European Consortium for Mathematics in Industry

28th ECMI Modelling Week
Final Report

19.07.2015—26.07.2015
Lisboa, Portugal

TÉCNICO LISBOA

Departamento de Matemática
Universidade de Coimbra

Group 10

Mathematical Modelling of Financial Data in Many Dimensions

Astrid Vestergrd
DTU Compute, Technical University of Denmark,
Anker Engelunds Vej 1, Building 101A, 2800 Kgs. Lyngby , Denmark
Camara Mamadou,
UFR Mathématiques Informatique,
University of Strasbourg 7 Rue Rene Descartes 67084 Strasbourg, France
Matilde Lopes Rosa
Instituto Superior Técnico, University of Lisbon,
Rua dos Arneiros n.º 26 2.º A 1550-059 Lisboa, Portugal
Ramona Maraia
Department of Mathematics, Faculty of Science and Technology, Univesity of Milan, via Saldini 50 Milan, Italy
Velislav Bodurov
Faculty of Mathematics and Informatics, Sofia University,
5 James Bourchier Blvd., Sofia, Bulgaria

Instructor: Janusz Gajda
Department of Pure and Applied Mathematics, Wrocaw University of Technology
ul. Wybrzee Wyspiaskiego 27 50-370 Wrocaw, Poland

Abstract

Forecasting stock price movements is not an easy task. The data is complex and each set of prices for a stock may have dependencies, therefore financial data may be better modelled as multidimensional data. In this report the data has been considered as time series, multidimensional time series. Three sets of data have been considered and modelled, that is, the stock prices of Tech companies, the price of precious metals, and finally, the behavior of the Portuguese inflation rate, consumption, unemployment rate and real exchange rate.

We find that for the first data set, the movements are random after removing trends and a multivariate modelling approach is not the best suited for the problem. The second data set was more correlated and well suited for multivariate modelling, however the predictions did not do well. The third data was also well suited for multivariate modelling, and the final model was able to predict the data well for the first time period but not so well for the second, as the financial crisis was not taken into account in the data used for modelling.

10.1 Introduction

The aim of this project is to build a multivariate time series model to model real financial data. A time series is an ordered sequence of values of a variable at equally spaced time intervals, which is useful to see how a given variable changes over time or how it changes compared to other variables over the same time period. The idea is to find correlated variables and build a model to predict the future behavior of those variables by using the dependencies between them. We built three models for different data sets. First we present a financial model to predict stock prices of Tech companies but which turned out to be stochastic. Secondly, we show a model to predict the price of some precious metals (gold, silver, palladium and platinum). The last model aims to predict the behavior of several Portuguese macroeconomic variables (inflation, consumption, unemployment and real exchange rate). Finally, we discuss further work, especially regarding the two latter data sets.

10.2 Theory

VAR(p) model in the multivariate case

The vector autoregression (VAR) model is one of the most successful, flexible, and easy to use models for the analysis of multivariate time series. It is an extension of the univariate autoregressive model to dynamic multivariate time series. The VAR model has proven to be especially useful for describing the dynamic behavior of economic and financial time series as well as for forecasting.

The model captures the relationship between different time series in time, time series observations are available for variables of interest.

VAR models are a specific case of the more general VARMA models. VARMA models for multivariate time series include the VAR structure above along with moving average terms for each variable. More generally yet, these are special cases of ARMAX models that allow for the addition of other predictors that are outside the multivariate set of principal interest.

The structure is that each variable is a linear function of past lags of itself and past lags of the other variables. In general, for a VAR(p) model, the first p lags of each variable in the system would be used as regression predictors for each variable.

Let

$$Y_t = (y_{1t}, y_{2t}, ..., y_{nt})',$$

denote an $(n \times 1)$ vector of n different time series variables.

The basic p-lag VAR(p) model has the form

$$Y_t = c + \Pi_1 Y_{t-1} + \Pi_2 Y_{t-2} + \cdots + \Pi_p Y_{t-p} + \epsilon_t, \quad t = p, \ldots, T,$$

where Π_i are $(n \times n)$ coefficient matrices and ϵ_t is an $(n \times 1)$ unobservable zero mean white noise vector process (uncorrelated or independent) with time invariant covariance matrix Σ. ϵ_t is Gaussian white noise, $\epsilon_t \sim N(0, \Pi\epsilon)$ $\forall t$ and ϵ_t and ϵ_s are independent for $s \neq t$.

In economic time series, the white noise series is often thought of as representing innovations, or shocks. That is, ϵ_t represents those aspects of the time series of interest which could not have been predicted in advance. In lag operator notation, the VAR(p) is written as

$$\Pi(L)Y_t = c + \epsilon_t,$$

where $\Pi(L) = I_n - \Pi_1 z - \cdots - \Pi_p z^p$.

The VAR(p) is stable if the root of

$$det(I_n - \Pi_1 z - ... - \Pi_p z^p) = 0,$$

lies outside the complex unit circle (have modulus greater than one),or, equivalently, if the eigenvalues of the companion matrix

$$F = \begin{bmatrix} \Pi_1 & \Pi_2 & \ldots & \Pi_n \\ I_n & 0 & \ldots & 0 \\ 0 & \ddots & 0 & \vdots \\ 0 & 0 & I_n & 0 \end{bmatrix},$$

have modulus less than one.
Assuming that the process has been initialized in the infinite past, then a stable VAR(p) process is stationary and ergodic with time invariant means, variances, and autocovariances.

Parameter estimation

Given a sample of size T, $y_1, \ldots, y_T$, and p presample vectors, $y_{p-1}, \ldots, y_0$, the parameters can be estimated efficiently by ordinary least squares (OLS) for each equation separately.

