Deriving an Empirical Model of the Canadian zero-coupon yield curves

Tsun Yan Chan

Major Paper Presented to the
Department of Economics of the University of Ottawa
In Partial Fulfillment of the Requirements of the M.A. Economics

Supervisor: Professor Kathleen M. Day

Ottawa, Ontario
May 2020

Abstract. This paper aims to derive an empirical model of the Canadian zero-coupon yield curve using a five-step linear regression method. By emulating from Adrian, Crump and Moench's (2013) model, I am able to fit the model-implied yields to the observable data on the zero-coupon yield curve. The zero-coupon yield curve dataset is constructed by the Bank of Canada and the sample period is January 1987 to December 2011. There are a total of $T = 300$ monthly observations and a cross section of $N = 12$ maturities. First, I compute the principal components of the observable yields and use the first five as state variables. Then, I compute the parameters and derive the model-implied zero-coupon yields. The results show that my model fits well, with very small errors especially for maturities of less than 5 years. I have also estimated the model-implied term premiums.

Acknowledgement

I would like to express my sincere gratitude to my supervisor, Professor Kathleen Day, for her helpful guidance, advice and encouragement throughout the research. She has been very welcoming and keen to respond to my inquiries in details through digital platform, amid the pandemic situation of the novel coronavirus.

Table of Contents

1. Introduction

The derivation of the government zero-coupon yield curve has been a topic of interest for economists, investors and researchers. This is primarily because the yield curve affects how much investors earn by buying government bonds. Its impact on interest rates also determines a person's or a household's borrowing power using credits and loans to buy essential goods and services. Therefore, the trend of the yield curve indeed contributes to the economy significantly. As government bond yields increase, the interest rates on consumer and business loans with similar term lengths will follow. For example, the 30-year government bond yield will affect 30-year mortgages, as banks will tend to charge higher interest on mortgages for a similar maturity duration. Therefore, investors and individuals will be less inclined to buy houses because they have become less affordable. Such a phenomenon will eventually slow GDP growth (Amadeo 2020). As a result, economists are interested in analyzing and deriving the government zero-coupon curve.

Following Adrian, Crump and Moench's approach (2013), in this paper I derive an empirical model of the Canadian zero-coupon yield curve using a five-step linear regression approach. This leads to a considerable amount of empirical work but a more accurate result than other approaches, as the method makes fewer distributional assumptions and produces identified structural parameters. Using Adrian et al.'s (2013) model, I fit the model-implied yields to the zero-coupon yield curve dataset that is constructed from Bank of Canada data. The sample period is from January 1987 to December 2011, with a quarterly maturities interval. The method involves an initial construction of a set of state variables, estimating a VAR(1) model in the state variables and an equation for excess holding returns and finally using these estimates to compute the remaining parameters of the model. Afterwards, I derive the zero-coupon yield curve implied by the underlying theory. The steps of the derivation are documented in Appendix B.

The main results of my study show that my model fits well, as the fitting errors of the model-implied yields are very small. For a maturity period smaller than 5 years, the mean of the absolute fitting errors is close to 0. For a maturity period longer than 10 years, the mean absolute fitting errors are less than 4 basis points. I then estimate the model-implied term premiums, which are shown in the graphs in Appendix A.

One application for the model is to multi-level stress-testing. By obtaining the vector of state variables X_t from the macro-scenario, one can use the same approach to derive the implied ZC curve at every point *t* of the stress-testing horizon. Then, the ZC-curve $y_t^{(n)}$ can be applied to derive the implied government zero-coupon curve. Also, since the method can be used with observable factors, one can use the typical components of macro-stress tests scenario variables such as the interest rate, GDP growth, CPI inflation and the unemployment rate, etc. to forecast the term structure of yield curve.

One limitation of this research relates to the data selected. As the code I constructed using the statistical program R assumes monthly maturity intervals, I must convert the quarterly yield intervals of the data set to monthly yield intervals. This potentially increases the error in the model and likely affects the results to some extent. However, since the implied zero-coupon yield curve is found to fit well with the observable zero-coupon yield curve, I conclude that the model I constructed is quite reliable.[1]

The rest of this paper is structured as follows: Section 2 reviews the literature on different estimation approaches for the model. Section 3 presents the data. Section 4 explains the methodology. Section 5 presents the results of my empirical work, and Section 6 concludes the paper.

[1] However, in principle, it is possible to adjust the R code for quarterly intervals. I was not able to do so within the time available for this research.

2. Literature Review

This section provides a general idea of what interest rates and the zero-coupon yield curve are, and what affine term structure model used in this paper is. I also review past literature on different estimation methods for the model.

