UNIVERSITY OF LJUBLJANA
FACULTY OF MATHEMATICS AND PHYSICS

Financial mathematics – 2. cycle

Živa Mitar

Modelling the term premium: Cochrane-Piazzesi measure

Master thesis

Advisor: prof. dr. Bojan Basrak

Ljubljana, 2015

UNIVERZA V LJUBLJANI
FAKULTETA ZA MATEMATIKO IN FIZIKO

Finančna matematika – 2. stopnja

Živa Mitar

Modeliranje terminske premije: Cochrane-Piazzesiina mera

Magistrsko delo

Mentor: prof. dr. Bojan Basrak

Ljubljana, 2015

Podpisana Živa Mitar izjavljam:

– da sem magistrsko delo z naslovom *Modeliranje terminske premije: Cochrane-Piazzesiina mera* izdelala samostojno pod mentorstvom prof. dr. Bojana Basraka in

– da Fakulteti za matematiko in fiziko Univerze v Ljubljani dovoljujem objavo elektronske oblike svojega dela na spletnih straneh.

Ljubljana, 19. 03. 2015 Podpis: ....

Contents

Work programme

The master thesis should present the concept of term premium and its modelling, with emphasis on the approach suggested in Cochrane and Piazzesi's article Bond Risk Premia from 2005 [8]. The thesis should include the implementation of the model and the presentation of results.

prof. dr. Bojan Basrak

Modelling the term premium: Cochrane-Piazzesi measure

Abstract

This thesis presents the definition and economic interpretation of the bond risk premium, also known as term premium. Its common interpretation is, that it presents the excess return that investors require for holding long-term bonds, instead of rolling over bonds with shorter maturities. However, this is not always the case. In the first half of the previous decade the US term premium was negative and that caused series of attempts to properly measure it. Cochrane and Piazzesi modelled it using regressions on forward rates and yields and discovered a single factor - a linear combination of forward rates, that predicts it with high coefficient of determination. In the thesis we first describe the tools behind the Cochrane and Piazzesi's model (with main focus placed on linear regression and its assumptions), followed by its presentation and implementation on German yield curve data. This extends the work of common studies, as they usually focus on US yield curve. At the same time this allows us to test the robustness of the measure. The end of the thesis includes results about the predictive power of the model in comparison to standard ARMA models.

Modeliranje terminske premije: Cochrane-Piazzesiina mera

Povzetek

V dani magistrski nalogi so predstavljene osnovne definicije in ekonomske interpretacije terminske premije oz. premije za tveganje pri nakupu obveznic. Običajno dojemamo terminsko premijo kot dodaten donos vlagatelja, kateri namesto večkratnega zaporednega nakupa obveznic krajših ročnosti, sredstva investira v obveznice z daljšim dospetjem. Terminska premija pa ni nujno pozitivna – v ZDA je bila v obdobju pred krizo negativna in s tem je pritegnila pozornost strokovne javnosti. Cochrane in Piazzesi sta jo modelirala z uporabo linearne regresije po metodi najmanjših kvadratov na terminskih obrestih merah. Odkrila sta obstoj linearne kombinacije le-teh, ki napoveduje vrednosti terminske premije različnih ročnosti z visokim koeficientom determinacije. V začetku naloge so opisana orodja uporabljena v njunem modelu, sledi implementacija modela na trgu nemških državnih obveznic. Le-to predstavlja odklon od standardne literature, kjer je najpogosteje predmet opazovanj terminska premija na ameriškem trgu, hkrati pa nam omogoča preverjanje robustnosti modela. Konec magistrskega dela vključuje tudi uspešnost modela pri napovedi vrednosti terminske premije v primerjavi s standardnim ARMA modelom.

Math. Subj. Class. (2010): 62J05, 62H12, 62H15, 91B84, 91G30, 91G70

Ključne besede: Terminska premija, Časovna struktura obrestnih mer, Večkratna (multipla) linearna regresija, Metoda najmanjših kvadratov, Cochrane-Piazzesiina mera, Avtokorelacija časovnih vrst

Keywords: Term premium, Yield structure, Multiple linear regression, Ordinary least squares, Cochrane-Piazzesi measure, Autocorrelation

1 Introduction to bond markets and term premium

1.1 About bonds

A bond is a debt security issued by a borrower, who is then required to repay to the lender (buyer of the bond) the amount borrowed plus interest, over a specified period of time, called maturity. Bonds are known as fixed income instruments as in past they were only paying fixed coupons once each year. Today there are many different types of bonds, the most common one is still the conventional (or plain vanilla or bullet) bond.

Key features determining the bond are issuer and its ability to repay the loan, principal and coupon rate and term to maturity (also known as maturity or term).

Bonds are most commonly issued by sovereign governments and their agencies, local government authorities, supranational agencies and larger corporations. The largest bond markets are those of sovereign borrowers: in Germany government bonds are known as bunds or Schatze, in UK they are known as gilts and in USA as Treasuries.

Term to maturity refers to a time period after which the bond principal will be repaid in full and over which the bondholder can expect to receive possible coupon payments. The exact date, presenting the end of such period, is known as maturity date. Another important dates are trade date and settlement date. First one is the date when the bond is traded and the later one is the date, when the bond is delivered to the buyer and the payment is received by bond issuer. Usually this occurs one or two days after the trade date.

The principal of a bond is the amount that the issuer agrees to repay the bondholder on the maturity date. This amount is also referred to as the redemption value, par value, nominal value or face amount. The coupon rate (or nominal rate) is the interest rate that the issuer agrees to pay each year. The annual amount of the interest payment made is called the coupon. Coupons are usually paid semi-annually or annually, occasionally also on quarterly or monthly basis. Depending on coupon payments we differentiate between zero-coupon bonds, fixed rate bonds, floating rate notes, index-linked bonds, etc.

On international markets the currency in which the bonds are issued play important role as the overall return of the bond may not be only determined by interest rate but also by currency exchange rates. Currency of bond denomination may also affect the liquidity of the bond; as bond ownership is transferable the ease of selling and buying the bond is defined as bond liquidity.

Next subsections will introduce some basic terminology related to bond pricing and yields, readers interested in this topic are further invited to have a look in Choudry [6] and Fabozzi [12].

1.1.1 Bond pricing

The fair price P of any bond is the present value of all its cash flows.

Let us first consider a bond with fixed coupon rate $c \in (0,1)$ and principal M. Coupons $C = cM$ are paid out at times $t_n = n\Delta t$, where $n = 1, \dots, N$ and

$t_n - t_{n-1} = \Delta t$ for each n. Amount $\tau = t_N$ represents the (residual) maturity (measured in interest periods) and $r_t(t_n)$ represents the discount rate at time t for period t_n. In case of discrete compounding the fair price P of such bond at time t is calculated as (Fabozzi [12]):

$$P_t(\tau) = \sum_{n=1}^{N} \frac{C}{(1 + r_t(t_n))^{t_n}} + \frac{M}{(1 + r_t(\tau))^{\tau}}. \quad (1)$$

Under assumptions that coupons C are paid out yearly and the yearly rate of return r is constant, the above formula is simplified (note that in such case $\Delta t = 1, r_t(t_n) = r$ and $\tau = N$):

$$P_t(N) = \sum_{n=1}^{N} \frac{C}{(1 + r)^n} + \frac{M}{(1 + r)^N}. \quad (2)$$

In case of continuous compounding (1) can be rewritten as:

$$P_t(\tau) = \sum_{n=1}^{N} Ce^{-r_t(t_n)t_n} + Me^{-r_t(\tau)\tau}. \quad (3)$$

The discount interest rate r is also known as the bond's yield. It is the return demanded by the investor (for buying the bond), therefore it is sometimes called the bond's return. As we can see from above formulas the yield and price are inversely related. Higher yields mean lower price - the investor wants more return on her investment and therefore believes that bond is not worth as much as before.

1.1.2 Bond yields and yield curves

1.1.2.1 Yield to maturity and yield to maturity curve

The internal rate of return on any investment is the interest rate that will make the present value of the cash flows from the investment equal to the initial cost (price) of the investment [6]. For bonds internal rate of return is equivalent to yield to maturity (YTM).

The formula for calculating YTM is essentially the same as the formula for calculating price. Under assumption that the coupon will be delivered exactly one interest period from now, YTM is calculated by solving (4) or (5) for y.

$$P_t(\tau) = \sum_{n=1}^{N} \frac{C}{(1 + y)^{t_n}} + \frac{M}{(1 + y)^{\tau}}, \quad (4)$$

$$P_t(\tau) = \sum_{n=1}^{N} Ce^{-r_t(t_n)t_n} + Me^{-r_t(\tau)\tau} = \sum_{n=1}^{N} Ce^{-yt_n} + Me^{-y\tau}. \quad (5)$$

YTMs calculated across more maturities construct so called YTM yield curve. The typical shapes of an YTM yield curves are normal (slightly positive), negative, flat or humped (Fig. 1).

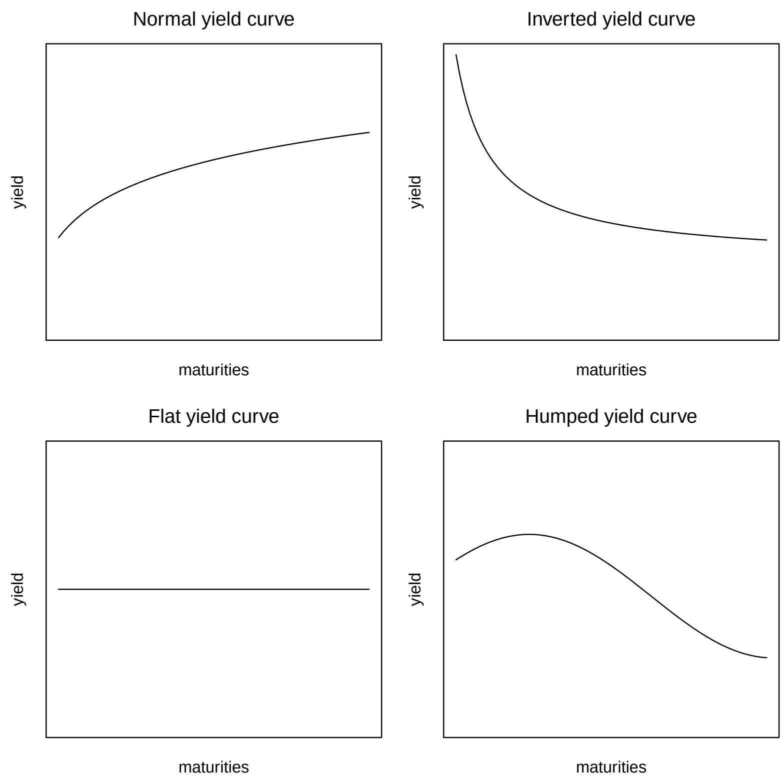


Figure 1: Different types of yield curve.

1.1.2.2 Spot and forward rates

For the zero-coupon bond ($C = 0$) with maturity N the expression to calculate bond price simplifies significantly:

$$P_t(N) \quad = \quad \frac{M}{(1 + r_t(N))^N} \approx e^{-N r_t(N)}. \tag{6}$$

Consequently one can easily calculate the yield to maturity for all zero-coupon bonds, which is also known as **spot rate**:

$$rs_{t,n} = rs_t(n) \quad = \quad \sqrt[n]{\frac{M}{P_t(n)}} - 1 \approx -\frac{\log \frac{P_t(n)}{M}}{n}. \tag{7}$$

By compounding multiple spot rates with different maturities we can construct spot rate curve, also called zero-coupon yield curve or term structure of interest rates.

In case spot rate is calculated at a forward time using data from current yield curve, this is a **forward rate**.

The calculation of forward rates is important for forward contracts. A buyer of forward contract agree at time t to purchase a bond with price $P(n_1)$ at time n_1 and get principal M in return at n_2.

If rf_{t,n_1,n_2} denotes a forward spot rate at time t between times n_1 and n_2, then the current price of a bond with yearly coupon C and maturity N is equal to:

$$\begin{aligned} P_t(N) &= \frac{C}{(1+rs_{t,1})} + \frac{C}{(1+rs_{t,1})(1+rf_{t,1,2})} + \dots \\ &+ \frac{C+M}{(1+rs_{t,1})(1+rf_{t,N-2,N-1})(1+rf_{t,N-1,N})}. \end{aligned} \tag{8}$$

Comparison of (2) with (8) also reveals the relationship between spot and forward rates:

$$(1+rs_{t,n})^n = (1+rs_{t,1})(1+rf_{t,1,2})\dots(1+rf_{t,n-1,n}) \tag{9}$$

and therefore

$$rs_{t,n} = \sqrt[n]{(1+rs_{t,1})(1+rf_{t,1,2})\dots(1+rf_{t,n-1,n})} - 1. \tag{10}$$

However, these equalities are valid only, when we believe that forward rates are in fact equal to future spot rates. A base for this assumption is found in **expectation hypothesis theory**, which will be further explained in the following section.

1.2 Expectation hypothesis and term premium

The theory used to derive the forward yield curve and under which the (9) and (10) are valid, is called **(pure) expectations hypothesis**. It is the basic theory of the term structure of interest rates and it states that current implied forward rates are unbiased estimators of future spot rates. However the observations of the market yields may not always seem to follow this hypothesis.

As it was written by Kim and Orphanides [20]:

> According to this (expectation) hypothesis, the expected return from holding a long bond until maturity is the same as the expected return from rolling over a series of short bonds with a total maturity equal to that of the long bond. That is, the long bond yield is the average of the expected short term rates. Equivalently, the forward rate is the expected future short-term rate.[1]
>
> Though the expectation hypothesis provides a simple and intuitively appealing interpretation of the yield curve, it ignores interest rates risk.[2] Except if calculated until maturity, the nominal return on a long bond is uncertain, and investors may require compensation for this risk. The 'term premium' refers to such compensation and any other sources of deviation from expectation hypothesis.

Note that risk compensation and therefore the term premium is not necessarily positive, although this is the direction we usually would think of.

[1]Comment ZM: In paragraph long bond should be read as bond with longer maturity and short bond as a bond with shorter maturity.

[2]Interest rate risk is the risk that arises for bond owners from fluctuating interest rates. How much interest rate risk a bond has depends on how sensitive its price is to interest rate changes in the market. The sensitivity depends on two things, the bond's time to maturity, and the coupon rate of the bond (Ross, Westerfield, Jordan [26]).

1.2.1 Term premium definitions

Kim and Orphanides [20] define three types of term premium. All definitions are associated with expectations of one-period rate and holding period returns.

Term premium - return premium $\phi_t(N)$ is equal to expected return on holding a zero-coupon bond with maturity N for one period minus the spot rate:

$$\phi_t(N) = E_t(R_{N,t+1}) - rs_{t,1}, \tag{11}$$

where $R_{n,t+1} = \log \frac{P_{t+1}(n-1)}{P_t(n)}$.

Term premium - forward premium is equal to forward rate minus the expected future spot rate:

$$\phi_t(N) = rf_{t,N-1,N} - E_t(rs_{t,N-1}). \tag{12}$$

Term premium - yield premium is equal to yield on zero-coupon bond with maturity N>1 minus the average of expected short rates from the present to the maturity of the bond:

$$\phi_t(N) = rs_{t,n} - \frac{1}{N}\sum_{n=1}^{N} E_t(rs_{t,n}). \tag{13}$$

In all of the above definitions $E_t(X)$ refers to the expectation of variable X at time t.

1.3 Measuring the term premium

As we see from above definitions, the measurement of term premium is fairly easy, provided that we know the markets' expectations about future interest rates over long horizons. These are however difficult to measure and the uncertainty of such measurements is also reflected in our term premium estimation.

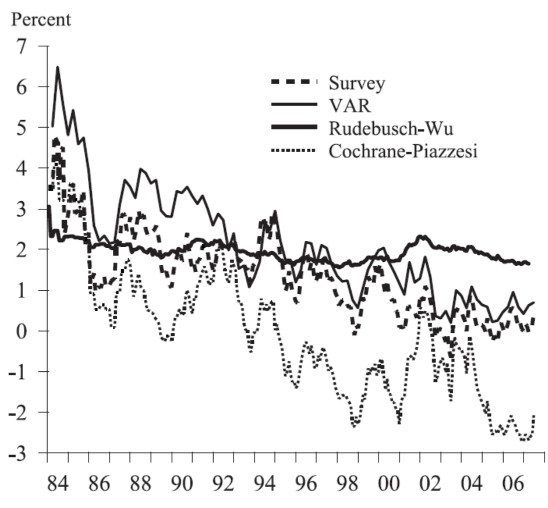


Figure 2: Different estimates of 10-year term premia (adapted from Swanson [32]).

This uncertainty is seen on Figure 2, where four different ways of estimating term premium are depicted. In the figure label *Survey* refers to survey based measure. In this case predictions about future rates are done based on market participants' answers regarding their expectations for future short-term interest rates. Label VAR is referring to VAR based measure, where VAR stands for Vector Autoregressive model. It is based on the VAR projection of the short rate on macroeconomic variables - unemployment rate, inflation, consumer price and additionally 3-month bill rate. The *Rudebusch-Wu measure* is in a way similar to VAR model, as it also uses macroeconomic variables, but the model used to forecast interest rates is New Keynesian macroeconomic model (Rudebusch, Sack and Swanson [27]).

The last mentioned measure is *Cochrane-Piazzesi measure*: a regression based model, which proposes to forecast the bond excess returns by initial forward rates. In such manner calculated expected one year excess return can be iterated in order to get the expected excess return for each of the wanted maturities. This measure is further presented and commented on in third section.

2 Linear regression

A classical statistical problem is to try to determine the relationship between a **dependent or explained variable** (also **regressand**) and one or more **independent or explanatory variables** (also **regressors**). If we believe the relationship between variables to be linear, we use linear regression model.

In cases where the behaviour of dependent variable is predicted by one independent variable, we speak of univariate linear regression. When there are two or more regressors, then we speak of multiple (univariate) linear regression.

The generic form of multiple linear regression model is

$$y = f(x_1, x_2, \ldots, x_K) + \epsilon = \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_K x_K + \varepsilon.$$

In this setting y is explained variable and $x_1, x_2, \ldots, x_K$ are explanatory ones and ε is the **disturbance** or the **error term**.

Suppose we have a sample with N observations of our dependant random variable Y and other independent variables $(X_1, X_2, X_3, \ldots, X_K)$, which can either be stochastic or non-stochastic.

As it is often the case, we will from now on assume one of the independent variables to be equal to a constant; without loss of generality we can proclaim X_1 to be such. The observations of Y and X's will be written as:

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix} \qquad \mathbf{X} = \begin{pmatrix} \mathbf{x}_1' \\ \mathbf{x}_2' \\ \vdots \\ \mathbf{x}_N' \end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} & \ldots & x_{1K} \\ x_{21} & x_{22} & \ldots & x_{2K} \\ \vdots & \vdots & \ldots & \vdots \\ x_{N1} & x_{N2} & \ldots & x_{NK} \end{pmatrix}. \tag{14}$$

In the $N \times K$ matrix $\mathbf{X}$, the n-th row refers to observation n and the k-th column refers to the k-th regressor. Element x_{nk} tells us the n-th observation of variable X_k. Analogically y_n refers to the n-th observation of regressand. This kind of model can also be referred to as the **population regression model.**

With this representations in mind, we can write linear model as:

$$y_n = \beta_1 + \beta_2 x_{n2} + \ldots + \beta_K x_{nK} + \varepsilon_n, \quad n = 1, 2, \ldots, N \tag{15}$$

or

$$y_n = \mathbf{x}_n' \boldsymbol{\beta} + \varepsilon_n, \quad n = 1, 2, \ldots, N \tag{16}$$

or in a matrix notation as:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}. \tag{17}$$

As already mentioned, here $\boldsymbol{\varepsilon}$ presents unobserved standard variable and is also referred to as the error vector. The term ε_n is it´s n-th component.

The coefficients β_k are unknown fixed population parameters, which are also called **regression coefficients**. They measure the expected change in dependent variable as consequence of the unit increase of a value x_{nk}, while other elements in $\mathbf{x}_n$ are held constant. Our goal is to find a good estimate for their values:

$$\begin{aligned} E(Y|X_1 &= x_1, X_2 = x_2 + 1, \ldots, X_K = x_K) - E(Y|X_1 = x_1, X_2 = x_2, \ldots, X_K = x_K) \\ &= \beta_1 + \beta_2(x_2 + 1) + \ldots + \beta_K x_K - \beta_1 - \beta_2 x_2 - \ldots - \beta_K x_K = \beta_2. \end{aligned} \tag{18}$$

The equality in (18) is supposed to hold for any possible observation, while we only observe a sample of N observations. The term $\mathbf{x}_n'\boldsymbol{\beta}$ is also known as **regression line**.

2.1 Classical linear regression model and its assumptions

Denote the population multiple regression by $y_n = \beta_1 + \beta_2 x_{n2} + \ldots + \beta_K x_{nK} + \varepsilon_n$ and assume that N sets of observations are available. Based on assumptions we make, we differentiate among various linear regressions models. In **classical linear regression (CLR) model** the standard assumptions, as nicely summoned in Kennedy [19, Ch. 3], are:

[**A1**] It is possible to calculate the dependent variable as a linear function of specific set of independent variables, plus an error term:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}.$$

Unknown coefficients $\boldsymbol{\beta}$ of mentioned linear function are assumed to be constants.
Violations: Wrong regressors, non-linearity and changing parameters $\boldsymbol{\beta}$.
Comment: Because the unknown coefficients $\boldsymbol{\beta}$ are assumed to be constants, classical linear regression model is fixed effects regression model. When one or more of the unknown coefficients arise from random causes, models are called random or mixed effects models.

[**A2**] **Explanatory variables are exogenous.** Observations on the independent variable can be considered fixed in repeated samples. Sometimes this is also stated as $\mathbf{X}$ is a deterministic non-stochastic matrix.
Violations: Measurement errors in independent variables, autoregression, simultaneous equation estimation (situations in which the dependant variable is determined by the simultaneous interaction of several relationships).
Comment: In other linear regression model we may also assume that $\mathbf{X}$ is a stochastic matrix. In that case we additionally assume that $\mathbf{X}$ and $\boldsymbol{\varepsilon}$ are independent.

[**A3**] The expected value of the disturbance term is zero. Under [**A2**] this means, that $E(\boldsymbol{\varepsilon}) = 0$ and consequently $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$.
Violations: Biased intercept problem.
Comment: Without any assumptions about $\mathbf{X}$, this condition should be equal to $E(\boldsymbol{\varepsilon}|\mathbf{X}) = 0$ and $E(\mathbf{y}|\mathbf{X}) = \mathbf{X}\boldsymbol{\beta}$.

[**A4**] Disturbances have uniform variance and are uncorrelated. Given the assumptions [**A2**] and [**A3**], this means following:

$$E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \Sigma = \sigma^2 I,$$

where I is identity matrix of the dimension $N \times N$. This assumption is also known as **homoskedasticity**.

In the cases where $\mathbf{X}$ is stochastic, we have: $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'|\mathbf{X}) = \Sigma = \sigma^2 I$.
Violations: Heteroskedasticity and autocorrelated errors. In first case Σ is diagonal, but the elements vary throughout the whole diagonal, whereas in the second case, Σ is not diagonal.
Comment: The distribution of $\boldsymbol{\varepsilon}$ is often assumed to be normal (especially by doing tests and calculating confidence intervals); the parameter σ is either known or unknown.

[A5] $\mathbf{X}\,(N \times K, N >> K)$ **has a full rank.** Typically there are more observations than there are independent variables. Besides that, there is no exact linear relationship between explanatory variables.
Violations: Multicollinearity or near-linear dependence (some regressors are nearly linear combinations of others). In some cases our variables may indeed be linearly independent, but the near-linear dependence is then introduced when sampling and rounding the data (approximate linear dependence).

2.2 Least Squares Estimator

The goal of regression is to find a good approximate value for $\boldsymbol{\beta}$. The rule which tells us how to translate a given sample in approximate value is called **estimator** and the result for the given sample is called **estimate**. First one is the vector of random variables, whilst the second one is just the vector of numbers, depending on sample. The most widely used estimator in econometrics is the **ordinary least squares (OLS)** estimator.

Let us again have a sample of N observations od dependent variable Y and independent variables $X_1, \ldots, X_K$. Because we want to find a best approximation for $\mathbf{y}$, we will be interested in the differences between estimated $\tilde{\mathbf{y}}$ and the given values of $\mathbf{y}$. For n-th observation the difference between y_n and an arbitrary linear combination of explanatory variables would look like:

$$y_n - \tilde{\beta}_1 x_{i1} + \tilde{\beta}_2 x_{i2} + \ldots + \tilde{\beta}_K x_{iK} = y_n - \mathbf{x}_n' \tilde{\boldsymbol{\beta}}. \tag{19}$$

Our task is to estimate the values of $\tilde{\beta}_1, \ldots, \tilde{\beta}_K$ in a way, that the sum of square differences between estimated $\tilde{\mathbf{y}}$ and true $\mathbf{y}$ is the smallest possible. We minimize the expression:

$$S(\tilde{\boldsymbol{\beta}}) := \sum_{i=1}^{N} (y_i - \mathbf{x}_i' \tilde{\boldsymbol{\beta}})^2. \tag{20}$$

For ease of use we will from now on in this section use the matrix notation. On some parts also the expended version will be given.

In matrix notation $S(\tilde{\boldsymbol{\beta}})$ is written as:

$$S(\tilde{\boldsymbol{\beta}}) = (\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}) = \mathbf{y}'\mathbf{y} - 2\mathbf{y}'\mathbf{X}\tilde{\boldsymbol{\beta}} + \tilde{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\tilde{\boldsymbol{\beta}}. \tag{21}$$

Differentiation with respect to $\tilde{\boldsymbol{\beta}}$ and then applying first order conditions to (21), gets us to so called normal equations:

$$\frac{\partial S(\tilde{\boldsymbol{\beta}})}{\partial \tilde{\boldsymbol{\beta}}} = -2(\mathbf{X}'\mathbf{y} - \mathbf{X}'\mathbf{X}\tilde{\boldsymbol{\beta}}) = 0. \tag{22}$$

If the matrix $\mathbf{X}'\mathbf{X}$ is invertible (there is no perfect multicollinearity), then the solution of equation (22) exists and will be denoted with $\mathbf{b}$:

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}.$$

By checking second order conditions we can verify that $\mathbf{b}$ indeed corresponds to a minimum:

$$\frac{\partial^2 S(\tilde{\boldsymbol{\beta}})}{\partial^2 \tilde{\boldsymbol{\beta}}} = 2\mathbf{X}'\mathbf{X} \geq 0.$$

Using these results, the approximation $\tilde{y}_n$ of y_n is equal to $\tilde{y}_n = \mathbf{x}_n' b$. The difference between the measured and approximated value is defined as **residual**:

$$e_n = y_n - \tilde{y}_n = y_n - \mathbf{x}_n'\mathbf{b}.$$

All residuals together build the N-dimensional vector of residuals $\mathbf{e} = [e_1, e_2, \dots, e_N]'$ and with help of this vector, we can decompose $\mathbf{y}$ as:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}.$$

In the matrix notation $\tilde{\mathbf{y}}$ is expressed as:

$$\tilde{\mathbf{y}} = \mathbf{X}\mathbf{b} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{P}_\mathbf{X}\mathbf{y}.$$

Matrix $\mathbf{P}_\mathbf{X} \equiv \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ is in linear algebra known as a projection matrix, which means nothing else but that the result we get is an orthogonal projection of the dependent variable on the independent ones.

