DECLARATION FOR THE PhD THESIS
To be inserted as the first page of the thesis

The undersigned

SURNAME	UHL
NAME	BJÖRN
Registration number	1094721

Thesis title:

ESSAYS IN FINANCIAL ECONOMETRICS

PhD in	ECONOMICS
Cycle	XXI
Candidate's tutor	PROF. CARLO AMBROGIO FAVERO
Year of discussion	2010

DECLARES

Under his responsibility:

1) that, according to the President's decree of 28.12.2000, No. 445, mendacious declarations, falsifying records and the use of false records are punishable under the penal code and special laws, should any of these hypotheses prove true, all benefits included in this declaration and those of the temporary embargo are automatically forfeited from the beginning;

2) that the University has the obligation, according to art. 6, par. 11, Ministerial Decree of 30th April 1999 protocol no. 224/1999, to keep copy of the thesis on deposit at the Biblioteche Nazionali Centrali di Roma e Firenze, where consultation is permitted, unless there is a temporary embargo in order to protect the rights of external bodies and industrial/commercial exploitation of the thesis;

3) that the Servizio Biblioteca Bocconi will file the thesis in its 'Archivio istituzionale ad accesso aperto' and will permit on-line consultation of the complete text (except in cases of a temporary embargo);

4) that in order keep the thesis on file at Biblioteca Bocconi, the University requires that the thesis be delivered by the candidate to Società NORMADEC (acting on behalf of the University) by online procedure the contents of which

must be unalterable and that NORMADEC will indicate in each footnote the following information:

- thesis ESSAYS IN FINANCIAL ECONOMETRICS ;
- by BJÖRN UHL ;
- discussed at Università Commerciale Luigi Bocconi – Milano in 2010 ;
- the thesis is protected by the regulations governing copyright (law of 22 April 1941, no. 633 and successive modifications). The exception is the right of Università Commerciale Luigi Bocconi to reproduce the same for research and teaching purposes, quoting the source;

5) that the copy of the thesis deposited with NORMADEC by online procedure is identical to those handed in/sent to the Examiners and to any other copy deposited in the University offices on paper or electronic copy and, as a consequence, the University is absolved from any responsibility regarding errors, inaccuracy or omissions in the contents of the thesis;

6) that the contents and organization of the thesis is an original work carried out by the undersigned and does not in any way compromise the rights of third parties (law of 22 April 1941, no. 633 and successive integrations and modifications), including those regarding security of personal details; therefore the University is in any case absolved from any responsibility whatsoever, civil, administrative or penal and shall be exempt from any requests or claims from third parties;

7) that the PhD thesis is not the result of work included in the regulations governing industrial property, it was not produced as part of projects financed by public or private bodies with restrictions on the diffusion of the results; it is not subject to patent or protection registrations, and therefore not subject to an embargo;

31 JANUARY 2010

BJÖRN UHL

Essays in Financial Econometrics

Dissertation in partial fulfillment of the requirements for the academic degree of Doctor of Philosophy in Economics (XXI cycle).

Björn Uhl (1094721)
Università Commerciale Luigi Bocconi, Milan

Thesis Committee :

Carlo A. Favero, IGIER and Department of Finance, Bocconi University (thesis advisor)
Francesco Corielli, IGIER and Department of Finance, Bocconi University
Vanessa Smith, CFAP and Judge Business School, University of Cambridge

Contents

Acknowledgments

This thesis was written during my graduate studies at Bocconi University, Milan, between 2005 and 2010. This work has benefitted from the advice of several scholars. First and foremost, I thank my supervisor Carlo A. Favero for his critical comments and academic guidance throughout my PhD studies. Moreover, I am grateful for the many discussions, suggestions and feedback of Francesco Corielli, Andrew Harvey, Alessandro Missale, Pietro Muliere, Sze Kim Pang, Sonia Petrone, Vanessa Smith and Claudio Tebaldi.

I would also like to thank the Departments of Economics and Engineering at the University of Cambridge and St. Edmund's College for their hospitality during my reseach stay in the United Kingdom.

Several improvements of parts of this thesis have resulted from discussions with participants of the following workshops and conferences: 10th Workshop on Quantitative Finance in Milan, 2nd PhD Conference in Economics in Athens, 2nd PhD Conference in Economics and Finance at Brunel University, West-London, 15th Young Researchers' Workshop in Statistics of the German Statistical Society in Merseburg and the 5th Nordic Econometric Meeting in Lund as well as two graduate workshops helt at Bocconi University.

I am also grateful for the generous financial support provided by Bocconi University.

Last, but certainly most importantly, I am deeply indebted to my parents. Without their continuous support, encouraging motivation and financial assistance this work would have been never accomplished.

Björn Uhl

Milan, January 2010

Introduction

This introduction outlines the common idea behind the following three essays and gives a brief introduction to each of them. If one abstracts from the specific field of application, all three essays deal with econometric modeling and statistical inference in different empirical financial contexts. Modeling and inference are closely related in the sense that modern statistical inference algorithms allow for more freedom in specifying an empirical model. Typical examples include data augmentation type algorithms that tackle the problem of complicated likelihood maximizations due to unobserved components like stochastic volatility or regime-switching. Another example are simulation-based filtering methods which allow to infer the latent states in nonlinear and non-Gaussian state-space models online. All three essays make use of such innovative econometric techniques. The empirical applications are pairs trading, unbiased estimation of implied volatility surfaces and term structure of interest rates modeling under particular consideration of public debt and time-varying market expectations. The general idea underlying each of the three essays is to provide a novel modeling framework by applying modern (mostly Bayesian) econometric methods. The ideas are diverse though.

The first essay deals with a novel approach to identify pairs trading opportunities. Defining a nonlinear state-space model in the spirit of Newtonian dynamics, an MCMC particle filter is applied to jointly infer the latent group structure and model parameters. If objects are grouped, they establish a center of gravitation which attracts them. The expected direction for future movements is hence towards the group mean. In that essay this idea is exploited to model asset price series which exhibit a tendency to move together even if this equilibrium is temporarily pertubated. Those short-termed spreads can be used to generate statistical arbitrage profits by executing an investment

strategy which is known as pairs trading.

The second essay takes a close look at implied volatility surface modeling. The main difficulty is to provide an estimator that simultaneously avoids either kind of bias induced by an insufficient fit through sparsely parameterized regression models or by potential dents in classical smoothing estimators that arise in areas of the domain of the surface where few, if any, options are observed. The proposed usage of a Bayesian model averaging estimator addresses this issue and does not suffer from neither of these biases. In addition, it does not render the computation of the state-price density as complicated as semiparametric methods do and it provides an easy way to compute pointwise confidence bands.

The third essay is a joint work with Carlo A. Favero. We analyze the effects of economic shocks to the Italian sovereign debt. The yield curve is not viewed independently but in relation to a reference yield curve. The yield spreads are explained by fiscal and monetary policy variables. Around this centerpiece we model the relation between the economic variables. Market expectations are explicitly modeled by a binary Markov-chain which induces two latent regimes.

Pairs Trading and State Space Models

Pairs trading is an investment strategy which could generally be labeled as an attempt to generate statistical arbitrage profits. It has been introduced by Wall Street quants at Morgan Stanley in the 1980s. Pairs in this context are defined as assets that exhibit a long-run equilibrium. Once found, the strategy suggests to exploit temporary deviations from that equilibrium. If two normalized assets are considered both assets trade at (nearly) the same price in the equilibrium. If a spread occured, the rich asset would be shorted and the cheap bought. Unwinding of the two positions when the equilib-

rium is re-established would generate a profit. Jegadeesh and Titman (1995) analyze contrarian investment strategies that exploit such spreads. They find that a possible explanation for the existence of such spreads is non-synchronous trading. This justifies spreads in intraday asset prices but not over several days. Their findings also support the conjecture that the prevailing reason may be an overreaction of investors to the release of individual corporate information that affect one of the stocks.

Methods to find pairs have been proposed in the literature. Many authors suggest to use cointegration analysis to find such assets. That approach can be understood as a linear, Gaussian state space model. The necessary portfolio weights follow from the cointegration vectors.

In the first essay I extend the linear, Gaussian to a nonlinear, non-Gaussian model and demonstrate how this can be used as a very flexible tool to identify pairs trading opportunities. The inference algorithms are a particle filters which allow on-line inference of the latent states. Online inference is paramount in algorithmic trading which could be used to execute pairs trading. This approach will prove to have many advantages. Most importantly, the time series that are considered do not need to be integrated which makes it suitable for volatility series and the like. But it also amplifies the flexibility in modeling the dependence and allows better for managing the risk of the positions in the portfolio.

Since particle filters are not yet among the standard tools in econometrics, this paragraph gives a brief introduction to the main ideas which are required for an understanding of the first essay. In general, a state space model can be summarized by a likelihood function for the observation y_t, a transition for a state vector x_t as well as priors for the initial state and the parameter θ,

$$p(y_t|x_t,\theta), \quad p(x_t|x_{t-1},\theta), \quad p(x_0|\theta), \quad p(\theta).$$

The aim of online inference is to sequentially compute the posterior distribution of the states and parameters, $p(x_{0:t}, \theta|y_{1:t})$. If the model was linear and Gaussian, as in the cointegration case, optimal inference could be achieved with the Kalman (1960) filter, see e.g. Harvey (1989) for an exhaustive review. In nonlinear and non-Gaussian models other techniques may yield better estimates. To sequentially obtain the filtering distribution of the states, particle filters have been introduced by Gordon et al. (1993). Particle filters are named after their particular approximation of the posterior distribution by a set of N_p random draws (called particles) $x_{0:t}^{(i)}$ with associated importance weights $w_t^{(i)}$,

$$\hat{p}(x_{0:t}|y_{1:t}) = \sum_{i=1}^{N_p} w_t^{(i)} \, \delta_{\{x_{0:t}^{(i)}\}}(x_{0:t}).$$

This distribution is recursively updated, that is when new data become available the past particle paths remain unchanged and are merely extended.

The critical question is how new particles are proposed. $p(x_t|y_t, x_{t-1}^{(i)}, \theta)$ is optimal but often difficult to compute. An easy way is to use the state transition which is then called Sample Importance Sampling. This will result in a very inefficient propagation of the particles and consequently in sample impoverishment as only a few particles receive a significant weight. This can be remedied by resampling the particles according to their importance weights which is known as Sample Importance Resampling. However, the problem remains that only a few particles represent the current posterior. More elaborate methods are the Resample-Move particle filter by Gilks and Berzuini (2001) and the Auxiliary particle filter by Pitt and Shephard (1999). The former moves all particles after resampling by a Markov-kernel that is invariant with respect to the filtering distribution. As a result the particle become more diversified without amending the distribution. The latter exchanges the order of resampling and proposing by extending only promising particle paths. Hereto the filtering distribution is decomposed into $p(x_t|y_{1:t}) \propto \sum_{i=1}^{N_p} w_{t-1}^{(i)} p(x_t|x_{t-1}^{(i)}) p(y_t|x_t)$. Setting $g_t^{(i)} \propto w_{t-1}^{(i)} p(y_t|\mu_t^{(i)})$ with

$\mu_t^{(i)} = E(x_t|x_{t-1}^{(i)})$ one samples an N_p indices $j^{(1)}, \ldots, j^{(N_p)}$ with probabilities proportional to $g_t^{(i)}$. The new particles are then proposed from the state transition conditional on the states indexed by $j^{(k)}$.

Yet it has been implicitly assumed the parameter θ is known. In most economic applications this is not the case. Prominent solutions that have been proposed are Storvik (2002) and Polson et al. (2008). Probably the most popular method has been developed by Liu and West (2001) which is to include the parameters in the state vector constraining its mean and variance to be constant. Since the particles for the parameter components in that case would be sampled from the prior and not be refreshed over time the degenerate distribution is regularized by assigning kernels to the point masses. Each kernel is shrunk towards the overall mean and the bandwidth is set as to maintain the original posterior variance.

Implied Volatility Surface Estimation

An inherent feature of financial assets is time-varying volatility. An understanding of an asset's volatility pattern is crucial for any investor. Volatility is an essential ingredient in asset allocation, in pricing derivative contracts and in risk management.

The Black&Scholes option pricing model, which is commonly used to price financial derivatives, can be inverted to retrieve the volatility. This has led to the market practice of quoting option prices in terms of percentage volatility. The implied volatility surface has to be estimated from the options that are actually traded. The data appears in strings along the moneyness axis with different maturities, see figure 1.

It should be stressed that the analysis of IV is restricted to plain vanilla options. It has been shown (e.g. Shaw, 1998) that for more complex derivative securities, such as Asian or Barrier options, the B&S equation cannot be uniquely inverted with respect to the

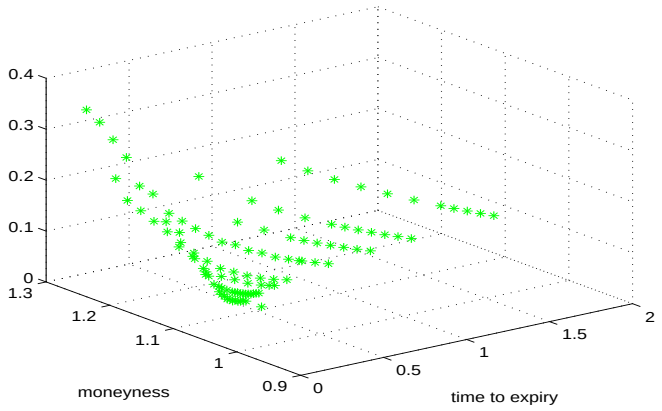

FIG. 1 An example of implied volatility data.

volatility and thus in such case no reliable IV could be inferred from the option prices. The aim of estimating the implied volatility surface can be considered a task of interpolation between observations. In practice, this is important to price over-the-counter options with strike and maturity different from options that publicly traded. The surface is hence an artificial construct. It is believed to be continuous and smooth.

A rich class of estimators for the implied volatility surface have been offered in the literature. Each of which has its advantages and disadvantages. The classical Nadaraya-Watson kernel estimator is defined as $\hat{\sigma}^{BS}(K,T) = \sum_i w_i(K,T)\sigma^{BS}(K_i,T_i)$ with weights w_i proportional to the (bivariate) kernel. This approach has sufficient flexibility to provide an excellent fit. It is easy to compute and implemented in virtually every statistical package. But, if the bandwidth is chosen too small, it can induce dents in areas of the domain where no or only few options are observed. A too large bandwidth smoothes out features of the surface. Furthermore, if for instance the state-price density is of interest, derivatives of the surface with respect to strike and maturity need to be taken which can be numerically imprecise.

More recently a strain of literature has focused on dynamic semiparametric factor models, $\hat{\sigma}^{BS}(K,T) = m_0(K,T) + \sum_l^L \beta_{t,l} m_l(K,T)$ where m_l are smooth basis functions. Yet

another method to interpolate the surface is a simple regression model where the implied volatilities are projected on moneyness and its square as well as on time to maturity. The disadvantage is its limited flexibility and hence insufficient fit with the data. Regression models that include more variables to overcome this problem lack statistical motivation.

The second essay makes use of another statistical approach which is known as Bayesian model averaging. It reduces the model selection to the selection of covariates in a linear regression $y = X\beta + e$. The standard approach is to choose a full specification which contains many regressors. George and McCulloch (1993, 1995) propose to draw submodels by MCMC techniques and then estimate each submodel. These steps are repeated and the estimates averaged over the drawn submodels. For variable selection, one usually defines a selection variable $\gamma \in \{0,1\}^K$ where $\gamma_i = 1$ means that the ith regressor is selected. The selection method is employed and leaves a smaller model which has to be estimated, that is given a model γ a linear regression $y = X_\gamma \beta_\gamma + e$ is estimated. Since the parameterization of the actual model is a priori not known and it is referred to as Bayesian nonparametric modeling. In a sense, the model is strictly-speaking parametric but nonparametric regarding its flexibility and ability to fit the data. The second essay discusses the advantages that this approach offers for estimating the implied volatility surface compared to the aforementioned methods.

Econometric Modeling of the Term Structure and Public Debt

The third essay deals with an econometric model of the Italian term structure, its relation to a reference term structure and the country's public debt. The focus is on the dynamics implied by economic shocks such as monetary policy or government spending shocks. In the literature the issue that there may exist different market expectations regimes and how these may affect the implied dynamics, has not been addressed.

A government will issue bonds to finance its outstanding debt. The yields it has to pay on those bonds depend on the yields on reference bonds. In the Euro-area, markets

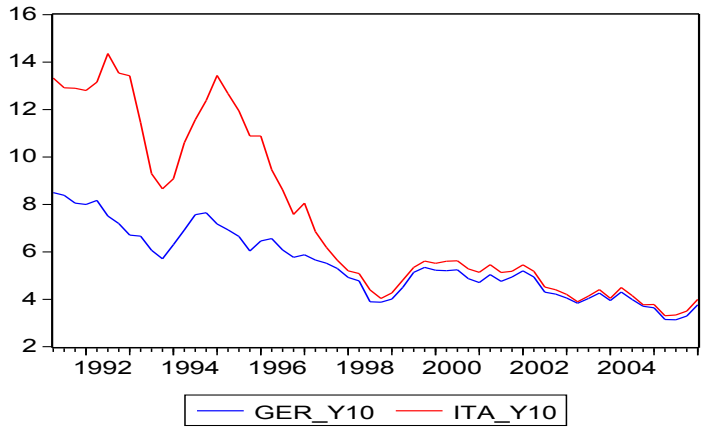

FIG. 2 Yields on German and Italian bonds with 10 years time to maturity.

charge the smallest risk premium on German government bonds and thus we explicitly consider the spread between those two yield curves. In figure 2 the yield time series on German and Italian government bonds with 10 years to maturity are depicted. Towards the EMU period the spread narrows significantly.

The yield differential is composed of three risk factors, which are exchange rate, liquidity and default risk, and is also determined by the bond's tax treatment. With the inception of a common currency, in the EMU regime the exchange rate risk diminishes. Bonds may be less liquidly traded if the country is small or has little public debt. The German and the Italian bond markets are the two largest in Europe and thus liquidity is likely to play a minor role. Consequently, the main focus can be directed towards the default risk. The risk premium hinges crucially on fiscal policy and market expectations. Fiscal policy in this context can be measured by the public debt ratio and its evolution. The effect of fiscal policy can be amplified in either way according to the market expectations about the sustainability of the fiscal policy. If they were pessimistic, they would charge a larger risk premium even if the public debt had not

changed. Expectations are not observed and thus incorporating them into the model is not straightforward. Another challenge is, however, to put all building blocks together, to link them properly and to simulate the model.

Motivated by the decomposition theorem of Wold (1954) that every covariance-stationary time series can be written as an infinite order moving average process, VAR are multivariate parsimonious approximations. They are particularly useful in computing the expected path of the variables if the system is pertubated by a shock. These functions are known as impluse responses. These are important to policy makers as well as to other market participants because they reveal how the projected path of economic variables deviates from their steady-state if a non-anticipated shock occurs. They do, however, not address the issue of instability in the behavioral assumption. See e.g. Favero (2001) for a discussion. Despite this drawback, impulse response functions have become the standard tool in analyzing VAR.

We use VAR to model the dynamics of the Italian macroeconomic variables, the fiscal reaction function and the German yield curve factors. In particular, we analyze the effects of policy shocks on the dynamics of those variables in dependence on the market expectations regime. The focus of our analysis is the dynamics of the Italian public debt.

References

Favero, C. (2001). *Applied Macroeconometrics*, Oxford University Press.

George, E. and McCulloch, R. (1993). Variable selection via gibbs sampling, *Journal of the American Statistical Association* **85**: 398–409.

George, E. and McCulloch, R. (1995). Approaches for bayesian variable selection, *Statistica Sinica* **7**: 339–373.

Gilks, B. and Berzuini, C. (2001). Following a moving target - monte carlo inference for dynamic bayesian models, *Journal of the Royal Statistical Society, Series B* **63**: 127–146.

Gordon, N., Salmond, D. and Smith, A. (1993). A novel approach to nonlinear and non-gaussian bayesian state estimation, *IEEE Proceedings* **140**: 107–113.

Harvey, A. (1989). *Forecasting, Structural Time Series Models and Prediction Theory*, Cambridge University Press.

Jegadeesh, N. and Titman, S. (1995). Overreaction, delayed reaction, and contrarian profits, *Review of Financial Studies* **8**: 973–993.

Kalman, R. (1960). A new approach to linear filtering and prediction problems, *Journal of Basic Engineering* **82**(1): 35–24.

Liu, J. and West, W. (2001). Combined parameter and state estimation in simulation-based filtering, *in* A. Doucet, N. DeFreitas and N. J. Gordon (eds), *Sequential Monte Carlo Methods in Practice*, Springer.

Pitt, M. and Shephard, N. (1999). Filtering via simulation: Auxiliary particle filters, *Journal of the American Statistical Association* **94**: 590–599.

Polson, N. G., Stroud, J. R. and Müller, P. (2008). Practical filtering with sequential parameter learning, *Journal of the Royal Statistical Society, Series B* **70**: 413–428.

Storvik, G. (2002). Particle filters in state space models with the presence of unknown static parameters, *IEEE Transactions of Signal Processing* **50**(2): 281–289.

Wold, H. (1954). *A Study in the Analysis of Stationary Time Series*, 2 edn, Almqvist& Wiksell, New Jersey.

Chapter 1

Particle-Based Clustering of Financial Time Series: A Novel Approach to Identify Pairs Trading Opportunities

Abstract

In this paper we present a novel approach to identifying financial assets which can be used for pairs trading. In contrast to the existing literature, we model the dependence of the considered time series explicitly. Hereto a latent grouping variable is introduced, which indicates membership to a group of assets. If an asset is assigned to a group, it will exhibit a long-run tendency to level with the other group members. The group structure model can be assumed either static or time-varying. To estimate the model in the latter case, we use the MCMC particle filter by Pang, Li and Godsill (2008) and extend it to static parameter learning. In the case of a constant group structure, we employ a regularized auxiliary particle filter. This proves to be computationally less efficient, but it is easily parallelizable. The problem of dimensionality, which occurs in the case that many assets are considered, is discussed and solved by proposing a pairs-only group model. The framework is flexible enough to accomodate more sophisticated strategies such as volatility pairs trading. In an empirical section the methodology is applied to a number of stock and implied volatility time series. It is shown how the algorithm tracks those that are suitable for pairs trading. An investment exercise illustrates how the proposed methods can be used to generate excess returns that are up to three percentage point larger than those obtained by standard techniques.

Key words: Pairs Trading, Online Group Tracking, Sequential Parameter Estimation
JEL classification: G11, C15

1 Introduction

In this paper we propose a novel flexible modeling framework for pairs trading and an inference algorithm which allows jointly to track on-line which assets are suitable for this investment strategy and to estimate the model parameters. As opposed to the current literature, which focuses on standard econometric techniques (see e.g. Gatev et al., 2006), the proposed method allows to model the dependence of time series explicitly and permits real time inference, which is particularly important in algorithmic trading.

Pairs trading is a contrarian investment strategy based on the relative prices of two assets. A comprehensive review of this strategy and traditional methods to identify pairs can be found in Vidyamurthy (2004) and Ehrman (2006). In short, pairs trading assumes that two assets have a joint long-run equilibrium which at some times is pertubated. The resulting spread can be exploited by short selling the rich asset and buying the cheap. Once both have leveled again, one unwinds both positions and thus realizes a profit. The profit depends on the time when the positions are taken because the cash flow occurs already at the beginning of the strategy, and not at the end. This point stresses the main difference to many other quantitative trading strategies. Pairs trading translates the common forecast of an asset's direction into a prediction about its relative position to another asset. That prediction is usually made in a static way. If two assets have been identified as a pair, it is assumed that this property will also apply in the future.

The principle of pairs trading can be generalized to more than two assets. If a group of assets exhibits an equilibrium, a portfolio can be formed that utilizes any occuring spread to generate a profit. Most of the algorithms in this paper are designed for groups of assets. An explicit constraint to pairs can be made to circumvent problems with the dimensionality.

The co-movements of certain assets can be found in numerous markets. In figure 1 a pair of normalized stock prices of two German companies is depicted. A common pattern and temporary deviations from it can be observed. Interestingly, the two companies do not share the same industry and thus will not be found by pure economic arguments. This emphasizes the need for statistical tools that help to find such pairs.

There exist several statistical approaches to identify pairs of assets. A trivial way to

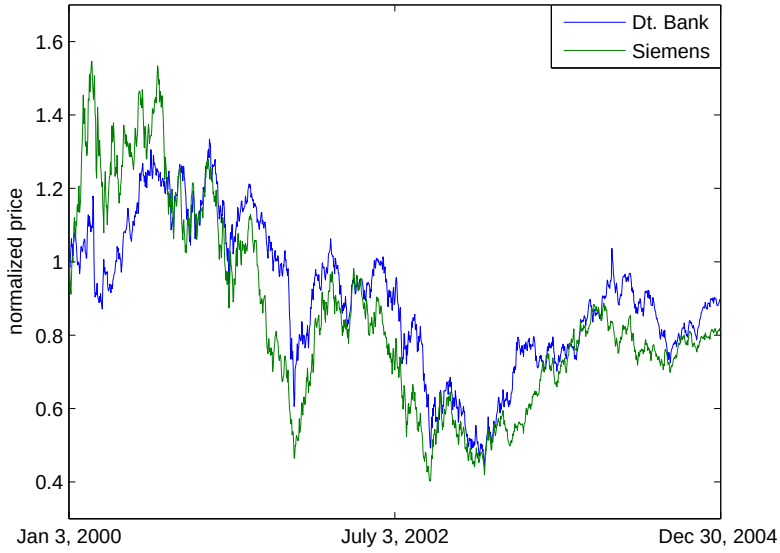

Fig. 1. Pair of Stock Prices: Deutsche Bank AG and Siemens AG from January 2000 until December 2004.

find an asset S_2 from a set $\mathcal{S}$ that forms a pair with another asset S_1 is to choose the one that minimizes the mean squared difference between the historical observations. More precisely, S_2 is assumed a candidate for establishing a pair with S_1 iff

$$S_2 \in \arg\min_{S \in \mathcal{S}} \sum_{t=0}^{T} (S_{1t} - S_t)^2. \tag{1}$$

This methodology has been applied for instance by Gatev et al. (2006) who claim that this approach is preferred by many traders in practice. Their findings support the ongoing success of pairs trading and report estimated excess returns of up to 11% using

(1) to identify pairs.

