1983), 1281–1304.

[8] M. Deistler, B. D. O. Anderson, A. Filler, C. Zinner, and W. Chen, *Generalized linear dynamic factor models – an approach via singular autoregressions*, European Journal of Control (2009).

[9] M. Forni, M. Hallin, M. Lippi, and L. Reichlin, *The generalized dynamic-factor model: identification and estimation*, Review of Economics and Statistics **82** (2000), 540–554.

[10] M. Forni, M. Hallin, M. Lippi, and L. Reichlin, *The generalized dynamic-factor model: consistency and rates*, Journal of Econometrics **119** (2004), 231–255.

[11] M. Forni, M. Hallin, M. Lippi, and L. Reichlin, *The generalized dynamic-factor model: one-sided estimation and forecasting*, Journal of the American Statistical Association **100** (2005), 830–840.

[12] J. Geweke, *The dynamic factor analysis of economic time series*, In Latent Variables in Socio-Economic Models, ed. Aigner, D.J. and Goldberger A.S., Chapter 19, Amsterdam, North-Holland, (1977).

[13] M. Hallin and R. Liska, *Determining the number of factors in the general dynamic factor model*, Journal of the American Statistical Association **102** (2007), 603–617.

[14] C. Lam and Q. Yao, *Large latent factor for time series I: model and estimation*, Manuscript, (2010).
[15] C. Lam, Q. Yao, and N. Bathia, *Large latent factor for time series II: nonstationary factors and forecasting*, Manuscript, (2010).
[16] J. Pan, D. Peña, W. Polonik, and Q. Yao, *Modelling multivariate volatilities via common factors*, Manuscript available at http://stats.lse.ac.uk/q.yao/qyao.links/paper/pppy.pdf, (2008).
[17] J. Pan and Q. Yao, *Modelling multiple time series via common factors*, Biometrika **95** (2008), 356–379.
[18] D. Peña and E. P. Box, *Identifying a simplifying structure in time series*, Journal of the American Statistical Association **82** (1987), 836–843.
[19] D. Peña and P. Poncela, *Nonstationary dynamic factor analysis*, Journal of Statistical Planning and Inference **136** (2006), 1237–1257.
[20] T. J. Sargent and C. A. Sims, *Business cycle modeling without pretending to have too much a priori economic theory*, In New methods in business cycle research, ed. Sims, C. et al, Minneapolis, Federal Reserve Bank of Minneapolis, 45–108, (1977).
[21] J. H. Stock and M. W. Watson, *New indexes of coincident and leading economic indicators*, NBER Macroeconomics Annual **4** (1989), 351–394.
[22] J. H. Stock and M. W. Watson, *Macroeconomic forecasting using diffusion indices*, Journal of Business and Economic Statistics **20** (2002), 147–162.
[23] G. C. Tiao and R. S. Tsay, *Model specification in multivariate time series (with discussion)*, Journal of the Royal Statistical Society, Series B **51** (1989), 157–213.
[24] Y. Wang, Q. Yao, P. Li, and J. Zou, *High dimensional volatility modeling and analysis for high-frequency financial data*, Manuscript available at http://stats.lse.ac.uk/q.yao/qyao.links/paper/highfreq.pdf, (2008).

On the thick-pen transformation for time series

Piotr Fryzlewicz
(joint work with Hee-Seok Oh)

Traditional visualisation of time series data consists of plotting the time series values against time and "connecting the dots". We propose an alternative, multiscale visualisation technique, motivated by the scale-space approach in computer vision. In brief, our method also "connects the dots", but uses a range of pens of varying thicknesses for this purpose. The resulting multiscale map, termed the Thick-Pen Transform (TPT) corresponds to viewing the time series from a range of distances. We hope that the resulting set of plots will provide interesting and useful information about the structure of the time series, not only in a heuristic, but also in a formal probabilistic sense.