The VAR model should include all variables which the theory indicates are relevant and we should choose the lag length which has a high likelihood of capturing all of the dynamics. Once these values have been set, either a general-to-specific search can be conducted or an information criteria (IC)

can be used to select the appropriate lag length.
In the VAR case, the main information criteria are:

Akaike information criterion (AIC) is a measure of the relative quality of statistical models. Given a set of candidate models for the data, the preferred model is the one with the minimum AIC value.

Bayesian information criterion (BIC) is a criterion for model selection among a finite set of models. The model with the lowest BIC is preferred; it is based, in part, on the likelihood function and it is closely related to the Akaike information criterion (AIC).

Hannan – Quinn information criterion (HQC) is a criterion for model selection. It is an alternative to the Akaike information criterion (AIC) and the Bayesian information criterion (BIC).

The lag length should be chosen to minimize one of these criteria, and the BIC will always choose a (weakly) smaller model than the HQC which in turn will select a (weakly) smaller model than the AIC.

Random Walk

A random walk is a mathematical formalization of a path that consists of a succession of random steps. Often, the walk is in discrete time, and indexed by the natural numbers, as in $X_0, X_1, X_2, \ldots$. However, some walks take their steps at random times, and in that case the position X_t is defined for the continuum of times $t \geq 0$. Specific cases or limits of random walks include the Levy flight. Random walks are related to the diffusion models and are a fundamental topic in discussions of Markov processes.

In economics, the "random walk hypothesis" is used to model shares prices and other factors. Empirical studies found some deviations from this theoretical model, especially in short term and long term correlations.

10.3 A financial model

The data used for this project has been gathered from Yahoo Finance. We started with 50 tech companies listed on NASDAQ from [5], and eliminated the companies that did not have data for the entire period we are considering, leaving 46 to be used. The chosen period is five years long and goes from 01/01/2010 to 31/12/2014 and the data is the adjusted daily closing prices. For this part of the data preparation we have used first Matlab to get the data from Yahoo and build a matrix with the stocks and the dates, and then we used R to analyse the data. In R we transformed the data into a time series object with the function `ts`.

After that we used the built-in R function `acf` to create a correlation matrix for all of the stocks and plotted the results. The `acf` function, which stands for Auto-Correlation Function, computes the correlation between each stock and the other stocks. We did this because we wished to use data that is highly correlated, in order to make a better model for forecasting future prices. In fact, the results showed that they were all highly correlated. However, by looking at the plotted time series we noticed they all had a clear trend so we decided to remove it by differencing the data. Then we filtered the correlation matrix by choosing the stocks with the highest absolute values of correlation (higher that 0.65) and also imposing the condition that the smallest values of the correlation should be above 0.05. This gave us 5 stocks, which are shown in Table 10.1. The plotted values

Companies	MA	PAYX	AMAT	ADP	TXN

Table 10.1: The final set of companies.

can be seen on Figure 10.1.

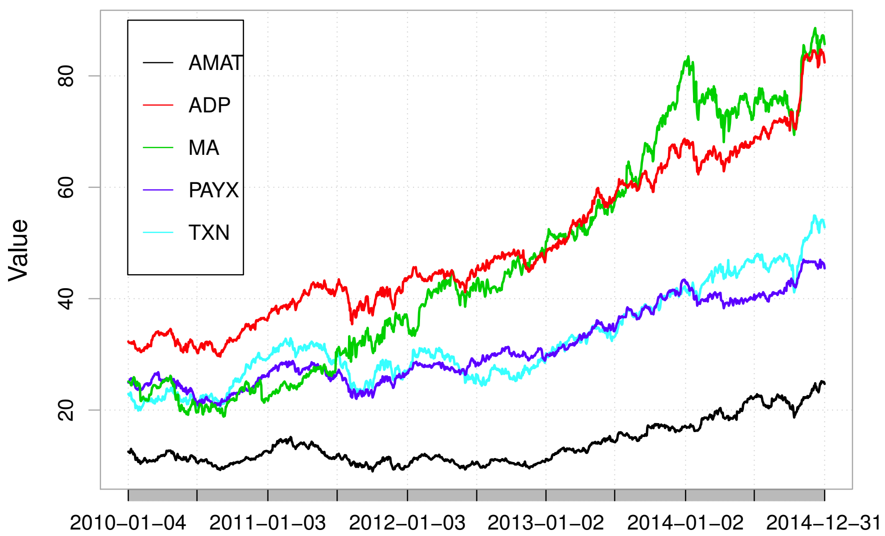

Figure 10.1: Stock values

Then we computed the acf again but this time there was almost no correlation.

Still, for simplicity first we tried to model the future behavior of just two of the company stocks. Our choice fell on MA and PAYX and what follows are the results for the autocorrelation of the two time series. On Figures 10.2 and 10.3 are shown the results before and after detrending the data.

