The Royal Institute of Technology (KTH) - Division of Mathematical Statistics

Macroeconomic multifactor model

An econometric study

Hulda Thorbjörnsdottir

January 2011

This paper is an econometric study of five publically attainable macroeconomic and market variables from the US market. The aim is to study the behaviour and dependencies of GDP growth, inflation, yield curve, equity indices and foreign exchange rates in order to determine a forecasting model with horizon of one year. Such a model could be used to study possible future macroeconomic scenarios as well as to increase the understanding of how these variables react one to another. Very little is assumed of the series prior to modelling and relations predicted by econometric theory are expected to reveal themselves through the process of data analysis. The forecasting model is kept as simple as possible and the complexity of a model is only increased if it results in significant improvements of forecasting accuracy. Univariate ARMA models as well as multivariate VAR models that allow for international variable dependencies are tested, assuming either normal, t(3) or GARCH-gauss distribution of error terms. Models are ranked by comparing RMSE forecasting values and error terms behaviour. Results indicate that it is hard to find models that outperform the random walk although diagnostic tests indicate that many series share similar patterns of historical movements. It is therefore of interest to study further whether the models can be improved by identifying a common cointegrating vector.

Table of Contents

1 Introduction 5

1.1 Disposition 6

2 Theoretical background 7

2.1 Univariate time series 7

2.2 Multivariate time series 14

2.3 State-space representation 16

3 Relevant results from the literature 19

3.1 Inflation 19

3.2 GDP growth 19

3.3 Equity Index 20

3.4 Foreign exchange rate 20

3.5 Yield curve 21

4 Data 23

5 Modelling procedure 25

5.1 In-sample model estimation 25

5.2 Out-of-sample forecasting 25

6 Model selection – Analysis and results 27

6.1 Real GDP growth 27

6.2 Inflation 30

6.3 Equity Index 32

6.4 FX rates 37

6.5 Treasury rates – Yields 43

6.6 The yield curve 54

7 Conclusions 57

8 Acknowledgements 59

9 References 61

1 Introduction

In 2002 the Swedish National Debt Office introduced a stochastic simulation model that simulates different macroeconomic variables in order to forecast the cost development of the Swedish national debt (Bergström, Holmlund, & Lindberg, 2002). The model was recently refined by Jonson (2008) and used to investigate inter-risk correlations in order to examine risk aggregation.

The macroeconomic model is interesting on its own and can be used to investigate possible future scenarios and modified to highlight specific issues, such as studying the consequences of adding a shock in various variables or analysing the dependencies of other economic variables based on the ones included in the macro-model. For example, by investigating and modelling the dependencies of credit default swap spreads index for given class of rating one could gain some understanding of how credit risk of a company of a certain rating should evolve with changes in the economic environment. This is desirable information for potential investors that might not have access to the internal data of the company at stake.

One could also use the macroeconomic model to explore how indexes for corporate bonds of specific rating classes depend on the economic factors and investigate whether there are any significant differences between the dependencies of the different rating levels. This could increase the understanding of how much bonds of different rating classes can be expected to evolve with changes in the economy.

In this thesis, I focus on modelling the factors of the macroeconomic model. I study the time series and propose a model that best describes the data, accounting for cross-market information as well as autorcorrelating qualities and behaviour of error terms.

I restrain the model to cover variables that can be reached through official and publicly attainable channels. The factors included are the annual inflation rate, quarterly changes in GDP, foreign exchange rates, equity indices and the yield curve with main focus on the American market. Data from the Euro area, UK and Japan is included in order to test the possibility of cross-market effects.

The series are modelled assuming as little as possible about the economic relationships within the data prior to modelling. The idea is to let the data reveal these relations through the models as inspired by Carriero, Kapetanios, & Marcellino (2009) and at the same time keep the modelling as simple as possible. A more sophisticated model is only chosen over a simpler one only if it performs significantly better than the latter.

The univariate models tested are the Random walk along with the linear ARMA model. The residuals are assumed to be either IID or follow a normal GARCH process. Cross-variable information is studied by applying vector-autoregressive models.

The yield curve is modelled in two ways. One is to apply the dynamic Nelson-Siegel model and the other is to model the yield changes with a VAR model and then interpolate between the nodes using a natural cubic spline.

Out-of-sample comparison is done at the forecasting horizon of 12 months.

1.1 Disposition

Chapter 2 provides a theoretical background and introduction to the time series models proposed and tested in this thesis, whereas chapter 3 includes a literature study that lists and summarizes recent findings. Chapter 4 contains a detailed description of the time series used and chapter 5 includes the process of modelling the provided set of data. The actual modelling takes place in chapter 6 along with results and discussions. Conclusions are found in chapters 7.

2 Theoretical background

This chapter provides a theoretical background and introduction to the time series models proposed and tested in this thesis. In 2.1 I start by describing the linear models of univariate time series along with a variety of test related to fitting the models to the data. I then move on to introducing the nonlinear GARCH and ARHC models that make use of time-varying volatility. Chapter 2.2 describes multivariate time series models and in 2.3 the state-space representations are introduced with applications on how to model the yield curve.

2.1 Univariate time series

2.1.1 Stationary time series

Most models rely on the time series being stationary upon modelling. Stationary is defined as follows:

> Stationary time series
>
> $\{X_t\}$ is a time series with $E(X_t^2) < \infty$. $\{X_t\}$ is (weakly) stationary if its mean function, $\mu_X(t)$ is independent of t and the covariance function $\gamma_X(t+h, t)$ is independent of t for each h.

There are several ways to make sure that the modelled data is stationary. The first step is to plot the data and by visual inspection determine if there are any signs of a stochastic- or a deterministic trend. Linear models assume constant volatility. If the data shows signs of exponential growth or variability that increases or decreases with time, the Box-Cox transformation can be applied to the data in order to achieve time series that are approximately linear.

> Box-Cox transformations[1]
>
> This class of transformation can be used to stabilise variability of data when the variability increases or decreases with time. The Box-Cox transformation is defined as:
>
> $$f_\lambda(y) = \begin{cases} \log(y), & \lambda = 0 \\ \dfrac{y^\lambda - 1}{\lambda}, & \lambda \neq 0 \end{cases}$$

Next step is to eliminate the linear trend from the data and test for stationary. By applying the Augmented Dickey-Fuller test for unit root one can determine whether or not the data is stationary and if not, decide whether the trend is deterministic or stochastic. An appropriate transformation of the data can then be made.

> Augmented Dickey-Fuller test (ADF)[2]
>
> The augmented Dickey-Fuller tests look at the model
>
> $$\Delta y_t = c + dt + \beta y_t + \alpha_1 \Delta y_{t-1} + \cdots + \alpha_m \Delta y_{t-m} + \varepsilon_t$$
>
> and test the null hypothesis of a non-stationary process against a stationary process.

[1] As defined in (Brockwell & Davis, 2002)
[2] (Alexander, 2001)

The first test assumes that the process has zero mean and is trend stationary. In this case, c and d are set to zero prior to conducting the test. The second version applies do data showing no sign of a trend but appears to have a nonzero mean setting $d = 0$. Finally, if the data seems include a linear trend, the model is tested as it is.

In all cases the hypothesis testing becomes equivalent to testing $H_0: \beta = 0$ against $H_1: \beta < 0$.

Now, there are a few possibilities. Given that the test was not able to reject the null hypothesis of non-stationarity the ADF test is altered to test whether the trend is deterministic or stochastic. If the test indicates a deterministic trend, one should react by removing a fitted trend line from the data. If the trend appears to be stochastic, the data is differentiated and tested again for stationary in order to confirm the order of integration.

After obtaining stationary data one turns focus at the autocorrelation of the observations of the time series. Just by looking at the plotted sample ACF one can make some assumptions of the data. If $|\hat{\rho}(h)|$ is slowly decaying with time then this is a sign indicating that the time series are still non-stationary.

Sample autocorrelation function (sample ACF)

Given the observations $x_1, \dots, x_n$of a time series the sample mean, sample autocovariance function and sample autocorrelation function of $x_1, \dots, x_n$ are defined as:

$\bar{x} = \frac{1}{n}\sum_{t=1}^{n} x_t$ is the sample mean)

$\bar{\gamma}(h) := n^{-1}\sum_{t=1}^{n-|h|}\left(x_{t+|h|} - \bar{x}\right)(x_t - \bar{x}),\ -n < h < n$
is the sample autocovariance function (sample ACF)

$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)},\ -n < h < n,$ is the sample autocorrelation function.

2.1.2 White noise

The simplest time series models assume that the variables are uncorrelated with identical mean and variance. The basic models of white noise, iid noise and NID noise are defined as follows:

White noise

A sequence of uncorrelated random variables is called white noise if each variable has zero mean and variance σ^2. The notation for this is $\{X_t\} \sim WN(0, \sigma^2)$.

IID noise

A sequence of independent and identically distributed random variables with mean 0 and variance σ^2 is called iid noise, written $\{X_t\} \sim IID(0, \sigma^2)$.

NID noise

A sequence of independent and normally distributed random variables with mean 0 and variance σ^2, is referred to as NID noise, notated $\{X_t\} \sim NID(0, \sigma^2)$.

The sample ACF indicates whether or not the data can be interpreted as observation from the processes above. The noise sequences are made up of uncorrelated random variables which should result in a sample ACF that is close to zero for all $h \neq 0$.

There are several other tests that can be used to determine whether the data may be seen as iid noise. Some of these tests are listed below.

The Ljung –Box Portmanteau test[3]

This test considers all sample autocorrelation functions $\hat{\rho}(h)$ simultaneously by studying Q_{LB}, defined as

$$Q_{LB} = n(n+2)\sum_{i=1}^{h} \hat{\rho}^2(j)/(n-j)$$

The sample ACFs of an iid sequence with finite variance become approximately iid with distribution $N(0,1/n)$ as long as n is large. For large n, Q_{LB} can therefore be approximated by the chi-squared distribution with h degrees of freedom.

A large value of Q_{LB} indicates that some of the sample autocorrelation functions are large, making it unlikely that the data to be iid random variables. The assumption of a iid sequence is therefore rejected at level α if $Q_{LB} > \chi^2_{1-\alpha}(h)$.

The turning point test[4]

This test counts the number of "turns" in the time series. A turning point is defined at time i if for the observation y_i it holds that either $y_{i-1} < y_i$ and $y_i > y_{i+1}$ or $y_{i-1} > y_i$ and $y_i < y_{i+1}$.

If the observations make a sequence of iid random variables the probability of a turning point at time i is $2/3$. The number of turning points, T, will therefore have an expected value $\mu_T = \frac{2}{3}(n-2)$ and variance $\sigma_T^2 = \frac{(16n-29)}{90}$. With n large T becomes approximately $N(\mu_T, \sigma_T^2)$.

If T is much larger than μ_T we conclude that the sequence is fluctuating too rapidly for being an iid sequence. A T much smaller than μ_T indicates a positive correlation between the neighbouring observations. The assumption of an iid sequence is therefore rejected at level α if $|T - \mu_T|/\sigma_T > \Phi_{1-\alpha/2}$

The difference-sign test[5]

This test treats the differenced series $y_i - y_{i-1}$ and counts the number of time it is positive. If the sequence is made up of iid random variables this number, S, has an expected value $\mu_S = \frac{1}{2}(n-1)$ and variance $\sigma_S^2 = \frac{(n+1)}{12}$. With n large S can be approximated as $N(\mu_S, \sigma_S^2)$.
A large value of $|S - \mu_S|$ indicates a trend in the series of data. The assumption of no trend in the time series is therefore rejected at level α if $|S - \mu_S|/\sigma_S > \Phi_{1-\alpha/2}$.
The difference-sign test is not guaranteed to detect seasonal fluctuations in the data. A sinusoidal wave would, for example, pass through the randomness test since its differenced series would be positive equally often as negative.

---

[3] As defined in (Brockwell & Davis, 2002)
[4] (Brockwell & Davis, 2002)
[5] (Brockwell & Davis, 2002)

The rank test[6]

P is the number of pairs (i,j) for which $i<j$ and $y_i<y_j$. Assuming an iid sequence the probability of $y_j-y_i>0$ equals $1/2$ and since the number of pairs (i,j) such that $i<j$ is $\binom{n}{2}=\frac{1}{2}n(n-1)$ the expected value of P is therefore $\mu_P=\frac{1}{4}n(n-1)$ with variance $\sigma_P^2=\frac{n(n-1)(2n+5)}{72}$.
When n is large the approximation $P\sim N(\mu_P,\sigma_P^2)$ becomes valid.
A large value of $|P-\mu_P|$ indicates a trend in the series of data. The assumption of no trend in the time series is rejected at level α if $|P-\mu_P|/\sigma_P>\Phi_{1-\alpha/2}$.

Data samples containing 100 IID random variables are simulated where the first sample is made up of uniformly distributed variables and the second is NID. Dependent data series are created from each sample using following definition a first order autoregressive process[7]: $X_t=0{,}5*X_{t-1}+Z_t$. $Z_t\sim U(0{,}0.1)$ for the uniform data sample and $Z_t\sim N(0{,}0.1)$ in the normal case. These four samples are now used to test the validity of the IID tests listed above. The results are shown in Table 1 and confirm what is expected. All tests (Ljung-Box, Turning point test, Difference-sign test and Rank test) reject IID for the AR processes but not for the uniform and normal data samples at 5 % level of significance.

N = 100 and α = 0.05

	IID				NID
Distribution	Ljung-Box	Turning point test	Difference - sign test	Rank test	Jarque-Bera test
Uniform	0	0	0	0	1
Normal	0	0	0	0	0
AR(1) Uniform	1	1	1	1	1
AR(1) Normal	1	1	1	1	1

Table 1 Test results for 4 simulated data samples with known distribution and sample size 100. Randomness tests as well as tests for NID are conducted at level of significance 0.05.

In order to determine if the time series might be normally distributed one can look at the histogram of the data and compare it with the curve expected from a normally distributed set of observations. Finally, one can look at the qq-plot and proceed with Jarque-Bera test.

The Jarque-Bera test[8]

This test is designed to check for normality by looking at the Jarque-Bera statistic, JB, where

$$JB=n\left[\frac{m_3^2}{6m_2^3}+\frac{\left(\frac{m_4}{m_2^3}-3\right)^2}{24}\right],\text{ and }\ m_{r=}\sum_{i=1}^{n}(Y_i-\bar{Y})^r/n$$

is distributed asymptotically as $\chi^2(2)$ if the residuals $\{Y_t\}$ are normally distributed IID random variables.

---

[6] (Brockwell & Davis, 2002)
[7] The AR process is discussed in detail in chapter 2.1.4.
[8] (Alexander, 2001)

The Jarque-Bera test for NID is used on the simulated data samples (see results in Table 1) and, as anticipated, rejects NID for all samples but the IID normal one. Knowing that the IID and NID tests are good enough to distinguish between IID, NID and non-IID is comforting for later purposes when these tests will be used on data series whose distributions are unknown.

2.1.3 Random walk

The random walk is another simple process. The random walk has proven hard to beat when it comes to modelling economic variables and its simple nature is feasible from a computing point of view. The drawback is that it contains no information of underlying relationships or developing of the markets. It has therefore become the benchmark model which other suggested models must be able to outperform in order to be considered a realistic modelling alternative.

Random walk

The random walk is a stochastic process obtained by tracing a sequence of independent and identically distributed (iid) random variables.

Thus, random walk is defined as $S_t = \sum_{i=0}^{t} X_i$, for $t = 0,1, \ldots$ where $\{X_t\}$ is iid noise.

The simplest way of testing for random walk is to look at the differentiated data and determine if this new set of data can be seen as iid noise.

2.1.4 ARMA models[9]

$ARMA$ models make a class of linear processes that are defined by linear difference equations that have constant coefficients. $ARMA$ stands for autoregressive moving-average and is a combination of just that, an autoregressive (AR) process and a moving-average (MA) process.

$AR(p)$ Process

$\{X_t\}$ is a autoregressive process of order p if

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + Z_t,$$

where $\{Z_t\} \sim WN(0, \sigma^2)$ and $\phi_1, \ldots, \phi_p$ are constants.

$MA(q)$ Process

A moving-average process of order q is defined as $\{X_t\}$ such that

$$X_t = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q},$$

where $\{Z_t\} \sim WN(0, \sigma^2)$ and $\theta_1, \ldots, \theta_q$ are constants.

$ARMA(p, q)$ Process

$\{X_t\}$ is stationary. For every t,

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q},$$

$\{Z_t\} \sim WN(0, \sigma^2)$ and the polynomials $(1 - \phi_1 z - \cdots - \phi_p z^p)$ and

$(1 + \theta_1 z + \cdots + \theta_q z^q)$ have no common factor. Then $\{X_t\}$ is known as an $ARMA(p, q)$ process.

---

[9] (Brockwell & Davis, 2002)

When fitting data to an ARMA model one starts by looking at the sample ACF to get an indication of what the correlation structure looks like.

An ARMA model is determined by using the parameter estimations derived from maximising the likelihood function of the system, a method known as Maximum likelihood estimation or MLE. The definition of the likelihood function is based on the distribution of the distribution of Z_t. The Gaussian likelihood for an ARMA process is given by the equation

$$L = (2\pi)^{-\frac{n}{2}}(det\Gamma_n)^{-\frac{1}{2}}\exp\left(-\frac{1}{2}\boldsymbol{X}_n'\Gamma_n^{-1}\boldsymbol{X}_n\right)$$

where $\boldsymbol{X}_n = (X_1, \ldots, X_n)'$ and $\Gamma_n = E(\boldsymbol{X}_n\boldsymbol{X}_n')$.

The log likelihood value of each model can then be used to compare the models against each other and decide which model is most desirable for the data. Several methods have been developed and the Akaike information Criterion is among those that are best known. The AIC value is computed for each model and the model with minimum AIC is preferred. However, studies have shown that the AIC statistics tend to overestimate the number of estimation parameters needed for modelling. By adding a penalty factor to the AIC value one can counteract this tendency to overestimate.

AIC (The Akaike Information Criterion)

The AIC value is a measure of the goodness of fit of a given model and is defined as

$$AIC := -2\ln(LFF) + 2(p + q + 1)$$

where LFF is the likelihood value. The AIC value provides a way to compare different models where the one with the lowest AIC value is assumed to be the model that best fits the data.

BIC (Bayesian Information Criterion)

The BIC statistics is closely related to AIC but the penalty factor for overfitting is stronger for BIC than AIC. BIC is defined as

$$BIC = -2\ln(LFF) + 2(\mathrm{p} + \mathrm{q} + 1)\ln(\mathrm{n})$$

where n is the number of observations.