We define the TPT of a real-valued univariate process $(X_t)_{t=1}^n$ as follows. Let $\mathcal{T}$ denote the set of thickness parameters. For each $\tau_i \in \mathcal{T}$, $i = 1, \ldots, |\mathcal{T}|$, let $U_t^{\tau_i}$ denote the upper boundary of the area covered by a pen of thickness τ_i while connecting the points $(t, X_t)_{t=1}^n$. Similarly, let $L_t^{\tau_i}$ denote its lower boundary. The TPT $TP_{\mathcal{T}}(X_t)$ is the sequence of all pairs of boundaries, i.e.

$$TP_{\mathcal{T}}(X_t) = \{(L_t^{\tau_i}, U_t^{\tau_i})_{t=1}^n\}_{i=1,\ldots,|\mathcal{T}|}.$$

The precise mathematical form of $TP_{\mathcal{T}}(X_t)$ depends on the shape of the pen used. For example, consider a pen which is a closed square of side length $\tau \in \mathcal{T}$,

positioned so that two of its sides are parallel to the time axis. For each point along the straight line connecting (t, X_t) with $(t+1, X_{t+1})$, we place the pen so that the given point is at the centre of the right-hand side of the pen. In this set-up, we have

$$U_t^\tau = \max(X_t, \ldots, X_{t+\tau}) + \frac{\tau}{2} \quad (1)$$

$$L_t^\tau = \min(X_t, \ldots, X_{t+\tau}) - \frac{\tau}{2}. \quad (2)$$

Other pen shapes are possible, in the same way that a variety of kernel shapes are possible in kernel smoothing. Considering $TP_{\mathcal{T}}(X_t)$ for a range of thickness values τ, a multiscale transform of the data X_t is obtained, with higher values of the thickness parameter bringing out coarser-scale features of the data, and vice versa.

The TPT is not the only multiscale tool in time series analysis. Wavelets, which provide linear, multiscale and local decomposition of data, have been used extensively in time series analysis (see e.g. [5]). SiZer (see [1]) is a linear data visualisation technique for displaying features of kernel-smoothed data as a function of location and bandwidth, simultaneously over a range of bandwidths. We note here that the TPT is not linear as it is based, effectively, on localised and weighted min/max operations. Self-similarity and (multi-)fractality are oft-recurring concepts in time series analysis, aiming to study parametric relationships between distributions of the process at different scales, particularly in the context of long-range dependent processes, see e.g. [2]. Besides using different methodology, the aims of the TPT are different: we regard it as a visualiser which can be applied to any time series and which can ultimately assist in solving tasks such as nonstationarity detection, classification or measuring dependence between time series.

In the TPT as described above, one thickness value τ generates two sequences: U_t^τ and L_t^τ. In some time series problems, it might be more convenient to use a single summary sequence, instead of a pair. Probably the simplest possible summary sequences involving U_t^τ and L_t^τ are

- Volume of the pen, defined as $V_t^\tau = U_t^\tau - L_t^\tau$;
- Mean of the pen, defined as $M_t^\tau = \frac{1}{2}\{U_t^\tau + L_t^\tau\}$.

Many more summary statistics are possible, also those combining U_t^τ and L_t^τ nonlinearly. The volume statistic V_t^τ deserves special attention as statistical literature has previously explored the concept of "the volume of a covering of data", albeit in other contexts. [7] derived the "tube formula" for calculating the volume of a tube surrounding a smooth manifold. This result has more recently been applied in various statistical contexts by a number of authors, see e.g. [6]. We are unaware of any applications of tube formulae in classical time series, where sample paths are often intrinsically non-smooth. On the other hand, in estimating the Hurst exponent or the fractal dimension of stochastic processes, two techniques involving statistics related to V_t^τ are the Rescaled Range Analysis (see [4]) and the "box-counting" method, whose statistical properties in estimating the fractal dimension of a stationary continuous-time Gaussian process were studied in [3].

By contrast, our V_t^τ statistic is not an estimator, and applies to discrete-time, also nonstationary processes.

We now state a discrimination property of the TPT, which implies, roughly speaking, that two differently distributed Gaussian time series have differently distributed TPTs, under the (mild) Assumption 1 below. This is an important result as it gives us hope that the TPT can serve as an effective discriminant for time series.