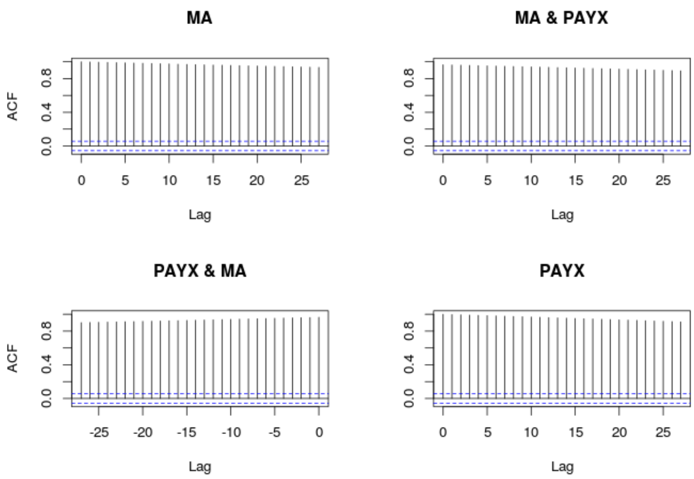

Figure 10.2: Autocorrelation plots for MA and PAYX before detrending

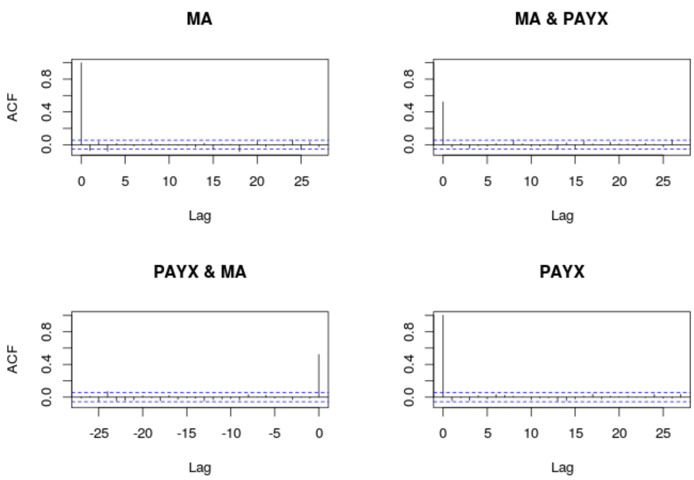

Figure 10.3: Autocorrelation plots for MA and PAYX after detrending

As there was no autocorrelation whatsoever (except at lag 0) we concluded that the behavior of the detrended series is completely random and should be modelled with a stochastic process.

We estimated the empirical cumulative distribution function and the empirical probability density function in order to try to find the distribution that the stocks follow. A plot of the empirical PDF of the ASML stock is shown and from it can be inferred that the distribution is actually a mixture of one-dimensional normal distributions. The same thing is generally true

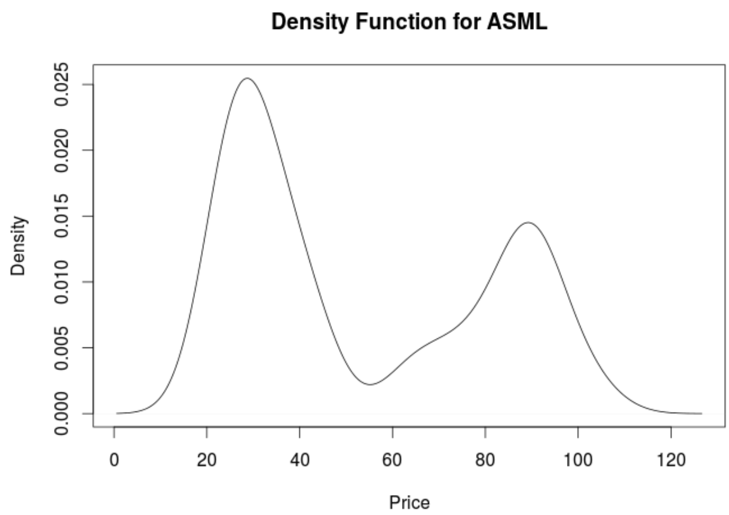

Figure 10.4: A density function for ASML stock.

for all the data we used and can be explored to make future predictions. Alas the natural choice for the task is a random walk model with drift. We used the built-in function `rwf` in R to make forecasts. The results for the ASML stock price are shown in Figure 10.5 below.

10.4 Modelling the Price of Precious Metals

The Data

Historically one of the most important (especially in times of economic downturn) commodities traded on the stock exchange are the precious metals. As such the behavior of their price is often an object of statistical modelling. In what follows we shall present an attempt at modelling the price of Gold, Silver, Palladium and Platinum on the NYSE using a VAR model and thus exploiting the relationship between their values.
We have collected data for the prices of those metals in US Dollars on the New York Stock Exchange from the beginning of 1995 till the middle of July 2015. For the actual modelling though we shall pick up only the data from 2010 to 2015 which contains about 1000 observations of the closing price for the commodity. The data was obtained from [6] and a plot can be seen on Figure 10.6.

From the plots it is quite obvious that the data doesn't have a constant expectation over time and even a constant variance. Since the VAR model

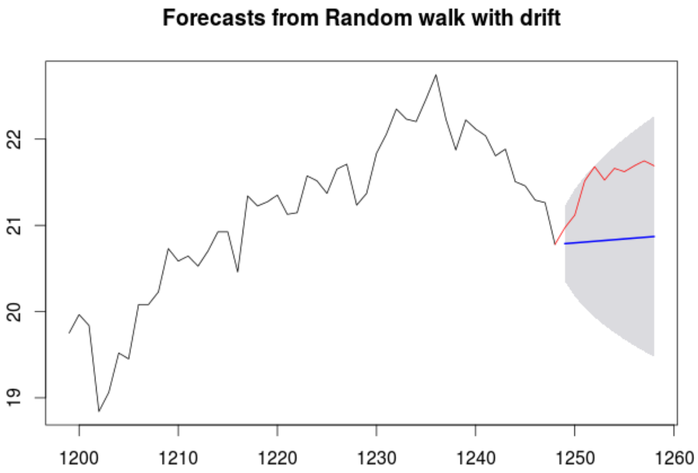

Figure 10.5: Random Walk Model for the ASML stock.

