A Forecast Evaluation of PCA-Based Adaptive Forecasting Schemes for the EURIBOR Swap Term Structure

als Inaugural-Dissertation
zur Erlangung des akademischen Grades eines Doktors
der Wirtschafts- und Sozialwissenschaften
der Wirtschafts- und Sozialwissenschaftlichen Fakultät
der Christian-Albrechts-Universität zu Kiel

vorgelegt von

M.Sc. (Statistics), Diplom–Volkswirt, Statisticien Économiste
Oliver Jim Blaskowitz
geb. 23.08.1974 in Berlin / Deutschland

Berlin, im September 2009

Gedruckt mit Genehmigung der
Wirtschafts- und Sozialwissenschaftlichen Fakultät
der Christian-Albrechts-Universität zu Kiel

Dekan: Professor Dr. Thomas Lux
Erstberichterstattender: Professor Dr. Helmut Herwartz
Zweitberichterstattender: Professor Dr. Thomas Lux

Tag der Abgabe der Arbeit: 15. September 2009
Tag der mündlichen Prüfung: 18. Dezember 2009

Acknowledgements

First of all, I want to thank Professor Dr. Helmut Herwartz very much for supervising this thesis. I want to underline that the numerous inspiring discussions with him, his continuing support, encouragement and patience were an important source for the progress of the thesis.

I am also very grateful to PD Dr. Bernd Droge for all the insightful conversations about econometrics, statistics and topics not related to any research issues.

Furthermore, I would also like to thank all current and former members of the Chair of Econometrics of the Humboldt–Universität zu Berlin with whom I shared more or less time working as a research and teaching assistant for the enjoyable environment they offered me.

Finally, but not least important, I want to express my deepest gratitude to my parents, my sister, my aunt and my friends for their support and understanding that I had to cope with so many things which they might have considered too abstract to be true.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

While term structure modelling has undergone extensive improvements, advances in term structure forecasting are comparatively small, yet the latter is particularly important for purposes of managing risk or hedging derivatives. Diebold and Li (2006) point out that the empirical performance of model based out–of–sample forecasts is rather poor. Reformulating the Nelson and Siegel (1987) model, they use autoregressive (AR) models for the factors to obtain encouraging results for long horizon ex–ante forecasts. While Diebold and Li (2006) find that forecasts based on vector autoregressive models (VARs) outperform forecasts implied by the random walk model, Duffee (2002) concludes that the random walk model is superior to standard affine term structure models.

Ang and Piazzesi (2003) model yield curves by means of traditional latent yield factors and observable macroeconomic variables. Forecast error variance decompositions show that macro factors explain up to 85% of the variation in bond yields. Taking a dynamic factor approach, Diebold, Rudebusch and Aruoba (2006) model the term structure by means of latent level, slope and curvature factors as well as macroeconomic variables like real activity, inflation and the federal funds rate. They find convincing evidence of macroeconomic effects on yields. Mönch (2008) forecasts the yield curve in a data rich environment. He uses a factor augmented VAR jointly with an affine term structure model and accounts for parameter restrictions implied by a no–arbitrage condition. The model turns out to outperform various benchmark models such as, for instance, a random walk, a standard VAR and the Diebold and Li (2006) approach, among others.

Though a large part of the term structure literature is concerned with factor models, a uniform conclusion with regard to the appropriate number of factors has not been achieved yet. Nelson and Siegel (1987) introduce a parsimonious three factor model for term structures and conclude that it is able to capture important yield curve characteristics. Numerous extensions of the Nelson–Siegel model exist. Inter alia, a two factor version is applied by Diebold, Piazzesi and Rudebusch (2005) and the four factor version from Svennson (1994) is frequently used by central banks (BIS, 2005). The principal component, respectively, factor analysis (PCA) of Steeley (1990) and Litterman and Scheinkman (1991) suggests that most of the term structure variation can be explained by three factors, interpreted as level, slope and curvature. Examining money market returns, Knez, Litterman and Scheinkman (1994) advocate a four factor model and argue that the additional component is related to private issuer credit spreads. Duffie and Singleton (1997) propose a multi factor model for interest rate swaps that accommodates counterparty default risk and liquidity differences between Treasury and Swap markets. They conclude that credit and liquidity effects are important sources to explain swap term structure dynamics. Within this framework, Liu, Longstaff and Mandell (2006) estimate a five factor model to analyze swap spreads.

To explain forecast failures of macroeconomic models, Clements and Hendry (2002), among others, argue that economies are subject to, e.g., institutional or technological changes. Neglecting changes in economic relations is a potential reason for the poor performance of model based out–

of–sample term structure forecasts. To allow for dynamic heterogeneity, data based adaptive forecasting procedures appear to be useful alternatives. Swanson and White (1997a,b) find that an adaptive approach yields promising results in forecasting macroeconomic variables. A particular issue in dynamic ex–ante forecasting is the stability of model parameters. Splitting a sample of US government interest rates covering the period January 1970 to December 1995 into three parts, Bliss (1997) concludes that factor loadings varied only slightly, although factor volatilities turned out to be relatively unstable. For US zero coupon bond yields, Audrino, Barone–Adesi and Mira (2005) find that loadings are time varying during the period from January 1986 to May 1995 in a three factor model allowing for conditional heteroscedasticity.

A large fraction of the term structure literature is concerned with the US treasury market. However, Remolona and Wooldridge (2003) point out that the EURO interest rate swap market has become one of the largest and most liquid markets world wide. The enormous increase in hedging and positioning activity tripled the turnover in Euro denominated interest rate swaps between 2000 and 2006 (ECB, 2007).

Due to the huge size of swap markets and the neglected attention paid to forecasting the term structure, we focus on forecasting the EURIBOR (European interbank offered rate) swap term structure. Employing a purely statistical factor model approach, the term structure of swap rates is decomposed by means of PCA, and AR models are applied for (adaptive) forecasting. Using various combinations of the number of factors, AR orders and rolling window sizes, we obtain a set of 100 alternative model specifications. These are evaluated in terms of various measures evaluating forecast performance.

1.1 Structure

This thesis is organized as follows. The factor model approach is given in the next Chapter. In Chapter 3 we present a first empirical case study in which we describe the data set and motivate the local heterogenous approach. To statistically assess the forecasting performance for particular rates and the level, slope and curvature of the swap term structure, we rely on the Henrikkson–Merton statistic. Economic performance is investigated by means of cash flows implied by alternative trading strategies. Moreover, an ANOVA based data adaptive model selection procedure is found to provide a promising forecast performance as compared to forecasting schemes that rely on global homogeneity of the term structure.

A deeper analysis of the suitability of data adaptive model selection strategies is conducted in Chapter 4. Within a unified loss functional framework, selected adaptive combination methods are subjected to an extensive forecast comparison experiment. To evaluate ex-ante forecasting performance for particular rates, distinct forecast features, such as mean squared errors, directional accuracy and directional forecast value, are considered. It turns out that, relative to benchmark models, the adaptive approach offers additional forecast accuracy in terms of directional accuracy and directional forecast value.

In Chapter 3 we use the Henrikkson–Merton statistic to measure the economic value of directional forecasts in the sense of Merton (1981). Common approaches to test for the latter value are based on the classical χ^2–test for independence, Fisher's exact test or the Pesaran and Timmerman (1992) test for market timing. These tests are asymptotically valid for serially independent observations. Yet, in the presence of serial correlation they are highly oversized as confirmed in a simulation study. We summarize serial correlation robust test procedures and propose a bootstrap approach in Chapter

5. By means of a Monte Carlo study we illustrate the relative merits of the latter. Two empirical applications demonstrate the relevance to account for serial correlation in economic time series when testing for the value of directional forecasts. By doing this, we provide further support for the preferred Median strategy introduced in Chapter 4.

Finally, we close with some concluding remarks in Chapter 6. We point out that directional forecasts can provide a convenient framework to assess the economic forecast value when loss functions (or success measures) are properly formulated to account for realized signs and realized magnitudes of directional movements. We discuss a general approach to evaluate (directional) forecasts which is simple to implement, robust to outlying or unreasonable forecasts and which provides an economically interpretable loss/success functional framework. As such, the measure of directional forecast value is a readily available alternative to the commonly used squared error loss criterion.

Chapter 2

A Factor Model for the Swap Rate Term Structure

Over a sample period of approximately 10 years, we consider daily quotes of swap rates for the following $M = 10$ maturities: 3m (3 months), 6m, 1yr (1 year(s)), 2yr, 3yr, 5yr, 7yr, 10yr, 12yr and 15yr. Due to the large dimension M factor models are in widespread use when modeling term structures. One may a–priori question the adequacy of structurally invariant dynamic models to hold over a sample period of 10 years. Therefore, a view of local structural invariance is adopted to implement the factor model conditional on time windows of size τ. The following model is mainly used to provide rolling ex–ante forecasts of the swap rate structure or the underlying factors:

$$\tilde{y}_t = \Gamma_K F_t + \xi_t, \quad t = T^* - \tau + 1, \ldots, T^*, \tag{2.1}$$

$$\Delta F_t = \nu + \Phi_1 \Delta F_{t-1} + \ldots + \Phi_p \Delta F_{t-p} + \eta_t. \tag{2.2}$$

In (2.1) $\tilde{y}_t = (\tilde{y}_{1t}, \tilde{y}_{2t}, \ldots, \tilde{y}_{Mt})'$ is a vector of swap rates over 10 maturities measured in terms of deviations from their local mean, $\tilde{y}_t = y_t - \bar{y}_{T^*}$, $\bar{y}_{T^*} = 1/\tau \sum_{t=T^*-\tau+1}^{T^*} y_t$. For notational convenience the dependence of the local mean on the window size τ is not indicated. F_t is a K–dimensional vector of factors, $F_t = (f_{1t}, ..., f_{Kt})'$, governing the term structure whose changes exhibit VAR dynamics. The $M-$ respectively $K-$dimensional error terms ξ_t and η_t are treated as white noise processes. If the covariance matrix of ξ_t is not diagonal and, instead, some weak correlation between elements of ξ_t is allowed, (2.1) is interpreted as an approximate factor model (Chamberlain and Rothschild, 1983). As argued below, potential contemporaneous dependence of η_t and ξ_t is not crucial for purposes of ex–ante forecasting.

We formalize ex–ante forecasts of the swap rates by means of conditional expectations implied by (2.1) and (2.2), i.e.

$$E[\tilde{y}_{T^*+h}|\Omega_{T^*,\tau}] = \Gamma_K E[F_{T^*+h}|\Omega_{T^*,\tau}], \tag{2.3}$$

$$\text{with } E[F_{T^*+h}|\Omega_{T^*,\tau}] = F_{T^*} + \sum_{j=1}^{h} E[\Delta F_{T^*+j}|\Omega_{T^*,\tau}]\,. \tag{2.4}$$

In (2.3) and (2.4) it is assumed that $E[\eta_{T^*+j}|\Omega_{T^*,\tau}] = 0$ for $1 \leq j \leq h$ and $E[\xi_{T^*+h}|\Omega_{T^*,\tau}] = 0$ such that potential contemporaneous dependence linking η_t and ξ_t is discarded. Swap rate forecasts are then given by

$$\hat{y}_{T^*+h} = \Gamma_K \hat{F}_{T^*+h} + \bar{y}_{T^*},$$

where factor implied forecasts are readjusted for the local in–sample mean.

The matrix Γ_K in (2.1) is obtained by means of PCA which uses a decomposition of the unconditional covariance matrix of $\tilde{y}_t$ observed over the period $t = T^* - \tau + 1, \ldots, T^*$, i.e.

$$\hat{\Sigma}_{T^*} = \frac{1}{\tau} \sum_{t=T^*-\tau+1}^{T^*} \tilde{y}_t \tilde{y}_t', \quad \hat{\Sigma}_{T^*} = \Gamma \Lambda \Gamma'. \tag{2.5}$$

In (2.5) Λ is a diagonal matrix of eigenvalues of $\hat{\Sigma}_{T^*}$ in decreasing order and the columns of Γ contain the corresponding eigenvectors, respectively. Then, the matrix Γ_K given in (2.1) contains the first K columns of Γ thereby accounting for the variation in $\tilde{y}_t$ driven by K principal components. Being aware of the differences in the concepts underlying principal component and factor analysis (Johnson and Wichern, 2002) we consider principal components as factors, and thus use both terms interchangeably.

Although the extraction of principal components from mean adjusted interest rates is common practice a word of caution seems appropriate. In the econometric literature interest rates are often diagnosed to be integrated processes while the covariance estimator in (2.5) is suitable only in case of stationary swap rates. The latter argument is particularly relevant since (2.1) describes the dynamics of persistent rates over (short) local time windows. Given the potential of nonstationarity, some of the extracted eigenvectors are likely to allow a similar interpretation as (unidentified) cointegration parameters (Johansen, 1995). In addition, in case of a nonstationary swap term structure PCA results in extracting at least some nonstationary factors.

At the first sight a VAR representation of variables which are orthogonal by construction appears unnecessarily general. Note, however, that orthogonality does not imply absence of serial cross correlation. From the perspective of ex–ante forecasting it might pay to condition estimates of a future factor not only on its own history but also on the remaining factors. The latter issue is empirically justified by means of Granger causality tests provided along with other descriptive features of the data in Section 3.2. When implementing the VAR, it turns out that for the purpose of forecasting F_{T^*+h} (and thus y_{T^*+h}) conditional on information contained in $\Omega_{T^*,\tau} = \{y_t \mid t = T^* - \tau + 1, \ldots, T^*\}$ a model specified in first differences of the factors yields more stable results in comparison with a level representation. Note that for the model in (2.2) ν is essentially a drift parameter, in turn implying a linear trend to govern the level of interest rates. Clearly such a property is at odds with empirical features of interest rates in the long run. In our local model, however, ν captures local trends that might improve the accuracy of ex–ante interest rate predictions.

A PCA/VAR approach is a simple and flexible way of modelling term structures. Representing factor dynamics by means of VAR models is common practice, and PCA yields unrestricted estimation of factor loadings.

Hence, it imposes less structure on the loadings than, for example, Nelson–Siegel type models. Moreover, PCA allows us to specify the number of factors in a data driven manner, as opposed to models where the number of factors is fixed by a priori reasoning.

Implementing the local model in (2.1) and (2.2) an analyst has to choose the parameters τ, K and p. In this study we employ a variety of window sizes $\tau \in \{42, 63, 126, 189, 252\}$ corresponding to trading periods of 2, 3, 6, 9 and 12 months. The number of relevant factors is varied over $K = 1, 2, 3, 4, 5$, and potential lag orders are $p = 0, 1, 2, 3$. Alternative selections of the latter parameters provide a battery of 100 competing PCA/VAR forecasting models. Distinct forecast horizons are $h = 1, 5, 10$ days. Since numerous alternative parameter selections are employed for the forecasting exercises we evaluate the overall statistical and economic performance by means of an analysis of variance (ANOVA, Johnson and Wichern, 2002). The latter is implemented by regressing some key measures of statistical and economic performance (Henrikkson–Merton statistic and cash flows) on a set of dummy variables corresponding to the choice of τ, K and p. Statistical and economic performance measures are now provided in turn.

Chapter 3

Empirical Analysis: Part 1

In this Chapter we measure forecast performance by the Henrikkson–Merton (1981) statistic and cash flows as realized via alternative trading strategies. We motivate that owing to dynamic heterogeneity of the term structure, one may hardly expect a particular PCA/VAR implementation to uniformly outperform the competing specifications. Instead, similar to Härdle, Herwartz and Spokoiny (2003) we argue in favor of local homogeneity of the term structure. A data driven procedure to 'predict the best forecasting model' is proposed. The latter is shown to outperform forecasting schemes building on global homogeneity of the term structure.

The remainder of the Chapter is organized as follows. The performance measures are discussed in the next Section. Section 3.2 introduces the data and offers a descriptive analysis to motivate the subsequent modeling approach. Sections 3.3 to 3.5 provide empirical results, the adaptive model selection strategy and a comparative discussion of forecasting performance. Section 3.6 concludes.

3.1 Performance Measures

3.1.1 Statistical measure of forecasting performance

We focus on changes of particular swap rates or of linear combinations of swap rates denoted $g_{T^*,h}(\theta) = \theta' y_{T^*+h} - \theta' y_{T^*}$, with $\theta \in R^M$. The corresponding forecast conditional on information available in T^* is $\hat{g}_{T^*,h}(\theta) = \theta' \hat{y}_{T^*+h} - \theta' y_{T^*}$. In the spirit of Merton (1981) a forecasting model is valuable if the distributional properties of the dichotomous forecasts $\hat{g}_{T^*,h}(\theta) \geq 0$ (or $\hat{g}_{T^*,h}(\theta) < 0$) come close to the characteristics of the corresponding realizations $g_{T^*,h}(\theta) \geq 0$ (or $g_{T^*,h}(\theta) < 0$). To summarize key features of the joint distribution of categorical variables contingency tables are often used in applied statistics. The Henrikkson–Merton (hm) statistic is the conditional probability of correctly forecasting a positive or negative value of $\hat{g}_{T^*,h}(\theta)$ given a positive or negative realization in time $T^* + h$, i.e.

$$\text{hm} = \Pr\left(\hat{g}_{T^*,h}(\theta) \geq 0 \mid g_{T^*,h}(\theta) \geq 0\right) + \Pr\left(\hat{g}_{T^*,h}(\theta) < 0 \mid g_{T^*,h}(\theta) < 0\right). \quad (3.1)$$

A successful forecasting scheme should deliver hm–statistics in excess of unity.

3.1.2 Economic measure of forecasting performance

To complement the statistical evaluation we measure economic forecasting performance in terms of cash flows achieved by six alternative trading strategies. Each strategy is implemented conditional on a (forecasted) rate or a factor based signal, thereby providing 12 strategies in total.

For the first trading strategy, consider the single 2yr swap rate. For example, if we proceed in time T^* from the expectation that the 2yr rate will increase we set up a *2yrTrade* by entering a 2yr payer swap agreement. This corresponds to a rate based trading signal $\hat{g}_{T^*,h}(\theta_2) \geq 0$, with

$\theta_2 = (0,0,0,1,0,0,0,0,0,0,0)$. Alternatively, factor based signals exploit factor forecasts and a presumption concerning the correlation between factors and rates. For instance, if the first factor is expected to increase and the correlation of the first factor with the 2yr rate is positive (negative), then enter a 2yr payer (receiver) swap agreement. Empirical correlations are estimated as

$$\hat{\rho}_{km}^{T^*} \quad = \quad \widehat{\mathrm{Cor}}(f_{kT^*}, y_{mT^*}) \quad = \quad \frac{\sqrt{\hat{\lambda}_k}\gamma_{km}}{\hat{\sigma}_m}. \tag{3.2}$$

In (3.2) $\hat{\lambda}_k$ is the k–th eigenvalue of the covariance matrix $\hat{\Sigma}_{T^*}$, γ_{km} is the loading of factor k on swap rate m and $\hat{\sigma}_m$ is the estimated standard deviation of swap rate m. For ease of notation we neglect in the upper definition the dependence of all quantities on window size τ. Positive correlation between the first principal component and a particular rate is indicated if $\gamma_{1m} > 0$. In complete analogy to the *2yrTrade* strategies for mid– and long–term maturities, a *5yrTrade* (θ_5) and a *10yrTrade* (θ_{10}) are also implemented.

In addition, we consider trades based on linear combinations of single rates. A *LevelTrade* is set up to exploit upward (downward) movements of the term structure level by entering payer (receiver) swap agreements with 2yr, 5yr and 10yr maturities. We rely on two signals to predict level increases. Firstly, using the forecasts of the 2yr, 5yr and 10yr swap rates, a rate based signal is computed by means of $\hat{g}_{T^*,h}(\theta_{level})$ with $\theta_{level} = (0,0,0,\frac{1}{3},0,\frac{1}{3},0,\frac{1}{3},0,0)$. Consequently, $\hat{g}_{T^*,h}(\theta_{level}) \geq 0$ indicates an expected level increase. Secondly, according to the interpretation of the first factor as measuring the level of the term structure (e.g. Steeley, 1990, Litterman and Scheinkman, 1991) VAR predictions of this factor are used as factor based signals. If the first factor is predicted to increase the term structure level is expected to move upwards.

Next, to initiate a *SlopeTrade* a 2yr receiver (payer) and a 10yr payer (receiver) swap is entered if the slope of the term structure is expected to increase (decrease). The corresponding rate based signal is $\hat{g}_{T^*,h}(\theta_{slope}) \geq 0$, $\theta_{slope} = (0,0,0,-\frac{1}{2},0,0,0,\frac{1}{2},0,0)$, whereas a forecasted increase in the second factor is interpreted to hint at a future slope increase. Similarly, a *CurvatureTrade* is characterized by $\hat{g}_{T^*,h}(\theta_{curve}) \geq 0$, $\theta_{curve} = (0,0,0,\frac{1}{4},0,-\frac{1}{2},0,\frac{1}{4},0,0)$ (rate based) or forecasts of the third principal component (factor based). In case the curvature of the term structure is expected to increase (decrease), enter 2yr and 10yr payer (receiver), and 5yr receiver (payer) swap agreements where the former swaps are weighted with 0.25 (-0.25) and the latter with -0.5 (0.5).

In order to compute cash flows implied by the portfolios described above, we implement the comparison swap valuation technique (Miron and Swannell, 1991). To illustrate this method consider, for example, the *LevelTrade* and a $h = 1$ day ($h = 5, 10$ days) ahead forecast. If in some instant of time, T^*, the term structure level is expected to increase, enter 2yr, 5yr and 10yr payer swap agreements. Next, in time $T^* + h$ close the positions by entering receiver swap agreements with the same reduced (by one, five, ten days) time to maturity. Since our data set does not contain swap rates for maturities other than 3m, 6m, 1yr, 2yr, 3yr, 5yr, 7yr, 10yr, 12yr and 15yr, the fixed rate for a, say, 2yr (5yr, 10yr) minus one day (five, ten days) swap is approximated as the 2yr (5yr, 10yr) rate observed in $T^* + 1$ ($T^* + 5, T^* + 10$). From the difference in both rates it is possible to compute the present value of the cash flows over the remaining time to maturity. We account for accrued interest but not for transactions costs. Note that entering a swap agreement at current market rates is costless. Hence, the interest–free bank account starts at 0 Euro. Daily swap agreements add up to a notional of 100 Euros. In case of the *LevelTrade* the 2yr, 5yr and 10yr swaps each have a notional of 33.33 Euros.

Distinguishing rate and factor based signals, trading strategies are denoted R1 to R6 and F1 to F6, respectively. Table 3.1 summarizes the strategies.

ID	TradeName	Signal	2yr	5yr	10yr
R1	*2yrTrade*	2yr up	1Pay	0	0
R2	*5yrTrade*	5yr up	0	1Pay	0
R3	*10yrTrade*	10yr up	0	0	1Pay
R4	*LevelTrade*	($\frac{1}{3}$2yr + $\frac{1}{3}$5yr + $\frac{1}{3}$10yr) up	$\frac{1}{3}$Pay	$\frac{1}{3}$Pay	$\frac{1}{3}$Pay
R5	*SlopeTrade*	($\frac{1}{2}$10yr - $\frac{1}{2}$2yr) up	$\frac{1}{2}$Rec	0	$\frac{1}{2}$Pay
R6	*CurvatureTrade*	($\frac{1}{4}$2yr - $\frac{1}{2}$5yr + $\frac{1}{4}$10yr) up	$\frac{1}{4}$Pay	$\frac{1}{2}$Rec	$\frac{1}{4}$Pay
F1	*2yrTrade*	first factor up & Cor$(f_1, 2yr) > 0$	1Pay	0	0
F2	*5yrTrade*	first factor up & Cor$(f_1, 5yr) > 0$	0	1Pay	0
F3	*10yrTrade*	first factor up & Cor$(f_1, 10yr) > 0$	0	0	1Pay
F4	*LevelTrade*	first factor up	$\frac{1}{3}$Pay	$\frac{1}{3}$Pay	$\frac{1}{3}$Pay
F5	*SlopeTrade*	second factor up	$\frac{1}{2}$Rec	0	$\frac{1}{2}$Pay
F6	*CurvatureTrade*	third factor up	$\frac{1}{4}$Pay	$\frac{1}{2}$Rec	$\frac{1}{4}$Pay

Table 3.1: Summary of trading strategies. Trading strategies are set out for an expected upward movement of single rates, level, slope or curvature, respectively. For trading on expected downward movements exchange payer (Pay) with receiver (Rec) and receiver with payer swap positions.

3.2 The data and stylized facts

So far empirical term structure modeling had a distinct focus on treasury and bond yield curves. For example, yield curves are basic building blocks in econometric ESG (Economic Scenario Generator) models such as Wilkie's stochastic investment model (Wilkie 1986, 1995). See Ziemba and Mulvey (1998) for an overview of econometric ESG approaches. Yet, empirical models of swap curves are rare. Dai and Singleton (2003) point out that U.S. swap and treasury markets share similar stylized features, although the institutional structure of both markets is different. They find, for instance, that

PCA yield similar ’level’, ‘slope’ and ‘curvature’ factors for both markets. To work out some stylized facts for the European swap market, we present in this section the data set and provide a brief description of swap rate and factor dynamics. The interpretation of the first three principal components as level, slope and curvature factors and the case of structural variation of model parameters are highlighted. Moreover, motivate the factor VAR approach.

The investigated data set comprises 2395 daily vector quotes of EURIBOR swap rates covering the period February 22, 1999 to April 25, 2008. The first 252 days are exclusively used for initial training samples. The data driven model selection procedure given in Section 3.4 requires additional 42 training observations. To ensure that all PCA/VAR forecasts at horizons $h = 1, 5, 10$ are compared over the same sample of swap rates the rolling window analysis starts in time $T^* = 303$ (April 19, 2000) and obtains 2134 ex–ante predictions. Before returning to the ex–ante content of the proposed model class this Section provides in–sample (summary) statistics to characterize the Swap term structure and motivate the adopted PCA/VAR model class. In this Section empirical characteristics are mostly conditional on time windows of size $\tau = 42$. Conclusions to be drawn from other window sizes are, however, qualitatively identical.

3.2.1 Descriptive analysis

The evolution of the daily term structure is shown in Figure 3.1. It displays the variability of the term structure shape over time. For example, the level of the swap term structure is higher in October/November 2000 (around week 100) than in March 2004 (around week 280). Similarly, the slope of the term structure is higher in March 2003 (around week 210) than, e.g. in November 2007 (around week 400). Moreover, the curvature in November

2003 (around week 250) exceeds the corresponding measure in October 2001 (around week 140).

A selection of descriptive statistics is provided in Table 3.2. The shape of the swap curve varies over time with respect to its level, slope and curvature. Table 3.2 documents that the average swap term structure is increasing and rates at short maturities are more volatile than those at the long end. Approximating the curvature as 0.25*2yr - 0.5*5yr + 0.25*10yr the evidence is less clear. Both the empirical mean and median signify a slightly positive (i.e. convex) term structure, although Figure 3.1 reveals also locally distinct curvatures.

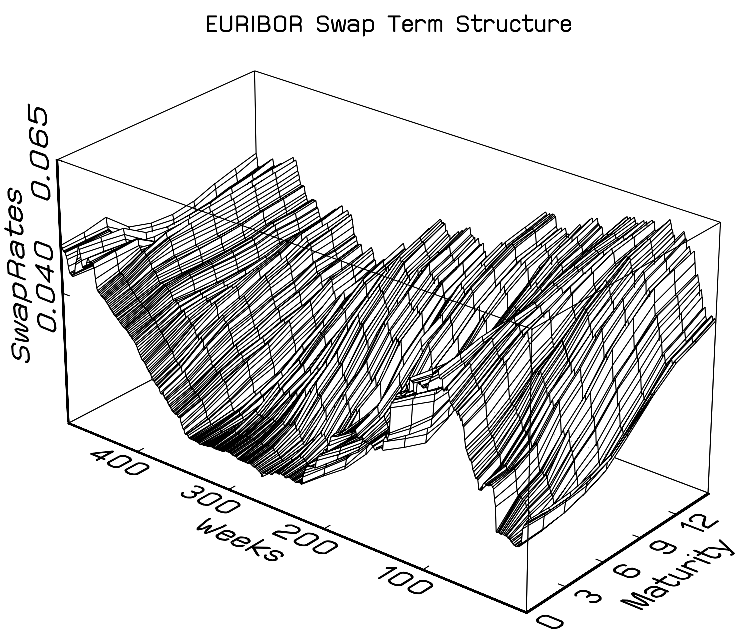

Figure 3.1: Evolution of the swap term structure from February 22, 1999 to April 25, 2008.

3.2.2 Factors and correlations

It is common practice to interpret principal components by means of component weights attached to original variables. For instance, at November 6,

	3m	6m	1yr	2yr	3yr	5yr	7yr	10yr	12yr	15yr	Level	Slope	Curvature
Mean	3.310	3.363	3.472	3.654	3.836	4.134	4.368	4.619	4.728	4.857	4.136	0.483	0.00150
Median	3.341	3.332	3.407	3.735	3.861	4.045	4.212	4.470	4.596	4.740	4.067	0.527	0.00425
Min	1.984	1.950	1.956	2.010	2.240	2.615	2.850	3.141	3.250	3.395	2.668	0.006	-0.08250
Max	5.211	5.274	5.415	5.583	5.698	5.805	5.900	6.031	6.150	6.295	5.774	0.900	0.09950
StD	0.975	0.975	0.978	0.902	0.850	0.783	0.756	0.724	0.722	0.718	0.776	0.265	0.03476

Table 3.2: Descriptive statistics of location and dispersion for actual swap rates and shape parameters for the period from February 22, 1999 to April 25, 2008. StD is the standard deviation. Statistics for the 7yr and 12yr swap rates are not reported to economize on space. Level, slope and curvature are measured by (2yr + 5yr + 10yr)$/3$, (10yr$-$2yr)$/2$ and (2yr$-2*$5yr$+$10yr)$/4$, respectively. Swap rates are multiplied by 100 for this Table only. In the remaining analysis swap rates are measured, e.g. as .0312 instead of 3.12.

2001 ($T^* = 707$) eigenvectors ($K = 3$) extracted from historic windows of size $\tau = 42$ and $\tau = 252$ are, respectively,

$$\Gamma'_{3,\tau=42} = \begin{bmatrix} .304 & .335 & .345 & .319 & .304 & .278 & .294 & .317 & .324 & .337 \\ -.564 & -.472 & -.302 & -.016 & .086 & .190 & .216 & .250 & .304 & .351 \\ .199 & .190 & -.033 & -.575 & -.478 & -.301 & .030 & .232 & .287 & .370 \end{bmatrix}$$

and

$$\Gamma'_{3,\tau=252} = \begin{bmatrix} .389 & .420 & .443 & .402 & .348 & .266 & .216 & .172 & .159 & .145 \\ -.398 & -.314 & -.193 & .007 & .101 & .245 & .327 & .395 & .422 & .441 \\ .406 & .293 & -.017 & -.449 & -.422 & -.304 & -.021 & .222 & .292 & .378 \end{bmatrix}.$$

From the composition of weights it is natural to interpret the first principal component, f_{1t}, to represent the level of the term structure since each maturity enters with some positive weight. Similarly f_{2t} is obtained giving positive weights for higher and negative weights for lower maturities thereby measuring the slope of the term structure. Using mostly negative weights for midterm maturities the third principal component, f_{3t}, approximates the curvature of the swap term structure. Although both matrices $\Gamma_{3,\tau=42}$ and $\Gamma_{3,\tau=252}$ are estimated from overlapping windows of observations the reported weighting coefficients show considerable differences. For instance, using the large time window the level of the term structure is determined with weights that are decreasing in maturity. From the smaller time window an almost constant weighting scheme over alternative maturities is obtained.