Assumption 1. *For a given fixed lag $\tau > 0$, a process X_t satisfies*

$$\exists\, \lambda_0, \delta \in [0,1) \quad \forall \lambda > \lambda_0 \quad \forall\, t$$
$$P\left(\bigcup_{\substack{t \le i,j \le t+\tau; \\ \{i,j\} \neq \{t,t+\tau\}}} |X_i - X_j| > |X_t - X_{t+\tau}| \;\middle|\; |X_t - X_{t+\tau}| > \lambda \right) \le \delta.$$

Discrimination theorem. *Let X_t, Y_t be two zero-mean Gaussian time series such that for some $s < t$, the distribution of $X_s - X_t$ is not the same as the distribution of $Y_s - Y_t$, and let both X_t and Y_t satisfy Assumption 1 with $\tau = t - s$. Let $TP_{\mathcal{T}}(X_t)$, $TP_{\mathcal{T}}(Y_t)$ be the TPTs of X_t, Y_t respectively, both with the square pen where the set $\mathcal{T}$ of thickness parameters is $\mathcal{T} = \{1, 2, \ldots\}$, and let $V_t^\tau(X), V_t^\tau(Y)$ be the corresponding volumes. Then, $TP_{\mathcal{T}}(X_t)$ and $TP_{\mathcal{T}}(Y_t)$ follow different probability distributions in the sense that the tri-variate random vectors $(V_s^{\tau-1}(X), V_{s+1}^{\tau-1}(X), V_s^\tau(X))$ and $(V_s^{\tau-1}(Y), V_{s+1}^{\tau-1}(Y), V_s^\tau(Y))$ are distributed differently.*

We have applied the TPT to testing for time series stationarity, and to quantifying dependence between two time series. We introduce here the former application. The key result is as follows.

Functional central limit theorem. *Let $\{X_t\}_{t=1}^n$ be a stationary process satisfying $\mathbb{E}|X_t|^r < \infty$ for some $r > 2$. In addition let X_t be α-mixing with the mixing coefficients α_m satisfying $\alpha_m = O(m^{-s})$ for some $s > \frac{r}{r-2}$. Let $TP_{\mathcal{T}}(X_t)$ be the TPT of X_t using an arbitrary pen but such that both U_t^τ and L_t^τ are functions of $X_{t-C\tau}, \ldots, X_{t+C\tau}$ only, for some $C > 0$. Further let the summary sequence K_t^τ be such that for each fixed τ, we have $n^{-1}\mathrm{Var}(\sum_{t=1}^n K_t^\tau) \to \sigma_\tau^2 < \infty$, and $|K_t^\tau| \le A + B|\max(X_{t-C\tau}, \ldots, X_{t+C\tau})|$ for some constants $A, B > 0$, possibly depending on τ. Under these conditions, the following functional central limit result holds for each fixed τ. Let $u \in [0,1]$ and denote $Y_n^\tau(u) = \sigma_\tau^{-1} n^{-1/2} \sum_{t=1}^{\lceil nu \rceil} K_t^\tau - \mathbb{E}(K_t^\tau)$. We have*

$$Z_n^\tau(u) := Y_n^\tau(u) - \frac{\lceil nu \rceil}{n} Y_n^\tau(1) \xrightarrow{d} B_u^0,$$

where B_u^0 is the standard Brownian bridge process on $[0,1]$.

Our stationarity test is based on the fact that under the null hypothesis of stationarity, the range of the empirical version of $Z_n^\tau(u)$ is distributed as the range of Brownian bridge. As the test is derived from the TPT, which is a visualiser, it can be regarded as a "visual" one. Hence, it should come as no surprise that it operates under low moment assumptions, and is equally valid for linear and nonlinear processes. It appears to offer very good empirical performance.