2.1. Interest Rates

An interest rate is a measure of the proportion of a loan that is charged to the borrower, as compensation for the risk of default borne by the lender over time. An interest rate is typically expressed as an annual percentage of the loan outstanding. Although the rate and its determinants are difficult to reverse-engineer, they provide valuable information regarding the economy. Collins (2013) explains this notion with the following three reasons. Firstly, the interest rate dictates the consumer's purchasing behavior. For example, if it is at a lower level, one pays lower installments on consumer products. This can encourage more spending on goods and services, ultimately driving economic growth. Secondly, it has a significant impact on governments' deficits. A hike in interest rates is likely to be followed by growing future budget deficits, discouraging domestic investment and the growth of output. In response, the government must devote more of its budget to servicing its debt. Lastly, the interest rate influences capital flows. The higher it becomes, the more foreign capital may be attracted due to the prospect of greater returns. All in all, the interest rate plays a critical role in the overall economy and financial markets.

2.2. The Zero-Coupon (ZC) Yield Curve

A zero-coupon bond is a debt security where the face value is repaid at the time of maturity. Unlike regular bonds, ZC bonds do not pay interest to bondholders. The zero-coupon (ZC) yield curve is a special type of yield indicator that maps interest rates on zero-coupon bonds to different maturities across time. For instance, it assigns the correct discount factors to cash flows that occur at a single point in time. It can also be utilized to price coupon bonds (as portfolios of single cash flows) and hence derive the implied yield on coupon-paying securities. Therefore, the term structure of the yield curve has long been a crucial research topic for economists and investors. Irturk (2006) suggests that the shape of the yield curve can be a reference for forecasting the future behavior of the economy; for example, the ZC curve is used to predict the market's expectations for interest rates.

In another study, Piazzesi (2010) states that given the known condition of an economy, a yield curve model helps us understand how short-term fluctuations project into long term yields. It thus allows the central bank to implement monetary policies accordingly. In the investment world, any spike in interest rates can also impact the return on investments, further emphasizing the potential role of the ZC yield curve in the global financial market.

2.3. The Affine Term Structure Model

The affine term structure (ATS) model is commonly used in finance and macroeconomics to study the behavior of the yield curve. The ATS model originated in Duffie and Kan (1996). They define critical restrictions on the stochastic model for the affine representation of the arbitrage-free model. This ensures that the yield of any zero-coupon bond at any given maturity can be seen to be a dependent affine function of the underlying state variables. They also analyze a simple multi-factor model of the term structure of interest rates using numerical techniques that render the model solvable.

Piazzesi (2010) reports some successes in the study of the affine structure model by showing that bond yield movements over time can be captured by simple vector autoregressions (VARs) in yields. The assumptions and details of the model will be explained later in this paper. In general, the affine model can be expressed by the following equation:

$$y^{(n)} = A(n) + B(n)'X \, ,$$

where $y^{(n)}$ is the yield of an n-period bond, *A(n)* and *B(n)* contain coefficients that depend on the maturity n, and *X* is a vector of state variables. Under varying values of the maturity n, the functions *A(n)* and *B(n)* and the state variables *X* ensure that the yield equations are consistent with each other.

Bolder (2001) also explains the ATS model by addressing two problems in finance: how the ZC interest rate curve is fitted to a set of cross-sectional bond price observations, where the relationship between the ZC interest rate and its term to maturity is called the "term structure of interest rates;" and how the term structure of interest rates behaves over time.

2.4. Different Estimation Methods for the Model

The standard way to estimate the ATS model is to use the Maximum Likelihood (ML) method. It is feasible as it takes advantage of the assumption that a subset of yields is priced with an independent and identically distributed (IID) error. This estimation method involves a numerical search for the parameter values to maximize a nonlinear likelihood function. However, it is well-known that the traditional ATS models can be unidentified, resulting in multiple values of structural parameters being associated with the same reduced-form parameter values (Hamilton and Wu 2012). Consequently, standard models estimated by the ML approach often feature unstable parameter values that are prone to errors due to scaling and initial conditions. It is thus preferable to adopt methods that result in identified structural parameters and require minimal distributional assumptions.

A difference between the approach of Adrian et al (2013) and that of Hamilton and Wu (2012) is that Adrian et al. do not constrain the principal components to be perfectly priced. Therefore, there may exist an inconsistency between the actual and model-implied principal components. However, Adrian et al. (2013) have shown in their paper that any inconsistency is numerically negligible when they use a five-factor specification. Hamilton and Wu (2012) also assume that the yield pricing errors are conditionally independent of lagged values of yield pricing errors. However, Adrian et al. (2013, 144) argue that this is because it may generate excess return predictability which is not captured by the pricing factors.

Several alternative estimation methodologies, which also suffer from problems of their own, have been proposed for pricing the term structure of interest rates. Chen and Scott (2003) apply the ML method to estimate the Cox-Ingersoll-Ross (CIR) model (Cox, Ingersoll and Ross 1985) of the term structure of interest rates. The CIR model estimates non-observable state variables which are generated by a Kalman filter. Although the filter model does not require additional restrictive assumptions, unlike maximum likelihood estimation, the model does not satisfy all the normality assumptions required for statistical consistency. Therefore, the filter for the CIR model is non-linear and may contain bias, and corrective modifications may be required.