The AICC value

The bias-corrected AIC value, the AICC value, is defined by

$$AICC = -2\ln(LFF) + \frac{2(p+q+1)n}{n-p-q-2}. \qquad (1)$$

2.1.5 Time varying correlation coefficients[10]

All the models above assume that the error terms are iid. This does not always seem reasonable so one might suspect that a model that allows the volatilities to vary with time could in some cases perform better.

Indeed, there exist a class of models that incorporate clustering of volatility; an attribute known as autoregressive conditional heteroscedasticity ($ARCH$). The $ARCH$ effect is defined as follows:

$ARCH(p)$

The autoregressive conditional heteroscedasticity model of order p defines the conditional variance of the stationary time series as a weighted average of the past unexpected squared returns:

[10] (Alexander, 2001)

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_p \varepsilon_{t-p}^2$$

Where $\alpha_0 > 0,\ \alpha_1, \dots, \alpha_p \geq 0$ and $\varepsilon_t | I_t \sim N(0, \sigma_t^2)$

The $ARCH$ models have been expanded to include autocorrelation of the squared conditional variance σ^2. These generalised models of order (p, q) are known as $GARCH$ models.

$GARCH(p, q)$

The generalised $ARCH$ model of order (p, q) is defined by the equation

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_p \varepsilon_{t-p}^2 + \beta_1 \sigma_{t-1}^2 + \cdots + \beta_q \sigma_{t-q}^2$$

Where $\alpha_0 > 0,\ \alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q \geq 0$ and $\varepsilon_t | I_t \sim N(0, \sigma_t^2)$

There are several versions of GARCH models. The asymmetrical GARCH model accounts for the asymmetrical behaviour of some data series that are more volatile following a sudden negative change in time series value than after a positive change of the same absolute value.

Normally, the GARCH(1,1) provides an adequate description of data. The general model is reduced accordingly and is defined as:

$GARCH(1,1)$

The generic GARCH(1,1) is defined by

$$\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2$$

Where $\omega > 0,\ \alpha, \beta \geq 0$ and $\varepsilon_t | I_t \sim N(0, \sigma_t^2)$

Sometimes it is reasonable to assume that the error terms might not be Gaussian. When the returns follow a t-distribution the corresponding $GARCH$ model is referred to as the $tGARCH$ model.

When fitting data to a GARCH model it is wise to start by looking at the sample ACF of the squared returns and check for signs of autocorrelation. If the squared returns do not indicate any autocorrelation the error terms can be regarded as independent and introducing a GARCH model is inappropriate. The Ljung-Box test defined earlier can be used test the squared returns for autocorrelation.

Just like earlier, the model parameters are estimated with MLE but the log likelihood function becomes

$$LLF = -1/2\left(T \log(2\pi) + \sum_{t=1}^{T} \log(\sigma_t^2) + \sum_{t=1}^{T} \frac{X_t^2}{\sigma_t^2}\right)$$

when the residuals are normally distributed and

$$LLF = T \log\left(\frac{\Gamma\left(\frac{\upsilon+1}{2}\right)}{\pi^{\frac{1}{2}}\Gamma\left(\frac{\upsilon}{2}\right)}(\upsilon - 2)^{-\frac{1}{2}}\right) - \frac{1}{2}\sum_{t=1}^{T} \log(\sigma_t^2) - \frac{\upsilon+1}{2}\sum_{t=1}^{T} \log\left(1 + \frac{X_t^2}{\sigma_t^2(\upsilon-2)}\right)$$

when they are t-distributed[11].

After fitting the GARCH model to data, the Ljung-Box test can be used again to check that the standardized squared returns, $r_t^{*2} = r_t^2 / \hat{\sigma}_t^2$, where $\hat{\sigma}_t^2$ is the estimate of the GARCH conditional variance, are free from autocorrelation.

---

[11] (Jonson, 2008)

2.2 Multivariate time series

It is now time to look at how the time series relate and depend on each other. By modelling the time series together one can include and investigate the correlation between the different variables and get a better understanding on dependencies within the economy as a whole.

2.2.1 VAR models

The vector autoregression model is a multivariate expansion of the univariate $ARMA$ model. It can be shown that every multivariate $ARMA$ model can be transformed into a multivariate AR model[12] which is why I only consider VAR models when expanding the univariate $ARMA$ models to fit multivariate data.

$VAR(p)$ model[13]

The vector autoregression of order p for an n-variate system of stationary time series is defined by

$$\boldsymbol{X}_t = \boldsymbol{\alpha}_0 + \boldsymbol{A}_1\boldsymbol{X}_{t-1} + \cdots + \boldsymbol{A}_p\boldsymbol{X}_{p-1} + \boldsymbol{\varepsilon}_t, \tag{1}$$

where $\boldsymbol{X}_t$, $\boldsymbol{\alpha}_0$ and $\boldsymbol{\varepsilon}_t$ are $n \times 1$ vectors and $\boldsymbol{A}_1, \dots, \boldsymbol{A}_p$ are $n \times n$ matrices of coefficients.

The model parameters are derived using MLE and the AICC can be altered to fit multivariate data.

Multivariate AICC

The multivariate expansion of the univariate AICC defined in (1) is

$$AICC = -2\ln(L) + \frac{2(pm^2+1)nm}{nm-pm^2-2}$$

Where p is the order of the VAR model, m is the number of observations and n is the number of variables in the system.

It is not always true that a full model with all parameters designed to fit the data is the best one. By restricting the model and presetting some parameters to zero before fitting the model one can achieve a simpler model with just as good, or even better, qualities as the full model. The likelihood ratio test is designed to compare a full model to a restricted one.

Likelihood ratio test

The likelihood ratio test (LR test) compares a model containing restrictions on the model parameters to the unrestricted model. The statistical significance of the difference in log likelihoods of the unrestricted and restricted parameter estimates is evaluated.

Let k be the difference in degrees of freedom between the restricted and the unrestricted model. The test statistics are then assumed to be asymptotically $\chi^2(k)$ distributed.

2.2.2 Error Correction Model (ECM)[14]

Two non-stationary time series are cointegrated if they share a stochastic trend. This means that the series are tied together in the long run even though they might drift apart in the short run.

---

[12] (Brockwell & Davis, 2002)
[13] (Alexander, 2001)
[14] (Alexander, 2001)

Cointegration

Two series x and y are said to be cointegrated if both are integrated of order one $x, y \sim I(1)$ if there exists a linear combination of x and y that is stationary. That is, there exists α such that $x - \alpha y \sim I(0)$.

Thus the cointegration trait of two time series can be used to improve the modelling of their stationary differentiated transformations by including a "disequilibrium term" that captures deviations from the long-run equilibrium relationship between the series. This corrected model is known as the error correction model (ECM).

ECM model

The error correction model defines the first difference of the non-stationary time series $\boldsymbol{X}_t$ as:

$$\Delta \boldsymbol{X}_t = \boldsymbol{\alpha}_0 + \boldsymbol{B}_1 \Delta \boldsymbol{X}_{t-1} + \cdots + \boldsymbol{B}_p \Delta \boldsymbol{X}_{p-1} + \Pi \boldsymbol{X}_{t-1} + \boldsymbol{\varepsilon}_t,$$

where every equation represents a variable in the system. $\boldsymbol{\alpha}_0$ is the constant term, $\boldsymbol{B}_1 \Delta \boldsymbol{X}_{t-1}, \dots, \boldsymbol{B}_p \Delta \boldsymbol{X}_{p-1}$ are the lagged terms up to order p and $\Pi \boldsymbol{X}_{t-1}$ represents the disequilibrium term. The parameters are estimated as usually using OLS method.

When estimating the ECM model, one starts by determining the disequilibrium term. This is done by studying the non-stationary data and search for stationary linear combinations.

Engle-Granger Methodology

The Engle-Granger process includes two steps. The first step is to use OLS regression to estimate the cointegration vector. Next step is to check the residuals for stationary in order to make sure that the linear combination is stationary. This can be done using the Augmented Dickey-Fuller test that was introduced earlier.

The Engle-Granger Methodology may be simple and convenient when modelling two variables but when the system becomes more complex and includes more than two time series questions arise. In larger system there may exist more than one cointegration relationship but the OLS regression will only identify one. Which relationship is then being identified and how many are still unaccounted for? Which variable should be considered to be the dependent one? In this case, the Johansen methodology is more appropriate.

Johansen Methodology

The Johansen test is the multivariate generalisation of the univariate unit root test and is based on a VAR model. Rewriting equation (1) returns

$$\Delta \boldsymbol{X}_t = \boldsymbol{\alpha}_0 + (\boldsymbol{A}_1 - \mathbf{I}) \Delta \boldsymbol{X}_{t-1} + \cdots + \left(\boldsymbol{A}_1 + \cdots + \boldsymbol{A}_{\mathrm{p}-1} - \mathbf{I}\right) \Delta \boldsymbol{X}_{t-p+1} \\ + \left(\boldsymbol{A}_1 + \cdots + \boldsymbol{A}_{\mathrm{p}} - \mathbf{I}\right) \boldsymbol{X}_{t-\mathrm{p}} + \boldsymbol{\varepsilon}_t,$$

$\Delta \boldsymbol{X}_t$ is stationary since $\boldsymbol{X}_t \sim I(1)$ which leads to the conclusion that each term in $\left(\boldsymbol{A}_1 + \cdots + \boldsymbol{A}_{\mathrm{p}} - \mathbf{I}\right)$ must be stationary. The rank of this matrix will therefore reveal the number of independent linear relationships. The eigenvectors of the non-zero eigenvalues are the cointegrating vectors of the system.

The Johansen test now uses a trace test to determine the rank of $\left(\boldsymbol{A}_1 + \cdots + \boldsymbol{A}_{\mathrm{p}} - \mathbf{I}\right)$. The statistic is

$$Tr = -T \sum_{i=R+1}^{n} \ln(1-\hat{\lambda}_i)$$

Where T is the sample size, R is the rank, n is the number of variables included in the system and $\hat{\lambda}_i$ are the estimated eigenvalues ordered decreasingly.

Then the test statistic for every $R = 0,..,n-1$ is calculated testing for $H_0: r \leq R$ against $H_1: r > R$. The critical values of the trace statistic are provided in Johansen & Juselius, (1990).

2.3 State-space representation[15]

A time series $\{Y_t\}$ is said to have a state-space representation if it can be described with an observation equation and a state equation. The w-dimensional observation Y_t is expressed as a linear function of a v-dimensional state variable X_t plus noise. The observation equation and the state equation are defined as follows:

Observation equation

$$\boldsymbol{Y}_t = \boldsymbol{G}_t \boldsymbol{X}_t + \boldsymbol{W}_t, \qquad t = 1,2,\ldots$$

where $\{\boldsymbol{W}_t\} \sim WN(\boldsymbol{0}, \{\boldsymbol{R}_t\})$ and $\{\boldsymbol{G}_t\}$ is a sequence of $w \times v$ matrices

State equation

$$\boldsymbol{X}_{t+1} = \boldsymbol{F}_t \boldsymbol{X}_t + \boldsymbol{V}_t, \qquad t = 1,2,\ldots$$

where $\{\boldsymbol{V}_t\} \sim WN(\boldsymbol{0}, \boldsymbol{Q}_t)$ and $\{\boldsymbol{F}_t\}$ is a sequence of $v \times v$ matrices. $\{\boldsymbol{V}_t\}$ is uncorrelated with $\{\boldsymbol{W}_t\}$. The initial state $\boldsymbol{X}_1$ is uncorrelated with all of the noise terms $\{\boldsymbol{V}_t\}$ and $\{\boldsymbol{W}_t\}$.

This is a flexible way of defining a time series model. In fact all the models from the previous sections can be transformed into state-space representation. The dynamic Nelson-Siegel model proposed by Diebold & Li, (2006) builds on state-space modelling. State-space models are more flexible than the previously defined ARMA models since they do not need the parameters to be constant, but can allow the parameters to vary with time and follow their own process. The observation equation and the state equation are estimated simultaneously by maximising the likelihood for the whole system.

The future behaviour of state-space modelled data can be predicted using a Kalman recursion.

Kalman prediction recursion

The one-step predictors $\widehat{\boldsymbol{X}}_t$ and their error covariance matrices Ω_t are defined as

$$\widehat{\boldsymbol{X}}_t := P_{t-1}(\boldsymbol{X}_t) \text{ and } \Omega_t = E[(\boldsymbol{X}_t - \widehat{\boldsymbol{X}}_t)(\boldsymbol{X}_t - \widehat{\boldsymbol{X}}_t)']$$

and are uniquely determined by the recursions

$$\widehat{\boldsymbol{X}}_{t+1} = F_t \widehat{\boldsymbol{X}}_t + \Theta_t \Delta_t^{-1}(\boldsymbol{Y}_t - G_t \widehat{\boldsymbol{X}}_t) \text{ and} \tag{2}$$

$$\Omega_{t+1} = F_t \Omega_t F_t' + Q_t - \Theta_t \Delta_t^{-1} \Theta_t', \qquad \text{t=1,...}$$

where

$$\Delta_t = G_t \Omega_t G_t' + R_t, \qquad \Theta_t = F_t \Omega_t G_t'$$

and Δ_t^{-1} is any generalised inverse of Δ_t and the initial conditions

[15] (Brockwell & Davis, 2002)

$$\widehat{\boldsymbol{X}}_1 = P(\boldsymbol{X}_1|\boldsymbol{Y}_0) \text{ and } \Omega_1 = E[(\boldsymbol{X}_1 - \widehat{\boldsymbol{X}}_1)(\boldsymbol{X}_1 - \widehat{\boldsymbol{X}}_1)']$$

Relying on the recursive nature of the one-step-ahead Kalman prediction it is now straight forward to extend the recursion for the h-step-ahead prediction. On finds that

$$PX_{t+h} = \mathrm{F_{t+h-1}P_t\,X_{t+h-1}} = \cdots = (F_{t+h-1}F_{t+h-2} \ldots F_{t+1})P_tX_{t+1}, \; h = 2,3, \ldots$$

and P_tX_{t+1} is defined i equation (2).

The h-step Kalman prediction

Define $\Omega_t^{(h)} = E[(\boldsymbol{X}_{t+h} - P_t\boldsymbol{X}_{t+h})(\boldsymbol{X}_{t+h} - P_t\boldsymbol{X}_{t+h})']$ and

$$\Delta_t^{(h)} = E[(\boldsymbol{Y}_{t+h} - P_t\boldsymbol{Y}_{t+h})(\boldsymbol{Y}_{t+h} - P_t\boldsymbol{Y}_{t+h})']$$

The recursion then becomes

$$P_t\boldsymbol{X}_{t+h} = (F_{t+h-1}F_{t+h-2} \ldots F_{t+1})P_tX_{t+1}, \qquad h = 2,3, \ldots$$

$$P_t\boldsymbol{Y}_{t+h} = G_{t+h}P_t\boldsymbol{X}_{t+h}, \quad h = 1,2, \ldots$$

$$\Omega_t^{(h)} = F_{t+h-1}\Omega_t^{(h-1)}F'_{t+h-q} + Q_{t+h-1}, \qquad h = 2,3, \ldots$$

$$\Delta_t^{(h)} = G_{t+h}\Omega_t^{(h)}G'_{t+h} + R_{t+h}, \qquad h = 1,2, \ldots$$

where $\Omega_t^{(1)} = \Omega_{\mathrm{t+1}}$.

After obtaining a h-step-ahead prediction one is interested in measuring the forecasting performance of the given model and compare it to the performance of other model candidates. The root mean squared error (RMSE) can be used as at tool for achieving just this.

RMSE

The root mean squared error is defined as the square root of the mean of the squared prediction errors.

$$RMSE = \sqrt{\left(\frac{E(\hat{\theta} - \theta)}{\sqrt{n}}\right)^2}$$

2.3.1 The dynamic Nelson-Siegel model

The dynamic Nelson-Siegel yield curve is defined as a function of maturity τ

$$y_t(\tau) = L_t + S_t\left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau}\right) + C_t\left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right)$$

The factors L_t, S_t and C_t are time-varying factors and represent the level, slope and curvature of the yield curve. Following (Diebold & Li, 2006) λ is fixed at $\lambda = 0.0609$.

The state-space representation of the model is now given by the observation equation:

$$\begin{pmatrix} L_t \\ S_t \\ C_t \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} L_{t-1} \\ S_{t-1} \\ C_{t-1} \end{pmatrix} + \begin{pmatrix} \eta_t(L) \\ \eta_t(S) \\ \eta_t(C) \end{pmatrix}$$

and the state equation

$$\begin{pmatrix} y_t(\tau_1) \\ \vdots \\ y_t(\tau_N) \end{pmatrix} = \begin{pmatrix} 1 & \dfrac{1-e^{-\tau_1\lambda}}{\tau_1\lambda} & \dfrac{1-e^{-\tau_1\lambda}}{\tau_1\lambda} - e^{-\tau_1\lambda} \\ \vdots & \vdots & \vdots \\ 1 & \dfrac{1-e^{-\tau_N\lambda}}{\tau_N\lambda} & \dfrac{1-e^{-\tau_N\lambda}}{\tau_N\lambda} - e^{-\tau_N\lambda} \end{pmatrix} \begin{pmatrix} L_t \\ S_t \\ C_t \end{pmatrix} + \begin{pmatrix} \varepsilon_t(\tau_1) \\ \vdots \\ \varepsilon_t(\tau_N) \end{pmatrix}$$

where the white noise disturbances of observation and state equations are orthogonal to the initial state and to each other.

$$\begin{pmatrix} \eta_t \\ \varepsilon_t \end{pmatrix} \sim WN\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} Q & 0 \\ 0 & H \end{pmatrix}\right] \text{ and } E[\eta_t(L_0 \quad S_0 \quad C_0)] = E[\varepsilon_t(L_0 \quad S_0 \quad C_0)] = 0.$$

3 Relevant results from the literature

Below is a summary of what has recently been written in the fields of modelling each variable included in the multifactor model.

3.1 Inflation

Jonson (2008) uses a first order autoregressive process to model inflation just as was done in the original paper from the Swedish national debt government (Bergström, Holmlund, & Lindberg, 2002). The idea is that since most countries aim at maintaining price stability one can assume that inflation is stable around a certain mean and that the state of the economy is represented through short term interest rates. For Jonson (2008) the process provided decent R^2 values that ranged between 85-96% for the different markets but for two of the regions, the assumption of normally distributed residuals was rejected with help of Jarque-Bera test.

Gonzáles, Hubrich, & Teräsvirta (2009) use a shifting-mean autoregressive model to forecast inflation in the euro area, UK and USA. By using penalised likelihood when estimating the model parameters they combine the information of the data series with exogenous information such as the inflation target of the central banks. In that sense the Gonzáles, Hubrich, & Teräsvirta (2009) provide a more flexible model than Bergström, Holmlund, & Lindberg (2002) but their model is designed to perform at medium and long-term forecasting. The forecasting period of my work however is 12 months only, indicating that different approach might be of more convenience.