References

[1] P. Chaudhuri and J.S. Marron, *SiZer for exploration of structures in curves*, Journal of the American Statistical Association **94** (1999), 807-823.
[2] P. Doukhan, G. Oppenheim and M.S. Taqqu, editors, *Theory and Applications of Long-Range Dependence*, Birkhäuser (2003).
[3] P. Hall and A. Wood, *On the performance of box-counting estimators of fractal dimension*, Biometrika **80** (1993), 246-252.
[4] H.E. Hurst, *Long term storage capacity of reservoirs*, Transactions of the American Society of Civil Engineers **116** (1951), 770-799.
[5] D.B. Percival and A.T. Walden, *Wavelet Methods for Time Series Analysis*, Cambridge University Press, (2000).
[6] J. Sun, *Tail probabilities of the maxima of Gaussian random fields*, Annals of Probability **21** (1993), 34-71.
[7] H. Weyl, *On the volume of tubes*, American Journal of Mathematics **61** (1939), 461-472.

Modelling Changes in the Unconditional Variance of Long Stock Return Series

Timo Teräsvirta

In this work we consider modelling daily return series that are long, that is, contain for example more than 20000 observations. A characteristic feature of such series is that they display not only volatility clustering that is often parameterised using GARCH models, but also smooth changes in the amplitude of the clusters. The latter type of fluctuations cannot be satisfactorily captured using stationary GARCH models. One solution to this problem has been to apply Fractionally Integrated GARCH (FIGARCH) models, see [2], but the standard stationary GARCH model may also be generalised in other ways. [4] constructed a locally stationary ARCH model to handle smooth changes. It is also possible to decompose the variance of the return process multiplicatively into two components: a stochastic stationary component and a deterministic nonstationary one. Examples include [7], [5] and [1]. [1] called their model the Time-Varying GARCH (TV-GARCH) model. It may also be mentioned that [3] considered a FIGARCH model with a deterministically time-varying intercept.

The deterministic component of the TV-GARCH model consists of a linear combination of logistic functions or sigmoids, whose argument is time. This linear combination of sigmoids is a very flexible functional form, which makes it useful in the present context. [1] developed a modelling strategy for building TV-GARCH models. It contains three stages: specification, estimation and evaluation. At the specification stage the form of the model is determined, which involves determining

the number of sigmoids which is unknown a priori. This can be done in various ways, but we prefer sequential testing. Maximum likelihood estimation is carried out by maximising the log-likelihood by parts as in [6]. At the evaluation stage the specified and estimated model is subjected to misspecification tests to find out whether or not it can be regarded as an adequate description of the time series under study.

This modelling strategy has been found to work very well when the time series are a few thousand observations long, in which case the number of sigmoids in the model is still relatively small. There is, however, some doubt concerning the functioning of the sequential tests when the number of sigmoids can potentially be large, for example around ten. We propose a modification of our strategy to fit into this situation. The idea is to divide the long series into subseries and apply the modelling strategy each of them separately. This means determining the number of sigmoids separately for each subseries and estimating the submodels, one for each series. After that has been done, the models are combined into one by 'tying the ends of the deterministic components together'. This can be done in different ways. One is to do it recursively, beginning with the last two subsets, re-estimating the model containing those, and working one's way backwards. How well this works is at best being investigated. Another interesting question demanding attention is how the stationary GARCH component of the model behaves when the time series are very long. Can one still assume that it has constant parameters? This problem will be studied in the near future. An example in this presentation is the long daily Dow-Jones stock index return series. The individual subset models can be specified and estimated without problems, and at the moment work is being done on how to put the pieces together into a single model.

References

[1] C. Amado and T. Teräsvirta, *Modelling volatility by variance decomposition*, Manuscript, (2009).

[2] R. T. Baillie, T. Bollerslev, and H.-O. Mikkelsen, *Fractionally integrated generalized autoregressive conditional heteroskedasticity*, Journal of Econometrics **74** (1996), 3–30.

[3] R. T. Baillie and C. Morana, *Modelling long memory and structural breaks in conditional variances: An adaptive FIGARCH approach*, Journal of Economic Dynamics and Control **33** (2009), 1577–1592

[4] R. Dahlhaus and S. Subba Rao, *Statistical inference for time-varying ARCH processes*, Annals of Statistics **34** (2006), 1075–1114.