To provide a more representative interpretation of the first three principal components, $f_{kt},\ k = 1, 2, 3$, Figure 3.2 displays their average correlations with observed swap rates,

$$\hat{\rho}_{km} = \frac{1}{2083} \sum_{T^*=303}^{2385} \hat{\rho}_{km}^{T^*} .$$

Time specific correlations $\hat{\rho}_{km}^{T^*}$ are defined in (3.2). Similar to stylized features of treasury and bond term structures Figure 3.2 confirms for the EURIBOR that the first principal component measures the level of the swap term structure since, on average, it is positively correlated with interest rates at all maturities. Moreover, correlations over maturities for the second and third principal component allow an interpretation to represent the slope and curvature of the swap term structure, respectively. It is worthwhile to mention that the latter findings do not hold uniformly for all trading days in the sample. There are time instances where, for particular sample windows, the first factor is characterized by a correlation profile similar to the average correlation of the slope factor illustrated in Figure 3.2. Opposite to the three principal components the correlation patterns obtained for the fourth and fifth factor do not allow any obvious interpretation that holds uniformly over the sample period.

To underscore the issue of structural variation Figure 3.3 displays the fraction of explained variances obtained from three rolling principal components ($\tau = 42,\ K = 3$). The fraction of data variability explained by the first factor is time varying between a lower and an upper bound of about 60% and 98%, respectively. Over periods with a relatively small degree of explanation achieved with the first factor the second factor is contributing

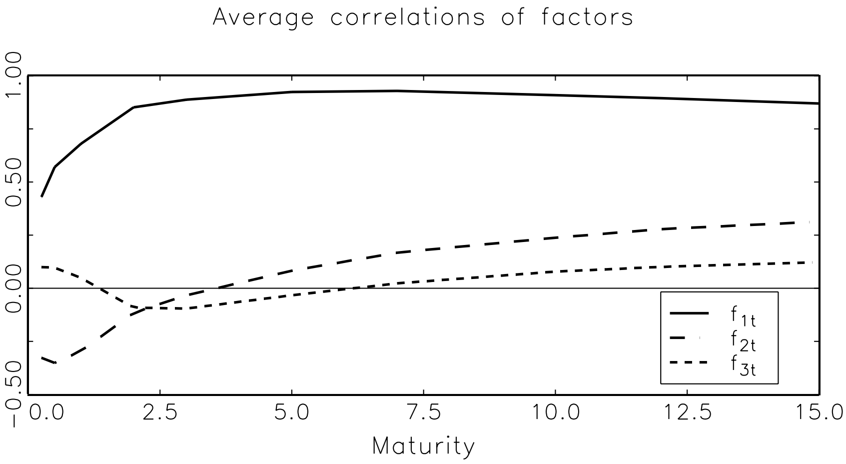

Figure 3.2: Average correlations of three principal components ($\tau = 42$) with swap rates of maturities 3 months to 15 years over the period from April 19, 2000 ($T^* = 303$) to April 11, 2008 ($T^* = 2385$).

up to 40% of explained data variation. Throughout, the contribution of the third factor is rather small. With respect to ex–ante forecasting of the swap term structure, one may conjecture from Figure 3.3 that over certain periods single factor models may provide sufficiently accurate forecasts whereas, for instance, in the second third of the forecasting period model implementations with more than one factor might give superior results. Since model parsimony is positively related with forecasting efficiency one may, moreover, expect that higher VAR orders for the dynamic factor model might be suitable in case the factor dimension is small ($K < 3$). In multiple factor models forecasting precision is likely superior if the autoregressive order of the factor VAR in (2.2) is small.

To justify the autoregressive model structure of factor differences, Table 3.3 documents some properties of autocorrelations for the first five factors. Based on a rule of thumb critical value, $2/\sqrt{42} = 0.309$, we consider the rejection frequency for the null hypothesis of no autocorrelation, as well as minimum and maximum autocorrelations at lags 1,2,5 and 10 over all 50 non–overlapping subsamples of size $\tau = 42$. It can be seen that the lev-

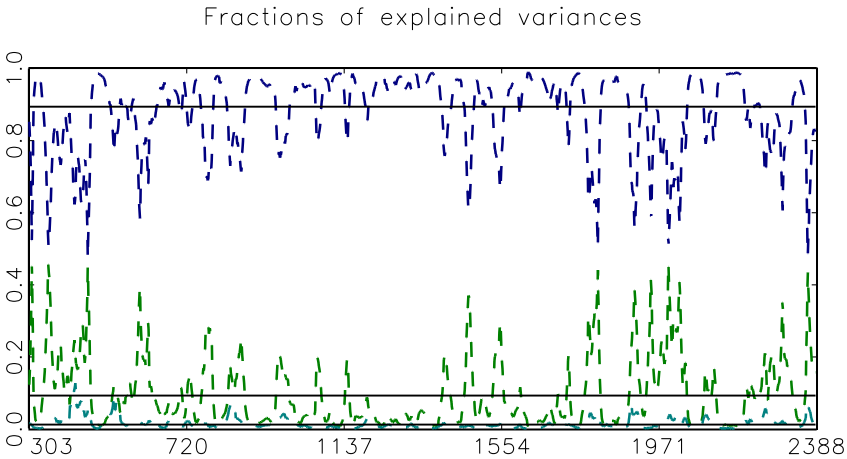

Figure 3.3: Fractions of explained variances for the period April 19, 2000 ($T^* = 303$) to April 11, 2008 ($T^* = 2385$). Dashed lines depict explained variances for three principal components obtained from rolling factor decompositions with $\tau = 42$. Solid lines show the corresponding fractions implied by the unconditional covariance matrix.

els of factors f_{1t}, f_{2t} are highly serially dependent as indicated by minimum and maximum autocorrelations. In particular, for the first two factors the hypothesis of a zero lag one resp. lag two correlation is rejected in more than 90% of the subsamples considered. Correlations at higher lag orders decrease slowly. Autocorrelations are smaller for factors f_{kt}, $k = 3, 4, 5$ and may become negative. The first differences of factors autocorrelate less. Yet, the null hypothesis of zero autocorrelation is rejected too frequently to ignore serial dependence. This is particularly true for the third, fourth and fifth factor. Moreover, from the minimum and maximum correlations it can be concluded that there are local time periods with significant serial correlation prevailing for changes of the first two factors.

Factor levels						Factor differences				
k	Autocorrelation lag					k	Autocorrelation lag			
	1	2	5	10			1	2	5	10
	.728	.459	.006	-.263	Min		-.326	-.336	-.372	-.365
1	1.0	1.0	.800	.080	α_{act}	1	.040	.020	.020	.060
	.946	.889	.688	.346	Max		.399	.254	.203	.204
	.420	.190	-.389	-.586	Min		-.435	-.399	-.498	-.332
2	1.0	.900	.400	.160	α_{act}	2	.160	.020	.100	.040
	.931	.853	.670	.335	Max		.304	.283	.359	.373
	-.063	-.301	-.289	-.478	Min		-.585	-.496	-.320	-.349
3	.640	.400	.060	.180	α_{act}	3	.420	.040	.040	.040
	.815	.671	.345	.345	Max		.177	.265	.313	.422
	-.302	-.408	-.320	-.460	Min		-.609	-.482	-.240	-.275
4	.500	.400	.020	.160	α_{act}	4	.580	.080	.060	.000
	.753	.596	.280	.206	Max		.141	.266	.451	.261
	-.231	-.284	-.423	-.322	Min		-.669	-.421	-.388	-.363
5	.280	.140	.080	.060	α_{act}	5	.760	.140	.080	.040
	.679	.493	.321	.367	Max		-.060	.341	.335	.236

Table 3.3: Rejection frequencies (α_{act}) for a zero lag j autocorrelation hypothesis (critical value given by $2/\sqrt{\tau}$), minimum and maximum autocorrelations of levels (left hand side panel) and first differences (right hand side panel) for the first five factors ($k \in \{1, ..., 5\}$) are provided. Autocorrelations are computed at lags $j = 1, 2, 5, 10$ for 50 non–overlapping windows of size $\tau = 42$ over the period April 19, 2000 to April 11, 2008.

VAR models are more general than univariate autoregressions in the sense that the former impose less structure on the factors and allow for interdependencies of distinct factors over time. A VAR approach for factor differences is supported by tests on Granger noncausality (Lütkepohl, 2005). Table 3.4 documents the outcome of Wald tests on Granger noncausality ($K = 3,\ p = 2$) applied to 50 non–overlapping time windows of size $\tau = 42$. It can be concluded that each differenced factor $\Delta f_{kt}, k = 1, 2, 3,$ is Granger caused by the two remaining factor changes at common significance levels. For example, with 5% significance the null hypothesis that Δf_{1t} and Δf_{3t} are not Granger causal for Δf_{2t} is rejected for more than 28% of the non–overlapping data windows.

	Δf_{2t} and Δf_{3t} do not cause Δf_{1t}			Δf_{1t} and Δf_{3t} do not cause Δf_{2t}			Δf_{1t} and Δf_{2t} do not cause Δf_{3t}		
α_{nom}	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10
α_{act}	.120	.140	.180	.140	.280	.400	.060	.220	.360

Table 3.4: Empirical rejection frequencies (α_{act}) of Granger noncausality tests using a VAR(2) model ($K = 3$) for first differences with alternative significance levels (α_{nom}) and non–overlapping time windows ($\tau = 42$). The total number of test decisions is 50.

Descriptive and in–sample characteristics of the EURIBOR swap term structure underpin the likelihood of time varying dynamic features in several respects as, for instance, prevalence and strengths of serial correlation or the dimensionality of the factor space. Given the in–sample support of a time varying term structure it is tempting to investigate the ex–ante content of competing PCA/VAR models in the next Sections.

3.3 Forecasting performance

This Section offers a forecast comparison for all employed PCA/VAR implementations. Since we implement 100 alternative model specifications to forecast six linear combinations of swap rates at time horizons ($h = 1, 5, 10$) we mostly refrain from providing model specific test results. Instead, by means of an ANOVA measures of statistical and economic performance are related to particular model features.

3.3.1 Statistical forecasting performance

ANOVA results for the hm–statistic are given in Tables 3.5 and 3.6 for rate and factor based trades, respectively. For completeness, the first row of both Tables gives the hm–statistic for $h = 5$ days ahead forecasts obtained for rolling implementations of local models using mostly $\tau = 42, K = 1, p = 0$ as benchmark specification. For factor based trades F5 and F6 benchmark specifications are characterized by $K = 2$ and $K = 3$, respectively, since the trading signals exploited for these strategies are directly taken from the second and third principal component. Measured in terms of averaged hm–statistics the forecasting accuracy is similar for rate and factor based signals.

The choice of locally homogeneous time windows is crucial from the viewpoint of ex–ante forecasting performance. For instance, relative to the benchmark choice ($\tau = 42$) time windows covering one year of daily quotes ($\tau = 252$) reduce average hm–statistics significantly in case of four rate based trading signals (R1, R2, R4, R6). For almost all strategies our results support the choice of small time windows ($\tau = 63$) to extract rate based trading signals. For only two factor based trading signals (F5 and F6) significant improvements of average hm–statistics are diagnosed for larger time windows of size $\tau = 189$.

The effect of including more than one factor ($K > 1$) on the average forecasting performance is mostly positive and significant for rate based signalling. The optimal number of factors varies from $K = 2$ (R1 and R5), $K = 3$ (R3 and R6) to $K = 5$ (R2 and R4). Slightly different results are obtained for factor based signals. Again, it turns out that choices of $K > 1$ improve the average performance of trading rules F1 to F6. Yet, compared with rate based signals performance improvements are smaller for all factor based trades except F5. For example, adding factors improves the hm–statistic between 0.0049 and 0.0104 units for F1, while the R1 performance increases between 0.0538 to 0.0732 units when using higher dimensional models. The statistical performance of the factor based slope trade (F5) deteriorates if $K > 2$.

Considering the preferable VAR order ANOVA estimates indicate that higher autoregressive orders involve positive (negative) impacts on the hm–statistic for the trading strategies R2, R3, R5, R6 and F2 (F3). Inconclusive results are obtained for R1, R4, F1 and F4 to F6. While it is difficult to detect a particular impact pattern over distinct lag orders, it appears that using more than one lag does not significantly improve forecasting accuracy in terms of hm–statistics. Most considered trading signals show stronger correspondence with the respective realized rates at the higher forecast horizon $h = 10$ in comparison with $h = 5$. Comparing medium with short term predictability the ANOVA documents higher accuracy of the former.

In sum, the obtained hm–statistics underscore forecasting ability of the class of PCA/VAR models. As a further indication of overall predictability it is noteworthy that average hm–statistics significantly exceed unity for all 12 trading strategies. Moreover, for all trading strategies except F6 the estimated intercept term of ANOVA regressions exceeds unity at conventional significance levels. Minimum hm–statistics, however, reveal that an unfavorable specification of the forecasting model has downside risk with respect to statistical performance.

	R1	R2	R3	R4	R5	R6
BenchHM	1.082	1.075	1.037	1.092	1.053	1.223
Const	1.063	1.044	1.014	1.054	1.016	1.149
	(11.52)	(7.85)	(2.07)	(9.27)	(2.73)	(17.76)
$\tau = 63$	**0.006**	**0.012**	**0.001**	**0.012**	**0.030**	**0.013**
	(1.25)	(2.52)	(1.65)	(2.52)	(5.90)	(1.87)
$\tau = 126$	−0.021	−0.002	−0.008	−0.007	0.014	0.009
	(−4.61)	(−0.42)	(−1.46)	(−1.34)	(2.67)	(1.25)
$\tau = 189$	−0.033	−0.040	−0.004	−0.042	0.014	−0.002
	(−6.99)	(−8.53)	(−0.78)	(−8.47)	(2.82)	(−0.33)
$\tau = 252$	−0.027	−0.031	0.008	−0.030	0.008	−0.034
	(−5.69)	(−6.67)	(1.43)	(−6.09)	(1.62)	(−4.81)
$K = 2$	**0.073**	0.037	0.022	0.019	**0.060**	0.066
	(15.75)	(7.85)	(3.86)	(3.91)	(11.80)	(9.30)
$K = 3$	0.059	0.037	**0.022**	0.021	0.005	**0.079**
	(12.70)	(7.94)	(3.93)	(4.30)	(1.05)	(11.20)
$K = 4$	0.054	0.040	0.017	0.034	0.007	0.068
	(11.58)	(8.46)	(3.06)	(6.82)	(1.37)	(9.66)
$K = 5$	0.055	**0.041**	0.021	**0.036**	−0.004	0.042
	(11.83)	(8.72)	(3.61)	(7.38)	(−0.83)	(5.93)
$p = 1$	**0.004**	**0.008**	0.003	**0.007**	**0.012**	0.030
	(0.85)	(1.93)	(0.64)	(1.59)	(2.60)	(4.67)
$p = 2$	−0.005	0.004	0.002	0.001	0.010	**0.030**
	(−1.09)	(0.86)	(0.45)	(0.06)	(2.25)	(4.73)
$p = 3$	−0.007	0.003	**0.003**	−0.002	0.005	0.029
	(−1.73)	(0.60)	(0.67)	(−0.51)	(0.99)	(4.58)
$h = 1$	−0.027	−0.000	**0.005**	−0.039	**0.023**	**0.004**
	(−7.10)	(−2.71)	(1.13)	(−10.29)	(5.71)	(0.73)
$h = 10$	**0.018**	**0.007**	−0.005	**0.013**	−0.002	−0.050
	(4.94)	(1.85)	(−1.03)	(3.39)	(−0.61)	(−9.14)
Min	0.989	0.964	0.924	0.962	0.990	1.025
$(\tau/K/p/h)$	(63/1/1/1)	(189/1/1/5)	(189/1/2/10)	(189/1/0/5)	(252/3/0/10)	(252/1/3/10)
Mean	1.092	1.065	1.034	1.055	1.057	1.204
$(t{-}stat)$	(21.64)	(17.81)	(10.36)	(13.21)	(14.44)	(35.57)
Max	1.202	1.164	1.120	1.154	1.167	1.288
$(\tau/K/p/h)$	(63/3/1/10)	(63/4/0/10)	(63/1/1/10)	(63/3/0/10)	(252/2/3/10)	(42/3/2/5)
ANOVA	1.194	1.101	1.050	1.127	1.098	1.261
$(\tau/K/p/h)$	(63/2/1/10)	(63/5/1/10)	(63/3/3/1)	(63/5/1/10)	(63/2/1/1)	(63/3/2/1)

Table 3.5: ANOVA results for hm–statistics for rate based signals for the period April 19, 2000 to April 11, 2008. BenchHM are the hm–statistics for $h = 5$ days ahead forecasts of the specification $\tau = 42, K = 1, p = 0$. Const represents the constant of the ANOVA regression. The t–statistic given in parentheses underneath is computed against an intercept of unity. The remaining ANOVA estimates are given with t–statistics for testing coefficient significance. Bold entries indicate model features providing the best forecasting results on average. The lower part shows minimum, mean and maximum hm–statistics for each trade. hm–statistics for the ANOVA implied specification (ANOVA) are given in the last line. For the Max, Min and ANOVA entries PCA/VAR characteristics are given in parentheses underneath. For average hm–statistics a t–statistic is given in parentheses for testing H_0 : hm = 1.

	F1	F2	F3	F4	F5	F6
BenchHM	1.092	1.057	1.018	1.052	1.016	0.965
Const	1.084 (17.13)	1.053 (10.40)	1.036 (7.97)	1.056 (10.71)	1.019 (3.58)	0.981 (−3.48)
$\tau = 63$	**0.035** (8.38)	**0.020** (4.58)	**0.018** (4.53)	**0.014** (3.09)	−0.004 (−0.80)	−0.012 (−2.29)
$\tau = 126$	0.004 (1.01)	−0.017 (−3.99)	−0.030 (−7.67)	−0.022 (−4.91)	−0.001 (−0.19)	0.017 (3.37)
$\tau = 189$	−0.019 (−4.67)	−0.031 (−7.23)	−0.036 (−9.24)	−0.021 (−4.83)	**0.042** (9.15)	**0.029** (5.72)
$\tau = 252$	0.001 (0.34)	−0.013 (−3.08)	0.008 (2.03)	0.003 (0.76)	0.003 (0.54)	0.028 (5.50)
$K = 2$	−0.000 (−0.04)	0.001 (0.18)	0.001 (0.30)	0.001 (0.16)	−	−
$K = 3$	0.005 (1.19)	0.013 (2.94)	**0.002** (0.53)	0.007 (1.63)	−0.001 (−2.32)	−
$K = 4$	**0.010** (2.52)	**0.020** (4.68)	0.000 (0.07)	**0.013** (3.04)	−0.011 (−2.66)	**0.009** (2.36)
$K = 5$	0.009 (2.16)	0.019 (4.37)	0.001 (0.20)	0.013 (2.90)	−0.014 (−3.33)	0.007 (1.84)
$p = 1$	**0.005** (1.40)	**0.013** (3.27)	−0.005 (−1.45)	**0.005** (1.26)	**0.005** (1.30)	−0.001 (−0.25)
$p = 2$	−0.005 (−1.33)	0.007 (1.67)	−0.011 (−3.06)	−0.004 (−1.04)	0.002 (0.39)	0.002 (0.33)
$p = 3$	0.001 (0.23)	0.011 (2.72)	−0.004 (−1.06)	−0.001 (−0.19)	−0.003 (−0.66)	**0.003** (0.68)
$h = 1$	−0.051 (−15.83)	−0.023 (−6.92)	−0.016 (−5.31)	−0.021 (−6.25)	−0.019 (−5.26)	0.001 (2.50)
$h = 10$	**0.018** (5.65)	**0.001** (2.85)	**0.014** (4.69)	**0.009** (2.73)	**0.024** (6.85)	**0.001** (2.51)
Min ($\tau/K/p/h$)	0.996 (63/2/2/1)	0.986 (42/1/3/1)	0.965 (189/2/3/10)	0.982 (42/1/3/1)	0.966 (252/2/0/5)	0.934 (63/4/1/10)
Mean ($t-stat$)	1.082 (20.12)	1.058 (17.39)	1.024 (7.43)	1.053 (17.27)	1.021 (5.66)	1.006 (1.74)
Max ($\tau/K/p/h$)	1.193 (63/2/3/10)	1.150 (63/1/0/10)	1.106 (63/3/1/10)	1.137 (63/1/0/10)	1.116 (189/2/3/10)	1.077 (252/4/3/10)
ANOVA ($\tau/K/p/h$)	1.183 (63/4/1/10)	1.147 (63/4/1/10)	1.105 (63/3/0/10)	1.126 (63/4/1/10)	1.108 (189/2/1/10)	1.033 (189/4/3/10)

Table 3.6: ANOVA results for hm–statistics for factor based signals for the period April 19, 2000 to April 11, 2008. For further notes see Table 3.5.

3.3.2 Economic forecasting performance

As a measure of economic forecasting performance of particular PCA/VAR implementations trading implied cash flows as described in Section 3.1.2 are cumulated over time to obtain total cash flow statistics. Tables 3.7 and 3.8 document ANOVA regressions of total cash flows on dummy variables representing specification parameters and forecast horizons for rate and factor based trading strategies, respectively.

ANOVA estimates indicate that choosing a small time window $\tau = 63$ appears preferable for most trading strategies (R1 to R4 and F1 to F4). For the remaining *SlopeTrade* (R5, F5) and *CurvatureTrade* (R6, F6) a significant improvement is diagnosed for conditioning on time windows comprising 189 trading days or more.

The impact of the number of factors on cash flows is similar as discussed for hm–statistics. Conditional on rate based trades, including more than one factor ($K > 1$) improves the outcome of R1 to R4 and R6, but deteriorates the performance of R5. Conditional on factor based signalling at least three factors are necessary to improve cash flows on average.

Concerning the VAR order ANOVA results are in favor of a model with one lag ($p = 1$) for all trades except R6, F3 and F6. For the *CurvatureTrade* (R6 and F6) autoregressive dynamics with more than one lag generally obtain positive excess cash flows. However, the improvement over the model with $p = 1$ is only marginal.

Furthermore, 10 trading strategies (all except R5 and F6) perform significantly better at a forecast horizon of $h = 10$. Assuming that the magnitude of rate changes increases with the forecast horizon, this result might be addressed to the fact that only the direction and not the size of a movement is important for the signalling.

Averaging over the full set of PCA/VAR models all trading strategies except R5 and F6 obtain significantly positive cash flows. Comparing minimum and maximum cash flows, however, PCA/VAR models have upside and downside potential. Regarding trade R3, for example, the specifications $\tau = 189, K = 1, p = 2$ and $\tau = 63, K = 3, p = 1$ yield minimum and maximum cash flows at the $h = 10$ forecast horizon of -106.08, respectively, 215.06 Euros.

As outlined ANOVA regressions are a valuable tool in figuring out particular model features that positively influence the forecasting performance of the PCA/VAR model. For instance, the ANOVA preferred model specification for the trading strategy R1 is given by a VAR model with one lag ($p = 1$) and $K = 2$ factors which are estimated conditional on $\tau = 63$ daily observations. In the following, we refer to such model compositions as the ANOVA preferred or ANOVA implied model specification. Overall, ANOVA preferred model specifications generate total cash flows that are close to the maximum total cash flow over all PCA/VAR implementations.

In two cases (F3, F5) the ANOVA implied model specification provides the maximum cash flow. For five (R1, R6, F1, F2, F4) out of 12 trading strategies the ANOVA implied forecasting models generate cash flows which are at least 80% of the maximum cash flow of respective unconditional models.

For 'ex–post' selection of PCA/VAR implementations one may be tempted just to choose the best performing model from the set of all PCA/VAR models. However, the outcome of such an approach neglects potential systematic influences of model features on forecasting performance. We regard an ANOVA as a suitable means to uncover model features that are essential for accurate predictions. The proximity of the ANOVA implied and the unconditionally best performing model documented in Tables 3.7 and 3.8 further motivates to employ ANOVA regressions for adaptive model selection as described in the following section.

	R1	R2	R3	R4	R5	R6
Const	28.80	40.77	13.50	43.45	−17.02	18.15
	(8.84)	(6.34)	(1.50)	(8.25)	(−3.86)	(16.61)
$\tau = 63$	**3.05**	**17.20**	**29.53**	**18.12**	13.11	−0.04
	(1.11)	(3.17)	(3.87)	(4.07)	(3.52)	(−0.04)
$\tau = 126$	−16.99	−6.61	11.63	−10.06	36.63	4.78
	(−6.17)	(−1.22)	(1.53)	(−2.26)	(9.84)	(5.18)
$\tau = 189$	−27.78	−53.34	−6.81	−55.27	30.87	**5.45**
	(−10.09)	(−9.82)	(−0.89)	(−12.42)	(8.29)	(5.90)
$\tau = 252$	−27.49	−51.43	11.29	−42.02	**38.56**	2.85
	(−9.99)	(−9.47)	(1.48)	(−9.44)	(10.36)	(3.09)
$K = 2$	**10.76**	4.40	31.74	−3.75	−6.67	**4.85**
	(3.91)	(0.81)	(4.16)	(−0.84)	(−1.79)	(5.26)
$K = 3$	5.40	10.59	**35.21**	3.50	−10.89	3.64
	(1.96)	(1.95)	(4.62)	(0.79)	(−2.92)	(3.94)
$K = 4$	5.35	12.31	28.29	6.95	−12.81	0.97
	(1.94)	(2.27)	(3.71)	(1.56)	(−3.44)	(1.05)
$K = 5$	7.04	**16.92**	33.52	**13.91**	−19.51	−3.27
	(2.56)	(3.11)	(4.40)	(3.13)	(−5.24)	(−3.55)
$p = 1$	**0.58**	**5.60**	**5.15**	**2.78**	**4.76**	3.99
	(0.24)	(1.15)	(0.76)	(0.70)	(1.43)	(4.83)
$p = 2$	−3.25	1.68	2.39	−3.41	2.24	4.06
	(−1.32)	(0.35)	(0.35)	(−0.86)	(0.67)	(4.92)
$p = 3$	−6.50	−0.73	2.98	−6.27	2.67	**4.09**
	(−2.64)	(−0.15)	(0.44)	(−1.57)	(0.80)	(4.95)
$h = 1$	−8.51	−14.09	−24.93	−23.48	**10.32**	−2.47
	(−3.99)	(−3.35)	(−4.22)	(−6.81)	(3.58)	(−3.46)
$h = 10$	**13.59**	**26.42**	**29.48**	**33.02**	−3.53	**1.80**
	(6.38)	(6.28)	(4.99)	(9.58)	(−1.23)	(2.51)
Min	−24.80	−53.56	−106.08	−41.37	−79.19	−6.17
$(\tau/K/p/h)$	(252/3/2/10)	(189/4/3/10)	(189/1/2/10)	(189/1/0/5)	(63/5/3/10)	(42/5/0/10)
Mean	20.07	36.53	52.52	31.18	1.52	24.81
$(t{-}stat)$	(15.65)	(14.22)	(18.26)	(12.44)	(0.99)	(64.76)
Max	95.33	180.49	244.49	166.01	56.95	39.79
$(\tau/K/p/h)$	(63/3/0/10)	(63/4/0/10)	(63/1/1/10)	(63/3/0/10)	(252/3/1/10)	(126/3/1/10)
ANOVA	84.04	103.71	126.73	118.86	3.25	34.23
$(\tau/K/p/h)$	(63/2/1/10)	(63/5/1/10)	(63/3/1/10)	(63/5/1/10)	(252/1/1/1)	(189/2/3/10)

Table 3.7: ANOVA results for cash flows from rate based signalling for the period April 19, 2000 to April 11, 2008. The upper part shows parameter estimates with t–statistics in parentheses underneath. The lower part gives minimum, mean and maximum cash flows for each trade obtained over all alternative forecasting models. Cash flows for the ANOVA implied specification (ANOVA) are given in the last line. For the Min, Max and ANOVA entries specifications are shown in parentheses below cash flows.

	F1	F2	F3	F4	F5	F6
Const	27.46	40.37	56.05	36.82	−6.07	−4.54
	(10.61)	(9.14)	(9.61)	(10.99)	(−2.28)	(−5.22)
$\tau = 63$	**10.68**	**21.28**	**33.81**	**13.99**	1.53	−4.04
	(4.88)	(5.70)	(6.86)	(4.94)	(0.66)	(−5.09)
$\tau = 126$	−9.30	−15.57	−16.90	−14.73	18.54	−2.33
	(−4.25)	(−4.17)	(−3.43)	(−5.20)	(7.95)	(−2.93)
$\tau = 189$	−27.53	−39.18	−47.37	−18.70	24.35	3.47
	(−12.58)	(−10.49)	(−9.61)	(−6.60)	(10.44)	(4.38)
$\tau = 252$	−17.53	−26.49	0.39	0.98	**26.47**	**6.06**
	(−8.01)	(−7.09)	(0.08)	(0.34)	(11.35)	(7.64)
$K = 2$	−1.33	−2.20	−2.77	−1.73	–	–
	(−0.61)	(−0.59)	(−0.56)	(−0.61)		
$K = 3$	**0.80**	5.06	**1.29**	2.38	**0.39**	–
	(0.37)	(1.35)	(0.26)	(0.84)	(0.19)	
$K = 4$	0.66	7.55	0.19	3.05	−1.22	**1.37**
	(0.30)	(2.02)	(0.04)	(1.08)	(−0.58)	(2.23)
$K = 5$	0.61	**7.74**	1.06	**3.55**	−3.26	1.06
	(0.28)	(2.07)	(0.21)	(1.25)	(−1.56)	(1.72)
$p = 1$	**1.24**	**5.48**	−3.41	**1.18**	**8.39**	1.56
	(0.63)	(1.64)	(−0.77)	(0.47)	(4.02)	(2.20)
$p = 2$	−2.55	−0.90	−12.40	−5.24	1.25	1.66
	(−1.30)	(−0.27)	(−2.81)	(−2.07)	(0.60)	(2.34)
$p = 3$	−2.66	−0.62	−11.17	−4.74	2.93	**1.83**
	(−1.36)	(−0.19)	(−2.53)	(−1.87)	(1.41)	(2.57)
$h = 1$	−12.76	−22.37	−43.09	−27.11	−6.00	**1.96**
	(−7.53)	(−7.73)	(−11.28)	(−12.36)	(−3.32)	(3.19)
$h = 10$	**19.46**	**32.87**	**62.40**	**36.01**	**4.28**	−1.16
	(11.48)	(11.37)	(16.34)	(16.42)	(2.37)	(−1.89)
Min	−17.93	−23.87	−27.69	−12.08	−30.32	−14.00
($\tau/K/p/h$)	(189/2/3/10)	(189/2/3/10)	(189/2/2/5)	(252/2/2/1)	(63/5/3/10)	(63/4/2/10)
Mean	20.12	36.50	49.68	35.35	9.65	−1.57
($t{-}stat$)	(15.70)	(17.02)	(14.98)	(18.91)	(8.93)	(−4.06)
Max	85.62	153.16	217.71	127.09	49.63	12.18
($\tau/K/p/h$)	(63/2/3/10)	(63/3/1/10)	(63/3/0/10)	(63/1/0/10)	(252/3/1/10)	(252/4/3/10)
ANOVA	85.19	139.04	217.71	108.93	49.63	5.51
($\tau/K/p/h$)	(63/3/1/10)	(63/5/1/10)	(63/3/0/10)	(63/5/1/10)	(252/3/1/10)	(252/4/3/1)

Table 3.8: ANOVA results for cash flows from factor based signalling for the period April 19, 2000 to April 11, 2008. For further notes see Table 3.7.

3.4 An adaptive modeling strategy

Implicitly, the investigation of the forecasting accuracy of particular PCA/VAR specifications in Section 3.3 proceeds from the view that the swap term structure is uniformly well approximated over the entire sample period. From the illustrations of potential structural variation of term structure dynamics in Section 3.2, however, it seems unlikely that a single model implementation performs homogeneously over rolling samples. There might be time periods for which dynamics are better captured by specific model features, such as, for example, parsimonious VAR orders or a low dimensional space of principal components. Moreover, structural variation could motivate the use of smaller time windows to extract the principal components. Conversely, using larger time windows is justified over long periods of structural homogeneity or slight structural variation. As a consequence, we propose a data driven adaptive model selection strategy to select the best forecasting model or at least a model which is likely to achieve positive cash flows in the near future.