Ang and Piazzesi (2003) describe the behavior of the yield curve for bond pricing, using the Gaussian Affine Term Structure model. The model is then reduced to a VAR with no-arbitrage assumptions. They find that observable macroeconomic variables explain up to 85% of the movement in the short and middle parts of the curve, but only around 40% of movement at the long end. Nonetheless, their representations of the model and their methodology are later criticized by Hamilton and Wu (2012) for being unidentified. This implies that unstable parameter values may lead to insignificant results. The same problem has also arisen in Pericoli and Taboga (2008). They derive a canonical representation for the affine model with both observable and unobservable state variables. However, Hamilton and Wu (2012) show that such a representation leads to an unidentified parameter result which can complicate the numerical search process.

In conclusion, traditional ATS models can be unidentified. As a result, standard models estimated by Maximum Likelihood often feature unstable parameter values that can be sensitive to scaling and initial conditions. Thus, it is desirable to follow Adrian, Crump and Moench's (2013) method that can result in identified structural parameters while making fewer distributional assumptions. Their method is outlined in the next section.

3. Methodology

3.1. The Model

This section provides a brief introduction to the affine term structure (ATS) model of Adrian, Crump and Moench (2013) that is used in my research. There are four assumptions made in the ATS model: (A1) Financial markets are free from arbitrage opportunities, which results in the existence of a unique pricing model; (A2) the pricing kernel is exponentially affine in a vector of "state" variables; (A3) the state of the economy can be summarized by a vector of variables with multivariate time series dynamics; and (A4) the logarithm of bond prices and the state variables are jointly normally distributed.

Firstly, I assume that the vector of state variables X_t evolves according to the following vector autoregression (VAR):

$$X_{t+1} = \mu + \Phi X_t + v_{t+1} \, , \tag{1}$$

where v_{t+1} is the factor innovations and μ and Φ are the parameters of the predictable component of X_{t+k}. The development of the state variables can be explained by the intertemporal general equilibrium asset pricing model developed by Cox, Ingersoll and Ross 1985. Specifically, they use the model to study the term structure of interest rates. Next, I assume that the factor innovations v_{t+1} follow a Gaussian distribution with variance-covariance matrix Σ:

$$v_{t+1} | \, \{X_s\}_{s=0}^{t} \sim \mathrm{N}(0, \Sigma) \, , \tag{2}$$

where $\{X_s\}_{s=0}^{t}$ denotes the history of X_t.

Then, I denote by $P_t^{(n)}$ the zero-coupon (ZC) treasury bond price with maturity *n* at time *t*. Under assumption (A1), a pricing kernel M_t (a random variable) is present such that prices are discounted expectations of future cashflows. In other words, today's price is the expectation of tomorrow's price with a discount factor. As such, the following recursive relationship thus holds:

$$P_t^{(n)} = E_t\left[M_{t+1} P_{t+1}^{(n-1)}\right], \tag{3}$$

where M_{t+1} is the pricing kernel in the next period (t+1) and $P_{t+1}^{(n-1)}$is the ZC treasury bond price with the maturity *n* - 1 at time *t* + 1.

Under assumption (A2), the pricing kernel M_{t+1} takes on the following exponentially affine form:

$$M_{t+1} = e^{-r_t - \frac{1}{2}\lambda_t' \lambda_t - \lambda_t' \Sigma^{-1/2} v_{t+1}}, \tag{4}$$

where λ_t is a vector that contains the market prices of risks, and $r_t = lnP_t^{(1)}$ is the continuously compounded risk-free rate. Therefore, if the vector of structural parameter can be estimated, as seen from the above equation, it is then feasible to deduce the pricing kernel.[2] Σ and v_{t+1} are also related to the state variables, as indicated above in equation (1).

The following affine form equation from Duffee (2002) is also assumed for the market prices of risk:

$$\lambda_t = \Sigma^{-\frac{1}{2}}(\lambda_0 + \lambda_1 X_t), \tag{5}$$

where X_t is the vector of state variables, λ_0 and λ_1 contain the "structural" parameters to be estimated, and Σ is the state variable variance-covariance matrix.

[2] The pricing kernel is also known as the stochastic discount factor (SDF). It is a random variable that satisfies the equation to compute the asset price.

The log excess holding return of a bond maturing in n periods, $rx_{t+1}^{(n-1)}$, is defined by the following equation:

$$rx_{t+1}^{(n-1)} = ln\, P_{t+1}^{(n-1)} - ln\, P_t^{(n)} - r_t\ . \tag{6}$$

In addition, yields are affine functions of the vector of state variables X_t. The following are the main equations of the model, which illustrate how the yields can be estimated:

$$ln\, P_t^{(n)} = \ y_t^{(n)}, \tag{7}$$

$$y_t^{(n)} = -\frac{1}{n}(A_n + B_n' X_t) + u_t^{(n)}\ , \tag{8}$$

where $y_t^{(n)}$ is the ZC yield, A_n and B_n are vectors of parameters, and $u_t^{(n)}$ is an error term. The log of model-implied bond prices can be calculated by substituting the observable yields into equation (7). Then, by substituting the log of model-implied bond prices into equation (6), I can compute the excess return of the bonds. The objective is to find the ZC yields. Since X_t are a known input of the model, only the two matrices of parameters, A_n and B_n', remain to be calculated. The state variables X_t are the pricing factors of the yields; they can be macroeconomic indicators such as the unemployment rate, inflation rate, GDP or the interest rate, etc.