[5] R. F. Engle and J. Gonzalo Rangel, *The spline GARCH model for low-frequency volatility and its global macroeconomic causes*, Review of Financial Studies **21** (2008), 1187–1222.

[6] P. X.-K. Song, Y. Fan and J.-D. Kalbfleisch, *Maximization by parts in likelihood inference*, Journal of the American Statistical Association **100** (2005), 1145–1158.

[7] S. van Bellegem and R. von Sachs, *Forecasting economic time series with unconditional time-varying variance*, International Journal of Forecasting **20** (2004), 611–627.

On the Univariate Representation of Multivariate Volatility Models with Common Factors

Franz C. Palm
(joint work with Alain Hecq, Sébastien Laurent)

Most financial econometrics textbooks start with univariate models to explain the presence of a time-varying volatility pattern in asset returns. Then a discussion of the properties of different models (e.g. GARCH, EGARCH, TARCH, APARCH, realized volatility, stochastic volatility) is undertaken together with their most important features in relation with stylized facts of financial data. It is also emphasized that to correctly account for the link between several series (e.g. contagion effects) and to study time-varying conditional correlations, a multivariate framework should be adopted. Indeed, understanding and predicting the dependence in the second order moments of asset returns is important for many issues in financial analysis and management.

It is not a secret however that unrestricted multivariate GARCH models and related processes of this kind suffer from the curse of dimensionality. An unrestricted GARCH$(0, q)$ for an n-dimensional set of financial assets is already of dimension $(n^2 + n)/2$ and implies $((n^2 + n)/2)^2 q$ unknown coefficients if one does not count the intercepts. As most portfolio analyses involve a large number of assets, the need for more parsimonious forms is obvious. Let us just mention the diagonal model, the constant conditional correlation (CCC), the dynamic conditional correlation (DCC), the dynamic equicorrelation (DECO, see [10]), the BEKK ([1]), the orthogonal GARCH or factor GARCH models as examples proposed in the literature to restrict the dimension of the multivariate setting to a manageable size as well as to impose the positive definitiveness of the covariance matrix (see *inter alia* the survey by [2]). To make their estimation feasible on large portfolios some of these models impose very strong restrictions on the dynamics of the covariance or correlation. One simple way to relax these constraints is to assume a block-diagonal structure where the dynamics is constrained to be the same only among groups of variables (see for instance the BLOCK-DCC and BLOCK-DECO of respectively [3], and [10]).

We take another route. We start with some general multivariate GARCH volatility models and study their implication for the underlying univariate processes. Indeed, we look at the final equation representation of multivariate GARCH models using a framework similar to that in Zellner and Palm (see *inter alia* [16],[17]) for the conditional mean. This framework allows us to derive the marginal (weak) GARCH representation for the conditional variances and conditional covariances of the multivariate GARCH model. However the implied univariate representation is far from being parsimonious. There exists a paradox between theoretically implied marginal volatility models derived from a multivariate model and the common empirical finding that low order univariate GARCH models often approximate single conditional volatilities GARCH rather well, a paradoxical issue

similar to the one observed for the VAR(p) and the marginal implied ARMA(p^*, q^*) models (see [5], [6]).

Indeed in empirical work, estimated univariate GARCH models are, too a large extent, very parsimonious. The GARCH$(1,1)$ specification for instance is able to capture the time-varying volatility in second moments of many symmetric asset returns. Consequently we look for multivariate representations that capture the multivariate dynamics and could at least be compatible with the univariate structure obtained in empirical work (like a GARCH$(1,1)$). The implied univariate models of familiar multivariate models such as the "unrestricted" BEKK (see [1]) do not reproduce the properties of individual asset returns. Therefore, we extend the analysis developed by [6] for vector autoregressive models to factor representations in the conditonal second moments. These are the Factor GARCH specification of [9] and the [12] pure variance model. These two specifications allow to explain why one might obtain such parsimonious univariate representations for potentially large multivariate systems. Alternatively, obtaining a very parsimonious univariate representation for individual returns might be an indication of the presence of co-movements or common factors in the volatility. On the contrary, evidence for the presence of long memory is a sign of absence of co-movements.