To illustrate the adaptive procedure consider, say, the *10yrSingleTrade* and a forecasting horizon $h = 10$. At a specific trading day T^* an analyst experiences the performance of each PCA/VAR implementation that has been used in time $T^* - h$ to forecast the 10yr swap rate in T^*. Using a second window of $\tilde{\tau} = 42$ forecasts over the period $T^* - h - \tilde{\tau} + 1$ to $T^* - h$ the adaptive strategy evaluates in T^* each model specification in terms of 'local' total cash flows. To be more precise, consider h–step forecasts of all factor model specifications $i = 1, \ldots, 100$, for $T^* - h - \tilde{\tau} + 1, \ldots, T^* - h$ given by $\hat{g}^i_{T^*-h-\tilde{\tau}+1,h}(\theta), \ldots, \hat{g}^i_{T^*-h,h}(\theta)$. If an upward movement is predicted, i.e. $\hat{g}^i_{T^*-h-\tilde{\tau}+1,h}(\theta) > 0$, then set up the corresponding 'swap portfolio' as defined in Table 3.1. For example, if the 10yr swap rate is forecasted to rise/fall, then enter a 10yr receiver/payer swap agreement. Closing this position h days later generates a cash flow $CF^{h}_{i,T^*-h-\tilde{\tau}+1}$. For each factor model specification i compute

$$SumCF_{i,h} = \sum_{t=T^*-h-\tilde{\tau}+1}^{T^*-h} CF^{h}_{i,t} \ .$$

Then perform an ANOVA regression of $SumCF_{i,h}$ on dummy variables representing τ, K and p. Those model features $i^* = \{\tau^*, K^*, p^*\}$ showing strongest systematic influences on local forecast performance correspond to the largest estimated dummy variable coefficients. Finally, the h–step forecast of the adaptive strategy in T^* is given by

$$\hat{g}^{\text{ANOVA}}_{T^*,h}(\theta) = \hat{g}^{i^*}_{T^*,h}(\theta) \ .$$

Table 3.9 documents the statistical and economic performance of adaptive model selection for each trading strategy and forecast horizon. The first part contains the hm–statistic and some statistics describing the cash flow. In the second part, we compare the economic performance of the adaptive strategy with the best, worst and average performance obtained from 'unconditional' PCA/VAR models discussed in Section 3.3.

Regarding the total cash flow (CF) the adaptive strategy outperforms the best 'unconditional' model specification in 8 out of 36 cases (F3, F6 for $h = 1$, F6 for $h = 5$ and R1, R4, F2, F4, F6 for $h = 10$). In 7 further cases (R6, F5 for $h = 1$, R1, R5, R6 for $h = 5$ and R2, F1 for $h = 10$) the adaptive approach comes close (at least 80 %) to the performance of the best 'unconditional' specification. In only one case (F3 for $h = 5$) the adaptive procedure earns a negative total cash flow. In addition, regarding the minimum bank account (MinBank) over the 2083 trading days no adaptive trading strategy appears to have a great downside potential. In 27 out of 36 cases the minimum bank account is not (or only slightly) below -10 Euros. In 8 further cases MinBank is between -32.61 and -10 Euros. Only trade F3 at $h = 5$ generates cash flows leading to a minimum bank account of -52.38 Euros.

	R1			R2			R3		
	$h=1$	$h=5$	$h=10$	$h=1$	$h=5$	$h=10$	$h=1$	$h=5$	$h=10$
hm	1.113	1.070	1.159	1.109	1.067	1.107	1.049	1.050	1.042
CF	12.24	38.84	100.18	33.49	71.00	164.41	36.42	78.94	118.88
MinBank	-1.29	-3.26	-7.74	-1.13	-8.12	-20.42	-1.13	-14.15	-29.30
MinCF	−1.00	−5.98	−24.80	−11.58	−47.13	−53.56	−12.07	−90.05	−106.08
($\tau/K/p$)	(63/1/1)	(189/3/2)	(252/3/2)	(189/1/0)	(189/1/1)	(189/4/3)	(189/1/3)	(189/1/2)	(189/1/2)
AvCF	9.86	18.38	31.97	18.33	32.42	58.84	26.08	51.01	80.49
MaxCF	22.31	46.91	95.33	45.05	96.51	180.49	48.23	138.51	244.49
($\tau/K/p$)	(63/3/1)	(63/3/0)	(63/3/0)	(189/4/2)	(63/3/0)	(63/4/0)	(126/3/1)	(63/1/0)	(63/1/1)

	R4			R5			R6		
	$h=1$	$h=5$	$h=10$	$h=1$	$h=5$	$h=10$	$h=1$	$h=5$	$h=10$
hm	1.056	1.056	1.134	1.053	0.988	0.993	1.231	1.239	1.177
CF	11.59	39.34	191.53	5.61	26.41	13.88	23.99	25.54	28.44
MinBank	-2.02	-8.73	-19.21	-0.44	-5.26	-9.75	-0.19	-0.04	-0.55
MinCF	−13.44	−41.37	−39.48	−6.91	−50.05	−79.19	7.60	2.62	−6.17
($\tau/K/p$)	(252/3/1)	(189/1/0)	(189/2/3)	(63/5/0)	(42/3/0)	(63/5/3)	(42/5/0)	(42/5/0)	(42/5/0)
AvCF	4.51	28.00	61.02	9.58	-0.75	-4.28	22.56	25.03	26.83
MaxCF	23.05	87.73	166.01	19.81	30.67	56.95	27.54	31.91	39.79
($\tau/K/p$)	(126/4/3)	(63/4/0)	(63/3/0)	(63/1/0)	(189/3/3)	(252/3/1)	(126/4/3)	(189/3/3)	(126/3/1)

	F1			F2			F3		
	$h=1$	$h=5$	$h=10$	$h=1$	$h=5$	$h=10$	$h=1$	$h=5$	$h=10$
hm	1.099	1.059	1.144	1.077	1.026	1.136	1.039	0.974	1.060
CF	10.58	20.34	78.59	24.30	22.02	177.59	28.19	-43.32	94.24
MinBank	-1.75	-7.16	-11.61	-1.40	-10.42	-18.84	-8.47	-52.38	-24.80
MinCF	−2.55	−11.96	−17.93	−9.52	−20.62	−23.87	−23.42	−27.69	−10.88
($\tau/K/p$)	(252/2/2)	(189/2/2)	(189/2/3)	(252/2/2)	(189/2/2)	(189/2/3)	(189/1/2)	(189/2/2)	(189/2/2)
AvCF	5.12	17.88	37.34	10.63	33.00	65.87	0.16	43.24	105.64
MaxCF	15.29	38.43	85.62	34.66	74.87	153.16	17.49	108.27	217.71
($\tau/K/p$)	(42/3/1)	(63/3/1)	(63/2/3)	(42/3/1)	(42/5/2)	(63/3/1)	(42/3/1)	(63/1/0)	(63/1/0)

	F4			F5			F6		
	$h=1$	$h=5$	$h=10$	$h=1$	$h=5$	$h=10$	$h=1$	$h=5$	$h=10$
hm	1.056	1.052	1.142	1.086	1.058	1.057	1.211	1.240	1.174
CF	7.03	14.78	157.65	15.43	5.87	2.34	18.76	23.00	24.95
MinBank	-1.61	-11.89	-32.61	-0.20	-6.35	-23.31	-0.11	-0.15	-0.36
MinCF	−12.08	−5.30	11.85	−7.16	−24.32	−30.32	−6.39	−11.84	−14.00
($\tau/K/p$)	(252/2/2)	(189/2/2)	(189/2/2)	(42/5/3)	(42/4/2)	(63/5/3)	(63/3/3)	(63/5/2)	(63/4/2)
AvCF	5.27	32.38	68.39	4.23	10.23	14.51	0.13	-1.83	-3.00
MaxCF	20.96	61.43	127.09	15.48	43.02	49.63	6.69	7.84	12.18
($\tau/K/p$)	(42/3/1)	(63/1/0)	(63/1/0)	(126/4/1)	(252/4/3)	(252/3/1)	(252/4/1)	(252/5/2)	(252/4/3)

Table 3.9: Results for the adaptive model selection strategy for the period April 19, 2000 to April 11, 2008. The first part contains the hm–statistic, the total cash flow (CF) and the minimum bank account (MinBank) over the 2083 trading days. The second part contains the minimium (MinCF), average (AvCF) and maximum (MaxCF) total cash flows over all the unconditional PCA/VAR models.

Finally and for completeness, we compare the adaptive strategy with the forecasting content of a 3–dimensional VAR(1) benchmark model of first differences of the 2yr, 5yr and 10yr swap rate, $\Delta y_t = (\Delta y_{2yr,t}, \Delta y_{5yr,t}, \Delta y_{10yr,t})'$. To account for the fact that the adaptive strategies switch between models with minimum, respectively, maximum window sizes of $\tau = 42$ and $\tau = 252$ days, the VAR(1) model is estimated for sample sizes of 42 and 252 days. From the level predictions of $\hat{y}_{T^*+h}$ forecasts of the level, slope and curvature measures are computed. To pick a particular example, the left hand side panel of Table 3.10 documents the performance of the VAR benchmark and the adaptive strategies for the *LevelTrade* (R4 and F4) and $h = 10$. It can be seen that for both rate and factor based signalling VAR forecasts are inferior in terms of statistical and economic performance. The right hand side panel of Table 3.10 shows the number of times a particular forecasting model outperforms the remaining 3 approaches in terms of statistical and economic performance over all 6 trades and 3 forecast horizons. While in terms of the hm–statistic the adaptive approach performs slightly better than the benchmark, the former has a clear advantage over the latter in terms of implied cash flows. In 17 out of 18 cases cash flows from the adaptive strategies are larger than those of the benchmark models.

	LevelTrade for $h = 10$				all trades and horizons			
	R4	F4	VAR42	VAR252	R	F	VAR42	VAR252
hm	1.134	1.142	1.114	1.066	5	6	3	4
CF	191.53	157.65	91.07	46.84	15	2	1	0

Table 3.10: The left hand side panel documents the forecasting performance (hm–statistics and total cash flows (CF)) of the VAR benchmark and the adaptive strategies (R4 and F4) for the *LevelTrade* and $h = 10$. The right hand side panel shows the number of times a particular strategy outperforms the remaining 3 strategies in terms of statistical and economic performance over all 6 trading strategies and 3 forecast horizons.

3.5 A note on the model selection risk of the adaptive strategy

Quantitative models are formulated to capture regularities governing economic phenomena. As economic relationships are generally complex, models that tempt to describe all aspects of the reality are complicated, too. In order to focus on the essential and to guarantee computational tractability models are formulated which are approximations to real world phenomena. In practice, an analyst frequently encounters a situation in which several candidate models provide reasonable approximations to reality. When selecting a particular model based on a distinct selection criterion, one might be tempted to include as many models as possible into the class of model candidates. Yet, such an approach entails the problem of selection bias. The chance that a model is best in terms of some selection criterion by pure luck increases with the cardinality of the candidate model class (Zucchini, 2000). Moreover, with a growing number of models considered the risk of choosing an excessively unfavorable candidate increases.

The proposed ANOVA based adaptive strategy does not choose a single model for the entire period. Instead, it adaptively selects a model in each time instant based on a particular criterion measuring historic ex–ante forecasting performance over a local time window. It is shown that if a forecasting framework is described by a large parameterized model class, the ANOVA based data driven procedure is a suitable tool for model selection which is neither exposed to selection bias nor to the risk of choosing excessively poor models. For illustration purposes, model performance is measured by trading implied cash flows. Moreover, we consider the strategies R1, R2 and R3 based on daily ex–ante forecasts for the 2yr, 5yr and 10yr swap rate. In the sequel, we compare the economic forecast performance of $h = 1, 5, 10$ days ahead ex–ante forecasts of the adaptive approach and the set of unconditional models.

From the top panel of Table 3.11 it can be seen that all adaptive strategies generate positive total cash flows over the 2083 trading days. In particular, total cash flows increase with the forecast horizon as swap rate movements are likely to be larger over multiple days. The performance of the adaptive model selection strategies compares favorably to the performance of the unconditional modeling approach. In four cases out of nine (R1 and R2 for $h = 5$ and $h = 10$ resp.) the adaptive strategy outperforms at least 90 out of 100 unconditional models. Moreover, for all forecast exercises the adaptive approach is better than at least 70 unconditional specifications. Furthermore, the total cash flow of the adaptive strategy exceeds the average cash flow over all unconditional PCA/VAR models for all trades considered. The class of unconditional PCA/VAR models considered contains specifications having considerable downside risk. The model specification with the minimum total cash flow (MinCF) always generates negative cash flows. Yet, the adaptive selection procedure seems to filter out unfavorable models as indicated by the corresponding total cash flows.

The above ranking of total cash flows is only a snapshot characterizing the last trading day. Figure 3.4 shows the ranking of the adaptive strategy with respect to unconditional models in terms of cumulated cash flows over time. The data driven ANOVA selection procedure performs strongly in 6 out of 9 (R2 and R3 for $h = 1$, R1 and R2 for $h = 5$ and $h = 10$ resp.) forecast exercises. In these cases the adaptive strategy cumulates higher cash flows than at least 50 unconditional models. Particularly convincing is the performance of the adaptive procedure for the R1 and R2 trades for horizons $h = 5, 10$. It performs better than at least 80 unconditional models. For the three remaining forecast exercises (R1 for $h = 1$, R3 for $h = 5, 10$) it can be seen that in terms of cumulated cash flows the ranking of the ANOVA selection procedure is more or less fluctuating around the median of undonditional models. The last line of Table 3.11 shows that the adaptive strategy's time average of the ranking is larger than the time average of the ranking of 50, 69 and 65 unconditional models resp.

	R1			R2			R3		
	$h=1$	$h=5$	$h=10$	$h=1$	$h=5$	$h=10$	$h=1$	$h=5$	$h=10$
CF	12.24	38.84	100.18	33.49	71.00	164.41	36.42	78.94	118.88
MinBank	-1.29	-3.26	-7.74	-1.13	-8.12	-20.42	-1.13	-14.15	-29.30
MinCF	-1.00	-5.98	-24.80	-11.58	-47.13	-53.56	-12.07	-90.05	-106.08
AvCF	9.86	18.38	31.97	18.33	32.42	58.84	26.08	51.01	80.49
RankCF	70	90	100	88	92	98	80	79	78
MaxCF	22.31	46.91	95.33	45.05	96.51	180.49	48.23	138.51	244.49
MinMinBank	-5.35	-21.72	-46.44	-12.41	-55.28	-71.45	-13.69	-105.53	-144.88
RankMinBank	51	77	74	48	68	74	73	54	53
MaxMinBank	0.03	-0.85	-3.16	0.20	-1.36	-10.46	0.15	-0.29	-6.37
AvRankBank	50	87	94	83	91	100	88	69	65

Table 3.11: Results for the adaptive strategy and some summary measures for the set of unconditional models for each forecast exercise over the period April 19, 2000 to April 11, 2008. CF and MinBank denote the total cash flow and the minimum bank account of the adaptive strategy. MinCF, AvCF and MaxCF give the minimum, average and maximum total cash flow, respectively, over all the unconditional PCA/VAR models. RankCF is the rank of the adaptive strategy within the set of unconditional models when ordered according to total cash flows. MinMinBank and MaxMinBank denote the minimum resp. maximum value of the minimum bank account over the 2083 trading days over all unconditional models. RankMinBank resp. AvRankBank are the ranks of the adaptive strategy within the set of unconditional models when ordered according to the minimum bank account value resp. when ordered according to the time average of ranks (see also Figure 3.4).

Table 3.11 also provides an insight into the evolution of the bank account for the adaptive and unconditional strategies over the 2083 trading days. The minimum bank account of the adaptive model selection procedure is negative for all forecast exercises considered as can be seen from the first panel of Table 3.11. However the adaptive approach is always better than at least 50 of the 100 unconditional models as shown in the third panel of Table 3.11 (RankMinBank).

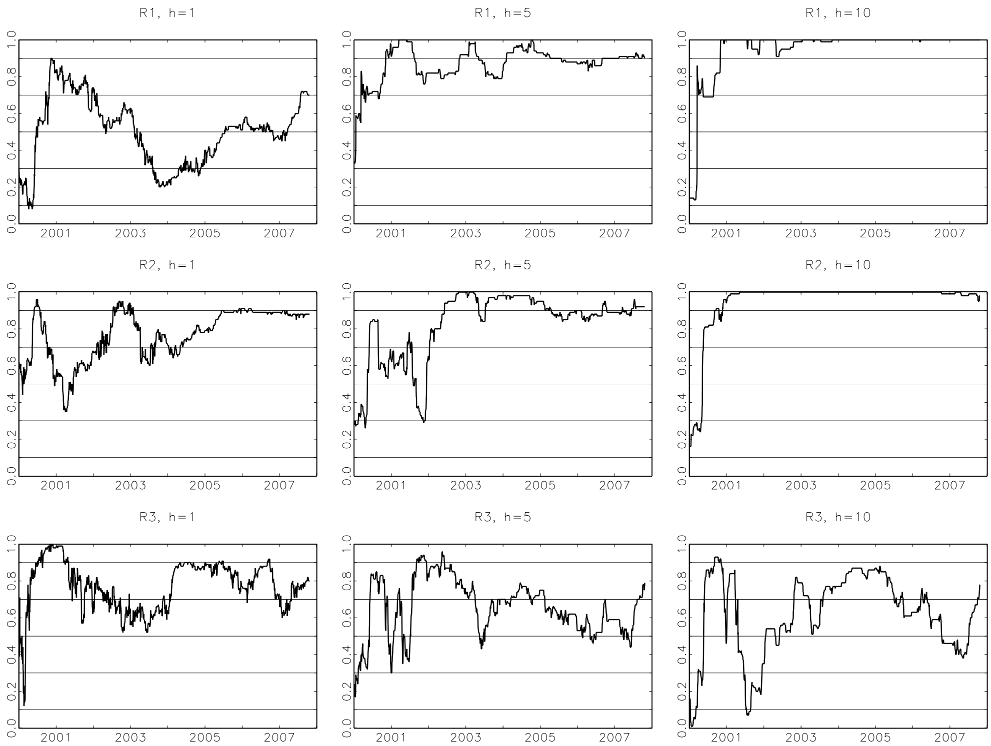

Figure 3.4: Rankings of the adaptive strategy with respect to the set of 100 unconditional models over the period April 19, 2000 to April 11, 2008. The value of the bank account at time T^* is used as the ranking criterion. Horizontal lines represent the 10th, 30th, ..., 90th quantile of unconditional models in time T^* according to the bank account value.

3.6 Conclusions

In this Chapter we focus on adaptive ex–ante forecasting of the EURIBOR swap term structure. By means of descriptive initial analyses we motivate that swap term structure dynamics are likely time varying. Evaluating the forecasting performance over a battery of 100 PCA/VAR implementations we fail to identify a uniformly dominating single specification in terms of both, statistical or economic performance. This finding may be seen to be at odds with global dynamic homogeneity of the swap term structure. Building upon the presumption of locally homogeneous dynamics a data driven adaptive strategy is motivated to ex–ante determine a particular factor model implementation which is likely to provide accurate trading signals. Evaluating the adaptive strategy in terms of economic performance it mostly outperforms static designs of trading the swap term structure. Moreover, opposite to the plentitude of implemented static designs the adaptive procedure shows only small downside risk but promises to achieve cash flows in excess of unconditional trading. In particular, the ANOVA based adaptive model selection is successful in reducing model selection risk when risk is measured in terms of trading implied cash flows. Furthermore, the adaptive strategy withstands a comparison with a VAR(1) benchmark model when considering trading implied cash flows. The proposed procedure is easy to implement and widely applicable. As a conclusion data–driven adaptive model selection based on local ex–ante economic forecasting performance merits consideration in order to account for evolving market conditions and to avoid model selection bias.

Chapter 4

Empirical Analysis: Part 2

In this Chapter we provide further support for structural variability of local swap term structures. Moreover, given the promising performance of the ANOVA based data adaptive strategy, we investigate various additional adaptive forecasting schemes. To evaluate ex-ante forecasting performance for particular rates, distinct forecast features, such as mean squared errors, directional accuracy and directional forecast value, are considered. It turns out that, relative to benchmark models, the adaptive approach offers additional forecast accuracy in terms of directional accuracy and directional forecast value.

Note that the data set we analyze in this Chapter comprises only 2100 daily observations from February 15, 1999, to March 2, 2007. While in Chapter 3 we motivated to model factor dynamics by means of a VAR(p) approach, in this Chapter we employ uncorrelated AR(p) models for each extracted factor. This might be seen as a robustness analysis of our model assumptions as we find that the conclusions with respect to the suitability of adaptive model selection procedures are not altered.

The remainder of the Chapter is organized as follows. In Section 4.1 we introduce the loss measures used to evaluate forecasting performance. Section 4.2 characterizes the unconditional approach to motivate adaptive model selection procedures. In Section 4.3 particular adaptive strategies are proposed, and we compare these adaptive strategies with unconditionally implemented factor models and some benchmark specifications. Section 4.4 concludes.

4.1 Loss functions

To introduce some notation, let a general loss function depend on the h–step ahead swap rate forecast, $\hat{y}_{m,T^*+h|T^*}$, the current swap rate, y_{m,T^*}, and the future (true, realized) swap rate, y_{m,T^*+h}, with maturity m, i.e.

$$L_{T^*}^{h,m} = L(\hat{y}_{m,T^*+h|T^*}, y_{m,T^*}, y_{m,T^*+h}) \, .$$

A common loss function is the quadratic loss

$$L_{1,T^*}^{h,m} \quad = \quad (y_{m,T^*+h} - \hat{y}_{m,T^*+h|T^*})^2 \, .$$

Diebold and Mariano (1995) point out, that in light of the variety of forecast based economic decision problems, statistical loss functions such as quadratic loss do not necessarily conform to economic loss functions. In an interest rate setting, Swanson and White (1995) show that the mean squared forecast error (MSFE) and profit measures are not closely linked. Similarly, Leitch and Tanner (1991) find that, in contrast to MSFE, the directional accuracy (DA) of forecasts, i.e., the ability to correctly predict the direction of change, is highly correlated with profits in a term structure analysis. As argued in Lai (1990), an investor can still gain profits even with statistically biased forecasts if they are characterized by significant DA. Ash, Smith and Heravi (1998) indicate that qualitative statements on the change of the economy in the near future are important prerequisites for an appropriate implementation of monetary and fiscal policy. Similarly, Öller and Barot (2000) emphasize the importance of DA for central banks, as a forecast of increased inflation (above target) would prompt central banks to raise interest rates. With $I(\bullet)$ denoting an indicator function, a loss function for DA is

$$\begin{aligned} L_{2,T^*}^{h,m} &= I\left((\hat{y}_{m,T^*+h|T^*} - y_{m,T^*})(y_{m,T^*+h|T^*} - y_{m,T^*}) > 0 \right) \\ &\quad -I\left((\hat{y}_{m,T^*+h|T^*} - y_{m,T^*})(y_{m,T^*+h|T^*} - y_{m,T^*}) < 0 \right) . \end{aligned}$$

Note that a model performs better the larger $L_{2,T^*}^{h,m}$. Hence, it is rather $-L_{2,T^*}^{h,m}$ which is a loss measure. Given this one–to–one correspondence to a loss function we use the terms 'loss function' or 'loss measure' for $L_{2,T^*}^{h,m}$.

Hatzmark (1991) investigates forecast ability by looking at DA and 'Big Hit Ability', with the latter taking into account the idea that a profit seeking trader might be better able to predict large price changes rather than small changes. In this case, forecast performance could depend on a small number of correct directional forecasts generating large profits and a large number of incorrect directional forecasts associated with negligibly small losses. Thus, the economic value of directional forecasts has to be distinguished from the directional accuracy $L_{2,T^*}^{h,m}$. A loss function for directional forecast value (DV) is

$$L_{3,T^*}^{h,m} = L_{2,T^*}^{h,m} |y_{m,T^*+h} - y_{m,T^*}| \, .$$

Analogous to previous arguments for $L_{2,T^*}^{h,m}$, $L_{3,T^*}^{h,m}$ is also termed as 'loss function'. The DV measure generalizes the DA statistic in that it takes the sign and the magnitude of the movement into account. It is noteworthy that $L_{3,T^*}^{h,m}$ is only approximately a profit function if y_{m,T^*} is a swap rate. The profit/loss from closing a swap position in $T^* + h$ is given by the swap value in $T^* + h$ since in T^* a swap with a fixed rate y_{m,T^*} has a value of zero. However, since a swap is a financial derivative, in $T^* + h$ the value of a swap with rate y_{m,T^*} is a non linear function of y_{m,T^*+h} (Miron and Swannell, 1991). Yet, as the second derivative of the swap value with respect to y_{m,T^*+h}, is often very small, upward/downward movements in y_{m,T^*+h} are almost proportional to changes in the profit/loss from closing the corresponding swap position.

4.2 Unconditional forecast models

We consider 4 forecast horizons ($h = 1, 5, 10, 15$ days) and focus on h–step forecasts of 2yr, 5yr and 10yr swap rates. Hence, overall there are 12 distinct forecast 'exercises' $FE_j = \{m^j, h^j\},\ j = 1, \ldots, 12$, where FE_j is a tuple from the cartesian set defined by $\{2, 4, 8\} \times \{1, 5, 10, 15\}$. To define the adaptive strategies, denote a particular model specification as $M_s = \{\tau^s, K^s, p^s\}$, where $\tau^s \in \Omega_\tau = \{42, 63, 126, 189, 252\}$, $K^s \in \Omega_K = \{1, 2, 3, 4, 5\}$ and $p^s \in \Omega_p = \{0, 1, 2, 3\}$. M_s is a three dimensional tuple from the cartesian set $\Omega_\tau \times \Omega_K \times \Omega_p$, the cardinality of which is 100. A forecast for a specification s at time T^* is $\hat{y}^s_{m,T^*+h|T^*}$. For a particular loss function $L_i^{h,m,s}, i = 1, 2, 3$, and model specification M_s, the average out–of–sample forecast performance over the time interval $[T_1^*; T_2^*]$ is

$$\frac{1}{T_2^* - T_1^* + 1} \sum_{T^*=T_1^*}^{T_2^*} L_{i,T^*}^{h,m,s} = \frac{1}{T_2^* - T_1^* + 1} \sum_{T^*=T_1^*}^{T_2^*} L_i(\hat{y}^s_{m,T^*+h|T^*}, y_{m,T^*}, y_{m,T^*+h}) \ .$$

We refer to the average loss associated with $L_i^{h,m,s}, i = 1, 2, 3$, respectively, as $\text{MSFE}_s^{h,m}$, mean directional accuracy as $\text{MDA}_s^{h,m}$, and mean directional forecast value as $\text{MDV}_s^{h,m}$.

To motivate an adaptive model selection approach, we first consider the 'unconditional' forecast performance, i.e., the average forecast performance of $M_s,\ s = 1, \ldots, 100$, for the period $T_1^* = 308$ (April 19, 2000) to $T_2^* = 2085$ (February 9, 2007). Table 4.1 shows the MSFEs obtained when forecasting the 2yr swap rate one day ahead ($h = 1$). The worst model features an MSFE measure that is by a factor of 8 larger than the corresponding quantity for the best model. The MSFE statistics documented for the 90th and 10th best model differ by a factor of 3. For MDA and MDV the overall picture is similar. Hence, choosing the wrong model may provide poor forecasts. Moreover, for MDA and MDV the latter conclusion holds throughout for all forecast exercises FE_j. For the MSFE criterion, however, the 'spread' between the best and worst models diminishes for forecast horizons $h > 5$.

When investigating relative model accuracy over all forecast exercises in terms of MSFE, a model specification with $\tau = 252$ and $K = 5$ reveals a robust performance for all lags considered. More precisely, the models $\tau = 252, K = 5$ and $p = 1, 2$ are among the ten best models in 11 out of 12 forecast exercises, while the specifications with AR orders $p = 3$ and $p = 0$ are among the 10 most preferable models in 10, respectively 9, forecast exercises. The remaining specifications reveal a high degree of variation in relative forecast accuracy. In terms of MDA and MDV no model performs systematically well over all forecast exercises. The most successful specification features $\tau = 63, K = 3, p = 1$, which is in 8 (MDA) and 7 (MDV) out of 12 forecast exercises among the 10 best models.

In addition to marked differences in relative model performance, the forecast accuracy of a particular factor model might vary over time. If there is structural variation, each model specification might be seen as an approximation to the true data generating process and the approximation accuracy depends on 'local' term structures. To describe time varying model performance we document transition probability matrices as in Camba–Mendez, Kapetanios and Weale (2002). Each of the 100 models is mapped to performance quartiles conditional on the first and second half of the forecasting period. The transition probabilities are obtained from counting the models that move from one quartile in the first to a particular quartile in the second subsample. While a diagonal transition matrix indicates performance stability, large off–diagonal entries hint at time varying performance.

	2yr				5yr				10yr			
	$h=1$	$h=5$	$h=10$	$h=15$	$h=1$	$h=5$	$h=10$	$h=15$	$h=1$	$h=5$	$h=10$	$h=15$
	MSFE$*10^6$											
1st	0.200 (126/5/0)	0.962 (252/5/0)	1.954 (252/3/1)	3.025 (252/3/1)	0.196 (252/5/0)	0.938 (252/5/0)	1.852 (252/5/0)	2.861 (252/5/3)	0.153 (126/5/0)	0.697 (252/4/3)	1.384 (252/5/3)	2.095 (252/2/3)
10th	0.203 (42/4/0)	0.971 (252/3/3)	1.969 (252/5/2)	3.040 (252/4/2)	0.201 (63/4/0)	0.953 (189/5/2)	1.882 (189/5/2)	2.903 (252/4/2)	0.155 (63/3/0)	0.699 (252/5/2)	1.390 (252/4/1)	2.142 (252/5/1)
50th	0.213 (126/3/2)	1.043 (42/3/0)	2.154 (126/2/2)	3.357 (63/3/1)	0.212 (42/2/0)	1.014 (63/3/1)	2.070 (63/5/0)	3.287 (63/3/0)	0.160 (63/4/2)	0.740 (63/4/0)	1.543 (63/5/0)	2.466 (63/2/0)
90th	0.654 (126/1/2)	1.556 (126/1/2)	2.774 (126/1/0)	4.128 (126/1/2)	0.655 (126/1/2)	1.550 (126/1/0)	2.740 (126/1/0)	4.064 (126/1/2)	0.578 (126/1/2)	1.170 (126/1/0)	1.976 (126/1/0)	3.019 (42/1/1)
100th	1.609 (252/1/0)	2.608 (252/1/0)	3.897 (252/1/0)	5.266 (252/1/0)	1.806 (252/1/0)	2.729 (252/1/0)	3.895 (252/1/0)	5.128 (252/1/0)	1.696 (252/1/0)	2.302 (252/1/3)	3.074 (252/1/0)	3.892 (252/1/0)
	MDA$*10$											
1st	1.130 (189/2/0)	1.828 (63/3/0)	2.222 (63/3/1)	2.469 (63/3/0)	0.793 (42/5/1)	1.243 (63/3/0)	1.710 (63/4/0)	1.789 (63/4/1)	0.973 (252/3/3)	0.827 (252/3/3)	1.035 (63/1/1)	1.389 (63/1/2)
10th	0.996 (189/2/2)	1.614 (63/2/2)	1.963 (63/3/2)	2.188 (63/2/0)	0.579 (63/2/1)	1.007 (63/3/2)	1.406 (63/2/1)	1.665 (63/5/0)	0.771 (252/4/0)	0.658 (189/3/3)	0.832 (189/2/1)	1.063 (42/1/2)
50th	0.726 (42/2/2)	0.973 (126/3/3)	1.187 (126/2/2)	1.254 (252/5/0)	0.343 (42/2/0)	0.771 (126/3/2)	1.024 (126/2/3)	0.967 (126/5/3)	0.523 (126/4/1)	0.377 (126/2/1)	0.551 (189/5/2)	0.771 (63/3/0)
90th	0.253 (252/1/1)	0.287 (189/1/2)	0.546 (189/4/3)	0.501 (189/3/2)	0.017 (252/3/3)	0.264 (189/3/3)	0.214 (189/4/2)	0.304 (252/2/1)	0.118 (189/2/2)	0.017 (63/2/2)	−0.157 (126/3/2)	−0.377 (126/5/0)
100th	−0.219 (63/1/1)	−0.039 (126/1/0)	0.231 (189/1/3)	0.197 (189/1/3)	−0.433 (189/1/2)	−0.399 (189/1/1)	−0.461 (189/1/3)	−0.067 (189/1/3)	−0.051 (42/2/3)	−0.579 (189/1/3)	−0.934 (189/1/2)	−1.063 (189/1/1)
	MDV$*10^4$											
1st	0.428 (252/5/1)	1.482 (63/2/0)	2.921 (63/3/1)	4.210 (63/3/0)	0.335 (252/5/3)	0.907 (63/3/0)	1.796 (63/4/0)	2.432 (63/3/2)	0.320 (252/3/3)	0.597 (63/1/0)	1.072 (63/1/1)	1.801 (63/1/0)
10th	0.401 (126/2/3)	1.306 (63/5/1)	2.622 (63/2/2)	3.923 (63/2/2)	0.272 (189/4/3)	0.721 (126/2/0)	1.420 (63/5/1)	2.124 (63/2/0)	0.249 (252/4/2)	0.409 (42/3/2)	0.771 (252/3/1)	0.989 (63/5/2)
50th	0.302 (126/3/3)	0.678 (189/5/2)	1.322 (126/3/3)	1.873 (126/2/3)	0.183 (126/3/2)	0.438 (63/1/0)	0.828 (42/4/0)	1.043 (42/1/1)	0.183 (63/5/3)	0.227 (126/1/0)	0.343 (189/4/0)	0.625 (42/5/1)
90th	0.049 (42/1/2)	0.239 (63/1/0)	0.405 (252/4/2)	0.420 (189/5/2)	0.026 (252/1/2)	0.008 (126/1/3)	0.102 (252/4/0)	−0.072 (189/4/0)	0.061 (42/2/3)	−0.066 (42/2/1)	−0.129 (252/1/3)	−0.160 (126/5/2)
100th	−0.108 (63/1/1)	−0.202 (126/1/0)	0.148 (189/1/3)	0.233 (189/1/3)	−0.166 (189/1/2)	−0.562 (189/1/1)	−0.611 (189/1/3)	−0.674 (189/1/2)	−0.048 (189/1/3)	−0.637 (189/1/2)	−1.161 (189/1/2)	−1.590 (189/1/1)

Table 4.1: Quantiles for MSFE$*10^6$, MDA$*10$ and MDV$*10^4$ out-of-sample forecast performance of $h = 1, 5, 10, 15$ day–ahead forecasts of 2yr, 5yr, 10yr swap rates for the forecast period April 4, 2000 to February 9, 2007 and 100 models $\{M_s\}_{s=1}^{100} = \{\tau^s, K^s, p^s\}_{s=1}^{100}$. Specifications are indicated in parentheses.