Marcellino (2008) writes a detailed analysis of the forecasting performance of univariate time series models. He compares several alternating specifications of autoregressive models for both GDP growth and Inflation, including normal AR models, time-varying AR models, smooth transition AR models and artificial neural networks. Marcellino (2008) concludes that quantitative gains from time-varying/non-linear models are non-existent and that it is hard to outclass carefully specified linear time series. Particularly, this applies to GDP growth.

Canova (2007) compares the forecasting performance of several models of inflation for G-7 countries. His results agree with those of Marcellino (2008) and suggest that very little improvement can be expected from choosing multivariate models, that are suggested by economic theory or statistical analysis, over univariate ones. Moreover, Canova (2007) explores the effect of international interdependencies and reckons that cross-sectional information can indeed improve the performance of a model but only when used in combination with time varying coefficients models.

3.2 GDP growth

When modelling the real GDP growth Jonson (2008) follows Bergström, Holmlund, & Lindberg (2002) and assumes a regime switching AR(1) process that shifts between the two states of the business cycle; recession and boom.

Hogrefe (2007) follows up on the documented leading properties of the yield spread on GDP growth. He examines the significance of the theory claiming that leading properties of the yield spread should be determined by monetary policy. Hogrefe (2007) finds some evidence that supports this hypothesis but none conclusive and the question on how the yield spread connects to GDP growth is left unanswered.

As mentioned earlier, Marcellino (2008) concludes that it may be hard to find a better model than a carefully specified linear time series model. A careful specification refers to taking into account possible existence of unit root and a right choice of deterministic component or information criteria for lag length selection.

3.3 Equity Index

Jonson (2008) follows the work of Ibbotson & Cheng from 2001[16] and applies the Building Block Method to model risk premium on equity index. This method defines the equity premium as a function of stock return, inflation and risk free premium and thus provides a theoretical connection between equity index and the Macroeconomy.

Akgiray (1989) studies equity indices from a pure time series perspective. By looking at historical daily CRSP[17] indices ranging between 1963 and 1986 he finds that the GARCH(1,1) model is well suited to model daily returns. He also concludes that although providing a good fit to daily data, the model is less appropriate to model weekly or monthly data. In fact, the monthly returns can be assumed to be independently normally distributed.

Mckenzie & Omran (2000) look at stock development of 50 UK companies and apply a GARCH model on daily stock returns and include the volume of trade in the conditional variance. They find that the GARCH model becomes superfluous in the model when the extra information of trade volumes is incorporated but there is still significant GARCH effect to be found in the autocorrelation of squared residuals.

Alexander (2001) finds strong signs of cointegration between Dutch, German and French equity indices. She also refers to a study from 1995 conducted by Alexander and Thillainathan[18] that discovers cointegration between Asian-Pacific equity markets, given that indices are expressed in local currencies.

3.4 Foreign exchange rate

The foreign exchange rate has long been modelled based on economic theories. Unfortunately, most studies of the last decades suggest that these models operate no better, or are in some cases outper-formed, by a simple random walk.[19] Jonson (2008) and Bergström, Holmlund, & Lindberg (2002) both employ a theoretical model based on GDP growth and long term interest rates of both currency areas.

Cuaresma & Hlouskova (2004) compare the performance of several vector autoregressive models when forecasting the exchange rates for Central and Eastern European currencies against the Euro and the US dollar. Their work is based on monetary economic theories for exchange rates. Some improvements over the random walk is detected but only within the range of long term forecasting and no evidence is found to support the choice of more refined models in favour of simple random walk when it comes to forecasting over shorter horizons.

Carriero, Kapetanios, & Marcellino (2009) deviate from the theoretical model and implement a pure time series approach instead. They use a panel of 33 exchange rates (where currencies are priced against the US dollar) to incorporate cross-dynamics between the exchange rates into their model.

---

[16] Ibbotson, R.G. & Chen, P. (2001) The Supply of Stock Market Return. Working Paper.
[17] CRSP stands for Center for Research in Security Prices
[18] Alexander, C.O. & Thillainathan, R. (1995). The Asian connection. *Emerging Markets Investor* **2**(6), 42-46.
[19] (Carriero, Kapetanios, & Marcellino, 2009)

By shrinking the VAR coefficients towards a random walk representation Carriero, Kapetanios, & Marcellino (2009) use a Bayesian Vector Autoregressive model (BVAR) to incorporate the prior assumption of a drift-less random walk to the cross-sectional information from the panel of data. They find that the BVAR model systematically outperforms random walk for most currencies and at all horizons.

3.5 Yield curve

When it comes to modelling the yield curve, Jonson (2008) deviates from the ways of Bergström, Holmlund, & Lindberg (2002). While Jonson (2008) models the yield curve as a whole, following the work of Diebold & Li (2006), Bergström, Holmlund, & Lindberg (2002) focus on short-term nominal interest rates separately from the long-term rates, nominal as well as real.

Much work has been done on modelling the yield curve. The articles described below can be seen as indications of recent development within the field.

Diebold & Li (2006) adjust the Nelson-Siegel framework to model the yield curve as a time varying three-dimensional parameter. The parameters are interpreted as the level, slope and curvature of the yield curve and the time variation of the parameters are modelled as an AR(1) process. Diebold & Li (2006) find that this model outperforms the Random walk at 1-year-ahead forecasting horizon.

Aruoba, Diebold, & Rudebusch (2006) refine the dynamic Nelson-Siegel model proposed by Diebold & Li (2006). The model is represented as a state-space representation that allows for interaction between the latent factors (level, slope and curvature) and macroeconomic factors. Aruoba, Diebold, & Rudebusch (2006) find that the yield curve is linked to inflation.

Moench (2008) uses a panel of 160 time series of various economic categories to predict the future development of the yield curve. With a factor-augmented VAR and a term structure based on an assumption of absence of arbitrage he claims to be able to outperform Diebold & Li (2006) at intermediate and long forecasting horizons.

Bowsher & Meeks (2008) propose a functional signal plus noise (FSN) model to model and forecast the yield curve. The stochastic behaviour of the signal function is described by the state equation, assuming a cointegrated VAR process for a subspace of the yields. These yields are then used to determine the observation equation of the state-space model. The observation equation has the form of a natural cubic spline. Bowsher & Meeks (2008) compare their work to Diebold & Li (2006) and find that the FSN-ECM model outperforms the dynamic Nelson-Siegel model when forecasting at 1-month-ahead horizon.

4 Data

Five variables are considered: GDP growth, foreign exchange rate, equity index and the yield curve.

Data is used from four currency areas: USA, Japan, UK and the Euro area.

The GDP growth is defined as quarterly percentage change of GDP volumes. The series come from the OECD database with collected volume series starting in March 1996. GDP growth data ranges from June 1996 to September 2009.

The inflation is determined using the monthly Consumers Price Index (CPI) for each currency area. Inflation is then defined as the percentage change in CPI Index over a 12 month period. The inflation series start in January 1997 and end in September 2009 with CPI data ranging from January 1996.

The exchange rates between the currencies of the three markets and the US dollar are included and presented in dollars. Series start in January 1999 and end in September 2009.

Three equity indices are used: the S&P 500 index, the S&P 400 Midcap index and the S&P 500 Smallcap index. Series start in January 1996 and end in September 2009.

U.S. Treasury yields with maturities of 3, 6, 12, 24, 36, 60, 84 and 120 months that are published by the Federal Reserve are used. The yields are then used to determine the yield curve. Series start in January 1996 and end in September 2009.

5 Modelling procedure

This chapter contains the process of modelling the provided set of data. Chapter 5.1 describes the strategy of choosing models that adequately explain the behaviour of the data. Chapter 5.2 explains the procedure of comparing the out-of-sample performances of the competing models in order to point out the most appropriate model for the given set of data.

5.1 In-sample model estimation

The time series are studied according to the theory described in chapter 2. After finding stationary transformations one proceeds to define and estimate the time series models.

The univariate models tested are the Random walk and the linear ARMA model. The residuals are assumed to be either IID or follow a normal GARCH process.

Next step is to investigate the influence of international data on GDP and inflation by allowing the USA data to be correlated to GDP and inflation of other countries. The relationships within the equity index data and foreign exchange rate data are explored. Vector autoregressive models are used.

The yield curve is modelled in two ways. The first approach is to apply the dynamic Nelson-Siegel model and focus on how to model the latent factors. The other solution is to apply a VAR model on yield changes and then interpolate between the nodes using a natural cubic spline. This method is, in many ways, similar to the scheme proposed by (Bowsher & Meeks, 2008).

The most promising models from this in-sample investigation are then let through to the second part of the study, the out-of-sample comparison where the forecasting performance of the selected models is analysed.

5.2 Out-of-sample forecasting

The out-of-sample forecasting is done using *a rolling window* forecasting technique. The estimation window is rolled over the whole estimation sample ranging from January 1996 to March 2009, delaying the start-date by one month for every new estimation and maintaining the sample size. The first estimation period is taken as from January 1996 to December 2004 followed by forecasting estimations with forecasting horizons of 12-months.

When modelling GDP growth, inflation and FX rates, the estimations are done recursively from the earliest data available to the date of forecast being made, with forecasting date starting in December 2004 and extending, one month at time, until the sample size is the same as of the estimation windows defined earlier. After that the estimation window is rolled as usual. The out-of-sample performance is measured using RMSE.

6 Model selection – Analysis and results

The modelling process of each macroeconomic variable is listed in a separate chapter along with results and relevant discussions (chapter 6.1 - 6.5). Modelling of the yield curve as a whole is described in chapter 6.6.

6.1 Real GDP growth

Data: Quarterly volume data from OECD database is used. Only US data is of interest. International series are used in order to improve US performance. Real GDP growth defined as quarter-on-quarter percentage change.

Series start in June 1996 and end in September 2009. (Volume data from Mar 1996)

In sample estimation: First the whole data window is used to estimate how models fit data. Promising models are then tested in an out-of-sample comparison. AICC value and residuals behaviour are used as measure of fit.

Out-of-sample estimation: The first forecasting window is from Jun1996-Dec2004. Window is then widened – one month at time – until March 2009 is reached. For each window, model is estimated for window data and a forecasting 12 months ahead is made. Performance is measured in RMSE of forecasting error.

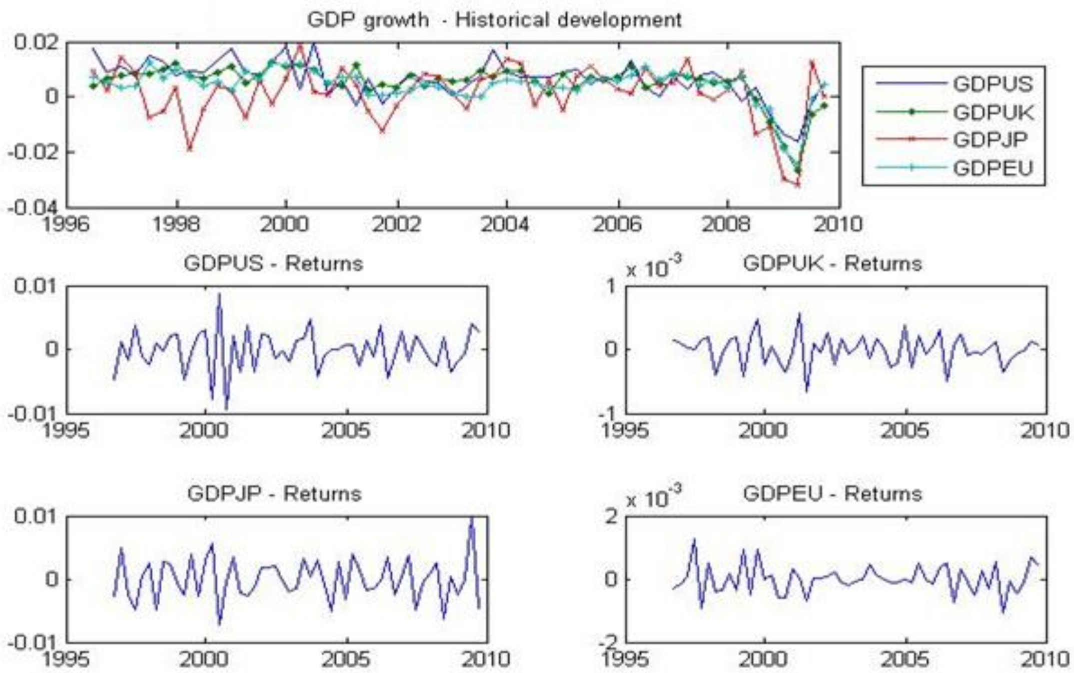

Fig 1 Plot at the top shows the historical development of real GDP growth in US, UK, JP and EU. Plots below show the derived series that will be modelled. In the case of real GDP growth, series have first been Box-Cox transformed and then differentiated.

Initial analysis: First step is to take a look at the historical development of the GDP growth. Fig 1 shows the historical GDP growth along with the transformed data. Transforming the series using Box-Cox transformation yields a set of data that is integrated of order one. The ADF test rejects stationarity at level 0.05 whereas the transformed increments are assumed stationary.

Correlation coefficients

	GDP-US	GDP-UK	GDP-JP	GDP-EU
GDP-US	1,00000	0,04337	-0,20325	0,04138
GDP-UK	0,04337	1,00000	0,13352	0,23503
GDP-JP	-0,20325	0,13352	1,00000	0,23539
GDP-EU	0,04138	0,23503	0,23539	1,00000

Table 2 Correlation coefficient matrix for GDP growth.

Table 2 shows the correlation coefficients for the different transformed GDP growth data series. One can see that there are almost no signs of a linear relationship between the US series and the European series respectively, with correlation coefficient at 0.04, and a vague negative relationship between US series and the Japanese data, correlation coefficient at -0.20. These results do not indicate relationships between the series.

Model selection: After having found a stationary representation of the data one now investigates whether or not this representation can be assumed to be IID. Since the goal is to correctly model GDP growth of USA, the focus is on US data only.

Fig 2 contains first test results for the US time series. The sample ACF indicates that a low order ARMA process might be more appropriate than random walk and the Ljung-Box Portmanteau test rejects the assumption of IID. Jarque-Bera test fails to reject normality at level 0.05 and there are signs of heteroscedasticity.

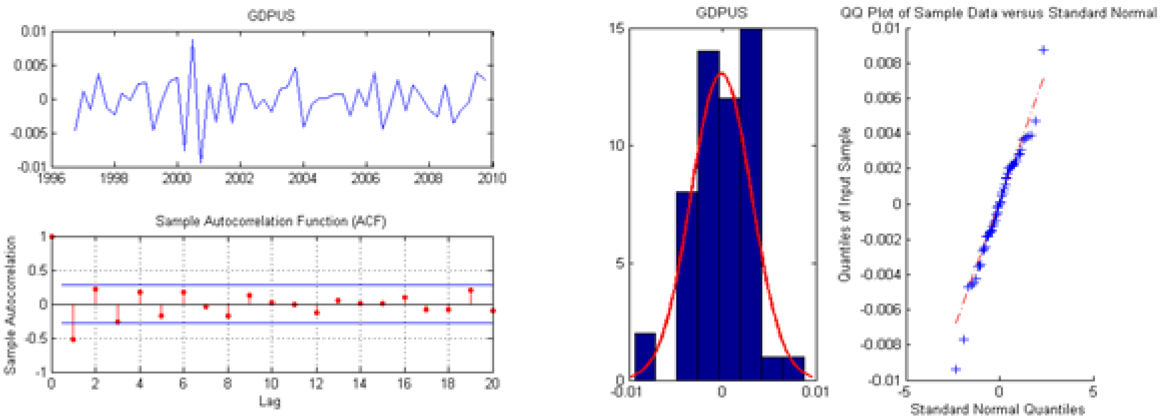

Fig 2 Test results for the changes in transformed real GDP growth series. The figure shows historical development, sample ACF, histogram including a normal curve for comparison and normal qq-plot.

The sample ACF indicates that an ARMA process might be appropriate to model the GDP growth. To test this several ARMA(p,q) models are fitted to the data. P and q are run between 0 and 6 and the error terms are assumed to have Gaussian, t(3) or GARCH(1,1)-Gaussian distribution. It turns out that the ARMA(0,1) leads to residuals that not only pass the IID tests described in detail in 2.1.2 but can also be assumed to be normally distributed.

Out-of-sample comparison (Table 3) reveals that there is little or none forecasting gain to be made from the error term distribution assumptions. Gaussian ARMA(0,1) is therefore chosen as the best and most simple model providing RMSE value at 0.0026.

GDPUS

Model	RMSE	IID Ljung-Box	IID Turning point test	IID Difference - sign test	IID Rank test	NID Jarque-Bera test	AICC
Data – initial analysis	-	1	0	0	0	0	-
ARMA(0,0) - Gaussian	0,00258	1	0	0	0	0	-
ARMA(0,1) - Gaussian	0,00258	0	0	0	0	0	-
ARMA(0,1) - GARCH-Gaussian	0,00258	0	0	0	0	0	-
ARMA(0,1) - T(3)	0,00258	0	0	0	0	0	-
VAR(1) - VAR full & Q full	0,00227	0	0	0	0	0	-2326
VAR(1) - VAR full & Q diag	0,00227	0	0	0	0	0	-2310
VAR(1) - VAR diag & Q full	0,00251	0	1	1	0	0	-2304

Table 3 Performance of best models measured in forecasting performance (RMSE) and residual behaviour (Tests for IID and NID) as well as initial analysis of the time series being modelled. Forecasting horizon is 12 months ahead and tests for independences are significant at level 95%.

The chosen ARMA model provides satisfying results for GDP growth in the US but, although it is not supported by covariance analysis, it is interesting to find out whether there is some advantage in including international data in the time series model. This is done by allowing the US series to correlate with GDP growth in Japan, UK and the Euro-area.

Vector autoregressive models of order between 1 and 8 are tested in four versions each where the VAR-matrix and correlation matrix are set to be either full or diagonal. It turns out that the high order models do not improve the forecasting performance of the first order models.

AICC values indicate that the model with both full VAR matrix and Q provides the best fit to data with more information stored in the VAR matrix than Q. The VAR(1) models with full VAR matrix provide NID residuals but both turning point test and difference-sign test reject IID for the diagonal VAR/full Q model. The best VAR models provide slightly better forecasting results than the best univariate model with RMSE at 0.0023.

Since the goal is to always use the simplest model possible, the first order VAR model with full VAR matrix and diagonal Q matrix is chosen to model GDP growth changes.

6.1.1 Summary and suggested improvements

Results indicate that a first order vector autoregressive model with diagonal correlation matrix and fully determined VAR matrix is the most suitable model for describing quarterly GDP growth. The model residuals can be considered NID adding support to the choice of model since the VAR model assumes NID error terms.