The definition of pairs leads to a more statistical approach based on time series methods. As described by Bossaerts (1988), the most natural way to model pairs of asset prices is to use cointegration regressions. For example, if only two assets are considered and both of them can be assumed to be I(1), a regression of the form

$$S_{1t} = a + bS_{2t} + v_t \tag{2}$$

is run. If the residual appears to be stationary, any spread that occurs between S_1 and S_2 should be of temporary nature. This procedure can be extended in a straight forward fashion to more than two assets by considering a vector error correction model,

$$\Delta S_t = c + \alpha\beta' S_{t-1} + \sum_{i=2}^{p-1} A_i S_{t-i} + v_t. \tag{3}$$

The estimate of the cointegration matrix β contains the information about the group structure in the data set. In particular, the number of cointegration relationships indicates the number of different portfolios that can be formed to perform pairs trading.

To identify groups of assets, we model the dependence of the time series explicitly conditional on a latent variable that indicates group membership. This provides a very general framework. For instance, in addition to a drift force that pulls the asset prices together, those time series that are assigned to the same group could share a common stochastic volatility component. This allows for a great degree of freedom in modeling. Moreover, in contrary to cointegration, in our approach the time series need not to be integrated in the first place, which becomes particularly relevant if volatility series are considered.

The paper is organized as follows. The second section outlines the statistical model. Some simulation results illustrate typical patterns that the proposed model can generate. In the third section, we present the inference algorithms and illustrate them on simulated data. A pseudo-code of the baseline algorithm is provided in the appendix.

A detailed discussion about the group structure and its dynamics is given. Moreover, possible extensions of the presented framework are discussed. In the fourth section, the proposed methodology is empirically tested by applying it to a number of stock price and implied volatility time series. It is shown how the algorithm identifies groups and in an investment exercise the relative performance of the method is compared to traditional approaches. The final section concludes.

2 Statistical Model

2.1 Asset Dynamics and Dependencies

A set of normalized asset price time series S_{it}, $i = 1, \ldots, N$, over a period $t = 1, \ldots, T$ and their associated returns r_{it} are considered. The latter are modeled as

$$r_{it} = \mu_i - F_{it} + e^{l_{it}/2}\varepsilon_{it} \tag{4}$$

with individual stochastic log-variances l_{it} which evolve according to

$$l_{it} \sim N(\alpha_i + \beta_i l_{it-1}, \sigma_i^2), \quad l_{i0} \sim N\left(\frac{\alpha_i}{1-\beta_i}, \frac{\sigma_i^2}{1-\beta_i^2}\right). \tag{5}$$

Thus far the time series are independent. To model the dependence between the time series a latent group variable $G_t = (g_{1t}, \ldots, g_{Nt}) \in \{1, \ldots, N\}^N$ is introduced which assigns every asset to a group. $\Lambda(G_t, d) = \{i : g_{it} = d\}$ denotes the index set of members of group d and $\lambda_{it} = \lambda(G_t, g_{it}) = \#\Lambda(G_t, g_{it})$ is the number of members in the group of asset i. By n_t the number of different groups is denoted. Note that a group may consist of a single member. Essentially, if $G = (1, \ldots, N)$ all assets are independent and the model reduces the N univariate stochastic volatility models.

The force that connects the returns of group members is F_{it} in (4). The mean return of

an asset, μ_i, is reduced by the group-correction term F_{it}. If asset i trades above the other group members, its mean return is reduced and thus it pulls the asset price towards the group mean of all members. Consequently, the force can be defined as to satisfy $S_{it}F_{it} = \eta_i(S_{it} - N^{-1}\sum_{j=1}^{N} S_{jt})$ or

$$F_{it} = \eta \left(1 - \frac{1}{\lambda_{it}\, S_{it}} \sum_{j \in \Lambda(G_t, g_{it})} S_{jt}\right) \tag{6}$$

with parameter $\eta > 0$ that controls the velocity of the adjustment and is equal for all group members.

The best way to understand the choice of F_{it} is to consider the expected asset dynamics

$$E(S_{it+1}|S_{it}) = (1+\mu_i)S_{it} + \eta_i(\lambda_{it}^{-1} \sum_{j \in \Lambda(G_t, g_{it})} S_{jt} - S_{it}).$$

The expected time-$t+1$ price is a mixture of the current asset price compounded by its mean return and a term whose sign depends on the relative location of the assets. Other functions for the drift force can be chosen (see Pang, 2009, for a discussion).

It is essentially a kind of mean reversion to a stochastic mean that is determined by all targets in the same group. The idea is similar to an Ornstein-Uhlenbeck-type process with a time-varying mean, which is frequently applied to model short-rate dynamics.

The instantaneous covariance of the drift forces of two group members i and j is given by

$$\operatorname{cov}(F_{it}, F_{jt}) = (\eta/\lambda_{it})^2 \sum_{k \in \Lambda(G_t, g_{it})} \sum_{l \in \Lambda(G_t, g_{jt})} \operatorname{cov}\left(\frac{S_{kt}}{S_{it}}, \frac{S_{lt}}{S_{jt}}\right).$$

Writing $\Delta S_{it} = S_{it} r_{it}$ and substituting (4) for r_{it} it can be seen that the model can be understood as a multivariate Euler-discretized diffusion process which in the continuous time limit, $\Delta t \to 0$, satisfies

$$dS_i(t) = S_i(t)\left[(\mu_i - F_i(t))\, dt + \exp(l_{it}/2)\, dW_i(t)\right],$$

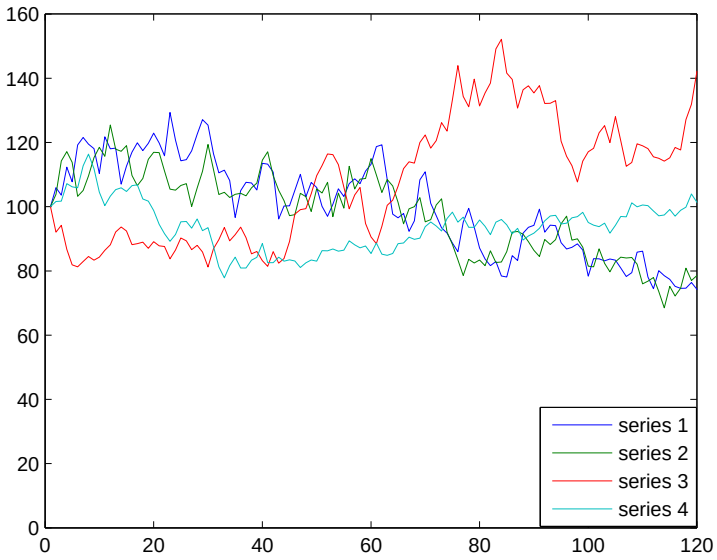

Fig. 2. Simulated time series with $G = (1, 1, 2, 3)$.

where $W_i(t)$, $i = 1, \ldots, N$, are independent standard Brownian motions.

Figure 2 illustrates the pattern of time series that can be generated by (4) and (5). In this example, the group structure is constant and equal to $G = (1,\ 1,\ 2,\ 3)$ which corresponds to group structure 5 (see appendix B). The first and second asset constitute a pair whereas the third and fourth behave individually. The time horizont is $T = 120$. The strength of the reversion towards the group mean is set to $\eta = 0.4$.

2.2 Group Dynamics

The general idea behind clustering in state space model is to introduce a latent variable G, which assigns each object to a group. In table B in the appendix, all possible group structures for four targets can be found. For identifiability the first target is always labeled 1, the second 1 or 2, and so on. This normalization has to be applied whenever a new group structure is drawn. For instance, if $G = (4, 1, 3, 1)$ has been sampled it needs

to be transformed into $(1, 2, 3, 2)$. Consequently, $n_t = \max G_t$.

It can be either assumed that the group structure of the objects is constant or that it may evolve over time resulting in a stochastic process $(G_t)_{t=1,\dots,T}$. The former may intuitively seem more appealing because an investor will not execute the strategy if the group structure can suddenly change. The latter can be justified by understanding that it also allows for staying in the same group structure. Short-termed deviations that stem from a temporary disequilibrium will not thward finding the correct group structure. Moreover, a time-varying group structure yields more significant results because its likelihood function has a larger degree of freedom. In the sequel, we study both approaches.

The inference algorithms are different but both are sequential Monte Carlo methods (see Doucet et al., 2001, for a review). For the static group structure model we apply the standard auxiliary particle filter (Pitt and Shephard, 1999) which can be easily extended to joint state and parameter learning (Liu and West, 2001). In case of a dynamic group structure we use an MCMC particle filter (Pang et al., 2008) and extend it by the aforementioned regularization to allow for joint state and parameter estimation.

To model the dynamics of the group structure, a distribution for the group transition, $p(G_t|G_{t-1})$, needs to be specified. It proofs to be convenient to introduce an auxiliary variable π_t which captures the distributional characteristics of G_t and hence to consider the joint dynamics $p(G_t, \pi_t|G_{t-1}, \pi_{t-1})$. The first to introduce this kind of group transition model using Dirichlet distributions were Wei et al. (2007). We use a version that has been further refined by Pang (2009). The probability of being in a certain group k is $p(g_{it} = k|\pi_t) = \pi_{t,k}$. These probability weights are random and Dirichlet distributed with base distribution α_t, $\pi_t \sim \mathcal{D}(\alpha_t)$.[1] The base parameter of the Dirichlet distribu-

[1] There exist different parameterizations of the Dirichlet distribution. We use a single parameter version where the base distribution is not required to be normalized. The normalizing constant determines the reciprocal of the variance.

tion, α_t, evolves over time as a mixture of the previous realization of the probability weights and the initial parameter α_0,

$$\alpha_t = \xi\alpha_{0,C}\pi_{t-1} + (1-\xi)\alpha_0.$$

The weights $\alpha_{0,i}$ initialize the prior distribution and $\alpha_{0,C} = \sum_i \alpha_{0,i}$ sets the variance of the initial distribution. The autoregressive parameter ξ controls the dependence of the dynamic Dirichlet distribution.

3 Sequential Inference

In this section the inference algorithms for both dynamic and static group learning are presented. This is followed by a proposal for solving the curse of dimensionality in this context. Moreover, it is demonstrated how the algorithm can be used to cluster assets according to their volatilities. The section concludes with a check for the robustness of the algorithm against false pairs.

3.1 Dynamic Group Learning

The model for the group structure is now assumed dynamic, $(G_t)_{t=1,\dots,T}$, and the transition $p(G_t, \pi_t|G_{t-1}, \pi_{t-1})$ is given by the dynamic Dirichlet distribution as introduced in the previous section.

3.1.1 MCMC Particle Filter

There exists a number of particle filters which use Markov chain Monte Carlo (MCMC) methods. The first to introduce MCMC in a sequential setting were Gilks and Berzuini (2001) who apply a Markov kernel, that is invariant with respect to the current posterior distribution, to the resampled particles in a standard Sampling Importance Resampling algorithm. This step further diversifies the location of the particles. To accomodate group dynamics efficiently in this context, Khan et al. (2005) consider the joint posterior of $t-1$- and t-states. They use an MCMC scheme to compute the integral in

$$p(Z_t|S_{1:t}) = \int p(Z_t, Z_{t-1}|S_{1:t})\, dZ_{t-1}$$

where $S_{1:t} = (S_1, \ldots, S_t)$ and $Z_t = (G_t, \pi_t, l_t)$ is the full set of states. This is computationally demanding because in every time step the predictive distribution has to be computed. Pang et al. (2008) propose a more efficient MCMC particle filter that requires less computational power. It is similar to the Resample Move particle filter but it differs in that the particles are not moved after resampling to achieve a larger number of distinct particles but the MCMC algorithm is used to propose new particles. Hereto a number of Metropolis-Hastings iterations are performed in a sequential fashion. The joint filtering distribution

$$p(Z_t, Z_{t-1}|S_{1:t}) \propto p(Z_t, S_t|Z_{t-1}, S_{t-1})\, p(Z_{t-1}|S_{t-1}). \tag{7}$$

is considered as the target distribution for the MCMC algorithm. This yields a sequence of draws $(Z_t^{(m)}, Z_{t-1}^{(m)})$. A burn-in sample is discarded from the MCMC output to ensure convergence and the remaining chain is thinned to reduce the autocorrelation between drawn particles. Only the first component $Z_t^{(p)}$ is kept,[2] which can be considered as a draw from the marginal distribution of (7). This distribution coincides with the desired

[2] The distinction between the indices m and p is made to distinguish iterates from the MCMC scheme and those draws that are kept as particles. Hence, $m = 1, \ldots, N_{MCMC}$, $p = 1, \ldots, N_p$

filtering distribution $p(Z_t|S_{1:t})$. A comparison of some variants of this algorithm can be found in Pang (2009, chapter 4).

All those algorithms allow for state filtering given parameters. In economic and financial practice, model parameters are unknown and need to be estimated. During the last decade, a number of methods to estimate parameters in a state-space framework online have been proposed, see Kantas et al. (2009) for a recent review. We implement the (static) parameter learning procedure originally proposed by Liu and West (2001) which uses a regularization of the parameter distribution and thereby generates an artificial parameter evolution.

Including the parameter vector θ in the target distribution (7) one obtains

$$\begin{aligned} p(Z_t, Z_{t-1}, \theta_t, \theta_{t-1}|S_{1:t}) \propto\, & p(S_t|G_t, l_t, \theta_t, S_{t-1})\, p(l_t|l_{t-1}, G_t, \theta_t)\times \\ & p(\theta_t|G_t, \theta_{t-1}, S_{1:t-1})\, p(G_t, \pi_t|G_{t-1}, \pi_{t-1})\, p(\theta_{t-1}, Z_{t-1}|S_{1:t-1}). \end{aligned} \tag{8}$$

The time-subscript on the parameters emphasizes the information set it is conditioned upon, that is, a particle θ_t approximates the distribution of θ given $S_{1:t}$. In other words, through its artificial evolution the estimator is time-varying whereas the estimand is not.

All terms that appear in the joint filtering distribution (8) can be easily evaluated and sampled from. The first term

$$p(S_t|G_t, l_t, \theta_t, S_{t-1}) = \prod_{d=1}^{n_t} p(S_{\Lambda(G_t,d),t}|l_t, S_{\Lambda(G_t,d),t-1}, \theta_t) \tag{9}$$

is a product of normal distributions

$$\begin{aligned} & p(S_{\Lambda(G_t,d),t}|l_t, S_{\Lambda(G_t,d),t-1}, \theta_t) \\ & \quad = N(S_{\Lambda(G_t,d),t-1}(1 + \mu_{\Lambda(G_t,d)} - F_{\Lambda(G_t,d),t-1}), S^2_{\Lambda(G_t,d),t-1} e^{l_{\Lambda(G_t,d),t}}) \end{aligned}$$

and $N_p = \frac{N_{MCMC} - N_{BurnIn}}{thin}$ where $thin$ is a thinning constant.

The transition of the log-variances, $p(l_t|l_{t-1}, G_t, \theta_t)$, is independent of the group structure. Also this distribution takes the form of a product of normal distributions

$$p(l_t|l_{t-1}, G_t, \theta_t) = \prod_{i=1}^{N} p(l_{it}|l_{i,t-1}, \theta_t) = \prod_{i=1}^{N} N(l_{it}|\alpha_i + \beta_i l_{i,t-1}, \sigma_i^2).$$

The term $p(\theta_t|G_t, \theta_{t-1}, S_{1:t-1})$ is the regularized distribution of the model parameters. It is normal with mean equal to the previous parameter particles shrunken towards their overall mean. The covariance equals the estimated covariation between the parameter particles at the previous time step multiplied by a factor that depends on the strength of shrinkage in the mean specification. This procedure maintains the first and second moments of the time-$t-1$ parameter particle approximation.

The regularization of the parameter particle distribution with normal kernels makes sense only if all parameter components have the whole real line as domain. Therefore we transform the parameters before updating their particles and vice versa before they are used. Standard deviations are considered in logarithms. We impose stationarity of the scedastic equation which has been justified e.g. by Jacquier et al. (1994). Also to our best knowledge, no empirical study has ever found a negative autoregressive coefficient in the stochastic log-variance equation. Hence, we transform β logistically and thereby constrain it on $(0, 1)$. When transformed on that domain, all parameters are assigned normal priors.

The term $p(G_t, \pi_t|G_{t-1}, \pi_{t-1})$ is the dynamic Dirichlet distribution as discussed above. The last term can be approximated by the previous set of particles

$$\hat{p}(\theta_{t-1}, G_{t-1}, \pi_{t-1}, l_{t-1}|S_{1:t-1}) = (1/N_p) \sum_{p=1}^{N_p} \delta_{\{(\theta_{t-1}, G_{t-1}, \pi_{t-1}, l_{t-1})^{(p)}\}}(\theta_{t-1}, G_{t-1}, \pi_{t-1}, l_{t-1}).$$

Finally, it should also be noted that in the current context the model parameters are not of particular interest *per se*. They could be labeled nuisance parameters since the only interesting result for a trader, who looks for suitable pairs, is the estimated group structure. However, obtaining precise parameter estimates will facilitate the algorithm

to track the 'true' group structure.

3.1.2 Simulation Experiment

We simulate a process from the (4) and (5) using $G = (1, 1, 2, 3)$, and run the algorithm on the simulated data. For the MCMC algorithm 13 000 iterations are used, a burn-in of 1 000 iterations is discarded and the chain is thinned by every other, resulting in 6 000 particles. The results are displayed in figure 3.

The algorithm finds the correct group structure 5 after an initial learning period of about 30 observations. Pairs, as shown in the right hand panels, are computed on the basis of the estimated group structure. [3] In practice, such strong evidence cannot be expected. This point results from the fact that when the algorithm is applied in an empirical context, the data generating process will differ from the one that is embedded in the algorithm. Moreover, the filtering results are not to be understood as a general statement about the distribution of the estimate. We run the filter only on a single realization. To obtain an estimate of the distribution of the filtering results, a Monte Carlo study would be necessary. Hereto a large number of realizations of the error term needed to be simulated and the algorithm to be run on each of the realizations. However, on a typical Intel-machine with 2 GHz a Matlab program that runs the filter on four assets using $13\,000$ MCMC iterations takes about 320 seconds per update and thus on a sample with 120 observation per series more than 10 hours. Consequently, a full Monte Carlo study with only $1\,000$ replications will take more than a year.

[3] It should be noted that only those groups are counted as pairs that have exactly two members. For instance, $G_t = (1, 1, 1, 2)$ contains no pair whereas $G_t = (1, 2, 2, 1)$ has two pairs. For first group structure uses three assets to compute the drift force which in general will not coincide with the corresponding pair means.)

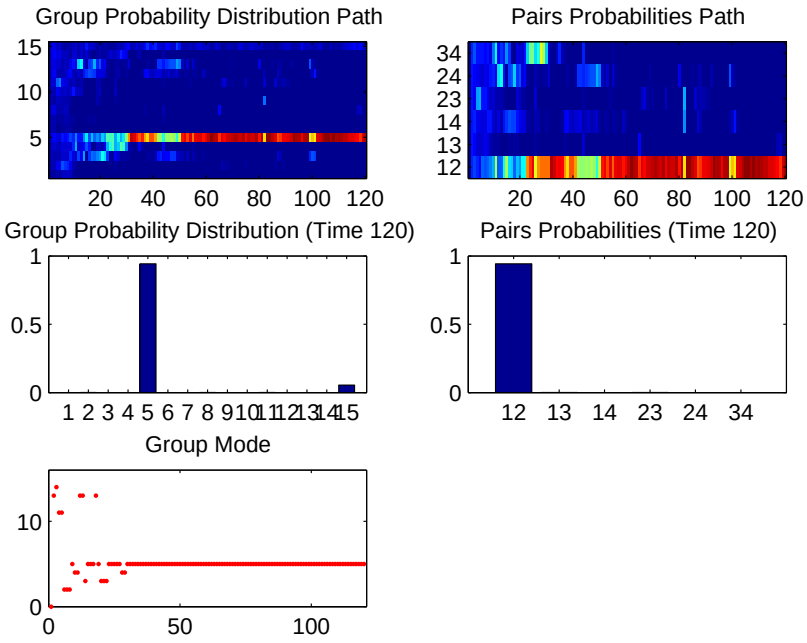

Fig. 3. Results with Drift-Only Clustering Algorithm.

3.2 Static Group Learning

Generally, the sequential estimation of static parameters in a nonlinear state space model imposes a number of difficulties. If the parameters are treated as states the most severe obstacle is that once the particles are sampled from the prior they are never updated but only their importance weights. If prior and posterior have different areas of concentration, the particles estimate might be biased. Fortunately, the group structure distribution is discrete and has only a finite number of possible realizations. Thus, when sampling from the prior, one needs to ensure that a sufficient number of particles are placed on each of the possible group structures. The updated importance weights yield an unbiased particle representation of the time-t posterior of the groups.

3.2.1 *Auxiliary Particle Filter*

The algorithm to infer a static group structure is simply the APF for a given initial group structure G. The joint time-t posterior of states and parameters is decomposed into

$$p(Z_t, \theta | S_{1:t}) \propto p(S_t | l_t, G, \theta, S_{t-1})\, p(l_t | G, \theta, S_{1:t-1})\, p(\theta | G, S_{1:t-1})\, p(G | S_{1:t-1}) \tag{10}$$

where $p(l_t | G, \theta, S_{1:t-1}) = \int p(l_t | G, \theta, l_{t-1})\, p(l_{t-1} | G, \theta, S_{1:t-1})\, dl_{t-1}$.
Since G does not evolve over time the particle filter does not reallocate particles between group structures. Therefore it is important that the prior on G allocates a sufficient number of particles to every possible group structure. A non-informative prior on the group structure G is to assume independent uniform distributions, that is $p(G) = \prod_{i=1}^{N} p(g_i)$ where $p(g_i) = U(\{1, \ldots, N\}) = N^{-1} 1_{\{1,\ldots,N\}}(g_i)$ which assigns every object independently to a group with equal probability. In practice, one will start with an equal number of particles for each group structure.
When time-t data become available, for each of the series the corresponding drift force is computed and the log-variance predicted. The likelihood takes the same form as in the dynamic group learning case (9). Hence, the likelihood ratios that determine the importance weights are easy to evaluate. Particles are resampled within each group structure, which prevents less likely group structures from not receiving any particles. This would hinder them to load up any probability when it should receive a higher posterior probability. Summing all normalized importance weights within each group structure yields its posterior probability.

3.2.2 *Simulation Experiment*

The results for the APF for static group learning are displayed in figure 4. We use $15 \cdot 4\,000 = 60\,000$ particles. The true group structure 5 receives the largest posterior

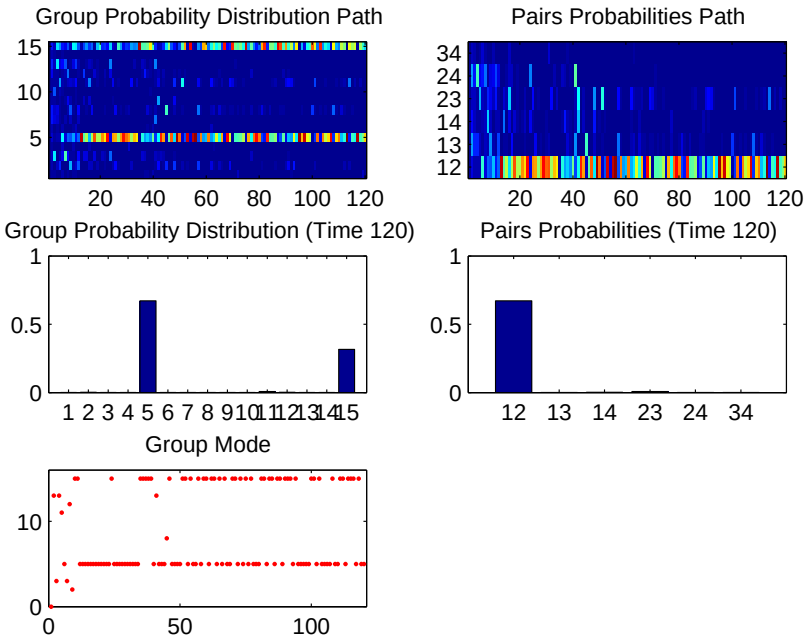

Fig. 4. Filtering results for the auxiliary particle filter.

probability. All other pairs possibilities have negligible probabilities. Apart from some noise, when the algorithm does not indicate the true group structure, it finds none, that is $\hat{G} = (1, 2, 3, 4)$, which can be interpreted as being conservative on investments. No grouping will support an investor's decision of not executing pairs trading. By contrast, finding an incorrect group structure would be hazardous for the investor.

This algorithm clearly requires a larger number of particles. The reason is that the number particles of very unlikely group structures is maintained as the algorithm does not allow for cross-model jumps.