In this paper, I will use the first five principal components of the ZC yield curve as the state variables. Hamilton's and Wu's study the pricing factors or state variables are often considered to be unobservable latent variables that are treated as observable. They are sometimes constructed from the observable yields by taking principal components. In using the first five principal components, I follow Adrian et al. (2013), whose method treats the principal components as observable variables. The first three principal components can also be further interpreted respectively as measures of the level of the term structure (also known as "level risk"), the slope of the term structure (also known as "slope risk") and the curvature of the term structure. A brief explanation of the principal components will be provided in the next section.

Now substitute equation (8) into equation (7), to get the following expression:[3]

$$ln\, P_t^{(n)} = -\frac{1}{n}(A_n + B_n' X_t) + u_t^{(n)} \; . \tag{9}$$

Next, we substitute equation (9) into equation (6):

$$rx_{t+1}^{(n-1)} = A_{n-1} + B_{n-1}' X_{t+1} + u_{t+1}^{(n-1)} - A_n + B_n' X_t - u_t^{(n)} + A_1 + B_1' X_t - u_t^{(1)}. \tag{10}$$

Adrian, Crump, and Moench (2013) show that under assumptions A1 – A4 the observed excess returns (i.e., the return generating process for log-excess holding period returns) are given by the following linear regression equation:

$$rx_{t+1}^{(n-1)} = \underbrace{\beta^{(n-1)'}(\lambda_0 + \lambda_1 X_t)}_{\substack{\text{Expected}\\ \text{return}}} - \underbrace{\frac{1}{2}\left(\beta^{(n-1)'}\Sigma\beta^{(n-1)} + \sigma^2\right)}_{\substack{\text{Convexity}\\ \text{adjustment}}} + \underbrace{\beta^{(n-1)'} v_{t+1}}_{\substack{\text{Priced return}\\ \text{innovation}}} + \underbrace{e_{t+1}^{(n-1)}}_{\substack{\text{Return pricing}\\ \text{error}}} \; , \tag{11}$$

where $\beta = \left[\beta^{(1)}\beta^{(2)} \cdots \beta^{(N)}\right]$ is a K × N matrix of factor loadings, and σ^2 is the variance of the $e_{t+1}^{(n-1)}$.

As shown, this equation can be divided into four parts: expected return, convexity adjustment, priced return innovation, and return of pricing error. After substituting equation (10) into equation (11), Adrian, Crump, and Moech (2013) show that the parameters A_n and B_n' obey the following recursions:

$$A_n = A_{n-1} + B_{n-1}'(\mu - \lambda_0) + \frac{1}{2}(B_{n-1}'\Sigma B_{n-1} + \sigma^2) - \delta_0 \; , \tag{12}$$

$$B_n' = B_{n-1}'(\Phi - \lambda_1) - \delta_1' \; , \tag{13}$$

[3] In implementing the model, I assume that $u_t^{(n)}$ is zero and do not calculate it.

$$A_0 = 0\ ,\ \ B_0' = 0\ , \tag{14), (15}$$

$$B^{(n)\prime} = B_n'\ . \tag{16}$$

δ_0 and δ_1 are the short rate parameters, which will be defined in step 4 of the estimation method. These equations indicate that the parameters μ, Φ, Σ, σ^2, λ_0, λ_1, δ_0, and δ_1 are necessary for the computation of A_n and B_n'. To determine their values, I will use Adrian, Crump and Moech's (2013) regression-based estimator.

3.2. Estimation Method

Adrian, Crump and Moech (2013) establish what they describe as a three-step ordinary least squares (OLS) estimator to compute the implied ZC yield curves. In this section, their estimation method is further explained in five steps.

Step 1 (obtain μ, Φ, Σ):

First, the vector of state variables X_{t+1} is regressed on its lagged value X_t as in equation (1), to obtain the parameters of the predictable component (μ , Φ) and the factor innovations v_{t+1}. After this VAR(1) has been estimated,, the vectors of shocks $\hat{v}_{t+1}$ are stacked to form the matrix $\hat{V}$, and then the variance-covariance matrix is estimated as follows:

$$\hat{\Sigma} = \hat{V}\hat{V}'/T\ . \tag{17}$$

Step 2 (obtain a, β, c, E, σ^2, $\hat{B}^*$ for λ_0 and λ_1):

Next, the excess returns are measured as shown in equation (6). Then, I regress them on the lagged levels of the state variables X_t and factor innovations $\hat{v}_{t+1}$:

$$rx = a\iota_T' + \beta'\hat{V} + cX_- + E\ , \tag{18}$$

where *rx* is the a matrix of excess returns of all bonds, $\beta = [\beta^{(1)}\beta^{(2)} \dots \beta^{(N)}]$ is a matrix of factor loadings, ι_T' is a vector of ones and $X_- = [X_0 X_1 \dots X_{T-1}]$ is a matrix of lagged state variables. Note that the β vector in equation (18) are identical to the ones appearing in equation (11). After that, I collect the matrix of residuals $\hat{E}$ from the regression, and estimate $\hat{\sigma}^2 = \frac{trace(\hat{E}\hat{E}')}{NT}$. Note that *N* indicates the number of maturities. Then, $\hat{B}^*$ is constructed from $\hat{\beta}$ as follows:

$$\hat{B}^* = [vec(\beta^{(1)}\beta^{(1)\prime}) \dots vec(\beta^{(N)}\beta^{(N)\prime})] \; . \tag{19}$$