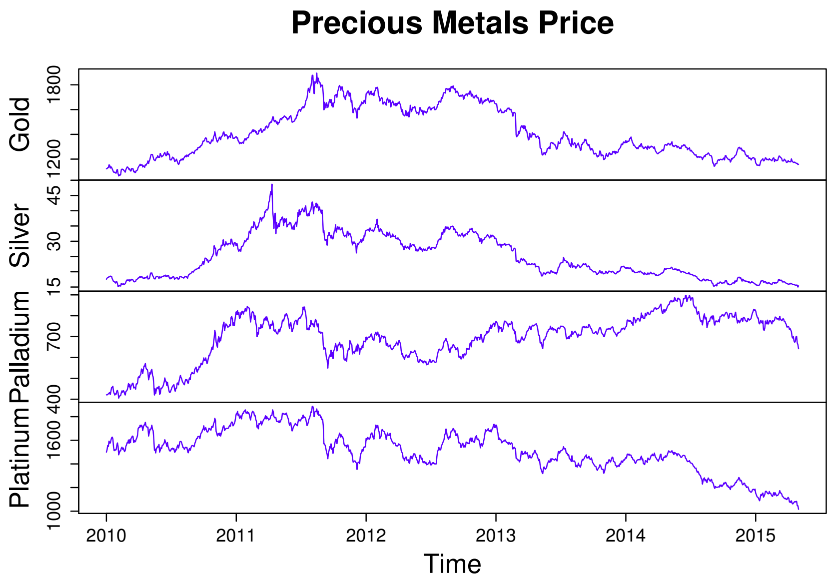

Figure 10.6: The time series from 2010 to 2015.

has stationarity as one of its main assumptions we can apply a transformation to the data (basically employing an integrated model). We difference the data once and these are the time series we shall work with from now on. It's worth mentioning that the data does not exhibit any seasonal effects.
As the price of gold (and respectively all other precious metals under con-

sideration) is highly influenced by the value of the US Dollar compared to other world currencies it seemed natural to try and incorporate that value into the model in some way. Overall though this turned out to be of almost no use as it did not improve the strength of our model in any meaningful way while introducing yet another set of parameters. For the sake of simplicity this idea was abandoned and we proceed to model only the prices of the four metals.

Model Selection

On Figure 10.7 the autocorrelation plots for the four time series can be observed.

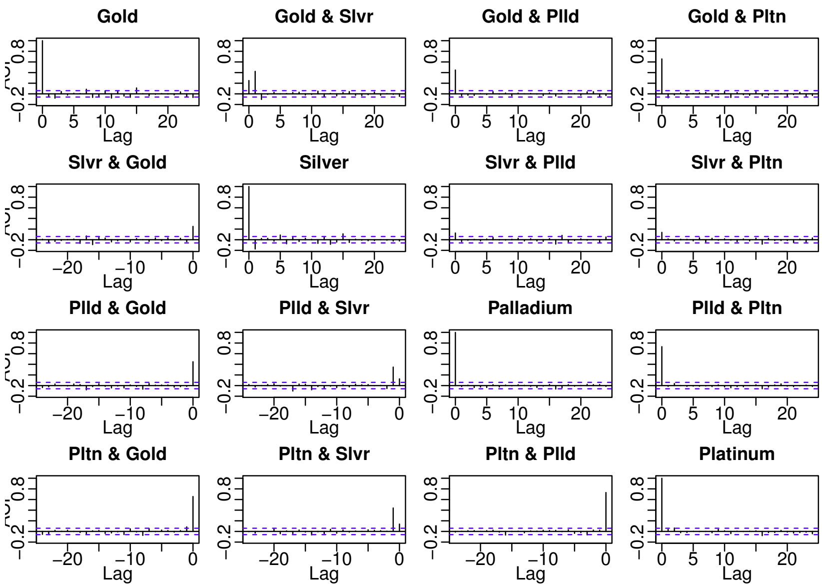

Figure 10.7: The time series from 2010 to 2015.

After applying the Dickey-Fuller test the following results were obtained: The low p-values suggest that the data is now stationary. The observed

	P-value	Test statistic
Gold Price	0.01	-11.262
Silver Price	0.01	-10.531
Palladium Price	0.01	-11.106
Platinum Price	0.01	-10.461

Table 10.2: Dickey - Fuller Test

results suggest that a two-lag VAR model is suitable.
Using the `VARselect` function from the `vars` package in `R` we can optimize the model selection based on four criteria. The following suggestions were given based on each criterion:

Criterion	Suggested Lag
AIC	2
HQ	2
SC	1
FPE	2

Table 10.3: Information Criteria

which once again confirms the output of the ACF plots.

The Model

We can write explicitly the form of the selected model:

$$\begin{pmatrix} Gold_t \\ Silver_t \\ Palladium_t \\ Platinum_t \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{pmatrix} + \Pi_1 \begin{pmatrix} Gold_{t-1} \\ Silver_{t-1} \\ Palladium_{t-1} \\ Platinum_{t-1} \end{pmatrix} + \Pi_2 \begin{pmatrix} Gold_{t-2} \\ Silver_{t-2} \\ Palladium_{t-2} \\ Platinum_{t-2} \end{pmatrix} + \begin{pmatrix} \varepsilon_t^{(1)} \\ \varepsilon_t^{(2)} \\ \varepsilon_t^{(3)} \\ \varepsilon_t^{(4)} \end{pmatrix},$$

where Π_1 and Π_2 are 4×4 matrices. Using the `VAR` function from the `vars` package in `R` we can fit a model to the data. The function uses ordinary least squares for fitting. The coefficient matrices of the model are as follows

$$c = \begin{pmatrix} 0.02 \\ 0.00 \\ 0.17 \\ -0.34 \end{pmatrix},$$

$$\Pi_1 = \begin{pmatrix} \color{red}{-0.14} & \color{red}{13.83} & \color{red}{-0.18} & 0.06 \\ 0.00 & \color{red}{-0.15} & 0.00 & 0.00 \\ \color{red}{-0.09} & \color{red}{8.17} & -0.04 & 0.03 \\ -0.01 & \color{red}{14.61} & -0.12 & 0.03 \end{pmatrix},$$

$$\Pi_2 = \begin{pmatrix} -0.06 & \color{red}{3.66} & 0.13 & -0.07 \\ 0.00 & 0.02 & 0.00 & 0.00 \\ 0.00 & 1.70 & 0.00 & 0.00 \\ 0.02 & \color{red}{2.91} & 0.02 & 0.01 \end{pmatrix},$$

where the coefficients significant at the 95% level are marked with red.