Further on we are interested on relationship between the vector of residuals and independent variables. First order conditions (22) imply that each column of matrix $\mathbf{X}$ (regressors) is orthogonal to vector of residuals:

$$\begin{aligned} \mathbf{X}'(\mathbf{y} - \mathbf{X}\mathbf{b}) &= 0 \\ \mathbf{X}'\mathbf{e} &= 0. \end{aligned} \tag{23}$$

We can also express $\mathbf{e}$ as $\mathbf{e} = \mathbf{y} - \mathbf{X}\mathbf{b} = \mathbf{y} - \mathbf{P}_\mathbf{X}\mathbf{y} = (\mathbf{I} - \mathbf{P}_\mathbf{X})\mathbf{y} = \mathbf{M}_\mathbf{X}\mathbf{y}$, i.e. the residual vector is a projection of $\mathbf{y}$ upon the space orthogonal to the one, spanned by regressors. Both matrices, $\mathbf{P}_\mathbf{X}$ and $\mathbf{M}_\mathbf{X}$, are symmetric an idempotent (${\mathbf{P}_\mathbf{X}}^2 = \mathbf{P}_\mathbf{X}$, ${\mathbf{M}_\mathbf{X}}^2 = \mathbf{M}_\mathbf{X}$).

Additional terms which can be mentioned at this point are also **Residual Sum Of Squares (RSS)**, **Explained Sum Of Squares (ESS)** and **Total Sum Of Squares (TSS)**.

Residual sum of squares measures how good can the differences in values of $\mathbf{y}$ be explained with regressors $\mathbf{X}$:

$$\text{RSS} = Q_e = \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} = (\mathbf{y} - \tilde{\mathbf{y}})'(\mathbf{y} - \tilde{\mathbf{y}}).$$

Weighted RSS is also used to estimate the variance σ^2, when this one is not given in advance:

$$\tilde{\sigma}^2 = \frac{(\mathbf{y} - \tilde{\mathbf{y}})'(\mathbf{y} - \tilde{\mathbf{y}})}{N-K} = \frac{Q_e}{N-K}.$$

Explained sum of squares measures how well a model represents the underlying data, i.e. how much variance is explained through regression:

$$\text{ESS} = (\tilde{\mathbf{y}} - \overline{\mathbf{y}})'(\tilde{\mathbf{y}} - \overline{\mathbf{y}}).$$

Total sum of squares describes the variation in the values of the observed variable itself:

$$\text{TSS} = (\mathbf{y} - \overline{\mathbf{y}})'(\mathbf{y} - \overline{\mathbf{y}}).$$

Total variance can be decomposed on variance, that is explained through regression and variance of the disturbance:

$$\begin{aligned}
(\mathbf{y} - \overline{\mathbf{y}})'(\mathbf{y} - \overline{\mathbf{y}}) &= (\tilde{\mathbf{y}} + \boldsymbol{\varepsilon} - \overline{\mathbf{y}})'(\tilde{\mathbf{y}} + \boldsymbol{\varepsilon} - \overline{\mathbf{y}}) \\
&= (\tilde{\mathbf{y}} - \overline{\mathbf{y}})'(\tilde{\mathbf{y}} - \overline{\mathbf{y}}) + 2\boldsymbol{\varepsilon}'(\tilde{\mathbf{y}} - \overline{\mathbf{y}}) + \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} \\
&= (\tilde{\mathbf{y}} - \overline{\mathbf{y}})'(\tilde{\mathbf{y}} - \overline{\mathbf{y}}) + 2\boldsymbol{\varepsilon}'(\mathbf{X}\mathbf{b} - \overline{\mathbf{y}}) + \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} \\
&= (\tilde{\mathbf{y}} - \overline{\mathbf{y}})'(\tilde{\mathbf{y}} - \overline{\mathbf{y}}) + 2\boldsymbol{\varepsilon}'(\mathbf{X}\mathbf{b} - \overline{\mathbf{y}}) + \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} \\
&= (\tilde{\mathbf{y}} - \overline{\mathbf{y}})'(\tilde{\mathbf{y}} - \overline{\mathbf{y}}) + 2\boldsymbol{\varepsilon}'\mathbf{X}\mathbf{b} - 2\boldsymbol{\varepsilon}'\overline{\mathbf{y}} + \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} \\
&= (\tilde{\mathbf{y}} - \overline{\mathbf{y}})'(\tilde{\mathbf{y}} - \overline{\mathbf{y}}) + \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}. \qquad (24)
\end{aligned}$$

Expression $2\boldsymbol{\varepsilon}'\mathbf{X}\mathbf{b}$ is zero due to first order conditions of optimization process (23). Expression $-2\boldsymbol{\varepsilon}'\overline{\mathbf{y}}$ is equal to zero because the mean of error term is zero (the necessary condition for this statement to be true is that our regression model includes an intercept).

2.2.1 Properties of the OLS estimator

Estimators are compared on the basis of their attributes. Attributes that can be compared regardless of the sample size are called **finite sample properties** and the ones that are seen or important only, when we have large 'unlimited' samples, are called **asymptotic properties.** First we will give some definitions and then we will describe the properties of the OLS estimator.

2.2.1.1 Finite sample properties of the OLS

Definition 1. An estimator $\tilde{\theta} = g(\mathbf{x}_1, \ldots, \mathbf{x}_N) = g(\mathbf{X})$ of parameter θ is *unbiased*, if the mean of its sampling distribution is θ.

Unbiasedness means that, if the samples are drawn independently and repeatedly and $\tilde{\theta}$ is calculated every time, than we expect the average value of these estimates to be very close to θ.

Although unbiasedness is a desirable property, it is usually not enough. We often wish that our estimator possess yet another property - efficiency.

Definition 2. An unbiased estimator $\tilde{\theta}_{\mathbf{1}}$ is more *efficient* than another unbiased estimator $\tilde{\theta}_{\mathbf{2}}$ if the sampling variance of $\tilde{\theta}_{\mathbf{1}}$ is less than that of $\tilde{\theta}_{\mathbf{2}}$. That is,

$$\text{var}(\tilde{\theta}_{\mathbf{1}}) < \text{var}(\tilde{\theta}_{\mathbf{2}}).$$

We can sum the properties of OLS estimator in Gauss-Markov Theorem.

Theorem 3. *(Gauss - Markov Theorem)*
Consider a classical linear model with standard assumptions:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta} \textit{ and } \operatorname{cov}(\mathbf{y}) = \sigma^2\mathbf{I}.$$

Then an OLS estimator $\mathbf{b}$ *for* $\boldsymbol{\beta}$ *satisfies the following:*

a) It is unbiased, $E(\mathbf{b}) = \boldsymbol{\beta}$*;*

b) $\operatorname{var}(\mathbf{b}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$*;*

c) $\mathbf{b}$ *is best linear unbiased estimator (BLUE), which means that among all unbiased linear estimators,* $\mathbf{b}$ *has the smallest variance.*

d) OLS estimator for σ^2 *is unbiased,* $E(\tilde{\sigma}^2) = \sigma^2$*.*

Proof of Theorem 3.

a) Observe that:

$$\begin{aligned}
E(\mathbf{b}) &= E((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}) \\
&= E((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon})) \\
&= E(\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}) \\
&= \boldsymbol{\beta} + E((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}) \\
&= \boldsymbol{\beta} + E((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')E(\boldsymbol{\varepsilon}) \\
&= \boldsymbol{\beta} + E((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')E(\boldsymbol{\varepsilon}) \\
&= \boldsymbol{\beta}.
\end{aligned}$$

The equality in last two steps holds, because $\mathbf{X}$ and $\boldsymbol{\varepsilon}$ are independent an $E(\boldsymbol{\varepsilon}) = 0$.

b) Observe that:

$$\begin{aligned}
\operatorname{var}(\mathbf{b}) &= E((\mathbf{b} - E(\mathbf{b}))(\mathbf{b} - E(\mathbf{b}))') \\
&= E((\mathbf{b} - \boldsymbol{\beta})(\mathbf{b} - \boldsymbol{\beta})') \\
&= E((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}) \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}')\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\sigma^2\mathbf{I}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\
&= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1} \\
&= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}.
\end{aligned}$$

c) We need to prove that $\operatorname{var}(\mathbf{b}) \leq \operatorname{var}(\tilde{\boldsymbol{\beta}})$ componentwise for any estimator $\tilde{\boldsymbol{\beta}}$ of form $\tilde{\boldsymbol{\beta}} = \mathbf{A}\mathbf{y}$ with $E(\tilde{\boldsymbol{\beta}}) = \boldsymbol{\beta}$.
As $E(\tilde{\boldsymbol{\beta}}) = E(\mathbf{A}\mathbf{y}) = AE(\mathbf{y}) = \mathbf{A}\mathbf{X}\beta = \beta$, then it has to hold $\mathbf{A}\mathbf{X} = \mathbf{I}$. Using this equality in variance calculations, we get:

$$\begin{aligned}
\operatorname{var}(\tilde{\boldsymbol{\beta}}) &= \operatorname{var}(\mathbf{A}\mathbf{y}) = \operatorname{var}(\mathbf{A}(\mathbf{X}\beta + \epsilon)) \\
&= \operatorname{var}(\mathbf{A}\mathbf{X}\beta) + \operatorname{var}(\mathbf{A}\epsilon) + 0 \\
&= \operatorname{var}(\mathbf{A}\epsilon) = \sigma^2\mathbf{A}\mathbf{A}'.
\end{aligned}$$

Consequentially it holds:

$$\begin{aligned}
\mathrm{var}(\tilde{\boldsymbol{\beta}}) - \mathrm{var}(\mathbf{b}) &= \sigma^2\mathbf{A}\mathbf{A}' - \sigma^2(\mathbf{X}'\mathbf{X})^{-1} \\
&= \sigma^2(\mathbf{A}\mathbf{A}' - (\mathbf{X}'\mathbf{X})^{-1}) \\
&= \sigma^2(\mathbf{A}\mathbf{A}' - \mathbf{A}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{A}') \\
&= \sigma^2\mathbf{A}\underbrace{(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')}_{\mathbf{M_X}}\mathbf{A}'.
\end{aligned}$$

As the $\mathbf{M_X}$ is idempotent with eigenvalues 0 or 1 (and as such positive semidefinite), it follows that $\mathbf{A}\mathbf{A}' - (\mathbf{X}'\mathbf{X})^{-1}$ is positive semidefinite as well. From this it follows that $\mathrm{var}(\tilde{\boldsymbol{\beta}}) - \mathrm{var}(\mathbf{b}) \geq 0$ componentwise.

d) Denote by tr the trace of a given matrix. Then

$$\begin{aligned}
E(\tilde{\sigma}^2) &= \frac{1}{N-K}E(Q_e) = \frac{1}{N-K}E(\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}) \\
&= \frac{1}{N-K}E(\mathbf{y}'\mathbf{M}'_{\mathbf{X}}\mathbf{M_X}\mathbf{y}) = \frac{1}{N-K}E(\mathbf{y}'\mathbf{M_X}\mathbf{y}) \\
&= \frac{1}{N-K}E(tr(\mathbf{y}'\mathbf{M_X}\mathbf{y})) = \frac{1}{N-K}E(tr(\mathbf{M_X}\mathbf{y}\mathbf{y}')) \\
&= \frac{1}{N-K}tr(\mathbf{M_X}E(\mathbf{y}\mathbf{y}')) = \frac{1}{N-K}tr(\mathbf{M_X}(\mathrm{var}(\mathbf{y}) + E(\mathbf{y})'E(\mathbf{y})) \\
&= \frac{1}{N-K}tr(\mathbf{M_X}\Sigma) + \frac{1}{N-K}E(\mathbf{y})\mathbf{M_X}E(\mathbf{y}) \\
&= \frac{1}{N-K}tr(\mathbf{M_X}\Sigma) + \frac{1}{N-K}\boldsymbol{\beta}'\mathbf{X}'\mathbf{M_X}\mathbf{X}\boldsymbol{\beta}) \\
&= \frac{1}{N-K}tr(\mathbf{M_X}\Sigma) = \frac{\sigma^2}{N-K}tr(\mathbf{M_X}).
\end{aligned}$$

As $tr(\mathbf{M_X}) = tr(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}') = tr(I) - tr((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}) = N - K$ it follows $E(\tilde{\sigma}^2) = \sigma^2$.

□

If we make some further assumptions about the distribution of errors $\boldsymbol{\varepsilon}$, we can derive also some distribution properties of the OLS estimator. The proof of the following can be found in Fahrmeir, Hamrle and Tutz [13].

Theorem 4. *(Distribution properties of the OLS estimator)*
Let it hold $\mathbf{y} = \mathbf{X}\boldsymbol{\beta}$ *and* $\boldsymbol{\varepsilon} \sim N(0, \sigma^2\mathbf{I})$. *Then*

a) $\mathbf{b} \sim N_K(\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$;

b) $\frac{(\mathbf{b}-\boldsymbol{\beta})'\mathbf{X}'\mathbf{X}(\mathbf{b}-\boldsymbol{\beta})}{\sigma^2} \sim \chi^2_k$;

c) $\mathbf{b}$ *is independent from* $\tilde{\sigma}^2$;

d) $\frac{Q_e}{\sigma^2} = (N-k)\frac{\tilde{\sigma}^2}{\sigma^2} \sim \chi^2_{N-K}$.

Additionally, under assumptions of Theorem 4, each element in $\mathbf{b}$ is normally distributed, i.e.

$$\mathbf{b}_k \sim N(\boldsymbol{\beta}_k, \sigma^2 c_{kk}), \tag{25}$$

where c_{kk} presents the (k,k) element in $(\mathbf{X}'\mathbf{X})^{-1}$.

2.2.1.2 Asymptotic properties of the OLS in classical linear regression model

In previous section we spoke about small sample properties of OLS. We observed them under the standard assumptions ([A1]-[A4]) and made some additional assumptions about the distribution of $\boldsymbol{\varepsilon}$ as well. The problem arises when these conditions are violated; in such cases the small sample properties become typically unknown. At that point we have to use an alternative approach: what properties do our estimators have when sample size grows infinitely large?

In following lines we deal with asymptotic behaviour of estimators and in order to make some conclusions about that, we need to use some results on convergence of random variables. The reader unfamiliar with those can find them explained in every fundamental probability book.

Definition 5. Let $\tilde{\theta}_{\mathbf{N}} = \mathbf{g}(\mathbf{x_1}, \ldots, \mathbf{x_N})$ denote an estimator of θ, where the sample size is equal to N. Then $\tilde{\theta}$ is an

i) *asymptotically unbiased* estimator of $\tilde{\theta}$, when $E(\tilde{\theta}_{\mathbf{N}}) \to \theta$ as $N \to \infty$;

ii) *asymptotically efficient* estimator of $\tilde{\theta}$, when its asymptotic variance (variance of the asymptotic distribution of θ) is smaller than the variance of any other consistent estimator.

Definition 6. Let $\tilde{\theta}_{\mathbf{N}} = g(\mathbf{x}_1, \ldots, \mathbf{x}_N)$ denote an estimator of θ where the sample size is equal to N. We say:

i) Estimator $\tilde{\theta}$ is a *weakly consistent* estimator of θ, when $\tilde{\theta}_{\mathbf{N}}$ converges to θ in probability, $\text{plim}_{N\to\infty}\tilde{\theta}_{\mathbf{N}} = \theta$.

ii) Estimator $\tilde{\theta}$ is a *strongly consistent* estimator of θ, when $\tilde{\theta}_{\mathbf{N}}$ converges to θ almost surely, $P(\lim_{N\to\infty}\tilde{\theta}_{\mathbf{N}} = \theta) = 1$.

How to check the consistency of the OLS estimator? It can be expressed as:

$$\begin{aligned}
\mathbf{b}_N &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = (\frac{1}{N}\mathbf{X}'\mathbf{X})^{-1}\frac{1}{N}\mathbf{X}'\mathbf{y} \\
&= \boldsymbol{\beta} + (\frac{1}{N}\mathbf{X}'\mathbf{X})^{-1}\frac{1}{N}\mathbf{X}'\boldsymbol{\varepsilon}.
\end{aligned}$$

For the consistency we need convergence in probability of the second term. We want to show that for any $\delta > 0$

$$\lim_{N\to\infty} P(|\mathbf{b}_N - \boldsymbol{\beta}| > \delta) = 0$$

or

$$\lim_{N\to\infty} P(|(\frac{1}{N}\mathbf{X}'\mathbf{X})^{-1}\frac{1}{N}\mathbf{X}'\boldsymbol{\varepsilon}| > \delta) = 0.$$

With result that sample averages converge to population mean all we need for the plim of the OLS estimator to exists is:

[A6] The value $\frac{1}{N}(\mathbf{X}'_N\mathbf{X}_N)^{-1}$ converges to finite nonsingular matrix Σ_{XX} when $N \to \infty$. This means that asymptotically there should be no multicollinearity between regressors.

[A7] Disturbances have mean 0 and are independent from explanatory variables, i.e. $E(\mathbf{X}'\boldsymbol{\varepsilon}) = 0$.

Similarly, under conditions above, also $\tilde{\sigma}^2$ is a consistent estimator for σ^2.

Definition 7. Let $\tilde{\theta}_{\mathbf{N}} = \mathbf{g}(\mathbf{x_1}, \ldots, \mathbf{x_N})$ denote an estimator of θ, where the sample size is equal to N. Then $\tilde{\theta}$ is *asymptotically normally distributed* when

$$\sqrt{N}(\tilde{\theta} - \theta) \to N(0, \text{cov}(\theta)),$$

where $\text{cov}(\theta)$ is a covariance matrix of θ.

It can be shown that, under conditions [A1] - [A6], it holds:

$$\sqrt{N}(\mathbf{b} - \boldsymbol{\beta}) \to N(0, \sigma^2\Sigma^{-1}_{\mathbf{XX}})$$

in distribution, as $N \to \infty$.

2.3 Model specification and hypothesis testing

In principle there are two main strategies in building the appropriate (OLS) model (Green [15]): **simple to general** and **general to simple** approach. In simple to general case we initially use a small set of regressors and add additional variables later on. The issue which may be triggered here is that due to underspecification of the initial model selection of additional variable may be biased. In general to simple case we first run regressions on a wide pool of variables and reduce their number in order to obtain the preferred model. One of the main concerns in such approach is that some variables appear significant by chance.[3]

In a process of defining a model, one can help herself with a wide variety of diagnostic checks. Hereby we will not go into details about them, but will only mention some categories of them: *Omitted variable tests, RESET tests, Tests for functional forms, Tests for structural break, Tests for outliers, Tests for nonspherical disturbances, etc.*

Once the model is better defined hypothesis testing is also used to clarify whether their covariates are really significant or not. A short overview is given in the next chapter.

2.3.1 Hypothesis testing and interval estimation

In regressions the main task is to estimate the unknown parameters and based on their estimation predict the value of dependant variable in a yet unobserved environment. Because the estimation is not precise there should be some degree of uncertainty attached to it.

[3]In relation to these two strategies Kennedy [19, p.82] nicely describes three methodologies: (1) Average economic regression, (2) Test, test, test (TTT) and Fragility analysis. In the same chapter he also outlines the general principles for the model specification.

Based on our knowledge we are usually able to define the **null hypothesis,** H_0 related to the value of unknown parameter(s). Depending on it, we differentiate between **one-** and **two-sided tests**. An example of a null hypothesis in one-sided test would be $H_0 : \theta \leq \theta^0$ with alternative hypothesis $H_1 : \theta > \theta^0$; in two-sided test the null hypothesis would be $H_0 : \theta = \theta^0$ with alternative hypothesis $H_1 : \theta \neq \theta^0$.

Once the null hypothesis is defined, the **test statistic** with known distribution (assuming the validity of H_0) using our observations is calculated. The null hypothesis is rejected if a realized value of the test statistic is unlikely to belong to the assumed distribution, i.e. the probability of observing the realized value of the test statistic is smaller than a given **significance level** α. Usually $\alpha = 0.05$ or $\alpha = 0.01$.

When deciding about the validity of the null hypothesis there are two types of errors, one can make. **Type I error** is the rejection of the true null hypothesis. Based on the above mention rule type I error would occur with probability of significance level α. The smallest significance level at which the null hypothesis is rejected is also called **probability value** or ***p*-value**.

Another possible error is to not reject null hypothesis, when it is in fact wrong. This is called **type II error** and the probability of accepting the false null hypothesis is commonly denoted with β. Then the probability of rejecting the false null hypothesis is $(1 - \beta)$, which is known as the **power** of the test.

A **confidence interval** can be defined as an interval of all values θ_0, for which the null hypothesis is not rejected. If a parameter θ is estimated with $\tilde{\theta}$, then interval estimation would look like

$$\tilde{\theta} \pm \text{ sampling variability.}$$

Here the value of the sampling variability is such, that the probability of the true value lying in the estimated interval is high enough (usually we choose probability of 95% or 99%, although any other value would be perfectly fine as well).

In the following subsections the focus will be in hypothesis testing and interval estimation in a context of the CLR model and OLS parameters. Readers interested in more general approach to hypothesis testing and interval estimation are welcomed to have a look in Newbold [23] and Green [15].

2.3.2 The *t*- and *F*-test for OLS regression parameters

Two most commonly used tests in testing of the CLR model are ***t*-test** and an ***F*-test**. First one is testing a single linear restriction, whereas the second one is used when testing more of them. The simplest case of a t-test would so be a test on an individual partial regression parameter, testing the effect of specific independent variable $\mathbf{X}_j$ on values of explained variable, whereas the simplest case of an F-test would test whether the explanatory variables add any value when explaining the value of dependant variable.

Example 1. Example of t-test
Test problem:

$$H_0 : \beta_j = \beta_j^0 \qquad H_1 : \beta_j \neq \beta_j^0. \tag{26}$$

Test statistic:

$$t = \frac{\mathbf{b}_j - \beta_j^0}{\tilde{\sigma}\sqrt{c_{jj}}}.$$

A special case of t-test is a case where $H_0 : \boldsymbol{\beta}_j = 0$.

From Theorem 4 and (25) we can conclude that, under the null hypothesis, the distribution of the test statistic is equal do Student's t-distribution with $N - K$ degrees of freedom, $t \sim t(N - K)$. In case that σ is known, then the distribution of the test statistic is normal.

Test: reject H_0 if $P(|t| > t(N - K; 1 - \alpha/2)) = \alpha$, i.e.

$$|t| > t(N - K; 1 - \alpha/2).$$

The $(1 - \alpha)$-confidence interval for the $\boldsymbol{\beta}_j$:

$$\begin{aligned} -t(N - K; 1 - \alpha/2) &< \frac{\mathbf{b}_j - \beta_j}{\tilde{\sigma}\sqrt{c_{jj}}} < t(N - K; 1 - \alpha/2) \\ \mathbf{b}_j - t(N - K; 1 - \alpha/2)\,\tilde{\sigma}\sqrt{c_{jj}} &< \beta_j < \mathbf{b}_j + t(N - K; 1 - \alpha/2)\,\tilde{\sigma}\sqrt{c_{jj}}. \end{aligned}$$

Example 2. Example of F-test
Test problem:

$$H_0 : \beta_2 = \ldots = \beta_K = 0 \qquad H_1 : \beta_j \neq 0 \textit{ for at least one } j \neq 1. \tag{27}$$

Test statistic:

$$f = \frac{R^2(N - K)}{(1 - R^2)(K - 1)} = \frac{\mathrm{ESS}(N - K)}{\mathrm{RSS}(K - 1)}.$$

It can be shown that under the H_0 f has an F distribution: $F \sim F(K - 1, N - K)$ [15, p.117-118].

Test: reject H_0 if $P(f > F(K - 1, N - K; 1 - \alpha)) = \alpha$, i.e.

$$f > F(K - 1, N - K; 1 - \alpha).$$

2.3.3 General linear hypothesis

Both sets of hypothesis in (26) and (27) could be, together with a variety of other hypothesis, expressed as a general linear hypothesis:

$$H_0 : \mathbf{C}\boldsymbol{\beta} = \xi \qquad H_1 : \mathbf{C}\boldsymbol{\beta} \neq \xi, \tag{28}$$

where $\mathbf{C}$ presents $s \times K-$ matrix with a full rank (otherwise the constrains would be linear combinations of each other) and ξ is a vector of length s.

In order to test the null hypothesis, one has to first find a solution $\mathbf{b}_0$ of CLR model under the H_0. Our goal is to minimize the total sum of squares, subject to conditions on $\boldsymbol{\beta}$:

$$\min_{\boldsymbol{\beta}} \, (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \quad \text{s.t. } \mathbf{C}\boldsymbol{\beta} = \xi.$$

Using the method of Lagrange-multipliers (Fahrmeier et. al [13, p.111]), this is equivalent to finding the minimum of an unconstrained problem:

$$\min_{\boldsymbol{\beta},\lambda} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + \lambda'(\mathbf{C}\boldsymbol{\beta} - \xi).$$

Partial differentiation and setting of partial derivatives to zero gives us two equations, $(\mathbf{X}'\mathbf{X})\mathbf{b}_0 + \mathbf{C}'\lambda = \mathbf{X}'\mathbf{y}$ and $\mathbf{C}\mathbf{b}_0 = \xi$. Solving them yields a solution:

$$\mathbf{b}_0 = \mathbf{b} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'[\mathbf{C}(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{C}']^{-\mathbf{1}}(\mathbf{C}\mathbf{b} - \xi).$$

The value of the residual sum of squares under the null hypothesis is then equal to:

$$\begin{aligned}\boldsymbol{\varepsilon}_0 &= \mathbf{y} - \tilde{\mathbf{y}}_0 = \mathbf{y} - \mathbf{X}\mathbf{b}_0 = \mathbf{y} - \mathbf{X}\mathbf{b} + \mathbf{X}\mathbf{b} - \mathbf{X}\mathbf{b}_0 \\ &= \boldsymbol{\varepsilon} + \mathbf{X}(\mathbf{b} - \mathbf{b}_0).\end{aligned}$$

The difference between the RSS under the null hypothesis $Q_{e,0}$ and the RSS of a general solution Q_e is then equal to

$$Q_{e,0} - Q_e = \boldsymbol{\varepsilon}_0'\boldsymbol{\varepsilon}_0 - \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} = (\mathbf{C}\mathbf{b} - \xi)'[\mathbf{C}(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{C}']^{-\mathbf{1}}(\mathbf{C}\mathbf{b} - \xi).$$

It can be shown [13, p.111] that $Q_{e,0} - Q_e$ is independent of Q_e and

$$(Q_{e,0} - Q_e)/\sigma^2 \sim \chi^2(s). \tag{29}$$

As a test statistics one can than define

$$f = \frac{(Q_{e,0} - Q_e)(N - K)}{Q_e \ \ s}, \tag{30}$$

which is due to point d) of Theorem 4 and (29) $F(s, N - K)$ distributed. The H_0 is rejected if $f > F(s, N - K; 1 - \alpha)$.