Table 4.2 shows the MSFE, MDA and MDV based transition probabilities for forecasting the 2yr rate ($h = 1$). MSFE based transitions are characterized by clustering within the two upper and lower quartiles, such that transitions crossing the subsample medians are relatively rare. With respect to the MDA (MDV) statistics, off–diagonal transition frequencies are between 0.08 and 0.42 (0.04 and 0.48). Again, the results are similar over all horizons $h = 1, 5, 10, 15$ and swap rates 2yr, 5yr, 10yr. Corresponding transition frequencies are not provided for space considerations but are available from the authors upon request. In summary, we diagnose marked heterogeneity of model specifications in terms of MDA and MDV performance. With respect to MSFE, model choice appears less crucial.

q	MSFE 1	MSFE 2	MSFE 3	MSFE 4	MDA 1	MDA 2	MDA 3	MDA 4	MDV 1	MDV 2	MDV 3	MDV 4
1	.600	.400	.000	.000	.308	.231	.231	.231	.600	.080	.160	.160
2	.400	.520	.080	.000	.083	.333	.417	.167	.280	.080	.280	.360
3	.000	.080	.800	.120	.464	.179	.143	.214	.040	.360	.280	.320
4	.000	.000	.120	.880	.091	.273	.318	.318	.080	.480	.280	.160

Table 4.2: Transition frequency matrices for one day–ahead forecasts of the 2yr swap rate. The forecast period is divided in two parts both comprising 889 forecasts. The first row contains the relative transition frequencies from the 1st quartile in the first sample half to the 1st, 2nd, 3rd and 4th quartile in the second sample half, etc. For further notes see Table 4.1.

In the light of time dependent forecast accuracy it is desirable to have a strategy at hand that ex–ante identifies locally preferable model specifications. In the next Section we describe and evaluate data driven model selection strategies.

4.3 Adaptive strategies

4.3.1 Data driven model selection

An unconditional modelling approach is inherently subjected to misspecification under changing relations between economic variables. The rolling window strategy allows the parameters of a model to evolve over time. Yet, if parameter values are exposed to variation one may conjecture that the quality of a model approximation is time specific as well. An adaptive selection/estimation strategy is a promising means to account for distinct relative forecasting performance. The adaptive model selection approach is based on a further time window of $\widetilde{\tau} = 42$ days in which the 'local' out–of–sample performance of specifications $M_s,\ s = 1, \ldots, 100$, is evaluated.

At each time point T^* the most recent $\widetilde{\tau}$ h–step forecast errors for swap rate m and model specification M_s are known. A local MSFE measure is

$$\text{MSFE}_{T^*}^{h,m,s} = \sum_{t=T^*-h-\widetilde{\tau}+1}^{T^*-h} L_{1,t}^{h,m,s}/\widetilde{\tau}.$$

The adaptive strategy, denoted MinMSFE, chooses the local MSFE minimizing specification

$$\hat{y}_{m,T^*+h|T^*}^{\text{MinMSFE}} = \hat{y}_{m,T^*+h|T^*}^{s^*}, \quad s^* = \underset{s=1,\ldots,100}{\text{argmin}}\{\text{MSFE}_{T^*}^{h,m,s}\}\ .$$

Selecting the best performing model out of the class of PCA/AR models might ignore potential systematic influences of model features $\{\tau, K, p\}$ on forecasting performance. A suitable tool to account for the latter is an ANOVA regression of the local MSFEs of factor models M_s on dummy variables representing τ, K and p. The AnoMSFE forecast is given by

$$\hat{y}_{m,T^*+h|T^*}^{\text{AnoMSFE}} = \hat{y}_{m,T^*+h|T^*}^{s^*}\ ,$$

where $s^* = \{\tau^*, K^*, p^*\}$ collects model features showing the smallest estimated (dummy variable) coefficients for τ, K and p.

Among others, Diebold and Pauly (1987) argue that in the presence of structural shifts composite forecasts can improve forecast precision. Numerous combining procedures have been proposed in the literature. We focus on both an equal weight scheme and a combination procedure that assigns distinct weights to individual forecasts. The Av10MSFE forecast is

$$\hat{y}^{\text{Av10MSFE}}_{m,T^*+h|T^*} = \frac{1}{10}\left(\hat{y}^{s_1^*}_{m,T^*+h|T^*} + \ldots + \hat{y}^{s_{10}^*}_{m,T^*+h|T^*}\right),$$

where $M_{s_1^*}, \ldots, M_{s_{10}^*}$ refer to the 10 best models in terms of local MSFEs. Conditional on $M_{s_1^*}, \ldots, M_{s_{10}^*}$ the BunnMSFE forecast is given by

$$\hat{y}^{\text{BunnMSFE}}_{m,T^*+h|T^*} = \hat{\theta}_{s_1^*}\hat{y}^{s_1^*}_{m,T^*+h|T^*} + \ldots + \hat{\theta}_{s_{10}^*}\hat{y}^{s_{10}^*}_{m,T^*+h|T^*},$$

where the weights $\hat{\theta}_{s_q^*},\ q = 1, ..., 10$, are proportional to the number of times (out of $\widetilde{\tau}$ forecast realizations) that model s_q^* outperforms all other 9 models in terms of smaller squared errors (Bunn, 1975).

Along similar lines as described for the MSFE criterion, adaptive forecasting is also based on the loss functions MDA and MDV. Finally, we employ two combining strategies that have found support in the empirical literature (Clemen, 1989). The average (Av) and median strategy (Med) take the average and median forecast of the 100 forecast models irrespective of past performance, i.e.

$$\hat{y}^{\text{Av}}_{m,T^*+h|T^*} = \frac{1}{100}\sum_{s=1}^{100}\hat{y}^{s}_{m,T^*+h|T^*} \quad \text{and} \quad \hat{y}^{\text{Med}}_{m,T^*+h|T^*} = \underset{s=1,\ldots,100}{\text{Median}}\left\{\hat{y}^{s}_{m,T^*+h|T^*}\right\}.$$

In summary, the set of adaptive strategies is

$$\begin{aligned}\Omega_{AS} \;=\; &\{\text{MinMSFE, Av10MSFE, AnoMSFE, BunnMSFE, MaxMDA,}\\ &\text{Av10MDA, AnoMDA, BunnMDA, MaxMDV, Av10MDV,}\\ &\text{AnoMDV, BunnMDV, Av, Med}\}.\end{aligned}$$

All forecast comparisons are performed over the same sample period comprising 1778 time instances. Accounting for the largest estimation window ($\tau = 252$), the highest forecast horizon ($h = 15$) and the model evaluation window ($\widetilde{\tau} = 42$), the rolling forecasting analysis starts in time point $T_1^* = 252+15+42-1 = 308$. Average losses of a particular adaptive strategy AS and forecast exercise FE_j are denoted by $\text{MSFE}_{AS}^{h,m}$, $\text{MDA}_{AS}^{h,m}$ or $\text{MDV}_{AS}^{h,m}$. To compare the performance of the adaptive strategies we normalize the average losses of adaptive strategies with respect to minimum and maximum average losses implied by unconditional models:

$$
\begin{aligned}
\text{nMSFE}_{AS}^{h,m} &= 1 - \frac{\text{MSFE}_{AS}^{h,m} - \min\limits_{s}\left\{\text{MSFE}_{s}^{h,m}\right\}}{\max\limits_{s}\left\{\text{MSFE}_{s}^{h,m}\right\} - \min\limits_{s}\left\{\text{MSFE}_{s}^{h,m}\right\}}, \\
\text{nMDA}_{AS}^{h,m} &= \frac{\text{MDA}_{AS}^{h,m} - \min\limits_{s}\left\{\text{MDA}_{s}^{h,m}\right\}}{\max\limits_{s}\left\{\text{MDA}_{s}^{h,m}\right\} - \min\limits_{s}\left\{\text{MDA}_{s}^{h,m}\right\}}, \\
\text{nMDV}_{AS}^{h,m} &= \frac{\text{MDV}_{AS}^{h,m} - \min\limits_{s}\left\{\text{MDV}_{s}^{h,m}\right\}}{\max\limits_{s}\left\{\text{MDV}_{s}^{h,m}\right\} - \min\limits_{s}\left\{\text{MDV}_{s}^{h,m}\right\}}.
\end{aligned}
$$

If an adaptive strategy performs better (worse) than the best (worst) unconditional model then normalized statistics are larger than 1 (smaller than 0).

For a given forecast horizon the sum of normalized losses for forecasts of the 2yr, 5yr and 10yr rates for the six best strategies are provided in Table 4.3. With respect to the MSFE criterion the Med strategy yields superior normalized losses for all horizons. The Av strategy performs slightly worse for horizons $h = 5, 10, 15$ and is overall the second best performing strategy. The Av10MSFE strategy is, for all horizons, among the best three adaptive strategies. In terms of MDA and MDV the Med strategy is again preferable overall. For $h = 1, 10, 15$ normalized losses are always better than the normalized losses of at least all but one adaptive strategy. Apart from the MSFE criterion, the Av10MDA and BunnMDA strategies are overall the second and third best competitor strategies.

Normalized MSFE

	$h=1$		$h=5$		$h=10$		$h=15$		all horizons	
1	Med	2.992	Med	2.972	Med	2.942	Med	2.904	Med	11.810
2	Av10MSFE	2.990	Av	2.935	Av	2.886	Av	2.832	Av	11.613
3	BunnMSFE	2.990	Av10MSFE	2.875	Av10MSFE	2.640	Av10MSFE	2.407	Av10MSFE	10.913
4	AnoMSFE	2.981	BunnMSFE	2.872	BunnMSFE	2.637	BunnMSFE	2.394	BunnMSFE	10.892
5	MinMSFE	2.981	AnoMSFE	2.857	BunnMDV	2.626	Av10MDA	2.367	AnoMSFE	10.763
6	BunnMDA	2.967	BunnMDA	2.835	Av10MDA	2.620	Av10MDV	2.345	Av10MDA	10.759

Normalized MDA

	$h=1$		$h=5$		$h=10$		$h=15$		all horizons	
1	AnoMSFE	2.378	AnoMSFE	2.662	Med	2.478	Av	2.379	Med	9.498
2	Med	2.321	BunnMSFE	2.538	Av	2.267	Med	2.370	Av10MDA	8.404
3	Av10MSFE	2.052	Av10MDA	2.487	BunnMDV	2.124	Av10MDA	2.047	BunnMDA	8.347
4	MinMSFE	2.019	BunnMDA	2.484	Av10MDV	2.101	Av10MDV	1.973	AnoMSFE	8.095
5	BunnMDA	2.002	Av10MSFE	2.471	AnoMDV	2.044	AnoMDA	1.855	BunnMSFE	7.851
6	BunnMSFE	1.872	BunnMDV	2.401	AnoMDA	2.038	BunnMDA	1.846	Av10MSFE	7.826

Normalized MDV

	$h=1$		$h=5$		$h=10$		$h=15$		all horizons	
1	AnoMSFE	2.595	AnoMSFE	2.543	Med	2.448	Av	2.330	Med	9.563
2	Med	2.491	Av10MDA	2.421	Av10MDV	2.274	Med	2.250	Av10MDA	8.570
3	MaxMDA	2.362	BunnMDA	2.417	BunnMDV	2.255	Av10MDV	2.064	BunnMDA	8.540
4	Av10MSFE	2.338	BunnMSFE	2.386	Av	2.221	Av10MDA	2.063	AnoMSFE	8.531
5	BunnMSFE	2.210	Med	2.373	AnoMDV	2.209	AnoMDA	1.941	BunnMDV	8.396
6	MinMSFE	2.160	BunnMDV	2.328	BunnMDA	2.197	BunnMDA	1.902	BunnMSFE	8.383

Table 4.3: MSFE, MDA and MDV comparison of adaptive strategies. For a given forecast horizon the sum of normalized losses for forecasts of the 2yr, 5yr and 10yr swap rates are provided. Normalization is accomplished with respect to the best and worst unconditional models in terms of MSFE, MDA and MDV. Results are shown for the six best adaptive strategies. To account for the fact that MDA and MDV are success measures, the ANOVA strategy is implemented by regressing -MDA and -MDV on the respective dummy variables. For further notes see Table 4.1.

4.3.2 Unconditional models vs. adaptive strategies

Having identified the overall best adaptive strategies, we analyze in this section how the forecasts from these adaptive strategies perform relative to unconditional models. Table 4.4 shows for each forecasting exercise normalized average loss estimates. Moreover, it provides, for a given adaptive strategy the number of worse performing unconditional model specifications M_s (columns labelled $\succ$), i.e.

$$\begin{array}{c} \sum_{s=1}^{100} \mathrm{I}(\mathrm{MSFE}_{AS}^{h,m} < \mathrm{MSFE}_{s}^{h,m}), \\ \sum_{s=1}^{100} \mathrm{I}(\mathrm{MDA}_{AS}^{h,m} > \mathrm{MDA}_{s}^{h,m}), \\ \sum_{s=1}^{100} \mathrm{I}(\mathrm{MDV}_{AS}^{h,m} > \mathrm{MDV}_{s}^{h,m}) \;. \end{array}$$

From the upper panel of Table 4.4 it can be seen that the adaptive strategies perform well in terms of MSFE. No adaptive strategy is worse than the 40th best unconditional model. Recall from Section 4.2 that the 50 best unconditional models perform similarly well and robustly over time. Hence it is not surprising that the best adaptive strategies outperform 40 unconditional specifications. The Med strategy is always better than at least 68 unconditional models. The Av strategy is in nine forecast exercises better than 62, and the Av10MSFE strategy is still in three cases better than 63 unconditional models. The relative performance in terms of MDA and MDV is documented in the two lower panels of Table 4.4. The Med strategy is always better than 66 unconditional models (except for the 5yr rate and $h = 5$ in terms of MDA). For the 10yr rate and $h = 10$ it is even better than the best unconditional model both in terms of MDA and MDV. For six forecasting exercises (the 2yr rate for $h = 5, 10, 15$, the 5yr rate, $h = 1, 15$, and the 10yr rate for $h = 5$) all three adaptive strategies are better than at least 60 unconditional models in terms of MDA. Regarding the MDV measure, all three adaptive strategies are better than at least 65 unconditional models, except for forecasting the 2yr and 10yr rate for $h = 1$ and the 10yr rate for $h = 10$.

An analysis of all adaptive strategies considered in Section 4.3.1 reveals that 'on average' adaptive model selection is more successful in terms of MDA and MDV than in terms of MSFE. Table 4.4 provides, for a given forecast exercise and loss function, the average number of underperforming unconditional models over all adaptive strategies. It shows that in nine out of twelve forecast exercises adaptive procedures outperform, on average, more unconditional models in terms of MDA and MDV than in terms of MSFE. These results can be viewed as an indication of the robustness of adaptive model selection in terms of MDA and MDV.

Choosing evaluation windows of length $\widetilde{\tau} = 42$ is thought to balance the needs of modeling flexibility under local heterogeneity on the one hand and statistical precision of parameter estimates on the other. A robustness analysis (results are available from the authors upon request) for the parameter $\widetilde{\tau}$ reveals that the general conclusions remain valid for $\widetilde{\tau} \in \{42, 63, 126, 189\}$. For $\widetilde{\tau} = 252$ the performance of the model adaptation procedures in terms of MDA and MDV deteriorates, indicating that this window size is too large to cope with the prevalence of local heterogeneity.

We conclude that adaptive model selection approaches offer promising forecast performance within the class of PCA/AR models. Furthermore, it is of interest to see how the adaptive procedures compare to some standard benchmark models. We note that the adaptive approach rarely leads to additional forecast accuracy in terms of MSFE when compared to the benchmark models. Hence, further results for the MSFE measure are not reported.

4.3.3 Adaptive forecasts vs. benchmark approaches

We compare the adaptive strategies with naive forecasts, autoregressive time series models, the Diebold and Li (2006) approach and the ex–post best unconditional specification in terms of MDA and MDV parameterized with $\{\tau, K, p\} = \{63, 3, 1\}$.

For the purpose of measuring DA and DV, the naive forecast is always a downward movement. Average losses of the naive strategy are denoted by $\text{MDA}^{h,m}_{\text{Naive}}$ and $\text{MDV}^{h,m}_{\text{Naive}}$. Next, time series forecasts for swap rate levels are based on a univariate AR(1) model for the first differences of the 2yr, 5yr and 10yr swap rates. This model is fitted recursively to time windows of length 42 or 252 days. For these benchmark forecasts average performance measures are denoted by $\text{MDA}^{h,m}_{\bullet}$ and $\text{MDV}^{h,m}_{\bullet}$, $\bullet =$ AR42, AR252. Using a decay parameter of $\lambda_t = 0.0609$, the Diebold and Li (2006) model is implemented by recursively fitting uncorrelated AR(1) processes for the first differences of factors using samples comprising 42 or 252 trading days. Average losses of iterated forecasts are denoted by $\text{MDA}^{h,m}_{\bullet}$ and $\text{MDV}^{h,m}_{\bullet}$, $\bullet =$ DL42, DL252.

Finally, we also consider the unconditional model $\{\tau, K, p\} = \{63, 3, 1\}$ and denote its average performance by $\text{MDA}^{h,m}_{\text{BestUnc}}$ and $\text{MDV}^{h,m}_{\text{BestUnc}}$. As mentioned in Section 4.2, this specification performs best, ex–post, in the class of PCA/AR models when compared over all forecast exercises. From the point of view of an analyst, selecting the best model from the set of PCA/AR models might be seen as the outcome of a random experiment. Hence, the ex–post best performing model is hardly a benchmark model which analysts attempt to outperform. Rather, it is the model that they try to approximate as closely as possible.

The two rightmost columns of Table 4.5 show that the adaptive strategies Med, Av10MDA and BunnMDA outperform the benchmark strategies

	2yr								5yr								10yr							
	$h=1$		$h=5$		$h=10$		$h=15$		$h=1$		$h=5$		$h=10$		$h=15$		$h=1$		$h=5$		$h=10$		$h=15$	
	nMSFE	$\succ$	nMSFE	$\succ$	nMSFE	$\succ$	nMSFE	$\succ$	nMSFE	$\succ$	nMSFE	$\succ$	nMSFE	$\succ$	nMSFE	$\succ$	nMSFE	$\succ$	nMSFE	$\succ$	nMSFE	$\succ$	nMSFE	$\succ$
Med	0.997	86	0.991	83	0.988	87	0.987	88	0.995	79	0.985	82	0.968	80	0.957	79	0.999	94	0.996	85	0.986	72	0.960	68
Av	0.982	41	0.973	62	0.967	76	0.969	81	0.987	41	0.973	64	0.947	64	0.929	74	0.991	42	0.988	75	0.971	68	0.934	68
Av10MSFE	0.996	83	0.951	51	0.894	48	0.816	43	0.996	85	0.958	50	0.873	45	0.808	49	0.998	63	0.967	45	0.874	36	0.784	44
Mean14	0.984	52	0.943	47	0.891	50	0.828	48	0.984	51	0.938	39	0.848	40	0.759	43	0.972	41	0.928	35	0.847	37	0.759	43
	nMDA	$\succ$	nMDA	$\succ$	nMDA	$\succ$	nMDA	$\succ$	nMDA	$\succ$	nMDA	$\succ$	nMDA	$\succ$	nMDA	$\succ$	nMDA	$\succ$	nMDA	$\succ$	nMDA	$\succ$	nMDA	$\succ$
Med	0.842	76	0.789	77	0.627	66	0.688	72	0.798	87	0.747	56	0.834	88	0.806	80	0.681	68	0.792	74	1.017	100	0.876	91
Av10MDA	0.650	37	0.801	77	0.565	60	0.733	77	0.826	90	0.822	77	0.731	64	0.764	74	0.374	29	0.864	88	0.726	41	0.550	27
BunnMDA	0.683	45	0.843	84	0.582	61	0.713	74	0.780	82	0.856	90	0.674	49	0.624	61	0.538	47	0.784	72	0.760	51	0.509	25
Mean14	0.627	39	0.674	63	0.559	59	0.627	65	0.676	60	0.830	78	0.720	61	0.616	59	0.440	38	0.769	68	0.675	39	0.594	37
	nMDV	$\succ$	nMDV	$\succ$	nMDV	$\succ$	nMDV	$\succ$	nMDV	$\succ$	nMDV	$\succ$	nMDV	$\succ$	nMDV	$\succ$	nMDV	$\succ$	nMDV	$\succ$	nMDV	$\succ$	nMDV	$\succ$
Med	0.845	67	0.745	76	0.611	67	0.568	68	0.906	92	0.831	77	0.826	87	0.798	81	0.740	77	0.797	83	1.010	100	0.884	96
Av10MDA	0.587	26	0.847	87	0.623	69	0.773	83	1.022	100	0.756	65	0.830	88	0.807	81	0.289	10	0.818	85	0.735	70	0.483	24
BunnMDA	0.689	36	0.900	92	0.655	75	0.793	83	0.950	95	0.762	67	0.786	83	0.665	66	0.387	17	0.754	69	0.756	74	0.444	18
Mean14	0.731	50	0.690	67	0.611	68	0.653	74	0.766	66	0.802	73	0.770	79	0.667	66	0.505	37	0.725	60	0.675	48	0.575	37

Table 4.4: MSFE, MDA and MDV comparison of adaptive and unconditional strategies. For each forecast exercise normalized average losses are provided in columns 'nMSFE', 'nMDA' and 'nMDV'. The number of unconditional models that perform (strictly) worse than the adaptive strategy given in the first column is shown in the column labeled $\succ$. 'Mean14' documents for a given forecast exercise the average of normalized losses and the average of number of underperforming unconditional models over all 14 adaptive strategies considered. For further notes see Table 4.1.

except the best unconditional model in a comparison over all forecast exercises. In particular, the Med strategy is overall best in terms of MDA and MDV. For the latter measure it is in eight forecast exercises (2yr rate for $h = 1, 5, 10$, 5yr for all horizons, 10yr rate for $h = 10$) better than all other strategies. With regard to the MDA measures, in eight forecasting exercises (2yr rate for $h = 1, 5, 15$, 5yr rate for all horizons, 10yr rate for $h = 10$) at least one of the three adaptive strategies outperforms all benchmark models. In terms of MDV, this is the case for ten forecast exercises (all but the 10yr rate and $h = 1, 5$).

With regard to the approximation of the best unconditional model, the last two columns of Table 4.5 reveal that the adaptive approach is more successful than the remaining benchmark approaches in terms of overall average losses.

A summary of bilateral model comparisons is provided in Table 4.6. Furthermore, using the Diebold and Mariano (1995) approach, we formally test if the expected loss of a particular adaptive strategy is significantly larger than the expected loss of the naive, the AR and BestUnc benchmark strategy (which outperform the Diebold–Li model). The number of forecast exercises $FE_j,\ j = 1, \ldots, 12$, in which an adaptive strategy $AS \in$ {Med, Av10MDA, BunnMDA} is (with 5% significance) better than the benchmark model $BM \in$ {Naive, AR42, AR252, BestUnc} is provided in the left hand side panels of Table 4.6. The right hand panels show how often a particular benchmark model BM (significantly) outperforms an adaptive strategy AS. Table 4.6 reveals that the BestUnc model is superior to any adaptive strategy in at least eight and seven forecast exercises in terms of MDA and MDV, respectively. Such a result is to be expected from an ex–post best model. Yet, in only at most three out of twelve forecast exercises it is significantly better than any adaptive strategy for MDA and MDV.

Normalized MDA

	2yr								5yr								10yr								SumRank	
	$h=1$		$h=5$		$h=10$		$h=15$		$h=1$		$h=5$		$h=10$		$h=15$		$h=1$		$h=5$		$h=10$		$h=15$			
	nMDA	$\succ$	nMDA	$\succ$	nMDA	$\succ$	nMDA	$\succ$	nMDA	$\succ$	nMDA	$\succ$	nMDA	$\succ$	nMDA	$\succ$	nMDA	$\succ$	nMDA	$\succ$	nMDA	$\succ$	nMDA	$\succ$		
Naive	0.325	9	0.163	10	0.102	7	0.119	8	0.798	87	0.637	39	0.627	46	0.661	64	0.516	44	0.984	98	0.971	99	1.046	100	6.949	(6)
AR42	0.425	19	0.675	63	0.650	71	0.703	73	0.514	27	0.801	74	0.772	79	0.788	79	0.165	10	0.872	89	0.800	66	0.844	89	8.008	(5)
AR252	0.633	35	0.500	43	0.401	38	0.366	36	0.697	67	0.466	14	0.399	23	0.297	21	0.879	98	0.808	79	0.766	52	0.638	35	6.850	(7)
DL42	0.400	15	0.404	21	0.542	59	0.614	68	0.037	3	0.267	4	0.523	44	0.533	48	0.319	26	0.536	18	0.800	66	0.867	90	5.842	(8)
DL252	0.458	22	0.337	17	0.390	38	0.446	48	0.147	4	-0.007	0	0.041	3	-0.218	0	0.253	20	0.776	70	0.674	30	0.573	28	3.871	(9)
BestUnc	0.758	65	0.988	95	1.000	99	0.946	97	0.835	92	0.904	93	0.927	96	0.988	97	0.824	92	0.608	30	0.777	56	0.803	69	10.358	(1)
Med	0.842	76	0.789	77	0.627	66	0.688	72	0.798	87	0.747	56	0.834	88	0.806	80	0.681	68	0.792	74	1.017	100	0.876	91	9.498	(2)
Av10MDA	0.650	37	0.801	77	0.565	60	0.733	77	0.826	90	0.822	77	0.731	64	0.764	74	0.374	29	0.864	88	0.726	41	0.550	27	8.404	(3)
BunnMDA	0.683	45	0.843	84	0.582	61	0.713	74	0.780	82	0.856	90	0.674	49	0.624	61	0.538	47	0.784	72	0.760	51	0.509	25	8.347	(4)

Normalized MDV

	2yr								5yr								10yr								SumRank	
	$h=1$		$h=5$		$h=10$		$h=15$		$h=1$		$h=5$		$h=10$		$h=15$		$h=1$		$h=5$		$h=10$		$h=15$			
	nMDV	$\succ$	nMDV	$\succ$	nMDV	$\succ$	nMDV	$\succ$	nMDV	$\succ$	nMDV	$\succ$	nMDV	$\succ$	nMDV	$\succ$	nMDV	$\succ$	nMDV	$\succ$	nMDV	$\succ$	nMDV	$\succ$		
Naive	0.265	10	0.227	8	0.087	8	0.113	18	0.453	17	0.602	33	0.528	42	0.567	53	0.347	14	0.851	92	0.898	96	0.866	96	5.803	(7)
AR42	0.458	23	0.611	57	0.596	66	0.686	75	0.543	28	0.773	67	0.659	63	0.719	73	0.132	4	0.819	86	0.679	56	0.642	44	7.319	(5)
AR252	0.399	21	0.468	41	0.303	33	0.300	35	0.364	8	0.376	10	0.295	10	0.300	31	0.719	71	0.769	76	0.700	62	0.655	52	5.648	(8)
DL42	0.348	16	0.307	13	0.422	50	0.604	68	0.231	4	0.497	21	0.568	44	0.610	59	0.751	80	0.799	83	0.773	77	0.704	75	6.614	(6)
DL252	0.383	20	0.491	47	0.395	49	0.412	50	0.286	6	0.176	4	0.176	5	0.064	8	0.802	89	0.924	96	0.811	83	0.709	77	5.628	(9)
BestUnc	0.993	97	0.925	94	1.000	99	0.987	98	0.778	70	0.947	96	0.915	96	0.970	98	0.704	68	0.753	68	0.780	78	0.755	86	10.508	(1)
Med	0.845	67	0.745	76	0.611	67	0.568	68	0.906	92	0.831	77	0.826	87	0.798	81	0.740	77	0.797	83	1.010	100	0.884	96	9.563	(2)
Av10MDA	0.587	26	0.847	87	0.623	69	0.773	83	1.022	100	0.756	65	0.830	88	0.807	81	0.289	10	0.818	85	0.735	70	0.483	24	8.570	(3)
BunnMDA	0.689	36	0.900	92	0.655	75	0.793	83	0.950	95	0.762	67	0.786	83	0.665	66	0.387	17	0.754	69	0.756	74	0.444	18	8.540	(4)

Table 4.5: MDA and MDV of benchmark strategies and comparison with adaptive strategies. For each forecast exercise the normalized losses are given in the columns labeled nMDA and nMDV. The number of unconditional models that perform (strictly) worse than the strategy given in the first column is shown in the columns '$\succ$'. The last column shows for a given strategy the sum of the normalized MDAs and MDVs over the 12 forecast exercises. For further notes see Table 4.1.

Ignoring the BestUnc model, it can be verified from the left hand side panels of Table 4.6 that any of the three adaptive strategies is better than a given benchmark model in at least six (out of twelve) forecast exercises in terms of MDA. For the MDV measure the results are even more compelling. Each adaptive strategy outperforms a given benchmark model in at least eight forecast exercises. In particular, the Med strategy is (significantly) better than the naive, AR42 and AR252 benchmark in 11 (2), 10 (1) and 12 (5) forecast exercises, respectively. It is noteworthy that no benchmark model except for the BestUnc model significantly outperforms any adaptive strategy AS in terms of MDV. Hence, we conclude that adaptive model selection/estimation within the class of PCA/AR models is preferable to standard benchmark models with Med being the most convincing adaptive approach.