3.3 Dimensionality and Pairs Dynamics

The algorithms of the previous two sections cluster time series into groups of arbitrary size. A problem, which will occur in financial practice, is the rapidly growing num-

ber of possible group structures. For N assets the number of possible group structures is the Nth Bell (1934) number which is recursively defined by $B_N = \sum_{n=0}^{N-1} \binom{N-1}{n} B_n$, $B_0 = 1$. For only ten assets there exist $B_{10} = 115\,975$ many group structures. This imposes an excessive number of particles. In practice, many assets will be considered so that numbers such as $N = 100$ will not be uncommon. This maligns the proposed method to be computationally infeasible for such practical purposes. If only pairs rather than general groups are considered the dimensionality can be reduced.

If only goups of at most two members, then named pairs, are considered the magnitude of the possible dependencies reduces from exponential to quadratic. For instance, if $N = 100$ only $(N-1)N/2 = 4\,500$ possibilities for pairs exist. To circumvent any problems from possible redundancies in the pairs structure, each asset can be assigned to at most one pair for every particle. Effectively, the reduction of the dimensionality comes at the cost of less generality. This becomes particularly relevant if it can be expected that more than two assets exhibit a grouping behavior as will be the case, for instance, if fixed income instruments are considered.

Let $\mathcal{P} \subset \{1, \dots, N\}^2$ denote the set of all possible pairs. We associate every pair with a number in the following order: $(1, 2)$ is always 1 and $(N-1, N)$ is $H := (N-1)N/2$. The mapping f_1, that maps $p = (p_1, p_2) \in \mathcal{P}$ to that number, is given by

$$\begin{aligned} f_1(p) &= \sum_{i=1}^{p_1} (N - i) - (N - p_2) \\ &= (p_1 - 1)N - \frac{p_1(p_1 + 1)}{2} + p_2. \end{aligned} \tag{11}$$

Similarly, if $\mathcal{I}$ denotes the power set of $\mathcal{P}$, then for $I \in \mathcal{I}$ define $f_2(I) = \sum_{i \in I} 2^{i-1}$, which is the unique number associated with this pairs structure. Every group structure G_t has the same dimension $N \times 1$ whereas the dimension of pairs is $m_t \times 2$ where $0 \leq m_t \leq H$. Hence, when proposing the new pairs structure, also the number of pairs has to be proposed. To this end, we specify a mixture distribution of a uniform distribution over all numbers and a point distribution in the previous number m_{t-1}. The autoregressive

parameter α_m controls the dynamics of the number of pairs whereas α_p determines how likely it is to propose the same assets that have been part of a pair in the previous time step. The inverse of f_1 can be easily sequentially computed by identifying the first component using its possible range and the second component by finding the difference in (11).

Summing up, given hyperparameters and a $t-1$ pairs structure (P_{t-1}, m_{t-1}) the new pairs structure is proposed in the following way

(1) $P(m_t = j | m_{t-1} = i) \propto \exp(-\alpha_m |j - i|), \quad j \in \{1, \dots, H\}$

(2) define $N_l = 1(\exists\, i : \; P_{i,t-1} = l\}$, $l = 1, \dots, N$

(3) create weights $w_l = 1 + \alpha_p N_l$

(4) for $i = 1 : m_t$ draw pair members:

$P(P_{it}^1 = l | P_{t-1}) \propto w_l$ and $P(P_{it}^2 = l | P_{t-1}) \propto w_l$, $l \in \mathcal{L}$

$\mathcal{L}$ is the set of indices, that is, initially, $\mathcal{L} = \{1, \dots, N\}$. Once P_{1t}^1 is drawn $\mathcal{L} = \{1, \dots, N\} \setminus \{P_{1t}^1\}$ and so on. The MH acceptance probabilities are analogous to A_1 to A_3 in $(\mathrm{A}.1)$ - $(\mathrm{A}.3)$ of the baseline algorithm.

As before, we demonstrate the pairs tracking algorithm on a simulated dataset with $N = 4$ assets with a single pair $(1, 2)$ which previously has been referred to as group 5 or $G_t = (1, 1, 2, 3)$. The results are shown in figure 5. It takes about 35 periods until it indicates the true pair and from about the 90th period the correct pairs structure is found with a posterior probability of mostly more than 50%. The only competing model that receives a significant weight is the model where all assets are independent. Again, this is not virulent because it will yield a more conservative investment rule.

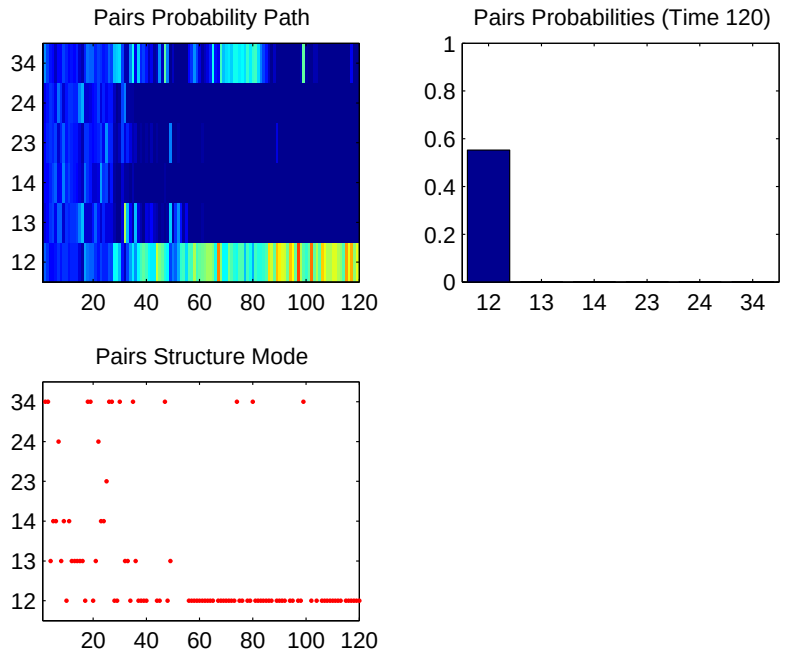

Fig. 5. Filtering results for the dynamic pairs structuring model for the pair $(1, 2)$.

3.4 Model Extensions and Robustness

3.4.1 Implied Volatility Pairs Trading

Pairs trading as described in the introduction exploits temporary deviations of an asset price from an equilibrium that is defined by its group members. Another way to conduct pairs trading is to consider an asset's volatility rather than its price.

In practice volatility pairs trading is executed by taking positions in derivative contracts on the underlying pairs (see e.g. Mougeot, 2005). The Vega of a European option on a stock is positive and given by $K\, e^{-r\tau}\phi(d_2)\sqrt{\tau}$ where ϕ denotes the pdf of the normal distribution and K the strike price. Hence, an option on a stock with a larger volatility will trade at a higher price. A spread in the volatility can thus be exploited by taking a long-short position in the respective options. For instance, if assets 1 and 2

establish a volatiltiy pair and if $\sigma_{S_1} > \sigma_{S_2}$ then the strategy is to buy a S_1 and sell a S_2 call or put option with equal moneyness and equal maturity. It should be noted that the other 'greeks' have to be hedged so that the only effect on the portfolio stems from the volatility (see e.g. McMillan, 2002).
Using the framework for dynamic and static group learning, volatility pairs trading means no challenge and it works just as for stock prices. It should be noted, however, that with cointegration techniques this task could not be addressed since implied volatilities are usually not found to be integrated and consequently cannot be cointegrated.
In appendix C a preliminary model to cluster assets with respect to their latent stochastic volatility is presented.

3.4.2 Alternative Contrarian Investment Strategies via Group Tracking

Contrarian investment strategies build upon the empirical observation that some stock prices exhibit a systematic pattern in relation to some other stocks. A typical example is a lag-1 cross-correlation between stock returns. This may be due, for instance, to nonsynchronous trading (see Tsay, 2002, chapter 5).
Lo and MacKinley (1990) discuss another case of contrarian investments which exploits the lead-lag-effect frequently found in some pairs of stocks. They find that some trade reactions on some stocks are faster than on others. This observation, which in effect is similar to non-synchronous trading, induces a stock price behaviour, which is known as lead-lag-effect. For instance, if good news is made public, some stock prices may increase before others. Thus a strategy that exploits this effect, could be to buy (short) an asset if its leader has increased (decreased) previously. What before has been the strengh of the reversion to the group mean, η, is here the length of the lag-period.

3.4.3 Robustness to False Pairs

If a potential pair is not found, it will not be exploited and no arbitrage profit will be realized. However, finding a pair that actually exhibits no long-term tendency to close the spread is virulent because it potentially generates losses.
A typical false pair will exhibit highly correlated returns but without any force to level. As such, we simulate a normally distributed return series

$$r_t \overset{iid}{\sim} N(\mu, \Sigma). \tag{12}$$

The false pairs are assigned a correlation of $\rho_{12} = 0.9$ and the other correlations are set to zero. A return series is simulated from (12) for $T = 120$ and the drift-only model with dynamic group learning is run on that simulated dataset. The results are displayed in figure 6. The algorithm initially finds pairs but for less than 20 periods. This occurs as a result of temporaneous comovements between the two prices of the 'fake pairs'. After the 75th observation the algorithm clearly indicates that there exists no pair in the dataset. Estimating the VECM (3) on the simulated data yields that the null of no cointegration between S_1 and S_2 is not rejected at 5% but at 10% with the maximum Eigenvalue test. Consequently, it appears that neither method finds the false pair. The rejection by standard cointegration seems, however, less convincing.

4 Empirical Analysis

In the previous sections we have presented alternative models and corresponding inference algorithms to identify pairs. On simulated data all algorithms have proven to work well. In this section we apply the statistical framework to real data. First, we demonstrate that we are able to track pairs both in price and in implied volatility series. Second, in an investment exercise we show how the output of the algorithm can

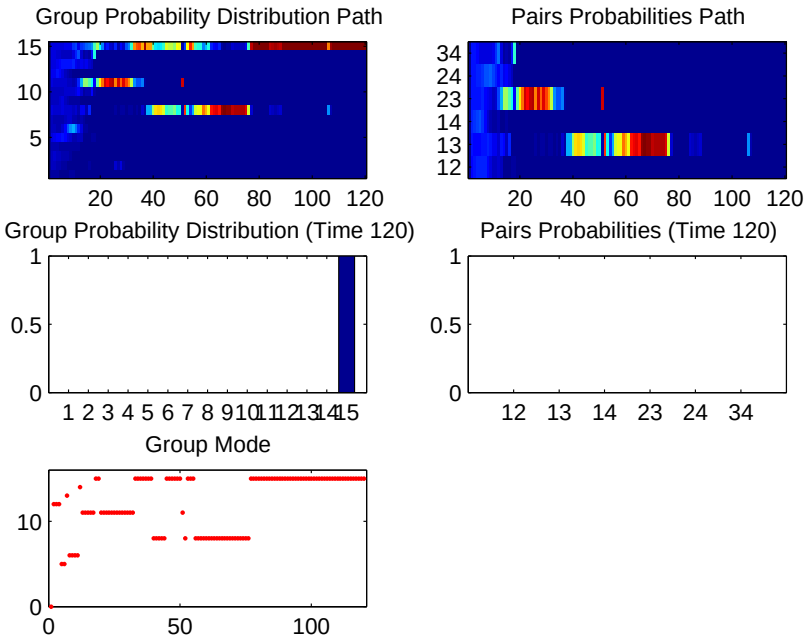

Fig. 6. Filtering results for dynamic group structure model with a simulated false pair [1 2].

be used for executing pairs trading and how the excess returns compare with the traditional approaches.

4.1 Identifying Pairs

In each of the following two examples three time series are considered. We choose the series as to have one pair and an obviously independent series.

4.1.1 Example: Pairs of Asset Prices

We consider three stock price series for the period from December 4, 2006 until April 20, 2007 sampled at a daily frequency. All prices are taken at market closing. The three stocks are of the two Italian banks Banca Intesa SanPaolo and Banca UniCredit, as well

Fig. 7. Stock price series of Banca Intesa SanPaolo (Ita), Banca UniCredit (Ita) and Google Corp (US).

as the US internet and software corporation Google. These are chosen since it can be economically justified that the first two stocks might be a pair since these two companies share the same industry, face the same market conditions and their stocks are traded on the same stock exchange. The normalized stock prices series are depicted in figure 7. By purely visual inspection it seems indeed reasonable to assume that Intesa and UniCredit establish a pair over the whole sample period.

We use the basic clustering algorithm with time-varying group structure. All priors are empirical Bayes. After the transformation, each parameter is assigned a normal distribution. We use a pre-sample period of six months. The means of the drifts equal the mean returns over that period. For the stochastic volatility component, we center the prior of the autoregressive term on 0.9 and the intercept so that the long-run log-variance $\alpha/(1-\beta)$ equals the pre-sample estimate of the variance for each of the series. The prior of the volatility of the error term in the log-variance equation is centered at

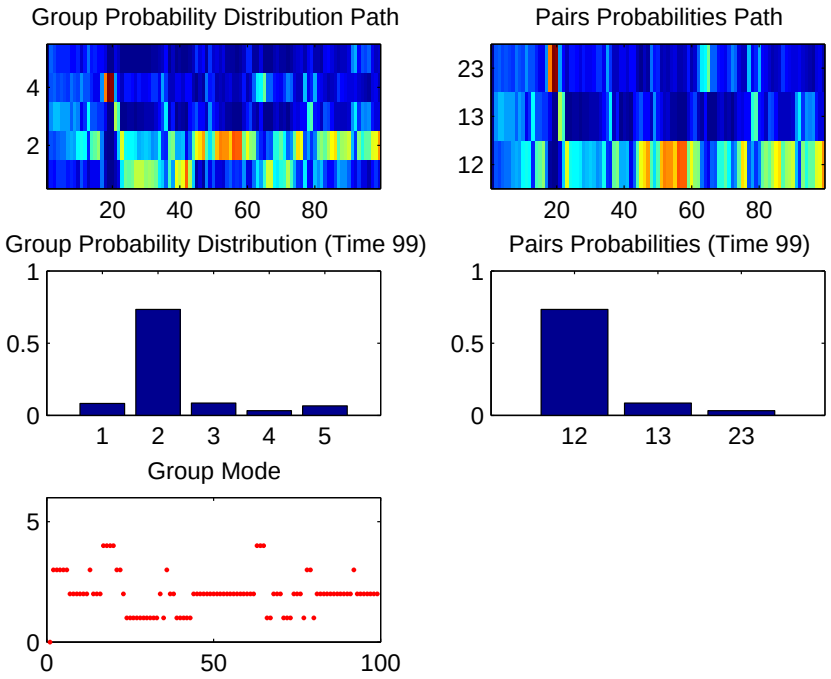

Fig. 8. Empirical results for the three stocks: Intesa SanPaolo, UniCredit and Google Corp.

0.1. For η we use a uniform prior distribution over $[0, 1]$.

Aside from some noise, from the 40th observation on, the algorithm indicates that assets 1 and 2 are likely to establish a pair.

4.1.2 Example: Pairs of Implied Volatilites

In this example we consider three German companies: Bayrische Motoren Werke (BMW) and Daimler Benz are automotive producers and Hypo Real Estate (HRE) is a mortgage bank. All series are sampled daily from January 1, 2009 until May 20, 2008 which, as before, corresponds to 100 trading days. The series are depicted in figure 9. Through the credit crunch turmoil the implied volatilities of all three stocks are large but even more extreme for the HRE. Both, economically and from a pure visual inspection of the data, it is clear that the first two volatility series establish a pair whereas the third

should be independent of the former. [4]

The filtering results are displayed in figure 10. It takes only a few periods until the

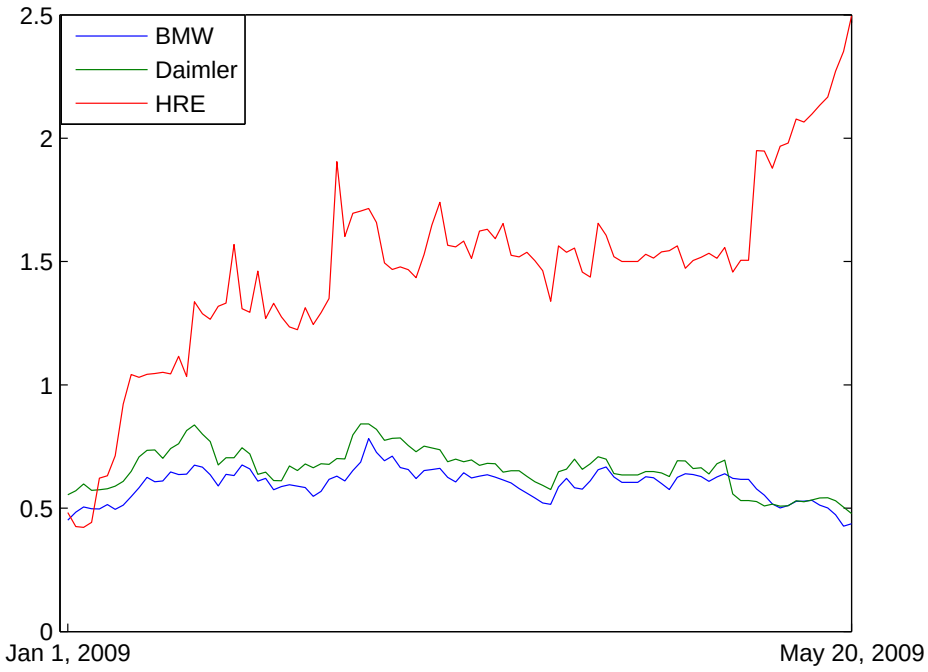

Fig. 9. At-the-money implied volatility time series of BMW, Daimler and HRE.

algorithm identifies the BMW - Daimler pair. The only competing group structure that receives a significant weight is 5, that is all volatilities are independent.

4.2 Investing in an Identified Pair

Suppose a pair has been identified as demonstrated in the previous subsection. The excess returns on executing the pairs trading strategy depend not only on which assets

[4] The BMW-Daimler pair has been discussed in Mougeot (2005) and is a typical example of an implied volatility pair.

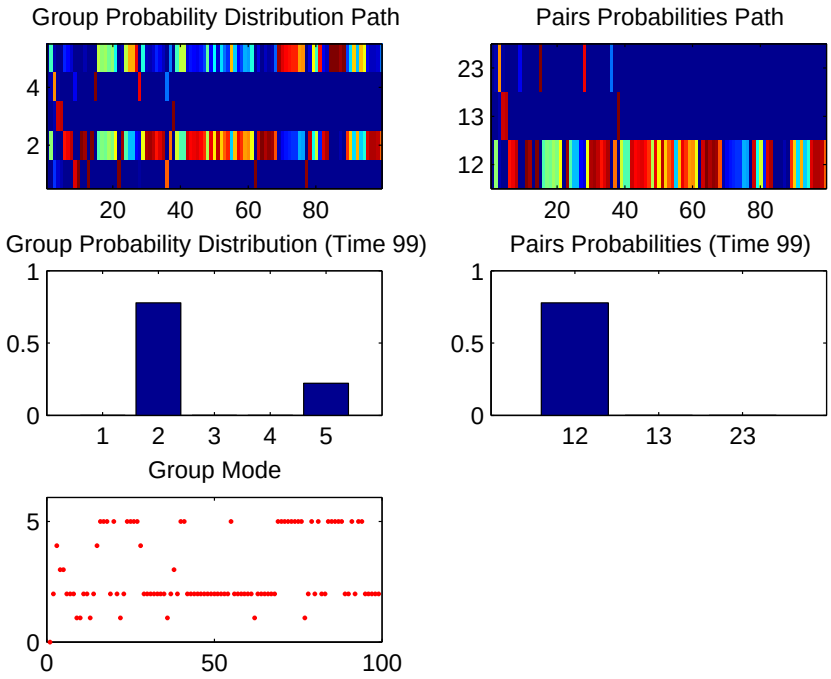

Fig. 10. Empirical results for the ATM-IV series of BMW, Daimler and HRE.

have been selected as a pair but also on the underlying statistical model. The practioners' investment rule is to exploit a spread if the deviation of both assets from the equilibrium is larger than twice the estimated standard deviation of the spread. In the bivariate cointegration framework (2), the estimated stationary residual

$$\hat{v}_t = S_{1t} - \left(\hat{a} + \hat{b}S_{2t}\right) \tag{13}$$

has a different variance. Thus, for the traditional approaches to pairs trading, the position is locked into if the spread exceeds twice the volatility either of the spread or of the residual of the cointegration regression (13).

The group tracking model allows for a different definition of an opening of the pair. We define the pair to be open if the probability of not being a pair is increasing and exceeds a threshold p_{thres}. This is motivated by the observation that this probability decreases if the assets depart from another. When both approach again the probability

increases. The lower this threshold probability p_{thres}, the larger the spread between the assets that the investor requires to take the positions. A larger value of p_{thres} will more frequently indicate that the pair is open.

By τ we denote the residual difference on the closing of the pair. The positions are unwound at the first time that $|S_{1t} - S_{2t}| \leq \tau$. The effect of τ is similar to the effect of choosing p_{thres}. A larger value of τ closes the pair more frequently and thus allows for more trades.

Computing the excess returns on exercising pairs trading is a non-trivial task. For the

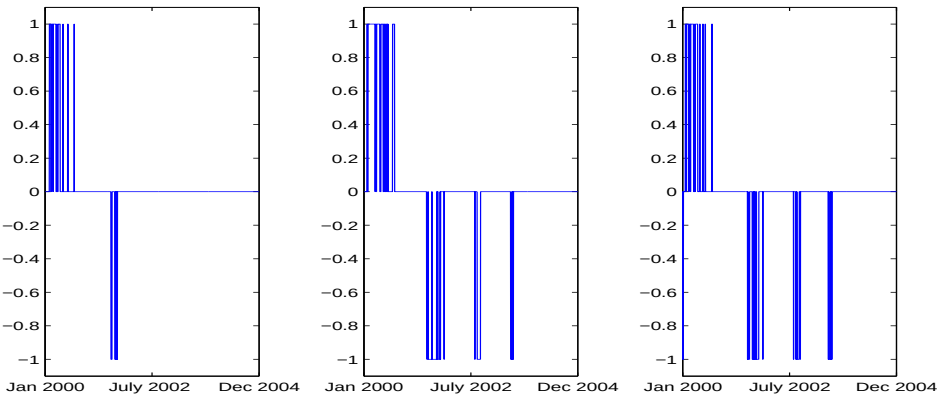

Fig. 11. Comparison of investment strategies with respect to the open-close identification of the pair.

details on the specific computations we refer to Andrade et al. (2005). Essentially, if the pair opens one dollar is invested in the pair. The resulting pay-off can be interpreted as an excess return. The pairs are found on the basis of the mid-price, whereas trading is executed using bid and ask prices. This accounts for the per trade transaction cost. The larger the number of trades, the higher the transaction costs that are associated with that strategy. This is particularly important because the number of trades may substantially differ across strategies. We neglect fixed costs as these can be assumed equal across strategies.

We illustrate this point empirically by considering two German stocks, Siemens and

Deutsche Bank, from January 2000 until December 2004. We consider end-of-day prices, resulting in 1270 observations per stock. The data series have been depicted in the introduction in figure 1.

In figure 11 the three investment rules are compared. A value different from zero in-

Strategy	t.costs	Basic Pairs	Cointegration	Group Tracking
$\tau = 0.001$, $p_{thres} = 0.25$	n	9.9%	13.1%	12.5%
	y	9.9%	13.0%	12.4%
Number of trades		2	4	4
$\tau = 0.005$, $p_{thres} = 0.25$	n	10.0%	13.3%	14.7%
	y	9.9%	13.0%	14.4%
Number of trades		2	4	5
$\tau = 0.005$, $p_{thres} = 0.30$	n	10.0%	13.3%	13.7%
	y	9.9%	13.1%	13.6%
Number of trades		2	4	5
$\tau = 0.005$, $p_{thres} = 0.40$	n	10.0%	13.3%	13.6%
	y	10.0%	13.1%	13.5%
Number of trades		2	4	5
$\tau = 0.01$, $p_{thres} = 0.25$	n	9.9%	13.0%	14.4%
	y	9.8%	12.8%	14.2%
Number of trades		2	4	5
$\tau = 0.025$, $p_{thres} = 0.25$	n	9.7%	12.6%	13.8%
	y	9.6%	12.3%	13.4%
Number of trades		2	4	5
$\tau = 0.025$, $p_{thres} = 0.40$	n	9.7%	12.6%	12.6%
	y	9.7%	12.3%	12.3%
Number of trades		2	4	5
$\tau = 0.05$, $p_{thres} = 0.35$	n	12.2%	13.8%	12.8%
	y	12.0%	13.4%	12.5%
Number of trades		3	5	6
$\tau = 0.05$, $p_{thres} = 0.40$	n	12.2%	13.8%	13.8%
	y	12.0%	13.5%	13.5%
Number of trades		3	5	6

Tab. 1. Excess returns on pairs trading strategies using the Deutsche Bank - Siemens pair between 2000 and 2004. In each cell, upper line: no transaction cost; lower line: including transaction costs.

dicates an open spread. The sign denotes the direction. If the spread is positive, a long position in Siemens and a short position in Deutsche Bank is taken, and vice versa if the spread is negative. Some similarities across the strategies appear immediately.
Table 1 shows the excess returns on executing different pairs trading strategies on the Deutsche Bank - Siemens pair. The first column refers to the model where the investment rule depends on the variance of the spread, the second to the cointegration approach and the third to the group tracking model. In each row, the excess returns for all three rules are shown without and with transaction costs. In addition, the numbers of trades are shown. It appears that group tracking suggests more frequently that the pair is open. Since the variance of the cointegration residual is smaller than the estimated variance of the spread, the former approach proposes more transactions than the latter. The results favour the group tracking methodology. Cointegration delivers excess returns that are about 2 percentage points higher than the ordinary investment rule. Pairs trading using group tracking yields even higher excess returns than cointegration. Indeed, only in a single case ($\tau = 0.05$, $p_{thres} = 0.35$), the performance of the group tracking pairs trading rule is worse than cointegration. However, even in that case, it still outperforms the basic pairs trading rule.