Import and export values are, in essence, a measure of interaction between markets and one might therefore expect inter-market dependencies for real GDP. These findings are consistent with the fact that GDP volumes do depend on the relation between the import and export.

To improve the out-of-sample performance even more, one could do some more tests. It is possible that other variables, such as treasury or inflation rates, contain additional information and even though the historical development of GDP changes in Fig 1 do not indicate a common cointegration vector, it cannot be ruled out completely without further research. For time being, the obtained first order VAR model is accepted.

6.2 Inflation

Data: Monthly CPI data from OECD database is used. Only US data is of interest. International series are used to try to improve US performance. Inflation defined as annual percentage change.

Series start in January 1997 and end in September 2009. (CPI data from Jan 1996

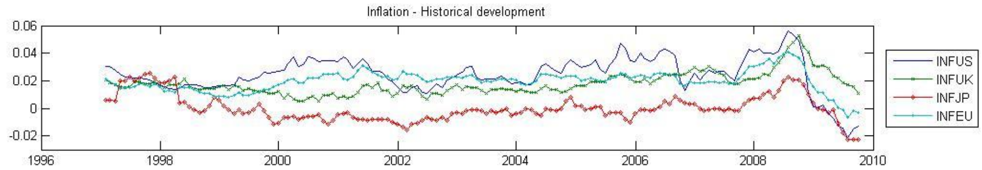

Fig 3 Plot showing the historical development of inflation in US, UK, JP and EU.

In sample estimation: First the whole data window is used to estimate how models fit data. Promising models are then tested in an out-of-sample comparison. AICC value and residuals behaviour are used as measure of fit.

Out-of-sample estimation: The first forecasting window is from Jan1997-Dec2004. Window is then widened – one month at time – until forecasting window is the same size as Jan1996-Dec2004. After that, window size is constant rolling one month at time until March 2009 is reached. For each window, model is estimated for window data and forecasting 12 months ahead is made. Performance is measured in RMSE of forecasting error.

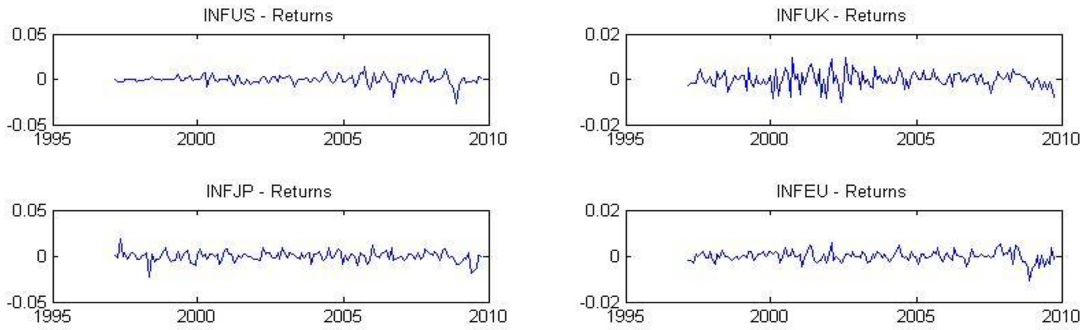

Fig 4 The derived return series that will be modelled. The inflation series have first been Box-Cox transformed and then differentiated.

Initial analysis: Just like in the case of GDP growth, a Box-Cox transformation is applied to the historical inflation data and the series is then differentiated. This results in stationary series for all currency areas. Fig 3 shows the historical development of US inflation series and the transformed return series are plotted in Fig 4.

Correlation coefficients

	Inf-US	Inf-UK	Inf-JP	Inf-EU
Inf-US	1,00000	0,27213	0,29282	0,67693
Inf-UK	0,27213	1,00000	-0,00776	0,39991
Inf-JP	0,29282	-0,00776	1,00000	0,17360
Inf-EU	0,67693	0,39991	0,17360	1,00000

Table 4 Correlation coefficient matrix for inflation

Table 4 shows the correlation coefficients of the transformed inflation series. Only the inflation of the EURO area seems to show some real signs of a positive linear relationship with the US inflation with correlation factor of 0.67.

Model selection: The ACF plot of US inflation indicates that it is unlikely that the inflation of the US should be modelled as random walk (Fig 5). There is a peak at lag 12, suggesting an element of yearly seasonality embedded in the data. Indeed, both Ljung-Box and turning point test reject the assumption of IID increments and Jarque-Bera rejects normality at level 0.05. There are also signs of heteroscedasticity.

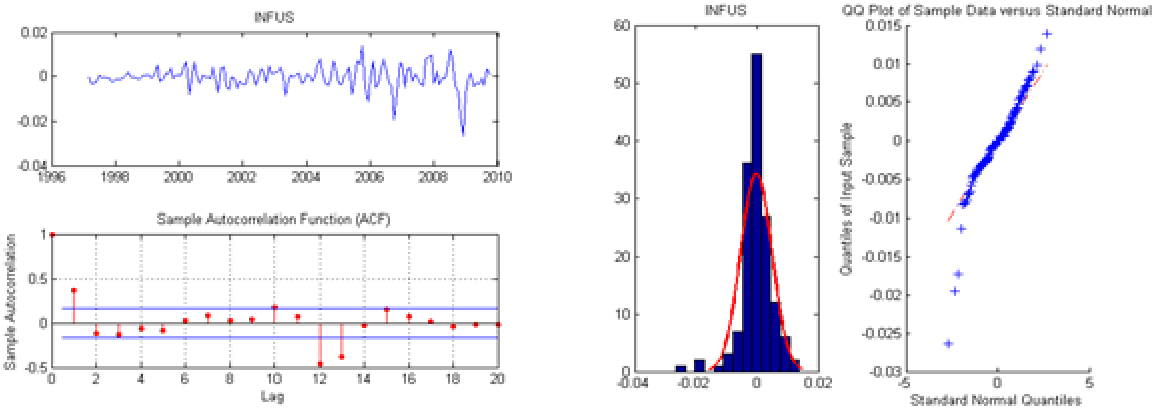

Fig 5 Test results for the changes in transformed inflation series. The figure shows historical development, sample ACF, histogram including a normal curve for comparison and normal qq-plot.

INFUS

Model	RMSE	IID				NID	AICC
		Ljung-Box	Turning point test	Difference - sign test	Rank test	Jarque-Bera test	
Data – initial analysis	-	1	1	0	0	1	-
ARMA(12,0) - Gauss - Seasonality	0,00616	1	0	0	0	1	-
ARMA(0,0) - Gaussian	0,00734	1	1	0	0	1	-
ARMA(0,0) - GARCH-Gaussian	0,00733	1	1	0	0	1	-
ARMA(0,0) - T(3)	0,00728	1	1	0	0	1	-
VAR(1) - VAR full & Q full	0,00749	1	0	0	0	1	-5138
VAR(1) - VAR full & Q diag	0,00749	1	0	0	0	1	-5022
VAR(1) - VAR diag & Q full	0,00748	1	0	0	0	1	-5110

Table 5 Performance of best models measured in forecasting performance (RMSE) and residual behaviour (Tests for IID and NID) as well as initial analysis of the time series being modelled. Forecasting horizon is 12 months ahead and tests for independences are significant at level 95%.

ARMA(p,q) models, where p and q are run between 0 and 6 are now fitted to the data assuming Gaussian, t(3) and GARCH(1,1)-Gaussian behaviour of the residuals. The results are not convincing (Table 5). There is no visible gain in preferring a more complicated model over a simpler one and the random walk model with normally distributed residuals turns out to be the best choice with RMSE value of 0.0073. Testing the residuals reveals that there are still signs of heteroscedasticity and Ljung-Box test as well as the turning point test and Jarque-Bera test reject assumption of IID and normality at level 0.05.

When searching for possible cross-effects from the inflation of the other markets VAR-models of order 1 to 13 are applied. Four versions of each model are tested; a full or a diagonal VAR matrix combined with full or diagonal Q matrix. Results indicate that a VAR model of order one with all parameters set should be chosen. Residual behaviour is slightly improved with the turning point test now failing to reject IID at level 0.05. Testing the out-of-sample performance returns RMSE = 0.0075 – almost the same value as obtained in the univariate case but using a much more complicated model. The cross-market effect will therefore be ignored in further modelling.

It was noted earlier that the sample ACF indicated a possible 12 month lag. An autoregressive twelfth lag is therefore added to an ARMA(0,0) process. Turning point test fails to reject the notion of IID but Ljung-Box and Jarque-Bera tests still reject IID and NID at level 0.05. The out-of-sample performance is improved with RMSE value of 0.0062. Assuming non-normal residuals does not improve the residual behaviour.

The chosen model for US inflation is therefore an autoregressive process lagged at level 12, assuming normally distributed error terms.

6.2.1 Summary and suggested improvements

Applying an ARMA model with autoregressive 12th lag to model the US inflation series improves the out-of-sample performance obtained by applying random walk. Cross-market inflation data does not affect modelling results. These findings are consistent with the fact that CPI Index is based on domestic price level and consumers expectations and should therefore be less sensitive to international influences than GDP.

The residual behaviour is still not satisfying. In search for a better fit, it is interesting to know whether there exist relations between the US inflation series and other variables such as treasury rates, GDP growth or even FX-rates. First step towards improving the model should nevertheless be to eliminate the yearly seasonality from the inflation series through subtraction before attempting to fit a model to the data.

6.3 Equity Index

Data: Monthly EI data in form of the S&P 500 index, the S&P 400 Midcap index and the S&P 500 Smallcap index are used. Series start in January 1996 and end in September 2009.

In sample estimation: First the whole data window is used to estimate how models fit data. Promising models are then tested in an out-of-sample comparison. AICC value and residuals behaviour are used as measure of fit.

Out-of-sample estimation: The first forecasting window is from Jan1996-Dec2004. After that, window size is constant- rolling one month at time until March 2009 is reached. For each window, model is estimated for window data and forecasting 12 months ahead is made. Performance is measured in RMSE of forecasting error.

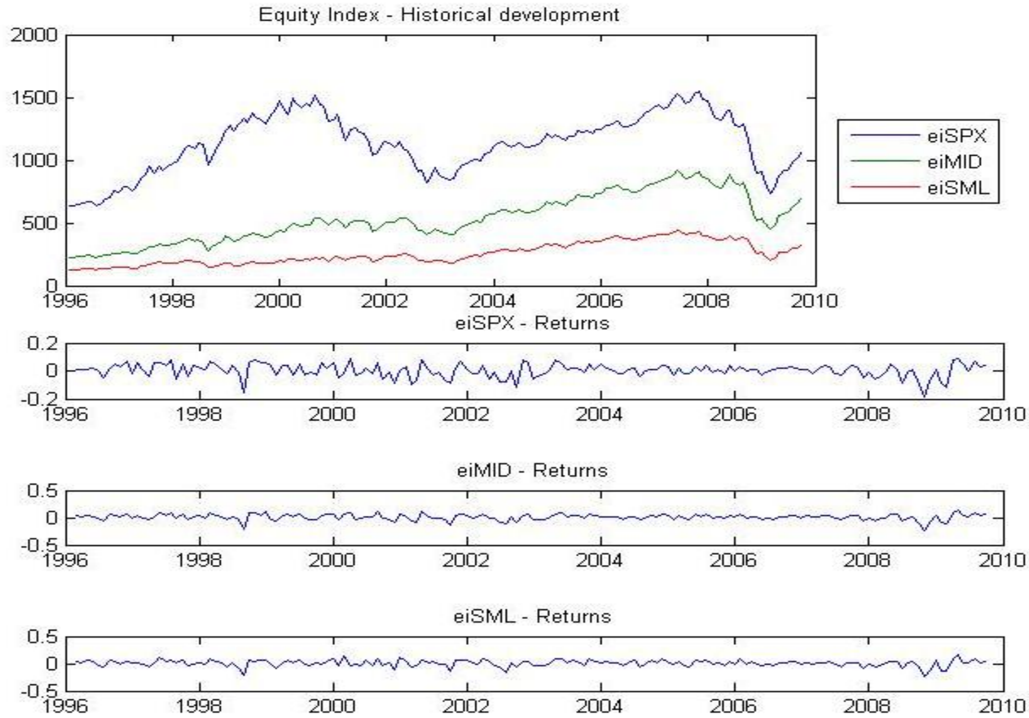

Fig 6 Plot at the top shows the historical development of equity indices SPX, MID and SML. Plots below show the derived series that will be modelled. In the case of equity indices, the logged series have been differentiated.

Initial analysis: Start by taking the logarithm of all equity index series. Both the historical development of the indices as well as transformed series is shown in Fig 6. ADF-test indicates that the logged data is integrated of the first order, that is, log-transformed series are non-stationary whereas the differentiated series are stationary. Each index is analysed individually in chapter 6.3.1, 6.3.2 and 6.3.3.

Correlation coefficients

	EI-SPX	EI-MID	EI-SML
EI-SPX	1,00000	0,89775	0,80326
EI-MID	0,89775	1,00000	0,92682
EI-SML	0,80326	0,92682	1,00000

Table 6 Correlation coefficient matrix for Equity Index

The correlation coefficient matrix (shown in Table 6) indicates that the transformed data is highly correlated, indicating that there exist relationships between the data series. In particular, one notices that the Midcap and Smallcap indices have the highest correlation coefficient at 0.93 whereas the SPX index and the Smallcap index have the lowest correlation value at 0.80. This could be explained by looking at how the indices are constructed. The SPX index is based on the stocks of 500 large publicly held companies traded in the US whereas the Smallcap index (SML) refers to smaller companies and the Midcap index (MID) covers the middle section. By simply considering the foundations of the indices, one therefore expects the SML and MID indices to depend more of the current state of the economy than the SPX index.

6.3.1 S&P 500 Index

When focusing on the S&P500 index (SPX) one can see from the sample ACF in Fig 7 that there is little evidence of a more appropriate model than the random walk. There is no sign of autocorrelation and none of the IID tests is able to reject the notion of IID. Fitting the histogram to a normal curve makes it unlikely that data should be normally distributed and the Jarque-Bera test does indeed reject normality at level 0.05.

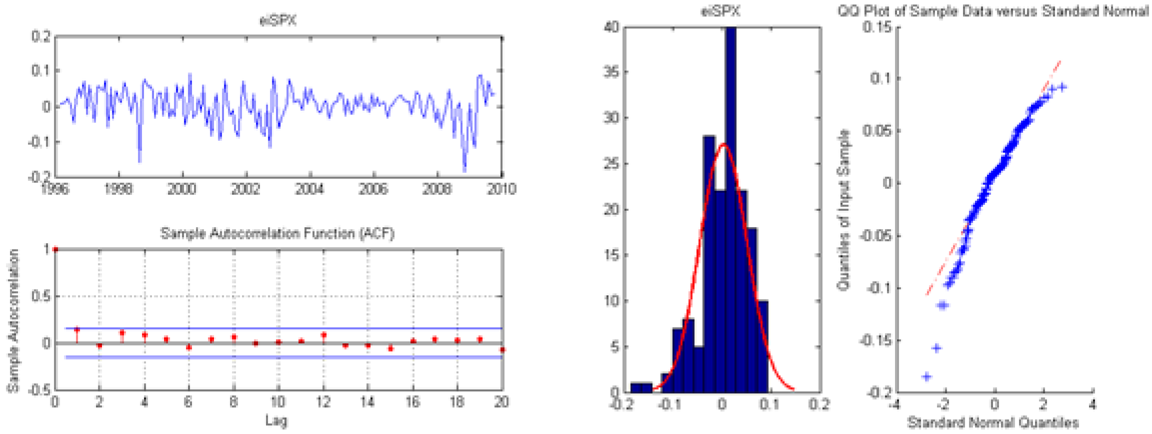

Fig 7 Test results for the logged SPX index returns. The figure shows historical development, sample ACF, histogram including a normal curve for comparison and normal qq-plot.

SPX index

Model	RMSE	IID				NID	AICC
		Ljung-Box	Turning point test	Difference - sign test	Rank test	Jarque-Bera test	
Data – initial analysis	-	0	0	0	0	1	-
ARMA(0,0) - Gaussian	0,05253	0	0	0	0	1	-
ARMA(0,0) - GARCH-Gaussian	0,05277	0	0	0	0	1	-
ARMA(0,0) - T(3)	0,05311	0	0	0	0	1	-
VAR(1) - VAR full & Q full	0,05253	0	0	0	0	1	-2065
VAR(1) - VAR full & Q diag	0,05253	0	0	0	0	1	-1468
VAR(1) - VAR diag & Q full	0,05253	0	0	0	0	1	-2046

Table 7 Performance of best models measured in forecasting performance (RMSE) and residual behaviour (Tests for IID and NID) as well as initial analysis of the time series being modelled. Forecasting horizon is 12 months ahead and tests for independences are significant at level 95%.

After testing down several ARMA(p,q) models, where p and q are run between 0 and 6 are now fitted to the data assuming Gaussian, t(3) and GARCH(1,1)-Gaussian behaviour of the residuals, one can conclude that more complicated models do not convincingly outperform a simpler one. Test results are listed in Table 7. The best model turns out to be an ARMA(0,0) process providing RMSE= 0.053.

Including cross-variable effect between the indices by applying VAR model is the next step. The order of the VAR model is run between 1 and 8 and four model variations are tested. Not much is achieved by applying a higher order model and fitting indicates that a model that has both full VAR matrix and correlating error terms is the most accurate one.

By comparing the AICC values one can even conclude that there is more information stored in the full variance matrix than in the vector autoregressive part. Out-of-sample comparison of the models reveals none or little difference between the model variations and no improvements are made on the residuals' behaviour.

The S&P500 index is modelled as an ARMA(0,0) process with normally distributed error terms.

6.3.2 S&P Midcap Index

The analysis for the Midcap index data (MID) is done just like in the case of the SPX index. First step is to look at the sample ACF in Fig 8 and conclude that there is no obvious sign of autoregressive behaviour within the data. The tests described in chapter 2 support the assumption of IID time series but Jarque-Bera test rejects normality and there are signs of heteroscedasticity. Test results for the transformed data are included in Table 8 along with a comparison of the best models.