4.4 Conclusions

Based on a factor model characterized by a dynamic autoregressive factor representation we forecast 2yr, 5yr and 10yr swap rates one day and one, two and three weeks ahead. We compare a set of 100 unconditional model specifications to a variety of adaptive model selection strategies. Additionally, the latter procedures are compared with a naive and autoregressive time series models and the Nelson and Siegel(1987) and Diebold and Li (2006) term structure models. Within a unified loss functional framework the comparison builds upon out–of–sample forecast performance measured by quadratic loss, directional accuracy and directional forecast value.

The measure for directional forecast value as defined in this paper may also be used to evaluate the profitability of trading systems. For basic financial instruments such as stocks it represents cash flows from an elementary buy/sell strategy. For quasi linear financial derivatives, such as swaps, it is

MDA

	Naive	AR42	AR252	BestUnc	Sum		Med	Av10 MDA	Bunn MDA	Sum
Med	9 (3)	8 (2)	10 (3)	4 (0)	31 (8)	Naive	3 (0)	4 (1)	5 (1)	12 (2)
Av10MDA	8 (2)	6 (0)	9 (0)	1 (0)	24 (2)	AR42	4 (0)	6 (1)	6 (1)	16 (2)
BunnMDA	7 (2)	6 (0)	8 (0)	1 (0)	22 (2)	AR252	2 (0)	3 (1)	4 (0)	9 (1)
						BestUnc	8 (1)	11 (2)	11 (3)	30 (6)

MDV

	Naive	AR42	AR252	BestUnc	Sum		Med	Av10 MDA	Bunn MDA	Sum
Med	11 (2)	10 (1)	12 (5)	5 (0)	38 (8)	Naive	1 (0)	4 (0)	3 (0)	8 (0)
Av10MDA	8 (4)	9 (1)	10 (3)	2 (0)	29 (8)	AR42	2 (0)	3 (0)	4 (0)	9 (0)
BunnMDA	9 (5)	8 (0)	9 (3)	2 (0)	28 (8)	AR252	0 (0)	2 (0)	3 (0)	5 (0)
						BestUnc	7 (2)	10 (2)	10 (2)	27 (6)

Table 4.6: MDA and MDV comparison of benchmark and adaptive strategies for 2yr, 5yr, 10yr swap rates and horizons $h = 1, 5, 10, 15$. Each panel provides the number of forecast exercises, out of 12, in which the strategy given in the first column (significantly) outperforms the strategy given in the first row in terms of MDA and MDV. A one-sided null hypothesis of forecast outperformance is tested by means of the Diebold–Mariano test statistic, Diebold and Mariano (1995). Significance is assessed at the 5% level. Standard deviations are computed using Bartlett's kernel and a truncation lag chosen as the integer part of $(4(1778/100)^{2/9})$, Newey and West (1994). Note that for the forecast exercise $h = 1$ and the 5yr rate the Med and Naive strategy have a normalized MDA of 0.798. For further notes see Table 4.1.

proportional to cash flows of a buy/sell strategy. Our definition of DV can easily be generalized using the cash flow function based on the 'exact' pricing function of the financial instrument or portfolio under consideration. Hence, in this framework it is possible to test for significant differences in profitability between two or more trading systems: see also Diebold and Mariano (1995) and West (2006).

We find that an adaptive model selection approach shows a robust performance within the class of PCA/AR models for all loss functions considered. Moreover, the adaptive procedures lead to additional gains in directional accuracy and directional forecast value when compared to the benchmark models. In particular, the median strategy and the strategies based on an equal/non–equal weighted average of the ten locally best performing unconditional models in terms of MDA (Av10MDA, BunnMDA) turn out to yield robust and highly accurate forecasts for distinct swap rates and forecast horizons. This result can be interpreted as evidence for an evolving economy characterized by changing underlying relations in economic variables. In the presence of dynamic heterogeneity, the risk of selecting a poor unconditional model is mitigated by locally adaptive model selection procedures. Hence, we conclude, again, that an adaptive approach is a promising and costless candidate for ex–ante forecasting that merits further consideration.

Chapter 5

Testing the Economic Value of Directional Forecasts in the Presence of Serial Correlation

Forecasts are produced in numerous areas as they are important tools for decision making. The implication of a decision based on a forecast can be evaluated by means of the (expected) gain/loss associated with the decision. A commonly used loss function for quantitative forecasts is the quadratic loss of the forecast error. Yet, the squared forecast error provides only a partial assessment of economic forecasts. Diebold and Mariano (1995) point out that in light of the variety of economic decision problems relying on forecasts, statistical loss functions such as quadratic loss need not necessarily conform to economic loss functions. Granger and Pesaran (2000) discuss relationships between statistical and economic measures of forecast accuracy and stress that the choice of the evaluation measure should be related to the objectives of the forecast user. Assessing the directional accuracy (DA) of predicted directions may provide valuable insights into forecast evaluation. Lai (1990) emphasizes that an investor can still gain profits even with statistically biased forecasts if they are on the correct side of the price change more often than not. Leitch and Tanner (1995) find that DA is highly correlated with profits in an interest rate setting. As standard measures such as mean squared/absolute forecast error (MSFE, resp. MAFE) are less correlated with profits, they conclude that DA is a better measure of forecast accuracy for profit maximizing firms. Ash, Smith and Heravi (1998) note that qualitative statements such as the economy is expanding or the economy is contracting in the near future are important pre–requisites for an appropriate implementation of monetary and fiscal policy. Öller and Barot (2000) point out that DA is of interest for central banks. A forecast of increased inflation (above target) would prompt central banks to raise interest rates.

An approach to assess directional forecasts which is linked but not equivalent to the loss functional approach is based on Merton (1981). He proposes an equilibrium theory for the economic value of market timing skills and provides a statistic to measure the value. Cicarelli (1982) uses the statistical measure to analyze turning point errors. Havenner and Modjtahedi (1988), Breen, Glosten and Jagannathan (1989), Schnader and Stekler (1990), Lai (1990) and Stekler (1994) were among the first to apply Merton's theory to evaluate the economic value of directional forecasts. More recent applications include, inter alia, Ash, Smith and Heravi (1998), Mills and Pepper (1999), Öller and Barot (2000), Pons (2001), Easaw, Garratt and Heravi (2005) and Ashiya (2003, 2006). Considering realized and forecasted directions as binary variables, Merton's theory implies that directional forecasts have no value if the directional outcomes and forecasts are independent. Henriksson and Merton (1981) propose statistical procedures for evaluating forecasting skills that are in fact related to Fisher's (1934) exact test for testing whether two binary variables are independent. Similarly, the classical asymptotic χ^2–test for independence and the asymptotic test for market timing introduced by Pesaran and Timmerman (1992, PT92 henceforth) can be used for testing the economic value of directional forecasts. Yet, these tests are derived under the assumption of serial independence. As we outline later, they are seriously oversized in the presence of serially correlated forecasted resp. realized directions.

Recently, Pesaran and Timmerman (2008, PT08 henceforth) have introduced statistics for testing dependence among serially correlated multi–category variables which can be used to test for the economic value of directional forecasts in the more realistic situation of serial correlation. However, their test procedures reveal some small sample size distortions in a Monte Carlo simulation study. In this paper, we summarize and analyze the size and power properties of a battery of tests for the economic value of directional forecasts in the presence of serial correlation. Furthermore, we propose a bootstrap test procedure to reduce size distortions in small samples. We show in a simulation study that the bootstrap test is robust to serial correlation and has appealing power properties. Our approach can be put in a more general framework, i.e. testing dependence of two binary variables in the presence of serial correlation. Moreover, it can be easily extended to multi–categorical data.

The remainder of the paper is organized as follows. We briefly review Merton's approach in the next Section. In Section 5.2 existing test procedures and the bootstrap approach are summarized. Section 5.3 documents a Monte Carlo study to analyze size and power properties of the tests. Section 5.4 provides two empirical applications and Section 5.5 concludes.

5.1 Merton's framework for evaluating directional forecasts

Merton (1981) proposes an equilibrium theory for the value of market timing skills. In the context of evaluating directional forecasts for a variable of interest Y_t, let realized upward resp. downward movements in Y_t be denoted by $\widetilde{Y}_t = 1$, respectively, $\widetilde{Y}_t = 0$. Forecasted upward resp. downward movements are denoted by $\widetilde{X}_t = 1$ resp. $\widetilde{X}_t = 0$. It is assumed that forecasts $\widetilde{X}_t$ are determined using only information up to time $t-1$. A directional forecast has no value in the sense of Merton (1981) if and only if

$$\mathbb{P}[\widetilde{X}_t = 1|\widetilde{Y}_t = 1] + \mathbb{P}[\widetilde{X}_t = 0|\widetilde{Y}_t = 0] \quad = \quad 1 \,. \tag{5.1}$$

In (5.1) $\mathbb{P}[\widetilde{X}_t = 1|\widetilde{Y}_t = 1]$ ($\mathbb{P}[\widetilde{X}_t = 0|\widetilde{Y}_t = 0]$) denote the conditional probability of a correct forecast of an upward (downward) movement. To alleviate notation, we define $HM = \mathbb{P}[\widetilde{X}_t = 1|\widetilde{Y}_t = 1] + \mathbb{P}[\widetilde{X}_t = 0|\widetilde{Y}_t = 0]$. For example, if $\widetilde{X}_t$ and $\widetilde{Y}_t$ are independent then $\mathbb{P}[\widetilde{X}_t = 1|\widetilde{Y}_t = 1] = \mathbb{P}[\widetilde{X}_t = 1]$ and $\mathbb{P}[\widetilde{X}_t = 0|\widetilde{Y}_t = 0] = \mathbb{P}[\widetilde{X}_t = 0]$. Consequently, $HM = 1$ and such directional forecasts have no value. In particular, naively forecasting only one direction, say $\widetilde{X}_t = 1 \; \forall t$, has no value.

Moreover, Merton (1981) points out that directional forecasts have positive value if and only if

$$HM > 1$$

and that the larger HM the larger the value. Noteworthy, it can be shown that

$$HM - 1 = \frac{\text{Cov}\left(\widetilde{X}_t, \widetilde{Y}_t\right)}{\mathbb{V}\left[\widetilde{Y}_t\right]} \,,$$

where $\text{Cov}(\widetilde{X}_t, \widetilde{Y}_t) = \mathbb{P}[\widetilde{X}_t = 1, \widetilde{Y}_t = 1] - \mathbb{P}[\widetilde{X}_t = 1]\mathbb{P}[\widetilde{Y}_t = 1]$ and $\mathbb{V}[\widetilde{Y}_t] = \mathbb{P}[\widetilde{Y}_t = 1] - \mathbb{P}[\widetilde{Y}_t = 1]^2$ denote the covariance between $\widetilde{X}_t$ and $\widetilde{Y}_t$ and the variance of $\widetilde{Y}_t$, respectively. Hence, the value of the forecasts can be assessed by means of the covariability of realized and forecasted directions. In particular, directional forecasts have (i) no value if and only if $\mathrm{Cov}(\widetilde{X}_t, \widetilde{Y}_t) = 0$ and (ii) have value if and only if $\mathrm{Cov}(\widetilde{X}_t, \widetilde{Y}_t) > 0$. Moreover, (iii) for a given process Y_t and hence $\widetilde{Y}_t$ (resp. $\mathbb{V}[\widetilde{Y}_t]$), it holds that the larger $\mathrm{Cov}(\widetilde{X}_t, \widetilde{Y}_t)$ the larger the value.

Furthermore, maximizing $\mathrm{Cov}(\widetilde{X}_t, \widetilde{Y}_t)$ is not equivalent to maximizing the probability of a correct directional forecast $\mathbb{P}[\widetilde{Z}_t = 1]$, where $\widetilde{Z}_t = I(\widetilde{X}_t = \widetilde{Y}_t)$ and $I(\bullet)$ denotes an indicator function. From the relationship

$$\mathrm{Cov}\left(\widetilde{X}_t, \widetilde{Y}_t\right) = \frac{1}{2}\mathbb{P}[\widetilde{Z}_t = 1] + \mathbb{P}[\widetilde{X}_t = 1]\left(\frac{1}{2} - \mathbb{P}[\widetilde{Y}_t = 1]\right) + \frac{1}{2}\left(\mathbb{P}[\widetilde{Y}_t = 1] - 1\right)$$

it can be seen that the correspondence between $\mathrm{Cov}(\widetilde{X}_t, \widetilde{Y}_t)$ and $\mathbb{P}[\widetilde{Z}_t = 1]$ is not monotone.

Consequently, for a given process Y_t, if the probability of a correct forecast $\mathbb{P}[\widetilde{Z}_t = 1]$ increases and the probability of an upward movement forecast $\mathbb{P}[\widetilde{X}_t = 1]$ changes, then

$$\Delta\mathrm{Cov}\left(\widetilde{X}_t, \widetilde{Y}_t\right) = \frac{1}{2}\Delta\mathbb{P}[\widetilde{Z}_t = 1] + \Delta\mathbb{P}[\widetilde{X}_t = 1]\left(\frac{1}{2} - \mathbb{P}[\widetilde{Y}_t = 1]\right) ,$$

with Δ denoting the total difference operator. Whether $\mathrm{Cov}(\widetilde{X}_t, \widetilde{Y}_t)$ increases depends on signs and magnitudes of $\Delta\mathbb{P}[\widetilde{X}_t = 1]$ and $\frac{1}{2} - \mathbb{P}[\widetilde{Y}_t = 1]$.

Moreover, the loss functional approach as defined below is not equivalent to the Merton approach. Frequently, loss functions to assess DA are defined as:

$$L_t = \begin{cases} a & \text{if } \widetilde{Z}_t = 1 \\ b & \text{if } \widetilde{Z}_t = 0, \end{cases}$$

where $(a,\ b) \neq 0$. Examples include Leitch and Tanner (1995), Greer (2005), Blaskowitz and Herwartz (2009b) where $(a,\ b) = (1, -1)$ or Swanson and White (1995, 1997a,b), Gradojevic and Yang (2006) and Diebold (2007) with $(a, b) = (1, 0)$. Hence, a correct directional forecast implies a loss of a (this is rather a gain if $a > 0$) and an incorrect directional forecast implies a loss of

b. In this case the expected DA is given by

$$\mathbb{E}[L_t] = (a-b)\mathbb{P}[\widetilde{Z}_t = 1] + b \ .$$

Consequently, maximizing expected DA is equivalent to maximizing the probability of a correct directional forecast (if $a > b$). See Pesaran and Skouras (2002) for a link between the HM statistic and a loss functional approach in a decision–based forecast evaluation framework. For test procedures using loss functions in the presence of serial correlation see, inter alia, Diebold and Mariano (1995) and West (2006).

Note that the value of directional forecasts in the sense of Merton does not take the magnitudes of realized and forecasted changes into account. Hence, the Merton framework is also different from the directional accuracy test proposed in Anatolyev and Gerko (2005) and from the notion of directional forecast value considered in Blaskowitz and Herwartz (2009b).

5.2 Testing for zero covariance

In this section, we first summarize some classical procedures to test for zero covariance between two categorical random variables when there is no serial dependence. Second, we describe tests for zero covariance in the presence of serial correlation and propose some bootstrap procedures to account for small sample size distortions. We consider tests of the null hypothesis

$$H_0 : \text{Cov}\left(\widetilde{X}_t, \widetilde{Y}_t\right) = 0 \ .$$

Notably, if $\widetilde{X}_t$ and $\widetilde{Y}_t$ are Bernoulli variables

$$\text{Cov}\left(\widetilde{X}_t, \widetilde{Y}_t\right) = 0 \quad \Leftrightarrow \quad \widetilde{X}_t \text{ and } \widetilde{Y}_t \text{ are independent} \ .$$

5.2.1 Testing for zero covariance under serial independence

In the framework outlined above it is straightforward to use 2×2 contingency tables whenever $\widetilde{X}_t$ and $\widetilde{Y}_t$ are serially independent. Testing H_0 can be accomplished using the asymptotic χ^2–test for independence. For small sample sizes Fisher's test (Fisher 1934) based on the hypergeometric distribution is exact and the uniformly most powerful unbiased (UMPU) test for H_0 when the marginals are fixed. If the latter condition does not hold, Fisher's test is no longer exact in finite samples but is asymptotically equivalent to the χ^2–test, see Agresti (1992) for a survey of exact inference for contingency tables.

PT92 proposed a test based on the difference of $\mathbb{P}[\widetilde{Z}_t = 1]$ under dependence and the probability of $\widetilde{Z}_t = 1$ under independence of $\widetilde{Y}_t$ and $\widetilde{X}_t$. In the former case it holds

$$\mathbb{P}[\widetilde{Z}_t = 1] = \mathbb{P}[\widetilde{Y}_t = 1, \widetilde{X}_t = 1] + \mathbb{P}[\widetilde{Y}_t = 0, \widetilde{X}_t = 0] \; .$$

If $\widetilde{Y}_t$ and $\widetilde{X}_t$ are independently distributed the probability of $\widetilde{Z}_t = 1$ is given by

$$\mathbb{P}_{indep}[\widetilde{Z}_t = 1] = \mathbb{P}[\widetilde{Y}_t = 1]\mathbb{P}[\widetilde{X}_t = 1] + \mathbb{P}[\widetilde{Y}_t = 0]\mathbb{P}[\widetilde{X}_t = 0] \; .$$

Hence, the test proposed by PT92 is based on

$$PT = \mathbb{P}[\widetilde{Z}_t = 1] - \mathbb{P}_{indep}[\widetilde{Z}_t = 1] = 2\text{Cov}\left(\widetilde{Y}_t, \widetilde{X}_t\right) \; .$$

Consequently, $\text{Cov}\left(\widetilde{X}_t, \widetilde{Y}_t\right) = 0$ if and only if $PT = 0$. Under the assumption of serial independence of $\widetilde{Y}_t$ resp. $\widetilde{X}_t$ and using a Hausman–type argument their proposed scaled test statistic is asymptotically Gaussian. Moreover, this test is asymptotically equivalent to the χ^2–test when two binary variables are considered. Granger and Pesaran (2000) and Pesaran and Skouras (2002) also derive a relationship between the HM statistic and the statistic proposed in PT92.

The three test procedures described above are frequently used within the context of directional forecast evaluation. The χ^2–approach is applied, inter alia, by Schnader and Stekler (1990), Artis (1996), Kolb and Stekler (1996), Swanson and White (1997a, 1997b), Ash, Smith and Heravi (1998), Mills and Pepper (1999), Öller and Barot (2000), Pons (2000, 2001), Easaw, Garratt and Heravi (2005) and Greer (2003, 2005). Applications of Fisher's test to analyse the value of directional forecasts include, among others, Havenner and Modjtahedi (1988), Lai (1990), Kuan and Liu (1995), Swanson and White (1995, 1997a, 1997b), Gençay (1998), Ash, Smith and Heravi (1998), Joutz and Stekler (1998, 2000), Easaw, Garratt and Heravi (2005) and Ashiya (2003, 2006). The test statistic proposed by PT92 is used, for example, by Pesaran and Timmerman (1995), Kuan and Liu (1995), Ash, Smith and Heravi (1998), Gençay (1998), Mills and Pepper (1999), Pons (2001), Schneider and Spitzer (2004) and Easaw, Garratt and Heravi (2005).

Another approach to test for zero covariance, which is useful when considering serial correlation over time, is given by the regression model

$$\widetilde{X}_t = \alpha + \beta \widetilde{Y}_t + \varepsilon_t \, , \tag{5.2}$$

where ε_t is a discrete zero mean random error. Note that for the population coefficient it holds $\beta = \text{Cov}\left(\widetilde{X}_t, \widetilde{Y}_t\right) / \mathbb{V}[\widetilde{Y}_t]$. Hence, testing H_0 amounts to standard significance tests for β in a linear regression model. Note, that we regard the regression model merely as a tool for testing purposes only. In our context the model in (5.2) does not have a 'causal' or 'economic' interpretation in the usual sense. Hence, it is also conceivable to regress $\widetilde{Y}_t$ on $\widetilde{X}_t$. These two approaches are asymptotically equivalent under the null hypothesis and differ only in terms of power (Anatolyev, 2006).

Moreover, consider the logistic regression model

$$\widetilde{X}_t = \frac{\exp\left(\alpha + \beta \widetilde{Y}_t\right)}{1 + \exp\left(\alpha + \beta \widetilde{Y}_t\right)} + \varepsilon_t \, ,$$

where ε_t is a discrete zero mean disturbance term. In this model with two binary variables it can be shown that

$$\frac{\text{Cov}\left(\widetilde{X}_t, \widetilde{Y}_t\right)}{\mathbb{V}\left[\widetilde{Y}_t\right]} = \left(e^{\beta} - 1\right) \mathbb{P}[\widetilde{X}_t = 1|\widetilde{Y}_t = 0]\mathbb{P}[\widetilde{X}_t = 0|\widetilde{Y}_t = 1]$$

(Cox and Hinkley, 1974). Again, it follows that Cov $\left(\widetilde{X}_t, \widetilde{Y}_t\right) = 0$ if and only if $\beta = 0$. Standard maximum likelihood estimation and likelihood ratio (LR) tests can be applied. The small sample distribution of the LR statistic is generally unknown but for Bernoulli variables $\widetilde{X}_t$ and $\widetilde{Y}_t$ the small sample LR test for $\beta = 0$ corresponds to Fisher's exact test (Cumby and Modest, 1987).

5.2.2 Testing for zero covariance in the presence of serial correlation

When there is serial dependence, the tests described above are no longer suitable. Bartlett (1951) and Patankar (1954) were among the first to show that for (Markov) dependent data the usual Pearson statistic for testing goodness of fit need not have common asymptotic χ^2–distribution. Within the framework of 2×2 contingency tables, Altham (1979) reports an inflated χ^2–statistic, $X^2_{I,T}$, when analyzing relationships between categorical variables observed over time and provides upper and lower bounds for the appropriate test statistic. Tavaré and Altham (1983) show that the classical χ^2 test statistic for independence is either inflated or deflated if $\widetilde{X}_t$ resp. $\widetilde{Y}_t$ are two–state Markov chains. For a general $r \times c$ contingency table Holt, Scott and Ewings (1980) and Tavaré (1983) establish that the asymptotic distribution of $X^2_{I,T}$ depends on unknown nuisance parameters under the null hypothesis if (in this case the multi–categorical variables) $\widetilde{X}_t$ resp. $\widetilde{Y}_t$ are arbitrary (but positive recurrent) Markov chains. More precisely, Holt, Scott and Ewings (1980) show that

$$X_{I,T}^2 \xrightarrow[T\to\infty]{\mathbb{D}} \sum_{i=1}^{(r-1)(c-1)} \rho_i Q_i^2 \ ,$$

where the Q_i are independent standard normal random variables and the ρ_i are the eigenvalues of a particular non–stochastic matrix which depends on the parameters of the underlying DGP. For a 2×2 contingency table and ρ_1 estimated by the Maximum Likelihood method (assuming a first–order Markov chain) the simulation experiment in Pesaran and Timmerman (2008) shows that this test procedure is oversized in small samples. Noteworthy, Tavaré (1983) also demonstrates that $X_{I,T}^2$ is still asymptotically distributed as χ^2 with $(r-1)(c-1)$ degrees of freedom when one process, say $\widetilde{X}_t$, is serially independent. Yet, if $\widetilde{Y}_t$ are directions of a serially correlated economic time series and $\widetilde{X}_t$ are reasonable directional forecasts of $\widetilde{Y}_t$ then both processes most likely exhibit serial correlation.

Furthermore, PT08 show in a simulation experiment that the test for market timing proposed in PT92 is seriously oversized in the presence of serial dependence. Finally, it is well known that coefficient tests in a regression model are size distorted if serial correlation is not taken into account. In the sequel we sketch some testing procedures that account for the more general situation of linear dependence over time.

Covariance test

The first robust approach is based on a classical covariance estimator and an estimator of its variance which accounts for serial correlation. Let $p_{\widetilde{Y}} = \mathbb{P}[\widetilde{Y}_t = 1]$ resp. $p_{\widetilde{X}} = \mathbb{P}[\widetilde{X}_t = 1]$ be constant over time, and decompose

$$\widetilde{Y}_t = p_{\widetilde{Y}} + \varepsilon_t^{\widetilde{Y}} \ \text{ resp. } \ \widetilde{X}_t = p_{\widetilde{X}} + \varepsilon_t^{\widetilde{X}} \ ,$$

where $\varepsilon_t^{\widetilde{Y}}$ resp. $\varepsilon_t^{\widetilde{X}}$ are binary zero mean random errors which may be serially correlated. Consequently, the null hypothesis that $\text{Cov}(\widetilde{Y}_t, \widetilde{X}_t) = 0$

is equivalent to $\mathbb{E}[\varepsilon_t^{\widetilde{Y}} \varepsilon_t^{\widetilde{X}}] = 0$. Under suitable assumptions (e.g. stationarity and weak dependence of $\{\varepsilon_t^{\widetilde{Y}} \varepsilon_t^{\widetilde{X}}\}_{t=1}^{T}$) a central limit theorem for $\frac{1}{T}\sum_{t=1}^{T} \varepsilon_t^{\widetilde{Y}} \varepsilon_t^{\widetilde{X}}$ holds (Lütkepohl, 2006):

$$\sqrt{T}\left(\frac{1}{T}\sum_{t=1}^{T} \varepsilon_t^{\widetilde{Y}} \varepsilon_t^{\widetilde{X}} - \mathbb{E}\left[\varepsilon_t^{\widetilde{Y}} \varepsilon_t^{\widetilde{X}}\right]\right) \xrightarrow[T\to\infty]{\mathbb{D}} N(0, S) \ ,$$

where $S = \sum_{j=-\infty}^{\infty} \text{Cov}\left(\varepsilon_t^{\widetilde{Y}} \varepsilon_t^{\widetilde{X}}, \varepsilon_{t-j}^{\widetilde{Y}} \varepsilon_{t-j}^{\widetilde{X}}\right)$ denotes the approximate asymptotic variance of $\sum_{t=1}^{T} \varepsilon_t^{\widetilde{Y}} \varepsilon_t^{\widetilde{X}}/T$. With the consistent estimators $\widehat{p}_{\widetilde{Y}} = \frac{1}{T}\sum_{t=1}^{T} \widetilde{Y}_t$ and $\widehat{p}_{\widetilde{X}} = \frac{1}{T}\sum_{t=1}^{T} \widetilde{X}_t$, the unobserved random errors can be estimated consistently by $\widehat{\varepsilon}_t^{\widetilde{Y}} = \widetilde{Y}_t - \widehat{p}_{\widetilde{Y}}$ resp. $\widehat{\varepsilon}_t^{\widetilde{X}} = \widetilde{X}_t - \widehat{p}_{\widetilde{X}}$.

Hence, letting $\overline{\varepsilon^{\widetilde{Y}} \varepsilon^{\widetilde{X}}} = \frac{1}{T}\sum_{t=1}^{T} \widehat{\varepsilon}_t^{\widetilde{Y}} \widehat{\varepsilon}_t^{\widetilde{X}} = \widehat{\text{Cov}}\left(\widetilde{Y}_t, \widetilde{X}_t\right)$ it follows that

$$CovNW_T = \sqrt{T}\frac{\left(\overline{\varepsilon^{\widetilde{Y}} \varepsilon^{\widetilde{X}}} - \mathbb{E}\left[\varepsilon_t^{\widetilde{Y}} \varepsilon_t^{\widetilde{X}}\right]\right)}{\sqrt{\widehat{S}_T^{NW}}} \xrightarrow[T\to\infty]{\mathbb{D}} N(0,1) \ . \tag{5.3}$$

In (5.3), $\widehat{S}_T^{NW}$ is the heteroscedasticity and autocorrelation consistent variance estimator (Newey and West, 1987) for $\widehat{\text{Cov}}\left(\widetilde{Y}_t, \widetilde{X}_t\right)$

$$\begin{aligned} \widehat{S}_T^{NW} = \widehat{\mathbb{V}}\left[\sqrt{T}\sum_{t=1}^{T} \varepsilon_t^{\widetilde{Y}} \varepsilon_t^{\widetilde{X}}\right] &= \widehat{\text{Cov}}\left(\varepsilon_t^{\widetilde{Y}} \varepsilon_t^{\widetilde{X}}, \varepsilon_t^{\widetilde{Y}} \varepsilon_t^{\widetilde{X}}\right) \\ &\quad + 2\sum_{q=1}^{Q} \omega(q,Q)\widehat{\text{Cov}}\left(\varepsilon_t^{\widetilde{Y}} \varepsilon_t^{\widetilde{X}}, \varepsilon_{t+q}^{\widetilde{Y}} \varepsilon_{t+q}^{\widetilde{X}}\right) \qquad (5.4) \\ \widehat{\text{Cov}}\left(\varepsilon_t^{\widetilde{Y}} \varepsilon_t^{\widetilde{X}}, \varepsilon_{t+q}^{\widetilde{Y}} \varepsilon_{t+q}^{\widetilde{X}}\right) &= \frac{1}{T}\sum_{t=1}^{T-q}\left(\widehat{\varepsilon}_t^{\widetilde{Y}} \widehat{\varepsilon}_t^{\widetilde{X}} - \overline{\varepsilon^{\widetilde{Y}} \varepsilon^{\widetilde{X}}}\right)\left(\widehat{\varepsilon}_{t+q}^{\widetilde{Y}} \widehat{\varepsilon}_{t+q}^{\widetilde{X}} - \overline{\varepsilon^{\widetilde{Y}} \varepsilon^{\widetilde{X}}}\right) \ , \end{aligned}$$

and the weighting function is defined as $\omega(q,Q) = (1 - \frac{q}{Q+1})$. The truncation lag Q can be chosen according to the integer part of $4(T/100)^{2/9}$ (Newey and West, 1994).

Note that under H_0 the squared statistic in (5.3) is equal to the Wald statistic discussed in Holt, Scott and Ewings (1980) or Rao and Scott (1981). The asymptotic covariance matrix of estimated cell proportions is determined by means of the Newey–West approach. We prefer the representation in (5.3) as it allows to test one–sided hypotheses which is particularly useful within the context of directional forecast evaluation.

Static/dynamic regression approach

A test of H_0 which accounts for serial correlation can also be accomplished in the linear regression model. First, consider the static regression model given in (5.2) where the disturbance term ε_t is allowed to be serially correlated. Then, the Newey–West corrected t–statistic for the OLS estimator $\widehat{\beta}_T^{OLS}$ is approximately Gaussian

$$\frac{\widehat{\beta}_T^{OLS} - \beta}{\sqrt{\widehat{\mathbb{V}}_T^{NW}[\widehat{\beta}_T^{OLS}]}} \dot{\approx} N(0,1)$$

(see Breen, Glosten and Jagannathan (1989) for an application).

Another possibility to allow for serial correlation is to dynamically augment model (5.2) with lagged dependent and explanatory variables $\widetilde{X}_t$ resp. $\widetilde{Y}_t$, i.e.:

$$\widetilde{X}_t \;=\; \gamma + \beta\widetilde{Y}_t - \sum_{j=1}^{m} \delta_j \widetilde{Y}_{t-j} + \sum_{j=1}^{m} \rho_j \widetilde{X}_{t-j} + u_t \,. \tag{5.5}$$

Testing H_0 in (5.5) amounts to a test of $\beta = 0$ after correcting for the effects of lagged dependent and explanatory variables. The number of lags m can be chosen according to some information criterion such as the Akaike Information Criterion (AIC). To account for remaining residual autocorrelation the Newey–West corrected t–statistic for $\widehat{\beta}_T^{OLS}$ can be computed (as in PT08). It is again approximately Gaussian. The truncation lag Q is chosen according to the integer part of $4(T/100)^{2/9}$. The tests based on (5.2) and (5.5) are called StatNW resp. DynNW.