5 Conclusions

In this paper we present an alternative, flexible real-time econometric model to identify pairs trading opportunities and inference algorithms to estimate the parameters and the group structure of a multivariate financial time series. Both static and dynamic group structure estimation appear successful on simulated data. The results with the dynamic group structure model seem even more convincing. When the model is applied to real data, we find that the algorithm selects reasonable pairs. Moreover, invest-

ment rules that are based on the output of the algorithm can generate excess returns that are larger than those generated by pairs trading rules based on cointegration. The group tracking methodology also allows to track non-integrated time series such as volatilities and thereby opens a number of new possibilities in pairs trading.

The only drawback of the proposed methodology is the curse of dimensionality when an asset is part of more than one pair, or when it is part of a group with more than two members. This will be the case, for instance, for fixed income instruments. This issue will be part of our future research agenda.

References

Andrade, S. C., di Pietro, V. and Seaholes, M. S. (2005). Understanding the profitability of pairs trading, *UC Berkeley Working Paper* .

Bell, E. T. (1934). Exponential numbers, *American Mathematical Monthly* **41**: 411–419.

Bossaerts, P. (1988). Common nonstationarity components of asset prices, *Journal of Economic Dynamics and Control* **12**: 347–364.

Doucet, A., de Freitas, N. and Gordon, N. (2001). *Sequential Monte Carlo Methods in Practice*, Springer.

Ehrman, D. S. (2006). *Handbook of Pairs Trading*, John Wiley & Sons, New Jersey.

Gatev, E., Goetzmann, W. N. and Rouwenhorst, K. G. (2006). Pairs trading: Performance of a relative-value arbitrage rule, *The Review of Financial Studies* **19**(3): 797–827.

Gilks, B. and Berzuini, C. (2001). Following a moving target - monte carlo inference for dynamic bayesian models, *Journal of the Royal Statistical Society, Series B* **63**: 127–146.

Jacquier, E., Polson, N. and Rossi, E. (1994). Bayesian analysis of stochastic volatility models, *Journal of the Business and Economic Statistics* **12**(4): 69–87.

Kantas, N., Doucet, A., Singh, S. S. and Maciejowski, J. M. (2009). A overview of se-

quential monte carlo methods for parameter estimation in general state-space models, *15th IFAC Symposium on Systems Identification, SYSID 2009*, p. (in press).

Khan, Z., Balch, T. and Dellaert, F. (2005). Mcmc-based particle filtering for tracking a variable number of interacting targets, *Georgia Institute of Technology Working Paper* .

Liu, J. and West, W. (2001). Combined parameter and state estimation in simulation-based filtering, *in* A. Doucet, N. DeFreitas and N. J. Gordon (eds), *Sequential Monte Carlo Methods in Practice*, Springer.

Lo, A. W. and MacKinley, A. C. (1990). When are contrarian profits due to stock market overreaction?, *Review of Financial Studies* **3**: 175–2005.

McMillan, L. G. (2002). *Options as a Strategic Investment*, 4th edn, New York Institute of Finance, New York.

Mougeot, N. (2005). Volatility investing handbook, *BNP Paribas: Equity & Derivatives Research* .

Pang, S. K. (2009). *Models and Algorithms for Group Tracking*, Ph.d. thesis, University of Cambridge.

Pang, S. K., Li, J. and Godsill, S. (2008). Models and algorithms for detection and tracking of coordinated groups, *Working paper*, University of Cambridge.

Pitt, M. K. and Shephard, N. (1999). Filtering via simulation: Auxiliary particle filters, *Journal of the American Statistical Association* **94**: 590–599.

Tsay, R. S. (2002). *Analysis of Financial Time Series*, Wiley, New Jersey.

Vidyamurthy, G. (2004). *Pairs Trading*, Wiley, New Jersey.

Wei, X., Sun, J., and Wang, X. (2007). Dynamic mixture models for multiple time series, *20th International Joint Conference on Artificial Intelligence*, pp. 2909–2914.

A Algorithm Details

Let $(Z_{t-1}, \theta_{t-1})^{(m)}$, $m = 1, \ldots, N_p$ be the $t-1$ filtering estimate. The MCMC algorithm to propose particle to estimate the updated filtering distribution is given by

(1) set $m = 1$.

(2) randomize: with probability P_L do

(a) propose $(G_t^*, \pi_t^*, l_t^*, \theta_t^*, G_{t-1}^*, \pi_{t-1}^*, l_{t-1}^*, \theta_{t-1}^*)$ jointly from [5]

$$
\begin{aligned}
(G_{t-1}^*, \pi_{t-1}^*, l_{t-1}^*, \theta_{t-1}^*) &\sim (1/N_p) \sum_{p=1}^{N_p} \delta(Z_{t-1} - Z_{t-1}^{(p)}), \\
(G_t^*, \pi_t^*) &\sim p(G_t, \pi_t | G_{t-1}^*, \pi_{t-1}^*), \\
\theta_t^* &\sim N(c_t^{(p^*)}, (1-a^2)V_t) \\
l_t^* &\sim p(l_t | l_{t-1}^*, G_t^*, \theta_t^*)
\end{aligned}
$$

where

$$
c_t^{(p^*)} = a\theta_{t-1}^{(p^*)} + (1-a)\bar{\theta}_{t-1} \quad \text{and} \quad V_t = \frac{1}{N_p - 1} \sum_{p=1}^{N_p} (\theta_{t-1}^{(p)} - \bar{\theta}_{t-1})(\theta_{t-1}^{(p)} - \bar{\theta}_{t-1})'
$$

(b) accept the joint proposal with probability $\rho_1 = \min(1, A_1)$ where

$$
A_1 = \frac{p(S_t | S_{t-1}, G_t^*, l_t^*, \theta_t^*)}{p(S_t | S_{t-1}, G_t^{(m-1)}, l_t^{(m-1)} \theta_t^{(m-1)})} \tag{A.1}
$$

and set $(Z_t^{(m)}, \theta_t^{(m)}) = (G_t^*, \pi_t^*, l_t^*, \theta_t^*)$ else keep $(Z_t^{(m-1)}, \theta_t^{(m-1)})$

(3) else do (with probability $1 - P_L$)

(a) set $(G_{t-1}^{(m)}, \pi_{t-1}^{(m)}) = (G_{t-1}^{(m-1)}, \pi_{t-1}^{(m-1)})$

(b) propose π_t^* from

$$
\pi_t^* \sim q_3(\pi_t | \pi_{t-1}^{(m-1)}) = p(\pi_t | \pi_{t-1}^{(m)})
$$

[5] The proposals are based on an initial proposal of a particle $p^* \sim U(1, \ldots, N_p)$.

and accept with $\rho_2 = \min(1, A_2)$ where

$$A_2 = \frac{\prod_{i=1}^{N} p(g_{i,t}^{(m-1)} | g_{i,t-1}^{(m-1)}, \pi_t^*)}{\prod_{i=1}^{N} p(g_{i,t}^{(m-1)} | g_{i,t-1}^{(m-1)}, \pi_t^{(m-1)})} \tag{A.2}$$

(c) propose $g_{i,t}^* \sim p(g_{i,t} | g_{i,t-1}^{(m)}, \pi_t^{(m)})$ and accept with $\rho_3 = \min(1, A_3)$ where

$$A_3 = \frac{p(S_t | S_{t-1}, l_t^{(m)}, G_{\setminus i,t}^{(m')}, g_{i,t}^*)}{p(S_t | S_{t-1}, l_t^{(m)}, G_t^{(m')})} \tag{A.3}$$

where $G_t^{(m')}$ is the vector of the current MCMC iterate with updated components and $G_{\setminus i}$ denotes group structure except the ith component. Such MH-step requires a Gibbs cycle.

(d) sample a random particle p and then $\theta_t^{(m)} \sim N(c_t^{(p)}, (1-a^2)V_t)$ where

$$c_t^{(p)} = a\theta_{t-1}^{(p)} + (1-a)\bar{\theta}_{t-1} \quad \text{and} \quad V_t = \frac{1}{N_p - 1} \sum_{p=1}^{N_p} (\theta_{t-1}^{(p)} - \bar{\theta}_{t-1})(\theta_{t-1}^{(p)} - \bar{\theta}_{t-1})'$$

(e) sample $l_t^{(m)} \sim p(l_t | l_{t-1}, G_t^{(m)}, \theta_t^{(m)})$

(4) until $m \geq N_{MCMC}$, increase $m \to m+1$ and return to (2)

(5) keep only $Z_t^{(p)} = Z_t^{(\lfloor m/thin \rfloor)}$

(6) compute particle representation of time-t filtering density

$$\hat{p}(Z_t, \theta_t | S_{1:t}) = (1/N_p) \sum_{p=1}^{N_p} \delta_{(Z_t, \theta_t)^{(p)}}(Z_t, \theta_t)$$

The algorithm with the presented hyperparameters yields an MH acceptance rate of 25% to 67% after the burn-in period. In all applications we have set $thin = 2$.

The randomization in steps (2) and (3) selects either the joint or the individual draw. It allows the algorithm to move out of region where joint proposals are likely to be rejected. This refinement also permits to explore regions of the state space that are unlikely to reach with a joint move. This produces more diversified particles.

In all setups we have set the discount factor in the Liu and West (2001) procedure to 0.95.

B Group and Pairs Structure

Structure Number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Asset 1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Asset 2	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2
Asset 3	1	1	2	2	2	1	1	1	2	2	2	3	3	3	3
Asset 4	1	2	1	2	3	1	2	3	1	2	3	1	2	3	4

Tab. A.1: Group Structure for 4 Assets.

Pair Number	1	2	3	4	5	6
Pair	(1,2)	(1,3)	(1,4)	(2,3)	(2,4)	(3,4)

Tab. A.2: Pairs Structure for 4 Assets.

C Clustering with Respect to the Latent Volatility

In the main text we describe how to group assets according to their implied volatility. The full flexibility of the framework can be understood by considering a grouping with respect to the latent stochastic volatility. Noting the close relation between the latent and realized volatility a possible trading strategy takes positions in variance swaps. Each asset i is assumed to satisfy

$$r_{it} = \mu_{it} + e^{l_{it}/2}\varepsilon_{it}. \tag{C.1}$$

The individual log-variances l_{it} are assumed to follow a stationary AR(1) process with long-run mean $\bar{l}_i = \frac{\alpha_i}{1-\beta_i}$. If the assets in a group have a similar long-term behavior with a mean location that is determined by the overall group average $\bar{l}_i$ can be replaced by $V_{i,t-1} = (1/\lambda_{it}) \sum_{j \in \Lambda(G_t, g_{it})} l_{j,t-1}$. The resulting volatility dynamics read

$$\Delta l_{it} = (1 - \beta_i)(V_{i,t-t} - l_{i,t-t}) + \sigma_{l_i} e_{it}. \tag{C.2}$$

It should be noticed that $V_{i,t-1}$ is computed for the time-t group structure G_t. This is for the reason that G_t must determine the dynamics of l_t whose mean evolution is computed on the basis of l_{t-1}. The group structure model is the same as in the main text and can be also substituted by the pairs dynamics. From $(C.2)$ it becomes clear that the

Fig. C.1. Simulated Volatility Pairs with Group Structure (1,1,2,3). Upper panel: normalized stock price series, lower panel: log-variance series.

intercept and the autoregressive coefficient of the scedastic equation have new meanings. The larger β_i the slower the speed of reversion towards the group volatility. If a group consists of only a single asset, $V_{i,t-1}$ is substituted by $\bar{l}_i$ to ensure the stationarity of the volatility equation. Otherwise, the volatility would be considered equal to its mean at every point in time and thus the process became a random walk.

Figure C.1 illustrates a pattern of sample paths that the model can generate. The upper panel shows the simulated normalized stock prices and the lower panel the log-variances with $G = (1, 1, 2, 3)$. In figure C.2 the results for these series are depicted. These appear weaker, but the algorithm clearly indicates the correct group structure.

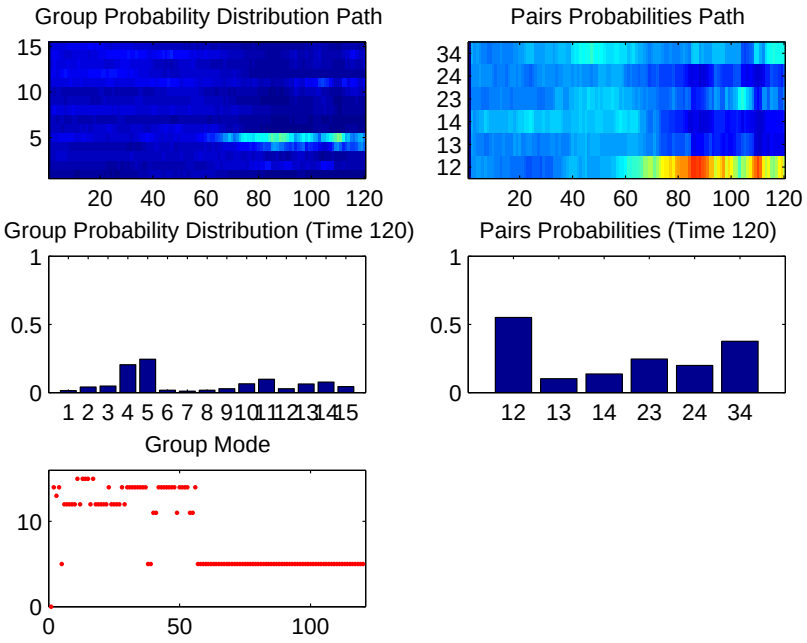

Fig. C.2. Filtering Results for Simulated Volatility Pairs with Group Structure (1,1,2,3).

Chapter 2

Flexible Bayesian Modeling of Implied Volatility Surfaces

Abstract

In their Bayesian approach to nonparametric bivariate regression, Smith and Kohn (1997) propose a regressor selection algorithm to fit a flexible surface to data. In this paper we demonstrate how this methodology can be applied to estimate implied volatility surfaces and why it should be considered an alternative to classical estimators. The implied volatility surface is modeled by a spline regression on many functions of moneyness and time to maturity. Covariates are selected by an MCMC algorithm and the estimate is obtained by averaging the estimates of each draw. In contrary to the majority of the literature on IVS modeling, we also assume a more financial perspective and consider option pricing as opposed to volatility fitting. Moreover, we show how the dependency of the surfaces over time can be modeled in this framework. Empirically the approach is illustrated through an application to plain vanilla European style options on the S&P500 index.

Key words: Implied Volatility Surface, Bayesian Model Averaging, Unbiased Estimation, Flexible Modeling
JEL classification: C51, G13

1 Introduction

Knowledge about the volatility of an asset is paramount to practioners in the financial industry. First, an asset's volatility is an essential parameter in pricing options which have that asset as underlying. And second, since implied volatilities are a forward-looking measure, they can be also used to predict future realized volatility which is a key factor in asset allocation.

Volatility is unobserved and thus has to be estimated. There are two distinct approaches. Either a statistical model for the volatility process is specified and subsequently estimated, or the traditional Merton (1973) and Black and Scholes (1973) (henceforth B&S) model is used to infer the volatility that has been used to price options that are currently traded. The volatility estimates that are obtained in the latter way are referred to as implied volatilities (IV). If these are considered on a domain of strike prices and maturity times an extrapolation of those points yields a "smile"- or "smirk"-shaped surface. An estimator of the IV surface (IVS) must be able to capture its nonlinear shape. One of the peculiarities of IVS data is that options are usually issued only for certain maturities and strike prices. IVS data thus look like a number of chains along the strike price axis with no data in between observed maturities. Also less liquid options are not traded every day and thus leave observational gaps. Implied volatility estimates in between such maturities are, however, extremely important since many over-the-counter derivatives are hedged using options with maturities and strikes that are not quoted.

This paper advocates the use of Bayesian regressor selection (George and McCulloch, 1993, 1995) and subsequent model averaging (Raftery et al., 1996, 1997) to estimate IVS. In Bayesian model averaging posterior inference is conducted in two steps. The posterior probability of the parameter is computed for each possible model. Thereafter the models are integrated out over their respective posterior probabilities. More precisely, given a full model $\mathcal{M}$ which can be written as a number of submodels $\mathcal{M} = \{M_i\}_{i=1}^{N}$

inference about the parameter of interest, say β, is based on the mixture estimate

$$p(\beta|y, \mathcal{M}) = \sum_{i=1}^{N} p(\beta|M_i, y) \cdot p(M_i|y). \tag{1}$$

Applied to the estimation of IVS there are three main advantages of this approach. First, it is nearly as flexible as kernel smoothers but, because of its parametric form, it yields a global estimate and thus does not suffer from biases in region of the surface where no data are observed. Second, in order to compute the state price density of the underlying one has to differentiate the implied volatility function twice with respect to moneyness. Further derivatives may be needed to test for jumps in the underlying. If a parsimonious model is used the estimate might be imprecise. If the model is non-parametric taking the derivatives numerically can cause instability in the subsequent computations. A similar argument applies to the computation of the local volatility surface, which can be easily computed for parametric models. Third, the procedure reveals which functions of moneyness and maturity are necessary to explain the shape of the surface. This provides further evidence that parsimonious regression models are miss-specified.

This paper is structured as follows. The second section gives a formal definition of the IVS and the statistical model that we use to estimate it. A short review of previous work on the estimation of the IVS is given and the issue of no-arbitrage in IV models is discussed. In the third section the MCMC algorithm by Smith and Kohn (1997) is briefly summarized and how the dependence between sequential surfaces can be modeled. The prior distributions for the Bayesian analysis are discussed. The fourth section deals with the empirical analysis. In- and out-of-sample results of the method when applied to European options on the S&P500 index are reported. The final section concludes. Some details on the practical implementation of the MCMC algorithm is given in the appendix.

2 Regression-type Modeling of Implied Volatility Surfaces

2.1 Implied Volatilities

Implied volatility is the volatility of an asset which is derived from options written on that asset. Throughout this paper we refer simply to implied volatilities but more precisely by this B&S implied log-normal volatilities are meant. These assume that the option from which the IV is computed has been priced with the B&S formula with a standard geometric Brownian motion describing the stochastic dynamics of the underlying. If the underlying, which is denoted by S_t, is log-normally distributed, it satisfies

$$\frac{dS_t}{S_t} = (r_t - \delta_t)\,dt + \sigma\,dW_t, \quad S_0 = s_0, \quad W_t \text{ a Brownian motion}$$

under the pricing measure.[1] The other variables are the risk-free interest rate r_t, the dividend rate δ_t and σ denotes the volatility of the process. For this dynamics of the underlying asset, the price of a European style call option C_t with a strike price K and maturity T on the underlying S_t is given by the B&S formula

$$C(S_t, \sigma, t, K, T, r_t, \delta_t) = e^{-\delta_t \tau} S_t\, N(d_1) - K e^{-r_t \tau} N(d_2) \tag{2}$$

with

$$d_1 = \frac{\log(S_t/K) + (r_t - \delta_t + \sigma^2/2)\tau}{\sigma\sqrt{\tau}}, \quad d_2 = d_1 - \sigma\sqrt{\tau} \text{ and } \tau = T - t. \tag{3}$$

[1] Implied volatilities are obtained by inverting the B&S pricing formula and hence apply under the risk-neutral measure. However, since log-normal asset dynamics are considered, Girsanov's theorem ensures that the volatility is the same under both measures.

N stands for the cumulative distribution function of the standard normal distribution. In the B&S formula (2) all parameters are known except for the volatility. [2] The crucial assumption in the B&S framework is that the volatility is constant. Formula (2) is bijective in the volatility and thus can be numerically inverted for all observed options, that is for different strikes and times-to-maturity. In this sense, implied volatilities rather than true volatilities are observed. The IVS at time t is defined by the mapping

$$\sigma_t^{BS} : (K, T) \longrightarrow \sigma_t^{BS}(K, T)$$

for a strike K and maturity T. Frequently a reparameterized version in relative coordinates is chosen

$$I_t(\kappa, \tau) := \log\left(\sigma_t^{BS}\left(\kappa S_t, t + \tau\right)\right) \tag{4}$$

where $\kappa = K/S_t$ denotes moneyness and τ time-to-maturity. This reparametrization is advisible since the trading frequency of an option is usually determined by its moneyness rather than its absolute strike price. The domain of the IVS shall be denoted by $D = [\tau_{min}, \tau_{max}] \times [\kappa_{min}, \kappa_{max}] \subset \mathbb{R}^2_+$. The logarithm is applied to stabilize the estimation in two respects. First, more extreme and less likely values for the IV are lowered. And second, the observations are less skewed (Fengler et al., 2005).

2.2 Statistical Model

Based on the definition of the IVS (4) an estimate must satisfy certain conditions. Most importantly, the IVS estimate must be smooth. One way to ensure a smooth surface estimate has been proposed by Smith and Kohn (1997) by assuming that the surface

[2] Actually, the dividends of an equity index are not quoted in the market. This can be estimated easily using the Put-Call-parity as shown in the appendix.

can be approximated by a function f which lies in the span of the tensor product of two univariate function bases with an additive error term,

$$I_t(\kappa, \tau) = f(\kappa, \tau) + \varepsilon_t, \quad f \in \text{span}\left[\{1, b_j^1(\kappa)|j \in I_1\} \otimes \{1, b_j^2(\tau)|j \in I_2\}\right]. \tag{5}$$

The terms in the tensor product are univariate function bases which here are taken to be splines. The span of the tensor product is a bivariate function basis in which the approximating function f lies. As univariate bases we use (truncated) cubic regression splines, but other specifications are possible as well.
To be more specific, if only one spline knot in each dimension is chosen, $K_\kappa = K_\tau = 1$, the full specification of the regression model (5) reads

$$\begin{aligned} I_t(\kappa, \tau) = {} & \beta_1 + \beta_2\kappa + \beta_3\kappa^2 + \beta_4\kappa^3 + \beta_5\tau + \beta_6\tau\kappa + \beta_7\tau\kappa^2 + \beta_8\tau\kappa^3 + \beta_9\tau^2 \\ & + \beta_{10}\tau^2\kappa + \beta_{11}\tau^2\kappa^2 + \beta_{12}\tau^2\kappa^3 + \beta_{13}\tau^3 + \beta_{14}\tau^3\kappa + \beta_{15}\tau^3\kappa^2 + \beta_{16}\tau^3\kappa^3 \\ & + \beta_{17}(\kappa - \kappa_1)_+^3 + \beta_{18}\tau(\kappa - \kappa_1)_+^3 + \beta_{19}\tau^2(\kappa - \kappa_1)_+^3 + \beta_{20}\tau^3(\kappa - \kappa_1)_+^3 \\ & + \beta_{21}(\tau - \tau_1)_+^3 + \beta_{22}\kappa(\tau - \tau_1)_+^3 + \beta_{23}\kappa^2(\tau - \tau_1)_+^3 + \beta_{24}\kappa^3(\tau - \tau_1)_+^3 \\ & + \beta_{25}(\tau - \tau_1)_+^3(\kappa - \kappa_1)_+^3 + \varepsilon_t. \end{aligned} \tag{6}$$

As standard in the financial literature, $(\cdot)_+ = \max\{\cdot, 0\}$.
In general, the number of parameters to be estimated in (5) is $q = (K_\kappa + 4)(K_\tau + 4)$ for cubic splines or $(K_\kappa + 3)(K_\tau + 3)$ for quadratic splines. This may seem rather overparameterized at a first glance. However, the posterior probabilities of those submodels that include many regressors will be small in practice. The model posterior probabilities also establish which terms are actually needed to model the IVS.
All 2^q submodels are identified by a binary selection variable $\gamma \in \{0, 1\}^q$. Let X be the matrix with all covariates in the regression (6).By X_γ we denote the submatrix that consists only of those columns of X which are included through γ.

The analysis of implied volatilities has gone in two directions. One strand of research, which has attracted attention more recently, considers market models for implied volatilities, see e.g. Jacod and Protter (2006) and Schweizer and Wissel (2008). These models are not primarily motivated by empirical modeling but by the possibility of enforcing no-arbitrage conditions. These are imposed by restricting the drift of a diffusion for the IV, similar to term structure of interest rate models.
The main strand of research has followed statistical specifications that range from parametric, e.g. Ané and Geman (2006) and Guidolin and Goncales (2006), over semiparametric, e.g. Fengler (2004), to fully nonparametric, e.g. Cont and da Fonseca (2002). The first class of models consists of parsimonious regression models that do not provide a fit with the data as good as the latter two classes do and thus induce a bias. To overcome that problem some authors, e.g. Fanone and Russo (2009), suggest to use regression models with many covariates. Such approaches lack statistical justification. Audrino and Colangelo (2007) apply tree-boosting to IVS data and find a significant improvement in its predictability. It has been shown by Smith and Kohn (1997) that their method outperforms boosting techniques in terms of efficiency. Modeling the IVS by a time-varying parameter cubic polynomial in forward moneyness, Hodges and Bedendo (2009) demonstrate that the Kalman filter can be applied to obtain accurate forcasts of the IVS. Cont and da Fonseca (2002) estimate the IVS nonparametrically with the kernel smoothing estimator that has been originally proposed by Nadaraya (1964) and Watson (1964), and show how its dynamics can be modeled by specifying a functional basis and decompose each surface into its functional principal components to reduce the dimensionality. The authors describe the dynamics of the IVS by its first three functional component time series. Kernel smoothing entails the difficulty of choosing the bandwidth. A too large bandwidth can smooth out interesting features

of the IVS whereas a too narrow kernel can induce dents in the surface. Fengler et al. (2005) use a dynamic semiparametric model and employ a VAR to model the dynamics of the time-varying functional principal components. Semiparametric models remedy the problem of biases but may also cause numerical problems when differentiating the implied state price density or when computing confidence intervals.