2.3.4 Confidence intervals

In this subsection we will mention how to build confidence intervals for a single regression parameter and how for a part of them.

As already mentioned in previous subsections $(1 - \alpha)$-confidence interval for single partial regression coefficient is equal to the following interval:

$$\mathbf{b}_j - t(N - K; 1 - \alpha/2)\, \tilde{\sigma}\sqrt{c_{jj}} \quad < \quad \beta_j \ < \mathbf{b}_j + t(N - K; 1 - \alpha/2)\, \tilde{\sigma}\sqrt{c_{jj}}.$$

In case we want to build the confidence region for more of them together, i.e. for $(\beta_{m+1}, \ldots, \beta_{m+s})$, there exists possible options:

1. Following (27) the confidence region is defined as a region containing all such $(\beta_{m+1}, \ldots, \beta_{m+s})$, for which the calculated f-statistics would be smaller than $F(s, N - K; 1 - \alpha)$. This is an s-dimensional elipsoid.

2. One can construct simultaneous confidence intervals - separate confidence interval for $\beta_{m+1}, \ldots, \beta_{m+s}$ using various methods. For instance Bonferroni method or Union-intersection-principle.

Both of them are described in [13].

2.3.5 Goodness of fit and R^2

One of the most often used (and misused) measures to decide about the quality of the regression is **coefficient of determination** or $\mathbf{R^2}$, which is a ratio between the variance of $\mathbf{y}$ that is explained through regression and the variance of $\mathbf{y}$.

If $\tilde{\mathbf{y}} = \mathbf{Xb}$ and $\overline{\mathbf{y}}$ denotes the sample mean of $\mathbf{y}$, than we define R^2 as:

$$\begin{aligned} R^2 &= \frac{\text{ESS}}{\text{TSS}} = \frac{\text{var}(\tilde{\mathbf{y}})}{\text{var}(\mathbf{y})} \\ &= \frac{(\tilde{\mathbf{y}} - \overline{\mathbf{y}})'(\tilde{\mathbf{y}} - \overline{\mathbf{y}})}{(\mathbf{y} - \overline{\mathbf{y}})'(\mathbf{y} - \overline{\mathbf{y}})} = \frac{\sum_{n=1}^{N} (\tilde{\mathbf{y}}_n - \overline{\mathbf{y}})^2}{\sum_{n=1}^{N} (\mathbf{y}_n - \overline{\mathbf{y}})^2}. \end{aligned} \tag{31}$$

With the help of the decomposition of total variance, shown in (24), R^2 can be also written as:

$$R^2 = 1 - \frac{\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}{(\mathbf{y} - \overline{\mathbf{y}})'(\mathbf{y} - \overline{\mathbf{y}})} = 1 - \frac{\text{RSS}}{\text{TSS}}. \tag{32}$$

From here on some properties of R^2 are obvious:

- $0 \leq R^2 \leq 1$.
- $R^2 = 1$ if and only if $\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} = 0$.
- $R^2 = 0$ if and only if $(\mathbf{y} - \overline{\mathbf{y}})'(\mathbf{y} - \overline{\mathbf{y}}) = \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}$.

Higher R^2 means that the variance of the observed variable is better explained with the variance of explanatory variables. In that cases we usually conclude that regressors and model are appropriate. When R^2 is small, than we conclude that the regressors do not provide us with information about regressand.

Due to this property R^2 is often used when analysing goodness of fit. But it also has some limitations (Kennedy [19]):

- It grows with the number of explanatory variables - more variables are used in regression, higher it gets.
- R^2 is a measure of linear dependence so it does not tell us anything about explanatory power of regressors if we assume another (say polinomial) type of dependance.
- Although R^2 higher then 0.9 seems high and R^2 lower then 0.4 seems low in absolute terms, this unfortunately is not a good basis for comparison. In using time series data high R^2 are expected, and when processing panel data, already R^2 of 0.5 seems pretty high.

In order to solve the problem with growing R^2 when adding variables, we can use **adjusted** $\mathbf{R^2}$. The presented equation is only one possible way of modifying R^2 and not the only or necessary the best one:

$$\overline{R}^2 = 1 - \frac{N-1}{N-K}(1 - R^2). \tag{33}$$

Whether $\overline{R}^2$ declines or rises depends on whether the contribution of the new variable is bigger then correction for losing additional degree of freedom.

Besides R^2 and $\overline{R^2}$ one can use also other criteria for model comparison. Some of them are listed in Asteriou and Hall [3]:

- **AIC**, Akaike Information Criterion:

$$\text{AIC} = e^{\frac{2K}{N}}\left(\frac{\text{RSS}}{N}\right).$$

- **FPE**, Finite prediction Error:

$$\text{FPE} = \frac{N+K}{N-K}\left(\frac{\text{RSS}}{N}\right).$$

- **SBC**, Schwarz Bayesian Criterion:

$$\text{SBC} = e^{\frac{K}{N}}\left(\frac{\text{RSS}}{N}\right).$$

- **HQC**, Hannan and Quin Criterion:

$$\text{HQC} = (\log n)^{\frac{2K}{N}}\left(\frac{\text{RSS}}{N}\right).$$

2.3.6 Functional forms of linear model

Despite its name, the classical linear regression model, is not limited to a linear relationship between the dependent and independent variables. [4]

Consider first the classical linear model in its very general form:

$$g(y_i) = \beta_0 + \beta_1 f_1(\mathbf{z}_i) + \beta_2 f_2(\mathbf{z}_i) + \ldots + \beta_K f_K(\mathbf{z}_i) + \epsilon_i,$$

where $\mathbf{z}_i = (\mathbf{z}_{i1}, \ldots, \mathbf{z}_{iK})$ is an i-th observation of K variables. Functions $f_1, \ldots, f_K$ are linearly independent functions of $\mathbf{z}$ that map the K-dimensional vectors into L-scalars and function g is a univariate function of y.[5]

If the usual assumptions about disturbances and transformed independent variabilities are satisfied, then this model can be rewritten as a common multiple linear regression model with $K+1$ regressors:

$$\begin{aligned} g(y_i) &= \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_K x_{iK} + \epsilon_i \\ &= \mathbf{x}_i'\boldsymbol{\beta} + \epsilon_i. \end{aligned}$$

Some examples of such linear forms would be:

[4] The main focus of this thesis is the linear regression model, which is neither the only one, neither necessarily the best. Reader interested in other models (nonlinear, semiparametric and nonparametric ones) is invited to read chapters on this in, for example, Green [15].

[5] Please note that this model may be non-linear in variables, but is linear in parameters.

- Models with logarithmic transformations:
  The most widely used is log-log model, which is sometimes also referred to as log-linear model:

$$\log y_i = \beta_0 + \beta_1 \log z_{i1} + \beta_2 \log z_{i2} + \ldots + \beta_K \log z_{iK} + \epsilon_i. \tag{34}$$

  This model is convenient as the coefficient provide the elasticity (measurement of how responsive an economic variable is to a change in another) of the dependent variable with respect to the independent variable (omitting index i): $\beta = \frac{\partial \log y}{\partial \log x} = \frac{\partial y/y}{\partial x/x}$.
  However this model should not be confused with another two semi-log models: log linear model and linear log model. First one is logarithmic only in dependent variable and second one only in independent ones.

- Models with polynomial transformations;

- Models with inverse transformations;

- Models with binary/dummy variables;

- Models with spline functions;

- Piecewise linear regression.

2.3.7 Prediction

For now we have only spoken about how good the estimation of regression coefficient is and whether their values are significant or not. But the goals of any econometric model is of course not only to understand the relationship between dependent and independent variables, but also to give an approximation for not yet observed values.

Given that the model $\mathbf{y} = \mathbf{x}'\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ holds for all potential observations, it also holds for any given value of regressors, $\mathbf{x}_o$. Based on $\mathbf{x}_o$ one can use results from CLR model in order to estimate y_o as

$$\tilde{y}_o = \mathbf{x}_o'\mathbf{b}.$$

The prediction error is in that case equal to $y_o - \tilde{y}_o = \tilde{e}_o = \mathbf{x}_o'(\boldsymbol{\beta} - \mathbf{b}) + \epsilon_o$, where $E(\tilde{e}_0) = 0$ and $\text{var}(\tilde{e}_0) = \sigma^2\Big(\mathbf{x}_o'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_o + 1\Big)$.

If we assume the normal distribution of errors and no correlation between prediction error and variance, then it follows that

$$\frac{\mathbf{y}_o - \tilde{\mathbf{y}}_o}{\sqrt{\tilde{\sigma}^2(\mathbf{x}_o'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_o + 1)}} \sim t(N-K).$$

Then $(1-\alpha)$ prediction interval is equal to

$$\tilde{\mathbf{y}}_o - t(N-K; 1-\alpha/2)\,\tilde{\sigma}\sqrt{(\mathbf{x}_o'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_o + 1)} < \mathbf{y}_o <$$
$$< \tilde{\mathbf{y}}_o + t(N-K; 1-\alpha/2)\,\tilde{\sigma}\sqrt{(\mathbf{x}_o'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_o + 1)}.$$

In case prediction variance is known, the distribution of the test statistic is normal.

2.3.7.1 Prediction in case the regression model is logaritmic in dependent variable

Consider the log log model from (34):

$$\begin{aligned}\log y_o &= \beta_0 + \beta_1 \log z_{o1} + \beta_2 \log z_{o2} + \ldots + \beta_K \log z_{oK} + \epsilon_i \\ &= \mathbf{x}_o'\boldsymbol{\beta} + \epsilon_o,\end{aligned} \tag{35}$$

where we assume (for sake of simplicity) that error terms are normally distributed with mean 0 and variance σ^2.

'A naive approach' when predicting y_o is to use $\tilde{y}_o = e^{\mathbf{x}_o'\mathbf{b}}$. However, the issue in that case is

$$\begin{aligned}E(y|\mathbf{x}_o) &= E\left(e^{\mathbf{x}_o'\boldsymbol{\beta}+\epsilon_o} \mid \mathbf{x}_o\right) \\ &= e^{\mathbf{x}_o'\boldsymbol{\beta}}\, E\left(e^{\epsilon_o} \mid \mathbf{x}_o\right) \\ &= e^{\mathbf{x}_o'\boldsymbol{\beta}}\, e^{\sigma^2/2} \\ &> e^{\mathbf{x}_o'\boldsymbol{\beta}}.\end{aligned}$$

This implies that naive approach consistently under-predicts the conditional mean function of y and if we believe that conditional mean is the desired predictor for our y we should indeed include also the additional term ($e^{\sigma^2/2}$).

2.4 Violating assumptions of linear regression: Autocorrelation

In section 2.1 we have stated main assumptions od the linear regression model and also touched upon their violations. Each of them is in further details covered in numerous books on econometrics, nice overview is given in e.g. Kennedy [19] and Auer [4]. As the main focus of the following chapters is using the linear regression as a tool in estimating and forecasting financial time series (term premia) this section objective will be to explain violations of the assumption [A4], especially autocorrelation. Economic time series namely often display a 'memory', so the disturbances are not necessarily independent through time.

2.4.1 The generalized linear regression model

In the case that disturbances violate the assumption [A4] multiple linear regression model is extended to **generalized linear regression model**, where assumption [A4] is replaced by [A4*].

> **[A4*]** Variance $\text{var}(\boldsymbol{\varepsilon}\,|\mathbf{X}) = \sigma^2\Omega$, where $\boldsymbol{\Omega}$ is a positive definite matrix, that may depend upon $\mathbf{X}$.

The two leading cases, that are now allowed for, are heteroskedasticity and autocorrelation. First one implies that error terms are not identically distributed, whereas the second one implies dependencies among observations.

Variance var($\boldsymbol{\varepsilon}$ |**X**) **- heteroskedastic disturbances**:

$$\sigma^2\boldsymbol{\Omega} = \sigma^2 \begin{pmatrix} \omega_1 & 0 & \dots & 0 \\ 0 & \omega_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \omega_N \end{pmatrix}.$$

Variance var($\boldsymbol{\varepsilon}$ |**X**) **- autocorrelated (homoskedastic) disturbances**:

$$\sigma^2\boldsymbol{\Omega} = \sigma^2 \begin{pmatrix} 1 & \rho_1 & \dots & \rho_{N-1} \\ \rho_1 & 1 & \dots & \rho_{N-2} \\ \vdots & \vdots & \dots & \vdots \\ \rho_{N-1} & \rho_{N-2} & \dots & 1 \end{pmatrix}.$$

Heteroskedasticity and autocorrelation do not result in OLS estimator being biased and inconsistent, but only implies different covariance matrix for it. Consequently OLS estimator becomes inefficient and is not any more BLUE. The impact of changed covariance matrix can be summarized as:

- For unbiasedness of an OLS estimator only the assumption about zero error mean was used. As the error means are the same even under heteroskedasticity and autocorrelation **b** remains the unbiased estimator of $\boldsymbol{\beta}$. Detailed proof is given in Green [15, p. 259].
- Based on $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\,\mathbf{y} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\,(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\,\boldsymbol{\varepsilon}$ conditional variance of the OLS estimator can be calculated as:

$$\begin{aligned} \text{var}(\mathbf{b}\,|\mathbf{X}) &= \text{var}((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}\,|\mathbf{X}) \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\,\text{var}(\boldsymbol{\varepsilon}\,|\mathbf{X})\,\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\,\Omega\,\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}. \end{aligned} \tag{36}$$

  Common statistics, like t- and F-statistics would in these case have different distribution as before and comparing their values to those of the t- or F-distribution would give misleading results.

There are more ways of dealing with heteroskedastic and autocorrelated errors:

- respecify the regression model,
- find an alternative - best and unbiased estimator and
- use the OLS estimator with standard errors adjustments.

2.4.2 Alternative estimator for $\boldsymbol{\beta}$

The idea of finding the best linear unbiased estimator for $\boldsymbol{\beta}$ under [A4*] assuming $\boldsymbol{\Omega}$ is known is based on calculation of BLUE estimator under assumptions [A1] to [A4]. The aim is to transform a model in such a way that those assumptions would hold again.

As $\mathbf{\Omega}$ is a positive definite symmetric matrix (see [15]), there exists an orthogonal matrix Γ such that

$$\mathbf{\Omega} = \mathbf{\Gamma}\mathbf{\Lambda}\mathbf{\Gamma}',$$

where $\mathbf{\Lambda}$ is a diagonal matrix with eigenvalues of $\mathbf{\Omega}$ and columns of $\mathbf{\Gamma}$ are an orthonormal set of eigenvectors of $\mathbf{\Omega}$.

Let us define $\mathbf{T} = \mathbf{\Gamma}\mathbf{\Lambda}^{\mathbf{1/2}}$ and $\mathbf{P}' = \mathbf{\Gamma}\mathbf{\Lambda}^{-\mathbf{1/2}}$. Then it holds $\mathbf{\Omega} = \mathbf{T}\mathbf{T}'$ and $\mathbf{\Omega}^{-1} = \mathbf{P}\mathbf{P}^{-\mathbf{1}}$. With this factorization in mind we can rewrite generalized linear regression model into

$$\mathbf{P}y = \mathbf{P}\mathbf{X}\boldsymbol{\beta} + \mathbf{P}\boldsymbol{\varepsilon}$$

or

$$y^* = \mathbf{X}^*\boldsymbol{\beta} + \boldsymbol{\varepsilon}^*. \tag{37}$$

It holds:

$$E(\mathbf{P}\epsilon|\,\mathbf{X}) = \mathbf{P}E(\boldsymbol{\varepsilon}|\,\mathbf{X}) = 0$$

and

$$\begin{aligned}
\mathrm{var}(\mathbf{P}\epsilon|\,\mathbf{X}) &= \mathbf{P}\,\mathrm{var}(\boldsymbol{\varepsilon}|\,\mathbf{X})\mathbf{P}' = \sigma^2\mathbf{P}\mathbf{\Omega}\mathbf{P}' \\
&= \sigma^2\mathbf{P}(\mathbf{P}'\mathbf{P})^{-\mathbf{1}}\mathbf{P}' = \sigma^2\mathbf{P}\mathbf{P}^{-\mathbf{1}}(\mathbf{P}')^{-\mathbf{1}}\mathbf{P}' \\
&= \sigma^2\mathbf{I}.
\end{aligned}$$

Now $\boldsymbol{\varepsilon}^*$ satisfies conditions [A1] to [A4] and the BLUE estimator for $\boldsymbol{\beta}$ in the transformed model (37) is equal to:

$$\begin{aligned}
\hat{\mathbf{b}} &= (\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{X}^{*\prime}\mathbf{y} \\
&= (\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{y}.
\end{aligned}$$

Value $\hat{\mathbf{b}}$ is called **generalized least square estimator (GLS)** or Aitken estimator of $\boldsymbol{\beta}$.

In case $\mathbf{\Omega}$ is unknown, then we have to estimate it first and then use its estimate in deriving the $\hat{\mathbf{b}}$. This new estimator is then usually referred to **feasible or estimated generalized least squares estimator (FGLS/EGLS).**

2.4.3 Autocorrelation and time series

In case of time series modelling the most commonly used models to estimate the autocorrelation structure of the data are:

- **Autoregressive processes - AR:** The current value od dependent variable depends only on its past realizations (and on error term.)
- **Moving average processes - MA:** The current value od dependent variable depends only on the present and past values of a white noise disturbance term.
- **Autoregressive moving average processes - ARMA:** This model is a combination of the above mentioned models. The current values of dependent variable are a linear combination of its past observations and white noise error terms.

Definition 8. The time series process $(X_t)_{t\in\mathbb{Z}}$ is said to be *(weakly) stationary* if

i) $E(X_t^2) < \infty, \forall t \in \mathbb{Z}$;

ii) $E(X_t) = \mu, \forall t \in \mathbb{Z}$;

iii) $\text{cov}(X_t, X_s) = \text{cov}(X_{t+r}, X_{s+r}), \forall t, s \in \mathbb{Z}, r \in \mathbb{N}$.

Example 3. The simplest stationary process is a **white noise process**. A process $(X_t)_{t\in\mathbb{Z}}$ is called white noise, if $E(X_t^2) = \text{var}(X_t) = \sigma^2$, $E(X_t) = 0$ and $\text{cov}(X_{t+r}, X_t) = 0$, for $\forall t \in \mathbb{Z}, r \in \mathbb{N} \setminus \{0\}$. We write $X_t \sim WN(0, \sigma^2)$.

Example 4. The stationary process $(X_t)_{t\in\mathbb{Z}}$ is an autoregressive process AR(p) if it satisfies the linear difference equation:

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

where $\phi_p \neq 0$ and $Z_t \sim WN(0, \sigma^2)$.

Example 5. The stationary process $(X_t)_{t\in\mathbb{Z}}$ is a moving average process MA(q) if it satisfies the linear difference equation:

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

where $\theta_q \neq 0$ and $Z_t \sim WN(0, \sigma^2)$.

Example 6. The stationary process $(X_t)_{t\in\mathbb{Z}}$ is an autoregressive moving average process AR(p) if it satisfies the linear difference equation:

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

where $\phi_p \neq 0$, $\theta_q \neq 0$ and $Z_t \sim WN(0, \sigma^2)$.

To familiarize with them in more detail user is recommended to have a look in Mills and Markellos [22] or Hamilton[18], although this topic should be sufficiently covered in any basic econometric book.

2.4.4 OLS estimator with adjusted standard errors

For inference in the OLS models it is essential that we have a consistent estimator for variance of $\boldsymbol{\beta}$ as defined in (36). In previous section we covered the choice of estimator when we either know $\boldsymbol{\Omega}$ or some assumptions about it are made.

In case we do not know structure of $\boldsymbol{\Omega}$ the simplest approach would be to use the common OLS estimator. As already mentioned this would conclude in biased variance of an estimator and therefore misleading results. One can avoid this using so called **HC (heteroskedasticity consistent)** or **HAC (heteroskedasticity and autocorrelation consistent)** estimators.

Let us first rewrite the $\text{var}(\mathbf{b}|\,\mathbf{X})$ from (36) as done in Zeileis [36]:

$$\text{var}(\mathbf{b}|\,\mathbf{X}) = \left(\frac{1}{n}\mathbf{X}'\mathbf{X}\right)^{-1} \frac{1}{n}\,\Phi\,\left(\frac{1}{n}\mathbf{X}'\mathbf{X}\right)^{-1}, \tag{38}$$

where $\Phi = \frac{1}{n}\mathbf{X}'\mathbf{\Omega}\mathbf{X}$. Here Φ is essentially the covariance matrix of the scores (first derivatives of the log-likelihood functions): $\mathbf{V_i}(\boldsymbol{\beta}) = \mathbf{x}_i(\mathbf{y}_i - \mathbf{x}_i\boldsymbol{\beta})$.

If error terms are correlated and no other specifications about $\mathbf{\Omega}$ are given, its estimation is not feasible, as it would need and estimation of $n(n-1)/2$ parameters. However, one can estimate the variance of $\mathbf{b}$ using the estimation for $\mathbf{\Phi}$ which is much simpler. This tehnique is used by HAC estimators. Estimator for $\mathbf{\Phi}$ can be defined as:

$$\hat{\mathbf{\Phi}} = \frac{1}{n}\sum_{i,j=1}^{n} \mathbf{w_{|i-j|}}\hat{\mathbf{V}}_\mathbf{i}\hat{\mathbf{V}}_\mathbf{j}{}',$$

where $\mathbf{w} = (w_0, \ldots, w_{n-1})'$ is a vector of weights. In case of finite sample the expression is multiplied by $\frac{n}{n-k}$. For time series it is reasonable to assume that the correlation should decrease with the distance in time (increasing lag $l = |i-j|$), which results in decreasing w_l.

There exists multiple options in defining weights. Newey-West devised an estimator with linearly decaying weights:

$$w_l = \begin{cases} 1 - \frac{m}{L+1}, & \text{if } m = l \\ 0, & \text{otherwise.} \end{cases}$$

Here L is the maximum lag. This estimator is widely used in econometrics and is known as **Newey-West estimator.**

Other possible estimators might be as well [36]:

- *Hansen and Hodrick*: $w_l = 1$, others weights are set to 0.
- *Andrews*: $w_l = K(l/B)$, where K is a kernel function and B a bandwith parameter used.
- *Lumley and Heagerty*: They suggest using truncated weights (weights at lag l are equal to 1 if the autocorrelation is still present otherwise set to 0) or smoothed weights (weights at lag l are either 1 or autocorrelation at that lag, depending on what is smaller).

2.4.4.1 HAC estimators and overlapping data

When we need to compare two different datasets, they should ideally be independent or, in time series context, serial independent. Sometimes, this is not the case and we have to deal with the data sets, where some data points are the same. This is referred to as an overlapping data problem.

In financial time series analysis, we often need to use the values expressed in annual terms (i.e. annual return, annual revenue, etc.). In order to estimate them without breaking any of serial dependence constraints one should use annual data at estimation. However, as the financial markets data are usually unavailable for a really long periods researchers often use monthly data in order to achieve greater efficiency.

With usage of monthly data in estimating annual changes (or any other multi-period changes) we use the data of one month twelve times. This creates a moving average error term and causes a bias in the OLS estimates.

The simplest way to solve this issue is to either use yearly data or to model monthly changes instead of yearly ones. Other approaches would be to use averaged data. In this case the yearly data are recreated from monthly ones without using overlapping periods. A result is a smaller data set which contains all the original information. Unfortunately this does not necessary save the moving average problem, but it can also introduce additional forms of autocorrelation.

A valid option is also to use the overlapping data and to account for the moving average error term in hypothesis testing by using HAC estimators mentioned in subsection 2.4.4, which can provide asymptotically valid hypothesis tests.

3 Cochrane Piazzesi Approach

In seminar paper *Bond Risk Premia* [8], written by John H. Cochrane and Monika Piazzesi published in American Economic Review in 2005, authors study time-varying risk premium by running regressions of one-year access returns on (five) different forward rates available at the beginning of the period. Their work extends the classical regressions done by Fama and Bliss in 1978 and Campbell and Shiller in 1991, where only the ability of forward-spot spread to predict future excess returns is tested.

3.1 The empirical framework

3.1.1 Notations

Throughout the next sections the following notations will be used:

$P_t^{(n)}$	The nominal price of an n-year zero-coupon bond at time t paying 1 unit at maturity.
$p_t^{(n)} = \log P_t^n$	The log price of an n-year zero-coupon bond at time t.
$y_t^{(n)} \equiv -\frac{1}{n} p_t^{(n)}$	The log yield of an n-year zero-coupon bond at time t.
$f_t^{(n)} \equiv p_t^{(n-1)} - p_t^{(n)}$	The forward rate at time t for loans between time $t+n-1$ and $t+n$.
$r_{t+1}^{(n)} \equiv p_{t+1}^{(n-1)} - p_t^{(n)}$	The log holding period return from buying an n-year bond at time t and selling it at time $t+1$ as an $(n-1)$-bond.
$rx_{t+1}^{(n)} \equiv r_{t+1}^{(n)} - y_t^{(1)}$	The excess log return of an n-year bond.

3.1.2 Regression equations

Cochrane and Piazzesi ran regressions of bond excess returns at time $t+1$ on forward rates at time t. They used forward rates with maturities $2, 3, 4$ and 5 years, although one could also analyse the dependency between excess returns and forward rates of different maturities:

$$\begin{aligned} rx_{t+1}^{(n)} &= \beta_0^{(n)} + \beta_1^{(n)} y_t^{(1)} + \beta_2^{(n)} f_t^{(2)} + \beta_3^{(n)} f_t^{(3)} \\ &+ \beta_4^{(n)} f_t^{(4)} + \beta_5^{(n)} f_t^{(5)} + \epsilon_{t+1}^{(n)}, \qquad n = 2,3,4,5. \end{aligned} \tag{39}$$

After running regressions for all the maturities (from 2 to 5 years), plotting the regression coefficients on the same graph revealed a pattern in regression coefficients (Fig. 3). This led them to the idea that excess returns at all maturities could be expressed as a single linear combination of forward rates, where longer maturities just have greater loadings:

$$\begin{aligned} rx_{t+1}^{(n)} &= b_n(\gamma_0^{(n)} + \gamma_1^{(n)} y_t^{(1)} + \gamma_2^{(n)} f_t^{(2)} + \gamma_3^{(n)} f_t^{(3)} \\ &+ \gamma_4^{(n)} f_t^{(4)} + \gamma_5^{(n)} f_t^{(5)}) + \epsilon_{t+1}^{(n)}, \qquad n = 2,3,4,5. \end{aligned} \tag{40}$$

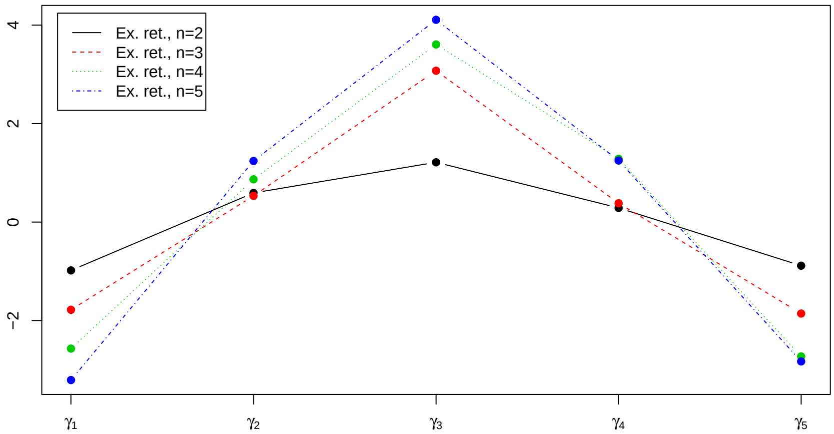


Figure 3: Coefficients from unrestricted regressions for different maturites for US yield curve building a 'tent' shape (adapted from Cochrane and Piazzesi [8]).