Pesaran and Timmerman (2008) test

PT08 propose a more general approach for multicategory variables. Reinterpreting (5.5) as a reduced rank regression, they propose test statistics based on canonical correlations. For the time points $t = 1, ..., T$ and m initial values for $\widetilde{X}_t$ resp. $\widetilde{Y}_t$ model (5.5) can be rewritten as

$$\widetilde{X} = \widetilde{Y}\beta + WB + U \; ,$$

with

$$\underset{T\times 1}{\widetilde{X}} = \begin{pmatrix} \widetilde{X}_1 \\ \vdots \\ \widetilde{X}_T \end{pmatrix} \; , \; \underset{T\times 1}{\widetilde{Y}} = \begin{pmatrix} \widetilde{Y}_1 \\ \vdots \\ \widetilde{Y}_T \end{pmatrix} \; , \; \underset{T\times 1}{U} = \begin{pmatrix} u_1 \\ \vdots \\ u_T \end{pmatrix} \; ,$$

$$\underset{(2m+1)\times 1}{B} = (\gamma, \delta_1, ..., \delta_m, \rho_1, ..., \rho_m)' \; ,$$

$$\underset{T\times(2m+1)}{W} = \begin{pmatrix} 1 & \widetilde{Y}_0 & \widetilde{Y}_{-1} & ... & \widetilde{Y}_{-m+1} & \widetilde{X}_0 & \widetilde{X}_{-1} & ... & \widetilde{X}_{-m+1} \\ \vdots & & & & & & & & \vdots \\ 1 & \widetilde{Y}_{T-1} & \widetilde{Y}_{T-2} & ... & \widetilde{Y}_{T-m} & \widetilde{X}_{T-1} & \widetilde{X}_{T-2} & ... & \widetilde{X}_{T-m} \end{pmatrix} \; .$$

PT08 show that under the null hypothesis

$$(T-2)S \mathrel{\dot{\approx}} \chi^2_{(1)} \; ,$$

where

$$S = S_{XX}^{-1} S_{YX}' S_{YY}^{-1} S_{YX}, \;\; S_{YY} = \tfrac{1}{T}\widetilde{Y}'M\widetilde{Y}, \;\; S_{YX} = \tfrac{1}{T}\widetilde{Y}'M\widetilde{X}$$
$$S_{XX} = \tfrac{1}{T}\widetilde{X}'M\widetilde{X}, \;\; M = I_T - W\left(W'W\right)^{-1}W', \;\; \widetilde{Y} = (\widetilde{Y}_1, ..., \widetilde{Y}_T)' \; .$$

In the binary case S is a scalar random variable. Generally, S is a $(c_x - 1) \times (c_x - 1)$–matrix, with c_x being the number of $\widetilde{X}_t$–categories. For finite samples PT08 simulate critical values under H_0 using multinomial sampling. They consider a static and a dynamic version in full analogy to the regression based testing outlined before.

Bootstrap approach

We implement the bootstrap procedure for the covariance test in Section 5.2.2 as it allows both one– and two–sided alternative hypotheses. Moreover, the adaptation to general $r \times c$ contingency tables is possible. The block bootstrap is nowadays commonly accepted as an appropriate bootstrap method if an analyst wants to avoid to impose parametric restrictions on the structure of the data generating process. Künsch (1989), Lahiri (1991), Liu and Singh (1992), Politis and Romano (1992) were among the first to consider the bootstrap for time series. They show that the block bootstrap for time series is a suitable tool to obtain asymptotically valid procedures to approximate distributions of a large class of statistics and weakly dependent data generating processes. Radulovic (1996) proves that consistency of the block bootstrap for the mean usually holds when the statistic is asymptotically normal for a strongly mixing stationary sequence. Götze and Künsch (1996) and Lahiri (1996) cover the asymptotic refinements over the classical normal approximation of the error in rejection probability (ERP) of one–sided tests. Results for two–sided tests are given by Hall and Horowitz (1996), Andrews (2002) and Inoue and Shintani (2006). They demonstrate that the block bootstrap is more accurate than the normal approximation in terms of ERP for two–sided tests if properly implemented. Various blocking procedures have been proposed. The non–overlapping (NBB) resp. overlapping moving block bootstrap (MBB) were considered by Hall (1985), Carlstein (1986) and Kuensch (1989). Politis and Romano (1992, 1994) introduced the circular block bootstrap (CBB) and the stationary bootstrap (SB). Lahiri (1999) concludes that for estimating the distribution of a studentized statistic the MBB and CBB procedures are more efficient than NBB and SB versions in terms of MSE. The bootstrap sample mean has an expectation equal to the sample mean of the observed series under the CBB which is not the case for the MBB scheme. Hence, for the CBB centering the bootstrap distribution to establish a zero mean distribution is accomplished in the usual way.

To perform a two–sided test of $H_0 : \text{Cov}\left(\widetilde{Y}_t, \widetilde{X}_t\right) = 0$ we investigate two bootstrap approaches for the studentized statistic

$$ST_T \;=\; \frac{\sqrt{T}\left(\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^{\widetilde{Y}}\varepsilon_t^{\widetilde{X}} - \mathbb{E}\left[\varepsilon_t^{\widetilde{Y}}\varepsilon_t^{\widetilde{X}}\right]\right)}{\sqrt{\widehat{\mathbb{V}}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_t^{\widetilde{Y}}\varepsilon_t^{\widetilde{X}}\right]}} \;,$$

where $\widehat{\mathbb{V}}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_t^{\widetilde{Y}}\varepsilon_t^{\widetilde{X}}\right]$, as given in (5.4), is a consistent estimator of the long run variance of the sample mean of $\varepsilon_t^{\widetilde{Y}}\varepsilon_t^{\widetilde{X}}$. Below we point out that care has to be taken with respect to the choice of the weighting function and the truncation lag. Note that for ease of exposition we do not distinguish between $\varepsilon_t^{\widetilde{Y}}, \varepsilon_t^{\widetilde{X}}$ and $\widehat{\varepsilon}_t^{\widetilde{Y}}, \widehat{\varepsilon}_t^{\widetilde{X}}$. First, consider the observed series $\{\varepsilon_t^{\widetilde{Y}}\varepsilon_t^{\widetilde{X}}\}_{t=1}^{T}$. The circular block bootstrap (CBB) exploits T overlapping blocks of length B given by

$$B_t^{\widetilde{Y}\widetilde{X}} = (\varepsilon_t^{\widetilde{Y}}\varepsilon_t^{\widetilde{X}}, ..., \varepsilon_{t+B-1}^{\widetilde{Y}}\varepsilon_{t+B-1}^{\widetilde{X}}),\;\; t = 1, ..., T \;.$$

Observations $\varepsilon_t^{\widetilde{Y}}\varepsilon_t^{\widetilde{X}}$ for $r > T$ are wrapped around in a circle, i.e. $\varepsilon_{T+b}^{\widetilde{Y}}\varepsilon_{T+b}^{\widetilde{X}} = \varepsilon_b^{\widetilde{Y}}\varepsilon_b^{\widetilde{X}}$ for $1 \leq b \leq B$. Let the integer part of T/B, $[T/B]$, be the number of blocks K which are drawn randomly with replacement from the set of blocks $B_t^{\widetilde{Y}\widetilde{X}}$. Each of the drawn blocks, $k = 1, ..., K$, is denoted by $\xi_k^{\widetilde{Y}\widetilde{X}} = (\xi_{k,1}^{\widetilde{Y}\widetilde{X}}, ..., \xi_{k,B}^{\widetilde{Y}\widetilde{X}})$. Concatenating all $\xi_{k,b}^{\widetilde{Y}\widetilde{X}}$ in a vector defines the bootstrap sample $V_1^*, ..., V_L^*$. Thus the length of the bootstrap sample is $L = KB \leq T$, and the bootstrap sample average is

$$\bar{V}_L^* = \frac{1}{L}\sum_{t=1}^{L} V_t^* = \frac{1}{K}\sum_{k=1}^{K}\left(\frac{1}{b}\sum_{b=1}^{B}\xi_{k,b}^{\widetilde{Y}\widetilde{X}}\right) \;.$$

Under CBB sampling (which implies a measure P_{CBB1}^*) it can be shown that

$$\mathbb{E}_{CBB1}^*\left[\bar{V}_L^*\right] = \frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^{\widetilde{Y}}\varepsilon_t^{\widetilde{X}} \;,$$

where $\mathbb{E}_{CBB1}^*$ is the expectation under P_{CBB1}^*. Davison and Hall (1993) demonstrate that the block bootstrap for studentized statistics provides an improvement in asymptotic accuracy when applied properly. In particular, the naive studentization based on plugging in the bootstrapped sample into the formula for the long run variance estimator $\widehat{\mathbb{V}}\left[\bullet\right]$, i.e.

$$\frac{\sqrt{L}\left(\bar{V}_L^* - \mathbb{E}^*_{CBB1}\left[\bar{V}_L^*\right]\right)}{\sqrt{\widehat{\mathbb{V}}\left[\sqrt{L}\bar{V}_L^*\right]}} \, ,$$

yields a bootstrap approximation which maybe less accurate than the classical normal approximation. For ERPs of one–sided tests asymptotic refinements are obtained when studentization is accomplished by means of the variance of the rescaled bootstrap average under P^*_{CBB1} (Lahiri, 1991 and 1996, Götze and Künsch, 1996). This is given by

$$\mathbb{V}^*_{CBB1}\left[\sqrt{L}\bar{V}_L^*\right] = \frac{B}{T}\sum_{t=1}^{T}\left[\left(\frac{1}{B}\sum_{b=1}^{B}\varepsilon^{\widetilde{Y}}_{b+t-1}\varepsilon^{\widetilde{X}}_{b+t-1}\right) - \mathbb{E}^*_{CBB1}\left[\bar{V}_L^*\right]\right]^2 \, .$$

For two–sided tests the studentization by means of $\mathbb{V}^*_{CBB1}\left[\sqrt{L}\bar{V}_L^*\right]$ does not yield a superior performance over the normal approximation. Lahiri (1992) and Hall and Horowitz (1996) introduce correction factors to obtain refinements for both one– and two–sided symmetric tests. In particular, they define the bootstrap statistic as

$$ST^*_{T,CBB1} \;=\; \frac{\sqrt{L}\left(\bar{V}_L^* - \mathbb{E}^*_{CBB1}\left[\bar{V}_L^*\right]\right)}{\sqrt{\widehat{\mathbb{V}}\left[\sqrt{L}\bar{V}_L^*\right]}}\sqrt{\frac{\mathbb{V}_{CBB1}\left[\frac{1}{\sqrt{L}}\sum_{t=1}^{L}\varepsilon^{\widetilde{Y}}_t\varepsilon^{\widetilde{X}}_t\right]}{\mathbb{V}^*_{CBB1}\left[\sqrt{L}\bar{V}_L^*\right]}} \, , \quad (5.6)$$

where $\mathbb{V}_{CBB1}\left[\frac{1}{\sqrt{L}}\sum_{t=1}^{L}\varepsilon^{\widetilde{Y}}_t\varepsilon^{\widetilde{X}}_t\right]$ is the bootstrap analog of $\mathbb{V}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon^{\widetilde{Y}}_t\varepsilon^{\widetilde{X}}_t\right]$. Hence, the former is given by (5.4) with a weighting function $\omega(q,Q) = 1$ and truncation lag $Q = T-1$.

Next, a bootstrap procedure explicitly accounting for the independence of $\varepsilon^{\widetilde{Y}}_t$ and $\varepsilon^{\widetilde{X}}_t$ under the null hypothesis is outlined. We randomly resample with replacement K circular blocks of $\varepsilon^{\widetilde{X}}_t$

$$B^{\widetilde{X}}_t = (\varepsilon^{\widetilde{X}}_t, ..., \varepsilon^{\widetilde{X}}_{t+B-1}), \;\; t = 1, ..., L \, .$$

Concatenating the resampled blocks $\xi_k^{\widetilde{X}} = (\xi_{k,1}^{\widetilde{X}}, ..., \xi_{k,B}^{\widetilde{X}})$ in a vector, the bootstrap sample average is given by

$$\bar{V}_L^* = \frac{1}{K}\sum_{k=1}^{K}\left(\frac{1}{B}\sum_{b=1}^{B}\varepsilon_{B(k-1)+b}^{\widetilde{Y}}\xi_{k,b}^{\widetilde{X}}\right) .$$

This resampling approach implies

$$\begin{aligned}
\mathbb{E}_{CBB2}^*\left[\bar{V}_L^*\right] &= \overline{\varepsilon_L^{\widetilde{Y}}}\,\overline{\varepsilon_L^{\widetilde{X}}} \\
\mathbb{V}_{CBB2}^*\left[\sqrt{L}\bar{V}_L^*\right] &= \frac{L}{K^2}\sum_{k=1}^{K}\left(\frac{1}{L}\sum_{t=1}^{L}\left[\left(\frac{1}{B}\sum_{b=1}^{B}\varepsilon_{B(k-1)+b}^{\widetilde{Y}}\varepsilon_{b+t-1}^{\widetilde{X}}\right) - \overline{\varepsilon_L^{\widetilde{X}}}S_k^{\widetilde{Y}}\right]^2\right) ,
\end{aligned}$$

where $\overline{\varepsilon_L^{\widetilde{Y}}} = (1/L)\sum_{t=1}^{L}\varepsilon_t^{\widetilde{Y}}$, $\overline{\varepsilon_L^{\widetilde{X}}} = (1/L)\sum_{t=1}^{L}\varepsilon_t^{\widetilde{X}}$ and $S_k^{\widetilde{Y}} = (1/B)\sum_{b=1}^{B}\varepsilon_{B(k-1)+b}^{\widetilde{Y}}$. Accordingly, the bootstrap statistic is

$$ST_{T,CBB2}^* = \frac{\sqrt{L}\left(\bar{V}_L^* - \mathbb{E}_{CBB2}^*\left[\bar{V}_L^*\right]\right)}{\sqrt{\widehat{\mathbb{V}}\left[\sqrt{L}\bar{V}_L^*\right]}}\sqrt{\frac{\mathbb{V}_{CBB2}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_t^{\widetilde{Y}}\varepsilon_t^{\widetilde{X}}\right]}{\mathbb{V}_{CBB2}^*\left[\sqrt{L}\bar{V}_L^*\right]}} . \quad (5.7)$$

Note that for this bootstrap scheme the bootstrap analog to $\mathbb{V}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_t^{\widetilde{Y}}\varepsilon_t^{\widetilde{X}}\right]$ is given by

$$\begin{aligned}
\mathbb{V}_{CBB2}\left[\frac{1}{\sqrt{L}}\sum_{t=1}^{L}\varepsilon_t^{\widetilde{Y}}\varepsilon_t^{\widetilde{X}}\right] &= \widehat{\mathrm{Cov}}\left(\varepsilon_t^{\widetilde{Y}},\varepsilon_t^{\widetilde{Y}}\right)\widehat{\mathrm{Cov}}\left(\varepsilon_t^{\widetilde{X}},\varepsilon_t^{\widetilde{X}}\right) \\
&\quad +2\sum_{q=1}^{T-1}\widehat{\mathrm{Cov}}\left(\varepsilon_t^{\widetilde{Y}},\varepsilon_{t+q}^{\widetilde{Y}}\right)\widehat{\mathrm{Cov}}\left(\varepsilon_t^{\widetilde{X}},\varepsilon_{t+q}^{\widetilde{X}}\right) \\
&\quad -L\left(\overline{\varepsilon_L^{\widetilde{Y}}}\,\overline{\varepsilon_L^{\widetilde{X}}}\right)^2 ,
\end{aligned}$$

which explicitly accounts for the independence of $\varepsilon_t^{\widetilde{Y}}$ and $\varepsilon_t^{\widetilde{X}}$.

Implementation of the bootstrap approach

The choice of the kernel function $\omega(q, Q)$ is crucial for the bootstrap to provide better approximations than the classical normal approximation. For one–sided tests, Götze and Künsch (1996) show that for all kernels but the

Bartlett kernel improvements in ERP's can be obtained when $B = Q = O(T^{1/4})$. Moreover, they point out that their results also hold for other choices of $B \leq Q$. Hall and Horowitz (1996) and Andrews (2002) consider approximation errors for two–sided symmetric tests when the truncated kernel is used. Inoue and Shintani (2006) extend these results to show that for kernels such as the truncated, trapezoidal or Parzen kernel the bootstrap yields refinements for two–sided symmetric tests when $B = Q = O(T^{1/3})$. They also point out that their results hold if block sizes B are proportional to the truncation parameter Q. Otherwise the rate of the bootstrap approximation error is determined by the faster rate of B and Q. Hence, in our analysis we choose the truncated kernel and set $B = Q$. The choice of the truncated kernel does not guarantee that the variance estimator is positive. Yet, for positively persistent time series this problem is not as crucial as compared to data exhibiting negative serial correlation.

In order to implement a block bootstrap the block length parameter B has to be specified. Various approaches to determine optimal block sizes have been proposed. Hall, Horowitz and Jing (1995) derive optimal block sizes based on an asymptotic mean squared error criterion for bias/variance estimation or one– and two–sided distribution estimation. They show that optimal block lengths are $O(T^{1/3}), O(T^{1/4})$ and $O(T^{1/5})$, respectively. Zvingelis (2001) determines an asymptotically optimal block length minimizing the asymptotic ERP of one– and two–sided tests. He concludes that the optimal block sizes are $O(T^{1/4})$ and $O(T^{1/3})$, respectively. The constants of proportionality depend on, e.g., the autocovariance function of the DGP. Politis and White (2004) derive for the CBB scheme an explicit expression of the optimal block length B_{opt} for an AR(1)–process when interest focuses on bias/variance or distribution function estimation. They show that the optimal block size increases with the autocorrelation coefficient.

Relying on the result of Zvingelis (2001) an adaptation of the data based block length selection procedure of Hall, Horowitz and Jing (1995) targeting the empirical ERP criterion is straightforward. In particular, we compute the empirical ERP of the bootstrap test for all subsamples of length $\widetilde{T} < T$ and a grid of selected block lengths. Given the block length $B_{\widetilde{T}}$, for which the empirical ERP is closest to the nominal significance level, the estimated optimal block length for a sample of size T is then obtained from $\widehat{B}_{\text{opt}} = (T/\widetilde{T})^{1/3} B_{\widetilde{T}}$ for a two–sided test.

5.3 Simulation results

In order to shed light on the small sample properties of the test procedures presented above, we carry out a simulation study. We document the MC design and describe the size and size–adjusted power results, in turn.

5.3.1 Design

To simulate Bernoulli serially correlated random variables, we consider the stationary 2–dimensional VAR(1) process

$$\begin{pmatrix} Z_{1t} \\ Z_{2t} \end{pmatrix} = \begin{pmatrix} \phi_{11} & 0 \\ 0 & \phi_{22} \end{pmatrix} \begin{pmatrix} Z_{1t-1} \\ Z_{2t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix},$$

with $|\phi_{ii}| < 1,\ i = 1, 2$, and

$$\begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \sim NID\left[0, \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}\right].$$

Defining $\sigma_i^2 = (1 - \phi_{ii}^2),\ i = 1, 2$, $\sigma_{12} = \rho(1 - \phi_{11}\phi_{22})$ and $\phi_{11} = \phi_{22}$, the univariate processes, Z_{1t} and Z_{2t}, have unit variance $\mathbb{V}[Z_{it}] = 1,\ i = 1, 2$ and serial correlation $\text{Corr}(Z_{it}, Z_{it-j}) = \phi_{ii}^j, i = 1, 2$. Moreover, the contemporaneous cross covariance/correlation is given by $\text{Cov}(Z_{1t}, Z_{2t}) = \text{Corr}(Z_{1t}, Z_{2t}) = \rho$. Hence, cross sectional dependence and serial correlation are specified by selection of $|\rho| < 1$ and ϕ_{11}, respectively. Finally, let

$$\widetilde{X}_t = 1\left(Z_{1t} > 0\right) \text{ and } \widetilde{Y}_t = 1\left(Z_{2t} > 0\right) .$$

For cross sectional independence ($\rho = 0.0$) and medium and strong cross sectional dependence ($\rho = 0.5$ and $\rho = 0.8$) we simulate 5000 Monte Carlo replications of the process with no, medium and strong serial dependence ($\phi_{11} \in \{0.0, 0.5, 0.8\}$). We consider samples of size $T \in \{20, 50, 100, 500, 1000\}$. For each Monte Carlo replication we use 100 initializing presample values. The Fisher test is implemented as described in Henrikkson and Merton (1981). For the dynamic regression approach a maximum lag of 4 is allowed when choosing the lag order by means of the AIC. The truncation lag in the Newey–West estimation procedure is given by the integer part of $4(T/100)^{2/9}$. Finally, for the bootstrap approach we choose B, naively, as the nearest integer to $T^{1/3}$.

In our simulations a naive choice of the block size leads to rejection frequencies smaller than the nominal level of 5% for $T \geq 100$. Thus, for $T \geq 100$ we also choose the block length using the data based selection approach of Hall, Horowitz and Jing (1995). More precisely, for $T = 100$ the subsample length and the grid of block lengths are $\widetilde{T} = 30$ and $B_{\text{grid}} = \{3, 4, ..., 7\}$. For $T = 500, 1000$ we set $\widetilde{T} = 100$ and $B_{\text{grid}} = \{3, 4, ..., 15\}$.

5.3.2 Rejection frequencies under H_0

First, we describe the results for the case of cross sectional independence ($\rho = 0$) and no serial correlation ($\phi_{11} = 0$). The nominal significance level is 5%. Notably, results for other nominal levels are qualitatively identical. From the upper panel of Table 5.2 it can be inferred that the classical χ^2, the PT92, the PT08 and the bootstrap test perform very well and have empirical rejection frequencies very close to the nominal 5% level for all sample sizes considered. Fisher's test is seriously oversized in small and medium sample sizes but rejection rates converge to the nominal level as T increases. The small sample size distortion is possibly due to the fact that the simulation design does not guarantee fixed row and column marginals. The CovNW, StatNW and DynNW test procedures are also markedly oversized in small samples but approach empirical rejection frequencies close to 5% for increasing T. Accounting for serial correlation when there is none, does not pay off in small samples. Yet, it is not surprising that correctly assuming serial independence leads to an improved performance.

The medium panel of Table 5.2 displays empirical rejection frequencies under moderate serial correlation ($\phi_{11} = 0.5$). It reveals some size distortions for all but the bootstrap test. The χ^2 and the PT92 tests share similar rejection frequencies between 7% and 8% for all sample sizes considered. Fisher's test is seriously oversized in small samples with a rejection frequency of $\approx$ 8.5%. Among the three test procedures relying on the Newey–West variance estimator, the CovNW approach uncovers smallest size distortions for small samples. Empirical rejection frequencies of robust tests converge to the nominal level of 5% for all of these tests. Using the PT08 test H_0 is oversized in small samples but for medium and large samples the test has the correct rejection rate.

Introducing strong serial correlation ($\phi_{11} = 0.8$), the lower panel of Table 5.2 indicates that size distortions are severe for those tests which do not account for serial correlation. The χ^2, Fisher and PT92 tests are massively oversized for all sample sizes considered. Relative rejection frequencies appear to converge to $\approx$ 20%. The rejection frequency of the CovNW, StatNW and DynNW approaches is far too high in small samples but stabilizes $\approx$ 7%. The PT08 test is for small samples oversized (>10%) but as $T \geq 100$ it has appropriate rejection frequency. Among all tests considered the bootstrap approach performs best. It reveals (if any) small size distortions with empirical rejection frequencies close to the nominal level. For $T \geq 100$ the simulations for the alternative block length selection reveal a robust performance with rejection frequencies around 5%.

In summary, the bootstrap approach turns out to offer a remarkably robust performance. Its implied empirical size is close to the nominal level under serial independence and in the presence of serial correlation for all sample sizes considered. The PT08 approach is robust to serial dependence for medium and large sample sizes but reveals size distortions if $T < 100$.

5.3.3 Size–adjusted power

Table 5.2 documents the size–adjusted power results for selected scenarios of serial correlation ($\phi_{11} = 0.0, 0.5, 0.8$) when the cross correlation is $\rho = 0.5$ resp. $\rho = 0.8$. Some general conclusions can be drawn for all tests considered. The power decreases with increasing serial correlation. In the presence of serial dependence concordant observations of $\widetilde{X}_t$ and $\widetilde{Y}_t$ are more likely. Hence, it is more difficult to isolate the effects of cross sectional and serial dependence. Furthermore and most reasonably, size–adjusted power increases with increasing cross correlation.

While for sample sizes larger than 100 the power performance is very similar across the various test procedures, there are some differences for smaller sample sizes. For $T = 20, 50$ the χ^2, Fisher's, the PT92 and the PT08 test are somewhat more powerful than the CovNW, StatNW, DynNW and the bootstrap test. For example, while the former tests have a power close to 80% the latter reject slightly less frequently in less than 75% of the cases when $\rho = 0.8,\ \phi_{11} = 0.8$ and $T = 50$. The power of the bootstrap test is despite its non–parametric nature very appealing. It is close to the power of the remaining tests. For example, when $\rho = 0.8,\ \phi_{11} = 0.8$ and $T = 50$, the rejection rate of the bootstrap approach is 60%.

$\phi_{11} = 0.0$

T	χ^2	FE	PT 92	Cov NW	Stat NW	Dyn NW	PT 08	CBB2
20	0.055	**0.172**	**0.063**	**0.120**	**0.111**	**0.122**	**0.067**	**0.057**
50	0.054	**0.103**	0.055	**0.079**	**0.075**	**0.075**	0.049	0.055
100	0.052	**0.082**	0.053	**0.063**	**0.061**	**0.061**	0.049	0.050
500	0.056	**0.067**	0.056	**0.058**	**0.057**	**0.057**	0.054	0.049
1000	0.050	**0.059**	0.050	0.054	0.054	0.054	0.051	0.045

$\phi_{11} = 0.5$

T	χ^2	FE	PT 92	Cov NW	Stat NW	Dyn NW	PT 08	CBB2
20	**0.081**	**0.246**	**0.090**	**0.135**	**0.150**	**0.161**	**0.091**	**0.064**
50	**0.085**	**0.148**	**0.088**	**0.098**	**0.094**	**0.093**	**0.074**	0.048
100	**0.081**	**0.117**	**0.082**	**0.072**	**0.071**	**0.067**	**0.057**	0.049
500	**0.086**	**0.102**	**0.086**	**0.063**	**0.063**	0.054	0.053	0.052
1000	**0.087**	**0.097**	**0.088**	**0.062**	**0.061**	0.053	0.053	0.053

$\phi_{11} = 0.8$

T	χ^2	FE	PT 92	Cov NW	Stat NW	Dyn NW	PT 08	CBB2
20	**0.155**	**0.507**	**0.163**	**0.161**	**0.263**	**0.226**	**0.141**	**0.060**
50	**0.198**	**0.301**	**0.204**	**0.152**	**0.168**	**0.099**	**0.092**	**0.043**
100	**0.219**	**0.276**	**0.220**	**0.129**	**0.131**	**0.068**	0.056	0.051
500	**0.239**	**0.262**	**0.240**	**0.097**	**0.097**	0.051	0.049	0.053
1000	**0.242**	**0.257**	**0.242**	**0.094**	**0.094**	0.051	0.053	0.045

Table 5.1: Empirical rejection frequencies under H_0 ($\rho = 0.0$) and nominal significance level 5%. Different serial correlation parameters $\phi_{11} \in \{0.0, 0.5, 0.8\}$ and sample sizes $T \in \{20, 50, 100, 500, 1000\}$ are considered. χ^2 and FE denote the χ^2– and Fisher's exact test, respectively. Moreover, CovNW, StatNW, DynNW denote the covariance test and the tests based on the static and dynamic regression approaches using the Newey–West variance estimator. Corresponding naively chosen block sizes are $3, 4, 5, 8, 10$ when $\rho = 0.0$. When $\rho = 0.5, 0.8$ block sizes are 3 and 4 for $T = 20$ and $T = 50$. For $T \geq 100$ block sizes are determined by means of the approach proposed by Hall, Horowitz and Jing (1995). Bold figures are not within the 95% confidence interval given by $[\alpha \pm 2\sqrt{\alpha(1-\alpha)/5000}]$, where $\alpha = 0.05$.

$\rho = 0.5, \phi_{11} = 0.0$								
T	χ^2	FE	PT 92	Cov NW	Stat NW	Dyn NW	PT 08	CBB2
20	0.296	0.268	0.296	0.265	0.259	0.239	0.270	0.167
50	0.681	0.675	0.681	0.638	0.637	0.633	0.674	0.439
100	0.931	0.931	0.931	0.925	0.926	0.925	0.931	0.844
500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
1000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

$\rho = 0.8, \phi_{11} = 0.0$								
T	χ^2	FE	PT 92	Cov NW	Stat NW	Dyn NW	PT 08	CBB2
20	0.775	0.742	0.775	0.703	0.699	0.663	0.740	0.443
50	0.995	0.995	0.995	0.989	0.990	0.990	0.995	0.931
100	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999
500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
1000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

$\rho = 0.5, \phi_{11} = 0.5$								
T	χ^2	FE	PT 92	Cov NW	Stat NW	Dyn NW	PT 08	CBB2
20	0.233	-	0.236	0.192	0.177	0.176	0.192	0.134
50	0.558	0.552	0.560	0.523	0.520	0.504	0.535	0.400
100	0.857	0.858	0.857	0.848	0.851	0.835	0.853	0.708
500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
1000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

$\rho = 0.8, \phi_{11} = 0.5$								
T	χ^2	FE	PT 92	Cov NW	Stat NW	Dyn NW	PT 08	CBB2
20	0.671	-	0.673	0.537	0.540	0.524	0.604	0.317
50	0.978	0.977	0.978	0.968	0.969	0.963	0.975	0.891
100	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.995
500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
1000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

$\rho = 0.5, \phi_{11} = 0.8$								
T	χ^2	FE	PT 92	Cov NW	Stat NW	Dyn NW	PT 08	CBB2
20	0.158	-	0.155	0.127	-	-	0.144	0.081
50	0.303	-	0.316	0.282	0.282	0.247	0.301	0.233
100	0.570	0.577	0.572	0.545	0.550	0.519	0.568	0.429
500	0.996	0.998	0.996	0.997	0.997	0.998	0.999	0.991
1000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

$\rho = 0.8, \phi_{11} = 0.8$								
T	χ^2	FE	PT 92	Cov NW	Stat NW	Dyn NW	PT 08	CBB2
20	0.453	-	0.449	0.339	-	-	0.412	0.160
50	0.818	-	0.827	0.734	0.766	0.699	0.796	0.628
100	0.979	0.980	0.980	0.970	0.976	0.968	0.981	0.909
500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
1000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 5.2: Size–adjusted power. Different cross sectional correlation parameters $\rho \in \{0.5, 0.8\}$, serial correlation parameters $\phi_{11} \in \{0.0, 0.5, 0.8\}$ and sample sizes $T \in \{20, 50, 100, 500, 1000\}$ are considered. Note, no size–adjusted power is reported for Fisher's, the StatNW and the DynNW test in some cases. Due to the discreteness of the data it happens that at a nominal significance level of 0.1% the empirical size is 8% or larger. For further notes see Table .

5.4 Empirical applications

To illustrate the application of the test procedures and highlight the importance of accounting for serial correlation in applied work, we consider two empirical examples.

5.4.1 A large sample case

We apply the χ^2, Fisher's, the PT92, the PT08 and the bootstrap test to analyze directional forecasts for selected EURIBOR swap rates. Blaskowitz and Herwartz (2009b) consider $h = 1, 5, 10, 15$ days ahead ex–ante forecast for the EURIBOR swap term structure. Based on a battery of factor models they adaptively combine models to produce 1778 daily forecasts for the 2yr swap rate from April 19th, 2000, till mid February / beginning of March 2007 (depending on the forecast horizon h). We consider forecasts obtained from the most preferable Median strategy.