2.4 No-Arbitrage Considerations

The regression model (6) imposes no constraints on the surface. Despite the empirical observation that there exist plenty of arbitrage opportunities, which can be observed in IV data, a financial model of the IVS should exclude severe violations of no-arbitrage (NA) to avoid misspricings.
The most obvious NA condition states that a call option price shall not exceed the dividend-corrected stock price and must be greater than the difference between stock and strike price. More precisely, the price of a call option must satisfy

$$(S_t e^{-\delta_t \tau} - K e^{-r_t \tau})_+ \leq C_t \leq S_t e^{-\delta_t \tau}. \tag{7}$$

This is a weak necessary but not sufficient condition. Stricter NA conditions for IV have been proposed in the more theoretical literature as mentioned before. The focus of this paper is, however, on achieving strong in- and out-of-sample fit while not allowing for unnecessary biases. Therefore we do not essentially enforce the estimate of the IVS to be arbitrage-free in a stricter sense because it will deteriorate the in-sample fit and, as shown in the context of term structure models, the evidence on the advantage in predictions is mixed (Niu et al., 2007; Christensen et al., 2007). We use the weak no-arbitrage bounds (7) to ensure that naive arbitrage is excluded in the estimated surface.

3 Estimation and Prediction

3.1 Prior Distributions

The prior distributions should reflect any pre-sample information that are available. The model parameters of (5) are the regression parameters β and σ as well as the selection-variable γ. The prior is hierachical and uses the decomposition $p(\beta, \sigma^2, \gamma) = p(\beta|\sigma^2, \gamma)\, p(\sigma^2|\gamma)\, p(\gamma)$. Since the regressors are not orthogonal, we remain uniformative about γ,

$$p(\gamma) = 2^{-q}.$$

We also assign flat prior distribution to the log-variance,

$$p(\sigma^2|\gamma) \propto \sigma^{-2}. \tag{8}$$

Conditional on σ^2 and γ, the regression parameter is normal distributed with mean zero and variance proportional to the Fisher information,

$$p(\beta_\gamma|\sigma, \gamma) = N_{q(\gamma)}(0, c\sigma^2(X_\gamma' X_\gamma)^{-1}) \tag{9}$$

where c is a constant to tune the impact of the prior distribution on the posterior estimates. The covariance matrix in (9) exhibits the desirable property of taking collinearity into account. Those components of β that are most affected by highly linearly dependent covariates are assigned a larger prior variance.

The prior distributions (8) and (9) are known as G-prior which has been introduced and developed by Zellner (1971, 1987). It is worth noting that this prior setup for the linear regression model has led to substantial criticism. Most critics argue that if X is not deterministic its use in the prior (9) may violate the likelihood principle because a part of the data is used twice. Robert (2001) finds that debate 'rather vacuous' for sev-

eral reasons. First, the whole regression model (5) is conditional on the covariates and the prior (9) can be viewed as a first-stage posterior with respect to X. Second, it avoids any unrealistic choice of the prior. And third, the prior suggests a constant distribution on the mean of the implied volatility rather than on the regression parameter. Moreover, Robert (2001) argues that the true drawback of this prior setup is not addressed in those criticisms which is that the prior does not depend on true prior information.
The last point is particularly true when a sufficiently large value of c is considered in which case the subsequent estimation is purely based on the likelihood. However, in our view this prior setup does allow to utilize a sufficient amount of prior information on the shape of the surface. Since the coefficients are shrunk towards zero, a larger weight of the prior implies smaller posterior mean coefficients and thus a flatter surface. This can be used, for instance, to impose the weak NA condition (7). If the data violate the condition and therefore imply too steep wings of the surface, a smaller value of c can pull them down. This is as much prior information as we may wish to include in the model.
Other non-sample information that could be used in this prior are e.g. that index options have a steeper smile than options on a single stock (see e.g. Branger and Schlag (2004) for a theoretical justification). In such case, if the IVS of an index has to be estimated and previous estimates with the same base model on a single stock IVS are available, the prior mean could be set accordingly and the variance parameter set small enough to assign it some weight.

3.2 Posterior Analysis á la Smith and Kohn (1997)

The estimation of regression model (6) follows Smith and Kohn (1997) and is carried out for each trading day individually. The authors propose the use of a Gibbs' sampler to obtain a sample from the posterior of the parameters:

(1) initialize the selection variable $\gamma^{[0]}$

(2) for $m = 1 : \tilde{M} + M$

for $i = 1 : q$ draw $\gamma_i^{[m]}$ from

$$p(\gamma_i|y, \gamma_{j\neq i}) \propto (1+c)^{-q(\gamma)/2} S(\gamma)^{-n/2} \tag{10}$$

where

$$S(\gamma) = y'y - \frac{c}{1+c} y' X_\gamma (X_\gamma' X_\gamma)^{-1} X_\gamma' y$$

and $q(\gamma) = \sum_i \gamma_i$ the number of selected regressors

(3) for $m = \tilde{M} + 1 : M$

estimate $y = X_{\gamma^{[m]}} \beta_{\gamma^{[m]}} + e^{[m]}$ to obtain a sequence of distributions for $\{\beta_{\gamma^{[m]}}\}$

The iterates $\gamma^{[m]}$ are drawn from $p(\gamma|y)$ and thus the general version of Bayesian model averaging (1) becomes

$$p(\beta|y) = \int p(\beta|y,\gamma)\, p(\gamma|y)\, d\gamma \approx (1/M) \sum_{m=\tilde{M}+1}^{\tilde{M}+M} p(\beta|y, \gamma^{[m]}).$$

Using this model averaging strategy, we obtain a smoothed IVS by evaluating the regression model (6) with $\hat{\beta} = \widehat{E(\beta|y)}$ and plugging κ and τ for an equidistant grid on the domain D.

3.3 On the Choice of the Loss Function

The inclusion probability in (10) is based on a residual sum of squares. For $c \to \infty$, $S(\gamma) = y'(I - P_{X_\gamma})y$. Hence, the loss function that is implied by the model setup above is $\|\widehat{I}_t - I_t^{BS}\|^2$. In financial practice not an optimal fit of the surface with the IVS data is sought, but the avoidance of option miss-pricings. Therefore, from a financial point of view, a loss function in terms of option prices is desired. This can be computationally

efficiently done by noting that pricing errors are the result of an unprecise estimate of the IVS in regions where the option price varies a lot with the volatility. That measure of sensitivity is frequently summarized by the greek letter Vega $\mathcal{V} = \partial C/\partial\sigma$. The Vega is analytically available for European style options. Accordingly, this leads to a modification of the loss function. We weight each observation with the option implied $\mathcal{V}$. Hereto the IV and the covariates are multiplied by the square root of the corresponding option's Vega. In this way, the surface will have a stronger fit near ATM since those options exhibit the largest $\mathcal{V}$s.

3.4 Dynamics and Prediction

In a Bayesian framework all forecasts are based on the predictive distribution of the random variable of interest. Generally, the predictive distribution of a set of observations $\beta_1, \ldots, \beta_T$, whose distribution depends on a parameter θ, is computed as $p(\beta_{T+1}|\beta_{1:T}) = \int p(\beta_{T+1}|\beta_{1:T}, \theta)p(\theta|\beta_{1:T})\, d\theta$ where $\beta_{s:t} := \beta_s, \ldots, \beta_t$. If no closed-form solution is available, one simulates draws of the parameter from its posterior and subsequently from the distribution of the $T+1$th observation given these draws and averages them out. The dependence over time $p(\beta_{T+1}|\beta_{1:T}, \theta)$ needs to be modeled explicitly. As in Guidolin and Goncales (2006), we model the dynamics of the posterior means by a VAR

$$\mathrm{vec}(\boldsymbol{\beta}) = (\mathbf{I}_{q_e} \otimes \boldsymbol{\beta}_{(-1)})\mathbf{b} + \mathbf{e},$$

where $\mathbf{b} = \mathrm{vec}(\mathbf{B})$, $\mathbf{I}_{q_e}$ denotes an identity matrix of dimension q_e, $\boldsymbol{\beta} = (\hat{\beta}_{p+1:T} - \bar{\beta})$ and $\boldsymbol{\beta}_{(-1)}$ is its lagged version as well as $\bar{\beta} = T^{-1}\sum_{t=1}^{T}\hat{\beta}_t$. The effective dimension $q_e \leq q$ denotes the size of the VAR. Only variables that are frequently included are labeled essential.

In this paper we use $p = 1$ for the sake of computational simplicity. Endogenizing p will come at the cost of many more iterations to either compute Bayes factors for a number

of possible lag lengths or for running a reversible jump MCMC algorithm.
Whether the posterior distribution of $\mathbf{B}$ is analytically tractable or not, depends on the prior specification. Despite the fact that the Minnesota prior (Litterman, 1986) has become a frequent choice among applied economists, we use the Normal-Wishart prior because Kadiyala and Karlsson (1997) find that it performs superior in forecasting. It is hierachical and given by

$$\mathbf{b}|\mathbf{S_e} \sim N(\underline{\mathbf{b}}, \mathbf{S_e} \otimes \underline{\mathbf{S}}_\beta), \quad \mathbf{S_e} \sim iW(\underline{\mathbf{S}}_\mathbf{e}, \alpha). \tag{11}$$

The marginal posterior of the parameters $\mathbf{B}$ is matrix-variate Student-t

$$p(\mathbf{B}|\hat{\beta}_{1:T}) = t(\bar{\mathbf{S}}_\beta^{-1}, \bar{\mathbf{S}}_\mathbf{e}, \bar{\mathbf{B}}, T + \alpha) \tag{12}$$

with

$$\bar{\mathbf{B}} = \bar{\mathbf{S}}_\beta \left(\underline{\mathbf{S}}_\beta{}^{-1} \underline{\mathbf{B}} + \boldsymbol{\beta}'_{(-1)} \boldsymbol{\beta}_{(-1)} \hat{\mathbf{B}} \right)^{-1} \text{ and } \bar{\mathbf{S}}_\beta = \left(\underline{\mathbf{S}}_\beta^{-1} + \boldsymbol{\beta}'_{(-1)} \boldsymbol{\beta}_{(-1)} \right)^{-1}.$$

The IVS at time $T+h$, $h > 0$, is reconstructed from the predicted parameters. It is useful to marginalize out the error covariance matrix since it does not appear in the chosen forecasting function. The resulting 1-step ahead predictor is

$$\hat{\beta}_{T+1|T} = \bar{\beta} + \int (\hat{\beta}_T - \bar{\beta})\, \mathbf{B} \cdot p(\mathbf{B}|\hat{\beta}_{1:T})\, d\mathbf{B} \tag{13}$$

for which the distribution $p(\mathbf{B}|\hat{\beta}_{1:T})$ has been stated above. In the sequel we focus on the posterior mean prediction.

4 Empirical Results

4.1 Data

For the estimation of the model we use European options on the S&P 500 index. The data have been provided by the Chicago Board of Options Exchange. The data series contain daily observations from January until December 2005. This corresponds to 252 trading days. The period does not contain any abnormal jumps or high-volatility periods. The time-to-maturity is computed using the 30/360-daycount convention.

In figure 1 the data together with the volatility index VIX [3] (lower panel) are illus-

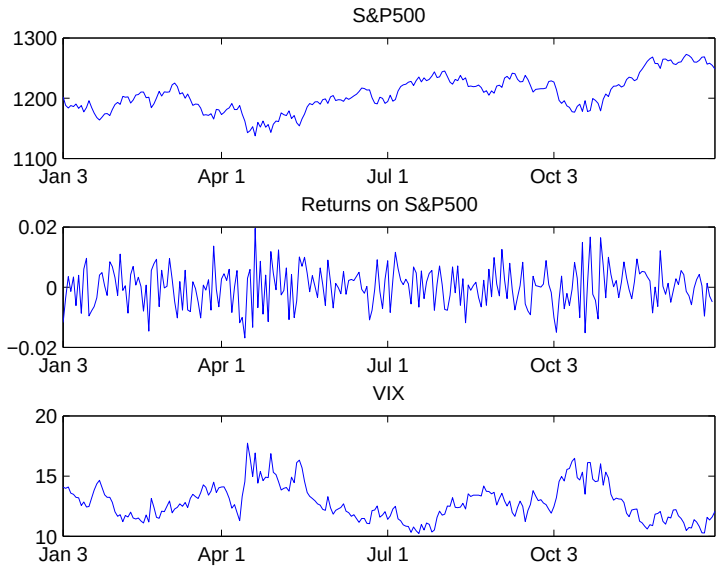

FIG. 1 upper panel: S&P500 (level), middle panel: S&P500 returns, lower panel: VIX. All data are from 2005.

[3] The VIX data are as of the most recent update in 2003 since when the VIX is based on options on the S&P500 rather than S&P100. More options are used which make it a more stable estimate of market volatility.

trated. The upper and middle panel show the S&P500 level and returns, respectively. Obviously, the VIX peaks up simultaneous with increased volatility of the returns of the S&P 500 in April/May and in October/November. Generally, it has been a tranquile period. All returns are within a $\pm 2\%$-margin. [4] The number of observations per day ranges from 119 to 224 and is 175 on average.

In figure 2 a sample estimate of the IVS on the closing of May 11, 2005 is depicted.

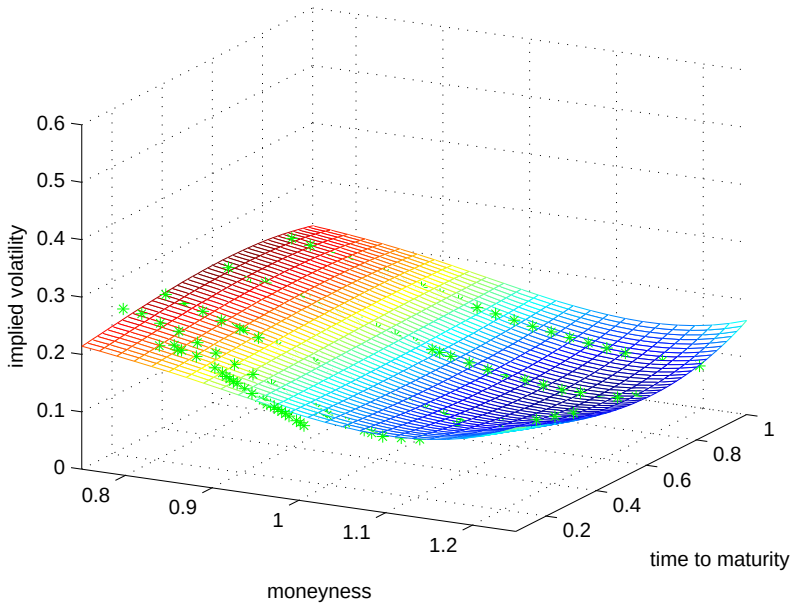

FIG. 2 Cubic Spline Regression Estimate of the S&P500 IVS on closing of May 11, 2005.

On the left axis is the moneyness on the right axis is time to expiry. One can easily see the smirk along the moneyness axis. Options were priced with approximately 15% to 30% volatility on the considered domain. Figure 3 shows the estimate of the IVS of the same day but with the Nadaraya-Watson kernel smoother. The bandwidth is chosen according to the rule-of-thumb proposed by Scott (1992).

[4] For a comparison, in 2002 the VIX reached over 40% and in the second half of 2008 above 80% whilst in the considered period the VIX never exceeded 20%. A daily volatility of 1% corresponds to about 16% annual volatility.

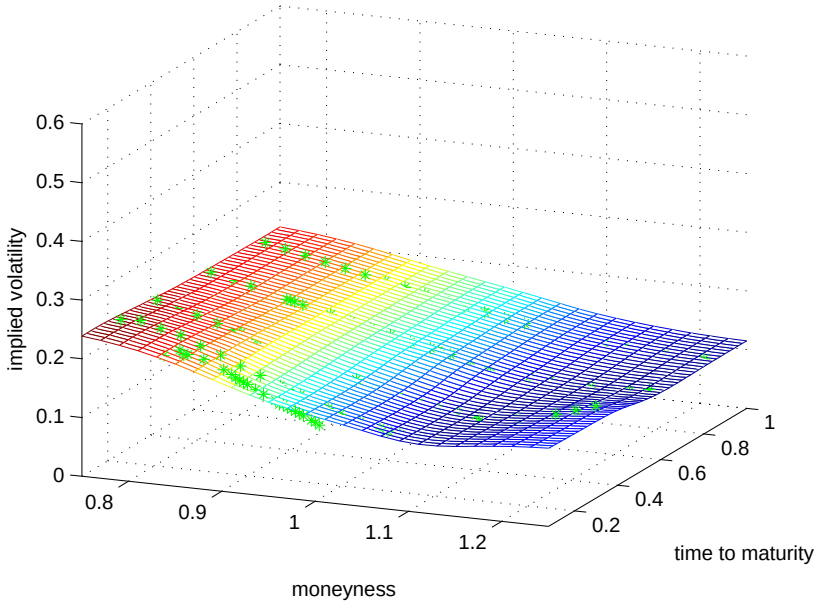

FIG. 3 Nadaraya-Watson Estimate of the S&P500 IVS on closing of May 11, 2005.

4.2 Model Specification Revisited

Besides yielding flexible estimates, the regressor-selection-algorithm has the advantage of revealing which regressors, i.e. functions of κ and τ are necessary to explain the IVS. This can be seen by inspecting the time series of inclusion probabilities for each regressor. Some functions are included nearly in every draw and some hardly ever. Figure 4 shows the 144 time series of inclusion probabilities for each regressor. It reveals that the constant, κ, κ^2, κ^3, τ, τ^2, τ^3, $\kappa\tau$, $\kappa^2\tau$, $\kappa^3\tau$, $\kappa^3\tau^2$, $\kappa^3\tau^3$, $(\kappa - k_1^\kappa)_+^3$, $(\kappa - k_2^\kappa)_+^3$, $(\kappa - k_3^\kappa)_+^3$, $(\kappa - k_4^\kappa)_+^3$, $\tau(\kappa - k_1^\kappa)_+^3$, $\tau(\kappa - k_2^\kappa)_+^3$ and $\tau^2(\kappa - k_1^\kappa)_+^3$.[5] All other terms are selected less than 5% over all draws and days. The terms which contribute the most to the explanation of the shape are primarily functions of the moneyness and interaction terms with time to maturity. The large number of regressors that contribute significantly to

[5] From right to left and up to down the numbers of these time series are 1, 2, 3, 4, 5, 6, 7, 8, 11, 14, 15, 16, 17, 18, 19, 20, 25, 26 and 33.

the explanation of the shape of the IVS indicate that simple parismonious models are miss-specified.

Rather than letting the algorithm pick the essential covariates, one can enforce the in-

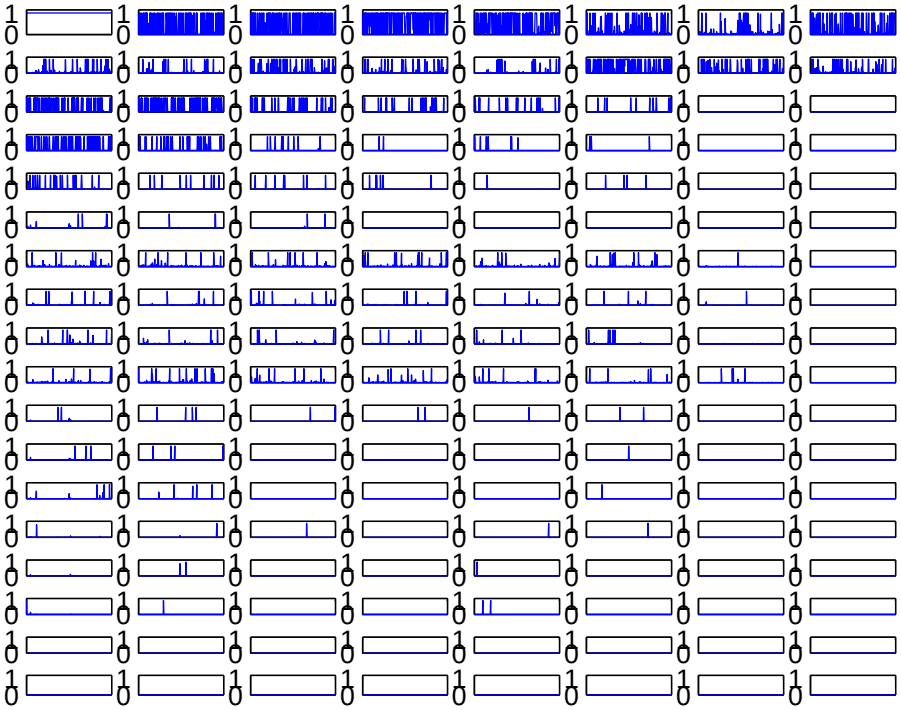

FIG. 4 Time series for inclusion probabilities for the 144 potential regressors.

clusion of certain variables. Such variables are referred to as focal variables (Pesaran and Timmermann, 2000). In the framework of thick modeling (Granger and Jeon, 2003), which is basically the frequentist counterpart to Bayesian model averaging, this has been employed. For instance, to forecast the returns on the S&P500 Aiolfi and Favero (2005) set those covariates as focal that define the long-run equilibria of the stock market. In all empirical setups, we set the constant focal to ensure a zero mean residual. Other covariates could be set focal, too, if desired. This reduces the dimensionality and thus the required computational effort. A typical example in the current context could

be to include the covariates of the basic model $\beta_1 + \beta_2\kappa + \beta_3\kappa^2 + \beta_5\tau + \beta_6\tau\kappa$ always.

4.3 In-Sample Performance

In the current context, the in-sample analysis is twofold as explained in section 3.3. From a statistical perspective, a good fit of the IVS with the observed IV is desired. From a financial perspective, small pricing errors are in the focus.

4.3.1 Statistical Perspective

From a purely statistical perspective, the in-sample fit of the IVS estimate at every trading day is measured by the average mean squared error of the posterior mean IVS in the domain D,

$$AMSE_t = \frac{1}{N_t} \sum_{\{i:(\kappa_{t,i},\tau_{t,i})\in D\}} \left(\widehat{\sigma_{t,i}^{BS}} - \sigma_{t,i}^{BS}\right)^2 \tag{14}$$

where N_t is the cardinality of the summation set. The posterior mean smoothed IV $\hat{\sigma}^{BS}$ are computed using $\hat{\beta}$. As before, σ^{BS} denote the observed IV. Averaging the AMSE over all trading days yields

$$AMSE = \frac{1}{T}\sum_{t=1}^{T} AMSE_t \tag{15}$$

which is used as a model evaluation criterion.

In table 1 the results for the in-sample fit for the regression model with $K_\kappa = 8$ and $K_\tau = 8$ knots in the full specification are reported. Throughout this paper $c = 10^6$ which in all constellations delivered the best results. In theory, the more the posterior is based on the likelihood the closer the predicted surface should be to observations. All of the estimated surfaces exhibit no arbitrage as defined in the weak no arbitrage condition (7). For a comparison, in the last row of table 1 the in-sample results for a

parsimonious model as advocated e.g. by Guidolin and Goncales (2006) are reported. [6] It should be noted that the fit of the Nadaraya-Watson estimator can be tuned as to de-

method	$\sqrt{AMSE}$	$\sqrt{AMSE_{25\%}}$	$\sqrt{AMSE_{50\%}}$	$\sqrt{AMSE_{75\%}}$
cubic spline	0.0157	0.0118	0.0142	0.0175
cubic spline (fin)	0.0258	0.0191	0.0232	0.0294
simple	0.0190	0.0143	0.0170	0.0215
Nadaraya-Watson	0.0126	0.0099	0.0115	0.0138

Tab 1. Comparison of model setups for in-sample fit with $K_\kappa = 8$ and $K_\tau = 8$.

liver an arbitrarily good fit by decreasing the bandwidth. However, this will increase the unmeasurable bias in between observations. To measure its insample fit, we use the rule-of-thumb bandwidth proposed by Scott (1992), $h_{\kappa/\tau} = n^{-1/6}\hat{\sigma}_{\kappa/\tau}$, which has become a standard method in practice. With the so-chosen bandwidth, the Bayesian model averaging spline-regression lies in between the performance of the simple regression model and the kernel-regression.

4.3.2 Financial Perspective

From a financial perspective, not the fit of the IVS estimate matters, but the fit with the corresponding option prices. Hence, the financial measure reads

$$OAMSE_t = \frac{1}{N_t} \sum_{\{i:(\kappa_{t,i},\tau_{t,i})\in D\}} \left(\frac{C(\widehat{\sigma_{t,i}^{BS}}) - C(\sigma_{t,i}^{BS})}{C(\sigma_{t,i}^{BS})} \right)^2 \tag{16}$$

[6] Such parsimonious model takes the form $I_t = \beta_1 + \beta_2\kappa + \beta_3\kappa^2 + \beta_4\tau + \beta_5\tau\kappa + \varepsilon_t$. Guidolin and Goncales (2006), however, use forward moneyness.

where, as before, N_t is the cardinality of the summation set and $OAMSE = \frac{1}{T}\sum_{t=1}^{T} OAMSE_t$. The results in table 2 reveal a similar picture to that in the previous section. The exception, however, is that the Vega-weighted spline-model yields significantly lower pricing errors than the unweighted model. The pricing errors are marginally larger than with the Nadaraya-Watson estimator and smaller than with the simple or the unweighted spline-selection model. Again, the kernel-smoother can suffer from miss-pricings for option maturities that are not observed. This emphasizes the point we

method	$\sqrt{OAMSE}$	$\sqrt{OAMSE_{25\%}}$	$\sqrt{OAMSE_{50\%}}$	$\sqrt{OAMSE_{75\%}}$
cubic spline	0.6902	0.6619	0.6947	0.7153
cubic spline (fin)	0.3104	0.2857	0.3054	0.3250
simple	0.6018	0.3886	0.4898	0.6605
Nadaraya-Watson	0.2698	0.2213	0.2542	0.3070

Tab 2. Comparison of model setups for in-sample fit with $K_\kappa = 8$ and $K_\tau = 8$.

made in the estimation section. A statistical model for the IVS is not necessarily a good financial model. A strong fit of the IVS does not imply small pricing errors and thus the model that weigths observations with respect to their sensitivity to the implied volatility is perferable.