As without constraints on b_n and γ_n this equation has infinite number of solutions, they introduced two-step restricted regression, imposing that the average value of b_n is one.

Step 1

In Step 1 we identify 'the common factor' used for forecasting:

$$\begin{aligned} \overline{rx}_{t+1} = \frac{1}{4}\sum_{n=2}^{n=5} rx_{t+1}^{(n)} &= \gamma_0^{(n)} + \gamma_1^{(n)} y_t^{(1)} + \gamma_2^{(n)} f_t^{(2)} + \gamma_3^{(n)} f_t^{(3)} + \\ &+ \gamma_4^{(n)} f_t^{(4)} + \gamma_5^{(n)} f_t^{(5)} + \epsilon_{t+1}^{(n)} \\ &= \boldsymbol{\gamma}' \boldsymbol{f_t} + \bar{\epsilon}_{t+1}. \end{aligned} \quad (41)$$

Step 2

In Step 2 we estimate the loadings on specific maturities b_n:

$$rx_{t+1}^{(n)} = b_n(\boldsymbol{\gamma}' \boldsymbol{f_t}) + \bar{\epsilon}_{t+1}, \qquad n = 2, 3, 4, 5. \quad (42)$$

In this setting $\boldsymbol{\beta} = \mathbf{b}\boldsymbol{\gamma}'$ and a state variable explaining the time-varying expected returns is equal to $\boldsymbol{\gamma}' \boldsymbol{f_t}$.

These regressions (either one step or two-step approach) will henceforth be sometimes referred to as CP model, CP approach or CP regressions.

3.1.3 Data

The initial focus of the analysis is to test the approach from Cochrane and Piazzesi also on one of the European yield curves as only this will also provide information about the robustness of the model related to different underlying data. To ensure the yield rates represent market-determined yields, the country selection was limited to the countries that have robust and liquid financial markets. The need for a sufficiently large time sample was additional restriction; in the end the analysis was done on the dataset describing German yield curve.

The dataset itself is smoothed using Svensson methodology, which has its pros and cons. On a positive side it enables us to summarize the yield curve by the yields of zero-coupon bonds at annually spaced maturities. It removes differences in coupon rates, bond maturities, and individual bond idiosyncrasies, which allows a clean comparison of the results from different data sets (M. Ehrmann, M. Fratzscher, R.S. Gürkynak and E.T. Swanson [11]).

In an nutshell, Svensson model is an extension of more common Nelson Siegel model, which is usually used when fitting yield curve for monetary policy purposes. In this model the interest rate is the sum of constant and various exponential terms. Considering the fix maturity m, the interest rate is presented as a function of six parameters, given as (Svensson [31]):

$$
\begin{aligned}
y(m,\lambda,\tau,\beta_0,\beta_1,\beta_2,\beta_3) \quad = \quad & \beta_0 + \beta_1\left(\frac{1-\exp(-m/\lambda)}{m/\lambda}\right) + \\
& +\beta_2\left(\frac{1-\exp(-m/\lambda)}{m/\lambda} - \exp(-m/\lambda)\right) \\
& +\beta_3\left(\frac{1-\exp(-m/\tau)}{m/\tau} - \exp(-m/\tau)\right).
\end{aligned}
$$

One of the arguments against using a smoothed data set is that smoothing across maturities could result in lesser predictability. To check this point the regressions were ran also on non-smoothed German yield (strip) curve.

To complete the picture the following chapters also present results from running regressions on smoothed US yield curve (data from Gürkaynak, Sack and Wright [17][6]), alongside with the regression results from Cochrane and Piazzesi's article [8], which are based on CRSP data.

All the calculations in this chapter are done with programming tool R [25], additional non-standard packages that were used are packages *xts* [28], *pastecs* [16], *strucchange* [38], *tseries* [34], *urca* [24], *car* [14] and *lmtest* [37].

3.1.3.1 German yield curve - smoothed data set

In practice German Government securities are coupon bonds with various maturities. In order to easily separate between the results, due to maturity differences, and the results, due to coupons and coupon's related policies (e.g. taxation), it is convenient to express the yield curve by the yields of zero-coupon bonds of different (annual) maturities. To perform this fit, researches are most commonly choosing between Nelson-Siegel or Svensson model. The data used in this thesis are smoothed with Svensson model and are together with parameters used for fitting provided on web pages of Deutsche Bundesbank.[7]

The original data cover the period between September 1972 and September 2014 and include monthly information on German zero-coupon bonds of 1 to 10 and 15, 20 and 30 years of maturity. We compute continuously compounding forward rates and

[6] Available at `http://www.federalreserve.gov/pubs/feds/2006/200628/200628abs.html`.

[7] Available at `http://www.bundesbank.de/Navigation/EN/Statistics/statistics.html`.

realized one year excess returns according to the definitions in section 3.1.1. As the calculation of excess returns is calculated using information of 12 months ahead, the yields and forwards used in calculations effectively span the period between September 1972 and September 2013, whereas the excess return starts in September 1973 and ends with September 2014.

Tables 1 and 2 show some of the important descriptive statistics, Figures 4 and 5 present the yield, forward rates and excess return through time. The average yield curve is upward sloping, standard deviations of yields generally decrease with maturity. Excess returns are on average higher in case of holding and selling bonds with larger maturities, however they are at the same time also the most volatile.

Our variables are highly cross- and autocorrelated, all of them non-normally distributed taking the whole time series into account (test is done using Shapiro-Wilk test statistics). Dependent variables are potentially non-stationary, whereas the excess returns are stationary according to the augmented Dicky-Fuller test (ADF test). ADF test is a test for a unit root in a time series data with the null hypothesis, that time series possesses unit root and is therefore non-stationary. For the independent variables the test returns p-values of 0.13 (yield), 0.10 (f_1), 0.9 (f_2), 0.8 (f_3) and 0.7 (f_4), for the dependent ones it returns p-values smaller than 0.01. Although one can argue that regressing stationary time series with non-stationary ones will bring spurious result, in some cases regression may be justified - when the regression residuals are stationary with the characterises similar to those of dependent variable (Baffes [5]). Same issue exists also in Cochrane and Piazzesi's paper as also CRSP data are non-stationary (Ang and Piazzesi [1]). Similar observations are reported also by other authors (Donati and Donati[10]).

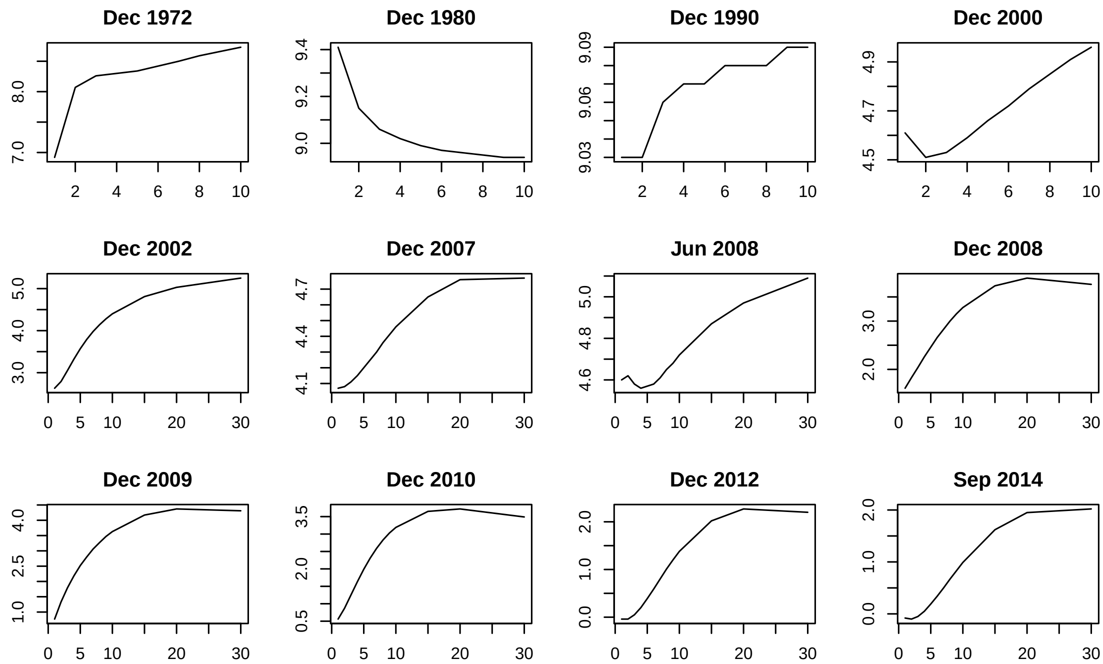


Figure 4: German yield curve through different time periods (yields vs. maturities); first row presents data before EUR currency, additional emphasis is given to the data in the beginning of the recent crisis (2007- 2009).

Table 1: German yield curve - descriptive statistics of the yield and forwards.

YIELD and FORWARD RATES (493 obs.)					
	yield 1 yr	fwd 1-2 yr	fwd 2-3 yr	fwd 3-4 yr	fwd 4-5 yr
mean	4.79	5.25	5.68	5.99	6.20
sd	2.66	2.51	2.41	2.30	2.18
skewness	0.28	-0.09	-0.26	-0.29	-0.28
kurtosis	-0.39	-0.68	-0.70	-0.67	-0.64
norm. test	0.98	0.98	0.98	0.98	0.98
test p-values	5.72 E-07	2.66 E-05	4.69 E-07	3.98 E-07	7.83 E-07
autocorr. (l=1)	0.99	0.98	0.99	0.99	0.99
autocorr. (l=12)	0.78	0.80	0.82	0.83	0.83
autocorr. (l=18)	0.64	0.70	0.74	0.74	0.74
correlations	1.00	0.97	0.93	0.90	0.87
		1.00	0.99	0.96	0.94
			1.00	0.99	0.98
				1.00	0.99
					1.00

Table 2: German yield curve - descriptive statistics of the realized excess returns.

REALIZED EXCESS RETURNS (493 obs.)					
	Ex ret 1 yr	Ex ret 2 yr	Ex ret 3 yr	Ex ret 4 yr	Mean ex ret
mean	0.64	1.27	1.77	2.18	1.47
sd	1.47	2.69	3.73	4.65	3.11
skewness	-0.35	-0.35	-0.42	0.48	-0.43
kurtosis	0.64	0.18	0.10	0.11	0.17
norm. test	0.99	0.99	0.99	0.98	0.99
test p -values	3.30 E-03	1.62 E-03	1.07 E-04	1.48 E-05	1.16 E-04
autocorr. (l=1)	0.93	0.94	0.94	0.95	0.95
autocorr. (l=12)	0.18	0.13	0.10	0.08	0.11
autocorr. (l=18)	0.03	0.04	0.03	0.01	0.03
correlations	1.00	0.98	0.95	0.92	1.00
		1.00	0.99	0.97	
			1.00	0.99	
				1.00	

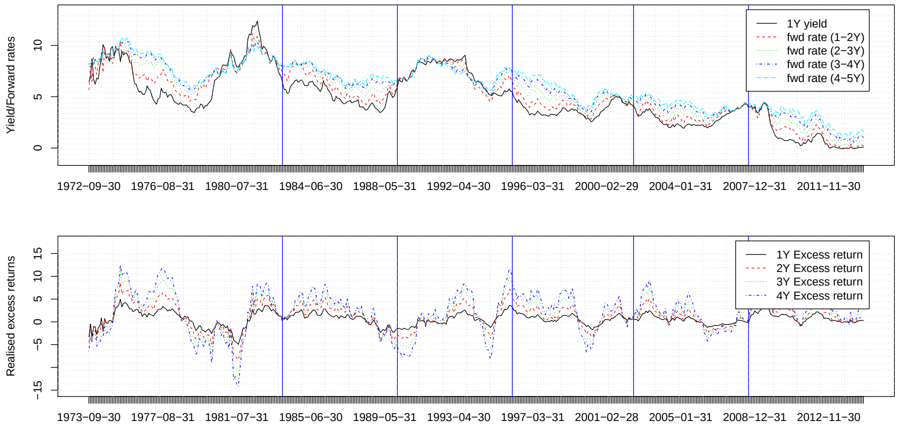


Figure 5: Evolution of the German yield and forwards (upper panel) and evolution of the term premia (lower panel).

3.1.3.2 German yield curve - stripped bonds

In order to check the influence of data smoothing (i.e. the usage of smoothed yields instead raw yields) this thesis includes also results obtained by running CP regression on non-smoothed data related to stripped German bonds. For strip bonds the principal amount is separated from coupon payments so they can be seen as equivalents of zero-coupon bonds.

Data span the time period from October 1997 to June 2014; calculations of the yields, forwards and excess returns are done in same fashion as for the German yield curve - smoothed data set. General statistics are provided in Tables 3 and 4.

Yields and forward rates still present the non-normality behaviour and high cross- and autocorrelations, however they seem to be less persistent. In contrast to the other data sets, realized excess returns appear to be normal (the test about normality cannot reject the null hypothesis that the data are normally distributed).

3.1.3.3 US yield curve

Similarly as for German yield curve data presenting zero-coupon bonds at annually spaced maturities for US market are estimated with Svensson model. The data presenting zero-coupon treasury yields are provided by Gürkaynak, Sack and Wright (2006) and are still frequently updated.

The original data cover the period between June 1961 and October 2014 and include monthly information on US zero-coupon bonds of $1, 2, 3, 4, 5, 7, 10, 25$ and 30 years of maturity. Tables 5 and 6 show some of the important descriptive statistics, Figure 6 presents the yield, forward rates and excess return through time.

The variables are highly cross- and autocorrelated, all of them non-normally distributed taking the whole time series into account. Dependent variables are potentially non-stationary, for excess returns ADF tests rejects null hypothesis about non-stationarity with p-value smaller than 0.01; regressors are however again non-stationary, with ADF test resulting p-values above 0.5. Similar observations are also valid for CRSP data set used by Cochrane and Piazzesi, as mentioned in Ang and Piazzesi [1].

Table 3: German yield stripped curve - descriptive statistics of yield and forwards.

YIELD and FORWARD RATES (189 obs.)					
	yield 1 yr	fwd 1-2 yr	fwd 2-3 yr	fwd 3-4 yr	fwd 4-5 yr
mean	2.49	2.86	3.28	3.63	3.77
sd	1.44	1.39	1.36	1.25	1.04
skewness	-0.28	-0.52	-0.77	-0.96	-0.63
kurtosis	-1.15	-0.65	-0.10	0.34	0.17
norm. test	0.94	0.95	0.93	0.91	0.96
test p-val	3.44 E-07	1.81 E-06	1.44 E-07	3.63 E-09	1.65 E-05
autocorr. (l=1)	0.98	0.97	0.97	0.96	0.95
autocorr. (l=12)	0.61	0.61	0.58	0.57	0.51
autocorr. (l=18)	0.42	0.48	0.47	0.45	0.43
correlations	1.00	0.96	0.91	0.86	0.82
		1.00	0.98	0.95	0.92
			1.00	0.98	0.96
				1.00	0.97
					1.00

Table 4: German yield stripped curve - descriptive statistics of realized excess returns.

REALIZED EXCESS RETURNS (189 obs.)					
	Ex ret 1 yr	Ex ret 2 yr	Ex ret 3 yr	Ex ret 4 yr	Mean ex ret
mean	0.59	1.27	1.89	2.27	1.51
sd	1.01	1.94	2.78	3.48	2.27
skewness	0.25	-0.01	-0.09	0.08	-0.06
kurtosis	-0.18	-0.56	-0.63	0.52	-0.61
norm. test	0.99	0.99	0.99	0.99	0.99
test p-val	3.66 E-01	5.45 E-01	1.35 E-01	4.08 E-01	2.03 E-01
autocorr. (l=1)	0.94	0.94	0.94	0.93	0.94
autocorr. (l=12)	0.01	-0.03	-0.07	-0.13	-0.08
autocorr. (l=18)	-0.06	-0.03	-0.02	-0.02	-0.03
correlations	1.00	0.98	0.94	0.88	1.00
		1.00	0.99	0.96	
			1.00	0.99	
				1.00	

Table 5: US yield curve - descriptive statistics of the yield and forwards.

YIELD and FORWARD RATES (629 obs.)					
	yield 1 yr	fwd 1-2 yr	fwd 2-3 yr	fwd 3-4 yr	fwd 4-5 yr
mean	5.29	5.72	6.01	6.23	6.41
sd	2.98	2.87	2.70	2.55	2.43
skewness	0.42	0.31	0.38	0.49	0.58
kurtosis	0.32	0.15	0.13	0.14	0.16
norm. test	0.97	0.98	0.98	0.97	0.96
test p-val	1.74 E-09	3.16 E-07	2.00 E-07	2.70 E-09	3.26 E-11
autocorr. (l=1)	0.99	0.99	0.99	0.99	0.99
autocorr. (l=12)	0.84	0.86	0.87	0.88	0.88
autocorr. (l=18)	0.74	0.79	0.81	0.82	0.82
correlations	1.00	0.98	0.95	0.93	0.90
		1.00	0.99	0.98	0.96
			1.00	0.99	0.98
				1.00	0.99
					1.00

Table 6: US yield curve - descriptive statistics of the realized excess returns.

REALIZED EXCESS RETURNS (629 obs.)					
	Ex ret 1 yr	Ex ret 2 yr	Ex ret 3 yr	Ex ret 4 yr	Mean ex ret
mean	0.48	0.82	1.09	1.29	0.9
sd	1.58	2.87	3.99	5.01	3.3
skewness	-0.06	-0.04	-0.01	0.02	0.0
kurtosis	0.67	0.50	0.41	0.38	0.4
norm. test	0.99	0.99	0.99	0.99	0.99
test p-val	1.06 E-04	3.33 E-04	8.78 E-04	3.70 E-03	8.63 E-04
autocorr. (l=1)	0.93	0.93	0.93	0.93	0.93
autocorr. (l=12)	0.20	0.16	0.14	0.11	0.14
autocorr. (l=18)	0.04	0.03	0.01	0.00	0.02
correlations	1.00	0.99	0.96	0.94	1.00
		1.00	0.99	0.98	
			1.00	0.99	
				1.00	

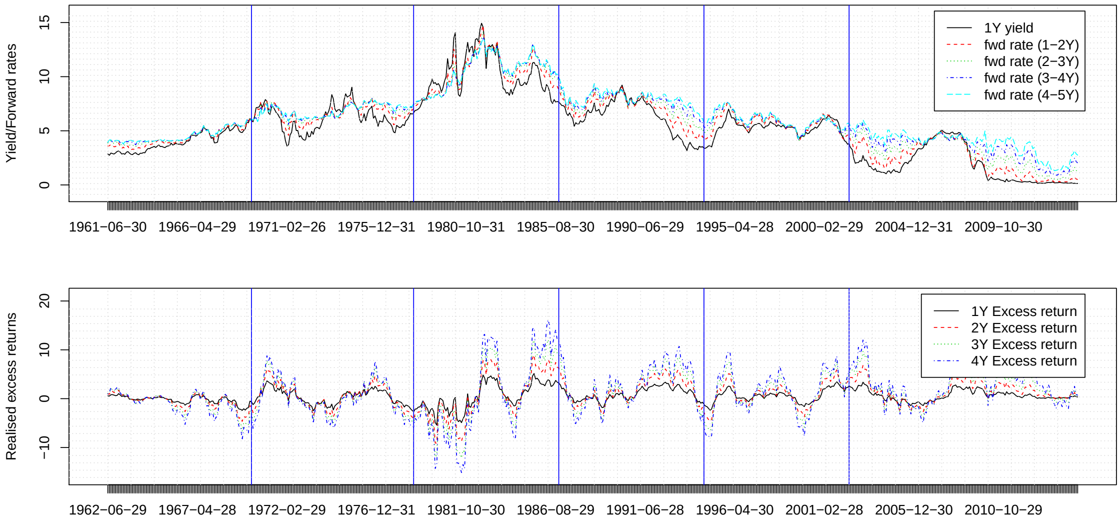


Figure 6: Evolution of the US yield and forwards (upper panel) and evolution of the term premia (lower panel).

3.2 CP Regressions on German yield curve data

3.2.1 Unrestricted regressions

First we ran unrestricted regressions for each of the excess returns (for $n = 2, 3, 4, 5$) variables on five independent variables. The results of regressions are presented in Table 7; first column provides betas from specific regression, following three provide standard errors and test statistics using common covariance matrix. In order to cater for correlation in used data, the results obtained by using Newey West covariance matrix with lag being equal to twelve (in the following chapters also referred to as NW covariance matrix) are also included. These results are presented in last three columns of the above mentioned table.

For all unrestricted regressions the t-statistics (which measure whether the specific coefficient is different from 0) calculated using standard covariance matrix are rather low. Effect of yield appears to be statistically significant in all cases, sometimes some significance is seen also for forward rates. On the other hand, F-statistics measuring joint significance of all variables appear to be rather high, higher that 99.9% critical value. This implies that indeed some of the variables do have explanatory power, which nevertheless, due to low adjusted R^2, appears to be rather weak. Using NW covariance matrix, we can conclude that chosen independent variables have little explanatory power, also joint significance as measured by F-statistics drops substantially. Same conclusions hold as well, if we are using Hansen and Hodrick corrected covariance matrix (results not presented).

In case of German yield curve the coefficients of regressions for different maturities follow the same pattern. This confirms the main message of the Cochrane and Piazzesi article, that the excess returns follow one-factor structure. On the Figure 7 one can clearly see that the coefficients form a pattern resembling 'W' shape. As in this case coefficients quickly jump between positive and negative values, one of the reasons behind the 'W' shape could be multicollinearity of dependant rates. To check this point we reviewed results of regressing the excess return only on yield, forward rate between 2 and 3 years and forward rates between 4 and 5 years. The results are similar to before, usually impact of the yield and one of the remaining forward rates appears to be significant (based on using non corrected covariance matrix). Corresponding F-statistics appear to be slightly higher, whereas R^2 remains practically unchanged. Based on this result we could indeed assume that additional variables do not add much explanatory value.

The distribution of the residuals appears to be non-normal, with slightly fatter tails. In our case it does not present the limitation per se, as our main goal is to generate the prediction which would minimize the (mean square) error. However, discrepancies from normal distribution can result in misspecified standard errors and test statistics. (Most probably quite big standard errors are mainly attributable to the non-normal distribution of residuals and multicollinearity effects.)

Table 7: German yield curve - unrestricted regressions results (using entire data set from 1972 to 2014).

UNRESTRICTED REGRESSIONS COEFICIENTS

Realized excess return, n=2

	Coef	SE	t	P(>|t|)		$SE^{(NW)}$	$t^{(NW)}$	$P(>\|t\|)^{(NW)}$
(Intc.)	-0.392	0.274	-1.434	0.152		0.734	-0.534	0.593
yield	-0.360	0.108	-3.334	0.001	***	0.313	-1.148	0.251
f1	0.842	0.390	2.159	0.031	*	0.80	1.053	0.293
f2	-1.944	1.174	-1.656	0.098	.	2.334	-0.833	0.405
f3	2.382	1.776	1.342	0.180		3.471	0.686	0.493
f4	-0.785	0.946	-0.830	0.407		1.931	-0.407	0.684
Adj R^2: 0.064, F-stat (5 and 487 DF): 7.734 and 1.400 $^{(NW)}$								

Realized excess return, n=3

	Coef	Std. Error	t value	Pr(>|t|)		$SE^{(NW)}$	$t^{(NW)}$	$P(>\|t\|)^{(NW)}$
(Intc.)	-0.429	0.50	-0.857	0.392		1.317	-0.326	0.745
yield	-0.575	0.198	-2.912	0.004	**	0.556	-1.034	0.301
f1	0.924	0.714	1.295	0.196		1.283	0.720	0.472
f2	-3.170	2.146	-1.477	0.140		4.010	-0.791	0.429
f3	5.181	3.247	1.595	0.111		6.194	0.836	0.403
f4	-2.158	1.730	-1.247	0.213		3.478	-0.621	0.535
Adj R^2: 0.067, F-stat (5 and 487 DF): 8.074 and 1.014 $^{(NW)}$								

Realized excess return, n=4								
	Coef	Std. Error	t value	Pr(>|t|)		$SE^{(NW)}$	$t^{(NW)}$	$P(>\|t\|)^{(NW)}$
(Intc.)	-0.317	0.690	-0.458	0.647		1.753	-0.181	0.857
yield	-0.797	0.273	-2.925	0.004	**	0.751	-1.062	0.289
f1	1.087	0.984	1.104	0.270		1.625	0.669	0.504
f2	-4.921	2.962	-1.661	0.097	.	5.334	-0.923	0.356
f3	8.619	4.481	1.923	0.055	.	8.380	1.029	0.304
f4	-3.775	2.387	-1.581	0.114		4.703	-0.803	0.422
Adj R^2: 0.073, F-stat (5 and 487 DF): 8.755 and 0.660 $^{(NW)}$								

Realized excess return, n=5								
	Coef	Std. Error	t value	Pr(>|t|)		$SE^{(NW)}$	$t^{(NW)}$	$P(>\|t\|)^{(NW)}$
(Intc.)	-0.332	0.857	-0.387	0.699		2.093	-0.159	0.874
yield	-1.007	0.338	-2.978	0.003	**	0.919	-1.096	0.273
f1	1.256	1.222	1.027	0.305		1.912	0.657	0.512
f2	-6.433	3.676	-1.750	0.081	.	6.495	-0.990	0.322
f3	11.136	5.562	2.002	0.046	*	10.234	1.088	0.277
f4	-4.731	2.963	-1.597	0.111		5.708	-0.829	0.407
Adj R^2: 0.080, F-stat (5 and 487 DF): 9.651 and 0.507 $^{(NW)}$								

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

5% and 1% critical val. of F-stat. on 5 and 487 DF are 2.23 and 3.05 respectively.