Step 3 (obtain λ_0 and λ_1):

As mentioned previously, the critical objective of this model is to estimate the parameters λ_0 and λ_1, as they are needed to estimate the pricing kernel. Once we have the estimates â, $\hat{\beta}$, $\hat{c}$, $\hat{B}^*$, $\hat{\sigma}^2$, $\hat{\Sigma}$, we can utilize the following expressions to obtain the market price of risk of the "structural" parameters λ_0 and λ_1:

$$\hat{\lambda}_0 = \left(\hat{\beta}\hat{\beta}'\right)^{-1}\hat{\beta}(\hat{a} + \tfrac{1}{2}\left(\hat{B}^* vec\left(\hat{\Sigma}\right) + \hat{\sigma}^2 \iota'_N\right) \; , \tag{20}$$

$$\hat{\lambda}_1 = (\hat{\beta}\hat{\beta}')^{-1}\hat{\beta}\hat{c} \; . \tag{21}$$

Step 4 (obtain A_n and B_n):

As mentioned earlier, I recursively calculate A_n and B_n using equations (12) – (16), where δ_0 and δ_1 are the short rate parameters. Short rate means that the yields mature in one month; in other words, the parameters of the yields when *n* = 1. The values of the two short rate parameters can be obtained by substituting *n* = 1 into equation (8) and regressing the one-month T-bill rate on the state variables. This means that:

$$\hat{y}_1^{(1)} = -\tfrac{1}{12}(\delta_0 \; + \delta_1' X_1) \; . \tag{22}$$

After setting $A_1 = -\hat{\delta}_0$ and $A_1 = -\hat{\delta}_1$, all *As* and *Bs* can be derived computationally.

Step 5 (obtain $y_t^{(n)}$):

Once we have A_n and B_n, we can use the equation $\hat{y}_t^{(n)} = -\frac{1}{n}(A_n + B_n' X_t)$ to derive the entire ZC-curve as a function of X_t.

The result of all this work is to produce the $\hat{y}_t^{(n)}$ values for all the *n*. A more detailed derivation of the model is provided in Appendix B.

4. Data

4.1. The Zero-coupon yields

The data used in this paper are the Yield Curves for Zero-Coupon Bonds, constructed by the Bank of Canada (BoC).[4] It contains daily yield curves for zero-coupon bonds, generated using pricing data for Government of Canada bonds and treasury bills. The maturity terms are quarterly based, ranging from 0.25 years (3 months) to 30 years (120 months). All numbers are expressed in decimals (e.g. 0.0100 => 1.00% yield).

The rationale for using the BoC data set is because it is the most recent and credible Canadian data available for this study. However, one limitation remains as the Bank measures yields quarterly, while the methodology and programming in this study is based on a monthly interval. In order to convert daily data to monthly data, I take the end-of-month value of each month. Since the original raw data yields of maturity are measured quarterly, I plot the graph with this raw data and obtain monthly yields

[4] The source of the data can be found at Bank of Canada (n.d.) *Yield curves for zero-coupon bonds,* https://www.bankofcanada.ca/rates/interest-rates/bond-yield-curves/

by interpolation. As such, errors in the model may increase as the quarterly yield data were converted to monthly yields.