4.4 Out-of-Sample Forecasts

Out-of-sample forecasts of the IVS can be used as an investment decision support tool: if accurate enough, IVS predictions can generate profits by taking a long (short) position in options whose IV is predicted to increase (decrease).
For the out-of-sample forecast exercise, we use a rolling window of 100 trading days.

A rolling window is advisable when structural breaks in the underlying process may be present. Over time the parameter estimates adjust faster to the new values. Here it is applied to maintain the precision of the parameter estimates and hence the forecast over time.

We concentrate on 1-step-ahead forecasts. After each, the next observation is added to and the oldest is discarded from the dataset. We are interested in whether the forecast is biased and what kind of IVS movement can be possibly predicted. The performance measure (15) serves us to compare to quality of the forecasted IVS.

As pointed out in e.g. Cont and da Fonseca (2002), market practioners use the sticky moneyness rule to predict the IVS. The rule claims that the IVS does not alter over short time horizonts,

$$I_t(\kappa, \tau) \approx I_{t+\Delta t}(\kappa, \tau) \qquad \forall (\kappa, \tau) \in D.$$

This rule applies both in the original and in relative coordinates (Cont and da Fonseca, 2002). It is implemented by using the same smoothed surface parameters for the forecast as for the previous trading day. Thereby it avoids the use of a time series model for predictions. More precisely, $X\hat{\beta}_{t-1}$ is compared to $X\hat{\beta}_{t|t-1}$ as predictors for $\log(\sigma_t^{BS})$. In table 3 the mean squared predicition errors and its quartiles are reported. Moreover, the size of the VAR for different inclusion threshold values and the corresponding results for the sticky-moneyness rule (last row) are shown.

The performance of the VAR prediction depends on the chosen effective dimension q_e. It appears that the best forecast is achieved for a value of about 5% or 19 variables. In that case the forecast using a VAR is still marginally inferior to sticky moneyness.

prob_i	q_e	$\sqrt{MSPE}$	$\sqrt{MSPE_{25\%}}$	$\sqrt{MSPE_{50\%}}$	$\sqrt{MSPE_{75\%}}$
10%	5	0.0481	0.0243	0.0307	0.0443
7.5%	9	0.0253	0.0223	0.0241	0.0275
5%	19	0.0229	0.0209	0.0226	0.0265
3.5%	26	0.0259	0.0226	0.0251	0.0285
2.5%	33	0.0265	0.0208	0.0260	0.0293
1%	53	0.0431	0.0220	0.0243	0.0300
sticky κ		0.0226	0.0209	0.0226	0.0243

Tab 3. Comparison of out-of-sample performances for difference effective dimension q_e.

4.5 Highest Posterior Bands

An intrinsictly strong point of Bayesian analysis is that posterior simulation methods deliver highest posterior intervals (HPI) as a by-product. The two inference stages require some attention though. At the first accounts for the uncertainty from the model and the second for the uncertainty associated with the regression parameter. The general procedure is to randomly pick a model and for that model simulate the regression parameter from its conditional posterior. With the drawn parameter a set of implied volatility can be predicted. Repeating the same procedure a large number of times and sorting the predicted IV, the 5th and 95th percentile are taken as pointwise highest posterior interval.

Figure 5 shows such 90%-HPI for the IVS estimate of May 11, 2005. For that particular day the interval ranges from 5% around ATM to about 20% deep ITM and OTM options. The surface is evaluated on the same grid that is spanned by the original observations.

If kernel estimators are used, only asymptotic confidence intervals are available.

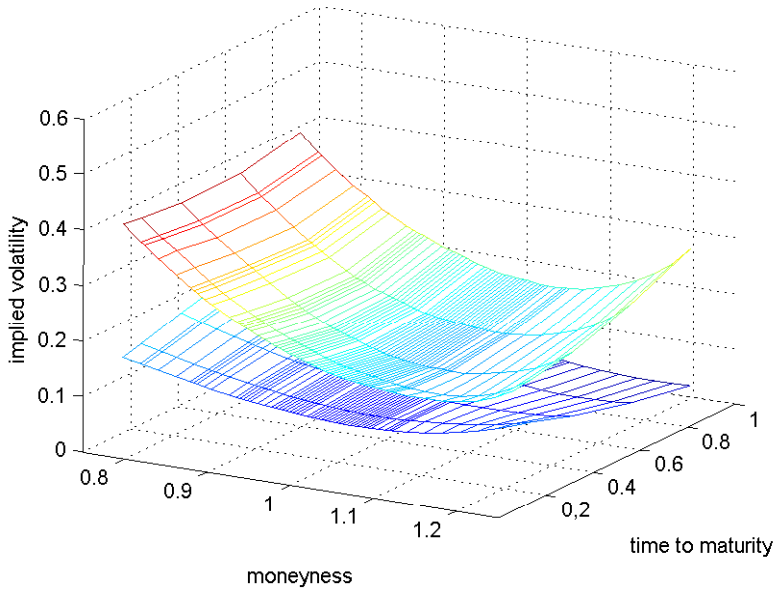

FIG. 5 Pointwise highest posterior intervals for the IVS of May 11, 2005.

4.6 State Price Density

A precise estimate of the state price density (SPD) can be useful for numerous reasons. It is mostly used for pricing exotic derivatives. It can also be employed to test whether the model for the underlying should include jumps. To this end Aït-Sahalia and Jacod (2009) derive a statistical test which is based on the second cross derivative of the SPD with respect to S_t and $S_{t+\tau}$ which, if estimated by classical nonparametric techniques, can be very unstable. The state price density can be obtained from option prices by computing $\rho(K,T|S_t,t) = e^{r\tau}\partial^2 C(K,T)/\partial K^2$ (Breeden and Litzenberger, 1978). For the B&S model the derivative can be expressed in terms of the volatility function (Shimko,

1993) and in relative coordinates

$$\rho(\kappa,\tau|S_t,t) = N'(d_2)\left(\frac{1}{\kappa S_t\sigma_t^{BS}\sqrt{\tau}} + \frac{2d_1}{\sigma_t^{BS}}\frac{\partial\sigma_t^{BS}}{\partial\kappa}\frac{1}{S_t} + \frac{d_1 d_2\kappa S_t\sqrt{\tau}}{\sigma_t^{BS}}\left(\frac{\partial\sigma_t^{BS}}{\partial\kappa}\right)^2\frac{1}{S_t^2} + \kappa S_t\sqrt{\tau}\frac{\partial^2\sigma_t^{BS}}{\partial\kappa^2}\frac{1}{S_t^2}\right). \tag{17}$$

N' denotes the pdf of the standard normal distribution. The partial derivatives of the IV with respect to moneyness can be analytically computed once the posterior model is chosen and estimated because it is parametric. Let the model $I_t(\kappa,\tau)$ be selected after discarding non-relevant regressors (fixing corresponding coefficients at zero). Obviously,

$$\frac{\partial\sigma_t^{BS}}{\partial\kappa} = \frac{\partial I_t(\kappa,\tau)}{\partial\kappa}\sigma_t^{BS} \text{ and } \frac{\partial^2\sigma_t^{BS}}{\partial\kappa^2} = \left(\frac{\partial^2 I_t(\kappa,\tau)}{\partial\kappa^2} + \left(\frac{\partial I_t(\kappa,\tau)}{\partial\kappa}\right)^2\right)\sigma_t^{BS}. \tag{18}$$

The derivative $\partial I_t(\kappa,\tau)/\partial\kappa$ is found by noting that $\frac{\partial(\kappa-\kappa_1)_+^3}{\partial\kappa} = 3(\kappa-\kappa_1)_+^2$ and $\frac{\partial^2(\kappa-\kappa_1)_+^3}{\partial\kappa^2} = 6(\kappa-\kappa_1)_+$. The use of cubic splines renders the resulting estimate twice continuously differentiable. In the quadratic spline case one would have $\frac{\partial^2(\kappa-\kappa_1)_+^2}{\partial\kappa^2} = 2\cdot 1_{\{\kappa\geq\kappa_1\}}$, which would introduce a discontinuity. Hence, the SPD (17) can be estimated from the smoothed IVS without resorting to numerical methods. Applying formulas (18) to $\hat{I}_t$ the resulting 8-week SPD (17) for March 30, 2005 is plotted in figure 6 together with the SPD estimate from a parsimonious model like that used by Guidolin and Goncales (2006). Since the in-sample performance of the regressor-selection method is better than the one achieved with the parsimonious model, the resulting SPD is also more accurate. Furthermore, in contrast to classical nonparametric methods, cross-derivatives with respect to start and end point as needed for the testing procedure in Aït-Sahalia and Jacod (2009) are computed with no numerical problems.

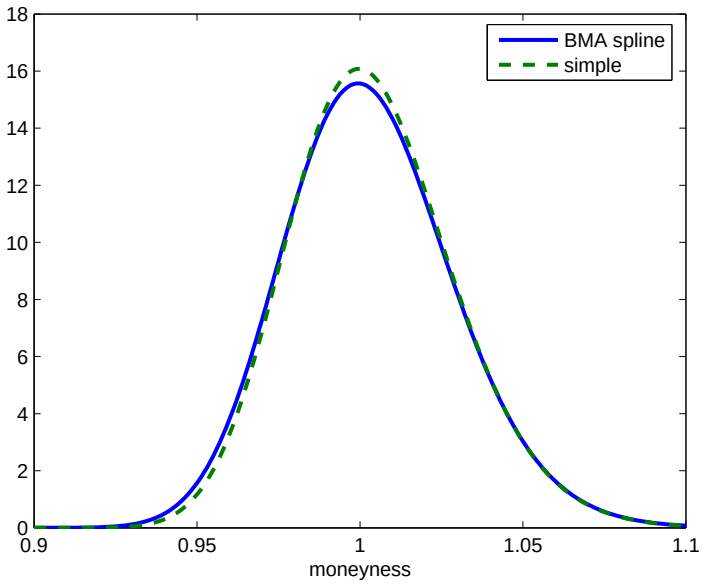

FIG. 6 15-days state price density of the S&P500 as of March 30, 2005. solid: regressor-selection, dotted: parsimonious regression

5 Conclusions

This paper advocates the use of Bayesian model averaging of spline-regressions to estimate the IVS because it avoids any kind of unnecessary bias in the estimation. The effect of the prior to shrink the wings of the estimated surface in order to exclude naive arbitrage in the surface has been discussed. Moreover, it does not suffer from numerical deficiencies when computing the state price densitiy or testing for jumps in the underlying. It also allows for finite sample pointwise highest posterior intervals.
Out-of-sample forecasts based on a Bayesian VAR reveal no improvement over a simple practioners' rule. Since Guidolin and Goncales (2006) find that there is a forecastable component in the IVS, there is space of further research. First, the trade-off between neglecting regressors and reducing the dimension of the VAR can be further investigated. Second, the Karhune-Loeve decomposition to extract functional principal components could be used. This has been proposed by Cont and da Fonseca (2002) but

they do not provide any empirical evidence of its power in forecasting.

References

Aiolfi, M. and Favero, C. A. (2005). Model uncertainty, thick modelling and the predictability of stock market returns, *Journal of Forecasting* **17**: 2789–2795.

Aït-Sahalia, Y. and Jacod, J. (2009). Testing for jumps in a discretely observed process, *Annals of Statistics* **37**(1): 184–222.

Ané, T. and Geman, H. (2006). Stochastic volatility and transaction time: An activity-based volatility estimator, *Journal of Risk* **2**(1): 57–69.

Audrino, F. and Colangelo, D. (2007). Forecasting implied volatility surfaces, *Discussion Paper 2007-43*, St.Gallen University.

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

Branger, N. and Schlag, C. (2004). Why is the index smile so steep?, *Review of Finance* **8**: 109–127.

Breeden, D. T. and Litzenberger, R. H. (1978). Prices of state contigent claims implicit in options prices, *Journal of Business* **51**: 621–651.

Christensen, J. H. E., Diebold, F. X. and Rudebusch, G. D. (2007). The affine arbitrage-free class of nelson-siegel term structure models, *NBER Working Paper* **13611**.

Cont, R. and da Fonseca, J. (2002). Dynamics of implied volatility surfaces, *Quantitative Finance* **2**: 45–60.

Fanone, E. and Russo, V. (2009). A parametric model for the implied volatility surface:evidence for equity and commodity options, *mimeo* .

Fengler, M. (2004). *Semiparametric Modelling of Implied Volatility*, Ph.d. thesis, Humboldt-Universität zu Berlin.

Fengler, M. R., Härdle, W. K. and Mammen, E. (2005). A dynamic semiparametric factor

model for implied string volatilities, *SFB 649 Discussion Paper 2005–020*, Humboldt-University, Berlin.

George, E. and McCulloch, R. (1993). Variable selection via gibbs sampling, *Journal of the American Statistical Association* **85**: 398–409.

George, E. and McCulloch, R. (1995). Approaches for bayesian variable selection, *Statistica Sinica* **7**: 339–373.

Geweke, J. (1992). Evaluating the accuracy of the calculation of posterior moments, *in* J. M. Bernado, M. J. Bayarri, J. O. Berger, A. P. David, D. Heckerman and A. F. M. Smith (eds), *Bayesian Statistics, Vol. 3*, Oxford University Press.

Granger, C. W. J. and Jeon, Y. (2003). Thick modelling, *Economic Modelling* **21**: 323–343.

Guidolin, M. and Goncales, S. (2006). Predictable dynamics of the s&p 500 implied volatility surface, *Journal of Business* **79**(3): 1591–1635.

Hodges, S. D. and Bedendo, M. (2009). The dynamics of the volatility skew: A kalman filter approach, *Journal of Banking and Finance* **33**: 1156–1165.

Jacod, J. and Protter, P. (2006). Risk neutral compatibility with option prices, *Preprint*, Université Paris IV and Cornell University.

Kadiyala, K., R. and Karlsson, S. (1997). Numerical methods for estimation and inference in bayesian var-models, *Journal of Applied Econometrics* **12**: 99–132.

Lindström, E., Ströjby, J., Brodén, M., Wiktorsson, M. and Holst, J. (2008). Sequential calibration of options, *Computational Statistics & Data Analysis* **52**: 2877–2891.

Litterman, R., B. (1986). Forecasting with bayesian vector autoregressions - five years of experience, *Journal of Business and Economic Statistics* (4): 25–38.

Merton, R. C. (1973). Theory of rational option pricing, *Bell Journal of Economics and Management Science* **1**: 141–183.

Nadaraya, E. A. (1964). On estimation regression, *Theory of Probablility and its Applications* **9**(1): 141–142.

Niu, L., Sala, L. and Favero, C. A. (2007). Forecasting the term structure of interest

rates: Small vs. large information set, *IGIER Working Paper* **340**.

Pesaran, M. H. and Timmermann, A. (2000). A recursive modeling approach to to predicting uk stock returns, *The Economic Journal* **110**: 159–191.

Raftery, A. E., Madigan, D. and Hoeting, J. A. (1997). Bayesian model averaging for linear regression models, *Journal of the American Statistical Association* **92**: 179–191.

Raftery, A. E., Madigan, D., Volinski, C. and Hoeting, J. A. (1996). Bayesian model averaging, *AAAI Workshop on Integrated Multiple Learned Models* pp. 77–83.

Robert, C. (2001). *The Bayesian Choice*, 2nd edition edn, Springer.

Roberts, G. O. and Sahu, S. K. (1997). Updating schemes, correlation structure, blocking and parameterization for the gibbs sampler, *Journal of the Royal Statistical Society* **59**(2): 291–37.

Schweizer, M. and Wissel, J. (2008). Term structures of implied volatilities: Abscence of arbitrage and existence results, *Mathematical Finance* **18**(1): 77–114.

Scott, D. W. (1992). *Multivariate Density Estimation: Theory, Practice, and Visualization*, John Wiley & Sons.

Shimko, D. (1993). Bounds on probability, *Risk* **6**: 33–37.

Smith, M. and Kohn, R. (1996). Nonparametric regression using bayesian variable selection, *Journal of Econometrics* (75): 317–343.

Smith, M. and Kohn, R. (1997). A bayesian approach to nonparametric bivariate regression, *Journal of the American Statistical Association* **92**(440): 1522–1535.

Watson, G. S. (1964). Smooth regression analysis, *Shankhya Series A* **26**: 359–372.

Zellner, A. (1971). *Introduction to Bayesian Inference in Econometrics*, John Wiley & Sons.

Zellner, A. (1987). On assessing prior distributions and bayesian regression analysis with g-prior distributions, *in* P. Goel and A. Zellner (eds), *Bayesian Inference and Decision Techniques*, Elsevier North-Holland, Amsterdam, pp. 233–243.

A Appendix: Practical Aspects of the Implementation

The spline knots k_i^κ and k_j^τ are set equidistantly set in D. For the MCMC algorithm, the number of iterations is $\tilde{M} + M = 120\,000$ of which nine tenth are taken as burn-in sample. The chain is thinned by taking only every 10th iterate to reduce the autocorrelation among draws.

A.1 Probability of Including a Regressor

As pointed out in Smith and Kohn (1996), the probability $p(\gamma_i|y, \gamma_{j\neq i})$ is computed by evaluating the expression for $\gamma_i = 0$ and $\gamma_i = 1$, and subsequent normalizing. Since the direct computation can become numerically unstable, it can be simplified to

$$\frac{1}{1 + (1+c)^{1/2}(S(\gamma_i = 1)/S(\gamma_i = 0))^{n/2})}$$

which behaves numerically more stable. The expression is easily found by computing the ratio of expressions (10) with $\gamma_i = 0, 1$.
As proposed in Roberts and Sahu (1997), it is advantagous to randomize the order of drawing in the Gibbs' sampler. Let π be a random permutation of the numbers $1, \ldots, q$ and π_i be its ith component. Then for $i = 1 : q$:

$$\gamma_{\pi_i}^{[m]} \sim p(\gamma_{\pi_i}|y, \gamma_{\pi_j \neq \pi_i}).$$

Another parameter in selecting a new regressor is the threshold level of the new moment-matrix's determinant. When a new regressor is highly correlated with another already included covariate, the resulting moment-matrix $X'_{\gamma_1} X_{\gamma_1}$ will be near-singular. Therefore, a threshold level $\tilde{t}$ should be introduced such that if $\det(X'_{\gamma_1} X_{\gamma_1}) < \tilde{t}$ the new potential regressor is not included. We set $\tilde{t} = 10^{-20}$ to ensure that it does not severly affect the algorithm.

A.2 MCMC Convergence Diagnostics

To test whether the sequence of draws of the selection variable $\gamma^{[i]}$ are from the stationary distribution, that is, whether the algorithm has converged, we apply the test statistic proposed by Geweke (1992) to test for the equality of two ergodic averages. The statistic is given by

$$CD = \frac{\bar{\gamma}_A - \bar{\gamma}_B}{\sqrt{\hat{\sigma}_A^2 + \hat{\sigma}_B^2}}$$

where $\bar{\gamma}_A$ denote the mean of n_A consequent draws at the beginning of the drawn sequence and $\hat{\sigma_A}^2$ its estimated variance. $\bar{\gamma}_B$ is defined analogously but with draws from the end of the sequence. Usually A and B are chosen to be the first and the last 30% of draws. Under the null of convergence the test statistic CD is asymptotically standard normal distributed for fixed ratios n_A/N and n_B/N. For the test we use the first and last 1000 draws after the burn-in sample has been discarded. With a few exceptions all values are within the 95%-bands.

A.3 Hyperparameters in the VAR Dynamics

The variables in the VAR can change from day to day since they depend on the estimation results over the selected period which here is a rolling window. Thus it is hardly possible to specify informative hyperparameters. We choose, as for the estimation of each surface, a shirkage prior, $\underline{\mathbf{B}} = \mathbf{0}$, with a large prior covariance matrix, $\underline{\mathbf{S}}_\beta = \underline{s}_\beta \mathbf{I}$ and $\underline{s}_\beta = 10^6$.

A.4 Dividends

In order to derive a reliable estimate of an IVS, also the dividend stream δ_t that the underlying asset pays is needed. Unfortunately, there exists no such market quotation for the S&P500 index. This quantity can be estimated from market data with a cross-section regression using the well-established put-call-parity (see e.g. Lindström et al., 2008), [7]

$$\delta_{i,t} = -(1/\tau_i)\left(\log(C_t(K_i,\tau_i) - P_t(K_i,\tau_i) + K_i e^{-r_t\tau_i}) - \log(S_t)\right) + e_{i,t}.$$

This equation can be estimated along the cross-section dimension by the sample mean. [8] In figure A.1 the estimate for the whole sample period is plotted. We use these raw estimates. Instead, one could also model them as constant and use their estimated average annual dividend of about 1.5%.

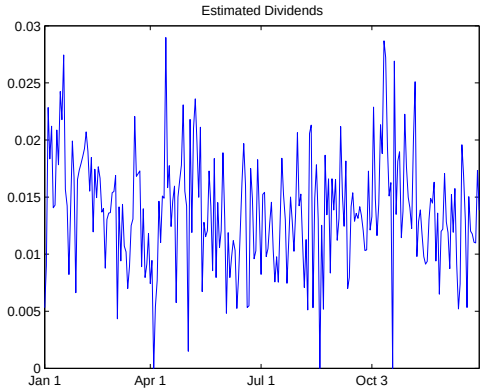

FIG. A.1 Estimated dividend stream on the S&P500 index.

[7] Actually, Lindström et al. (2008) do not estimate the dividends themselves but the dividend-corrected price of the underlying, $S_t e^{-\delta_t \tau}$.

[8] We do not take a simple average but truncate the lower and upper 20% of the ordered dividends to avoid outliers.

Chapter 3

An Econometric Framework for Simulating the Effects of Policy Shocks on the Dynamics of the Italian Sovereign Debt

This chapter is a joint work with Professor Carlo A. Favero,
CEPR, IGIER and Department of Finance, Universitá Bocconi, Milan, Italy.

Abstract

In this paper we develop an econometric model for simulating the Italian sovereign debt in a European context. We model the dependence between the sovereign debt and the term structure of interest rates as well as policy variables. The Italian yield curve is modeled throough the spread to the German yield curve. The factor spreads between the two curves are explained by monetary and fiscal policy variables as in the novel approach proposed by Favero and Giglio (2006). Noting that the yields at medium and long maturities are significantly affected by market expectations about the sustainability of the Italian fiscal policy, we distinguish an optimistic and a pessimistic market expectations regime. The Italian public debt dynamics are predicted under different policy and market expecations scenarios. The estimation results indicate that the monetary and the expectations regimes virtually coincide. Solving the whole model numerically forward, we provide impulse response functions for a European monetary and an Italian fiscal policy shock in either regime. We find that the public debt absorbs shocks better in the optimistic regime which allows us to conclude that joining the EMU has also brought fiscal stability to Italy.

Key words: sovereign debt prediction, yield curve dependence, market expectations
JEL classification: E60, C50

1 Introduction

A government that faces a primary deficit can finance it through the issuance of bonds. Auctions that place these bonds in the market are usually held at the end of each fiscal year. The price for such bonds, or equivalently their yields, depend on their term but also on country specific risk factors. For most countries these interest rates are quoted in terms of the spread to a reference yield curve. That spread reflects the market view on the bonds effective return, that is after an adjustment for default and currency risks as well as for the bonds' liquidity and how these are taxed. Most of these factors cannot or only marginally be influenced by a government. Among those that the government can target the default risk is often the most important. The risk of a credit event becomes larger when the market sees an increasing chance that the government does not follow a sustainable debt path which is determined by the government's fiscal decisions. In this paper we develop an econometric model that describes the interdependencies between the yield curve, public debt and policy variables. We focus our attention on the Italian economy. The model is tailored to simulate the effects of economic shocks on the main aggregate Italian economic variables, in particular the sovereign debt.

Before the inception of the common currency EURO in the European Economic and Monetary Union (EMU) in 1999, yield spreads on government bonds were determined by the exchange rate risk, the tax treatment in the corresponding country, the sovereign credit risk and the liquidity of these bonds. In the EMU period only the latter two factors remain as all bonds are denominated in the same currency and differences in tax treatments have been eliminated well before 1999. Gómez-Puig (2008) provides empirical evidence that in the EMU such domestic factors dominate in explaining the yield spreads rather than international risk factors which are commonly referred to as the main drivers of yield spreads on sovereign bonds issued by emerging economies. In

Europe German government bonds are commonly referred to as the benchmark. Germany is the largest European economy and has been mostly in line with the Maastricht criteria. [1] The country risk premium that is charged on those bonds is negligibly small. Since 1999 the European Central Bank (ECB) conducts a unified monetary policy for the whole Euro area. The ECB overnight-rate applies to all Euro-member countries and thus the short-ends of the term structures are equal for all Euro-member countries since then. Consequently, we will refer to the European-German (EG) as the reference yield curve at times. Even before January 1999 Banca d'Italia could not set the Italian overnight-rate entirely independently. It had to keep the exchange rate in a target zone. That period is known as the European Monetary System (EMS). Hence, we consider two monetary regimes.