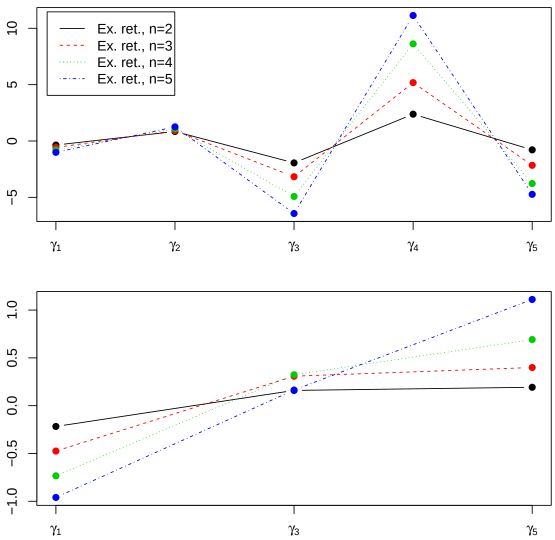


Figure 7: Unrestricted regression coefficients for German data set (1972-2014). Figure shows the repeating pattern of coefficients, with increasing loadings on higher maturities, regardless whether we regress on five (upper panel) or three (lower panel) independent variables.

As one could be surprised over the little explanatory value provided by results of unrestricted regressions, we have run regressions also on subsets of data, which were originally determined by checking for structural breaks. Structural breaks are points where regression coefficients change substantially. These (structural breaks) are depicted with blue lines on Figure 5. Coefficients obtained by running unrestricted regressions for different time periods are presented in Figure 8. One can see the main 'shape of coefficients' and how it is changing through time. Mainly there are four repeating patterns: first one before the fall of Berlin Wall, second one before the introduction of EUR in physical shape, third one before the crisis in 2007/08 and the last one since crisis' beginning until the present time.

Table 8 provides R^2 and NW F-statistics for those subsets. The resulting conclusions are now more similar to those of Cochrane and Piazzesi, although these depend on the small size of the subsamples.

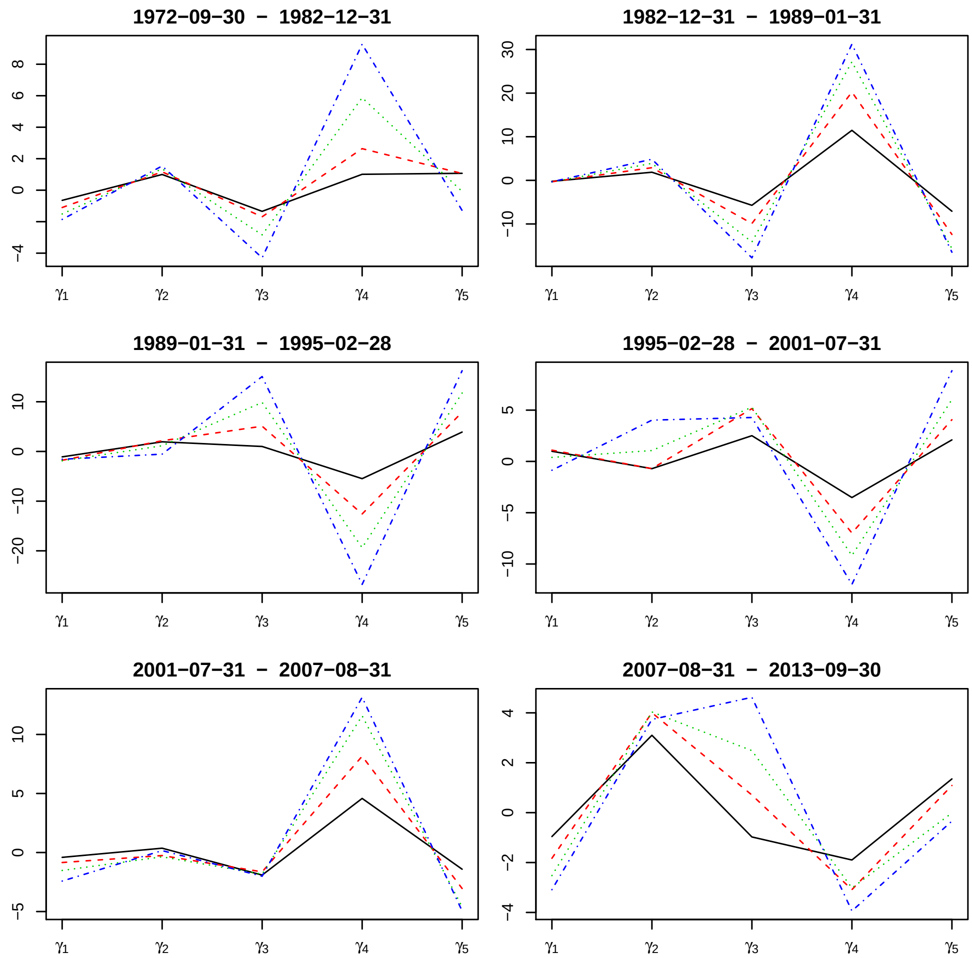


Figure 8: Changing pattern of the unrestricted regression coefficients for different time periods.

Table 8: German yield curve - unrestricted regressions results, split by periods.

UNRES. REGRESSION - RESULTS BY TIME PERIOD									
	R^2				F-stat (NW)				Obs.
Period	n=2	n=3	n=4	n=5	n=2	n=3	n=4	n=5	
1972-1982	0.352	0.430	0.456	0.460	10.103	14.753	16.214	16.112	124
1982-1989	0.456	0.391	0.343	0.305	0.981	0.840	1.422	2.781	74
1989-1995	0.096	0.101	0.128	0.160	0.187	0.358	0.644	1.090	74
1995-2001	0.712	0.731	0.717	0.691	62.242	43.582	25.523	16.026	78
2001-2007	0.860	0.853	0.848	0.841	140.85	157.62	172.130	180.380	74
2007-2013	0.696	0.618	0.530	0.445	26.205	14.522	8.286	5.037	74
1972-2014	0.064	0.067	0.073	0.080	1.400	1.014	0.660	0.507	487

3.2.2 Restricted regressions

One of the key messages from the Cochrane and Piazzesi article was, that the term premium can be partially explained also by a common factor (revealed by the pattern of the coefficients). As this is also seen in regressions for the German yield curve data, this section presents results of the restricted regressions, despite outcome of the unrestricted ones is not so promising.

Based on the observations of the pattern of coefficients through time the series is split into three parts: first part before 1989, second one from 1989 until 2001 and last one since 2001. This follows the logic of splitting set on data for West Germany, data for Bundesrepublik Deutschland with Deutsche Marke curve and EUR curve. For sake of space efficiency this section provides only the stepwise analysis of the first part, analysis of the remaining two we will touch upon later.

First step of the two-step regression is to find a common single factor, which could drive all the excess returns. Running the regressions, as described in Section 3.1.2, we get the result in Table 9. Besides the coefficients, which present a single factor, table also includes standard errors and values of t-statistics. These are calculated with common covariance matrix and with Newey West (NW) and Hansen Hodrick (HH) corrected covariance matrix, both using lag equal to 12. Second step of the restricted regression, which finds how much single factor affects excess return of a specific maturities are presented in Table 10.

Single factor forms a similar pattern, which was observed in unrestricted regressions. Corresponding t-statistics and F-statistics computed using common covariance matrix are significant, but one should note that these are calculated assuming normal distribution of errors. As the residuals are not normally distributed, but a distribution is slightly skewed and flattened , the results are to be interpreted with due caution (Fig. 10).[8] Additionally residuals present also high levels of autocorrelation, as seen on ACF and PACF plots (Fig. 11). This again does not affect

[8]Despite non-normality flags some possible misspecification, the issue of non-normality of residuals was left a bit aside, as the goal of this thesis was mainly to estimate the coefficients and generate predictions using regressions as proposed by Cochrane and Piazzesi.

coefficients of the regression, but only implies smaller standard errors and larger test statistics. To overcome these issues, Table 9 presents standard errors and test statistics calculated using NW and HH covariance matrix as well. We can not rule out the null hypothesis that each standalone coefficient is equal to zero (except for the intercept), however all together they appear to be significant (F-statistic presenting the joint significance is higher as 1% critical values).

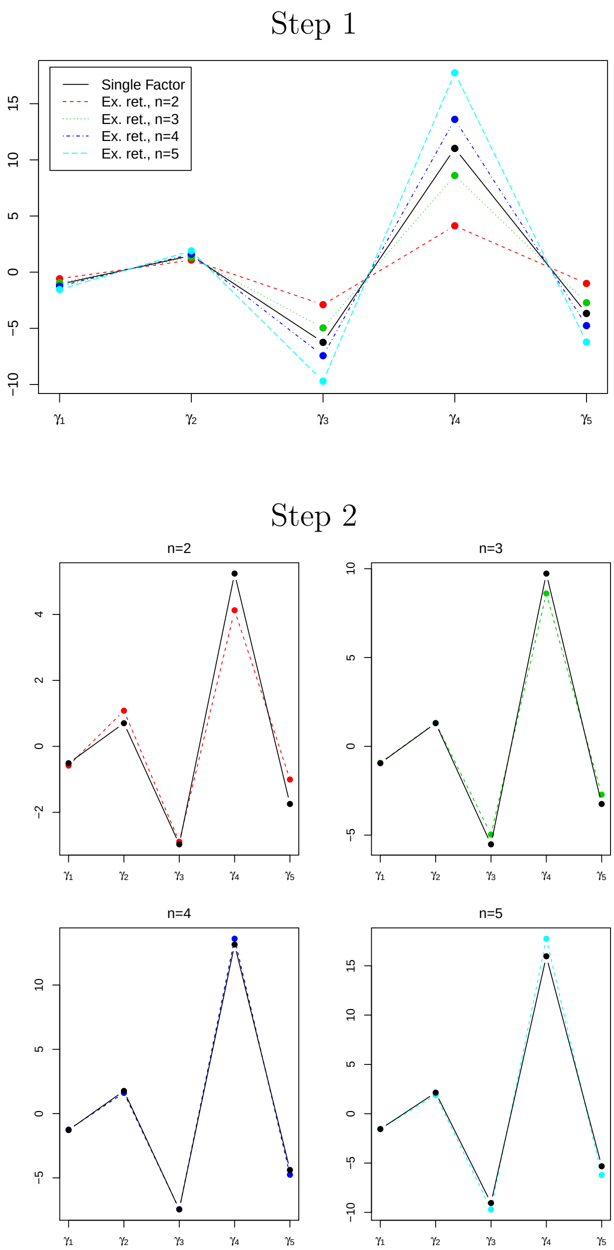


Figure 9: Comparison of restricted and unrestricted regression results for German yield curve (1972-1989). Upper panel depicts coefficients retrieved by unrestricted regressions alongside with the coefficients from Step 1 of the restricted CP model. Lower panel consist of four smaller figures. Each of them plots $b_n\boldsymbol{\gamma}$ and compares it with β_n.

Table 9: Coefficients of restricted regressions results for Step 1 regression on German yield curve data (1972-1989). Coefficients present the single factor which drives excess return of the bonds with different maturities. First third of the table contains results using standard covariance matrix, second third uses Newey West corrected covariance matrix and the remaining third presents results calculated with Hansen Hodrick covariance matrix. These two last cases should cater for the autocorrelation of residuals.

RESTRICTED REGRESSION - STEP 1								
	(Intc.)	yield	f1	f2	f3	f4	R^2	F-stat.
Coef	-11.688	-1.075	1.478	-6.249	11.021	-3.675	0.293	
SE	1.714	0.274	0.927	3.048	4.835	2.678		16.26
t	-6.818	-3.927	1.594	-2.050	2.279	-1.373		
Pr(>|t|)	0.0	0.0	0.113	0.042	0.024	0.172		
$SE^{(NW)}$	3.981	0.598	1.431	5.409	7.991	4.226		5.098
$t^{(NW)}$	-2.936	-1.797	1.032	-1.155	1.379	-0.870		
$Pr(>\|t\|)^{(NW)}$	0.004	0.074	0.303	0.250	0.170	0.386		
$SE^{(HH)}$	4.573	0.677	1.581	6.021	8.783	4.543		4.570
$t^{(HH)}$	-2.556	-1.588	0.934	-1.038	1.255	-0.809		
$Pr(>\|t\|)^{(HH)}$	0.011	0.114	0.351	0.301	0.211	0.420		
5% and 1% critical val. of F-stat. on 5 and 179 DF are 2.265 and 3.121 respectively.								

Table 10: Coefficients of restricted regressions results for Step 2 regression on German yield curve data (1972-1989). Coefficients present loadings b_n on the single factor $\boldsymbol{\gamma f_t}$. They slowly increase with maturity.

RESTRICTED REGRESSION - STEP 2				
	n=2	n=3	n=4	n=5
Coef	0.475	0.882	1.193	1.449
SE	0.014	0.010	0.005	0.020
$SE^{(NW)}$	0.036	0.029	0.010	0.055
$SE^{(HH)}$	0.044	0.035	0.012	0.067
R^2	0.249	0.292	0.318	0.331

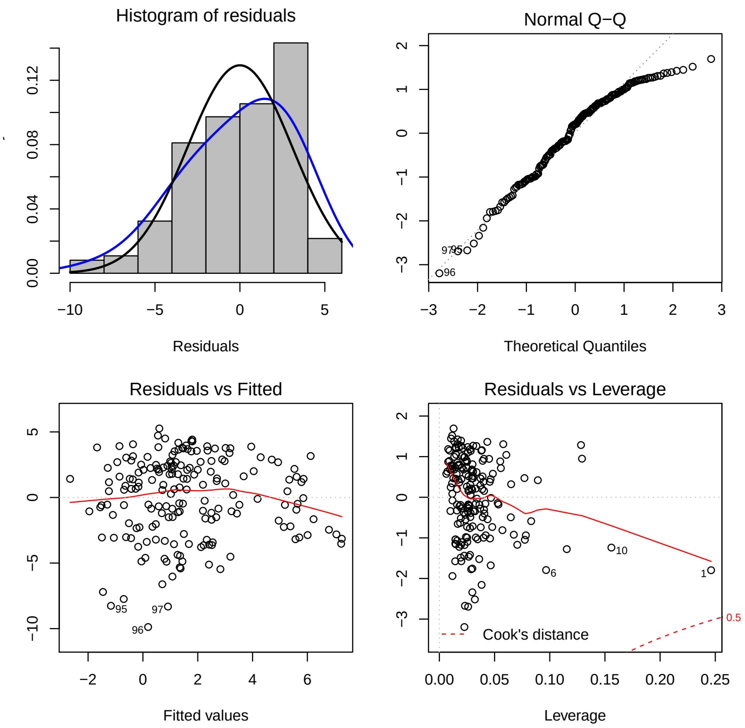


Figure 10: The distribution of the residuals is slightly negatively skewed and with heavier left tail than in case of normal distribution. This is not an issue per se, however it is harder for us to interpret the t-values and statistical significance of particular coefficient.

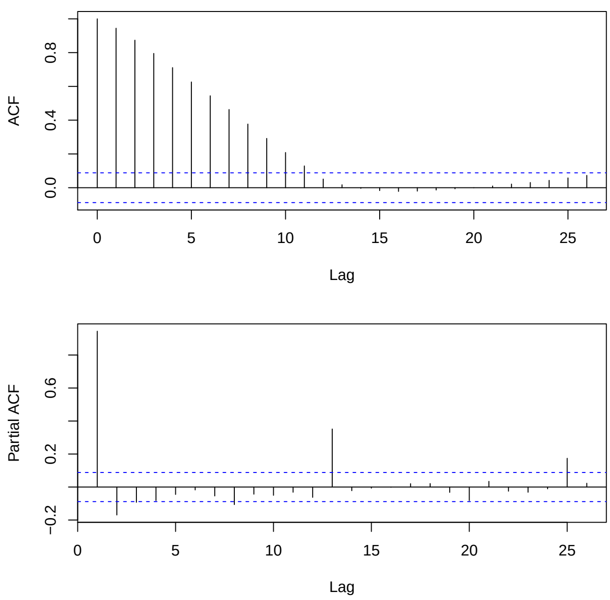


Figure 11: ACF and PACF of the residuals (restricted regression step 1).

As interesting additional information Table 11 presents F-statistics of different models. It includes the results from 'No overlap model' and results from regressions with lagged independent variables. In 'No overlap model' only yearly data are taken into account; regressions are ran on January data, February data, etc. Once these twelve regressions are ran the average covariance matrix of the twelve individual covariance matrices is used to calculate the F-statistics. In regressions with lagged variables the F-statistics is calculated with NW corrected covariance matrix. By using lagged data we omit the issue of having the same variable (yield at time t) present on both side of the regression equation.

There is obviously large difference between standard case and 'No overlap' case as the second covariance matrix is quite larger- we namely throw away the information by skipping overlapping data. The results calculated by lagged data are the same or even better as those calculated with non-lagged data. This points out that the high statistics are not necessarily connected to measurements errors.

Table 11: F-statistics, measuring joint significance presented for different models (German yield curve, 1972-1989).

MODEL COMPARISON - F-STAT.							
	Std.	No overlap	HH	NW	1m lag*	6m lag*	1yr lag*
F-stat.	16.261	0.839	4.570	5.098	4.656	9.097	6.844

*In models with lags F-stat. is calculated using NW covariance matrix.

3.2.3 Comparison of the regression results for different datasets

In order to properly check robustness of the model one has to run the model on different underlying data and see whether the conclusions based on different data sets remain the same.

As already mentioned, one of the advantages and at the same time drawbacks of the smoothed German yield curve dataset is in fact smoothing itself. On one side cleaning the data by removing some data specific idiosyncrasies is in general positively perceived, as it enables better comparison across different data sources, but on other side it potentially removes also the information needed for forecasting.

Additionally it is also know that smoothing also increases levels of multicollinearity (J. Annaert, A.G.P. Claes, M.J.K. De Ceuster and H.Zhang [2]). It is reasonable to say that some level of the multicollinearity when predicting with the forward rates is due to the nature of the problem itself; however, in case of smoothed datasets, multicollinearity may be even more heavily expressed, as we fit the yield curve at different maturities with the same coefficients. For these reasons some additional analysis is done also on US yield curve data set and German stripped bonds curve (Table 12).

Comparing the same data set - smoothed German yield curve data - throughout different periods reveals the dependency of the model on the movement of interest rates in the time period itself. For the whole period, R^2 is small and F-statistics is smaller than 95% and 99% critical values (which are equal to 2.232 and 3.054 respectively). The result gets better once splitting the whole set in smaller subsets: R^2 gets substantially higher across all subsets, F-statistics are now all above 95% critical values, value for first and last period even above corresponding 99% critical values.[9] One of the reasons behind underperformance of the model on the whole time period relative to the performance of the model on the subsets of data can also be that the coefficients differ quite a lot depending on the observed period, meaning that the underlying factor explaining excess returns is changing through time significantly.

Our initial guess was that some of the forecast may be lost due to the nature of the data (smoothing). Comparing the results of smoothed data set with the interpolated data set can not confirm that. F-statistics is not higher then its critical values, also R^2 is either in line or lower in comparison with our previous observations. However, the point that smoothed data possess larger multicollinearity as interpolated data is more clear; observing the pattern of coefficients for interpolated data we can see a tent shaped pattern, values not jumping from negative to positive values.

The third and fourth panel of the Table 12 present results for smoothed US yield curve and CRSP data. The values of both sets are really similar, with a bit larger discrepancies in the beginning of the time series and an average difference of 3%. The conclusions, which one can make using CRSP data (the data set used by Cochrane and Piazzesi) are much clearer than by using smoothed yield curve data. The R^2 for interpolated zero-coupon US curve is twice as high and F-statistic is also more convincing. Here we need to mention, that F-statistic is in both cases above the 99% critical value and the null hypothesis, stating that there is no dependency of excess return from forward rates can be rejected. Looking also at coefficients, we can, similarly to the case of German bonds, observe two different pattern of coefficients. Smoothed yield curve coefficients again form a 'zig-zag' pattern, whereas interpolated curve forms a tent-shaped pattern of coefficients (Fig. 12).

Comparing all sets one can say, that in case of the US data there are some empirical evidence against the expectation hypothesis, whereas there is less rejection (no rejection) in case of German bonds. In case we focus on smaller subsets then our conclusions could as well be against the expectation hypothesis, however we are not really convinced due to small samples and non-homogeneous result.

3.3 Prediction

In this subsection we will try to evaluate how good is CP model by predicting the excess returns and compare it to simple ARMA model. The fitting/forecasting the values will be done using both, restricted and unrestricted version of the CP approach, for ARMA model we have chosen MA(12) process as this was implied by overlapping data issue. For both models (CP and ARMA) in-sample and out-of-sample predictions using full data sample and rolling data sample will be presented. As in subsection 3.2.2 the forecasting exercise will be performed on German smoothed yield curve.

[9] Critical values of F-test statistics are: $F(0.95, 5, 191) = 2.261, F(0.99, 5, 191) = 3.114,$ $F(0.95, 5, 144) = 2.277, F(0.99, 5, 144) = 3.147, F(0.95, 5, 141) = 2.278, F(0.99, 5, 141) = 3.150$

Table 12: Comparison of coefficients, R^2 and F-statistics for different data sets (Line *US: CP Results* presents coefficients from [8]).

COMPARISON OF RESULTS AMONG DIFFERENT DATA SETS

	(Intc.)	yield	f1	f2	f3	f4	R^2	F	Obs.
Smoothed German yield curve data - different time periods									
1972-2014	-0.367	-0.685	1.027	-4.117	6.829	-2.863	0.072	2.151	505
1972-1989	-11.688	-1.075	1.478	-6.249	11.021	-3.675	0.293	5.098	197
1989-2001	-5.773	-2.579	2.717	4.839	-11.770	7.752	0.242	2.880	150
2001-2014	-2.583	-1.969	5.575	-5.232	-2.404	4.889	0.130	3.551	147
German stripped bonds curve data									
1997-2014	0.442	-1.040	-0.061	0.908	1.516	-1.235	0.113	2.209	189
Smoothed US yield curve data									
1961-2014	-1.869	-2.671	7.038	-10.359	6.660	-0.404	0.161	5.119	629
US: CP results									
1964-2004	-3.244	-2.135	0.808	3.001	0.801	-2.076	0.344	22.157	468

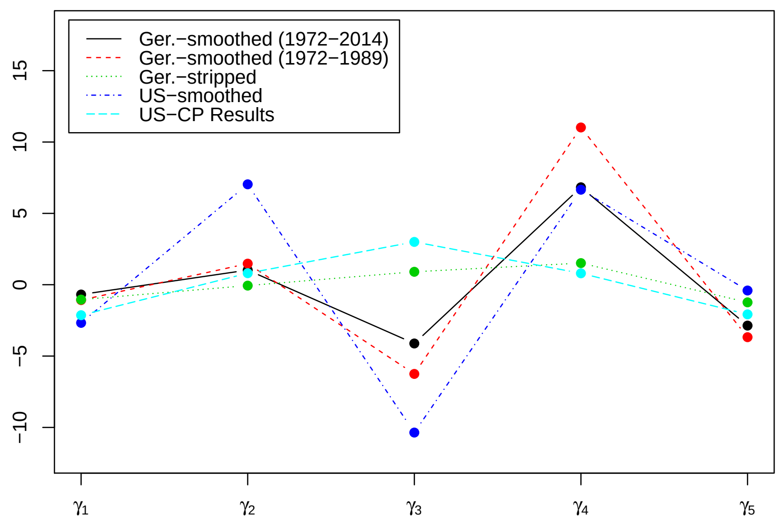


Figure 12: Pattern of coefficients (single factor) for different data sets (*CP Results* presents coefficients from [8]).

First kind of in-sample prediction (fit) will be based on results of regressing excess return on forward rates for each subset separately. In case we will be fitting the value as of 30.09.1995 we will use the coefficients of regression for subset 1989 -2001, in case we will be fitting the value of the curve as of 30.09.1985 we will use the regression results for subset 1972-1989. MA(12) fit will be calculated in similar fashion. In Tables 13, 14, 15 and 16 you can find those models marked as *Restricted m.- whole dataset, Unrestricted m.- whole dataset* and *MA(12)- whole dataset*.

For rolling in-sample prediction, we will only be using the data which were available until and including the time point, for which we are predicting the value of the excess return. The window will either span all past observations or only 5 years of past observations. In this later case the main idea is to try to capture 'one business cycle of data'.[10] In previously mentioned tables the two variations are marked with - *rolling since beginning* and - *rolling window of 5 yr*.

For the out-of-sample forecasts the procedure will be the same as in the rolling in-sample predictions with only one difference. Regressions are ran only on data points observed either one, six or twelve months in advance of the prediction point. The results for one month, six months and twelve months forecasts are to be found in Tables 14 to 16.

For the measures of (forecast) accuracy we have selected mean absolute error, root mean square error and coefficient of 'Direction symmetry'. They are defined as:

- *Mean absolute error* is (as name implies) an average of the absolute errors:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{i=N} |e_i|.$$

- *Root mean square error* is a quadratic measure based on an average magnitude of the error:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{i=N} (e_i)^2}.$$

- *Direction symmetry coefficient* measures in how many cases prediction moves in the same direction as observation:

$$\text{DS} = \frac{1}{n}\sum_{i=1}^{i=N} \mathbb{1}_{\{prediction\ and\ model\ move\ in\ same\ direction\}}.$$

Looking at the Table 13, panel A and B, it is clear that for German yield curve the best fit as measured by MAE and RMSE is obtained by MA(12) model, which is best calibrated when using complete dataset. Unrestricted and restricted model appear to be similarly successful at fitting the curve, in both cases there is small outperformance when calibrating with whole dataset as opposed to rolling window of five years (Fig. 13 and Fig. 14). As both error terms, MAE and RMSE lie close together for majority of the alternatives, it is fair to assume that the residuals (errors) are quite evenly distributed (of similar magnitude) across the whole sample.

[10] The choice for length of business cycle was arbitrary. According to the average length of business cycle in modern history is 5 years.

Interesting enough, when trying to analyse how often the excess return and its fit move into the same direction, one can see all models as similarly effective. The rate of success varies between 60% and 70%, so slightly better then simple guessing. The MA(12) models still appear to be better at fitting even when using DS as a measure of performance, however the differences between all models are really small.

When doing predictions for the one month in the future conclusions[11] are similar as when accessing the quality of in-sample fit (Table 14). Using MAE and RMSE the MA(12) models seem to provide the best fit, followed by restricted and unrestricted model calibrated used 5 years rolling window (Fig. 15). Also in this case restricted and unrestricted model behave quite similar, i.e. values of error terms are more or less the same regardless of whether we use restricted or unrestricted version of the model.

The DS coefficient is smaller than when fitting (which is perfectly understandable), but still above 50%. It is the highest for restricted- and unrestricted- rolling window of 5 yr models with average DS ratios of 63% and 62% (respectively).

Tables 15 and 16 present all three measures of accuracy for six and twelve months predictions, predictions are presented on Figures 16 and 17. The MAE and RMSE error terms are here by far the largest for restricted model, almost doubling with increasing maturity (for $n = 5$ the error terms are around eight times larger as for the $n = 2$ case). Albeit this trend is also observed in unrestricted and MA(12) model, it appears there in much lesser scale. Both models provides similar results, with MA(12) error terms being slightly smaller.

The restricted model, despite being 'the worst' model when using MAE and RMSE, provides really good guesses, whether the excess return will increase or decrease. It correctly assumes the increase/decrease in more than 60% of the cases, sometimes even in more than 70%. Figure 17 clearly depicts that restricted model usually overshoots the excess return, however it follows the similar shape. Opposite to before, the MA(12) model appears to be 'the worst' in this respect, it correctly predicts the direction in around or less than 50% - the result which we would obtain by simple coin tossing.