For comparison purposes some benchmark models are also considered. Namely, an AR(1) model and a variant of the term structure model proposed by Diebold and Li (2006) are fitted by means of rolling windows of 42 daily observations (see Blaskowitz and Herwartz, 2009b) for details. The benchmark strategies are denoted by AR resp. DL.

Table 5.3 illustrates the extent of serial correlation present in realized and forecasted directions (up–/downward movements) of the 2yr EURIBOR swap rate. Apart from the realized directions of the 2yr swap rate for one day ahead forecasts, all remaining series are highly and significantly serially correlated. Moreover, the higher the horizon, the stronger the serial dependence. For forecast horizons $h = 5, 10, 15$ first order correlations for outcomes in directions are high, about 0.6, 0.7, 0.8, respectively. Correlations decrease to less than 0.1 at lag 20. For forecasted directions, first order correlations are between 0.75 and 0.93 for $h = 5, 10, 15$ and remain high (above

$\approx$ 0.4) at all lags considered. This evidence suggests that commonly applied procedures to test for the value of directional forecasts in the sense of Merton (1981) are inadequate for all but the one day ahead forecasts of the 2yr swap rate.

	$h=1$	$h=5$	$h=10$	$h=15$	$h=1$	$h=5$	$h=10$	$h=15$
	Realized Directions				MedStrat			
1	-0.040	**0.605**	**0.749**	**0.802**	**0.365**	**0.782**	**0.896**	**0.931**
5	0.015	**0.070**	**0.393**	**0.524**	**0.323**	**0.741**	**0.841**	**0.868**
10	-0.009	**0.056**	**0.063**	**0.272**	**0.270**	**0.699**	**0.805**	**0.828**
15	0.015	0.044	**0.050**	**0.086**	**0.239**	**0.657**	**0.754**	**0.775**
20	-0.012	**0.060**	**0.083**	**0.076**	**0.208**	**0.609**	**0.700**	**0.721**
	AR				DL			
1	**0.546**	**0.815**	**0.851**	**0.864**	**0.097**	**0.676**	**0.832**	**0.863**
5	**0.503**	**0.693**	**0.714**	**0.720**	**0.097**	**0.529**	**0.710**	**0.726**
10	**0.410**	**0.603**	**0.616**	**0.625**	**0.098**	**0.409**	**0.596**	**0.629**
15	**0.407**	**0.560**	**0.564**	**0.573**	0.010	**0.272**	**0.523**	**0.546**
20	**0.368**	**0.489**	**0.499**	**0.501**	**0.068**	**0.211**	**0.418**	**0.458**

Table 5.3: Serial correlations of realized and forecasted directions of EURIBOR swap rates. Bold numbers are significant at a 5% significance level. Critical values are $\pm 2/\sqrt{1778} \approx \pm 0.047$.

To analyze the value of EURIBOR swap rate forecasts, Table 5.4 shows empirical estimates of covariances and HM statistics for various forecast exercises. It can be seen that the forecasts of all models have positive value. Moreover, Table 5.4 provides the results for testing H_0 : Cov$(\widetilde{Y}_t, \widetilde{X}_t) = 0$ against H_1 : Cov$(\widetilde{Y}_t, \widetilde{X}_t) \neq 0$ for various significance levels $\alpha \leq 0.20$. Using traditional test procedures the null of no value is rejected at a 1% significance level for all forecast exercises except for $h = 1$ AR and DL forecasts. For the latter, H_0 is rejected at nominal levels between 11% and 15%. Conclusions drawn from the serial correlation robust test procedures are different in some cases. The discrepancy becomes more apparent the larger the serial correlation. Test decisions for $h = 1$ generally agree for all procedures. Yet, striking differences in significance are obtained for the 5 day ahead forecasts for the DL model as well as for the 10 day ahead forecasts for the Median strategy and the DL model. Note, for the bootstrap test we used a the data based block length selection method of Hall, Horowitz and Jing (1995) as described in Section 5.2.2. Using a naive block choice $B = [1778^{1/3}] = 12$ does not change the conclusions.

	MedStrat				AR				DL			
	$h=1$	$h=5$	$h=10$	$h=15$	$h=1$	$h=5$	$h=10$	$h=15$	$h=1$	$h=5$	$h=10$	$h=15$
Cov	0.023	0.035	0.036	0.042	0.009	0.029	0.038	0.045	0.002	0.013	0.029	0.037
HM	1.091	1.139	1.143	1.170	1.036	1.118	1.152	1.182	1.010	1.052	1.115	1.148
χ^2	1%	1%	1%	1%	14%	1%	1%	1%	15%	1%	1%	1%
FE	1%	1%	1%	1%	13%	1%	1%	1%	11%	1%	1%	1%
PT92	1%	1%	1%	1%	14%	1%	1%	1%	15%	1%	1%	1%
PT08	1%	2%	NR	1%	14%	1%	5%	13%	15%	NR	NR	2%
CBB2	1%	1%	6%	6%	NR	3%	5%	4%	NR	10%	11%	7%

Table 5.4: Covariances, HM statistics and test results for various significance levels $\alpha \leq 0.2$ are provided. NR indicates that H_0 cannot be rejected at the 20% significance level.

5.4.2 A small sample case

Moreover, we investigate the stock return predictions analyzed in Herwartz and Morales (2008). Based on a panel asset pricing model they determine $h = 3, 6$ month ahead forecast of returns of Germany's DAX30, Italy's MIB30 and Norway's OBX25. We focus on the most recent 50 forecasts which cover the period 06/2000 to 01/2005 (depending on the forecast horizon). Positive/negative realized resp. forecasted returns are considered as up–/downward movements.

As can be seen from Table 5.5 the covariance and HM statistic for the 6 months ahead forecasts of Norway's OBX25 returns are quite large, around

0.15 resp. 1.6. Even if serial correlations are significant at least up to lags 4 and 6 for realized and forecasted directions all test procedures reject at low significance levels. As the test statistic is high, any test should reject the null and the impact of serial correlation should be negligible.

For Germany's DAX30 both the 3 and 6 month ahead forecasts have a rather low value. Covariances and HM statistics are about 0.05 resp. 1.2. Moreover, there is no marked serial correlation beyond lag 4. Thus, similar decisions are inferred from all tests. The null hypothesis is not rejected at conventional significance levels.

The 3 and 6 month ahead forecasts of the MIB30 and the 3 month ahead forecasts of the OBX25 have a moderate value, with covariances between 0.08 and 0.09 and HM statistics between 1.3 and 1.4. Serial correlations are significant up to lags 4 and 6. In such a situation accounting for serial correlation is important when testing for the value of directional forecasts. While all the classical tests reject the null hypothesis, the serial correlation robust procedures yield a downgrading of the forecast's economic value. The bootstrap test is carried out using $B = 4$. Alternative choices of $B = 2$ and $B = 6$ provide the same results.

Serial correlations

$h=3$							$h=6$					
realized directions			forecasted directions				realized directions			forecasted directions		
Ger	Ita	Nor	Ger	Ita	Nor	lag	Ger	Ital	Nor	Ger	Ita	Nor
0.533	**0.619**	**0.497**	**0.394**	**0.628**	**0.661**	1	**0.760**	**0.650**	**0.661**	0.071	**0.694**	**0.740**
0.151	0.199	**0.314**	0.251	**0.550**	**0.529**	4	0.236	**0.466**	**0.325**	-0.019	**0.458**	**0.560**
0.039	0.117	0.075	0.259	**0.453**	**0.371**	6	0.214	**0.389**	0.204	0.113	0.261	**0.360**
0.063	0.135	0.053	0.065	0.258	0.270	9	0.013	0.206	-0.015	0.022	0.143	0.060
-0.202	-0.045	-0.049	0.125	0.063	0.089	12	-0.187	0.104	0.045	-0.131	-0.095	-0.040

Test statistics and test results

$h=3$				$h=6$			
	Ger	Ita	Nor		Ger	Ita	Nor
Cov	0.047	0.082	0.078	Cov	0.055	0.090	0.150
HM	1.200	1.350	1.312	HM	1.254	1.403	1.600
χ^2	17%	2%	3%	χ^2	9%	1%	1%
FE	10%	1%	2%	FE	4%	1%	1%
PT92	17%	2%	3%	PT92	8%	1%	1%
PT08	NR	NR	NR	PT08	9%	NR	1%
CBB2	NR	9%	NR	CBB2	12%	16%	4%

Table 5.5: Upper panel shows serial correlations of realized and forecasted directions of European stock market returns. Bold numbers are significant at a 5% significance level. Critical values are $\pm 2/\sqrt{50} \approx \pm 0.283$. The lower panel provides covariances, HM statistics and test results for significance levels $\alpha \leq 0.2$. NR indicates that H_0 cannot be rejected at the 20% significance level.

5.5 Conclusions

Commonly applied procedures to test for the value of directional forecasts suffer from marked size distortions in the presence of serial correlation. As this issue is highly relevant for economic applications, we summarized existing procedures and proposed a simple statistic for which we implement a bootstrap approach. By means of a Monte Carlo simulation we find that the bootstrap test reveals only minor size distortions in small samples as opposed to traditional procedures and retains appealing power. For medium and large sample sizes, the dynamically augmented maximum correlation test proposed in Pesaran and Timmerman (2008) represents an alternative approach with correct size and promising power. In two empirical examples we illustrate the relevance and application of serial correlation robust test procedures for small as well as for large sample sizes.

A particular merit of the investigated test statistic is that it allows for both one–sided and two–sided alternative hypotheses. Moreover, since its square is equal to a Wald statistic under the null hypothesis the test procedure can be easily extended to general $r \times c$ contingency tables. In this framework, the generalized test of market timing as proposed in Pesaran and Timmermann (1992) can be dealt with readily. In principle, the remaining test procedures summarized in this paper can be subjected to resampling. Yet, for the reasons outlined above we focuse on the covariance test statistic and leave it for further research to develop bootstrap algorithms for the other tests.

An appropriate choice of the block length is important for a proper bootstrap test. Our simulations reveal that a naive choice based on the fact that the optimal block length is $O(T^{1/3})$ results in a slightly undersized bootstrap scheme for large sample sizes. Adapting the data based block length selection procedure of Hall, Horowitz and Jing (1995) yields empirical rejection frequencies close to the nominal size. Additional improvements can be expected by a block size selection procedure that accounts for size and power considerations. We regard the latter issue to merit further reflection.

Chapter 6

A Note on the Economic Evaluation of Directional Forecasts

It is commonly accepted that information is helpful if it can be exploited to improve the decision making process. Frequently, the available information set is used to produce forecasts. Hence, information is useful if the forecasts help to make decisions that reduce losses/costs or increase gains/utility (see also Armstrong and Collopy 1992, Granger and Pesaran 2000, Pesaran and Skouras 2002).

Diebold and Mariano (1995) and Granger and Pesaran (2000), among others, point out that in order to evaluate the usefulness of forecasts, measuring the realized economic value is more suitable than assessing a realized 'statistical value' by means of criteria such as mean squared forecast errors. Other loss functions based on forecast errors exist and find some support when evaluating the accuracy of various forecast methods across many series. These are, for example, the geometric mean of the relative absolute error, the mean absolute scaled error or the log mean squared error ratio (e.g. Thompson 1990, Armstrong and Collopy 1992, Hyndman and Koehler 2006). However, such forecast criteria generally suffer from lack of economic interpretability. Moreover, criteria based on forecast errors are not suitable whenever a forecast method is akin to produce unreasonable forecasts which are far away from the realizations of the variable of interest. Robustness to outliers is particularly relevant in applied research when numerous (econometric) forecast procedures have to be compared (e.g. Armstrong and Collopy 1992, Makridakis 1993).

A forecast evaluation criterion should be related to decision making as put forward, for example, by Armstrong and Collopy (1992), Granger and Pesaran (2000) and Pesaran and Skouras (2002). In economics, decisions are often based on forecasts of directional up– or downward movements of the variable of interest. This note focuses on some aspects of the economic evaluation of directional forecasts (DFs). We argue that commonly used approaches to evaluate DFs relying on signs are mostly incomplete measures of the economic value. We point out that DFs can, nevertheless, provide a convenient framework to assess the economic forecast value. This is accomplished when loss functions (or success measures) are properly formulated to account for realized signs and realized magnitudes of directional movements. An easily interpretable success measure is advocated which is an important issue when analyzing forecast performances (e.g. Ahlburg 1992). In addition, such an evaluation framework is simple to implement and robust to outlying forecasts.

In the next section we review the evaluation of DFs when considering directional signs only. In Section 6.2 we present the general framework to assess the economic value of DFs and provide an illustration in Section 6.3. Section 6.4 contains some concluding remarks.

6.1 Directional forecasts: signs only

In economic applications the forecast user is often interested in directional (up–/downward) movements of the variable of interest denoted by Y_t henceforth. A prominent macroeconomic example is given by a monetary authority who raises interest rates if inflation is forecasted to rise. In finance, the speculator buys the stock if the stock price is expected to rise. Various other examples exist.

To formalize the forecast evaluation procedure we let h denote the forecast horizon. The forecast for Y_{t+h} using the information available in t is given by X_t^h. Using the indicator function $I(\bullet)$, the realized and forecasted directions are given by $\widetilde{Y}_t = I(Y_{t+h} - Y_t > 0)$ and $\widetilde{X}_t = I(X_t^h - Y_t > 0)$. Directions can also be determined using a non–zero threshold. In principle, DFs need not necessarily be derived from forecasted and current levels X_t^h and Y_t. Any other forecast method producing $\widetilde{X}_t$ is allowed. For example, DFs can be based on probability forecasts of changes in Y_t. In–/correct DFs are defined by $\widetilde{Z}_t = I(\widetilde{X}_t = \widetilde{Y}_t)$.

A commonly used loss function for DFs is given by

$$L_t^{\mathrm{DA}} = \begin{cases} a & \text{if } \widetilde{Z}_t = 1 \\ b & \text{if } \widetilde{Z}_t = 0, \end{cases}$$

where $(a, b) \neq (0, 0)$. In this framework, a correct DF has a 'value' of a and an incorrectly forecasted direction a 'value' of b. Frequently $(a, b) = (1, -1)$ or $(a, b) = (1, 0)$. Hence, it makes more sense to call L_t^{DA} a success function. Leitch and Tanner (1995), Greer (2005), Blaskowitz and Herwartz (2009b) employ $(a,\ b) = (1, -1)$. Other authors use $(a, b) = (1, 0)$, e.g. Swanson and White (1995, 1997a,b), Gradojevic and Yang (2006) and Diebold (2007). Note that $\mathbb{E}[L_t^{\mathrm{DA}}] = (a - b)\mathbb{P}[\widetilde{Z}_t = 1] + b$. Consequently, using this loss function amounts to considering the number of correct, respectively, incorrect DFs. While L_t^{DA} is robust to outlying forecasts X_t^h, it ignores the size of realized directional movements. Therefore, it does not measure the economic value to the forecast user whenever correctly predicted small respectively large realized directional changes have different benefits/losses to the forecast user.

Merton's (1981) theory implies that DFs have no value if the forecast user's subjective probability function for $\widetilde{Y}_t$ given the forecast user's information set does not change when the user obtains a forecast $\widetilde{X}_t$. Since, the forecaster's information set is useless to the user, the latter would not be willing to pay for such a forecast. Within the framework of Merton (1981) it holds that DFs have no value if and only if

$$HM = \mathbb{P}[\widetilde{X}_t = 1|\widetilde{Y}_t = 1] + \mathbb{P}[\widetilde{X}_t = 0|\widetilde{Y}_t = 0] = 1 \ .$$

of Henriksson and Merton (1981).) Moreover, DFs have positive value if and only if

$$HM > 1 \ .$$

In this case, the subjective probability function of the forecast user changes such that he considers up–/downward movements more likely when the forecast is an up–/downward movement. For an application of the HM statistic, see Schnader and Stekler (1990), Mills and Pepper (1999) and Ashiya (2006), among others. Merton's framework is not equivalent to the loss functional approach described earlier as pointed out, for instance, in Merton (1981) and Blaskowitz and Herwartz (2008). Notably, it is easily verified that

$$HM - 1 = \frac{\text{Cov}\left(\widetilde{X}_t, \widetilde{Y}_t\right)}{\mathbb{V}\left[\widetilde{Y}_t\right]} \ ,$$

where $\text{Cov}\left(\widetilde{X}_t, \widetilde{Y}_t\right)$ and $\mathbb{V}\left[\widetilde{Y}_t\right]$ denote the covariance between realized and forecasted directions respectively the variance of realized directions. Hence, DFs have no value if and only if $\text{Cov}\left(\widetilde{X}_t, \widetilde{Y}_t\right) = 0$. Equivalently, $\widetilde{X}_t$ and $\widetilde{Y}_t$ are independent in this case. DFs have positive value if and only if

$\text{Cov}\left(\widetilde{X}_t, \widetilde{Y}_t\right) > 0,$ i.e. $\widetilde{X}_t$ and $\widetilde{Y}_t$ are positively correlated. A prominent naive benchmark strategy for DFs is given by forecasting always an upward (or downward) movement. Such naive DFs have no value in the sense of Merton. Hence, HM measures the additional value of a DF when compared to naive predictions. Consequently, the HM measure is not only robust to outlying forecasts, it also has a sensible and intuitive economic interpretation. Yet, it considers only the sign and neglects the magnitude of changes in the movement of Y_t.

The nonparametric test of predictive performance presented in Pesaran and Timmermann (1992) tests the null hypothesis that predicted and realized signs $\widetilde{X}_t$ and $\widetilde{Y}_t$ are independent. The latter hypothesis is equivalent to the null hypothesis implied by the Merton framework. Applications include, for instance, Pesaran and Timmermann (1995), Pons (2001), Schneider and Spitzer (2005).

While the DA criterion (as well as criteria based on forecast errors) does not measure the economic value of (directional) forecasts, the HM and the Pesaran and Timmermann (1992) approaches provide an 'all–purpose' measure for an economic value of DFs in a rather restrictive sense. A more appropriate context–specific assessment of the economic value of DFs is explained in the next Section.

6.2 The economic value of directional forecasts

To formalize the economic evaluation of DFs we define

$$L_t^{\text{DV}} = \begin{cases} H_t^{UU} = H^{UU}(Y_{t+h}, Y_t) & \text{if correct upward prediction} \\ H_t^{DD} = H^{DD}(Y_{t+h}, Y_t) & \text{if correct downward prediction} \\ H_t^{UD} = H^{UD}(Y_{t+h}, Y_t) & \text{if incorrect upward prediction} \\ H_t^{DU} = H^{DU}(Y_{t+h}, Y_t) & \text{if incorrect downward prediction.} \end{cases} \tag{6.1}$$

In (6.1) H_t^{UU} resp. H_t^{DD} denote the benefit/gain/value to the forecast user when he believes in a directional up– resp. downward forecast and an up–

resp. downward movement realizes. Similarly, H_t^{UD} resp. H_t^{DU} denote the cost/loss/value to the forecast user in case of an incorrect directional prediction. As L_t^{DV} depends only on the DF $\widetilde{X}_t$ and not on the exact value of X_t^h it is robust to forecasts which are far apart from Y_{t+h}. Testing hypothesis about $\mathbb{E}\left[L_t^{\mathrm{DV}}\right]$ is readily accomplished within the framework of Diebold and Mariano (1995), as long as L_t^{DV} is stationary. Moreover, testing equality in prediction accuracy of alternative methods, such as naive DFs, can be implemented easily. Note that for the special case $H_t^{UU} = H_t^{DD} = a$ and $H_t^{UD} = H_t^{DU} = b$ it holds $L_t^{\mathrm{DV}} = L_t^{\mathrm{DA}}$.

We illustrate the flexibility of L_t^{DV} by means of some examples. Let $H_t^{UU} = H_t^{DD} = |Y_{t+h} - Y_t|$ and $H_t^{UD} = H_t^{DU} = -|Y_{t+h} - Y_t|$. Then L_t^{DV} captures the ability to forecast the sign and the magnitude of realized changes. See Blaskowitz and Herwartz (2009b) for an application. Such a property is particularly interesting when Y_t is a stock price and the DFs are used to make buy/sell decisions. In this case L_t^{DV} is the realized cash flow from the position set up based on the DFs. In the framework of Skouras (2001) a risk–neutral artificial technical analyst chooses from a set of competitive directional forecasting methods the one which maximizes expected utility. The latter is accomplished by maximizing expected cash flows. Note also that numerous loss functions are scaled in arbitrary units. The scale of L_t^{DV} is in the units of the forecast variable allowing an immediate interpretation of the forecast value.

An obvious modification measuring realized returns derived from DFs is given by

$$L_t^{\mathrm{DV}} = \begin{cases} |(Y_{t+h} - Y_t)/Y_t| & \text{if } \widetilde{Z}_t = 1 \\ -|(Y_{t+h} - Y_t)/Y_t| & \text{if } \widetilde{Z}_t = 0\,, \end{cases}$$

where we assume that $Y_t > 0$. See Gencay (1998) or Anatolyev and Gerko (2005) for an application. Note that in this case L_t^{DV} is unit–free which is particularly useful when comparing forecast methods for various series with different scale (Armstrong and Collopy 1992). The excess profitability test of Anatolyev and Gerko (2005) can be viewed as a test of the null hypothesis that $\mathbb{E}\left[L_t^{\text{DV}}\right]$ is greater than the expected profits from an artificial benchmark strategy. While the buy/sell signal frequencies of the benchmark and the trading strategy under investigation are equal, the artificial strategy generates buy/sell signals randomly.

More general functions of Y_{t+h} and Y_t can be accommodated within this framework. For example, let Y_t denote a fair value swap rate at time t. Furthermore, let $RSW(Y_t, K, \tau)$ be the value of a receiver swap agreement with fixed rate K and termination date τ when the current fair value swap rate is Y_t. Similarly $PSV(Y_t, K, \tau)$ denotes the value of a payer swap. For simplicity, we neglect the dependence of the swap value on other variables (see e.g. Miron and Swannell 1991). The current value of a payer swap with fixed rate $K = Y_t$ is zero, $PSV(Y_t, Y_t, \tau - h) = 0$. If $K < Y_t$ then $PSV(Y_t, K, \tau) > 0$. Thus, in swap trading, a speculator enters a payer swap agreement if he expects the fair value swap rate to rise. On the other hand, if the fair value swap rate is expected to fall a receiver swap agreement is entered. Consequently, a success measure is given by

$$L_t^{\text{DV}} = \begin{cases} PSV(Y_{t+h}, Y_t, \tau - h) & \text{if } PSV(X_t^h, Y_t, \tau - h) > 0 \\ & \text{and } PSV(Y_{t+h}, Y_t, \tau - h) > 0 \\ RSV(Y_{t+h}, Y_t, \tau - h) & \text{if } RSV(X_t^h, Y_t, \tau - h) > 0 \\ & \text{and } RSV(Y_{t+h}, Y_t, \tau - h) > 0 \\ PSV(Y_{t+h}, Y_t, \tau - h) & \text{if } PSV(X_t^h, Y_t, \tau - h) > 0 \\ & \text{and } PSV(Y_{t+h}, Y_t, \tau - h) < 0 \\ RSV(Y_{t+h}, Y_t, \tau - h) & \text{if } RSV(X_t^h, Y_t, \tau - h) > 0 \\ & \text{and } RSV(Y_{t+h}, Y_t, \tau - h) < 0 \, . \end{cases}$$

Notably, $PSV(X_t^h, Y_t, \tau - h)$ and $RSV(X_t^h, Y_t, \tau - h)$ can be any signal that indicates rising or falling values of swap agreements. Moreover, $PSV(Y_{t+h}, Y_t, \tau - h)$ and $RSV(Y_{t+h}, Y_t, \tau - h)$ can be theoretical or observed market prices. See also Blaskowitz and Herwartz (2009a) for an application.

The measure defined in (6.1) can deal with numerous other specifications. For example, instead of assessing the value of directional swap rate forecasts any financial derivative such as stock options can easily be analyzed similarly. H_t^{ij} could also be determined by a utility function such as the negative exponential utility function as in West, Edison and Cho (1993). Furthermore, the framework of DF evaluation is not restricted to financial applications. Business applications include decisions of a company whether to increase production by, say, 3% or not, conditional on forecasts about changes in economy wide output levels such as BIP. The DF value could be determined by incremental sales or revenues. In macroeconomics, monetary authorities who have to decide whether to increase or decrease interest rates by 25 basis points given DFs for inflation could use a social welfare/cost function to measure the economic value of DFs. The DF measure (6.1) also accommodates situations in which directional costs/benefits are asymmetric. For example, consider a strategy to short put options until maturity when the market is predicted to go up or to invest in the cash market when it is expected to go down. In this case, an incorrect upward prediction might be more expensive than an incorrect downward movement, $H_t^{UD} < H_t^{DU}$.

6.3 Empirical illustration

To demonstrate the issues discussed above we provide an empirical example. We consider $h = 5$ day ahead forecasts for the 2yr EURIBOR swap rate using the principal components analysis (PCA) based approach analyzed in Blaskowitz and Herwartz (2009a). They estimate K principal components (or factors) from τ observations for the EURIBOR swap term structure defined by the 3 and 6 month EURIBOR rates, and the 1yr (year), 2yr, 3yr, 5yr, 7yr, 10yr 12yr, 15yr swap rates. Factors are forecasted using a vector autoregressive (VAR) model with p lags. Note that Blaskowitz and Herwartz (2009a) consider 100 different models by combining five estimation windows $\tau \in \{42, 63, 126, 189, 252\}$, five factor choices $K \in \{1, 2, 3, 4, 5\}$ and four lag orders $p \in \{0, 1, 2, 3\}$. For illustrative purposes we focus on the model specification defined by an estimation window of $\tau = 252$ observations, $K = 4$ factors and $p = 1$ autoregressive lag (we abbreviate the model by $252/4/1$). Altogether 80 forecasts are produced for the period September 3, 2001 to December 21, 2001.

Results are reported for the mean squared forecast error (MSFE, multiplied by 10^6) and for the mean DF value (MDV, multiplied by 100). The latter is defined by the cash flows derived from a swap trading strategy which enters 2yr payer (receiver) swap agreements if an increase (decrease) in the 2yr swap rate is forecasted. Five days later the cash flows are computed using the comparison swap valuation technique (Miron and Swannell 1991) and the realized 2yr swap rate. Note that to determine cash flows the forecasted swap rate is not needed making the evaluation measure insensitive to outlying forecasts. Its economic interpretability is obvious as opposed to the economic content of the MSFE criterion. The factor model specification $252/4/1$ implies a MSFE of 3.80 and a MDV of 3.27. From Table 6.1, left panel, it can be seen that it is 65th and 47th best model in terms of MSFE and MDV. Inspection of the time series plot of forecasts and actuals for the above model, given in Figure 6.1, reveals that the 14th forecast is somewhat unreasonably far away from, both, forecasts and actuals. In order to separate the impact of the outlying forecast from the comparison, we delete it from all models. Then, the model $252/4/1$ has a MSFE of 2.59 and a MDV of 3.01, see the right panel of Table 6.1. With respect to the MSFE criterion it is now 4th best model and remains 47th in terms of MDV. Removing the outlier leads to a 30% reduction in MSFE and a substantial improvement in the model ranking relative to the remaining

99 specifications, while the MDV comparison remains unaffected.

	with outlier		outlier removed	
	MSFE$*10^6$	MDV$*100$	MSFE$*10^6$	MDV$*100$
1st	2.53	8.77	2.56	8.63
10th	2.61	8.23	2.63	8.07
30th	2.75	4.94	2.78	4.75
40th	2.84	4.06	2.84	3.85
60th	3.34	2.51	3.36	2.28
70th	3.88	2.04	3.93	1.81

Table 6.1: Quantiles for MSFE$*10^6$ and MDV$*100$ out–of–sample forecast performance of 5 day–ahead forecasts of 2yr swap rates for the period of September 3, 2001 to December 21, 2001 of 100 PCA–VAR models.

Deleting outliers from the forecast evaluation is not necessarily the best choice for several reasons. First, it is a delicate matter to define outliers. It might be that large observed forecast errors belong to the tail of the forecast error distribution in which case a removal boils down to truncating the latter distribution. Second, deleting predictions from all models leads to a loss of information. This is particularly relevant when relatively few forecasts are available as in numerous macroeconomic applications. Next, taking the evolution of actuals and forecasts into account an applied analyst would doubt the exact value of the outlier(s) but he would probably admit that a further directional movement is not unreasonable. For example, in the case of the 14th forecast as shown in Figure 6.1 an analyst might believe in a further downward movement. The directional prediction content of the 14th forecast may still be of value. Moreover, visual inspection of the corresponding plots for the all 100 models reveals that there are more outliers from time to time. Accounting for the widespread use of PCA–VAR approaches, especially in term structure modelling, it would be inappropriate to discuss the suitability of the forecast method itself. Given the large number of models, a manual outlier removal is time consuming and subjective. Applying an 'insanity filter' based on ad–hoc rules to define and delete outliers reduces the workload but still remains subjective, see, for instance, Elliot and Timmermann (2008). All in all, in the presence of outliers a forecast comparison in terms of MSFE is distorted, whether outliers are deleted or not. The robust DF measure represents a meaningful tool for forecast evaluations as it is readily interpretably in economic terms and circumvents all the above problems.

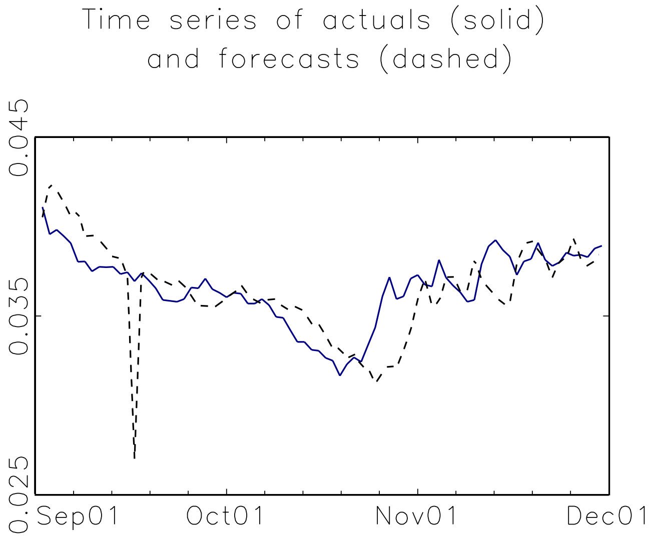

Figure 6.1: Time series of actuals and out–of–sample 5 day–ahead forecasts for the 2yr swap rate for the period of September 3, 2001 to December 21, 2001 of the model specification $252/4/1$.

6.4 Conclusion

We discuss a general approach to evaluate (directional) forecasts which is simple to implement, robust to outlying or unreasonable forecasts and which provides an economically interpretable loss/success functional framework. As such, the measure of directional forecast value presented here, is a readily available alternative to the commonly used squared error loss criterion.

Christoffersen and Diebold (1996, 1997), Granger and Pesaran (2000) and Skouras (2007), among others, argue in favor of an integrated approach to allow for general loss functions in modelling, estimation, model selection, prediction and forecast evaluation. By focusing only on the evaluation of forecasts, we account for the fact that frequently only the predictions are available without knowing the method used to produce the latter (e.g. survey/analysts/judgemental forecasts). The underlying rationale is that even if such forecasts are not produced optimally within the above integrated framework, they may contain valuable information with respect to a distinct loss function.

Armstrong and Collopy (1992) argue that a forecast evaluation criterion should be related to decision making. Granger and Pesaran (2000) and Pesaran and Skouras (2002) put forward a decision–theoretic approach to forecast evaluation. It requires the specification of the decision environment of individual agents and distributional assumptions about the underlying DGP. Furthermore, in practical applications for most decision problems complex numerical optimizations are necessary. Pesaran and Skouras (2002) note that: "A widespread application of the decision–based approach in economics is likely to take decades rather than years before becoming a reality."

The framework we investigate is related to decision making as it provides the economic value of DFs in a very simple decision problem (buy/sell stocks, increase interest rates or not, etc.). Even if it does not encompass all possible decision problems, it can be seen as a compromise between an individualized decision–theoretic framework and a generalized loss functional approach in a decision making environment.