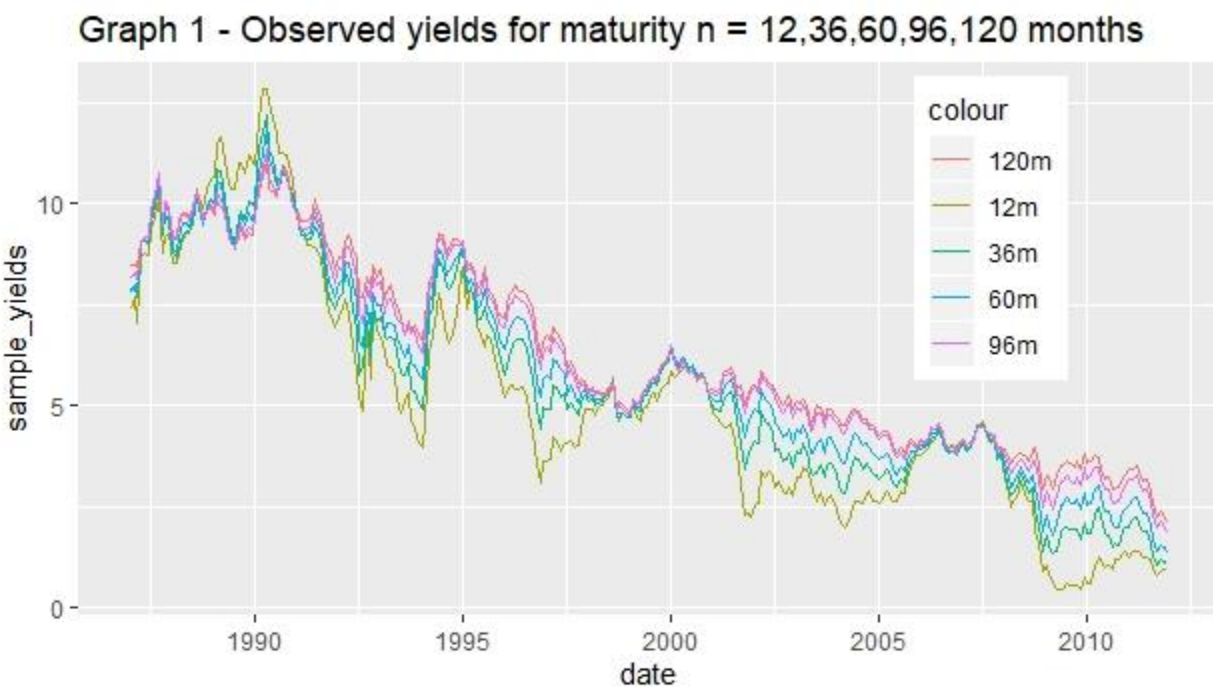
My sample period is from January 1987 to December 2011. This is the same period chosen in Adrian, Crump, and Moench (2013). There are a total of T = 300 monthly observations, and a cross section of N = 12 maturities. The yields of the maturities are n = 6, 12, 18, 24, 36, 48, 60, 72, 84, 96, 108, and 120 months. I then extract the principal components from these yields. Table 1 presents the summary statistics of the yields.

Table 1. Summary Statistics: Means and standard deviations

Yields to maturity (months)	Mean (sd)
n = 6	5.009 (3.125)
n = 12	5.127 (2.992)
n = 18	5.250 (2.879)
n = 24	5.363 (2.785)
n = 36	5.097 (2.655)
n = 48	5.290 (2.570)
n = 60	5.369 (2.506)
n = 72	5.967 (2.452)
n = 84	6.067 (2.407)
n = 96	6.151 (2.370)
n = 108	6.225 (2.341)
n = 120	6.294 (2.322)

As shown in table 1, the mean yields climb gradually as the time to maturity lengthens. For example, the mean for a six-month term is approximately 5.0%, while that for a ten-year term is approximately 6.3%. The percentage increase in the yield is 26%. On the other hand, the standard deviation of the yields decreases as the time to maturity increases.

Graph 1 shows the observed yields for maturities n = 12, 36, 60, 96, and 120 months. The red, purple, blue, green and yellow lines indicate the yields for maturities of different durations: ten years, eight years, five years, three years and one year respectively. We can see that yields are highly correlated across maturities, with an overall decreasing trend in the percentage yield over the time period. During the first subperiod, covering the time between 1987 and 1996, yields are relatively high and volatile, but a considerable drop is observed between 1990 and 1994. It is also interesting that during the time between 1989 and 1991, yields for longer terms of maturity are lower than those for shorter terms. This type of yield curve is rare and is considered a predictor of an economic recession; indeed, a recession occurred after 1990. During the second subperiod, covering the time between 1996 and 2011, yields are low but more stable.



4.2. The Principal Components

I follow Adrian, Crump and Moech (2013) in choosing the state variables X_t to be the first five principal components of the cross-section of N = 12 maturities. Principal components (PC) is a data reduction technique that allows statisticians to extract the independent sources of variation from a group of highly correlated variables. The principal components are normalized orthogonal linear combinations of the N = 12 yields. The core of a principal components analysis is represented by the eigenvectors and eigenvalues of a covariance matrix. The eigenvectors (the principal components) determine the directions of the data, while the eigenvalue is a number that explains how much variance there is in the data along that eigenvector.

In other words, a larger eigenvalue indicates that the principal component explains a larger amount of variance in the data. The first principal component explains the largest percentage of the variance of the component yields, whereas the second principal component explains the second largest percentage variation of the component yields, etc. Thus, using principal component analysis, I can create a reduced set of variables that still contains most of the important information in the larger set. These new variables have the advantage of being uncorrelated with each other.

Since twelve maturities are used to construct the principal components, there will be twelve principal components. The usual procedure is to return a subset of the principal components, such that the selected principal components account for a high proportion of the variation in the original data. Table 2 presents the contribution of the first five principal components of the observable ZC yields.

Table 2. The Contribution of the first Five Principal Components

PC1	PC2	PC3	PC4	PC5
81.108	17.580	0.103	0.013	0.006

The table shows the percentage of the total variation explained by each component. The contribution of the first principal component is the eigenvalue of that principal component divided by the sum of all eigenvalues.[5] The results show that the first principal component explains 81% of the variance of the component yields, while the second principal component explains approximately 17% of the variance of

[5] In Table 2, the proportions have been multiplied by 100.

the component yields. The fifth principal component only explains 0.008% of the variance of the yields. The sum of the principal components is the total variation of the data.