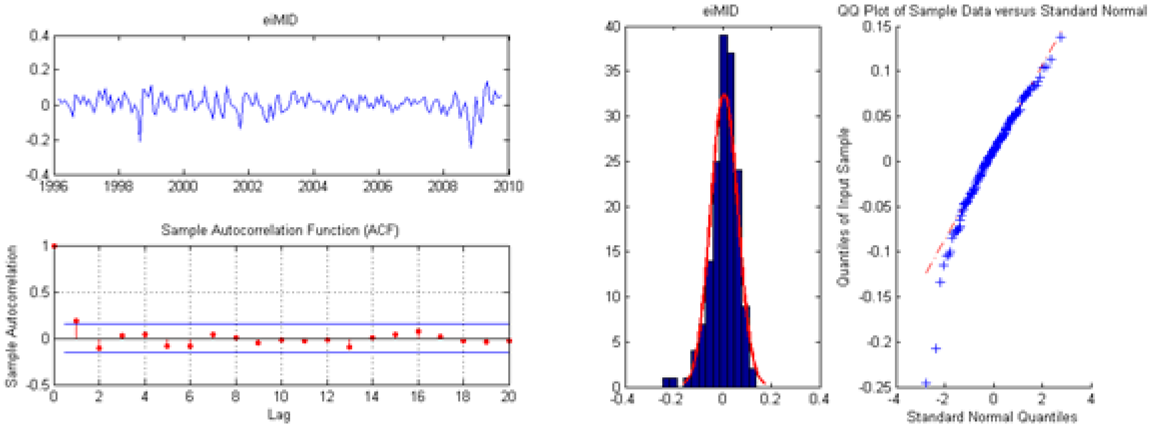

Fig 8 Test results for the logged MID index returns. The figure shows historical development, sample ACF, histogram including a normal curve for comparison and normal qq-plot.

MID index

Model	RMSE	IID				NID	AICC
		Ljung-Box	Turning point test	Difference - sign test	Rank test	Jarque-Bera test	
Data – initial analysis	-	0	0	0	0	1	-
ARMA(0,0) - Gaussian	0,06356	0	0	0	0	1	-
ARMA(0,0) - GARCH-Gaussian	0,06377	0	0	0	0	1	-
ARMA(0,0) - T(3)	0,06359	0	0	0	0	1	-
VAR(1) - VAR full & Q full	0,06356	0	0	0	0	1	-2065
VAR(1) - VAR full & Q diag	0,06356	0	0	0	0	1	-1468
VAR(1) - VAR diag & Q full	0,06356	0	0	0	0	1	-2046

Table 8 Performance of best models measured in forecasting performance (RMSE) and residual behaviour (Tests for IID and NID) as well as initial analysis of the time series being modelled. Forecasting horizon is 12 months ahead and tests for independences are significant at level 95%.

Now, a variety of ARMA models is fitted to the MID series. Lead-lag factors p and q are run between 0 and 6 and the distribution of error terms is assumed to be either Gaussian, t(3) or GARCH(1,1)-Gaussian. Just like expected, random walk performs just as well or even better than more complex autoregressive models. Since neither out-of-sample performance nor behaviour of residuals is improved by assuming other distribution than the normal one for the error terms, there is little reason to opt for other model than the Gaussian ARMA(0,0) model with RMSE = 0.064.

Testing for cross-index effects between MID series and the other equity indices is done in exactly the same way as described earlier when analysing the SPX series. And, as in the case of the SPX data, one concludes that a VAR model of first order with all parameters set provides best fit to data. AICC values indicate that a full Q matrix is more important than a full VAR matrix. Out-of-sample comparison of the models reveals that there is very little difference between them and no model outperforms the much simpler ARMA model.

6.3.3 S&P Smallcap Index

The sample ACF for Smallcap index data does not raise suspicion of underlying autoregressive process (Fig 9). The series pass all IID tests listed in chapter 2.1.2 but Jarque-Bera test rejects normally distributed observations.

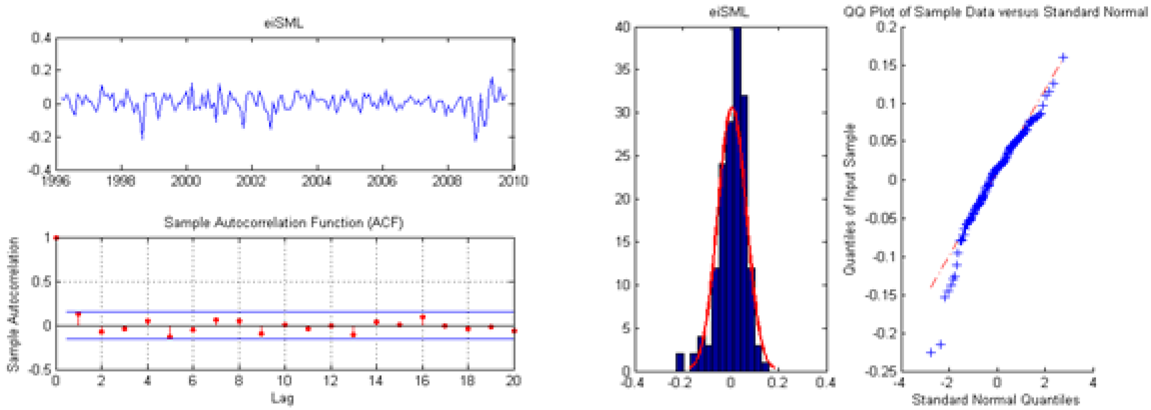

Fig 9 Test results for the logged SML index returns. The figure shows historical development, sample ACF, histogram including a normal curve for comparison and normal qq-plot.

SML index

Model	RMSE	IID				NID	AICC
		Ljung-Box	Turning point test	Difference - sign test	Rank test	Jarque-Bera test	
Data – initial analysis	-	0	0	0	0	1	-
ARMA(0,0) - Gaussian	0,06662	0	0	0	0	1	-
ARMA(0,0) - GARCH-Gaussian	0,06706	0	0	0	0	1	-
ARMA(0,0) - T(3)	0,06674	0	0	0	0	1	-
VAR(1) - VAR full & Q full	0,06663	0	0	1	0	1	-2065
VAR(1) - VAR full & Q diag	0,06663	0	0	1	0	1	-1468
VAR(1) - VAR diag & Q full	0,06663	0	0	0	0	1	-2046

Table 9 Performance of best models measured in forecasting performance (RMSE) and residual behaviour (Tests for IID and NID) as well as initial analysis of the time series being modelled. Forecasting horizon is 12 months ahead and tests for independences are significant at level 95%.

Fitting ARMA models in same manner as for the other indices brings familiar results. Since residual behaviour and out-of-sample performance is not improved using more complex models or more specific distribution of error terms there is no reason for preferring other model than Gaussian ARMA(0,0) with RMSE = 0.067 (see results in Table 9 above).

To check for cross-effect from other indices the VAR model is used with model order ranging from 1 to 8 and both VAR- and Q-matrices are allowed to be either full or diagonal. The Smallcap series seem to be relatively uncorrelated with SPX and MID indices and the best model is a full VAR(1) process. AICC values indicate that more information is contained in Q than in the vector-autoregression part of the model. The best VAR models do not outdo the chosen univariate model.

6.3.4 Summary and suggested improvements

Aall three equity indices are best modelled individually as random walks, ignoring any dependencies between the series; just as predicted by Akgiray (1989).

The residuals of the fitted random walk model can be assumed to be IID but show sign of being heteroscedastic. Heteroscedasticity is a known quality of single stock movements and it is logical to expect that this property of stocks is inherited by the stock indices. Nevertheless, allowing for GARCH behaviour does not improve equity index models.

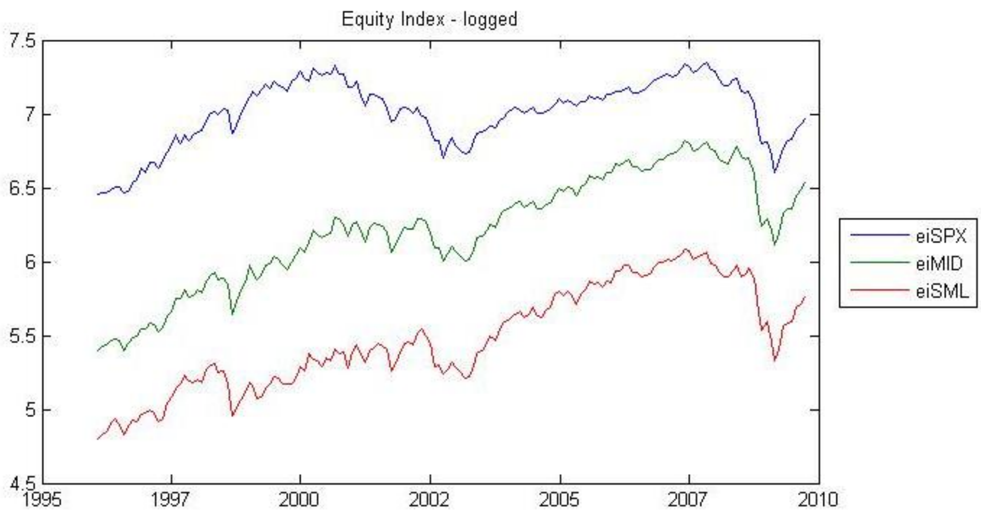

Fig 10 Plot shows the development of the logged equity indices.

A quick look at the historical development of the logged indices plotted together in Fig 10 reveals that the series look cointegrated or, at least, highly correlated. This seems obvious from data, and is supported by the correlation factors in Table 6, but the forecasting performance of the VAR model turned out to be only just as good as the ARMA model and results indicated that the indices should be modelled as three individual and uncorrelated series.

It is therefore justified to spend more time on testing for the possibility of cointegration within the equity index data series.

Other possible area of improvement lies in testing the effect of other domestic variables on the equity indices. One might expect more sensitivity toward changes in the economy in smaller businesses than bigger and therefore expect more prominent cross-variable dependencies in the Smallcap index than in S&P500 index.

6.4 FX rates

Data: Monthly FX data in form of historical EUR, GDP and JPY series is used. Series start in January 1999 and end in September 2009.

In sample estimation: First the whole data window is used to estimate how models fit data. Promising models are then tested in an out-of-sample comparison. AICC value and residuals behaviour are used as measure of fit.

Out-of-sample estimation: The first forecasting window is from Jan1999-Dec2004. Window is then widened – one month at time – until forecasting window is the same size as Jan1996-Dec2004. After that, window size is constant rolling one month at time until March 2009 is reached. For each window, model is estimated for window data and forecasting 12 months ahead is made. Performance is measured in RMSE of forecasting error.

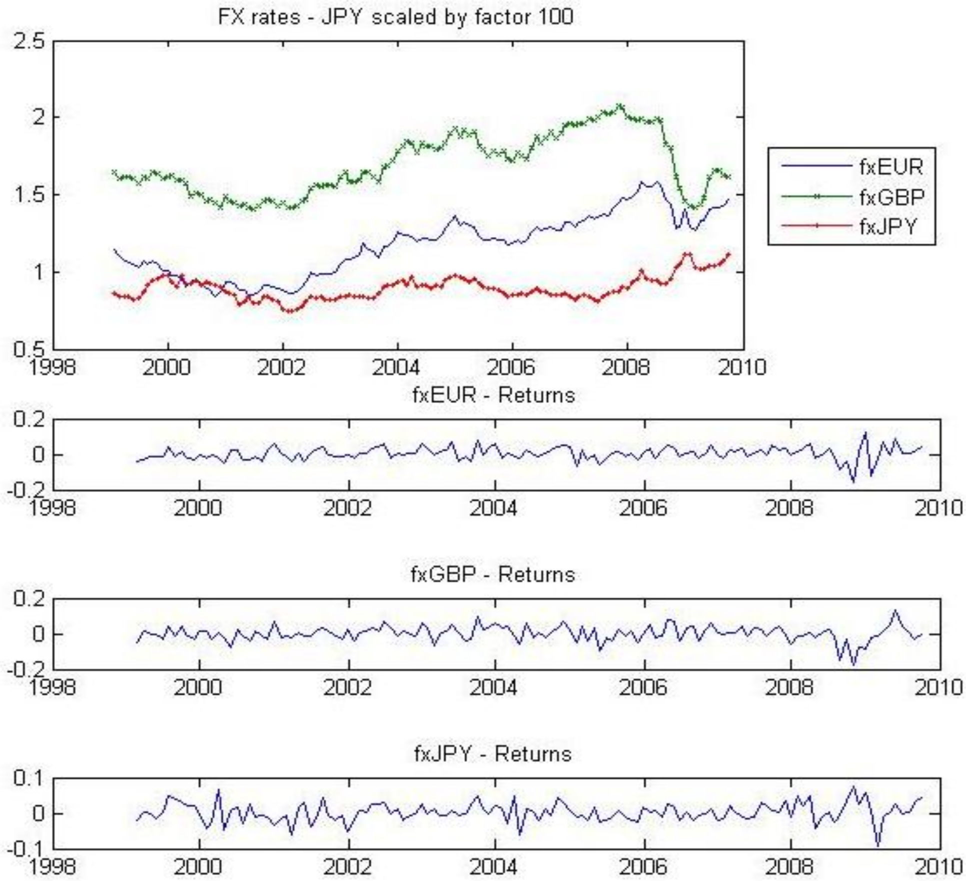

Fig 11 Plot at the top shows the historical development of FX rates EUR, GBP and JPY. Plots below show the derived series that will be modelled. In the case of FX rates, JPY is scaled up by factor 100 before all series are differentiated.

Initial analysis: Prior to fitting a model to the FX rates the JPY series is scaled up by factor 100 to match GDP and EUR series. Both the historical development of the indices as well as transformed series is shown in Fig 11. ADF test indicates that all series are I(1) – data series are non-stationary while differentiated data is stationary. Each index is analysed individually in chapters 6.4.1, 6.4.2 and 6.4.3.

Correlation coefficients

	FX-EUR	FX-GBP	FX-JPY
FX-EUR	1,00000	0,64036	0,27913
FX-GBP	0,64036	1,00000	0,10007
FX-JPY	0,27913	0,10007	1,00000

Table 10 Correlation coefficient matrix for FX-rates

The correlation coefficient matrix (shown in Table 10) indicates that the JPY evolves almost independently of the other currencies whereas the Euro and the British Pound share a positive relationship. This is consistent with the special development of the Japanese market for the past decades as well as the close relations between the British market and the Euro area.

6.4.1 EUR

When looking at the test results one can quickly assume that there is little reason to believe that an autoregressive process should be an appropriate model for the development of the Euro. The sample ACF in Fig 12 shows no signs of autoregression and when testing for IID, the assumption is accepted at level 0.05 with Ljung-Box turning point test but rejected with both difference-sign test and rank test. Jarque-Bera test rejects normality.

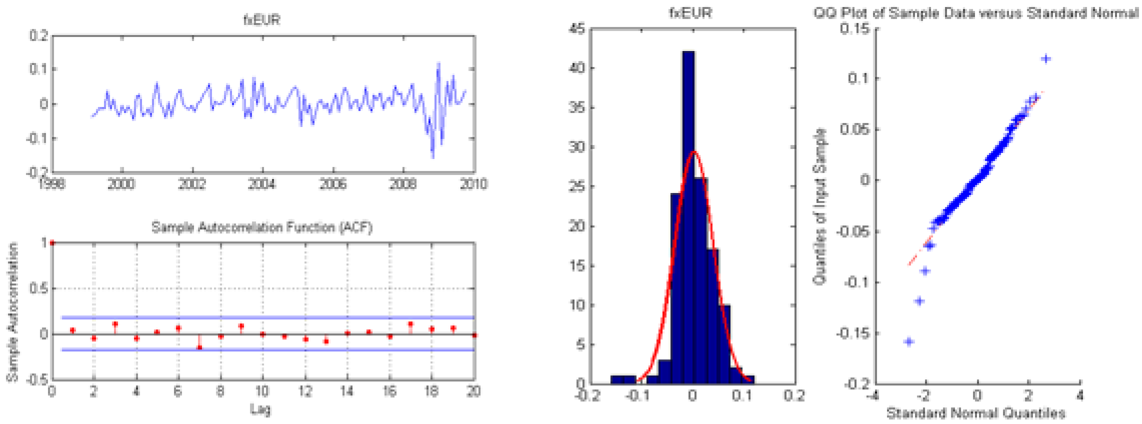

Fig 12 Test results for the rate changes of the EURO. The figure shows historical development, sample ACF, histogram including a normal curve for comparison and normal qq-plot.

EUR

Model	RMSE	IID				NID	AICC
		Ljung-Box	Turning point test	Difference - sign test	Rank test	Jarque-Bera test	
Data – initial analysis	-	0	0	1	1	1	-
ARMA(0,0) - Gaussian	0,04604	0	0	1	1	1	-
ARMA(0,0) - GARCH-Gaussian	0,04604	0	0	1	1	1	-
ARMA(0,0) - T(3)	0,04616	0	0	1	1	1	-
VAR(1) - VAR full & Q full	0,04623	0	0	1	1	1	-1555
VAR(1) - VAR full & Q diag	0,04623	0	0	1	1	1	-1473
VAR(1) - VAR diag & Q full	0,04623	0	0	1	1	1	-1544

Table 11 Performance of best models measured in forecasting performance (RMSE) and residual behaviour (Tests for IID and NID) as well as initial analysis of the time series being modelled. Forecasting horizon is 12 months ahead and tests for independences are significant at level 95%.

Although ARMA models do not appear to be appropriate for modelling the EUR, several such models are fitted to the data. The order parameters, p and q, are set to range between 0 and 6 while the error terms distribution is assumed Gaussian, t(3) or GARCH-Gaussian. Several levels for the GARCH model are tested.

The results are as anticipated and listed in Table 11. The ARMA(0,0) model provides the best fit and no convincing gain is made from assuming error terms distribution other than normal.

Fig 11 shows that the FX rates move in much similar ways raising questions about whether or not the series are correlated. To test for co-movements a VAR model is used. The order of the VAR models runs between and 8 and VAR and Q matrices are either full or diagonal.

The fully specified VAR(1) model provides best fit to data with more information stored in Q than in the VAR matrix. Out-of-sample comparison does not indicate that any of the VAR models outperform the ARMA(0, 0) and the residual behaviour is unchanged.

The simple ARMA(0, 0) is therefore chosen to model EUR rate with RMSE at 0.046.

6.4.2 GBP

The test results are even more distinct for the GDP series. All tests fail to reject IID and the sample ACF in Fig 13 does not indicate autoregression. Jarque-Bera test rejects normality at level 0.05.