Predictions

The predictions are made with the `predict` function from the `vars` package. The last five observations from the data were removed so that they can be used for testing the model. The plots of the predicted vs. real values can be seen on 10.8.

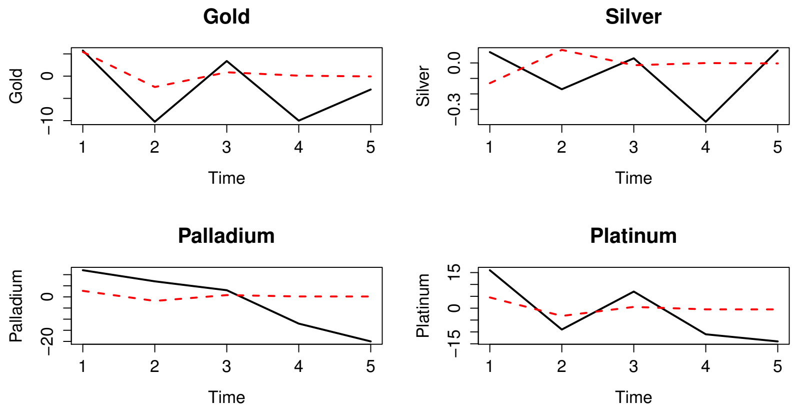

Figure 10.8: Predicted Values for five days in the future.

From the graphics above we can conclude that the model did not perform well in all of the cases and that prediction for more than two days into the future is a worthless endeavour (which is a drawback of the VAR model since the values tend to go to a constant value very fast).

10.5 A macro-economical analysis of Portugal

The Data

The goal is to build a model to forecast 4 macroeconomic variables for Portugal, based on time series with 1 year as frequency, from 1984 and up until 2009, as we have not been able to find a longer dataset. The variables we have chosen are the inflation rate, the consumer consumption in total, the real effective exchange rate, and the unemployment rate. The time series are plotted in Figure 10.9.

We see the scales are quite different, especially is the consumption large as it is the only non-rate series, therefore we start by normalising the data respectively to the series to be able to compare the rates and the consumption.

To see if they are correlated enough for us to be able to build and multivariate time series model, we compute the auto-correlation(acf) for the set. The acf can be seen in Figure 10.10.

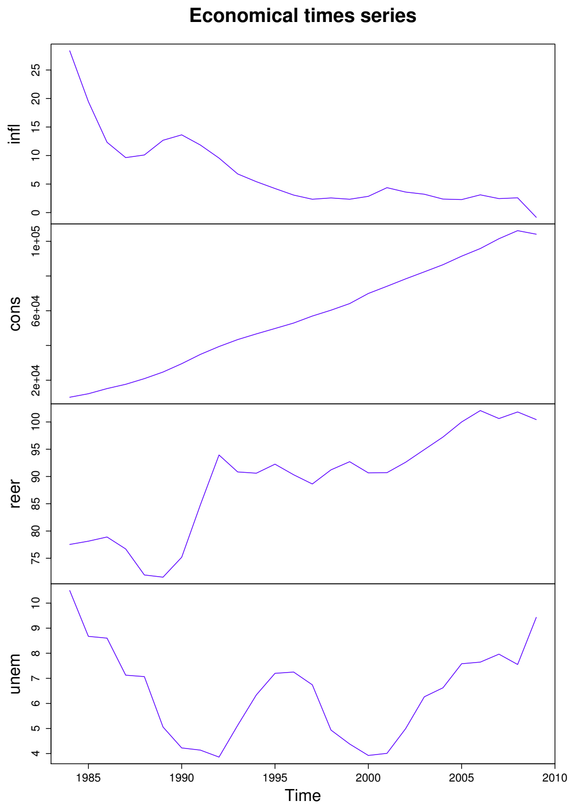

Figure 10.9: The time series from 1984 to 2009.

Looking at 10.10 we notice that exchange rate, consumption and inflation are highly correlated, clearly showing some pattern. Therefore we continue the modelling process.

Macroeconomic Model Analysis

The Var Model assumes stationarity, thus we test this assumption using the Dickey-Fuller Test on each series individually. We looked at the p-value and the compared the test-statistic with the critical values. We can observe the results in the tables below (Table 10.4 and Table 10.5).

	1pct	5pct	10pct
tau1	-2.66	-1.95	-1.6

Table 10.4: Critical values of the DF test statistics

The results show that, based on the p-value we cannot reject the null-hypothesis of non-stationarity only for the consumption , but by looking at