Multivariate volatility systems with common factors are not the only ones however that can imply parsimonious GARCH orders. The diagonal BEKK which implies no contagion effects can lead to similar univariate GARCH orders. A related paper to ours is [15]. Indeed these authors also study the marginal models derived from a multivariate GARCH process with a particular focus to the aggregation of individual series. They also sketch that the presence of a factor model (the FGARCH in their case) for the variance might be important in the marginalization of multivariate systems but without giving the orders of the univariate representations nor the proofs. We study the issue of marginalization of vector conditional volatlity processes with a factor structure in more detail and determine the orders of the implied weak Garch processes. Weak GARCH processes (see [8]) are in fact the Wold representation of single conditional second moments given their own past (only). This means that the innovations are white noise but not (necessarily) i.i.d. random variables or martingale difference sequences. We consider these implications of marginalization for empirical work.

In particular a likelihood procedure (assuming for instance normality) to estimate the marginal processes is in fact a quasi-likelihood procedure. This estimator is strongly consistent and asymptotically normally distributed but not efficient under the usual regularity conditions, and its standard-errors can be obtained using the so-called sandwich estimator(see for instance [13], [14], and [4]).

An important point of our paper is the need for methods to analyze subsets of highly correlated (contagion effects) assets but likely with a small number of common transmission mechanisms for the volatility and which exhibit similar univariate time series properties. We propose simple reduced rank procedures and we study their properties in a Monte Carlo simulation. We compare the behavior of the likelihood ratio test based on canonical correlation analysis on squared returns

(see [12], [11]) with a new method that uses combinations of series obtained by partial least squares. In this later case we use Box-Pierce tests for the null of no-ARCH on these combinations of squared returns and cross-returns to determine the presence of common factors. The Monte Carlo results we provide illustrate that these procedures are sensitive to the number of variables that are considered, i.e. their number and the inclusion or not of the covariances in the test statistics, as well as to the log transformation of the squared returns and cross-products.

In an emprical analysis of the daily returns of the 50 larges stocks listed at the NYSE leads to the conclusion that 6 return series exhibit very similar univariate GARCH(1,1) schemes, the remaining 44 series show signs of the presence of long memory.Various tests for the presence of common factors indicate the presence of one factor in the conditional second moments of the 6 series, a finding that is in line with results reported by [12].

Issues for future research are a more extended analysis of the properties of methods for statisitcal inference for (marginal) weak GARCH processes.

References

[1] Y. Baba, R. F. Engle, D. F. Kraft, and K. F. Kroner, *Multivariate Simultaneous Generalized ARCH*, Working paper DP 89-57, San Diego, (1989).

[2] L. Bauwens, S. Laurent and J. V. K. Rombouts, *Multivariate GARCH models: a survey*, Journal of Applied Econometrics **21** (2006), 79–109.

[3] M. Billio, M. Caporin, and M. Gobbo, *Block dynamic conditional correlation multivariate GARCH models*, Working paper 0303 GRETA, Venice, (2003).

[4] F. Comte and O. Lieberman, *Asymptotic theory for multivariate GARCH processes*, *Journal of Multivariate Analysis* **84** (2003), 61–84.

[5] G. Cubadda, A. Hecq, and F. C. Palm, *Macro-panels and reality*, Economics Letters **99**(2008), 537–540.

[6] G. Cubadda, A. Hecq, and F. C. Palm, *Studying interactions without multivariate modeling*, Journal of Econometrics **148** (2009), 25–35.

[7] G. Cubadda and A. Hecq, *Testing for common autocorrelation in data rich environments*, Manuscript, Maastricht University, (2009).

[8] F. C. Drost and T. E. Nijman, *Temporal aggregation of GARCH processes*, Econometrica **61** (1993), 909–927.

[9] R. F. Engle, V. Ng, and M. Rothschild, *Asset Pricing With a Factor ARCH Covariance Structure: Empirical Estimates of Treasury Bills*, Journal of Econometrics **45** (1990), 213–238.