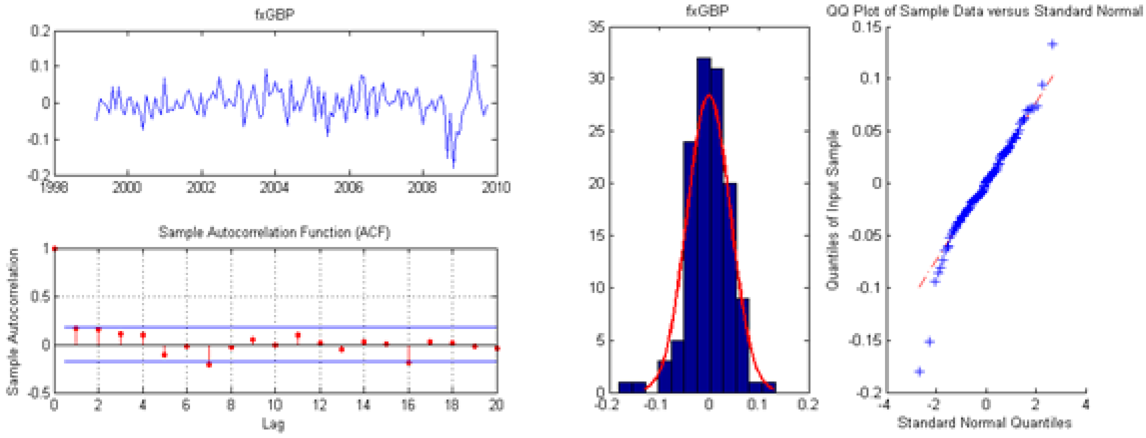

Fig 13 Test results for the rate changes of the Great British Pound. The figure shows historical development, sample ACF, histogram including a normal curve for comparison and normal qq-plot.

With same argumentation as for the Euro, same ARMA and VAR models are applied to fit the GDP series. Even higher levels of freedom are tested for the Student´s distribution of error terms in the ARMA models.

As expected, there is nothing to be gained from using higher order ARMA models or more complicated error assumptions than a normal distribution. ARMA(0,0) turns out to be the best model with no improvement on the residual behaviour.

GBP

Model	RMSE	IID				NID	AICC
		Ljung-Box	Turning point test	Difference - sign test	Rank test	Jarque-Bera test	
Data – initial analysis	-	0	0	0	0	1	-
ARMA(0,0) - Gaussian	0,05319	0	0	0	0	1	-
ARMA(0,0) - GARCH-Gaussian	0,05321	0	0	0	0	1	-
ARMA(0,0) - T(3)	0,05320	0	0	0	0	1	-
VAR(1) - VAR full & Q full	0,05239	0	0	0	0	1	-1555
VAR(1) - VAR full & Q diag	0,05239	0	0	0	0	1	-1473
VAR(1) - VAR diag & Q full	0,05237	0	0	0	0	1	-1544

Table 12 Performance of best models measured in forecasting performance (RMSE) and residual behaviour (Tests for IID and NID) as well as initial analysis of the time series being modelled. Forecasting horizon is 12 months ahead and tests for independences are significant at level 95%.

The same applies for the GBP as for the Euro when considering the cross-market effect. Although Fig 11 indicates that the series are highly correlated – especially EUR and GBP – the forecasting performance of the best VAR model that includes cross-market effects does not outperform the univariate simple random walk. Investigating the VAR models reveals that although there is no significant difference in out-of-sample performance of the three best VAR models, listed in Table 12, AICC values indicate that more information is stored in the correlation matrix Q than in the autocorrelation matrix VAR(1).

6.4.3 JPY

The initial tests on JPY indicate that a random walk is the best choice of model for the JPY. A tests confirm IID, Jarque-Bera tests does not reject normality and there is no hint of autoregression in the sample autocorrelation function (Fig 14).

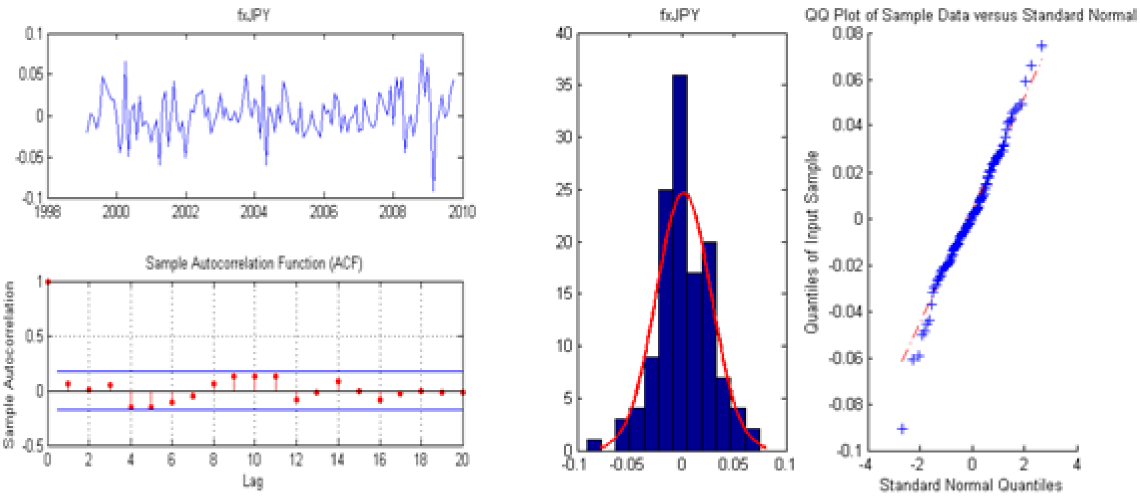

Fig 14 Test results for the rate changes of the scaled JPY. The figure shows historical development, sample ACF, histogram including a normal curve for comparison and normal qq-plot.

JPY

Model	RMSE	IID Ljung-Box	IID Turning point test	IID Difference - sign test	IID Rank test	NID Jarque-Bera test	AICC
Data – initial analysis	-	0	0	0	0	0	-
ARMA(0,0) - Gaussian	0,02820	0	0	0	0	0	-
ARMA(0,0) - GARCH-Gaussian	0,02821	0	0	0	0	0	-
ARMA(0,0) - T(3)	0,02819	0	0	0	0	0	-
VAR(1) - VAR full & Q full	0,02829	0	0	0	0	0	-1555
VAR(1) - VAR full & Q diag	0,02829	0	0	0	0	0	-1473
VAR(1) - VAR diag & Q full	0,02828	0	0	0	0	1	-1544

Table 13 Performance of best models measured in forecasting performance (RMSE) and residual behaviour (Tests for IID and NID) as well as initial analysis of the time series being modelled. Forecasting horizon is 12 months ahead and tests for independences are significant at level 95%.

Staying true to the methodology of testing down models all the same ARMA and VAR models as before are tested. Again, the best results are obtained by using a normal ARMA(0,0) or VAR models of first order where one or both VAR and Q matrices are full. From the cross-market investigation it is noted that a full Q matrix along with diagonal VAR(1) matrix provides a better fit to data than a full VAR(1) with diagonal Q. Since the goal is to find a model that is both appropriate and simple the random walk model is the best choice. Results are listed in Table 13.

6.4.4 Summary and suggested improvements

Consistent with known findings from earlier research done on FX rates, model comparison reveals no improvement made from applying complex models in favour of the simple normal random walk. Tests confirm that all three currencies return series are IID and, in the case of JPY, even normal. Including cross-variable information in a VAR model or by including cointegration does not lead to better results.

Nevertheless, by looking at the data in Fig 11 one can see that the series seem to evolve and react in much similar manner. GBP and EUR are obviously correlated whereas JPY follows behind with a lag. The relations between the British Pound and the Euro are also supported with a high correlation factor.

This could mean that although the currencies do not react directly to each other, they react to the same information with increase or decrease in variation. Since all three currencies are measured in US dollar this information might be connected to the US market. Testing against the yields or inflation is therefore of interest as well as testing for cointegration between the FX series.

Adding more currencies to the model as proposed by Carriero, Kapetanios, & Marcellino (2009) could possibly improve the forecasting accuracy.

6.5 Treasury rates – Yields

Data: Monthly zero coupon treasury rates from Federal Reserve are used with maturities 3M, 6M, 1Y, 2Y, 3Y, 5Y, 7Y & 10Y. Series start in January 1996 and end in September 2009.

In sample estimation: First the whole data window is used to estimate how models fit data. Promising models are then tested in an out-of-sample comparison. AICC value and residuals behaviour are used as measure of fit.

Out-of-sample estimation: The first forecasting window is Jan1996-Dec2004. After that, window size is constant rolling one month at time until March 2009 is reached. For each window, model is estimated for window data and forecasting 1, 3 & 6 months ahead is made. Performance is measured in RMSE of forecasting error.

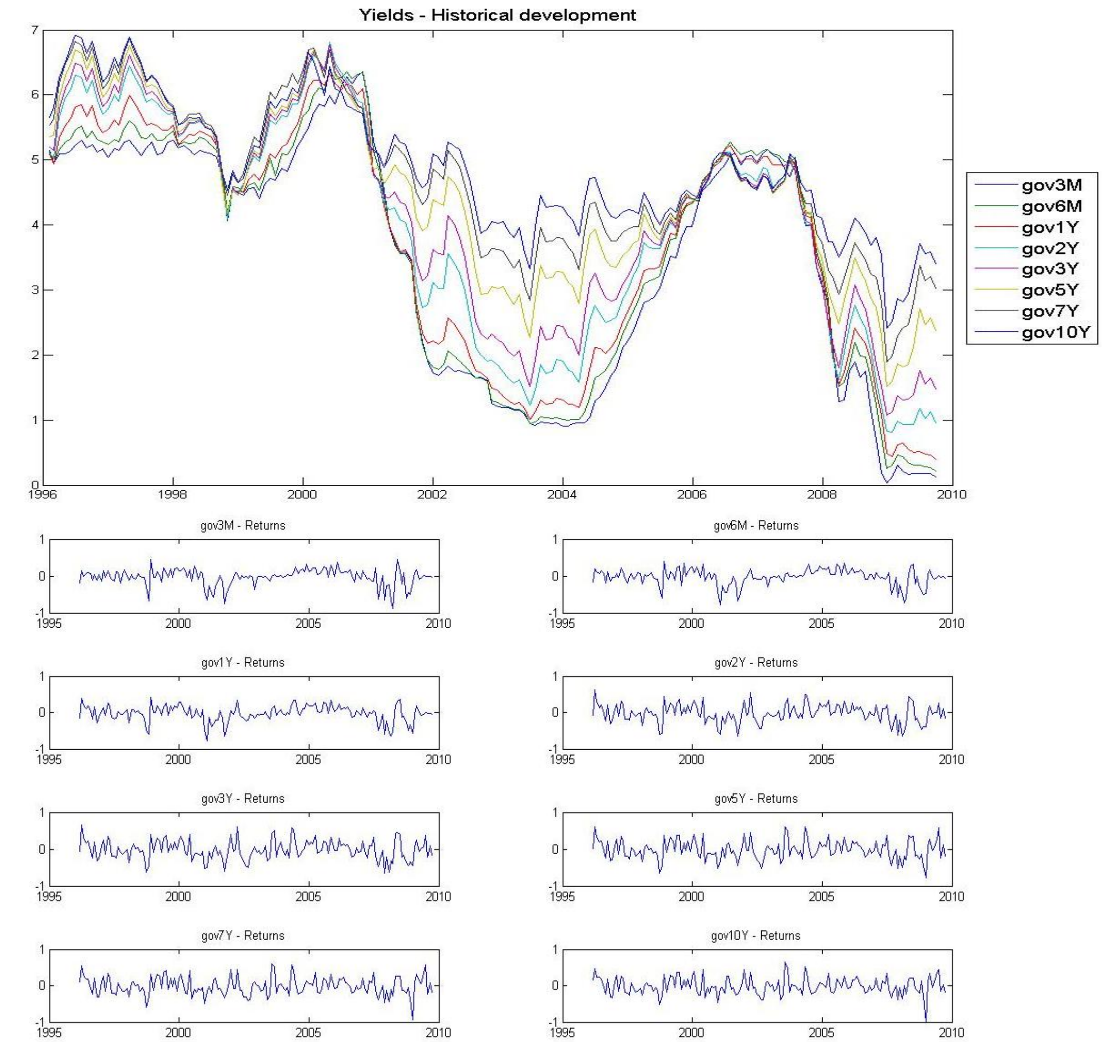

Fig 15 Plot at the top shows the historical development of treasury yields of maturities 3M, 6M, 1Y, 2Y, 3Y, 5Y, 7Y & 10Y. 3M series shows the largest variability whereas 10Y series is least variable. Plots below show the derived series that will be modelled. The derived series are obtained by differentiating the yields.

Initial analysis: Data is not transformed. Both the historical development of the indices as well as return series is shown in Fig 15. ADF-test indicates that all yields are I(1) – the original series are non-stationary while the differentiated time series are stationary.

Correlation coefficients

	3M	6M	1Y	2Y	3Y	5Y	7Y	10Y
3M	**1,00000**	**0,92979**	0,84311	0,68159	0,60350	0,49470	0,41987	0,34678
6M	**0,92979**	**1,00000**	**0,94635**	0,80871	0,73426	0,63991	0,57077	0,49929
1Y	0,84311	**0,94635**	**1,00000**	**0,92191**	0,86789	0,78196	0,71720	0,64546
2Y	0,68159	0,80871	**0,92191**	**1,00000**	**0,98474**	**0,92664**	0,87330	0,80135
3Y	0,60350	0,73426	0,86789	**0,98474**	**1,00000**	**0,96492**	**0,92468**	0,86137
5Y	0,49470	0,63991	0,78196	**0,92664**	**0,96492**	**1,00000**	**0,98509**	**0,95074**
7Y	0,41987	0,57077	0,71720	0,87330	**0,92468**	**0,98509**	**1,00000**	**0,97995**
10Y	0,34678	0,49929	0,64546	0,80135	0,86137	**0,95074**	**0,97995**	**1,00000**

Table 14 Correlation coefficient matrix for the yields. Values that exceed 0.9 are marked in bold.

From Table 14 one can see that the yields of different maturities are highly correlated. A closer look reveals a diagonal shape of correlation factors that exceed the value of 0.9, a revelation of the fact that the yields share stronger positive relations with maturities close to their own.

6.5.1 3M Yields

The sample ACF indicates autocorrelation in the 3 month yield return (Fig 16). Looking at the historical development of the data series one can see that there are periods of high volatility followed by periods of small volatility. An autocorrelating model allowing for GARCH–effects might correctly describe the time series. Ljung-Box test rejects IID and Jarque-Bera test normality at level 0.05.

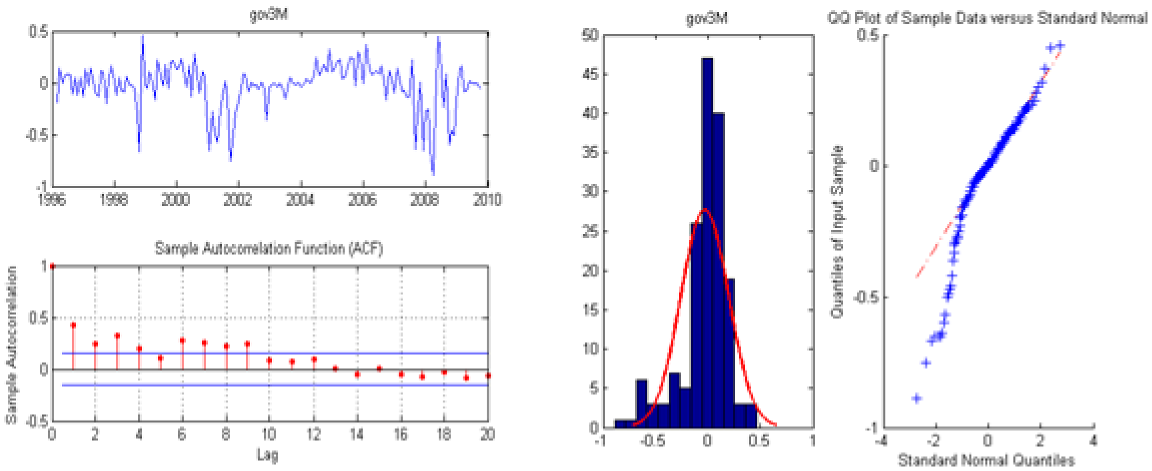

Fig 16 Test results for the 3M yield returns. The figure shows historical development, sample ACF, histogram including a normal curve for comparison and normal qq-plot.

To choose the ARMA model with best fit to data, order parameters p and q are run between 0 and 6 and the error term distribution is assumed Gaussian, t(3) or GARCH(1,1)-Gaussian.

It turns out that no tested ARMA model is good enough to be preferred over Random walk. ARMA(0,0) models gives the best results with unchanged residual behaviour for all assumptions of error term distribution. No significant difference between these models leads to favouring the most simple normal ARMA(0,0) model with RMSE = 0.2920.

3M Yield

Model	RMSE	IID				NID	AICC
		Ljung-Box	Turning point test	Difference - sign test	Rank test	Jarque-Bera test	
Data – initial analysis	-	1	0	0	0	1	-
ARMA(0,0) - Gaussian	0,29198	1	0	0	0	1	-
ARMA(0,0) - GARCH-Gaussian	0,30171	1	0	0	0	1	-
ARMA(0,0) - T(3)	0,30539	1	0	0	0	1	-
VAR(1) - VAR full & Q full	0,29152	1	0	0	0	1	-3493
VAR(1) - VAR full & Q diag	0,29152	1	0	0	0	1	-56
VAR(1) - VAR diag & Q full	0,29200	1	0	0	0	1	-3310

Table 15 Performance of best models measured in forecasting performance (RMSE) and residual behaviour (Tests for IID and NID) as well as initial analysis of the time series being modelled. Forecasting horizon is 12 months ahead and tests for independences are significant at level 95%.

From Fig 15 it is obvious that the yields are highly related. Several VAR models are therefore fitted to data in hope for more accurate modelling. The order of vector autoregression was tested for values between 1 and 8 and, in line with previous variable modelling for each level, four scenarios are tested. The VAR matrices and correlation matrix Q are set to be full or diagonal, respectively.

First order VAR models with at least one full VAR or Q matrix provide best fit to data where AICC values indicate that more information is stored in Q than in the VAR matrix (see Table 15). No VAR model improves the residual behaviour of the series. Ljung-Box does still reject IID at level 0.05 and Jarque-Bera rejects normality at same level.

The out-of-sample tests reveal little difference between the first order models and failure to outperform the random walk.

6.5.2 6M Yields

The historical plot of the 6 months yield returns (Fig 17) shows heteroscedastic behaviour with periods of high volatility squeezed between periods of low volatility. Sample ACF shows stronger signs of autoregression than spotted by 3M yields. The histogram and qq-plot do not match the patterns expected from a normally distributed set of observation and normality is rejected with Jarque-Bera test, just as anticipated. Ljung-Box rejects IID at level 0.05.