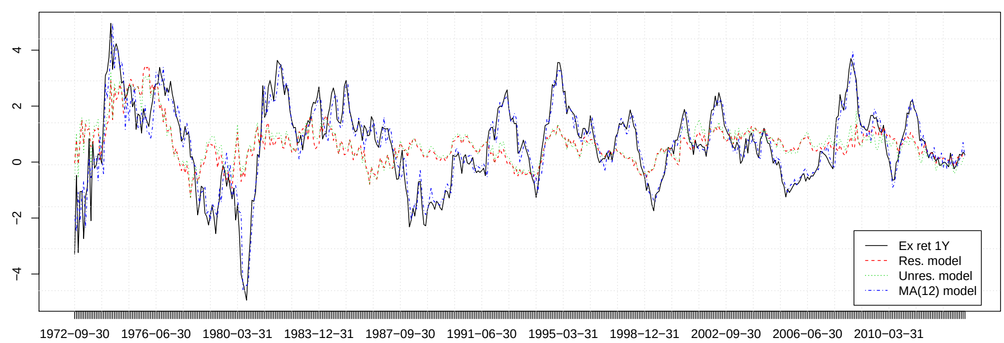


Figure 13: In-sample prediction of 1-year ($n = 2$) excess return using complete data set, as calculated by one and two-step CP model and MA(12) model.

[11]Even-dough the conclusions are similar, one should note that the ratio between the largest and smallest error decreased from ≈ 4 to ≈ 2.

Table 13: Goodness of fit of individual models as measured by MAE, RMSE and DS.

IN SAMPLE PREDICTIONS (FITTING)				
Panel A: Mean Absolute Error (MAE)				
	n=2	n=3	n=4	n=5
Restricted m. - whole dataset	1.030	1.901	2.621	3.240
Restricted m. - rolling since beginning	1.220	2.272	3.158	3.928
Restricted m. - rolling window of 5 yr	1.045	1.928	2.698	3.411
Unrestricted m. - whole dataset	1.033	1.909	2.618	3.221
Unrestricted m. - rolling since beginning	1.244	2.283	3.143	3.909
Unrestricted m. - rolling window of 5 yr	1.043	1.931	2.691	3.377
MA(12) - whole dataset	0.319	0.513	0.682	0.848
MA(12) - rolling since beginning	0.527	0.885	1.189	1.478
MA(12) - rolling window of 5 yr	0.460	0.823	1.124	1.426
Panel B: Root Mean Square Error (RMSE)				
	n=2	n=3	n=4	n=5
Restricted m. - whole dataset	1.304	2.346	3.209	3.962
Restricted m. - rolling since beginning	1.545	2.849	3.910	4.819
Restricted m. - rolling window of 5 yr	1.393	2.524	3.462	4.301
Unrestricted m. - whole dataset	1.292	2.342	3.208	3.953
Unrestricted m. - rolling since beginning	1.560	2.840	3.896	4.818
Unrestricted m. - rolling window of 5 yr	1.397	2.520	3.452	4.275
MA(12) - whole dataset	0.450	0.693	0.894	1.094
MA(12) - rolling since beginning	0.921	1.609	2.162	2.651
MA(12) - rolling window of 5 yr	0.886	1.576	2.128	2.626
Panel C: Direction symmetry (DS)				
	n=2	n=3	n=4	n=5
Restricted m. - whole dataset	62.9%	64.3%	64.7%	65.7%
Restricted m. - rolling since beginning	59.3%	60.2%	61.8%	60.8%
Restricted m. - rolling window of 5 yr	64.2%	64.4%	64.0%	64.0%
Unrestricted m. - whole dataset	64.5%	63.5%	64.7%	65.9%
Unrestricted m. - rolling since beginning	63.0%	62.6%	61.8%	60.4%
Unrestricted m. - rolling window of 5 yr	63.2%	63.6%	64.0%	64.8%
MA(12) - whole dataset	62.3%	66.1%	66.7%	66.5%
MA(12) - rolling since beginning	63.4%	63.6%	63.8%	65.4%
MA(12) - rolling window of 5 yr	65.0%	69.1%	69.3%	68.7%

Table 14: Predicting 1 year excess return for 1 month in advance: successfulness as measured by MAE, RMSE and DS.

OUT OF SAMPLE PREDICTIONS - 1 MONTH				
Mean Absolute Error				
	n=2	n=3	n=4	n=5
Restricted m. - rolling since beginning	1.246	2.321	3.223	4.009
Restricted m. - rolling window of 5 yr	1.170	2.179	3.051	3.848
Unrestricted m. - rolling since beginning	1.274	2.332	3.209	3.991
Unrestricted m. - rolling window of 5 yr	1.181	2.185	3.045	3.815
MA(12) - rolling since beginning	0.561	0.941	1.268	1.591
MA(12) - rolling window of 5 yr	0.560	0.989	1.360	1.717
Root Mean Square Error				
	n=2	n=3	n=4	n=5
Restricted m. - rolling since beginning	1.578	2.911	3.995	4.923
Restricted m. - rolling window of 5 yr	1.530	2.810	3.873	4.815
Unrestricted m. - rolling since beginning	1.600	2.903	3.979	4.920
Unrestricted m. - rolling window of 5 yr	1.551	2.811	3.862	4.786
MA(12) - rolling since beginning	0.946	1.648	2.213	2.729
MA(12) - rolling window of 5 yr	0.942	1.663	2.268	2.796
Direction symmetry				
	n=2	n=3	n=4	n=5
Restricted m. - rolling since beginning	58.2%	59.0%	60.1%	59.7%
Restricted m. - rolling window of 5 yr	63.1%	62.3%	62.5%	62.1%
Unrestricted m. - rolling since beginning	62.9%	61.7%	60.7%	58.8%
Unrestricted m. - rolling window of 5 yr	59.7%	61.9%	63.1%	63.7%
MA(12) - rolling since beginning	61.7%	60.5%	61.7%	62.7%
MA(12) - rolling window of 5 yr	57.4%	61.1%	60.5%	60.1%

Table 15: Predicting 1 year excess return for 6 month in advance: successfulness as measured by MAE, RMSE and DS.

OUT OF SAMPLE PREDICTIONS - 6 MONTH

Mean Absolute Error

	n=2	n=3	n=4	n=5
Restricted m. - rolling since beginning	1.809	5.715	10.574	15.729
Restricted m. - rolling window of 5 yr	1.732	5.570	10.679	16.187
Unrestricted m. - rolling since beginning	1.382	2.514	3.451	4.290
Unrestricted m. - rolling window of 5 yr	1.655	3.069	4.275	5.339
MA(12) - rolling since beginning	0.988	1.798	2.475	3.102
MA(12) - rolling window of 5 yr	1.112	2.020	2.802	3.526

Root Mean Square Error

	n=2	n=3	n=4	n=5
Restricted m. - rolling since beginning	2.375	7.094	12.872	18.951
Restricted m. - rolling window of 5 yr	2.278	6.953	12.992	19.481
Unrestricted m. - rolling since beginning	1.743	3.133	4.284	5.293
Unrestricted m. - rolling window of 5 yr	2.127	3.919	5.420	6.713
MA(12) - rolling since beginning	1.325	2.379	3.240	4.024
MA(12) - rolling window of 5 yr	1.467	2.620	3.592	4.495

Direction symmetry

Restricted m. - rolling since beginning	70.9%	67.3%	63.4%	61.4%
Restricted m. - rolling window of 5 yr	69.9%	67.1%	63.2%	61.6%
Unrestricted m. - rolling since beginning	61.8%	61.0%	59.8%	58.7%
Unrestricted m. - rolling window of 5 yr	59.8%	60.2%	60.8%	60.6%
MA(12) - rolling since beginning	55.7%	57.5%	58.5%	59.1%
MA(12) - rolling window of 5 yr	49.4%	53.7%	56.3%	57.5%

Table 16: Predicting 1 year excess return for 12 month in advance: successfulness as measured by MAE, RMSE and DS.

OUT OF SAMPLE PREDICTIONS - 12 MONTH

Mean Absolute Error

	n=2	n=3	n=4	n=5
Restricted m. - rolling since beginning	1.784	5.615	10.380	15.434
Restricted m. - rolling window of 5 yr	1.713	5.480	10.488	15.875
Unrestricted m. - rolling since beginning	1.445	2.631	3.592	4.453
Unrestricted m. - rolling window of 5 yr	2.033	3.768	5.218	6.498
MA(12) - rolling since beginning	1.221	2.282	3.150	3.923
MA(12) - rolling window of 5 yr	1.395	2.579	3.556	4.399

Root Mean Square Error

	n=2	n=3	n=4	n=5
Restricted m. - rolling since beginning	2.354	6.999	12.686	18.671
Restricted m. - rolling window of 5 yr	2.273	6.885	12.812	19.155
Unrestricted m. - rolling since beginning	1.800	3.234	4.419	5.457
Unrestricted m. - rolling window of 5 yr	2.585	4.802	6.641	8.200
MA(12) - rolling since beginning	1.586	2.909	3.985	4.932
MA(12) - rolling window of 5 yr	1.792	3.246	4.435	5.465

Direction symmetry

Restricted m. - rolling since beginning	69.9%	67.1%	63.2%	61.2%
Restricted m. - rolling window of 5 yr	70.3%	67.3%	63.2%	61.4%
Unrestricted m. - rolling since beginning	60.8%	60.4%	59.6%	59.3%
Unrestricted m. - rolling window of 5 yr	58.9%	58.1%	58.5%	56.7%
MA(12) - rolling since beginning	39.4%	39.4%	35.6%	38.6%
MA(12) - rolling window of 5 yr	41.3%	36.4%	37.6%	38.8%

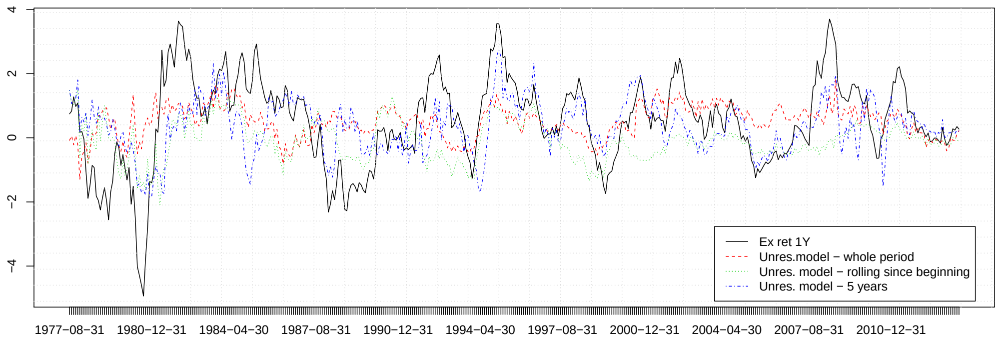


Figure 14: In-sample prediction of 1-year excess return ($n = 2$); comparing fits of one step CP model calculated on 5 year rolling window of data, rolling since beginning data and whole set of data.

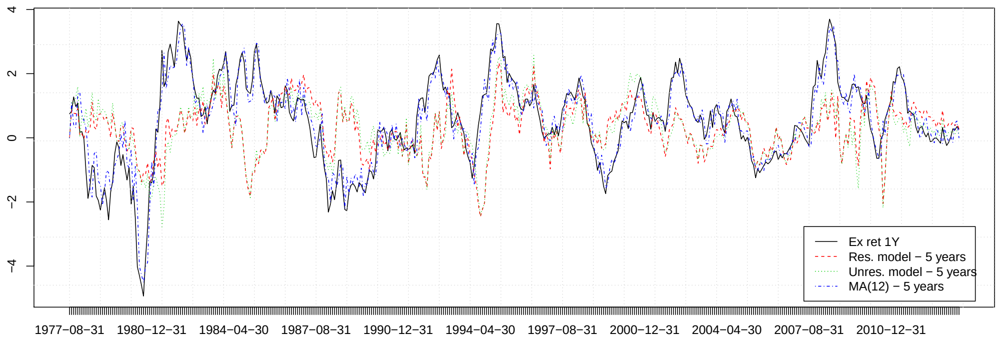


Figure 15: Out-of-sample prediction using rolling window of 5 years. Figure compares results obtained by one and two-step CP model and MA(12) model when fitting 1 year excess return ($n = 2$) for 1 month in advance.

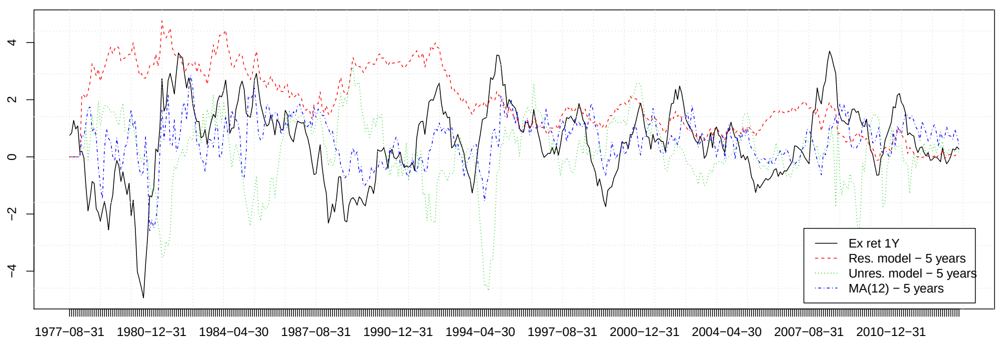


Figure 16: Out-of-sample prediction using rolling window of 5 years. Figure compares results obtained by one and two-step CP model and MA(12) model when fitting 1 year excess return ($n = 2$) for 6 months in advance.

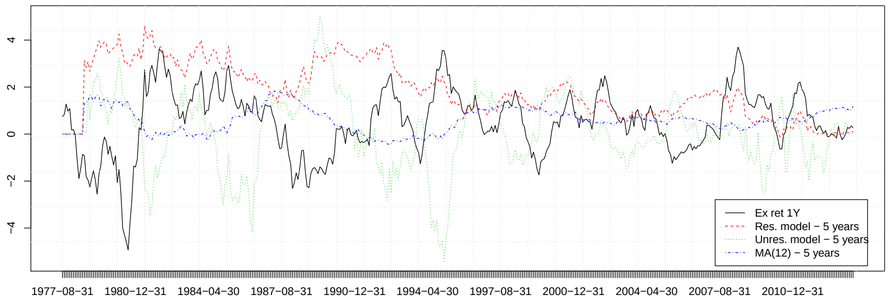


Figure 17: Out-of-sample prediction using rolling window of 5 years. Figure compares results obtained by one and two-step CP model and MA(12) model when fitting 1 year excess return ($n = 2$) for 12 months in advance.

3.4 Reasoning behind the single return factor

3.4.1 Principal component analysis of excess returns

In yield curve modelling it is quite common to use factor based models, i.e. to define small number of factors that drive movements in all yields. Factor models are based on principal component analysis, a statistical tool that converts a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables using eigenvalue decomposition.

Applied on yield curve movements this means that we first need to determine the covariance matrix of yield shifts and then determine the space that they span. This is determined by calculating eigenvectors and corresponding eigenvalues. Eigenvectors determine the main directions and the eigenvalues provide the relative weights or importance - the bigger are the corresponding eigenvalues of the eigenvectors, the greater is the loading that eigenvector has on the curve shift. Most commonly we do not consider all the factors spanning the space, but only those with greater loadings. For the yield curve that are usually first three factors, also referred to as yield, slope and curvature.

Despite the principal component analysis (PCA) is often done on yield curves, it is usually not done on expected excess returns. Cochrane and Piazzesi claim that the single forecasting factor is nothing else than the first component for the expected excess return. In the following pages this thesis will provide some high level information why this could be so and compare results obtained with PCA analysis with those of CP model.

The eigenvector decomposition of the covariance matrix of expected excess return (Cochrane and Piazzesi [7]), where W is orthogonal matrix (i.e. $\mathbf{W}'\mathbf{W} = \mathbf{W}\mathbf{W}' = \mathbf{I}$), is:

$$\text{cov}(E_t(\boldsymbol{rx_{t+1}}), (E_t(\boldsymbol{rx_{t+1}})') = \boldsymbol{\beta}\text{cov}(\boldsymbol{f_t}, \boldsymbol{f_t}')\boldsymbol{\beta}' = \mathbf{W}\boldsymbol{\Lambda}\mathbf{W}'. \tag{43}$$

If we define $\mathbf{\Gamma}' = \mathbf{W}'\boldsymbol{\beta}$ then the unconstrained regression can be rewritten as:

$$\begin{aligned} \boldsymbol{rx_{t+1}} &= \boldsymbol{\beta}\boldsymbol{f_t} + \boldsymbol{\epsilon} = \mathbf{W}\mathbf{W}'\boldsymbol{\beta}\boldsymbol{f_t} + \boldsymbol{\epsilon} \\ &= \mathbf{W}\mathbf{\Gamma}'\boldsymbol{f_t} + \boldsymbol{\epsilon} \end{aligned} \tag{44}$$

and

$$\mathbf{W}' \boldsymbol{rx_{t+1}} = \boldsymbol{\Gamma}' \boldsymbol{f_t} + \mathbf{W}' \boldsymbol{\epsilon}. \tag{45}$$

From equation (44) we see that the first column of **W** has the similar role as our coefficients in second step of the restricted model. It tells us how much the first expected return eigenvector (factor) load on expected return of bonds of different maturities. Similarly, is it easier to understand from the equation (45) the role of $\boldsymbol{\Lambda}'$. Its first row is actually an equivalent of our single factor.

Table 17 presents the principal component analysis for German yield curve data between 1972 and 1989 and compare them with results from Cochrane Piazzesi analysis. Panel A presents matrix of loadings **W** - factors of PCA, together with loadings on specific maturities as calculated by CP model and their ratio. Panel B provides values of matrix $\boldsymbol{\Gamma}$, which is an equivalent of single factor, and the ratio among them both.

The first factor of **W** captures almost all of the variation in expected excess returns and accounts for 99.5% of explained variance. Values-wise it is roughly equal to $-0.45\, b_n$, therefore it should also come as no surprise that when looking at panel B that the first column of $\boldsymbol{\Gamma}$ is roughly $2.2\,\boldsymbol{\gamma}$. When plotting them alongside on the same chart they display similar patterns of coefficients (Fig. 18).[12]

Similar conclusions can be drawn when performing principal component analysis also on German yield curve in following time periods, between 1989-2001 and between 2001-2014.

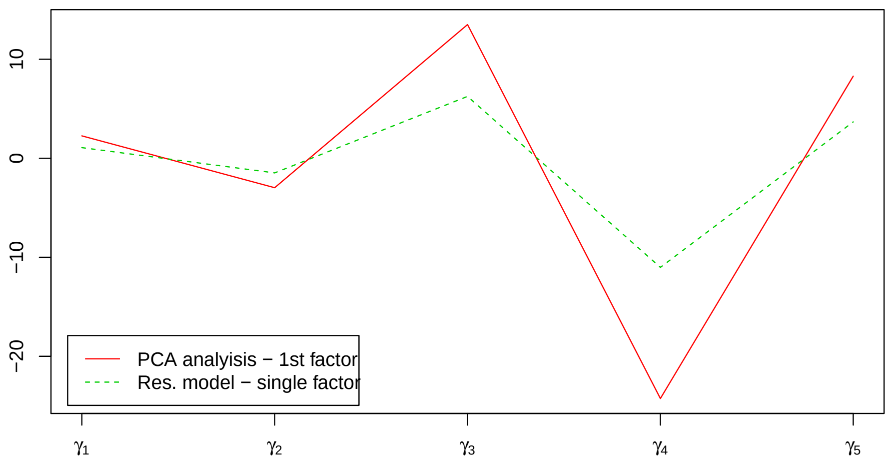


Figure 18: Single factor of CP model drawn alongside with first factor from PCA. They follow the same pattern, only with the difference in loadings. This is netted out when calculating expected excess returns of specific maturities. (German yield curve data, 1972-1989.)

[12]Note that the sign of first column of $\boldsymbol{\Gamma}$ is exactly opposite to the sign of single factor from CP model and therefore plotting original values would display two mirror patterns.

Table 17: Comparison of results from PCA with those of CP model.

PCA vs. CP results: German yield curve data, 1972-1989

A: Matrix of loadings W and its eigenvalues vs. CP regression Step 2

Maturity	**W columns (factors)**				**CP Reg.:**	**PCA\CP**
	1	2	3	4	$\mathbf{b_n}$	**Ratio***
n=2	**-0.209**	0.711	0.629	-0.237	0.475	-0.439
n=3	**-0.395**	0.501	-0.465	0.614	0.882	-0.448
n=4	**-0.556**	0.001	-0.450	-0.699	1.193	-0.466
n=5	**-0.701**	-0.494	0.431	0.279	1.449	-0.484
eigenvalues	**19.966**	0.080	0.006	2.86 E-04	-	-
explained var.	**99.5%**	0.39%	0.029%	0.01%	-	-

B: Single factor and its PCA equivalent Γ'

Variable	**Γ columns (forecasting portfolios)**				**CP Reg.:**	**PCA\CP**
	1	2	3	4	$\boldsymbol{\gamma}$	**Ratio***
y_t	**2.262**	-0.117	-0.040	0.010	-1.075	-2.105
f_t^2	**-2.973**	0.494	0.160	-0.040	1.478	-2.012
f_t^3	**13.500**	0.247	-0.352	0.133	-6.249	-2.160
f_t^4	**-24.264**	-1.513	0.130	-0.261	11.021	-2.202
f_t^5	**8.289**	0.994	0.087	0.162	-3.675	-2.255

*We calculate PCA\CP ratios by dividing first columns of $\mathbf{W}$ or $\boldsymbol{\Gamma}$ by $\mathbf{b_n}$ and $\boldsymbol{\gamma}$.

3.4.2 Role of yield factors in modelling excess returns

As forward rates are calculated from yields, one would assume that the main factors for yield curve (slope, level and curvature) would play a significant role also when estimating the movements of excess returns and would therefore also reveal the return forecasting factor. We will try to explore this possibility in the following paragraphs.

To understand the impact of principal components of yields on return-forecasting factor, we first have to calculate them and understand how much variance of yields do they explain. Once knowing these values, we should try to understand how much variance of excess return do they explain. All of these is presented in Table 18.

How much of variance of yields is explained by yield curve factors is reported in column *explained var.* However, these factors do not play exactly the same role by explaining variance of excess returns, which reader can see in the column *explained var. ER*. The most important factor appears to be slope (explaining around 72% of variance of return-forecast factor), whose loading increased on the account of decreased value of the level loading. Importance of level and curvature factor is more or less the same, where they explain 10% of variance each. However, now the role of fourth and fifth factor appears to be quite more important as before, as they explain a measurable fraction of variance.[13]

Similar results are valid also for periods from 1989-2001 and from 2001 until today. In the first of these periods the slope factor explain 83.1% of variance, the common fraction of variance explained by fourth and fifth factor is around 1.3% (around 1.0% on the account of fifth factor). From 2001 onwards slope explains 63.2%, the loading of level is by explaining 26.5% a bit higher relative to previous time periods. The fourth and fifth factor together now explain 10.3% of excess return variance (9.72 % on the account of fourth factor).

Table 18: Despite main three yield factors explain almost all variance of yield curves, the fourth and fifth factor play significant role when explaining variance of excess returns.

Yield factors and their impact; German yields 1972-1989					
Maturity	**Factors**				
	1 (level)	2 (slope)	3 (curvature)	4	5
n=1	-0.546	0.706	-0.439	0.101	0.016
n=2	-0.477	0.122	0.657	-0.544	-0.172
n=3	-0.429	-0.210	0.349	0.569	0.572
n=4	-0.394	-0.406	-0.106	0.358	-0.735
n=5	-0.368	-0.527	-0.492	-0.492	0.320
eigenvalues	1.49 E-03	3.00 E-05	2.90 E-06	1.33 E-07	3.16 E-09
explained var.	97.8%	2.0%	0.2%	0.0%	0.0%
explained var. ER	10.8%	71.8%	10.6%	2%	4.8%

3.4.3 Interpretations

Cochrane and Piazzesi's core result was that single return forecasting factor, which they got through regressions, is very much related to the dominant component of eigenvalue decomposition of the covariance matrix of expected excess returns. It used to be unnoticed, as majority of studies in the past focused mainly on level, slope and curvature components when fitting the yield curve and excess return. Although some of the variance of excess return is of course explained by these three factors, one should add also fourth and fifth factor[14] to achieve optimal results (similar to those using single return forecasting factor).[15]

[13] The result is in a way similar to those of Cochrane and Piazzesi [7]. They found that the most variance is explained by slope factor (58.7%), followed by fourth factor (24.3%).

[14] According to Cochrane and Piazzesi taking into account only the fourth factor should suffice, however we believe that for German yield curve also the fifth factor plays an important role.

[15] Their thesis is partially analysed by the paper from Xuyang [35], where she further explores the connection between CP factor and fourth component of variance decomposition of yield curves, which she refers to as *S-shape* factor. It namely displays how much of *S-shape* is the yield curve like.

In the research literature in the past ten years, there are several articles trying to interpret the economic meaning of the CP factor. Koijen, Lustig and Nieuweburgh [21] believe that the factor is proportional to transitory component of stochastic discount factor (SDF).[16]

3.5 Critiques of the CP model

Cochrane and Piazzesi's approach is most often criticised for providing spurious results as it is based on the regressions on bond yields data (including forward rates), which are highly cross- and autocorrelated. Additionally this issue may be emphasized further by using small sample data. Although up to now the usage of forward rates in prediction was discussed already in couple of articles (e.g. Dai, Singleton and Yang[9], Singleton[29]), the content of this subsection is based mainly on the analysis done by Thornton and Valente [33].

They re-wrote the equation (39) using only bond prices (overall result would be the same using yields):

$$p_{t+1}^{n-1} - p_t^n + p_t^1 = \beta_0 + \beta_1(-p_t^1) + \beta_2(p_t^1 - p_t^2) + \ldots + \beta_5(p_t^4 - p_t^5) + \epsilon_{t+1}^n. \quad (46)$$

As some of the variables appear on both sides of the equation this even reinforce the supposition of spurious regression issue.

In order to test their doubts Thornton and Valente ran several data generating processes, each of them using different assumptions about the time-series properties of bond prices.

Their initial assumption was that bond yields follow vector autoregressive process of order p - VAR(p) process, i.e. that they could be modelled as:

$$\mathbf{y_t} = \mathbf{y_0} + \sum_{i=1}^{p} \Gamma_i \mathbf{y_{t-i}} + \Phi \mathbf{e_t}.$$

In the above equation $\mathbf{y_t}$ is a vector of bond yields of different maturities, $\mathbf{y}_0$ presents initial state/intercept terms and $\mathbf{e_t}$ are assumed to be NIID(0,1) [17] residuals.

Based on different assumptions about matrices of parameters Γ and Φ they defined four cases:

Case 1: $\mathbf{y_t} = \mathbf{y_0} + \Phi \mathbf{e_t}$, with Φ diagonal. Bond yields are independent in mean and variance.