[10] R. F. Engle and B. Kelly, *Dynamic equicorrelation*, Manuscript, Stern School of Business, New York University, (2008).

[11] R. F. Engle and S. Kozicki, *Testing for common features (with comments)*, Journal of Business and Economic Statistics **11** (1993), 369–395.

[12] R. F. Engle and J. Marcucci, *A Long-run pure variance common features model for the common volatilities of the Dow Jones*, Journal of Econometrics **132** (2006), 7–42.

[13] C. Francq and J.-M. Zakoian, *HAC estimation and strong linearity testing in weak ARMA models*, Journal of Multivariate Analysis **98** (2007), 114–144.

[14] T. Jeantheau, *Strong consistency of estimators for multivariate ARCH models*, Econometric Theory **15** (1998), 70–86.

[15] T. Nijman and E. Sentana, *Marginalization and contemporaneous aggregation in multivariate GARCH processes*, Journal of Econometrics **71** (1996), 71–87.
[16] A. Zellner, and F. C. Palm, *Time series analysis and simultaneous equation econometric models*, Journal of Econometrics **2** (1974), 17–54.
[17] A. Zellner and F. C. Palm, *The structural econometric time series analysis approach*, Cambridge University Press, (2004).

Nonparametric Time Series with Markov Switching Dynamics

Jürgen Franke
(joint work with J.-P. Stockis, J. Tadjuidje-Kamgaing, W. K. Li)

We start from nonlinear autoregressive-ARCH models

$$X_t = m_k(X_{t-1}, \ldots, X_{t-p}) + \sigma_k(X_{t-1}, \ldots, X_{t-p})\eta_t, \quad k = 1, \ldots, K, \tag{1}$$

where η_t are i.i.d. innovations with mean 0 and variance 1. The order p has been chosen the same for all k and for the autoregressive resp. ARCH-components only for sake of simplicity.

Controlled by a non-observable Markov chain Q_t with finite state space $\{1, \ldots, K\}$, the process Z_t, which we are interested in, switches between those K data-generating mechanisms, i.e. writing $S_{tk} = 1_{\{k\}}(Q_t)$ for the state-indicators at time t we assume

$$Z_t = \sum_{k=1}^{K} S_{tk} \Big(m_k(Z_{t-1}, \ldots, Z_{t-p}) + \sigma_k(Z_{t-1}, \ldots, Z_{t-p})\eta_t \Big). \tag{2}$$

We are interested in estimating the autoregressive and volatility functions m_k, $\sigma_k, k = 1, \ldots, K$, nonparametrically and in estimating the parameters of the Markov chain, i.e. essentially its transition matrix A, as well as in reconstructing the hidden state sequence Q_t. The key for developping an asymptotic theory for those estimates are short memory conditions for processes of the form (2). In [3], we derive conditions for Z_t being geometrically ergodic which depend in a rather simple manner on A and on the limit behaviour of the functions $m_k(x), \sigma_k(x)$ for $||x|| \to \infty$ related to well-known conditions for the stationarity of processes of the form (1). In particular, the single state processes (1) do not have to correspond to a stationary regime for any k. The switching model may still be stationary if some states are explosive provided they do not occur too frequently.

As nonparametric estimates, we first consider kernel smoothers for the autoregressive functions $m_k(x)$ assuming the conditional variances σ_k to be constant. In [4], we develop an EM algorithm for calculating those local Gaussian quasi maximum likelihood estimates numerically in case where the Q_t are i.i.d., i.e. in the special case of an independent mixture of autoregressive schemes. The single steps in that algorithm are explicit and do not require further numerical approximation. We prove consistency of the kernel estimates as well as convergence of the EM algorithm for an increasing number of iterations. We illustrate the performance

with some simulations and an example with ECG data.

Next, we have a look at the general case where Q_t is a Markov chain with arbitrary transition matrix A. We now consider approximations of $m_k(x), \sigma_k^2(x)$ by single-layer feedforward neural networks as a special case of sieve estimates. In [5], we describe an EM algorithm for calculating those estimates as well as estimates of A. Using this estimates of the system functions and parameters, a Viterbi algorithm allows to calculate an approximation of the hidden states as well. We investigate the limit behaviour of those estimates for increasing sample size, and we illustrate its application with a portfolio management problem where the trading strategy is based on forecasts of stock prices using model (2).