Bibliography

Agresti, A., 'A survey of exact inference for contingency tables', *Statistical Science*, **7**, 131–153 (1992).

Ahlburg, D. A., 'A commentary on error measures', *International Journal of Forecasting*, **8**, 99–100 (1992).

Altham, S., 'Serial dependence in contingency tables', *Journal of the Royal Statistical Society* **Ser. B, 45**, 100–106 (1983).

Anatolyev, S., 'A unifying view of some nonparametric predictability tests' *Working Paper, New Economic School, Moscow, Russia* (2006).

Anatolyev, S. and A. Gerko, 'A trading approach to testing for predictability' *Journal of Business and Economic Statistics*, **23**, 455–461 (2005).

Andrews, D.W.K., 'Higher–order improvements of a computationally attractive k–step bootstrap for extremum estimators' *Econometrica*, **70**, 119–162 (2002).

Ang, A. and M. Piazzesi, 'A no–arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables', *Journal of Monetary Economics*, **50**, 745–787 (2003).

Armstrong, J. S. and F. Collopy, 'Error measures for generalizing about forecasting methods', *International Journal of Forecasting*, **8**, 69–80 (1992).

Artis, M.J., 'How accurate are the IMF's short–term forecasts? Another examination of the world economic outlook', *International Monetary Fund, Working Paper No. 96/89* (1996).

Ash, J.C.K., D.J. Smith and S.M., Heravi, 'Are OECD forecasts rational and useful?: A directional analysis', *International Journal of Forecasting*, **14**, 381–391 (1998).

Audrino, F., G. Barone–Adesi and A. Mira, 'The stability of factor models of interest rates', *Journal of Financial Econometrics*, **3**, 422–441 (2005).

Ashiya, M., 'The directional accuracy of 15–months–ahead forecasts made by the IMF', *Applied Economics Letters*, **10**, 331–333 (2003).

Ashiya, M., 'Are 16–month–ahead forecasts useful?: A directional analysis of japanese GDP forecasts', *Journal of Forecasting*, **25**, 201–207 (2006).

Bartlett, M.S., 'The frequency goodness of fit test for probability chains', *Proceedings of the Cambridge Philosophical Society*, **47**, 86–95 (1951).

BIS, 'Zero–coupon yield curves: Technical documentation', *BIS Papers No. 25, Bank for International Settlements*, (2005).

Blaskowitz, O. and H. Herwartz, 'PCA based ex–ante forecasting of swap term structures', *International Journal of Theoretical and Applied Finance*, **12**, 465-489 (2009a).

Blaskowitz, O. and H. Herwartz, 'Adaptive forecasting of the EURIBOR swap term structure', forthcoming in *Journal of Forecasting* (2009b).

Blaskowitz, O. and H. Herwartz, 'Testing for the economic value of directional forecasts in the presence of serial correlation', *SFB 649, Discussion Paper 2008–73* (2008).

Bliss, R., 'Movements in the term structure of interest rates', *Federal Reserve Bank of Atlanta, Economic Review*, **4**, 16–33 (1997).

Breen, W., L.G. Glosten and R. Jagannathan, 'Economic significance of predictable variations in stock index returns', *Journal of Finance*, **44**, 1177–1189 (1989).

Bunn, D.W., 'A bayesian approach to the linear combination of forecasts', *Operational Research Quarterly*, **3**, 325–329 (1975).

Camba–Mendez, G., G. Kapetanios and M.R. Weale, 'The forecasting performance of the OECD composite leading indicators for France, Germany, Italy and the UK', In *Clements, M.P. and D.F. Hendry (eds), A Companion to Economic Forecasting*, Blackwell Publishing, Oxford (2002).

Carlstein, E., 'The use of subseries methods for estimating the variance of a general statistic from stationary data', *Annals of Statistics*, **14**, 1171–1179 (1982).

Christoffersen P. F. and F. X. Diebold, 'Further results on forecasting and model selection under asymmetric loss', *Journal of Applied Econometrics*, **11**, 561–571 (1996).

Christoffersen P. F. and F. X. Diebold, 'Optimal prediction under asymmetric loss', *Econometric Theory*, **13**, 808–817 (1997).

Cicarelli, J., 'A new method for evaluating the accuracy of economic forecasts', *Journal of Macroeconomics*, **4**, 469–475 (1982).

Clemen, R.T., 'Combining forecasts: A review and annotated bibliography', *International Journal of Forecasting*, **5**, 559–583 (1989).

Clements, M.P. and D.F. Hendry, 'Explaining forecast failure in macroeconomics', In *Clements, M.P. and D.F. Hendry (eds), A Companion to Economic Forecasting*, Blackwell Publishing, Oxford, (2002).

Chamberlain, G. and M. Rothschild, 'Arbitrage factor stucture, and mean–variance analysis of large asset markets', *Econometrica*, **51**, 1281–1304 (1983).

Cox, D.R. and D.V. Hinkley, *Theoretical statistics*, Chapman and Hall, London (1974).

Cumby, R.E. and D.M. Modest, 'Testing for market timing ability', *Journal of Financial Economics*, **19**, 169–189 (1987).

Dai, Q. and K.J. Singleton, 'Term structure dynamics in theory and reality', *Review of Financial Studies*, **16**, 631–678 (2003).

Davison A.C. and P. Hall, 'On studentizing and blocking methods for implementing the bootstrap with dependent data', *Australian Journal of Statistics*, **35**, 215–224, (1993).

Diebold, F.X., *Elements of forecasting*, Cincinnati, South-Western College Publishing (2007).

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

Diebold, F.X. and R. Mariano, 'Comparing predictive accuracy', *Journal of Business & Economic Statistics*, **13**, 253–263 (1995).

Diebold, F.X. and P. Pauly, 'Structural change and the combination of forecasts', *Journal of Forecasting*, **6**, 21–40 (1987).

Diebold, F.X., M. Piazzesi and G.D. Rudebusch, 'Modeling bond yields in finance and macroeconomics', *American Economic Review*, **95**, 415–420 (2005).

Diebold, F.X., G.D. Rudebusch and S.B. Aruoba, 'The macroeconomy and the yield curve: A dynamic latent factor approach', *Journal of Econometrics*, **131**, 309–338 (2006).

Duffee, G.R., 'Term premia and interest rate forecasts in affine models', *Journal of Finance*, **57**, 405–443 (2002).

Duffie, D. and K. Singleton, 'An econometric model of the term structure of interest rate swap yields', *Journal of Finance*, **52**, 1287–1321 (1997).

Easaw, J.Z., D. Garratt and S.M. Heravi, 'Does consumer sentiment accurately forecast UK household consumption? Are there any comparisons to be made with the US?', *Journal of Macroeconomics*, **27**, 517–532 (2005).

Elliot, G. and A. Timmermann, 'Economic Forecasting', *Journal of Economic Literature*, **46**, 3–56, (2008).

European Central Bank, 'Euro money market study 2006', *Press Release*, download from http://www.ecb.eu/press/pr/date/2007/html/pr070213.en.html on August 16, 2008.

Fisher, R.A., *Statistical methods for research workers*, Oliver and Boyd, Edinburgh (1934).

Gençay, R., 'Optimization of technical trading strategies and the profitability in security markets', *Economics Letters*, **59**, 249–254 (1998).

Götze F. and H. Künsch, 'Second–order correctnes of the blockwise bootstrap for stationary observations', *Annals of Statistics*, **24**, 1914–1933 (1996).

Gradojevic, N. and J. Yang, 'Non–linear, Non–fundamental exchange rate forecasting', *Journal of Forecasting*, **25**, 227–245 (2006).

Granger, C.W.J. and M.H. Pesaran, 'Economic and statistical measures of forecast accuracy', *Journal of Forecasting*, **19**, 537–560 (2000).

Greer, M.R., 'Directional accuracy tests of long–term interest rate forecasts', *International Journal of Forecasting*, **19**, 291–298 (2003).

Greer, M.R., 'Combination forecasting for directional accuracy: An application to survey interest rate forecasts', *Journal of Applied Statistics*, **32**, 607–615 (2005).

Hall P., 'Resampling a coverage process' *Stochastic Process Applications*, **19**, 259–269 (1985).

Hall P. and J.L. Horowitz, 'Bootstrap critical values for tests based on generalized method of moments estimators' *Econometrica*, **64**, 891–916 (1996).

Hall P., J.L. Horowitz and B. Jing, 'On blocking rules for the bootstrap with dependent data' *Biometrika*, **82**, 561–574 (1995).

Härdle, W., H. Herwartz and V.G. Spokoiny, 'Time inhomogeneous multiple volatility modeling', *Journal of Financial Econometrics*, **1**, 55–95 (2003).

Hatzmark, M.L., 'Luck versus forecast ability: Determinants of trader performance in futures markets', *Journal of Business*, **64**, 49–74 (1991).

Havenner, A. and B. Modjtahedi, 'Foreign exchange rates: A multiple currency and maturity analysis' *Journal of Econometrics*, **37**, 251–264 (1988).

Henriksson, R.D. and R.C. Merton, On market timing and investment performance II: Statistical procedures for evaluating forecast skills, *Journal of Business*, **54**, 513–533 (1981).

Herwartz, H. and L. Morales–Arias, 'In–sample and out–of–sample properties of international stock return dynamics conditional on equilibrium pricing factors' *Universität Kiel, Economics Working Paper* (2008).

Holt, D., A.J. Scott and P.D. Ewings, 'Chi–squared tests with survey data' *Journal of the Royal Statistical Society*, **Ser. A, 143**, 303–320 (1980).

Hyndman, R. J. and A. B. Koehler, 'Another look at measures of forecast accuracy', *International Journal of Forecasting*, **22**, 679–688 (2006).

Inoue A. and M. Shintani, 'Bootstrapping GMM estimators for time series' *Journal of Econometrics*, **133**, 531–555 (2006).

Johansen, S., *Likelihood–based inference in cointegrated vector autoregressive models*, Oxford University Press, Oxford (1995).

Johnson, R.A. and D.W. Wichern, *Applied multivariate statistical analysis*, Prentice Hall, 5th ed. (2002).

Joutz, F. and H.O. Stekler, 'Data revisions and forecasting', *Applied Economics*, **30**, 1011–1016 (1998).

Joutz, F. and H.O. Stekler, 'An evaluation of the predictions of the federal reserve', *International Journal of Forecasting*, **16**, 17–38 (2000).

Knez, P.J., R. Litterman and J. Scheinkman, 'Explorations into factors explaining money market returns', *Journal of Finance*, **49**, 1861–1882 (1994).

Kolb, R.A. and H.O. Stekler, 'How well do analysts forecast interest rates', *Journal of Forecasting*, **15**, 385–394 (1996).

Kuan, C.M. and T. Liu, 'Forecasting exchange rates using feedforward and recurrent neural networks', *Journal of Applied Econometrics*, **10**, 347–364 (1995).

Künsch, H.R., 'The jackknife and the bootstrap for general stationary observations', *Annals of Statistics*, **17**, 1217–1241 (1989).

Lahiri, S.N., 'Second order optimality of stationary bootstrap', *Statistics & Probability Letters*, **11**, 335–341 (1991).

Lahiri, S.N., 'Edgeworth correction by moving block bootstrap for stationary and nonstationary data', *In LePage R. and L. Billard (eds.), Exploring the Limits of Bootstrap*, Wiley, New York (1992).

Lahiri, S.N., 'On edgeworth expansion and moving block bootstrap for studentized m–estimators in multipe linear regression models', *Journal of Multivariate Analysis*, **56**, 42–59 (1996).

Lahiri, S.N., 'Theoretical comparisons of block bootstrap methods', *Annals of Statistics*, **27**, 386–404 (1999).

Lahiri, S.N., K. Furukawa and Y.D. Lee, 'A nonparametric plug–in rule for selecting optimal block lengths for block bootstrap methods', *Statistical Methodology*, **4**, 292–321 (2007).

Lai, K.S., 'An evaluation of survey exchange rate forecasts', *Economics Letters*, **32**, 61–65 (1990).

Leitch, G. and J.E. Tanner, 'Economic forecast evaluation: Profits versus the conventional error measures', *American Economic Review*, **81**, 580–590 (1991).

Leitch G. and J.E. Tanner, 'Professional economic forecasts: Are they worth their costs?', *Journal of Forecasting*, **14**, 143–157 (1995).

Litterman, R. and J. Scheinkman, 'Common factors affecting bond returns', *Journal of Fixed Income*, **June**, 54–61 (1991).

Liu, J., F.A. Longstaff and R.E. Mandell, 'The market price of risk in interest rate swaps: The roles of default and liquidity risks', *The Journal of Business*, **79**, 2337-2359 (2006).

Liu, R.Y. and K. Singh, 'Moving blocks jackknife and bootstrap capture weak dependence', *In LePage R. and L. Billard (eds.), Exploring the Limits of Bootstrap*, Wiley, New York (1992).

Lindqvist, B., 'A note on bernoulli trials with dependence', *Scandinavian Journal of Statistics*, **5**, 205–208 (1978).

Lütkepohl, H., *New introduction to multiple time series analysis*, Springer Verlag, Berlin (2005).

Makridakis, S., 'Accuracy measures: theoretical and practical concerns', *International Journal of Forecasting*, **9**, 527–529 (1993).

Merton, R.C., 'On market timing and investment performance I: An equilibrium theory of value for market forecasts', *Journal of Business*, **54**, 363–406 (1981).

Mills, C.T. and G.T. Pepper, 'Assessing the forecasters: An analysis of the forecasting records of the treasury, the London Business School and the National Institute', *International Journal of Forecasting*, **15**, 247–257 (1999).

Miron, P. and P. Swannell, *Pricing and hedging swaps*, Euromoney Publications PLC, London (1991).

Mönch, E., 'Forecasting the yield curve in a data–rich environment: A no–arbitrage factor–augmented VAR approach', *Journal of Econometrics*, **146**, 26–43 (2008).

Nelson, C.R. and A.F. Siegel, 'Parsimonious modeling of the yield curve', *Journal of Business*, **60,**, 473–489 (1987).

Newey, W.K. and K.D. West, 'A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix', *Econometrica*, **55**, 703-708 (1987).

Newey, W.K. and K.D. West, 'Automatic lag selection in covariance matrix estimation', *Review of Economic Studies*, **61**, 631–653 (1994).

Öller, L.E. and B. Barot, 'The accuracy of european growth and inflation forecasts', *International Journal of Forecasting*, **16**, 293–315 (2000).

Patankar, V.N, 'The goodness of fit of frequency distributions obtained from stochastic processes', *Biometrika*, **41**, 450–462 (1954).

Pesaran, M.H. and S. Skouras, 'Decision–based methods for forecast evaluation', *In Clements, M.P., and D.F. Hendry (eds.), A Companion to Economic Forecasting*, Oxford, Blackwell Publishing (2002).

Pesaran, M.H. and A.G. Timmerman, 'A simple nonparametric test of predictive performance', *Journal of Business & Economic Statistics*, **10**, 461–465 (1992).

Pesaran, M.H. and A.G. Timmerman, 'Predictability of stock returns: Robustness and economic significance', *Journal of Finance*, **50**, 1201–1228 (1995).

Pesaran, M.H. and A.G. Timmerman, 'Testing dependence among serially correlated multi–category variables', forthcoming in the *Journal of American Statistical Association* (2008).

Politis, D.N. and J.P. Romano, 'A circular block–resampling procedure for stationary data', *In LePage R. and L. Billard (eds.), Exploring the Limits of Bootstrap*, Wiley, New York (1992).

Politis, D.N. and J.P. Romano, 'The stationary bootstrap', *Journal of the American Statistical Association*, **89**, 1303–1313 (1994).

Politis, D.N. and H. White, 'Automatic block–length selection for the dependent bootstrap', *Econometric Reviews*, **23**, 53–70 (2004).

Pons, J., 'The accuracy of IMF and OECD forecasts for G7 countries', *Journal of Forecasting*, **19**, 53–63 (2000).

Pons, J., 'The rationality of price forecasts: A directional analysis', *Applied Financial Economics*, **11**, 287–290 (2001).

Rao, J.N.K. and A.J. Scott, 'The analysis of categorical data from complex sample surveys: Chi–squared tests for goodness of fit and independence in two-way tables', *Journal of the American Statistical Association*, **76**, 221–230 (1981).

Radulovic D., 'The bootstrap of the mean for strong mixing sequences under minimal conditions', *Statistics & Probability Letters*, **28**, 65–72 (1996).

Remonola, E.M. and P.D. Wooldridge, 'The euro interest rate swap market', *Bank for International Settlements*, **March**, 53–64 (2003).

Schnader, M.H. and H.O. Stekler, 'Evaluating predictions of change', *Journal of Business*, **63**, 99–107 (1990).

Schneider, M. and M. Spitzer, 'Forecasting austrian GDP using the generalized dynamic factor model', *Oesterreichische Nationalbank (Austrian Central Bank), Working Paper No. 89* (2004).

Skouras, S., 'Financial returns and efficiency as seen by an artificial technical analyst', *Journal of Economic Dynamics & Control*, **25**, 213–244 (2001).

Skouras, S., 'Decisionmetrics: A decision–based approach to econometric modelling', *Journal of Econometrics*, **137**, 414–440 (2007).

Steeley, J.M., 'Modeling the dynamics of the term structure of interest rates', *Economic and Social Review*, **21**, 337–361 (1990).

Stekler, H.O., 'Are economic forecasts valuable?', *Journal of Forecasting*, **13**, 495–505 (1994).

Stekler, H.O. and G. Petrei, 'Diagnostics for evaluating the value and rationality of economic forecasts', *International Journal of Forecasting*, **19**, 735–742 (2003).

Svensson, L.E.O., 'Estimating and interpreting forward interest rates: Sweden 1992–1994', *NBER Working Paper No. 4871, National Bureau of Economic Research* (1994).

Swanson, N.R. and H. White, 'A model selection approach to assessing the information in the term structure using linear models and artificial neural networks', *Journal of Business & Economic Statistics*, **13**, 265–275 (1995).

Swanson, N.R. and H. White, 'A model selection approach to real–time macroeconomic forecasting using linear models and artificial neural networks', *Review of Economics and Business Statistics*, **79**, 540–550 (1997a).

Swanson, N.R. and H. White, 'Forecasting economic time series using flexible versus fixed specification and linear versus non linear econometric models', *International Journal of Forecasting*, **13**, 439–461 (1997b).

Tavaré, S. and P.M.E. Altham, 'Dependence in goodness of fit tests and contingency tables', *Biometrika*, **70**, 139–144 (1983).

Thompson, P. A., 'An mse statistic for comparing forecast accuracy across series', *International Journal of Forecasting*, **6**, 219–227 (1990).

West, K.D., 'Forecast evaluation', In *Elliot G., C.W.J. Granger and A. Timmermann (eds), Handbook of Economic Forecasting*, Elsevier B.V., 99–134 (2006).

West K. D., H. J. Edison and D. Cho, 'A utility based comparison of some models of exchange rate volatility', *Journal of International Economics*, **35**, 23–45 (1993)

Wilkie, A.D., 'A stochastic investment model for actuarial use', *Transactions of the Faculty of Actuaries*, **39**, 391 – 403 (1986).

Wilkie, A.D., *More on a stochastic investment model for actuarial use*, Presented to the Institute of Actuaries and Faculty of Actuaries, London (1995).

Ziemba, W.T. and J.M. Mulvey, *Worldwide asset and liability modeling*, Cambridge University Press, Cambridge (1998).

Zucchini, W., 'An introduction to model selection', *Journal of Mathematical Psychology*, **44**, 41–61 (2000).

Zvingelis, J.J., 'On bootstrap coverage probability with dependent data', *In D. Gilles (ed.), Computer–Aided Econometrics*, New York: Marcel Dekker (2001).

Ich erkläre hiermit an Eides Statt, dass ich meine Doktorarbeit “A Forecast Evaluation of PCA-Based Adaptive Forecasting Schemes for the EURIBOR Swap Term Structure” selbständig und ohne fremde Hilfe angefertigt habe und dass ich alle von anderen Autoren wörtlich übernommenen Stellen, wie auch die sich an die Gedanken anderer Autoren eng anlehnenden Ausführungen meiner Arbeit, besonders gekennzeichnet und die Quellen nach den mir angegebenen Richtlinien zitiert habe.

Berlin, September 2009

Curriculum Vitae

Professional Experience

- Landesbank Berlin AG, Asset Management Research, since Feb 2009
  Quantitative Analyst
- Humboldt-Universität zu Berlin, Department of Statistics and Econometrics, Sep 2002 – Jan 2009
  Research and Teaching Assistant
- Deutsche Börse, Market Development Cash Markets (Internship), Frankfurt/Main, Feb-Apr 2003
  Telefon survey of investment companies, multivariate analysis of the emipirical relationship of the costs of raising capital and transaction costs in stock trading, analysis of Asset Backed Securities, producing statistics for the trading activity of competing stock exchanges.
- District Administration of Berlin-Schöneberg/Tempelhof, Nov 2002 and Sep 2001
  Creation of a modul to process personal data in Access 2.0 and later in Access 97
- Société Générale, Debt Finance (Internship), London, Jun-Aug 2002
  Reporting to the Head of Sales Team, analysing strategies to hedge interest rate risk
- National Research Center (SFB) 373 of Humboldt-Universität zu Berlin, Jan-May 2002
  2 Chapters in Applied Quantitative Finance (ebook): „Trading on Deviations of Implied and Historical Distributions“ and „Long Memory Effects Trading Strategies“
- Humboldt-Universität zu Berlin, Department of Statistics and Econometrics, Oct 2000-Aug 2001
  Managing the double diploma exchange program with ENSAE, Paris
- Dresdner Bank AG, Risk Methodology Trading (Internship), Frankfurt/Main, Jul-Oct 2000
  Implementing and analysing stochastic volatility models using C++, Matlab,Visual Basics
- Mediatop, Media Marketing Consulting (Internship), Paris, Jul/Aug 1999
  Creation of a software on FilemakerPro and Excel for exploiting market research
- Mercedes Benz AG (Internship), Berlin, Mar/Apr 1998
  Departments: Purchasing, Accounting, Commercial Vehicles Disposal, Controlling
- GAPmbH, Oct 1996-Aug 1998, car services for Sixt AG
- Klinik Berlin (Civilian, in place of military, Service), Jul 1995-Jul 1996
  rehabilitation centre, speaker of all persons doing civilian service in Klinik Berlin

Education

- Dr. sc. pol, Dec 2009 (Econometrics)
  Christian-Albrechts-Universität zu Kiel
  Topic: A Forecast Evaluation of PCA-Based Adaptive Forecasting Schemes for the EURIBOR Swap Term Structure
  Supervisor: Prof. Dr. Helmut Herwartz
- Master of Science (Statistics), April 2008
  Humboldt-Universität zu Berlin
  Master Thesis: Testing for the Value of Directional Forecasts in the Presence of Serial Correlation
- Diplom-Volkswirt (M.A. in Economics), Oct 2001
  Specialisation: Econometrics, Finance, Statistics
  Diploma Thesis: Trading on Deviations of Implied and Historical Distributions
- Statisticien Économiste (M.A. in Statistics and Economics), Jun 2001
  Ecole Nationale de la Statistique et de l’Administration Economique (ENSAE), Paris
- A-Levels (Abitur) at Carl-Friedrich-von-Siemens Gymnasium, Berlin, Jul 1995

Oliver Blaskowitz

Fields of Interest

- Forecasting Methods, Forecast Evaluation, Applied Quantitative Finance, Econometrics, Statistics

Publications / Discussion Papers

- Blaskowitz O., Herwartz H. (2009): Adaptive Forecasting of the EURIBOR Swap Term Structure, Journal of Forecasting, 25, 575-594.
- Blaskowitz O., Herwartz H. (2009): PCA Based Ex-Ante Forecasting of Swap Term Structures, International Journal of Theoretical and Applied Finance, Vol. 12, No.4, 465-489.
- Blaskowitz O., Herwartz H. (2009): On Economic Evaluation of Directional Forecasts, SFB 649 Discussion Paper No. 2009-52, Humboldt-Universität zu Berlin.
- Blaskowitz O., Herwartz H. (2008): Testing for the Value of Directional Forecasts in the Presence of Serial Correlation, SFB 649 Discussion Paper No. 2008-73, Humboldt-Universität zu Berlin.
- Blaskowitz O., Herwartz H. (2008): A Note on the Model Selection Risk for ANOVA Based Adaptive Forecasting of the EURIBOR Swap Term Structure, SFB 649 Discussion Paper No. 2005-64, Humboldt-Universität zu Berlin. .
- Blaskowitz O., Herwartz H., de Cadenas-Santiago S. (2005): Modeling the FIBOR/EURIBOR Swap Term Structure: An Empirical Approach, SFB 649 Discussion Paper No. 2005-24, Humboldt-Universität zu Berlin.
- Blaskowitz O., Härdle W., Schmidt P. (2004): Skewness and Kurtosis Trades. Appeared in: Rachev S. T.: Handbook of Computational and Numerical Methods in Finance, Birkhäuser, Boston.
- Blaskowitz O., Merk A. (2004): Schätzung künftiger Aktienkursverteilung aus Optionspreisen, Börsen-Zeitung, Edition 76 at 21.04.2004, page 19.
- Blaskowitz O., Schmidt P. (2002): Trading on Deviations of Implied and Historical Densities. Appeared in: Härdle W., Kleinow T., Stahl G.: Applied Quantitative Finance, Springer.
- Blaskowitz O., Schmidt P. (2002): Long Memory Effects Trading Strategies, Appeared in: Härdle W., Kleinow T., Stahl G.: Applied Quantitative Finance, Springer.

Presentations

- A Note on the Economic Evaluation of Directional Forecasts, Econometric Seminar for Master/Ph.D. Students, Humboldt-Universität zu Berlin, Dez 2008.
- Testing for the Value of Directional Forecasts in the Presence of Serial Correlation, Haindorf Seminar, Heijnice, Czech Republic, Feb 2008.
- Testing for the Value of Directional Forecasts in the Presence of Serial Correlation, Econometric Seminar for Master/Ph.D. Students, Humboldt-Universität zu Berlin, Jan 2008.
- Adaptive Forecasting of the EURIBOR Swap Term Structure, Erich-Schneider-Seminar, Christian-Albrechts-Universität zu Kiel, Mai 2007.
- Adaptive Forecasting of the EURIBOR Swap Term Structure, Econometric Seminar for Master/Ph.D. Students, Humboldt-Universität zu Berlin, Apr 2007.
- A Forecast Comparison for the FIBOR/EURIBOR Swap Term Structure, Conference on Recent Developments in Econometrics, Florenz, Italy, Sep 2006.
- Determinants of the Xetra Liquidity Measure, Econometric Seminar for Master/Ph.D. Students, Humboldt-Universität zu Berlin, Feb 2006.
- Adaptive Forecasting of the FIBOR/EURIBOR Swap Term Structure: An Empirical Approach, 16th (EC^2) Conference (poster session), Istanbul, Turkey, Dec 2005.
- Modelling the FIBOR/EURIBOR Swap Term Structure: An Empirical Approach, Haindorf Seminar, Heijnice, Czech Republic, Feb 2005.

- Modelling the FIBOR/EURIBOR Swap Term Structure: An Empirical Approach,
  Erich-Schneider-Seminar, Christian-Albrechts-Universität zu Kiel, Nov 2004.
- Probability Trading,
  Haindorf Seminar, Heijnice, Czech Republic, Feb 2004.
- Probability Trading,
  Conference on Quantitative Methods in Finance, Sydney, Australia, Dec 2003.
- Probability Trading,
  Erich-Schneider-Seminar, Christian-Albrechts-Universität zu Kiel, Nov 2003.
- Skewness and Kurtosis Trades,
  Helenau-Seminar, Helenau, Germany, Aug 2003.
- Trading on Deviations of Implied and Historical Densities,
  Research Seminar Mathematical Statistics, Weierstrass Institute for Applied Analysis and Stochastics, Berlin, Apr 2003.
- Trading on Deviations of Implied and Historical Densities,
  Haindorf Seminar, Heijnice, Czech Republic, Feb 2003.

Teaching Experience

- Econometric Forecasting: winter term 2010.
- Introduction to Econometrics (Exercise session): summer terms 2003 to 2008.
- Econometric Methods (Exercise session): winter terms 2004 to 2007.
- Seminar Econometric Projects: winter terms 2002 (with Herwartz), 2005 (with Krätzig), 2007.
- Selected Topics in Econometrics (joint lecture with Droge, Karaman, Krätzig): winter term 2006.
- Selected lecture replacements in Econometric Methods and Econometric Analysis of Financial Market Data.

Bachelor-, Master-, Diploma Theses (Supervised)

- Incorporating Analysts' Forecasts Into the Black-Litterman Model. Grigory Bolotin, June 2009.
- Forecasting the Term Structure of Market Impacts. Steffen Richter, Sep 2009.
- A Bootstrap Approach to Testing Dependence Among Multicategorical Variables. Varvara Simakova, April 2009.
- Order Aggressiveness on the ASX Market. Ying Xu, Aug 2008.
- Forecasting the EURIBOR Swap Term Structure Using Dynamic Factor Analysis. Gleb Korolkov, Sep 2007.
- Forecasting French Macroeconomic Data Using the Stock and Watson Framework. Martin Weber, Feb 2007.
- Logistic Regression: A General Overview. Jana Riedel, Dec 2006.
- An Approach to the Analysis of Credit Data with Logistic Regression and Bootstrap. Kaiyuan Li, Aug 2006.
- Logistic Regression Based Scoring Models. Romain Alloiteau, Jul 2004.
- Does the Introduction of Deutsche Börse's CCP Increase Order Book Activities? Linlin Niu, Nov 2003.
- A Factor Vector-Autoregressive Model for the Term Structure. Gonzalo de Cadenas-Santiago, Oct 2003.
- Determinants of the Xetra Liquidity Measure. Yunus Avsar, Sep 2003.

Refereering

- Quantitative Finance, Computational Statistics

Skills

- Languages:
  German (native speaker)
  English (fluent)
  French (good)
  Spanish (beginner)
- Computer skills:
  Statistical software: Eviews, GAUSS
  Business software: Microsoft Office, LaTeX
  Other software (some experience): C, C++, FileMakerPro, MatLab, Mathematica, R, S+, SAS, VBA-Excel, XploRe

Miscellaneous

- Member of the local organization team for the Humboldt-Copenhagen Conference 2009 on „Recent Development in Financial Econometrics“
- Member of the local organization team for the CASE – Distinguished Lecture Series
  2009: Torben Andersen and Tim Bollerslev on “Recent Developments in Measuring and Modeling Financial Market Volatility”
  2006: Rudi Zagst on “Integrated Risk Management”
  2005: Steward Hodges on “Dynamic Models of Implied Volatility”
  2004: Ludger Overbeck und Marlene Müller on “Credit Risk Modeling”
  2003: Rama Cont on “Inverse Problems in Financial Modeling”
- Uhlenbruch Seminar “Professional portfolio protection strategies” (Prof. Dr. Thomas Zimmerer, Dr. Thorsten Rühl), Frankfurt am Main, Dec 2009
- Inhouse Seminar Landesbank Berlin AG on “Asset Allocation and Modern Bondmanagement in Practice” (Dr. Jochen Kleeberg, Prof. Dr. Thomas Zimmerer), Berlin, Sep 2009
- Robecco Seminar on “Behavioral Finance and Quant Innovations”, Frankfurt am Main, Sep 2009
- Seminar on “Introduction to Trading on Stock Exchanges”, Berlin Stock Exchange, Dec 2007
- Seminar on “Rhetorical Techniques”, Humboldt-Universität zu Berlin, Apr 2005
- Research Visit at Universidad Carlos III de Madrid, Spain, Aug/Sep 2003
- Grant from German-French University, 1998-2000
- Member of “Organisations-Team” of Carl-Friedrich-von-Siemens-Gymnasium, 1993-1994
- Sports: Football, Jogging and Fitness