As the CP factor is much related with slope and *S-shape* factor she estimates excess returns using level, slope and curvature alone and together with *S-shape*. Similarly to Cochrane and Piazzesi her analysis confirm that the four factor is statistically significant predictor of excess return for USD curve.

[16]As nicely described by Smith and Wickens [30] according to SDF model the price of the asset in period t is the expected discounted value of the asset's pay-off in period $t + s$ based on the information available in period t: $P_t = E_t(M_{t+s}X_{t+s})$. In this equation P_t is the price of the asset in period t, X_{t+s} is the pay-off of the asset in period $t + s$, M_{t+s} is the discount factor for period $t + s(0 \leq M_{t+s} \leq 1)$. Thus P_t is the current value of the period $t + s$ income X_{t+s}. In general this income will be not known in period t and will be a random variable.

[17]NIID - independent and identically distributed with normal distribution.

Case 2: $\mathbf{y_t} = \mathbf{y_0} + \Gamma_p \mathbf{y_{t-1}} + \Phi \mathbf{e_t}$, with Γ_1 and Φ diagonal. Bond yields are independent in variance, but due to introduction of Γ_1 persistent in mean.

Case 3: $\mathbf{y_t} = \mathbf{y_0} + \Gamma_p \mathbf{y_{t-1}} + \Phi \mathbf{e_t}$, with Γ_1 and Φ full rank. Same as Case 2, with the difference that in this case the residuals are correlated as well.

Case 4: $\mathbf{y_t} = \mathbf{y_0} + \sum_{i=1}^{12} \Gamma_i \mathbf{y_{t-i}} + \Phi \mathbf{e_t}$, with Γ_i and Φ full rank. In this case the yields are dependent from up to 12 previous observations (persistent in mean) with correlated error terms.

Using Monte Carlo simulations[18], they found out that in case of independent yields (Case 1) the predictive power of regression (46) according to R^2 are quite large, decreasing from 0.829 (for excess return for $n = 2$) to 0.613 (for excess return for $n = 5$). This is quite intriguing in relation to the fact that in this case bond yields are *IID*. Supporting doubts expressed by Thornton and Valente, coefficients of determination increase for all maturities once we introduce serial correlation (Case 2). Once the cross-correlation is introduced (full rank of Γ_1 in Case 3) the levels drop significantly and are in line with those reported by Cochrane and Piazzesi. When adding further lags to the data (Case 4) the R^2 increased slightly in comparison to Case 3 results.

As Case 3 and Case 4 are much more in line with results provided by Cochrane and Piazzesi, one cannot strictly rule out the predictive power of forward rates in the simulations. However the results (especially from Case 2) suggest that some of the high R^2 in CP results may be indeed due to correlations and the fact that the same variables appear on the both sides of equation (Thornton and Valente [33]).

[18]Data used to parametrize the initial distributions used in the Monte Carlo simulations by Thornton and Valente correspond to those used by Cochrane and Piazzesi, i.e. CRSP data set.

4 Concluding remarks

In the article *Bond Risk Premia* by Cochrane and Piazzesi, authors explain a simple regression method to model the bond term premia and excess returns. They run two-step regression, which they claim it reveals the return forecasting factor. In their view this factor is related to main component in the principal component analysis of covariance matrix of excess returns. The return forecast factor was unnoticed for so long due to the focus of studies on the first three yield factors as opposed to either considering direct excess return factors or considering four or more yield factors.

While there is already plenty of literature providing results on US yield curve data, this thesis main objective was to analyse the Cochrane and Piazzesi's approach on non-US yield curve data. This allows one to test the robustness of the model in different circumstances.

We ran regressions on German yield curve data spanned from 1972 until present day. Running regressions on the whole time period provides low coefficient of determination and results in small predictive power of forward rates on the term premium of German yield curve.

However, as the whole period was quite turbulent for the German yield curve, we have decided to additionally run regression on shorter subsets. For them the coefficients of determination increase quite substantially. Corresponding F-statistics are higher than 95% critical value and reject the null hypothesis, stating that there is no dependency of excess return on forward rates. Consequentially this could be seen as an evidence against the expectations hypothesis, as according to it the excess returns should not be predictable at all.

Contrary to the original results, single forecasting return factor in our analysis is mostly not tent-shaped, but has a 'W-shape'. This reveals that the main results could in fact be driven by the correlations in the data. The shape appear to be more 'tent-like' once we decrease the number of the dependent variables. Nevertheless, the pattern still reveals the common factor behind regressions for different maturities. When doing the principal component analysis this factor was indeed close to the first eigenvector of the covariance matrix of the excess returns.

Overall my personal judgement would be, that in case of the German yield curve, we can be much less conclusive about the findings from the article. There is little evidence of ability of forward rates to predict future excess returns for a whole dataset, higher R^2 and higher F-statistics are observed only when running regressions on smaller subsets. To determine whether in those cases the results in fact reveal the true underlying relationship or are only driven by measurement error, could be the further extension of the currently done work. Additionally we could extend the model by including some macroeconomic factors (e.g. inflation).

References

[1] A. Ang and M. Piazzesi, *A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables*, Journal of Monetary economics **50(4)** (2003) 745-787; available at `http://web.stanford.edu/~piazzesi/AP.pdf`.

[2] J. Annaert, A.G.P. Claes, M.J.K. De Ceuster, H.Zhang, *The Estimation of Svensson Model Term Structures and Their Volatilities*, [seen 10.02.2015], available at `http://ssrn.com/abstract=2054693`.

[3] D. Asteriou and S. G. Hall, *Applied econometrics*, 2nd ed., Palgrave Macmillan, Basingstoke, 2011.

[4] L. von Auer, *Oekonometrie: Eine Einfuehrung*, 3. Aufl., Springer, Berlin, 2005.

[5] J. Baffes, *Explaining stationary variables with non-stationary regressors*, Applied Economics Letters **4(1)** (1997) 69-75.

[6] M.Choudry, *Bond and money markets: strategy, trading, analysis*, Butterworth-Heinemann, Oxford, 2003.

[7] J. H. Cochrane and M. Piazzesi, *Appendix to Bond Risk Premia*, [seen 10.01.2015], available at `http://web.stanford.edu/~piazzesi/cpapp.pdf`.

[8] J. H. Cochrane and M. Piazzesi, *Bond Risk Premia*, American Economic Review **95** (2005) 138–161; available at `http://web.stanford.edu/~piazzesi/cp.pdf`.

[9] Q. Dai, K. Singleton and W. Yang, *Predictability of bond risk premia and affine term structure models*, [seen 10.01.2015], available at `http://faculty.chicagobooth.edu/john.cochrane/research/Papers/cpn.pdf`.

[10] P. Donati and F. Donati, *Modelling and Forecasting the Yield Curve Under Model Uncertainty*, European Central Bank Working Paper Series **917** (2008); available at `http://www.ecb.europa.eu/pub/pdf/scpwps/ecbwp917.pdf`.

[11] M. Ehrmann, M. Fratzscher, R.S. Gürkynak and E.T. Swanson, *Convergence and anchoring of yield curves in the euro area*, European Central Bank Working Paper Series **817** (2007); available at `http://www.ecb.europa.eu/pub/pdf/scpwps/ecbwp817.pdf`.

[12] F.J. Fabozzi, *The handbook of fixed income securities*, 6th ed., McGraw-Hill, New York, 2000.

[13] L. Fahrmeir, A. Hamerle and G. Tutz, *Multivariate statistische verfahren*, 2. Aufl., Walter de Gruyter, Berlin, 1996.

[14] J. Fox and S. Weisberg, *An R Companion to Applied Regression*, 2nd ed., Thousand Oaks, Sage, 2011.

[15] W.H. Greene, *Econometric analysis*, 7th ed., Prentice Hall, New Jersay, 2011.

[16] P. Grosjean and F. Ibanez, *pastecs: Package for Analysis of Space-Time Ecological Series*, R package version **1.3-18** (2014); available at `http://CRAN.R-project.org/package=pastecs`.

[17] R. S. Gürkaynak, B.P. Sack and J.H. Wright, *The U.S. Treasury Yield Curve: 1961 to the Present*, Journal of Monetary Economics **54(8)** (2007) 2291-2304; available at `http://www.federalreserve.gov/pubs/feds/2006/200628/200628pap.pdf`.

[18] J.D. Hamilton, *Time series analysis*, Vol. 2, Princeton University Press, Princenton, 1994.

[19] P. Kennedy, *A guide to econometrics*, 5th ed., MIT Press, Cambridge, 2003.

[20] D.H. Kim and A. Orphanides, *The bond market term premium: what is it, and how can we measure it?*, BIS Quarterly Review (2007) 27-40; available at `http://www.bis.org/publ/qtrpdf/r_qt0706e.pdf`.

[21] R. S. Koijen, H. Lustig and S. Van Nieuwerburgh, *The cross-section and time-series of stock and bond returns*, NBER Working Paper **No. 15688** (2012); available at `http://www.nber.org/papers/w15688.pdf`.

[22] T.C. Mills and R.N. Markellos, *The econometric modelling of financial time series*, Cambridge University Press, Cambridge, 2008.

[23] P. Newbold, *Statistics for Business and Economics*, 3rd ed., Prentice Hall, New Jersay, 1991.

[24] B. Pfaff, *Analysis of Integrated and Cointegrated Time Series with R*, 2nd ed., Springer, New York, 2008.

[25] R Core Team, *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, 2015; available at `http://www.R-project.org/`.

[26] S.A. Ross, R.W. Westerfield and B.D. Jordan *Fundamentals of Corporate Finance*, 10th ed., McGraw Hill, New York, 2013.

[27] G. D. Rudebusch, B.P. Sack and E. T. Swanson, *Macroeconomic implications of changes in the term premium*, Federal Reserve Bank of St. Louis Review **89** (2007) 241-269; available at `https://research.stlouisfed.org/publications/review/07/07/Rudebusch.pdf`.

[28] J.A. Ryan and J.M. Ulrich, *xts: eXtensible Time Series*, R package version **0.9-7** (2014); available at `http://CRAN.R-project.org/package=xts`.

[29] K. J. Singleton, *Empirical dynamic asset pricing: model specification and econometric assessment*, Princeton University Press, Princeton, 2009.

[30] P. Smith and M. Wickens, *Asset pricing with observable stochastic discount factors*, Journal of Economic Surveys **16(3)** (2002) 397-446; available at `http://web.cenet.org.cn/upfile/76975.pdf`.

[31] L.E.O. Svensson, *Estimating and interpreting forward interest rates: Sweden 1992-1994*, NBER Working Paper **No. 4871** (1994); available at http://www.nber.org/papers/w4871.pdf.

[32] E. Swanson, *What we do and don't know about the term premium*, FRBSF Economic Letter **2007-21** (2007); available at http://www.frbsf.org/economic-research/publications/economic-letter/2007/july/term-premium/el2007-21.pdf.

[33] D.L Thornton and G.Valente, *Revisiting the Predictability of Bond Risk Premia*, Research Division Federal Reserve Bank of St. Louis Working Paper Series **2009-009** (2009); available at http://research.stlouisfed.org/wp/2009/2009-009.pdf.

[34] A. Trapletti and K. Hornik, *tseries: Time Series Analysis and Computational Finance*, R package version **0.10-34** (2015); available at http://CRAN.R-project.org/package=tseries.

[35] M.Xuyang, *The S shape factor and bond risk premia*, [seen 12.01.2015], available at https://www.k-state.edu/economics/seminars/XuyangMaPaper.pdf.

[36] A. Zeileis, *Econometric computing with HC and HAC covariance matrix estimators*, Journal of Statistical Software **11(10)** (2004) 1–17; available at http://www.jstatsoft.org/v11/i10/paper.

[37] A. Zeileis, T. Hothorn *Diagnostic Checking in Regression Relationships*, R News **2(3)** (2002) 7-10. http://CRAN.R-project.org/doc/Rnews/.

[38] A. Zeileis, F. Leisch, K. Hornik and C. Kleiber, *strucchange: An R Package for Testing for Structural Change in Linear Regression Models*, Journal of Statistical Software **7(2)** (2002) 1-38; available at http://www.jstatsoft.org/v07/i02/.

Razširjeni povzetek v slovenskem jeziku

O terminski premiji

V začetku tega tisočletja, bolj natančno v obdobju pred krizo, je terminska premija v ZDA spodbudila zanimanje mnogih finančnih ekonomistov. Na trgu so bila takrat namreč prisotna visoka nihanja donos obveznice z 10-letnim dospetjem (6.8% na začetku tisočletja, 2.3% junija 2003, do nad 5% v letu 2007), poleg tega se smer nihanja ni nujno ujemala s spremembo kratkoročne obrestne mere. Ker dolgoročna obrestna mera vpliva na vrednost dolga in tudi zadolževanja, se je povečala potreba po razumevanju obnašanju le-te. Čeprav na nihanje dolgoročnih obrestnih mer vpliva mnogo dejavnikov, terminska premija predstavlja pomembno komponento tega gibanja.

Obstaja več definicij terminske premije, intuitivno pa terminska premija predstavlja dodaten donos vlagatelja ob nakupu obveznice z daljšim dospetjem v primerjavi z večkratnim zaporednim nakupom obveznic s krajšim dospetjem. Kim in Orphanides [20] sta definirala tri sorodne tipe terminske premije:

- Terminska premija kot razlika med pričakovanim donosom, ki ga ima kupec brezkuponske obveznice z ročnostjo N-let v obdobju prvega leta in trenutno enoletno obrestno mero:

$$\phi_t(N) = E_t(R_{t+1,N}) - rs_{t,1}, \tag{47}$$

  kjer $R_{t+1,N} = \log \frac{P_{t+1}(n-1)}{P_t(n)}$.

- Terminska premija kot razlika med terminsko obrestno mero med obdobjem od $N-1$ do N in pričakovano enoletno obrestno mero v času $N-1$:

$$\phi_t(N) = rf_{t,N-1,N} - E_t(rs_{t,N-1}). \tag{48}$$

- Terminska premija kot razlika med donosom N-letne brezkuponske obveznice in povprečjem pričakovanih enoletnih obrestnih mer v obdobju N-let:

$$\phi_t(N) = rs_{t,N} - \frac{1}{N}\sum_{n=1}^{N} E_t(rs_{t,n}). \tag{49}$$

Notacije uporabljene v zgornjih enačbah so sledeče: $rs_{t,n}$ je trenutna obrestna mera v času t (ang. *spot interest rate*) za obdobje med t in $t+n$, $P_t(n)$ je cena brezkuponske obveznice z dospelostjo n v času t, $E_t(.)$ pa označuje pričakovano vrednost (matematično upanje) spremenljivke v času t.

Vse zgoraj opisane definicije predstavljajo odstopanje od hipoteze pričakovanj (ang. *expectation hypothesis*), ki trdi da je trenutna dolgoročna obrestna mera vsota trenutne in pričakovane obrestne mere. Skladno s to teorijo so potemtakem vse zgoraj opisane terminske premije enake nič. [19]

[19] Obstajata dve različici hipoteze pričakovanj, čista hipoteza pričakovanj in (klasična) hipoteza pričakovanj. Prva trdi, da je terminska premija enaka nič, druga pa, da je konstantna v času.

Načinov merjenja terminske premije je več (npr. DGSE modeli, VAR modeli, 'novo-kenezijanski' modeli), v tem delu pa sem sledila pristopu Johna Cochrana in Monike Piazzesi, ki temelji na klasični linearni regresiji po metodi najmanjših kvadratov. Pred podrobno predstavitvijo njunega modela in prikazom rezultatov implementacije le-tega na donose nemških obveznic, pa bom opisala osnove lastnosti linearne regresije.

O linearni regresiji

Model linearne regresije lahko zapišemo kot

$$y_n = \beta_1 + \beta_2 x_{n2} + \ldots + \beta_K x_{nK} + \varepsilon_n \tag{50}$$

oz. v matrični prezentaciji kot

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}. \tag{51}$$

V obeh zgornjih enačbah predstavlja $\mathbf{y}$ opazovano odvisno spremenljivko, spremenljivke $X_1, X_2, \ldots, X_K$ pojasnjevalne spremenljivke, $\boldsymbol{\varepsilon}$ pa slučajen šum. Koeficienti $\beta_i \in \mathbb{R}$ merijo pričakovano spremembo opazovane spremembe, ko se vrednost pojasnjevalne spremenljivke spremeni za eno enoto.

Brez omejitev ima linearni model neskončno rešitev, saj vedno lahko najdemo take vrednosti šuma $\boldsymbol{\varepsilon}$, za katere enačbi (50) in (51) držita. V klasičnem modelu linearne regresije so omejitve naslednje:

- Odvisno spremenljivko je mogoče izraziti kot vsoto linearne funkcije neodvisnih spremenljivk in šuma;
- Neodvisne spremenljivke so deterministične (eksogene);
- Matematično upanje šuma je enako nič, $E(\boldsymbol{\varepsilon}) = 0$;
- Motnje (šum) imajo konstantno varianco in so med seboj neodvisne, $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \Sigma = \sigma^2 I$;
- Matrika $\mathbf{X}$ ima poln rang: na voljo imamo vsaj toliko opazovanj kot je neodvisnih spremenljivk. Neodvisne spremenljivke so linearno neodvisne.

Cilj regresije je pod danimi omejitvami poiskati ocene regresijskih koeficientov $\boldsymbol{\beta}$. Pri tem obstaja več različnih načinov, najboljša linearna nepristranska cenilka pa je pod danimi pogoji cenilka po metodi najmanjših kvadratov $\mathbf{b}$, v nadaljevanju tudi OLS (ang. *Ordinary least squares*) cenilka:

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}.$$

Izrek 1. *(Gauss, Markov)*
Naj velja (51) skupaj s standardnimi predpostavkami:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta} \text{ in } \operatorname{cov}(\mathbf{y}) = \sigma^2\mathbf{I}.$$

Tedaj za $\mathbf{b}$*, cenilko po metodi najmanjših kvadratov velja:*

a) OLS cenilka za $\boldsymbol{\beta}$ *je nepristranska,* $E(\mathbf{b}) = \boldsymbol{\beta}$*;*

b) Velja $\mathrm{var}(\mathbf{b}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$*;*

c) Cenilka $\mathbf{b}$ *je najboljša linearna nepristranska cenilka (ang. best linear unbiased estimator or BLUE), tj. ima med nepristranski linearnimi cenilkami najmanjšo varianco;*

d) OLS cenilka variance σ^2 *je nepristranska,* $E(\tilde{\sigma}^2) = \sigma^2$.

Kadar uporabljamo linearno regresijo na časovnih vrstah, ponavadi kršimo katero od standardnih predpostavk, najpogosteje je to predpostavka o neodvisnosti šuma skozi čas in njegovi konstanti varianci. Najpogosteje so napake (šumi) namreč bodisi heteroskedastične, bodisi avtokorelirane (bodisi oboje):

Varianca $\mathrm{var}(\boldsymbol{\varepsilon}\,|\mathbf{X})$ - heteroskedastične napake:

$$\sigma^2\mathbf{\Omega} = \sigma^2 \begin{pmatrix} \omega_1 & 0 & \dots & 0 \\ 0 & \omega_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \omega_N \end{pmatrix}. \tag{52}$$

Varianca $\mathrm{var}(\boldsymbol{\varepsilon}\,|\mathbf{X})$ - avtokorelirane (homoskedastične) napake:

$$\sigma^2\mathbf{\Omega} = \sigma^2 \begin{pmatrix} 1 & \rho_1 & \dots & \rho_{N-1} \\ \rho_1 & 1 & \dots & \rho_{N-2} \\ \vdots & \vdots & \dots & \vdots \\ \rho_{N-1} & \rho_{N-2} & \dots & 1 \end{pmatrix}. \tag{53}$$

Čeprav avtokoreliranost in heteroskedastičnost ne vplivata na nepristranskost cenilke, pa spremenita njeno kovariančno matriko. Posledično OLS cenilka ni več najbolj učinkovita. Problem avtokoreliranosti in heteroskedastičnosti lahko rešujemo na več načinov - lahko na novo definiramo model z drugimi pojasnjevalnimi spremenljivkami, lahko uporabimo drugo cenilko, ali pa uporabimo OLS cenilko ter pri tem upoštevamo spremenjeno kovariančno matriko.

Kadar spremenjeno kovariančno matriko Ω poznamo, za OLS cenilko velja:

$$\hat{\mathbf{b}} \;=\; (\mathbf{X}^{*\prime}\mathbf{X}^{*})^{-1}\mathbf{X}^{*\prime}\mathbf{y} = (\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{y}. \tag{54}$$

V primeru, da je Ω nepoznana, pa jo ocenimo s tako-imenovanimi HC in HAC matrikami (matrike skladne s heteroskedastičnostjo oz. matrike skladne s heteroskedastičnostjo in avtokorelacijo). Med širše uporabljeni cenilkami spremenjene kovariančne matrike so Newey-West cenilka, Hansen-Hodrick cenilka in Andersow razred cenilk.

O Cochrane-Piazzesiini meri

Cochrane-Piazzesiina mera je prvič predstavljena v članku *Bond Risk Premia*, katerega sta John H. Cochrane in Monika Piazzesi leta 2005 objavila v American Economic Review [8]. V članku sta avtorja vrednosti terminske premije napovedovala z regresijami na terminskih obrestnih merah, ki bodo podrobneje opisane v nadaljevanju.

Notacije

$P_t^{(n)}$	Cena brezkuponske obveznice z ročnostjo n-let v času t z izplačilom 1 enote ob dospetju.
$p_t^{(n)} = \log P_t^n$	Logaritem cene n-letne brezkuponske obveznice v času t.
$y_t^{(n)} \equiv -\frac{1}{n} p_t^{(n)}$	Logaritem donosnosti n-letne brezkuponske obveznice v času t.
$f_t^{(n)} \equiv p_t^{(n-1)} - p_t^{(n)}$	Terminska obrestna mera v času t za posojila med $t+n-1$ in $t+n$.
$r_{t+1}^{(n)} \equiv p_{t+1}^{(n-1)} - p_t^{(n)}$	Logaritem donosnosti, ob nakupu n-letne obveznice v času t in njeni prodaji v $t+1$.
$rx_{t+1}^{(n)} \equiv r_{t+1}^{(n)} - y_t^{(1)}$	Presežni donos enoletnega držanja n-letne obveznice nad trenutno enoletno obrestno mero.

Regresijske enačbe

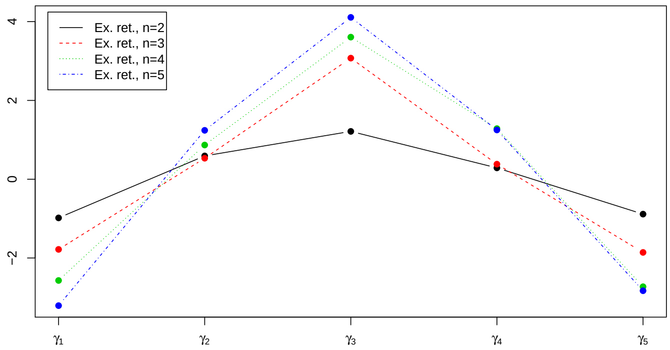
Cochrane in Piazzesi sta presežne donose obveznic poskusila aproksimirati z uporabo terminskih obrestnih mer. Kot odvisno spremenljivko sta uporabila presežne donose (ob nakupu/prodaji 2-, 3-, 4- in 5- letne obveznice) v času $t+1$, kot pojasnjevalne pa trenutno obrestno mero in terminske obrestne mere za obdobja 2, 3, 4 in 5 let v času t:

$$\begin{aligned} rx_{t+1}^{(n)} &= \beta_0^{(n)} + \beta_1^{(n)} y_t^{(1)} + \beta_2^{(n)} f_t^{(2)} + \beta_3^{(n)} f_t^{(3)} \\ &+ \beta_4^{(n)} f_t^{(4)} + \beta_5^{(n)} f_t^{(5)} + \epsilon_{t+1}^{(n)}, \qquad n = 2,3,4,5. \end{aligned} \tag{55}$$

Ob opazovanju regresijskih koeficientov za različna dospetja sta opazila ponavljajoči vzorec v obliki 'šotora' (Slika 1). Posledično sta predvidevala, da presežne donose obveznic različnih ročnosti lahko izrazimo kot večkratnik enoznačne linearne kombinacije terminskih mer, katero sta poimenovala *enoten (skupni) faktor*.

$$\begin{aligned} rx_{t+1}^{(n)} &= b_n(\gamma_0^{(n)} + \gamma_1^{(n)} y_t^{(1)} + \gamma_2^{(n)} f_t^{(2)} + \gamma_3^{(n)} f_t^{(3)} \\ &+ \gamma_4^{(n)} f_t^{(4)} + \gamma_5^{(n)} f_t^{(5)}) + \epsilon_{t+1}^{(n)}, \qquad n = 2,3,4,5. \end{aligned} \tag{56}$$

Da bi se izognila neskončnemu številu rešitev, sta problem (40) dodatno omejila s pogojem, da je povprečna vrednost b_n enaka ena. Svoj postopek sta tako razdelila na dva koraka.



Slika 1: Regresijski koeficienti pri modeliranju posameznih presežnih donosov na CRSP podatkih (prilagojeno po Cochrane, Piazzesi [8]).

Korak 1
V prvem koraku identificiramo skupen faktor:

$$\begin{aligned}\overline{rx}_{t+1} = \frac{1}{4}\sum_{n=2}^{n=5} rx_{t+1}^{(n)} &= \gamma_0^{(n)} + \gamma_1^{(n)} y_t^{(1)} + \gamma_2^{(n)} f_t^{(2)} + \gamma_3^{(n)} f_t^{(3)} + \\ &+ \gamma_4^{(n)} f_t^{(4)} + \gamma_5^{(n)} f_t^{(5)} + \epsilon_{t+1}^{(n)} \\ &= \boldsymbol{\gamma}' \boldsymbol{f_t} + \bar{\epsilon}_{t+1}. \end{aligned} \tag{57}$$

Korak 2
V drugem koraku ocenimo uteži b_n:

$$rx_{t+1}^{(n)} = b_n(\boldsymbol{\gamma}' \boldsymbol{f_t}) + \bar{\epsilon}_{t+1}, \qquad n = 2, 3, 4, 5. \tag{58}$$

Terminska premija nemških državnih obveznic

Primarni cilj magistrske naloge je vrednotenje Cochrane-Piazzesiine mere za obveznice evrskega območja, s čimer bi dodatno tudi testirali obnašanje modela ob uporabi različnih podatkov. Analiza je bila izvedena za nemške državne obveznice, saj je nemški trg vrednostnih papirjev eden izmed redkih trgov evrskega območja, kateri je dovolj likviden ter hkrati ponuja zadostno količino preteklih podatkov.[20]

Podatki

Regresije so bile izračunane na donosih nemških obveznic z dospetjem od 1 do 10 let ter 12, 20 in 30 let. Podatki so mesečni in obsegajo obdobje med septembrom 1972 in septembrom 2014, njihove opisne statistike pa so podane v Tabelah 1 in 2.

[20] Magistrska naloga sicer obsega analize na trgu nemških državnih obveznic na podlagi dveh različnih tipov podatkov, poleg tega pa vsebuje še primerjavo z rezultati na ameriškem trgu vrednostnih papirjev.