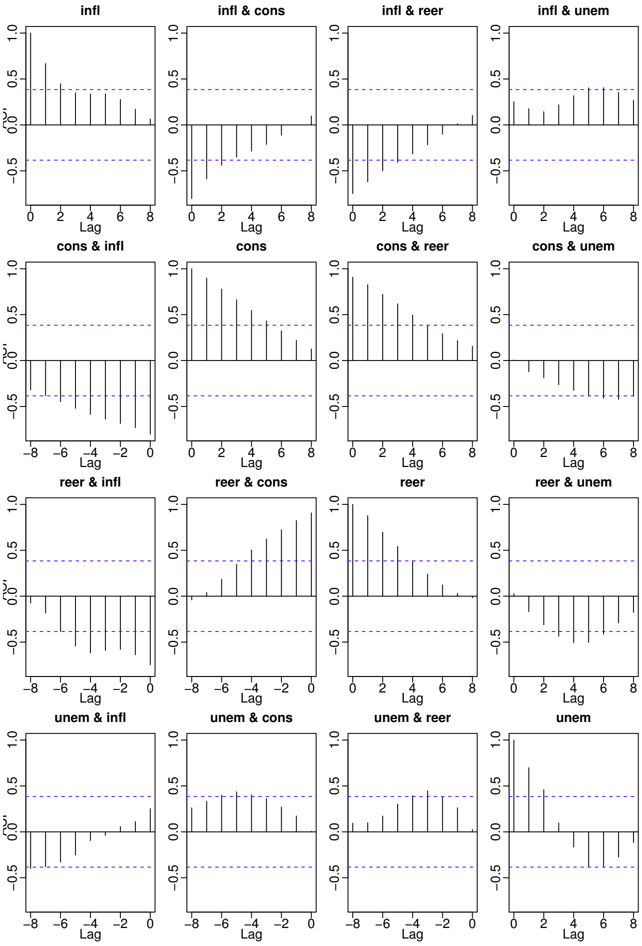

Figure 10.10: The acf for the four time series.

	P-value	Test statistics
Inflation rate	0.0003	-4.3533
Consumption	0.5439	-1.0759
REER	0.0068	-3.5671
Unemployment rate	0.0153	-1.7194

Table 10.5: Results from a Dickey fuller test on each times series

the test statistic we cannot reject the null hypothesis for all the variables. In this sense, we decided to remove the trend by differentiating the data. In Figure 10.11 we can see a plot of the differentiated time-series. We now see the series look more stationary.

After the differentiation we computed the acf function again and observed that the correlation pattern that was present before removing the

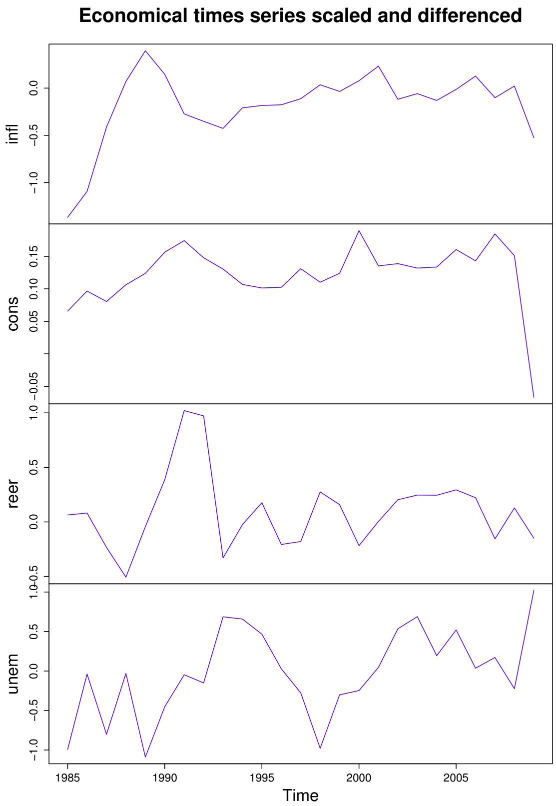

Figure 10.11: The time series from 1984 to 2009 differentiated.

trend is gone, but there were still some dependencies, specially between the exchange rate and the inflation rate and also between the inflation rate and the unemployment rate. Figure 10.12 shows the acf plot.

After checking the model assumptions we used the built-in function in R, `VARselect{vars}`, to find the optimal number of lags for the model. By looking at the AIC criteria, we decided to build a model with 4 lags.

We then used the VAR function, from the package "vars", in R, to compute the VAR model. Table 10.6 shows the results for the fitted model.

	Inflation		Consumption		REER		Unempl.		
	Adj. R^2	P-val	Adj. R^2	P-val	Adj. R^2	P-val	Adj. R^2	P-val	**AIC**
M1	0.5572	0.1864	0.1838	0.4458	0.2187	0.4221	0.2946	0.3695	-106.38

Table 10.6: Model 1: Differentiated once and lag 4, adjusted R squared values, p-values and the AIC for the model.

As the p-values of each variable all are very high, we conclude that none

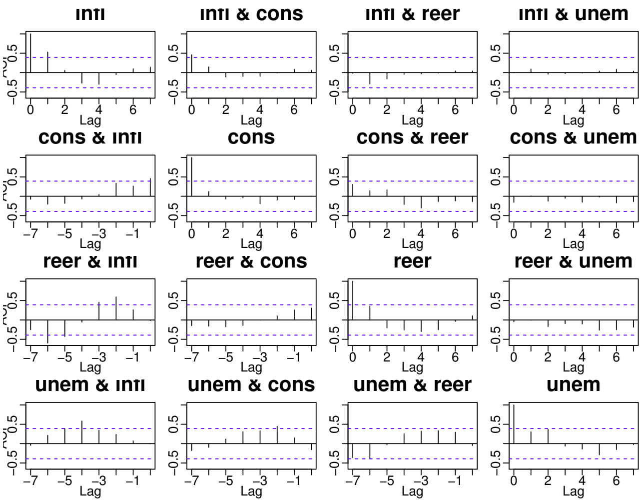

Figure 10.12: The acf for the four differeniated time series.

of them is significant, so we decided to try to differentiate only some of the variables instead of all of them. So, next we only differentiated the inflation rate, the effective exchange rate and the consumption and let the unemployment rate unchanged.

	Inflation		Consumption		REER		Unempl.		
	Adj. R^2	P-val	Adj. R^2	P-val	Adj. R^2	P-val	Adj. R^2	P-val	AIC
M2	0.5165	0.2139	0.1794	0.4488	0.3255	0.3477	0.7871	0.0535	-119.88

Table 10.7: Model 2: Inflation, consumption and unemployment are differentiated and lag 4.