5. Results

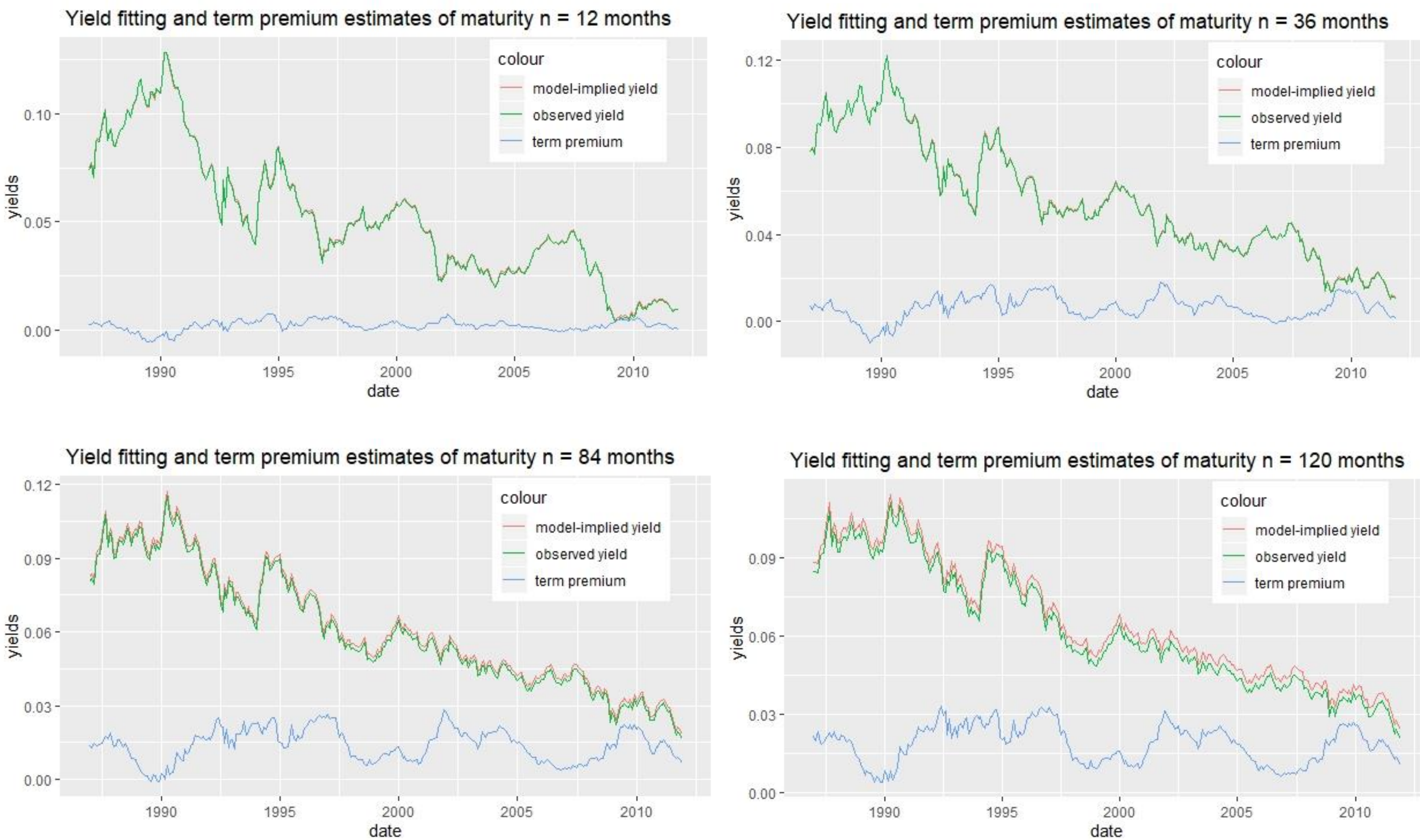
As discussed in section 3, the objective of all this work is to produce the $\hat{y}_t^{(n)}$ values for all the *n*.[6] First, I estimate the VAR(1) model, equation (1), to obtain its parameters and residuals. The residuals from this model are used to construct the matrix $\hat{V}$. Then, I estimate equation (18) using OLS. This requires first using equation (7) to compute the prices for each *n*, and then substituting them into equation (6) to compute the excess returns. Then equation (18) is estimated using the newly computed excess returns as the dependent variables. After that, I obtain the market price of risk "structural" parameters λ_0 and λ_1 by substituting the estimated parameters from equation (18) into equation (20) and (21). Next, I recursively calculate A_n and B_n using equations (12) – (16) and obtain the $\hat{y}_t^{(n)}$ values using equation (22).