A main assumption made in the papers refered to above is that the evolution of the Markov chain is largely independent of the observed process Z_t, i.e. more precisely that the conditional distribution of Q_t given the past only depends on Q_{t-1} and not on $Z_s, s < t$. In [2], a parametric switching model is presented which allows for a more general form of dependence between the processes Q_t and Z_t.

References

[1] J. Franke, J.-P. Stockis, J. Tadjuidje-Kamgaing, *A note on the identifiability of the conditional expectation for the mixtures of neural networks*, Statistics and Probability Letters **78** (2008), 739–742 .

[2] J. Tadjuidje-Kamgaing, H. Ombao, R.A. Davis, *Autoregressive processes with data-driven regime switching*, Journal of Time Series Analysis **30** (2009), 505–533 .

[3] J. Franke, J.-P. Stockis, J. Tadjuidje-Kamgaing, *On geometric ergodicity of CHARME models*, Journal of Time Series Analysis (2010), forthcoming.

[4] J. Franke, W. K. Li, J.-P. Stockis, J. Tadjuidje-Kamgaing, *Mixture of nonparametric autoregressions*, Journal of Nonparametric Statistics (2010), forthcoming.

[5] J. Franke, J.-P. Stockis, J. Tadjuidje-Kamgaing, *Nonparametric estimation in Markov switching autoregressive models*, Manuscript, (2010).

Reporter: Matteo Barigozzi

Participants

Dr. Matteo Barigozzi
Universite Libre de Bruxelles
CP114
Avenue F.D. Roosevelt 50
B-1050 Bruxelles

Prof. Dr. Luc Bauwens
CORE
Voie du Roman Pays, 34
B-1348 Louvain-la-Neuve

Prof. Dr. Manfred Deistler
Institut für Ökonometrie und
Operations Research
Technische Universität
Argentinierstraße 8
A-1040 Wien

Prof. Dr. Jürgen Franke
Fachbereich Mathematik
T.U. Kaiserslautern
Erwin-Schrödinger-Straße
67653 Kaiserslautern

Prof. Dr. Piotr Fryzlewicz
Department of Statistics
London School of Economics
Houghton Street
GB-London WC2A 2AE

Prof. Dr. Christian M. Hafner
Institut de Statistique
Universite Catholique de Louvain
Voie du Roman Pays, 20
B-1348 Louvain-La-Neuve

Prof. Dr. Gary Koop
Department of Economics
University of Strathclyde
130 Rottenrow
GB-Glasgow G4 0GE

Prof. Dr. Oliver Linton
Department of Economics
London School of Economics
and Political Science
Houghton Street
GB-London WC2A 2AE

Prof. Dr. Franz Palm
Department of Mathematics
University of Limburg
P. O. Box 616
NL-6200 MD Maastricht

Prof. Dr. Jeroen Rombouts
HEC Montreal
3000, chemin de la Cote-Sainte-Catherine
Montreal (Quebec) H3T 2A7
CANADA

Prof. Dr. Rainer von Sachs
Institut de Statistique
Universite Catholique de Louvain
Voie du Roman Pays, 20
B-1348 Louvain-La-Neuve

Richard Song Song
Wirtschaftswissenschaftl. Fakultät
Lehrstuhl für Statistik
Humboldt-Universität zu Berlin
Spandauer Str. 1
10178 Berlin

Prof. Dr. Timo Teräsvirta
Aarhus University
School of Economics and Management
Building 1322
DK-8000 Aarhus C

Prof. Dr. Sebastien Van Bellegem
GREMAQ
Toulouse School of Economics
University of Toulouse
21 Allee de Brienne
F-31000 Toulouse

Prof. Dr. Qiwei Yao
Department of Statistics
London School of Economics
Houghton Street
GB-London WC2A 2AE