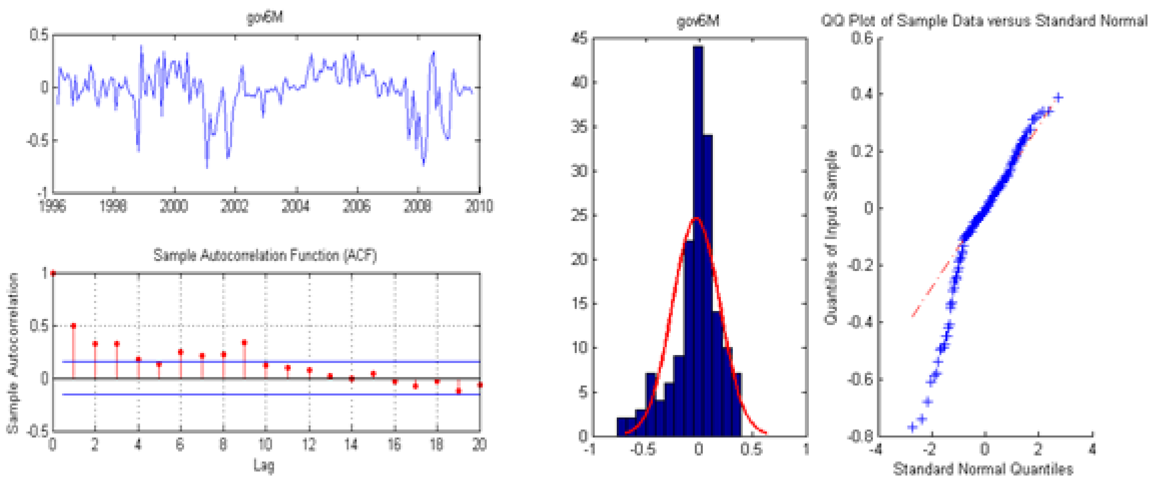

Fig 17 Test results for the 6M yield returns. The figure shows historical development, sample ACF, histogram including a normal curve for comparison and normal qq-plot.

6M Yield

Model	RMSE	IID				NID	AICC
		Ljung-Box	Turning point test	Difference - sign test	Rank test	Jarque-Bera test	
Data – initial analysis	-	1	0	0	0	1	-
ARMA(0,0) - Gaussian	0,26097	1	0	0	0	1	-
ARMA(0,0) - GARCH-Gaussian	0,27001	1	0	0	0	1	-
ARMA(0,0) - T(3)	0,27517	1	0	0	0	1	-
VAR(1) - VAR full & Q full	0,26048	1	0	0	0	1	-3493
VAR(1) - VAR full & Q diag	0,26048	1	0	0	0	1	-56
VAR(1) - VAR diag & Q full	0,26096	1	0	0	0	1	-3310

Table 16 Performance of best models measured in forecasting performance (RMSE) and residual behaviour (Tests for IID and NID) as well as initial analysis of the time series being modelled. Forecasting horizon is 12 months ahead and tests for independences are significant at level 95%.

Testing the same ARMA models as in the case of 3M yields brings out familiar results (shown in Table 16). The best results are obtained with ARMA(0,0) models and assumptions on error term distribution provide the same results in out-of-sample comparison with RMSE=0.2610. The residual behaviour remains the same regardless of error term distribution. Ljung-Box still rejects IID and Jarque-Bera rejects normality.

There is no reason for choosing anything other than a normal random walk to model 6M yield returns.

First order VAR models with at least one full VAR or Q matrix provide best fit to data with AICC values indicating that more information is stored in Q than in the VAR matrix. Residuals' behaviour is not improved from VAR. Ljung-Box and Jarque-Bera reject IID and normality at significance level 0.05.

Out-of-sample comparison reveals that none of the VAR models outclass the random walk.

6.5.3 1Y Yields

The sample autocorrelation function indicates a low order autocorrelation and there are signs of heteroscedasticity in the historical yield returns plot (Fig 18). Ljung-Box Portmanteau test rejects IID at level 0.05 and as does Jarque-Bera test for normality.

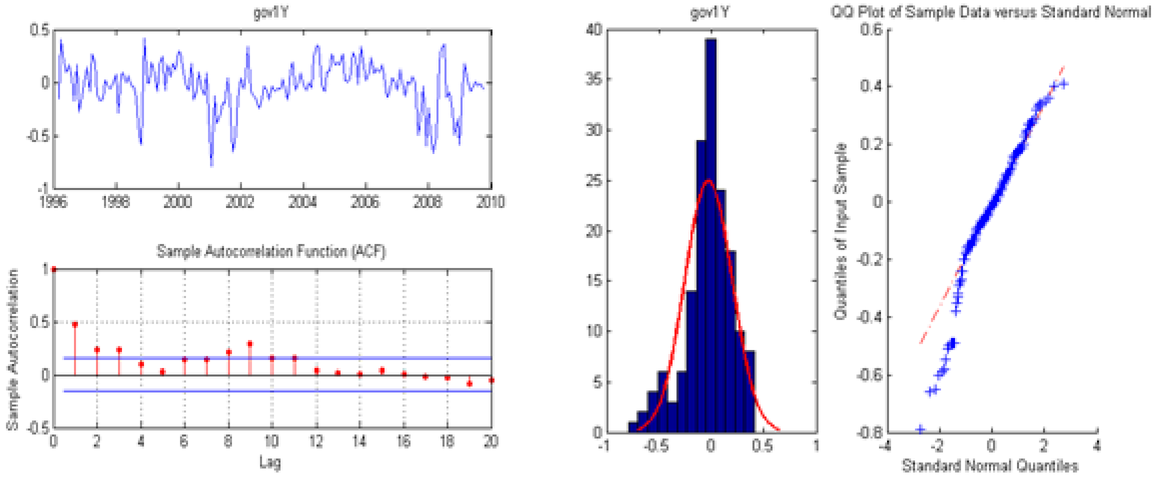

Fig 18 Test results for the 1Y yield returns. The figure shows historical development, sample ACF, histogram including a normal curve for comparison and normal qq-plot.

1Y Yield

		IID				NID	
Model	RMSE	Ljung-Box	Turning point test	Difference - sign test	Rank test	Jarque-Bera test	AICC
Data – initial analysis	-	1	0	0	0	1	-
ARMA(0,0) - Gaussian	0,25888	1	0	0	0	1	-
ARMA(0,0) - GARCH-Gaussian	0,26448	1	0	0	0	1	-
ARMA(0,0) - T(3)	0,26868	1	0	0	0	1	-
VAR(1) - VAR full & Q full	0,25852	0	0	0	0	1	-3493
VAR(1) - VAR full & Q diag	0,25852	0	0	0	0	1	-56
VAR(1) - VAR diag & Q full	0,25884	1	0	0	0	1	-3310

Table 17 Performance of best models measured in forecasting performance (RMSE) and residual behaviour (Tests for IID and NID) as well as initial analysis of the time series being modelled. Forecasting horizon is 12 months ahead and tests for independences are significant at level 95%.

When testing the ARMA procedures that where described in earlier cases, the most simple ARMA(0,0) cannot be beaten with RMSE at 0.2589. Ljung-Box test rejects IID and Jarque-Bera rejects normality (results are listed in Table 17). Since there is no improvement to be made from assuming more advanced error term distributions, neither in out-of-sample performance nor residual behaviour, the normal random walk stands out as the most reasonable choice of model.

First order VAR models with both full VAR matrix and correlating error terms is the most accurate model with AICC values indicating that more information is stored in Q than VAR matrix. Out-of-sample comparison reveals that there is almost no difference in forecasting performance between a first order VAR model with a full Q- and VAR matrix and a model with diagonal VAR matrix. Best model has RMSE=0,2585 – almost the same value as obtained with the univariate ARMA(0,0) model.

However, the residual behaviour is improved with cross-yield information. The residuals are now assumed to be IID since Ljung-Box fails to reject IID at significance level 0.05. Jarque-Bera test does still reject normality. Full VAR(1) model with full or diagonal covariance matrix is preferred over random walk.

6.5.4 2Y Yields

Signs of heteroscedasticity from the historical development of the 2 year yield returns are not as easily detectable as for the previous yields. From the histogram and qq-plot one can see that the distribution of the yield returns does coincide with the normal distribution (Fig 19). Indeed, Jarque-Bera fails to reject normality at level 0.05. Sample ACF indicates a low order autocorrelation and both Ljung-Box test and the Turning point test reject IID.

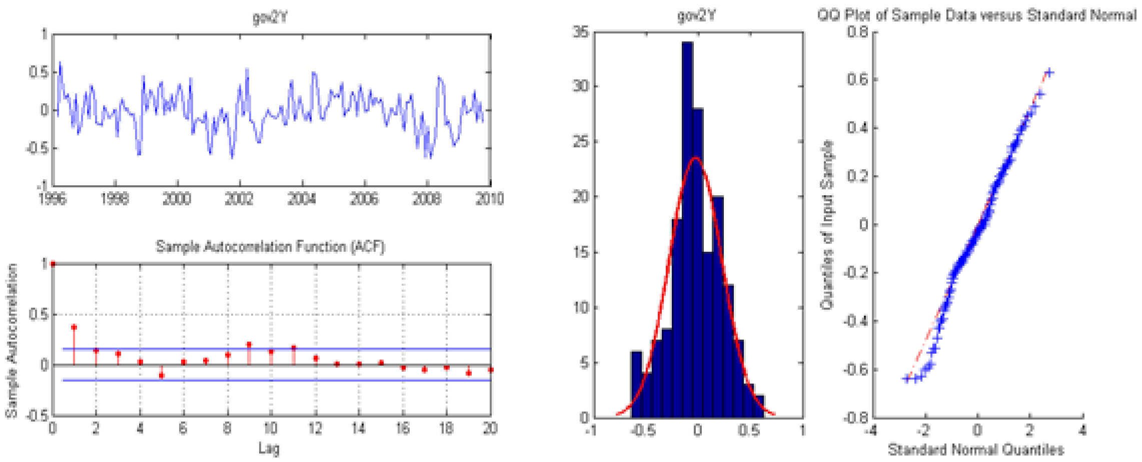

Fig 19 Test results for the 2Y yield returns. The figure shows historical development, sample ACF, histogram including a normal curve for comparison and normal qq-plot.

2Y Yield

Model	RMSE	IID				NID	AICC
		Ljung-Box	Turning point test	Difference - sign test	Rank test	Jarque-Bera test	
Data – initial analysis	-	1	1	0	0	0	-
ARMA(0,0) - Gaussian	0,26768	1	1	0	0	0	-
ARMA(0,0) - GARCH-Gaussian	0,26837	1	1	0	0	0	-
ARMA(0,0) - T(3)	0,27429	1	1	0	0	0	-
VAR(1) - VAR full & Q full	0,26730	0	0	0	0	0	-3493
VAR(1) - VAR full & Q diag	0,26730	0	0	0	0	0	-56
VAR(1) - VAR diag & Q full	0,26768	1	0	0	0	0	-3310

Table 18 Performance of best models measured in forecasting performance (RMSE) and residual behaviour (Tests for IID and NID) as well as initial analysis of the time series being modelled. Forecasting horizon is 12 months ahead and tests for independences are significant at level 95%.

The task of choosing the best ARMA model reveals a familiar result. Higher order models fail to outperform the simple ARMA(0,0) in out-of-sample comparison and, as anticipated, there is no gain from assuming other error term distribution than the normal one (see results in Table 18).

In the multivariate case, a first order model with both full VAR matrix and correlating error terms is the best choice with AICC values indicating that more information is stored in Q than VAR matrix.

Out-of-sample comparison results in RMSE=0.2673 for a first order model with full VAR(1) matrix and either full or diagonal Q matrix. The residuals pass the test of being NID. All test fail to reject IID and Jarque-Bera accepts normality at level 0.05.

Although there is no improvement in forecasting performance when cross-yields information is included in the model, the full VAR(1) model with full or diagonal Q is chosen over the simpler random walk due to improvement in residual behaviour.

6.5.5 3Y Yields

From the histogram and qq-plot in Fig 20 one can see that the 3 years yield returns look normally distributed while the historical plot shows vague signs of heteroscedasticity. Sample ACF indicates autoregression. Jarque-Bera test accepts normality and Ljung-Box rejects IID.

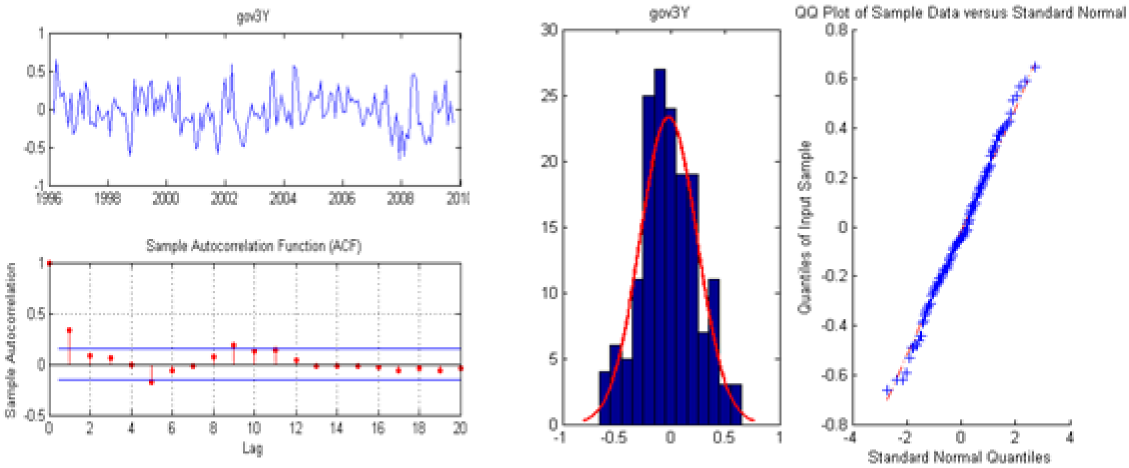

Fig 20 Test results for the 3Y yield returns. The figure shows historical development, sample ACF, histogram including a normal curve for comparison and normal qq-plot.

3Y Yield

Model	RMSE	IID				NID	AICC
		Ljung-Box	Turning point test	Difference - sign test	Rank test	Jarque-Bera test	
Data – initial analysis	-	1	0	0	0	0	-
ARMA(0,0) - Gaussian	0,27483	1	0	0	0	0	-
ARMA(0,0) - GARCH-Gaussian	0,27472	1	0	0	0	0	-
ARMA(0,0) - T(3)	0,27919	1	0	0	0	0	-
VAR(1) - VAR full & Q full	0,27450	0	0	0	0	0	-3493
VAR(1) - VAR full & Q diag	0,27450	0	0	0	0	0	-56
VAR(1) - VAR diag & Q full	0,27482	1	0	0	0	0	-3310

Table 19 Performance of best models measured in forecasting performance (RMSE) and residual behaviour (Tests for IID and NID) as well as initial analysis of the time series being modelled. Forecasting horizon is 12 months ahead and tests for independences are significant at level 95%.

Out-of-sample comparison points out the normal ARMA(0,0) model as the most appropriate model providing unchanged residual behaviour and RMSE=0.2748 (see Table 19). GARCH-Gaussian and t(3) assumptions of error term distribution do not lead to improved results.

Fitting a VAR model to data indicates that a first order model with both full VAR matrices and correlating error terms is the best choice with AICC values indicating that more information is stored in Q than in the VAR matrix. Out-of-sample comparison results in RMSE=0.2745 for a first order model with full VAR(1) matrix and either full or diagonal Q matrix. All tests accept IID and normality at level 0.05 providing NID residuals.

A vector autoregressive model is favoured over random walk due to the improved residual behaviour. Staying true to the philosophy of never choosing a more complicated model over a simpler one unless there are significant benefits, the full VAR(1) model with diagonal covariance matrix is picked.

6.5.6 5Y Yields

Comparing the basic analysis of the 5 year yield returns in Fig 21 to those of the 3 year data in Fig 20 reveals some obvious similarities between the two series. There are some indistinct signs of heteroscedasticity and autocorrelation and Ljung-Box rejects the notion of IID. The histogram and qq-plot agree with normally distributed data and normality is indeed accepted by Jarque-Bera test.

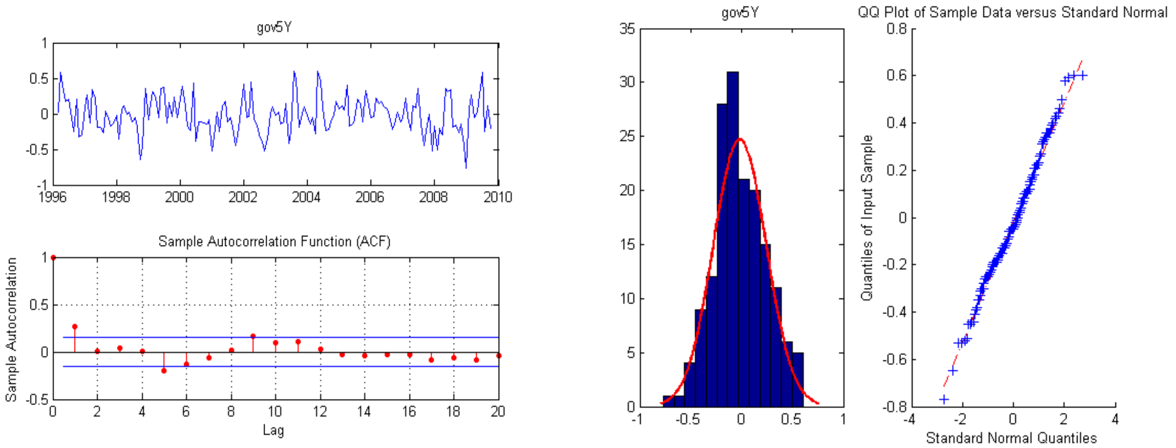

Fig 21 Test results for the 5Y yield returns. The figure shows historical development, sample ACF, histogram including a normal curve for comparison and normal qq-plot.

Out-of-sample comparison points out the normal ARMA(0,0) model as the most appropriate model providing unchanged residual behaviour and RMSE=0.2691 (see Table 20). GARCH-Gaussian and t(3) assumptions on error term distribution do not lead to improved results.

First order model with both full VAR matrices and correlating error terms is the best choice with AICC values indicating that more information is stored in Q than VAR matrix. Out-of-sample comparison results in RMSE=0.2688 for a first order model with full VAR(1) matrix and either full or diagonal Q matrix. All tests accept IID and normality at level 0.05 providing NID residuals.

5Y Yield

Model	RMSE	IID				NID	AICC
		Ljung-Box	Turning point test	Difference - sign test	Rank test	Jarque-Bera test	
Data – initial analysis	-	1	0	0	0	0	-
ARMA(0,0) - Gaussian	0,26911	1	0	0	0	0	-
ARMA(0,0) - GARCH-Gaussian	0,26896	1	0	0	0	0	-
ARMA(0,0) - T(3)	0,27203	1	0	0	0	0	-
VAR(1) - VAR full & Q full	0,26888	0	0	0	0	0	-3493
VAR(1) - VAR full & Q diag	0,26888	0	0	0	0	0	-56
VAR(1) - VAR diag & Q full	0,26913	1	0	0	0	0	-3310

Table 20 Performance of best models measured in forecasting performance (RMSE) and residual behaviour (Tests for IID and NID) as well as initial analysis of the time series being modelled. Forecasting horizon is 12 months ahead and tests for independences are significant at level 95%.