Tabela 1: Obrestne mere nemških obveznic (1972-2014) - opisne statistike.

TRENUTNA IN TERMINSKE OBRESTNE MERE (493 opazovanj)					
	yield 1 yr	fwd 1-2 yr	fwd 2-3 yr	fwd 3-4 yr	fwd 4-5 yr
mat. upanje	4.79	5.25	5.68	5.99	6.20
std. odklon	2.66	2.51	2.41	2.30	2.18
koef. simetrije	0.28	-0.09	-0.26	-0.29	-0.28
sploščenost	-0.39	-0.68	-0.70	-0.67	-0.64
norm. test	0.98	0.98	0.98	0.98	0.98
test p-vrednost	5.72 E-07	2.66 E-05	4.69 E-07	3.98 E-07	7.83 E-07
avtokor. (l=1)	0.99	0.98	0.99	0.99	0.99
avtokor. (l=12)	0.78	0.80	0.82	0.83	0.83
avtokor. (l=18)	0.64	0.70	0.74	0.74	0.74
korelacije	1.00	0.97	0.93	0.90	0.87
		1.00	0.99	0.96	0.94
			1.00	0.99	0.98
				1.00	0.99
					1.00

Tabela 2: Presežni donosi nemških obveznic med 1972 in 2014 - opisne statistike.

REALIZIRANI PRESEŽNI DONOSI (493 opazovanj)					
	Ex ret 1 yr	Ex ret 2 yr	Ex ret 3 yr	Ex ret 4 yr	Mean ex ret
mat. upanje	0.64	1.27	1.77	2.18	1.47
std. odklon	1.47	2.69	3.73	4.65	3.11
koef. simetrije	-0.35	-0.35	-0.42	0.48	-0.43
sploščenost	0.64	0.18	0.10	0.11	0.17
norm. test	0.99	0.99	0.99	0.98	0.99
test p-vrednost	3.30 E-03	1.62 E-03	1.07 E-04	1.48 E-05	1.16 E-04
avtokor. (l=1)	0.93	0.94	0.94	0.95	0.95
avtokor. (l=12)	0.18	0.13	0.10	0.08	0.11
avtokor. (l=18)	0.03	0.04	0.03	0.01	0.03
korelacije	1.00	0.98	0.95	0.92	1.00
		1.00	0.99	0.97	
			1.00	0.99	
				1.00	

V analizi so bili uporabljeni interpolirani podatki, saj se trenutna dospetja nemških obveznic med seboj razlikujejo, poleg tega pa ne obstaja 'točna' definicija nemške n-letne obveznice. Hkrati z interpolacijo omejimo tudi vpliv različnih kuponov obveznic ter njihovih ostalih posebnosti (M. Ehrmann, M. Fratzscher, R.S. Gürkynak in E.T. Swanson[11]). Podatki so interpolirani z uporabo Svenssonove metode, katera je razširitev klasičnega Nelson-Siegel modela za modeliranje časovne strukture obrestnih mer (v nadaljevanju je namesto časovna struktura obrestnih mer uporabljen tudi termin krivulja donosnosti). [21]

Rezultati regresij

Regresije so bile najprej ločeno izračunane za presežne donose različnih ročnosti (ob nakupu/prodaji 2-, 3-, 4- in 5- letne obveznice). V tem povzetku so predstavljeni le rezultati za preprodajo 2-letne obveznice (Tabela 3).[22] V tabeli so podani koeficienti posameznih regresij ter standardne napake izračunane z uporabo 'klasične' in Newey-West kovariančne matrike, katera omeji učinke avtokorelacije. Poleg tega tabela vsebuje t- in F-statistike, katere merijo statistično značilnost posameznega faktorja oz. skupno statistično značilnost vseh faktorjev.[23]

Vse omenjene regresije poročajo nizke koeficiente determinacije R^2 ter razmeroma nizke t-statistike.[24] Če ne upoštevamo avtokoreliranosti podatkov, imajo nasprotno vse regresije visoke F-statistike, kar nakazuje na skupno zmožnost napovedi. Vrednosti F-statistik pa se ob uporabi Newey-West kovariančne matrike občutno znižajo, s čimer lahko zaključimo, da je povezava med vrednostmi donosov in vrednostmi terminskih obrestnih mer šibka.

Čeprav rezultatih posameznih regresij takoj ne potrdijo Cochrane-Piazzesiinega modela, koeficienti posameznih regresij vseeno sledijo skupnemu vzorcu (Slika 2). Le-ta nima oblike 'šotora', ampak obliko črke 'W', kar implicira multikolinearnost neodvisnih spremenljivk. Multikolinearnost je potrjena z regresijo na manj neodvisnih spremenljivkah, saj rezultati regresij in R^2 ostanejo praktično nespremenjeni. Dodatno imajo regresije tudi 'ne-normalno' porazdelitev ostankov, njihova verjetnostna porazdelitvena funkcija ima širše repe. Čeprav je to načeloma nezaželeno pri uporabi regresij, v našem primeru to ni nujno problem. Naš osnovni motiv je namreč le napoved presežnih donosov, kjer smo minimizirali vsoto kvadratov komponent residuala.

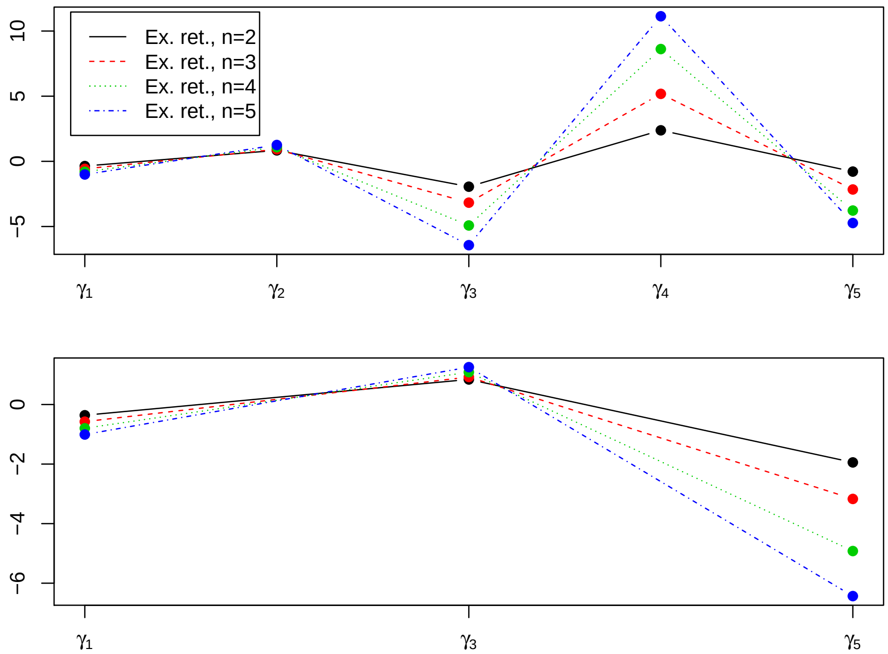
Ker rezultati posameznih regresij tako zelo odstopajo od pričakovanih, so bile regresije dodatno izračunane na manjših pod-vzorcih, kateri so bili določeni z iskanjem strukturnih prelomov. Koeficienti pridobljeni s temi regresijami so predstavljeni na Sliki 3. Opazimo lahko štiri prevladujoče vzorce, ki se pojavijo v obdobju pred padcem Berlinskega zidu, obdobju združitve Vzhodne in Zahodne Nemčije in uporabo nemške marke DEM, obdobju vpeljave evra pred krizo leta 2007 in obdobju po njej. Pregled rezultatov po posameznih obdobjih je podan v Tabeli 4.

[21] Podatki so skupaj s parametri interpolacije objavljeni na spletni strani nemške centralne banke; `http://www.bundesbank.de/Navigation/EN/Statistics/statistics.html`.

[22] Preostali rezultati so podani v Tabeli 7 glavnega dela magistrske naloge.

[23] Definicija testnih statistik je podana v 2.3.2.

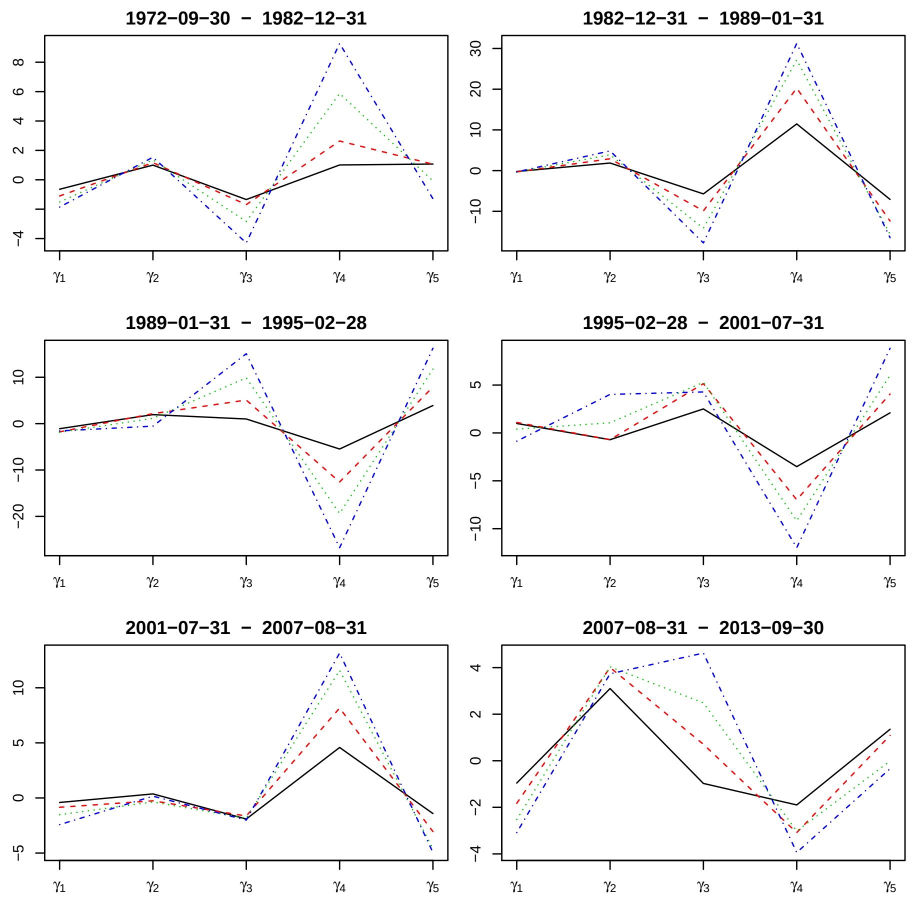
[24] Koeficient deteriminacije R^2 je osnovni element, s katerim ocenjujemo, kdaj se naš model dobro prilega podatkom, njegova definicija je podana v 2.3.5.



Slika 2: Slika prikazuje ponavljajoč vzorec, katerega tvorijo koeficienti posameznih regresij. Vzorec se pojavi neodvisno od števila neodvisnih spremenljivk. Vrednosti izračunane za nemške obveznice (1972-2014).

Tabela 3: Koeficienti regresije in njene statistike pri modeliranju presežnega donosa ob preprodaji 2-letne obveznice. Vrednosti izračunane za nemške obveznice (1972-2014).

REGRESIJA BREZ OMEJITEV								
Realiziran presežni donos, n=2								
	Coef	SE	t	P(>|t|)		$SE^{(NW)}$	$t^{(NW)}$	$P(>\|t\|)^{(NW)}$
(Intc.)	-0.392	0.274	-1.434	0.152		0.734	-0.534	0.593
yield	-0.360	0.108	-3.334	0.001	***	0.313	-1.148	0.251
f1	0.842	0.390	2.159	0.031	*	0.80	1.053	0.293
f2	-1.944	1.174	-1.656	0.098	.	2.334	-0.833	0.405
f3	2.382	1.776	1.342	0.180		3.471	0.686	0.493
f4	-0.785	0.946	-0.830	0.407		1.931	-0.407	0.684
Adj R^2: 0.064, F-stat. (5 in 487 DF): 7.734 in 1.400 $^{(NW)}$								



Slika 3: Vzorec koeficientov regresij skozi čas.

Tabela 4: Koeficient determinacije in F-statistike za posamezne regresije v različnih obdobjih.

REGRESIJA BREZ OMEJITEV - REZULTATI SKOZI ČAS									
	R^2				F-stat.$^{(NW)}$				Obs.
Obdobje	n=2	n=3	n=4	n=5	n=2	n=3	n=4	n=5	
1972-1982	0.352	0.430	0.456	0.460	10.103	14.753	16.214	16.112	124
1982-1989	0.456	0.391	0.343	0.305	0.981	0.840	1.422	2.781	74
1989-1995	0.096	0.101	0.128	0.160	0.187	0.358	0.644	1.090	74
1995-2001	0.712	0.731	0.717	0.691	62.242	43.582	25.523	16.026	78
2001-2007	0.860	0.853	0.848	0.841	140.85	157.62	172.130	180.380	74
2007-2014	0.696	0.618	0.530	0.445	26.205	14.522	8.286	5.037	74
1972-2014	0.064	0.067	0.073	0.080	1.400	1.014	0.660	0.507	487

Glede na rezultate posameznih regresij je bila regresija z omejitvami izvedena na pod-vzorcih podatkov. V sledečih odlomkih bodo predstavljeni rezultati za obdobje med leti 1972 in 1989, ki vsebuje podatke veljavne za Zahodno Nemčijo (ko so bile v uporabi še DEM).

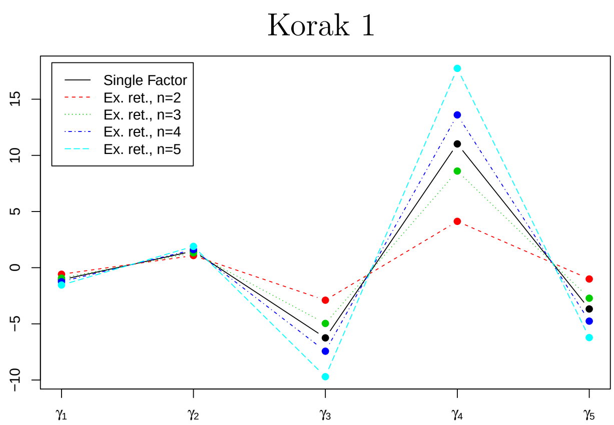
Kot že omenjeno, je regresija z omejitvami izvedena v dveh korakih. Najprej želimo poiskati skupen faktor ter potem določiti, kako je z njim povezano gibanje posameznega presežnega donosa. Skupen faktor, izračunan z regresijami, je predstavljen na Sliki 4 , rezultati obeh korakov pa v Tabelah 5 in 6.

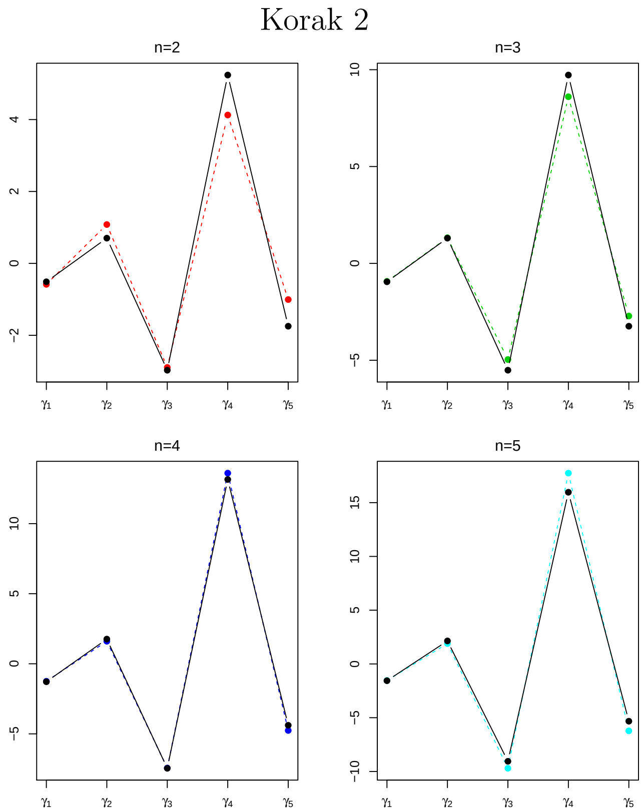
Čeprav t- in F-statistike, izračunane z uporabo 'klasične' kovariančne matrike, poročajo statistično značilnost vpliva neodvisnih spremenljivk, moramo biti pri interpretaciji pozorni na 'ne-normalno' porazdelitev napak, katere so prav tako avtokorelirane. Delno to popravimo z uporabo Newey-West kovariančе matrike: kljub temu, da v tem primeru težko govorimo o statistični značilnosti posamezne spremenljivke, pa lahko strnemo, da je njihov skupen vpliv statistično značilen.

Rezultati za posamezna krajša obdobja so sicer podobni rezultatom, ki sta jih Cochrane in Piazzesi dobila z regresijami na podatkih za ameriške obveznice, so pa zaradi majhnosti vzorca precej manj prepričljivi. Toda tudi, če bi za majhne vzorce lahko 'ovrgli' hipotezo pričakovanj, tega ne moremo trditi za celotni vzorec. Sodeč po rezultatih na trgu nemških obveznic, v nasprotju z rezultati za ameriški trg, hipoteze pričakovanj torej ne moremo ovreči.

Tabela 5: Koeficienti regresije z omejitvami, izračunani na podlagi Koraka 1, predstavljajo enoten skupen faktor. Standardne napake in testne statistike so izračunane s klasično in Newey-West kovariančno matriko. Za izračun so uporabljeni podatki za trg nemških obveznic med 1972 in 1989.

REGRESIJA Z OMEJITVAMI - KORAK 1								
	(Intc.)	yield	f1	f2	f3	f4	R^2	F-stat.
Coef	-11.688	-1.075	1.478	-6.249	11.021	-3.675	0.293	
SE	1.714	0.274	0.927	3.048	4.835	2.678		16.26
t	-6.818	-3.927	1.594	-2.050	2.279	-1.373		
Pr(>|t|)	0.0	0.0	0.113	0.042	0.024	0.172		
$SE^{(NW)}$	3.981	0.598	1.431	5.409	7.991	4.226		5.098
$t^{(NW)}$	-2.936	-1.797	1.032	-1.155	1.379	-0.870		
$Pr(>\|t\|)^{(NW)}$	0.004	0.074	0.303	0.250	0.170	0.386		
5% in 1% kritične vred. F-stat. s 5 in 179 DF sta 2.265 in 3.121.								





Slika 4: Primerjava regresijskih koeficientov, če uporabljamo ločene regresije za presežne donose oz. če uporabljamo regresijo z omejitvami. Na zgornji sliki vidimo koeficiente posameznih regresij in skupen faktor, na spodnji pa korak 2, ki primerja $b_n\boldsymbol{\gamma}$ z β_n.

Tabela 6: Koeficienti regresije z omejitvami, izračunani na podlagi Koraka 2, ki predstavljajo uteži b_n skupnega faktorja za različne ročnosti.

REGRESIJA Z OMEJITVAMI - KORAK 2				
	n=2	n=3	n=4	n=5
Coef	0.475	0.882	1.193	1.449
SE	0.014	0.010	0.005	0.020
$SE^{(NW)}$	0.036	0.029	0.010	0.055
R^2	0.249	0.292	0.318	0.331

Napovedovanje terminske premije

Ustvarjanje donosov je cilj običajnega investitorja. Ko se le-ta odloča o nakupu obveznic različnih dospetij, mu pri odločanju o primernem nakupu zagotovo koristi tudi vedenje o pričakovani vrednosti terminske premije. V magistrski nalogi so predstavljeni rezultati napovedovanja terminske premije za en mesec, šest mesecev ter eno leto vnaprej in sicer z:

- uporabo ločenih regresij za presežne donose različnih ročnosti;
- uporabo 'skupnega faktorja', kot je ta definiran s Cochrane-Piazesiino mero;
- uporabo ARMA oz. natančneje MA(12) modela.

V delu so predstavljeni rezultati napovedi znotraj in izven vzorca (ang. *in-sample*, *out-of-sample*). Pri napovedih znotraj vzorca smo za kalibriranje modelov uporabili bodisi celotno časovno vrsto, bodisi celotno časovno vrsto do vključujoč trenutka napovedi, bodisi obdobje petih let pred trenutkom napovedi (zopet vključno s trenutkom napovedi). Slednji možnosti sta bili prav tako uporabljeni za kalibriranje modelov v primeru napovedi izven vzorca.

Za uspešnost modela smo uporabili tri cenilke oz. kazalnike napake:

- *Povprečno absolutno napako* (ang. *mean absolute error*):

$$MAE = \frac{1}{n}\sum_{i=1}^{i=N} |e_i|;$$

- *Koren povprečne kvadratne napake* (ang. *Root mean square error*):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{i=N} (e_i)^2};$$

- *Koeficient, kateri meri, v koliko primerih se napoved in realizacija gibljeta v isto smer:*

$$DS = \frac{1}{n}\sum_{i=1}^{i=N} \mathbb{1}_{\{napoved\ in\ realizacija\ se\ gibljeta\ v\ isto\ smer\}}.$$

V primeru napovedi znotraj vzorca so rezultati glede na MAE in RMSE skoraj ne razlikujejo; najmanjše napake opazimo pri napovedovanju terminske premije z MA(12) modelom. Rezultati ostalih dveh metod so podobni, kar nakazuje da ima skupni faktor enako moč napovedovanja kot posamezne regresije. Zanimivo je, da se modeli med seboj ne razlikujejo bistveno pri ugibanju smeri terminske premije. Večinoma napovedujejo premik v pravo smer v približno 60% primerov.

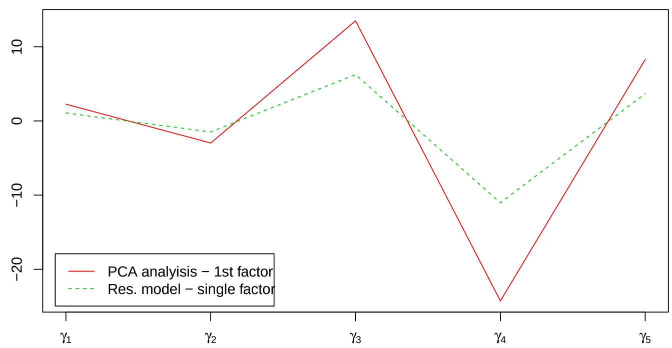
Pri napovedih izven vzorca so rezultati podobni, le-da v tem primeru, regresije občutno bolje napovedujejo rast oz. padec premije, kot to napoveduje MA(12) model. Pri napovedovanju terminskih premij, za obdobje enega leta vnaprej, je uspešnost regresij celo do 70%, medtem ko MA(12) uspešno predvidi smer gibanja le v približno 40%.

Pomen skupnega faktorja

Pri napovedovanju obnašanja časovne strukture obrestnih mer se v ekonometriji pogosto uporabljajo trije faktorji, t.i. nivo, naklon in ukrivljenost (ang. *level, slope, curvature*). Izračunamo jih z uporabo *metode glavnih komponent*, katere glavni cilj je poiskati komponente, ki kar najbolje razložijo razpršenost podatkov. Posledično se s tem tudi zmanjša število neodvisnih spremenljivk uporabljenih v analizi.

Čeprav je bila kovariančna matrika donosov (trenutnih obrestih mer) pogosto preučevana, pa so raziskovalci do sedaj manj pozornosti namenili kovariančni matriki terminskih obrestnih mer, katerih funkcija so tudi presežni donosi. Rezultati nakazujejo, da je prvi faktor, dobljen z metodo glavnih komponent kovariančne matrike terminskih obrestnih mer, soroden skupnemu faktorju, dobljenemu z uporabo Cochrane-Piazzesiine dvokoračne metode (Slika 5).

Ker so terminske premije po definiciji same funkcije trenutnih donosov, lahko presežne donose izrazimo tudi s faktorji, kateri determinirajo donose (torej z nivojem, naklonom in ukrivljenostjo). Za nemške obveznice velja, da čeprav so omenjeni trije faktorji pomembni pri napovedi presežnih donosov, sta pri njihovih napovedih hkrati pomembna tudi četrti in peti faktor (Tabela 7).



Slika 5: Slika prikazuje 'skupen faktor' dobljen s Cochrane-Piazzesiino mero skupaj s prvo komponento, dobljeno z metodo glavnih komponent. Oba faktorja sledita skupnemu vzorcu, pri čemer je en 'večkratnik' drugega. Ta učinek se izniči ob izračunu vrednosti posameznih presežnih donosov (Tabela 17 v magistrskem delu).

Tabela 7: Čeprav glavne tri komponente kovariančne matrike donosov razložijo skoraj celotno varianco donosov, imata tretja in četrta komponenta pomembno vlogo pri interpretaciji razpršenosti presežnih donosov. Rezultati so podani za trg nemških obveznic med 1972 in 1989.

Glavne komponente krivulje donosnosti in njihov učinek					
Pojasnjena varianca	**Komponente**				
	1 (nivo)	2 (naklon)	3 (ukrivljenost)	4	5
trenutni donosi	97.8%	2.0%	0.2%	0.0%	0.0%
presežni donosi	10.8%	71.8%	10.6%	2%	4.8%

Zaključek

V svojem delu *Bond Risk Premia* Cochrane in Piazzesi opišeta preprost regresijski model za napovedovanje terminske premije. Model je sestavljen iz dveh korakov, pri čemer prvi korak identificira t.i. skupen faktor, kateri je enak za presežne donose različnih ročnosti. Skupen faktor naj bi bil soroden prvemu lastnemu vektorju kovariančne matrike terminskih obrestnih mer, dobljenemu z metodo glavnih komponent.

Kot je običajno za strokovno literaturo na tematiko obrestnih mer, je bila njuna analiza narejena za krivuljo donosnosti ameriških trgov. Poglavitni namen magistrske naloge je bila zato analiza modela na podatkih za trge evrskega območja, s čimer bi hkrati tudi testirala robustnost modela.

Čeprav rezultati dobljeni z regresijami na obrestnih merah nemških obveznic potrjujejo obstoj skupnega faktorja in njegovo povezanost s prvim lastnim vektorjem kovariančne matrike terminskih obrestnih mer, so le-ti precej manj prepričljivi od rezultatov za ameriške trge. Koeficient determinacije je občutno nižji in zavzame vrednosti, podane v zgoraj omenjenem članku, le v primeru manjših vzorcev. Prav tako oblika skupnega faktorja ne sledi obliki 'šotora', ampak izriše obliko 'W'.

Cochrane in Piazzesi sta svoje rezultate uporabila tudi za prikaz neveljavnosti hipoteze pričakovanj. Rezultati dobljeni za nemški trg obveznic tega ne morejo potrditi oz. lahko to potrdijo le za krajša časovna obdobja. Možno nadaljevanje obstoječega magistrskega dela je zato dodatno preverjanje veljavnosti regresij na le-teh pod-vzorcih. Dodatna razširitev pa bi bila tudi vključitev nekaterih makroekonomskih spremenljivk (npr. inflacije).