Also at the medium and long maturities the Italian yield curve does not move independently of the German. However, as pointed out e.g. by Dai and Philippon (2005) at these maturities the yields are dominated by fiscal rather than by monetary policy. A sustainable fiscal policy will *ceteris paribus* narrow the spread whereas running up a large public debt ratio will increase it. In this sense, fiscal policy variables are used to proxy the likelihood of a credit event.

In figure 1 we sketch the building blocks and their interdependencies. One-sided arrows indicate simple dependencies whereas double-sided arrows denote an interaction. We take the German yield curve as exogenous and model the spread to the Italian yield curve. The yields on Italian government bonds determine the cost of financing its primary deficit and thus affect the public debt. This has an impact on the medium and long term yields. But it also affects the public spending and taxation. Moreover, its fiscal policy decisions depend also on GDP and inflation. The model is closed by noting that the output gap and inflation explain the short term spread of the yield curves

[1] These basically comprise targets for inflation, annual deficit and sovereign debt, exchange rate and long-term interest rates.

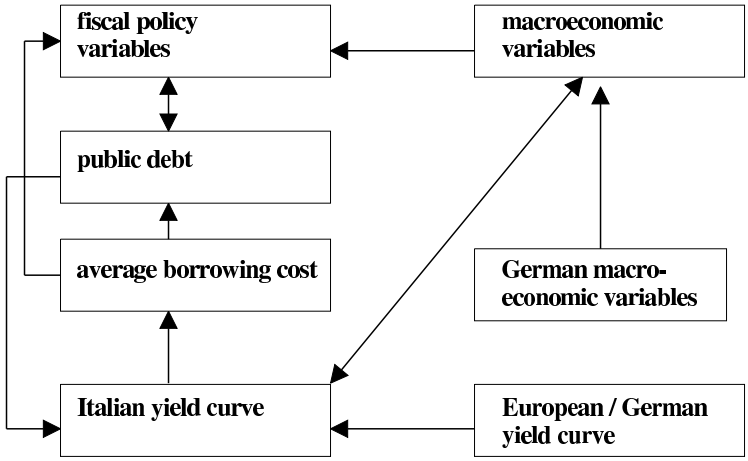

Fig. 1. Sketch of the economic variables and their interdependences within our model.

as these are typical Taylor (1993) rule variables. To model the dependence between the Italian yield curve and the German term structure of interest rates, we employ the novel approach proposed by Favero and Giglio (2006), in which the authors explain the three factor spreads by monetary and fiscal policy variables. The medium term and the long term factors are dominated by fiscal policy and the market expectations about it. In this context they use the debt ratio and its first difference to capture its evolution. Market expectations enter the specification through regime-dependent coefficients. The NS approach is a convenient way to model the spreads between the two yield curves. One might claim that the NS approach leaves the possibility of arbitrage opportunities, but we do not aim at bond pricing but to model and predict the curve's dynamics and its impact on the public debt. And, as shown in Diebold and Li (2006) and Niu et al. (2007), the NS approach to modeling the yield curve performs well regarding these issues. Furthermore, the NS parameterization permits to disentangle short-, medium- and long-term factors which allows to model market expectations in the equations for the latter two factor spreads with Markov-switching coefficients. This is used to capture time-varying effects of fiscal variables on the term structure of interest rates.

The aim of this paper is to simulate different economic shocks to our model and to compute impulse responses (IR) of the main Italian aggregates to these shocks. We are particularly interested in the effects across the different regimes.
Among others, recently published empirical models for a country's public debt can be found in Bergström and Holmlund (2000) for the Swedish economy and Pick and Anthony (2006) for the United Kingdom. In contrast to those models, we have to account for the fact that Italy joined the EMU and thus in 1999 has introduced the Euro. The relation between the term structure of interest rates and macroeconomic policy has been analyzed, for instance, by Turnovski (1989) using an IS-LM type model and by Fisher and Turnovski (1992) using a representative agent model. Our model, however, has a more empirical focus.
The paper is structured as follows. In the second section the model for the German yield curve and its dynamics are described. The third and fourth section discuss the model for the Italian yield spreads and the economy. This is followed by an impulse response analysis. The last section concludes. The majority of estimation results is deferred to the appendix.

2 German Yield Curve

In the Euro-area the risk premium on German government bonds is the smallest and, hence, yields on bonds of other countries are quoted in terms of the spread between the corresponding yields. The number of different maturities of bonds can be large and thus it is beneficial to summarize the yield curve by a model that represents the curve with a low number of factors. Nelson and Siegel (1987), or NS for short, use an unrestricted model with three factors and one loading-related variable. At a given time t the NS model of the term structure of interest rates for yields on bonds with maturity

τ is given by

$$r_t^{EG}(\tau) = L_t^{EG} + \frac{1-e^{-\tau\lambda}}{\tau\lambda} S_t^{EG} + \left(\frac{1-e^{-\tau\lambda}}{\tau\lambda} - e^{-\tau\lambda}\right) C_t^{EG} + e_{t,\tau}^{EG} \tag{1}$$

where the three factors L_t, S_t and C_t refer to the level, slope and curvature of the term structure. We use a reparameterized version of (1), which has been proposed by Diebold and Li (2006). The new factors are $f_{1t}^{EG} = L_t^{EG} + S_t^{EG}$, $f_{2t}^{EG} = L_t^{EG}$ and $f_{3t}^{EG} = C_t^{EG}$. It is easy to see that $\lim_{\tau\to 0} r_t(\tau) = f_{1t}^{EG}$ and $\lim_{\tau\to\infty} r_t(\tau) = f_{2t}^{EG}$. Thus f_{1t}^{EG} is referred to as the monetary policy instrument and f_{2t}^{EG} to as long-term factor. The curvature f_{3t}^{EG} loads medium-term maturities. Specification (1) becomes

$$r_t^{EG}(\tau) = \frac{1-e^{-\tau\lambda}}{\tau\lambda} f_{1t}^{EG} + \left(1 - \frac{1-e^{-\tau\lambda}}{\tau\lambda}\right) f_{2t}^{EG} + \left(\frac{1-e^{-\tau\lambda}}{\tau\lambda} - e^{-\tau\lambda}\right) f_{3t}^{EG} + e_{t,\tau}^{EG}. \tag{2}$$

The loadings of these factors are depicted figure B.1 in the appendix.

The loading parameter λ controls the exponential decay of the slope and curvature in the original specification (1). It could be jointly estimated using nonlinear least squares or simply be calibrated a priori. In the literature (e.g. Diebold and Li, 2006) it has been proposed to set it equal to the maximizing argument of the medium-term factor at a medium maturity such as $\tau = 30$ months and it keep it constant over time. This results in $\lambda = 0.0609$. We adobt this approach since cross-sectional estimates of λ may not be very precise since usually a realtively small number of bond prices in observed. Also, some authors (e.g. Yu and Zivot, 2008) claim that the factor loadings are not extremely sensitive to different values of the loading parameter. More recently, Koopman et al. (2007) have focussed on the dynamic nature of λ and model it as a stochastic process. However, in that case the factor loadings become time-varying, too, which renders their economic interpretation more difficult.

To model the dynamics of yield curve we specify a time series model for its three factors. All three factors are integrated of order one. Applying the Johanson (1988) test for cointegration we find one cointegrating relation (equally with the maximum eigen-

value and the trace test) for either regime.

Eventually, we do not discriminate the monetary regimes for the reference term structure due to the fact that both the Bundesbank and the ECB pursued similar monetary policies. In an analysis of comparing the costs of four countries of joining the EMU, Fair (1998) concludes that it can be assumed that 'the behavior of the European monetary authority is (...) the same as the historically estimated behavior of the Bundesbank, except the response is now to the total variables of the four countries rather than just the German variables'.

As the factors are integrated, we cannot not solve the model forward to compute impulse response functions using a plain VAR for the factors. We consider the factors in differences and specify a vector error correction model (VECM). Hence, stacking the factors as $\mathbf{f}_t^{EG} = (f_{1t}^{EG}, f_{2t}^{Ger}, f_{3t}^{Ger})'$,

$$\Delta \mathbf{f}_t^{EG} = \mathbf{a}_1 + \boldsymbol{\alpha}_f \boldsymbol{\beta}_f' \mathbf{f}_{t-1}^{EG} + \sum_{i=1}^{3} \mathbf{A}_i \Delta \mathbf{f}_{t-i}^{EG} + \mathbf{e}_t^{EG}. \tag{3}$$

The cointegration matrices $\boldsymbol{\alpha}_f$ and $\boldsymbol{\beta}_f$ are both of dimension or 3-by-1 (estimation results in table 4). The lag-length is chosen such as to avoid residual autocorrelation (table 5). The VECM is not extended by Taylor (1993) rule variables but is a simple projection of the three factors. The forecasting performance of this specification without additional macroeconomic variables has been superior to models that included such variables.

In figure 2 the impluse response functions (IR) of the three factors of the German term structure to a monetary policy shock are shown. Since all factors are I(1) changes appear permanent. The first factor is increased. Its new steady state level is reached after about 8 quarters and is about twice as large as the original shock. The second factor needs about the same time as the first to stabilize at its new level. It is, however, negative. In summary, given a shock to the first factor, short and medium term interest rates increase and long term rates decrease. Thus a monetary policy shock turns the

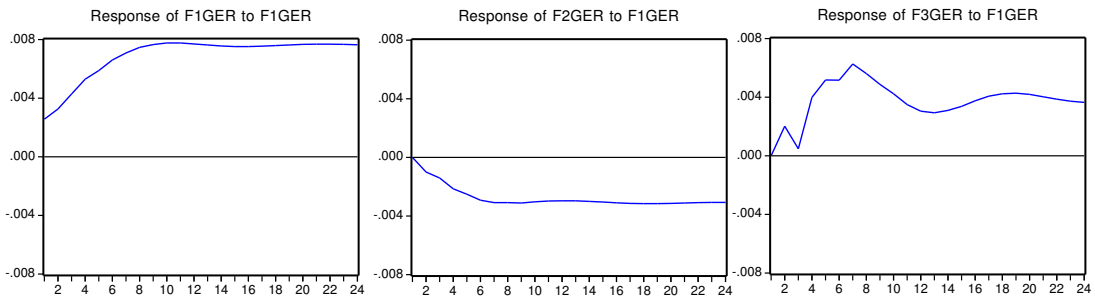

Fig. 2. Impulse Response Functions of the German Yield Curve Factors to a Monetary Policy Shock.

yield curve.

3 Italian Yield Curve and Market Expectations

The Italian yield curve model is exactly the same as for the German yield curve. The estimates of the three yield curve factors for the two countries are depicted in figures B.3 - B.5 in the appendix. The relation between both curves is modeled by the spreads between the three factors which are explained by economic variables. This part of our model is adopted from Favero and Giglio (2006) where details on the MCMC algorithm to estimate the equations jointly can be found.

At the short end the two monetary regimes ard labeled by $R_t \in \{0, 1\}$ for the EMS and EMU, respectively. The spread is explained by the Taylor (1993) rule variables inflation and output gap. It is assumed that Banca d'Italia targeted Germany's inflation. The specification reads

$$f_{1t}^{Ita} - f_{1t}^{EG} = \alpha^R + \gamma_1^R (f_{1t-1}^{Ita} - f_{1t-1}^{EG}) + \gamma_2^R (\pi_{t-1}^{Ita} - \pi_{t-1}^{EG}) + \gamma_3^R y_{t-1}^{Ita} + e_{1t}. \tag{4}$$

Similarly, at medium and long terms two regimes are distinguished. These regimes are labeled pessimistic and optimistic market expectations regime. All coefficients are super-scripted by S to denote their dependence on the regime which, since unobserved, is modeled as a latent Markov chain with stationary transition probabilities $P(S_t = i|S_{t-1} = j) = p_{ij}$, $i, j \in \{0, 1\}$. Hence, each coefficient can take two values depending on the current regime. In economic terms, time-varying coefficients account for the market's opinion about the sustainability of the government's fiscal policy. Over most of the sample period government spending exceeded tax revenues (see figure B.2 in the appendix).

For the long-term and for the medium-term factor Favero and Giglio (2006) specify

$$f_{2t}^{Ita} - f_{2t}^{Ger} = \delta_1^S(d_t - d_{t-1}) + \delta_2^S(d_t - 0.6) + e_{2t} \tag{5}$$

and

$$f_{3t}^{Ita} - f_{3t}^{Ger} = \theta_1^S(d_t - d_{t-1}) + \theta_2^S(d_t - 0.6) + e_{3t}, \tag{6}$$

respectively. No intercept is needed in (5) and (6) because any constant difference is already captured by the first factor as it already contains the level of the yield curve.

Liquidity is an important factor in determining the yield spread (see e.g. Gómez-Puig, 2008). However, here it plays no role because the Italian and German sovereign debt markets are the two largest in the Euro-area. For this reason the yield spread model (4) - (6) contains no factors such as the size of the market or the bid-ask spread.

The estimated factor spreads are shown in the appendix in figures B.3 - B.5. The coefficient estimates for the two regimes are displayed in table 1. The regimes are identifiedby the size of the coefficient estimates. If the regime shows larger estimates that translates into a larger risk premium that is charged on Italian government bonds for a given public debt ratio and its first difference and hence is labeled pessimistic. In the optimistic regime the coefficients of the f_1-spread are constrained to zero since the

short rates are equal by definition. This is empirically confirmed by observing the convergence of the short-term rates in figure B.3.

In figure 3 the filtered probabilities of being in the pessimistic regime are depicted.

coefficient on	pessimistic	optimistic
α	0.0079	0
γ_1	0.7641	0
γ_2	0.1025	0
γ_3	-5.971·10^{-4}	0
δ_1	0.1818	0.0268
δ_2	0.0654	0.0177
θ_1	0.1897	0.0575
θ_2	0.0998	-0.0069

Tab 1. Coefficient estimates the term structure factors.

The figure shows that until 1996 markets were pessimistic and turned optimistic afterwards.

It should be noted that in (5) and (6) $d_t - 0.6$ is used which is the deviation from the Maastricht criterion. In the Maastricht treaty it is determined that the threshold ratio of debt to GDP is 60% and should not be exceeded. Furthermore, it should be noted that Δd_t can be interpreted as the difference between the actual deficit, $g_t - \tau_t$, and the debt stabilizing deficit, $-i_t d_{t-1}$. [2]

The importance of modeling market expectations not constant but time-varying has

[2] The latter can be easily seen by solving $\Delta d_t = 0$ in the debt dynamics (9) for the surplus $g_t - \tau_t$.

Fig. 3. Probability of being in the pessimistic regime.

been pointed out by Engen and Hubbard (2004) who argue that the effect of fiscal policy on yield curve cannot be assumed constant over time.

In the sequel we distinguish only two rather than four regimes (pre-EMU/pes, EMU/pes, pre-EMU/opt, EMU/opt). These two regimes are called pessimistic and optimistic and refer to the time of the pre-EMU when markets were pessimistic and to the EMU-era with optimistic expectations. In our dataset these two regimes are found until mid-1996 and from 1999 onwards. This is motivated by the fact that the period of jointly observing pre-EMU and optimistic markets is short (only 10 quarters) and the other combination is not observed at all. Also the switch of the expectations from pessimistic to optimistic in 1996 could be interpreted as an anticipation of joining the EMU.

4 Macroeconomy and Public Debt Dynamics

The model for the Italian economy consists of three parts: macroeconomic fundamentals, fiscal policy and public debt.

First, inflation and output gap respond to short and long term rates and are autore-

gressive. Moreover, they may depend on their reference counterparts. More precisely, the extended VECM for the Italian macroeconomic variables inflation and output gap reads

$$\begin{pmatrix} \Delta y_t \\ \Delta \pi_t \end{pmatrix} = \mathbf{b}_1 + \mathbf{B}_1 \begin{pmatrix} y_{t-1} \\ \pi_{t-1} \end{pmatrix} + \mathbf{B}_2 \begin{pmatrix} \Delta y_{t-1} \\ \Delta \pi_{t-1} \end{pmatrix} + \mathbf{b}_2 f_{1t}^{Ita} + \mathbf{b}_3 f_{2t}^{Ita} + \mathbf{B}_3 \begin{pmatrix} y_t^{EG} \\ \pi_t^{EG} \end{pmatrix} + \mathbf{e}_t^{mp}, \tag{7}$$

where $\mathbf{B}_1, \mathbf{B}_2$ and $\mathbf{B}_3$ are 2×2 parameter matrices and $\mathbf{b}_1, \mathbf{b}_2$ and $\mathbf{b}_3$ are 2×1 parameter vectors (estimation results in tables 7 and 8). The output gap is stationary by construction whereas inflation is not (tables 3 and 4). In specification (7) the variables appear on the left-hand side in differences. This allows the nonstationary inflation to cointegrate with its reference counterpart π_t^{EG}.

Second, the fiscal policy variables react to the average borrowing cost and to its debt ratio as well as to its first difference. The government is assumed to respond to macroeconomic and fiscal information. It does not consider yield curve movements themselves. It only takes a single variable summary of the yield curve, i_t, into account which we will refer to as the average borrowing cost. To avoid the problem of simultaneity, we use its lag. Consequently, for the government we specify the following fiscal reaction function

$$\begin{pmatrix} \Delta g_t \\ \Delta \tau_t \end{pmatrix} = \mathbf{c}_1 + \mathbf{C} \begin{pmatrix} \Delta g_{t-1} \\ \Delta \tau_{t-1} \end{pmatrix} + \mathbf{c}_2 i_{t-1} + \mathbf{c}_3 (d_{t-1} - 0.6) + \mathbf{c}_4 \Delta d_{t-1} + \mathbf{e}_t^{fp} \tag{8}$$

with parameters $\mathbf{C}$ and $\mathbf{c}_1, \ldots, \mathbf{c}_4$ of dimensions 2×2 and 2×1, respectively (estimation results in tables 9 and 10). The fiscal policy variables are written in differences as they

appear to be integrated of order 1 if written in levels. No cointegration has been found neither with the maximum eigenvalue nor with the trace test in either regime.

In the literature (see e.g. Bohn, 2005) single equation models for the primary surplus are employed. Indeed, the fiscal reaction enters only the public debt dynamics in form of the deficit. Our specification includes the single equation model for the first difference of the primary surplus as a special case if for the entries of C held $c_{11} - c_{21} = c_{12} - c_{22}$. We test that null hypothesis with a likelihood-ratio test and find that it can be rejected in both the EMS and EMU regime.

When setting fiscal policy variables, the government faces a budget constraint identity.[3] In terms of ratios it states that the debt ratio d_t equals the present value of the previous debt ratio plus the current deficit ratio,

$$d_t = \frac{1 + i_t}{(1 + \eta_t)(1 + \pi_t)} d_{t-1} + g_t - \tau_t \tag{9}$$

where η_t denotes the real growth rate and $(1 + \eta_t)(1 + \pi_t) = (Y_t - Y_{t-1})/Y_{t-1}$ is the nominal growth rate. Each of the following building-blocks can be estimated seperately as a standard VAR/VECM with exogenous variables without considering the debt dynamics and hence without the average borrowing cost equation and the term structure dynamics. However, simulating the model without the debt dynamics yields biased estimates of the impulse response functions (Favero and Giavazzi, 2007).

The actual development of the debt ratio is depicted in figure 4. During the whole sample period the Italian public debt ratio clearly exceeds the Maastricht reference criterion. Furthermore, the data do not support the view that the Italian government aimed at achieving that target. In 1995 the debt ratio peaked at almost 120%. Since 2000 a stabilization towards a level of slightly more than 100% is visible.

The term structure of interest rates enters the public debt dynamics through the average borrowing cost i_t. This is the average interest rate that the Italian government has

[3] Identity refers to not including an error term in the econometric specification.

Fig. 4. Evolution of the Italian public debt ratio.

to pay on the bonds that it has issued to finance to debts. It depends on previous yields because the government must still serve interest payments on bonds which it has sold at times $s < t$. More formally, let n denote the number of bond issuances with different times to maturity τ_i and their proportions in the current debt structure ρ_i then

$$i_t = f(r_s^{Ita}(\tau_i), s \leq t, \rho_i, i = 1, \ldots, n) \tag{10}$$

with $r_s^{Ita}(\tau_i)$ being the time-s yield on a τ_i-time-to-maturity bond. Equation (10) has a complex input and is thus difficult to compute. For policy recommendations it will be useful to distinguish between short and long term bond issuance strategies. Here our focus is different and the financing structure is ignored. We therefore approximate its dependence on previous yields by adding an autoregressive term and write $i_t \approx f(i_{t-1}, f_{1t}^{Ita}, f_{2t}^{Ita}, f_{3t}^{Ita})$. To further simplify the structure we estimate a first-order Taylor expansion of it which results in the following linear regression as approximation to equation (10)

$$i_t = a_i + b_i i_{t-1} + c_i f_{1t}^{Ita} + d_i f_{2t}^{Ita} + e_i f_{3t}^{Ita} + \varepsilon_t. \tag{11}$$

We estimate the average borrowing cost equation by least squares. The estimates for this regression are reported in table 2. An extended version of this specification that includes inflation yields an insignificant coefficient and hence we use the specification (11).

coefficient on	pessimistic	optimistic
$const$	0.005850	0.004012
i_{t-1}	0.962379	0.802180
f_{1t}^{Ita}	0.064456	0.132175
f_{2t}^{Ita}	0.021153	0.022279
f_{3t}^{Ita}	-0.019617	-0.018894

Tab 2. Coefficient estimates for the average borrowing cost approximation.

A remarkable property of the estimation results is that the coefficients almost exactly sum to unity in both regimes. Hence, as one would have expected, a parallel shift of the yield curve will lead one-to-one to a shift of the average borrowing cost in the long-run.

5 Impulse Response Analysis

In this section the model that we have set up in the previous sections is simulated with exogenous shocks. In empirical studies simulated exogenous shocks are given to an econometric model in order to analyze the effects of deviations from the estimated rule. Alternatively, one could consider deviations from the systematic rule. We focus on the former which avoids problems arising from reduced form VAR. See Favero (2001) for a discussion.

Our simulation model consists of ten equations: two in each of the two VAR, the debt dynamics, the HP-filter relation for the output gap and GDP, the average borrowing cost and the three factor spreads. The IR for the model are carried out using generalized impulse response functions as introduced by Pesaran and Shin (1998). Its use is advantageous if no clear order of the variables is given by economic theory and thus it avoids the typical identification problem of the Choleski-decomposition. Generalized impulse response functions start from the original definition of ordinary IR. The only assumptions are normality and no serial correlation in the error terms. The former can be relaxed. Formally, the generalized IR for a variable x_t to a shock to a variable z_t at some horizont h is defined as

$$GI_x(h) = E(x_{t+h}|\varepsilon_{z,t} = 1, \mathcal{I}_t) - E(x_{t+h}|\mathcal{I}_t) \tag{12}$$

where $\mathcal{I}_t$ denotes the time.t information set and ε_z the error term in the equation that determines z_t. In practice equation (12) can be computed by adding a unit shock to one equation and a corresponding shock to all other equations in the same VAR. The size of those shocks equals the covariance between the error terms over the variance of original shock term. This procedure takes into account the contemporaneous correlations between the series in the VAR/VECM to which the shock is added.

The general procedure for computing the IR in this kind of model has been taken from Favero and Giavazzi (2007). First, the VARs, the yield spread equations and the average borrowing cost equation are estimated. Then the model is numerically solved forward and the residual series are saved. Second, the model is solved two more time with the calibrated series: once with a shock a described above and once with no shocks at all. Last, the difference between both solutions is computed. These differences correspond to the IR functions. The confidence bands are estimated by standard bootstrapping of the residuals. At every iteration each resampled residual series is added to the corresponding data series. The VARs are re-estimated on the bootstrapped data and

thereafter the aforementioned procedure is run again for one thousand times. For each observation over time the 5- and 95- percentile are taken as lower and upper bound and thus represent 90% pointwise confidence intervals. In all of the following figures the green line marks the original solution and the two blue lines are the confidence bands. The same procedure has been applied to both regimes accounting for the different covariations of the VAR and VECM residuals in both periods.

5.1 European Monetary Policy Shock

A European monetary policy shock is simulated by giving a unit percentage point shock to the first equation of the VAR (3) which we earlier referred to as the European / German monetary policy instrument. The other two factors in the same equation are also assigned shocks according to the methology explained above. Since we have modeled that VAR independent of the monetary regime, there is no change in the statistical model or the data and thus not in the residual covariance matrix. Consequently, the shocks are the equal across both regimes.

The IR are depicted in figures 5 and 6. In the optimistic regime the first and third yield curve factor are increased whereas the second factor decreases. That fall does not offset the increment of the other two factors and thus the average borrowing cost rises permanently. Consequently, also the public debt ratio elevates slightly over time. Government spending and taxation do not change significantly. The output falls marginally below its natural level. Inflation remains unaffected.

For the pessimistic regime the impulse responses are depicted in figure 6. All three yield curve factors are increased. As a result, the average borrowing cost rises by a quarter percentage point within four years. Since the predicted fall in tax revenues is larger than the decrement of government spending, the public debt rises sharply. After four years its mean predicted increment is almost 10%. Output and inflation remain

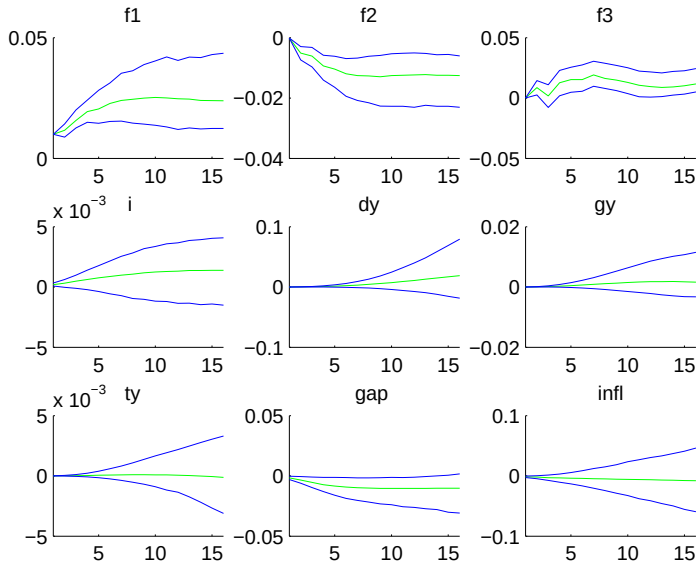

Fig. 5. IR to a European monetary policy shock in the optimistic expectations regime.