A vector autoregressive model is favoured over random walk due to the improved residual behaviour. Using same arguments as before the full VAR(1) model with diagonal covariance matrix is the preferred model for the 5 year yield returns.

6.5.7 7Y Yields

There is no sign of heteroscedasticity in the historical development of the 7 year yield returns, shown in Fig 22. Jarque-Bera test fails to reject normality at level 0.05 and both the histogram and qq-plot support assumption of normally distributed yield returns. Sample ACF is inconclusive.

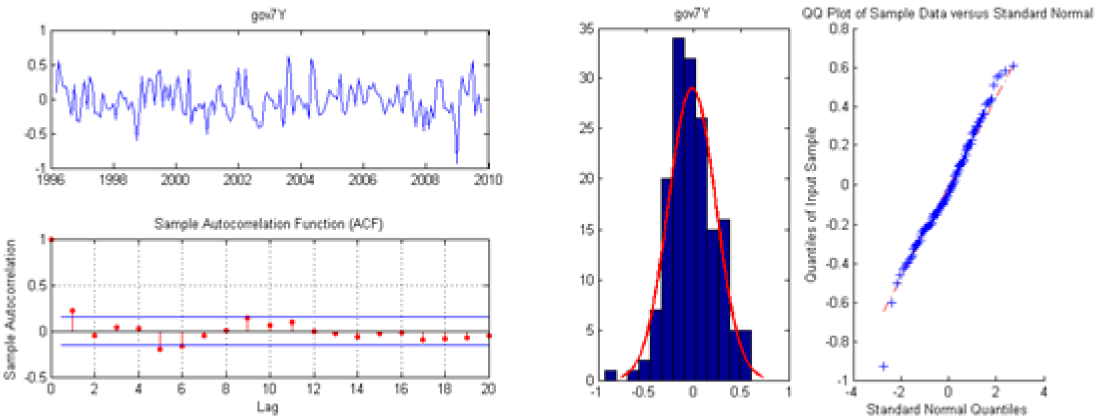

Fig 22 Test results for the 7Y yield returns. The figure shows historical development, sample ACF, histogram including a normal curve for comparison and normal qq-plot.

Out-of-sample comparison points out the normal ARMA(0,0) model as the most appropriate model providing unchanged residual behaviour and RMSE=0.2610 (see Table 21). GARCH-Gaussian and t(3) assumptions on error term distribution do not lead to improved results.

First order model with both full VAR matrices and correlating error terms is the best choice of model with AICC values indicating that more information is stored in Q than VAR matrix. Out-of-sample comparison results in RMSE=0.2608 for a first order model with full VAR(1) matrix and either full or diagonal Q matrix. All tests accept IID and normality at level 0.05 providing NID residuals.

7Y Yield

Model	RMSE	IID				NID	AICC
		Ljung-Box	Turning point test	Difference - sign test	Rank test	Jarque-Bera test	
Data – initial analysis	-	1	0	0	0	0	-
ARMA(0,0) - Gaussian	0,26099	1	0	0	0	0	-
ARMA(0,0) - GARCH-Gaussian	0,26082	1	0	0	0	0	-
ARMA(0,0) - T(3)	0,26189	1	0	0	0	0	-
VAR(1) - VAR full & Q full	0,26084	0	0	0	0	0	-3493
VAR(1) - VAR full & Q diag	0,26084	0	0	0	0	0	-56
VAR(1) - VAR diag & Q full	0,26101	1	1	0	0	0	-3310

Table 21 Performance of best models measured in forecasting performance (RMSE) and residual behaviour (Tests for IID and NID) as well as initial analysis of the time series being modelled. Forecasting horizon is 12 months ahead and tests for independences are significant at level 95%.

A vector autoregressive model is favoured over random walk due to the improved residual behaviour. The full VAR(1) model with diagonal covariance matrix is chosen to model the 7 year yield returns.

6.5.8 10Y Yields

Fig 23 shows the initial analysis of the 10 year yield returns. A visual inspection of the plot reveals no distinctive pattern within the historical movements of the time series. The sample ACF does not indicate autocorrelation and the data looks almost normal in the histogram and in the qq-plot. However, normality is rejected at significance level 0.05 with Jarque-Bera test. All tests accept IID. From these first findings it seems unlikely that the data will be modelled more accurately using a complex process than by simply applying random walk.

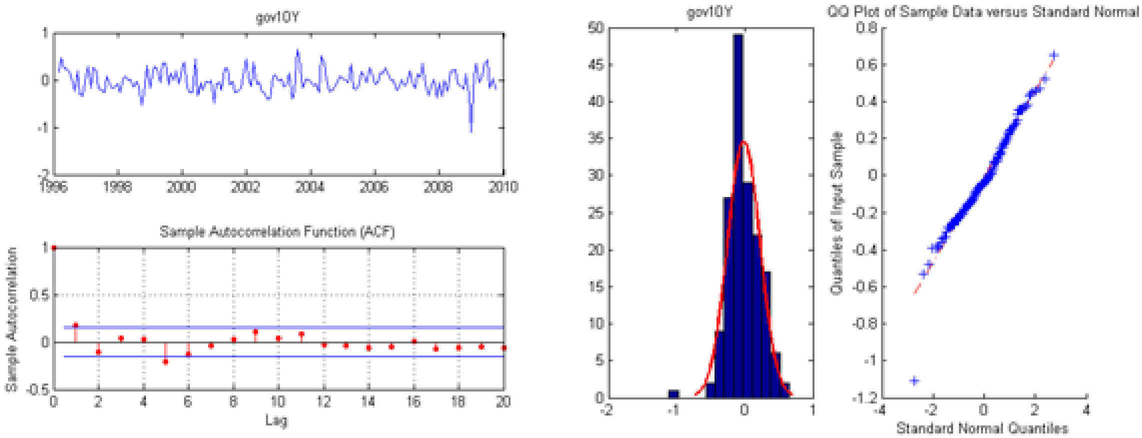

Fig 23 Test results for the 10Y yield returns. The figure shows historical development, sample ACF, histogram including a normal curve for comparison and normal qq-plot.

Much anticipated, the out-of-sample comparison points out the normal ARMA(0,0) model as the most appropriate with RMSE=0.2533. Jarque-Bera test still rejects normality at level 0.05. GARCH-Gaussian and t(3) assumptions on error term distribution do not lead to improved results.

10Y Yield

Model	RMSE	IID				NID	AICC
		Ljung-Box	Turning point test	Difference - sign test	Rank test	Jarque-Bera test	
Data – initial analysis	-	0	0	0	0	1	-
ARMA(0,0) - Gaussian	0,25330	0	0	0	0	1	-
ARMA(0,0) - GARCH-Gaussian	0,25345	0	0	0	0	1	-
ARMA(0,0) - T(3)	0,25378	0	0	0	0	1	-
VAR(1) - VAR full & Q full	0,25324	0	0	0	0	1	-3493
VAR(1) - VAR full & Q diag	0,25324	0	0	0	0	1	-56
VAR(1) - VAR diag & Q full	0,25332	0	0	0	0	1	-3310

Table 22 Performance of best models measured in forecasting performance (RMSE) and residual behaviour (Tests for IID and NID) as well as initial analysis of the time series being modelled. Forecasting horizon is 12 months ahead and tests for independences are significant at level 95%.

First order model with both full VAR matrices and correlating error terms are the best choice with more information stored in Q than VAR matrix. Out-of-sample comparison results in RMSE=0.2532 for a first order model with full VAR(1) matrix and either full or diagonal Q matrix. The residuals are IID but Jarque-Bera test rejects normality at level 0.05.

Since there are no observed improvements in either forecasting accuracy or residual behaviour, there is no apparent reason for abandoning the univariate random walk.

6.5.9 Summary and suggested improvements

One can now conclude that for five of eight yield series it is beneficial to include cross-yield information in the modelling. Three series - 3M, 6M and 10Y yields – are seemingly unaffected by other yields. Since the VAR model provides results for all maturities at the same time and modelling the 3M, 6M and 10Y yields separately as univariate series the effect would be that these three series were modelled twice. For simplicity, all eight yields are modelled together applying a full VAR(1) process with a diagonal covariance matrix, Q.

A quick look at the yield data in Fig 15 reveals that the data is highly correlated. Indeed, five series out of eight are improved by including cross-variable information in the modelling and the data seems to follow some common underlying process. This could indicate the presence of a cointegrating process within the yield series.

6.6 The yield curve

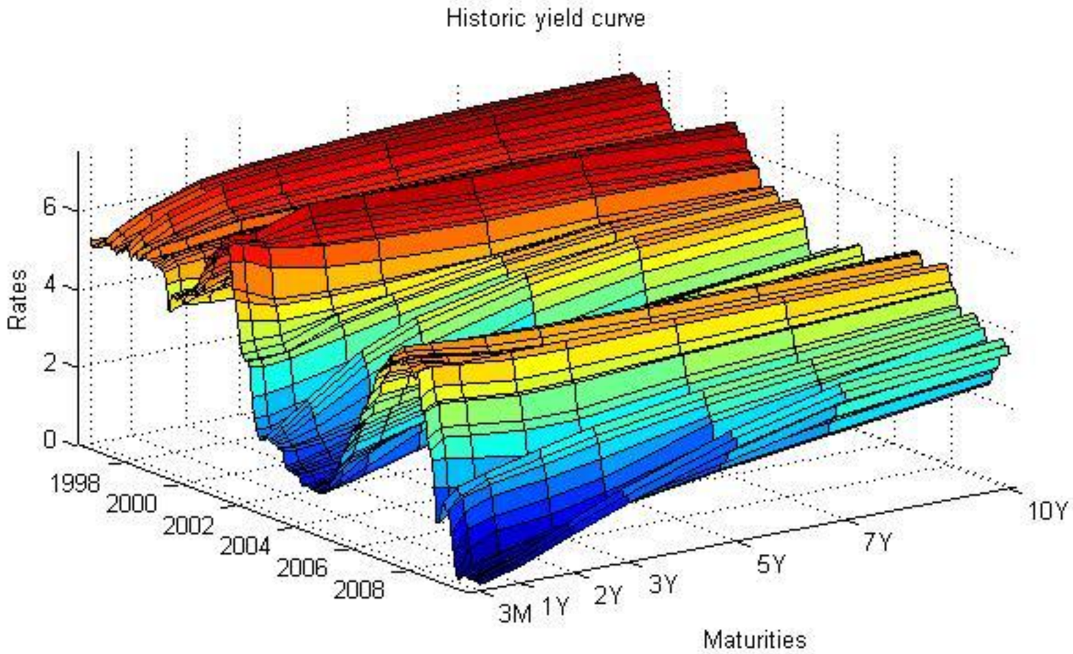

Fig 24 The historical development of the yield curve as a whole.

Fig 24 shoes the historical development of the whole yield curve in a 3-D plot. From the plot it is easy to identify the signature behaviour of a yield curve with the short term yields normally being lower than long term yields and with the rates stabilizing over higher maturities. One can also see how the gap between long term and short term yields increases during times of recession and decreasing during times of expansion, sometimes leading to the extreme situation of short term yields being more expensive than long term yields.

In order to model the whole yield curve, and thus obtain values for arbitrary yield maturities, two techniques are considered. One is to apply the findings from the previous chapters and simply interpolate a natural cubic spline between the directly forecasted yield maturities.

Yield returns modelled with a diagonal VAR(1) process and a full correlation matrix results in following out-of-sample RMSE values.

RMSE_VAR = 2.0152 1.9303 1.7779 1.5040 1.3301 1.0451 0.8605 0.6818

The other technique is to apply a Dynamic Nelson-Siegel model. The Nelson-Siegel equation for yields with maturities τ is defined as:

$$y_t(\tau) = L_t + S_t\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau}\right) + C_t\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right)$$

where L_t, S_t and C_t are constants that describe level, slope and curvature of the yield curve. The dynamic version proposed by Diebold & Li (2006) allows these three values to vary with time.

The Dynamic Nelson-Siegel model is applied by using the two-step approach from Diebold & Li (2006). To begin with, the time series for L_t, S_t and C_t are derived from the data sample of yields using OLS estimation. The factors are then modelled and predicted using a first order VAR process. Following Diebold&Li and setting $\lambda = 0.0609$ and estimating $L_t, S_t \& C_t$ from the data provides the predicted yield curve.

Modelling $L_t, S_t \& C_t$ as VAR(1) results in out-of-sample RMSE tabulated below:

$$RMSE\_DNS = 2.3199 \quad 2.1407 \quad 1.9988 \quad 1.8544 \quad 1.7343 \quad 1.4874 \quad 1.3153 \quad 1.1102$$

A comparison of the results from different yield curve models reveals that modelling yield returns as a fully determined VAR(1) model with diagonal correlation matrix, Q, provides better yields forecasting results than the two-step DNS model. The method of combining natural cubic spline with correlating yield returns is chosen as yield curve model.

6.6.1 Summary and suggested improvements

The yield curve is modelled as a whole using eight maturities as nodes in a natural cubic spline interpolation. The nodes are modelled as a diagonal VAR(1) process with a full covariance matrix, Q. The VAR(1) process is chosen for all eight maturities.

Improvements in modelling the nodes will improve the accuracy of the spline model. As described earlier in chapter 6.5.9 this could be achieved by accounting for cointegration between the yields and thus modelling the nodes with an error correction model instead of the simple VAR(1) model.

The maturities were chosen in order to maximise the performance of the dynamic Nelson-Siegel model that has now been discarded in favour of an approach inspired by Bowsher & Meeks (2008). It is therefore of interest to even further lean on that approach and maybe revaluate which maturities series are needed as nodes to model the yield curve.

Another possibility is to utilize the state-space representation of DNS defined in chapter 2.3.1. This is a more refined version of DNS modelling where the state-space system can be solved simultaneously and predicted directly using Kalman recursion. Such an approach would erase the model uncertainty connected to modelling in two steps instead of one.

7 Conclusions

The model comparison shows that it is not easy to find a model that outperforms the simple random walk. Vector autoregressive models improve forecasting performance of GDP and the model fit of the yield curve and a univariate ARMA process with one year AR lag is the best fit for inflation. Each equity index and FX rate is modelled independently as random walk.

There are several ways to potentially improve the macroeconomic multifactor model. Feasible actions for each variable were listed earlier in this paper and following suggestions refer to overall improvements of the multifactor model.

To start with, one can focus more on the relations between the variables and search for some co-movements. A study of the impulse responses of the system should then be done to confirm the established cross-variable relationships.

Careful investigation of the behaviour of the variables during and following market shocks is the next step. Improvements can be made by including dummy variables that are designed to respond to the occurrence of a certain scenario. A shock that simultaneously affects several variables can be modelled by temporarily allowing for higher correlation of the error terms. This could be achieved with a multidimensional GARCH model.

Including the economic cycle and allowing for a regime switching AR process as done by both Bergström, Holmlund, & Lindberg (2002) and Jonson (2008) is another approach that might possibly improve the results.

Finally, the model should be extended to include more international data. Some studies have indicated that, indeed, this could be the case. Diebold, Li, & Yue (2008) find that global yields improve the forecasting prediction of the yield curve and Carriero, Kapetanios, & Marcellino (2009) get encouraging results from modelling a panel of 33 exchange rates.

In this thesis I have made some initial model analysis and thoroughly examined the behaviour of the series of interest and compared the modelling results to random walk. In many cases I found a way to improve the model performance beyond random walk and thus both identify and utilize information affecting the data series. In other cases I conclude that there is not enough information present to justify applying a more advanced model than random walk.

The model I propose is ready to serve as basis for further analysis.

8 Acknowledgements

Thanks to Filip Lindskog and Fredrik Sandström for endless patience and support. Sara Jansson for wise comments and thoughts.

9 References

Akgiray, V. (1989). Conditional Heteroscedasticity in Time Series of Stock Returns: Evidence and Forecasts. *Journal of Business , 62*, 55-80.

Alexander, C. (2001). *Market Models: A Guide to Financial Data Analysis.* Chichester: John Wiley & Sons, Ltd.

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

Bergström, P., Holmlund, A., & Lindberg, S. (2002). *The SNDO's Simulation Model for Government Debt Analysis.* The Swedish National Debt Office.

Bowsher, C. G., & Meeks, R. (2008). The Dynamics of Economic Functions: Modelling and Forecasting the Yield Curve. *Journal of the American Statistical Association , 103* (484).

Brockwell, P. J., & Davis, R. A. (2002). *Introduction to Time Series and Forecasting.* New York: Springer Science.

Canova, F. (2007). G-7 Inflation forecasts: Random walk, Phillips curve or what else? *Macroeconomic Dynamics* , 1-30.

Carriero, Kapetanios, & Marcellino. (2009). Forecasting exchange rates with a large Bayesian VAR. *International Journal of Forecasting* , 400-417.

Cuaresma, J. C., & Hlouskova, J. (2004). Forecasting exchange rates in transition economies: A comparison of multivariate time series models. *Empirical Economics , 29*, 787–801.

Diebold, F. X., Li, C., & Yue, V. Z. (2008). Global yield curve dynamics and interactions: A dynamic Nelson-Siegel approach. *Journal of Econometrics , 146*, 351-363.

Diebold, F., & Li, C. (2006). Forecasting the term structure of government bond yields. *Journal of Econometrics* , 337-364.

Gonzáles, A., Hubrich, K., & Teräsvirta, T. (2009). *Forecasting inflation with gradual regime shifts and exogenous information.* CREATES Research Paper, Univesity of Aarhus, School of Economics and Management.

Hogrefe, J. (2007). *The yield spread and GDP growth - time varying leading properties and the role of monetary policy.* Working Paper, Christian-Albrechts-Universität Kiel, Department of Economics.

Johansen, S., & Juselius, K. (1990). Maximum Likelihood Estimation and Inference on Cointegration - With Applications to the Demand for Money. *Oxford Bulletin of Economics and Statistics* , 169-210.

Jonson, M. (2008). *Modellering och analys av beroende mellan riskslag.* Masters Thesis, KTH, Division of Mathematical Statistics.

Marcellino, M. (2008). A Linear Benchmark for Forecasting GDP Growth and Inflation? *Journal of Forecasting , 27*, 305-340.

Mckenzie, E., & Omran, M. (2000). Heteroscedasticity in stock returns data revisited: volume versus GARCH effects. *Applied Financial Economics , 10*, 553-560.

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