With this new approach the model was still not significant we see in Table 10.7, we then tried to use 2 lags instead of 4, because with 2 lags, although the model in general had a higher AIC than the second model, individually, the sub models were significant, except for the consumption. The results are presented in Table 10.8. We choose the later model which

	Inflation		Consumption		REER		Unempl.		
	Adj. R^2	P-val	Adj. R^2	P-val	Adj. R^2	P-val	Adj. R^2	P-val	AIC
M3	0.6555	0.0016	0.2371	0.1491	0.4357	0.0301	0.82	2.265e-5	-67.869

Table 10.8: Model 3: Inflation, consumption and unemployment are differenced and lag 2.

consists of

$$\begin{pmatrix} infl_t \\ cons_t \\ reer_t \\ uenm_t \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{pmatrix} + \Pi_1 \begin{pmatrix} infl_{t-1} \\ cons_{t-1} \\ reer_{t-1} \\ unem_{t-1} \end{pmatrix} + \Pi_2 \begin{pmatrix} infl_{t-2} \\ cons_{t-2} \\ reer_{t-2} \\ unem_{t-2} \end{pmatrix} + \begin{pmatrix} \varepsilon_t^{(1)} \\ \varepsilon_t^{(2)} \\ \varepsilon_t^{(3)} \\ \varepsilon_t^{(4)} \end{pmatrix}, \quad (10.1)$$

where Π_1 and Π_2 are 4×4 matrices and $\varepsilon \overset{iid}{\sim} \mathcal{N}(0, 1)$. To check the model's validity, we look at the residuals from the model. The residuals are plotted with plus/minus 1 standard deviation in Figure 10.13. They look rather

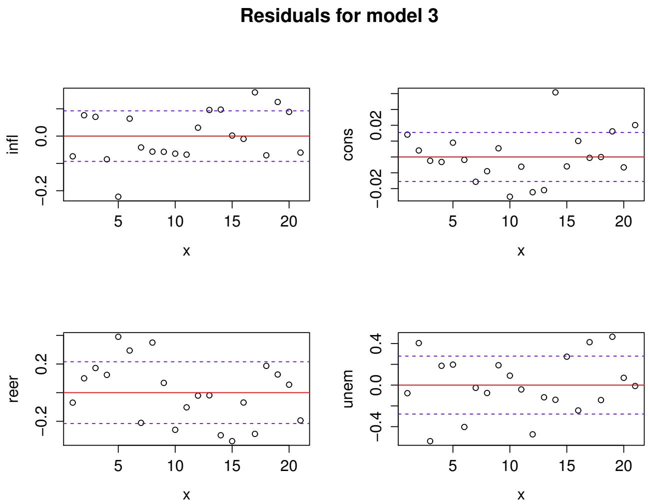

Figure 10.13: The residual plot of the chosen model.

random, but to be certain we can perform a test of randomness of sign changes for the residual. Ideally we want the probability of the residuals being positive or negative to be equal. We assume as written in equation (6.104, [4])

$$\text{Number of sign changes } \in \mathrm{B}\left(N-1, \frac{1}{2}\right),$$

where N is the number of residuals. For large values of N the binomial distribution can be approximated with a normal distribution such that we get

$$\text{Number of sign changes } \in \mathrm{B}\left(N-1, \frac{1}{2}\right) \simeq \mathcal{N}\left(\frac{N-1}{2}, \frac{N-1}{4}\right).$$

We can use the binom.test in R, where the H_0 test is that half of the trials are positive. The p-values, probability and confidence interval is seen in Table 10.9. For all submodels we get a p-value larger than 0.05 which is the critical number. Since it is larger than 0.05 we cannot reject the null hypothesis. From the output we also get the 95 % confidence interval, which is the 95 % CI for where the true probability for success lies. We get a probability of success on 0.5 and 0.6. This probability lies within the confidence interval and is therefore accepted. We notice it is close to the 0.5 that we assumed.

Parameter	Lower bound	Upper bound	Probability of success	P-value
Infl.	0.2719578	0.7280422	0.5	1
Cons.	0.3605426	0.8088099	0.6	0.5034
REER	0.3605426	0.8088099	0.6	0.5034
Unem	0.3605426	0.8088099	0.6	0.5034

Table 10.9: Result of test for sign change using N.

From the residual analysis we conclude that the residuals are random and the assumptions about this are kept.

The final model coefficients are

$$c = \begin{pmatrix} 0.18 \\ 0.15 \\ -0.59 \\ -0.83 \end{pmatrix},$$

$$\Pi_1 = \begin{pmatrix} 0.84 & 2.13 & -0.13 & 0.04 \\ 0.05 & 0.06 & 0.01 & -0.01 \\ -0.82 & 5.72 & 0.11 & -0.04 \\ -0.01 & -5.01 & -0.04 & 1.19 \end{pmatrix},$$

$$\Pi_2 = \begin{pmatrix} -0.35 & -3.93 & 0.11 & -0.06 \\ 0.04 & -0.12 & 0.03 & 0.02 \\ 0.76 & 0.51 & -0.63 & 0.05 \\ -0.04 & 11.08 & 0.20 & -0.35 \end{pmatrix},$$

where the numbers in red are significant at a 95%-level.

Predicting

When the model has been validated a prediction can be made. We have used the predict function in the "vars" package in R. We remove the last two observations from the data to use as a "test" set, leaving us with only 23 observations for predicting. The prediction and a 95% confidence interval can be seen in Figure 10.14.