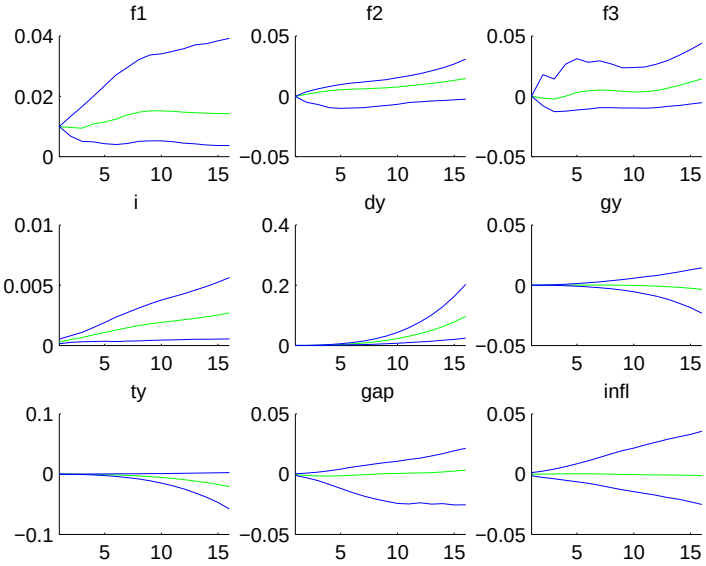

Fig. 6. IR to a European monetary policy shock in the pessimistic expectations regime.

unaffected.
In summary, the European/German monetary policy shock scenario reveals a clear difference across expectations regimes: in the pessimistic regime the predicted increase in public debt is larger, which stems from a larger rise in the average borrowing cost and a widening spread between taxes and government spending.

5.2 Government Spending Shock

To simulate the effects of an Italian government spending shock, we assign a unit percentage increment to the residual of Δg_t in the VAR for the Italian fiscal policy variables (8). In contrary to the IR analysis of the monetary policy shock, in the fiscal shock setup the shock itself changes. The Italian macroeconomy and fiscal policy are estimated seperately for both regimes. Hence, the covariance matrices of the error terms change and entails different shocks in both regimes. The IR are shown in figures 7 and 8.
In the optimistic regime the first yield curve factor is fixed as we have assumed that the European monetary authority does not respond to Italian economic variables. Effectively all other variables change insignificantly. The marginal increase in government spending is offset by an equal increase in tax revenues. All other effects are not statistically significant.
In the pessimistic regime, the short end of the Italian term structure is allowed to react to changes in the macroeconomic variables and so it does. The monetary policy instrument decreases slightly whereas the medium and long term factor reveal an upwards trend. The two effects on the average borrowing cost almost offset each other and it increases only slightly. However, the predicted spread between tax revenues and government spending increases and therefore also the public debt ratio rises. Again, output gap and inflation remain unchanged.

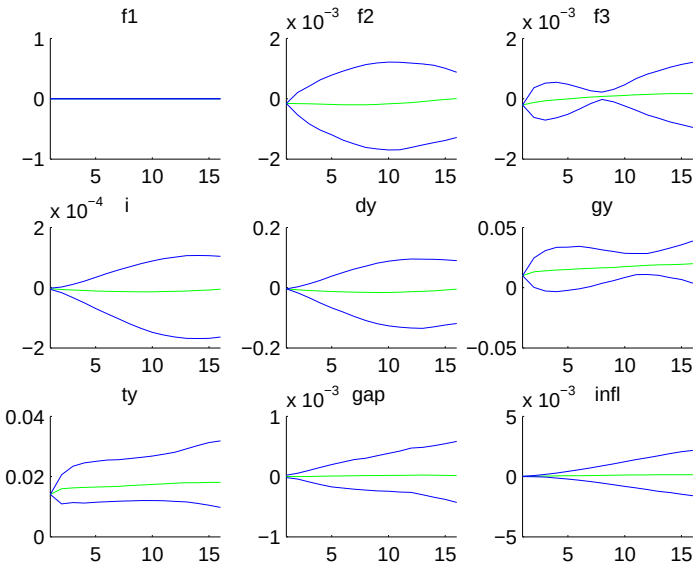

Fig. 7. IR to a government spending shock in the optimistic expectations regime.

In summary, the public debt ratio stabilizes in the optimistic regime whereas it does not in the pessimistic regime. This can be mostly ascribed to the predicted widening spread between expenditures and taxes. But also the medium and long term yield curve factors increase in the pessimistic regime and hence affect the cost of financing the debt.

6 Conclusions

In this paper we have set up an econometric model that allows to analyze the effects of economic shocks on the Italian public debt. Binary latent market expectations are considered and two monetary regimes are distinguished. As these roughly coincide we merge them into two distinct regimes. The comparison of the impulse responses in these two regimes, named optimistic and pessimistic, reveals significant differences. Under both simulated scenarios the impulse respose analysis shows that in the optimistic regime the public debt ratio evolves more stable than in the pessimistic ex-

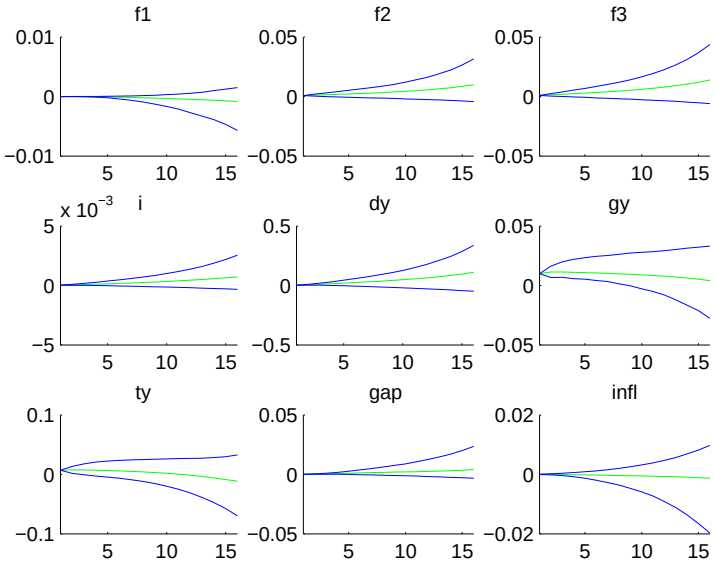

Fig. 8. IR to a government spending shock in the pessimistic expectations regime.

pectations regime. This indicates that being a member of the Euro area has brought economic robustness to the Italian economy.

Further research could extend the paper regarding the set of scenarios. For instance, simulating an expectations shock, that is an increase in the long-term interest rates, or an inflation shock might be of particular interest. Another strain of research could augment the model to different maturity bonds and thus to an optimal issuance strategy for the government. Clearly, the strategy will depend on the market expectations on the future fiscal policy as those affect the factors of the Italian yield curve. Hence, this strain of research will further analyze the dependencies between the variables in the model.

References

Bergström, P. and Holmlund, A. (2000). A simulation model framework for government debt analysis, *The Swedish National Debt Office Research* .

Bohn, H. (2005). Sustainability of fiscal policy in the u.s., *University of California Santa Barbara Working Paper* .

Dai, Q. and Philippon, T. (2005). Fiscal policy and the term structure of interest rates, *NBER Working Paper* **11574**.

Diebold, F. X. and Li, C. (2006). Forecasting the term structure of government bonds, *Journal of Econometrics* **130**(4): 337–364.

Engen, E. and Hubbard, R. G. (2004). Federal government debt and interest rates, *NBER Working Paper* **10068**.

Fair, R. (1998). Estimated stabilization costs of the emu, *National Institute Economic Review* (164): 90–99.

Favero, C. (2001). *Applied Macroeconometrics*, Oxford University Press.

Favero, C. A. and Giavazzi, F. (2007). Debt and the effect of fiscal policy, *IGIER Working Paper* **317**.

Favero, C. A. and Giglio, S. (2006). Fiscal policy and the term structure of interest rates: Evidence from the case of italy in the ems and emu periods, *IGIER Working Paper* **312**.

Fisher, W. and Turnovski, S. (1992). Fiscal policy and the term structure of interest rates: An intertemporal optimizing analysis, *Journal of Money, Credit and Banking* **24**(1): 1–26.

Gómez-Puig, M. (2008). Monetary integration and the cost of borrowing, *Journal of International Money and Finance* **27**: 455–479.

Hodrick, R. and Prescott, E. (1997). Post-war u.s. business cycles: An empirical investigation, *Journal of Money, Credit and Banking* **29**(1): 1–16.

Johanson, S. (1988). Statistical analysis of cointegration vectors, *Journal of Economic Dynmamics and Control* **12**: 231–254.

Koopman, S., Mallee, M. and van der Wel, M. (2007). Analyzing the term structure of interest rates using the dynamic nelson-siegel model with time-varying parameters, *Tinbergen Institute Discussion Paper* **TI 2007 - 095/4**.

Nelson, C. R. and Siegel, A. E. (1987). Parsimonious modeling of yield curves, *Journal of Business* **60**(4): 473–489.

Niu, L., Sala, L. and Favero, C. A. (2007). Forecasting the term structure of interest rates: Small vs. large information set, *IGIER Working Paper* **340**.

Pesaran, H. M. and Shin, Y. (1998). Generalized impulse response analysis in linear multivariate models, *Economics Letters* **58**: 17–29.

Pick, A. and Anthony, M. (2006). A simulation model for the analysis of the uk's sovereign debt strategy, *UK Debt Management Office Research* .

Taylor, J. B. (1993). Discretion versus policy rules in practice, *Carnegie Rochester Conference Series on Public Policy* **39**: 195–214.

Turnovski, S. (1989). The term structure of interest rates and the effects of macroeconomic policy, *Journal of Money, Credit and Banking* **21**(3): 321–347.

Yu, W. and Zivot, E. (2008). Forecasting the term structure of treasury and corporate bonds: Dynamic nelson-siegel models evaluated, *University of Washington Working Paper* .

A Data

For the estimation and simulation of our model we use quarterly data for the European, German and Italian variables. All times series are taken from the OECD Country Statistical Profiles. The debt data are provided by the Italian Treasury. All ratios are

computed with respect to the nominal GDP. The expenditure data do not contain the interest payment on current public debt. The average borrowing cost, that is the artificial interest rate the Italian government has to pay for its debts, is computed as interest payment over lagged debt.

The output gap is computed with the Hodrick and Prescott (1997) filter with a smoothing parameter $\lambda = 1600$ as usual for quarterly series, that is $y_t = Y_t - Y_t^{HP}$ where Y_t^{HP} is the long-run trend of the GDP. [4] All data have been sampled from 1991:1 to 2005:4 at the end of each period.

For the short-end of the yield curve which we denote by f_{1t}^{EG} it holds

$$f_{1t}^{EG} = \begin{cases} f_{1t}^{Eur} & \text{if } t \geq \text{1999:1}, \\ f_{1t}^{Ger} & \text{if } t < \text{1999:1} \end{cases}$$

and analogously for the inflation π_t^{EG} and the output gap y_t^{EG}. In either monetary regime, German term structure of interest rates has been reference to the Italian and only the short-end is regime-dependent.

B Data Statistics, Estimation Results and Figures

In this appendix all estimation results from the models that are not presented in main text are given. Values in squared brackets denote t-statistics.

[4] Since we model inflation and output gap jointly in a VAR the output gap is re-scaled as to have approximately the same magnitude as inflation.

variable	p-value	conclusion	p-value	conclusion
y_t	0.0012	I(0)	0.0005	I(0)
π_t	0.5619	I(1)	0.2008	I(1)
g_t	0.3664	I(1)	0.3467	I(1)
τ_t	0.1064	I(1)	0.6040	I(1)

Tab 3. p-values from testing the hypothesis of nonstationarity in the optimistic (left) and pessimistic (right) regime with augmented Dickey-Fuller test.

$f^{EG}_{1,t-1}$	1
$f^{EG}_{2,t-1}$	-0.4827
	[-3.2756]
$f^{EG}_{3,t-1}$	-0.7949
	[-13.6912]
c	-0.028483
	[-3.2457]

Tab 4a. Estimates for error correction part of the EG-factor VECM.

variable	Δf_1^{EG}	Δf_2^{EG}	Δf_3^{EG}
$\boldsymbol{\alpha}_f$	-0.08896	0.5459	0.5864
	[-0.7626]	[4.5483]	[0.82253]
$\Delta f_{1,t-1}^{EG}$	0.2963	0.2121	-0.4275
	[1.2009]	[0.8357]	[-0.2835]
$\Delta f_{1,t-2}^{EG}$	-0.06400	0.2383	-1.7689
	[-0.2676]	[0.9685]	[-1.2103]
$\Delta f_{1,t-3}^{EG}$	0.3022	-0.2986	1.2894
	[1.49926]	[-1.4394]	[1.0466]
$\Delta f_{2,t-1}^{EG}$	-0.2980	-0.1155	-2.3485
	[-1.5709]	[-0.5918]	[-2.0257]
$\Delta f_{2,t-2}^{EG}$	0.2019	-0.1390	1.0022
	[1.2437]	[-0.8325]	[1.0104]
$\Delta f_{2,t-3}^{EG}$	0.28064	0.09497	0.9482
	[1.79609]	[0.59070]	[0.99310]
$\Delta f_{3,t-1}^{EG}$	0.05949	0.23142	0.37211
	[0.6929]	[2.6197]	[0.7093]
$\Delta f_{3,t-2}^{EG}$	0.01722	0.1219	0.0004666
	[0.2746]	[1.8877]	[0.00122]
$\Delta f_{3,t-3}^{EG}$	0.04039	0.1066	0.3203
	[0.7965]	[2.0438]	[1.0334]

Tab 4b. Estimates for lagged variables of the EG-factor VECM.

lag	test-statistic	p-value	lag	test-statistic	p-value
1	8.7385	0.4618	11	7.7217	0.5624
2	13.7864	0.1301	12	9.2804	0.4118
3	8.1827	0.5158	13	8.2252	0.5116
4	7.8263	0.5517	14	6.7522	0.6629
5	3.8227	0.9227	15	10.0748	0.3445
6	8.0725	0.5269	16	13.1622	0.1554
7	6.4023	0.6991	17	9.4532	0.3965
8	13.0390	0.1608	18	10.4554	0.3149
9	7.5087	0.5843	19	8.9244	0.4443
10	7.0488	0.6320	20	14.5988	0.1026

Tab 5. p-values from testing the hypothesis of no autocorrelation for the residual of the *EG*-VECM.

variable	$\Delta\tau_t$	Δg_t
$\Delta\tau_{t-1}$	0.292653	-0.098654
	[1.03029]	[-0.15746]
Δg_{t-1}	0.018331	0.477389
	[0.16145]	[1.90619]
c	0.002750	0.006258
	[2.06840]	[2.13358]
i_{t-1}	0.044203	0.256647
	[0.74245]	[1.95430]
Δd_{t-1}	0.004331	0.001534
	[1.16699]	[0.18735]
$d_{t-1} - 0.6$	-0.007107	-0.019449
	[-2.12430]	[-2.63559]

Tab 6. Estimated for the fiscal policy VECM (optimistic regime).

variable	$\Delta\tau_t$	Δg_t
$\Delta\tau_{t-1}$	0.761425	0.192958
	[1.29743]	[0.39707]
Δg_{t-1}	-0.005712	0.498249
	[-0.01018]	[1.07194]
c	0.011949	-0.002658
	[0.92788]	[-0.24924]
i_{t-1}	-0.240110	0.113757
	[-0.83924]	[0.48018]
Δd_{t-1}	-0.008456	-0.022711
	[-0.47067]	[-1.52664]
$d_{t-1} - 0.6$	-0.009518	-0.000568
	[-0.92001]	[-0.06636]

Tab 7. Estimated for the fiscal policy VECM (pessimistic regime).

variable	variable/estimate	estimate
y_{t-1}	1	
π_{t-1}	-0.000633	
	[-0.132]	
c	-0.0832	
	[-1.40]	

	Δy_t	$\Delta \pi_t$
$\boldsymbol{\alpha}_m$	-0.3516	-24.49
	[-3.94]	[-2.76]
Δy_{t-1}	0.4868	-6.940
	[2.698]	[-0.388]
$\Delta \pi_{t-1}$	-0.00219	-0.228
	[-1.086]	[-1.138]
f_{1t}^{Ita}	-0.00232	-0.1060
	[-2.15]	[-0.992]
f_{2t}^{Ita}	$2.45 \cdot 10^{-5}$	-0.0316
	[-0.0218]	[-0.284]
y_t^{EG}	0.0649	3.940
	[2.28]	[1.40]
π_t^{EG}	0.0270	0.1880
	[3.74]	[2.63]

Tab 8. Estimates for the macroeconomic VECM (optimistic regime).

variable	variable/estimate	estimate
y_{t-1}	1	
π_{t-1}	0.1417	
	[5.44]	
c	0.01206	
	[3.54]	

	Δy_t	$\Delta \pi_t$
$\boldsymbol{\alpha}_m$	-0.00899	-21.05
	[-2.08]	[-3.48]
Δy_{t-1}	0.8720	24.397
	[5.95]	[1.19]
$\Delta \pi_{t-1}$	0.00088	0.3216
	[0.695]	[1.82]
f_{1t}^{Ita}	0.00089	0.1327
	[0.885]	[0.939]
f_{2t}^{Ita}	$-6.06 \cdot 10^{-6}$	-0.00249
	[-0.081]	[-0.0238]
y_t^{EG}	-0.0091	1.113
	[-0.164]	[0.144]
π_t^{EG}	0.00021	0.0455
	[2.17]	[3.35]

Tab 9. Estimates for the macroeconomic VECM (pessimistic regime).

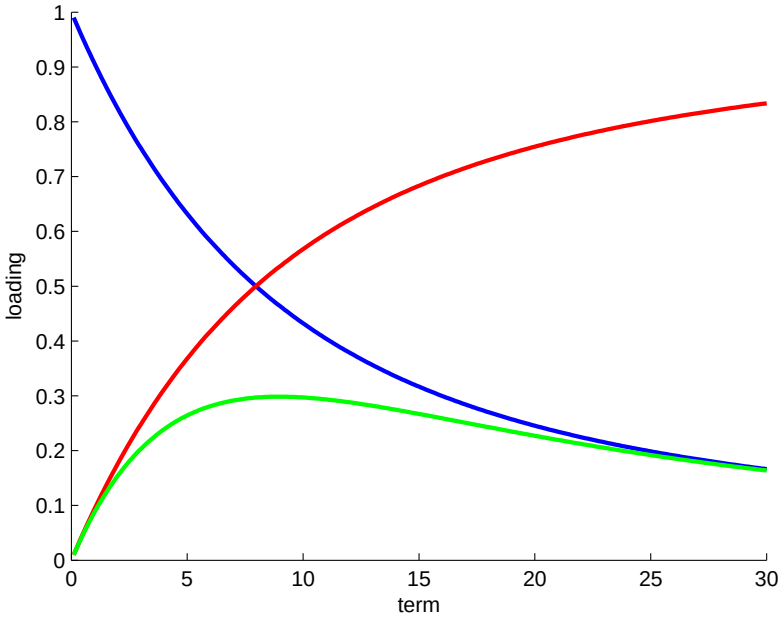

Fig. B.1. Diebold-Li-factor loadings: : f_1 (blue), f_2 (red), f_3 (green).

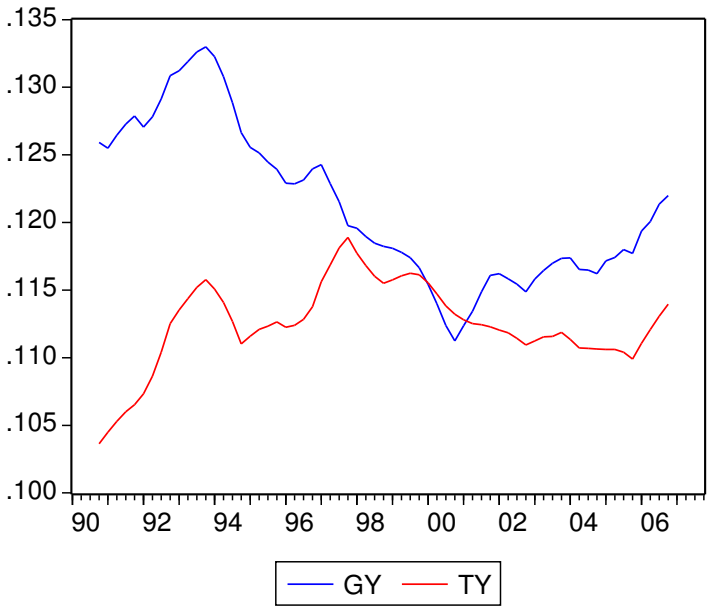

Fig. B.2. Italian government spending (red) and tax revenue (blue) to GDP ratios.

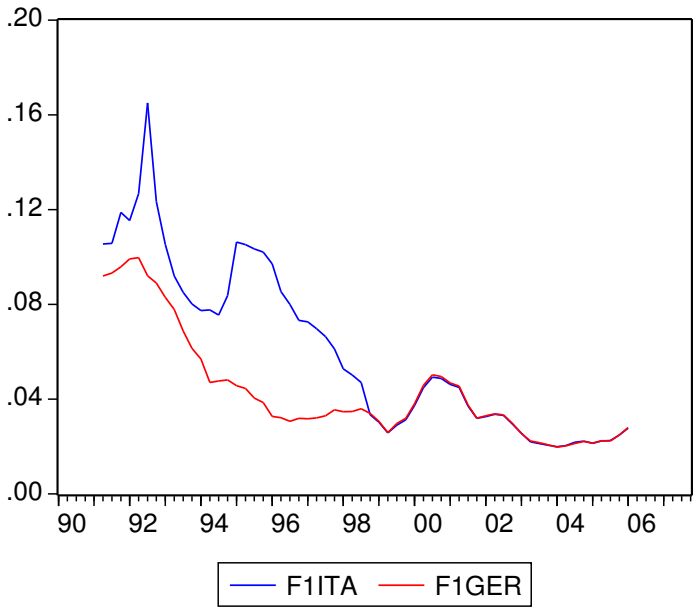

Fig. B.3. Short-term interest rate spread between Italian and German yields.

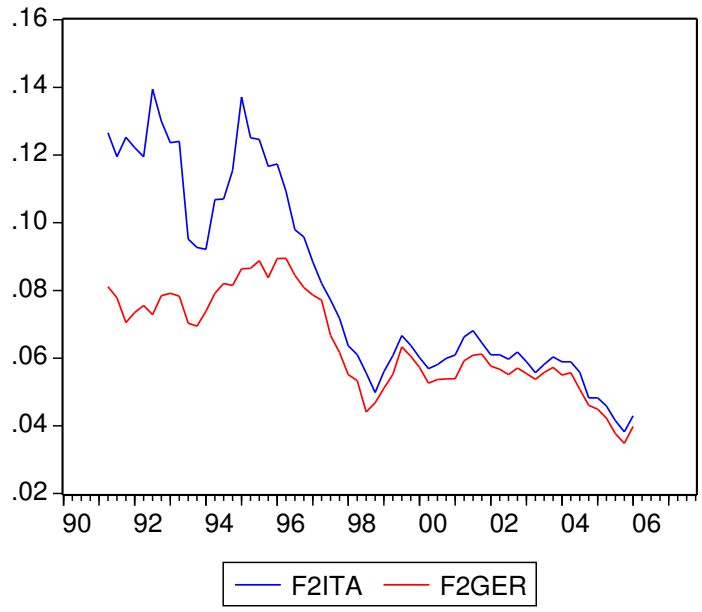

Fig. B.4. Long-term interest rate spread between Italian and German yields.

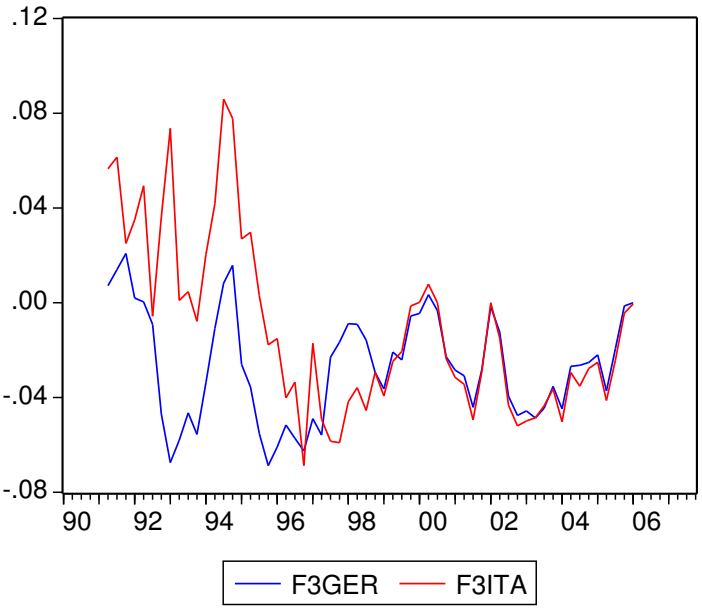

Fig. B.5. Medium-term interest rate spread between Italian and German yields.