We notice that the model predicts the inflation rate, the consumption and the unemployment rate are predicted quite well for the first period. The second prediction is only close for the inflation rate. We predict for the

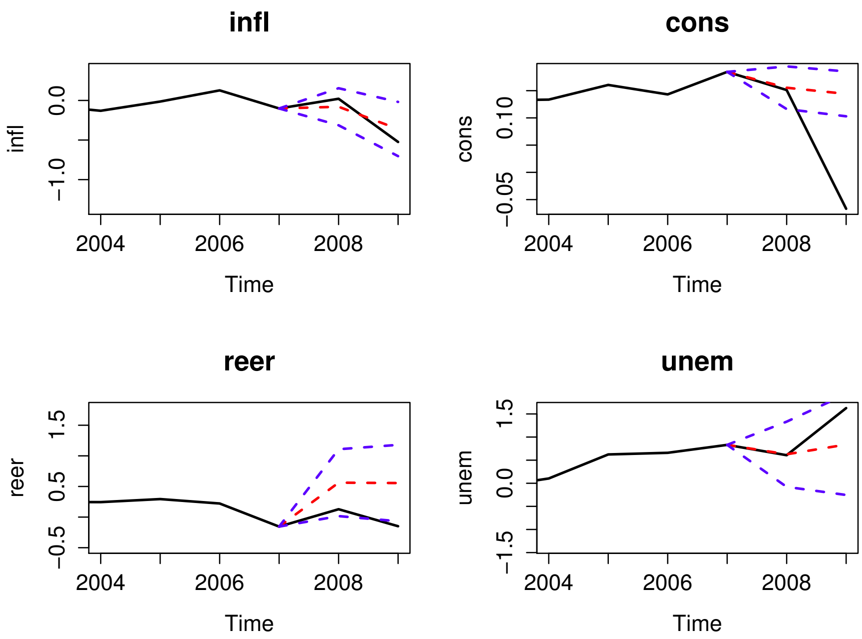

Figure 10.14: The predictions of the chosen model.

years 2008 and 2009, which is when the financial crisis of 2008 starts having and impact on the economies around the world. The model is not advanced enough to model a crisis, and unfortunately we did not have data for more recent year to model the recovery.

10.6 Further Work

Transforming the Data

Let $\{X_t\}_{t\in\mathbb{N}\cup\{0\}}$ be a time series taking values in R. In particular we will want to apply the subsequent construction to one of the financial time series appearing in the report. The basic idea behind the construction is to model the movement of the price regardless of the amount by which it changes (i.e. just whether it moves up or down). From the given series we can construct a new one taking values in $\{-1, 1\}$ as follows:

$$Y_t = \begin{cases} -1 & if\ X_{t-1} > X_t \\ 1 & if\ X_{t-1} \leq X_t \end{cases},$$

Additionally we can define $Y_0 = 1$, although it is nonessential. Later on we can define Y_0 as -1 or 1 with probability $\frac{1}{2}$. Our approach is to model

$\{Y_t\}_{t\in\mathbb{N}\cup\{0\}}$ instead of $\{X_t\}_{t\in\mathbb{N}\cup\{0\}}$.

The Poisson Process and the Telegraphic Process

Let N_t be a Poisson process with intensity λ. An appropriate model for Y_t is something in the lines of $(-1)^{N_t}$. Now the parameter λ is what remains to be estimated from the data. Having done that we can build a model of the stock price and make predictions employing a monte carlo simulation.
If we already have an estimate of the parameter λ we can consider as a very crude approximation to the stock price at time t given by the following process

$$Z_t = v \cdot \sum_{i=0}^{\lfloor t \rfloor} (-1)^{N_i},$$

where v is the amount by which the value of the time series changes (in our case $v = 1$). Further we can consider the stochastic process

$$A_t = A_0(\mu t + \sigma Z_t).$$

The parameters μ and σ can be estimated based on our knowledge of Z_t at certain points in time using OLS. This process can be used as a suitable model for the stock price.

Estimation of the Parameter λ

According to [2] such an estimate can be obtained explicitly.
Let $\{X_i\}_{i=0}^{n}$ be the n available observations (we shall assume observations equidistant in time). Let $m_2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - X_{i-1})$ be the quadratic variation of the series up to time n. Then the following estimator can be considered

$$\tilde{\lambda} = argmin_\lambda \left\{ m_2 - \frac{v^2}{2}\left(\Delta - \frac{1 - e^{-2\lambda\Delta}}{2\lambda}\right)\right\},$$

where v is the constant amount with which the price moves and Δ is the step size (in our case 1 day). Using a numerical procedure we can get an approximation to the real value of $\tilde{\lambda}$ which can subsequently be used in the models discussed.

10.7 Appendix

Packages and functions used in R

Package	**Function**
MTS	ts
	diff
	act
fUnitRoots	urdfTest
vars	VARselect
	VAR
	predict

Table 10: Packages and functions used.

Bibliography

[1] BROCKWELL, PJ AND DAVIS, RA. *Introduction to Time Series and Forecasting*. Springer Texts in Statistics. Springer, 2002.

[2] DE GREGORIO, A AND IACUS, S. *Parametric Estimation for the Standard and Geometric Telegraph Process Observed at Discrete Times*. 2013.

[3] LTKEPO, H. *New Introduction to Multiple Time Series Analysis*. Springer, 2005.

[4] MADSEN, H. *Time Series Analysis*. Texts in Statistics Science. Chapman and Hall, 2008.

[5] NAVELLIER, L. Stocks to Buy and Sell: Ratings for the Top 50 Tech Stocks kernel description, 2010.
URL `http://www.nasdaq.com/article/stocks-to-buy-and-sell-ratings-for-the-top-50-tech-stocks-cm18960`

[6] QUANDL.
URL `https://www.quandl.com/collections/markets`