Figure 1 illustrates the results of the yield fitting model and term premium estimates for maturities of *n* = 12, 24, 60 and 120 months respectively. The green, red and blue lines represent the observed yields, model-implied yields, and model-implied term premiums respectively. The model-implied fitting errors are the difference between actual yields $y_t^{(n)}$and the model-implied yields $\hat{y}_t^{(n)}$. As shown in the graphs in Figure 1, I observe that the fitting errors of the model-implied yields $\hat{y}_t^{(n)}$ are very small. For $n \leq 5$ years, the mean of the absolute fitting errors is close to 0. For $n \geq 10$ years, the mean absolute fitting errors are less than 4 basis points. Therefore, the model fit is comparable to that of other ATS models as it performs well within an acceptable range of pricing error. For example, Piazzesi (2010) states that the

[6] All calculations were programmed in R.

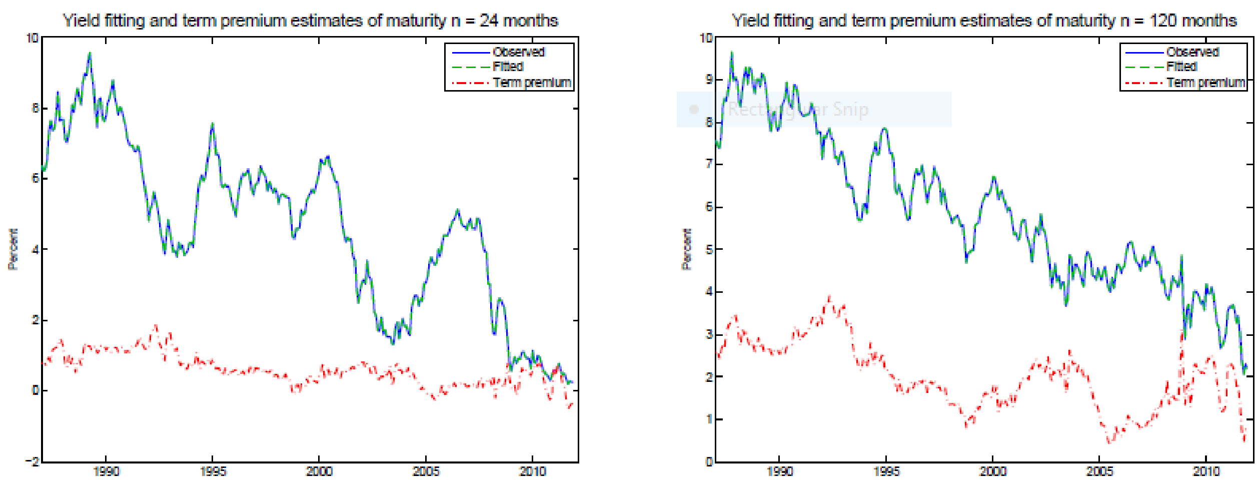
mean absolute value of fitting errors for a 5-year maturity yield is 8 basis points. All fitting errors of my model are smaller than the range as stated in Piazzesi (2010)'s paper. The term premium is given by the difference between the risk neutral yield and model-implied yield. The risk neutral yield assumes there is no price for risk in the yield curve. It is thus measured by setting the price of risk parameters λ_0 and λ_1 to zero in equations (12) and (13). The results generated for the risk-free yields are then subtracted from the model-implied fitted yield to get the model-implied term premium.

Figure 1 Five-factor model: observed, model-implied yields and term premiums



In order to compare my results with Adrian, Crump, and Moench's model. I present the graph below shown in Figure 2. The dataset used in their paper is the U.S. zero coupon yield data over the sample period 1987:01 – 2011:12. Under the five-factor specification, the yields and term premiums for the two-year and ten-year maturities are as shown in Figure 2. The observed yields, model-implied yields and model-implied term premiums are represented by the solid, dashed and dash-dotted lines respectively. I can see that there is no visible pricing error in their model at either maturity. This may be due to the fact that they have computed the log bond pricing errors ($u_t^{(n)}$) in their model, which I didn't.

Figure 2 Five-factor model: observed, model-implied yields and term premiums



Source: Figure 5 of Adrian, Crump, and Moench (2013)

Table 3 presents the results of the VAR (1) model. It shows the estimates of μ and beta Φ in equation (6) respectively. PC1, …, PC5 denote the first five principal components of the government zero-coupon yields, which are used as the state variables.

Table 3. The estimates of μ and Φ

	PC1	PC2	PC3	PC4	PC5
μ	-0.0216	-0.0065	0.0005	0.0001	0.0007
(t-statistic)	(-1.1348)	(-0.3086)	(0.05451)	(0.0312)	(0.2050)
Φ_{i1}	**0.9955**	-0.0028	0.0321	-0.1549	0.0683
(t-statistic)	(136.0099)	(0.4788)	(0.8607)	(0.1554)	(-0.0887)
Φ_{i2}	0.0039	**0.9635**	0.1382	-0.1219	-0.2653
(t-statistic)	(-0.1962)	(60.2158)	(-0.7275)	(-0.3650)	(0.5206)
Φ_{i3}	0.0034	-0.0057	**0.8343**	**0.0078**	0.0743
(t-statistic)	(0.5394)	(2.0906)	(25.9009)	(-2.3204)	(-0.5781)
Φ_{i4}	0.0003	-0.0012	-0.0320	**0.7305**	0.0726
(t-statistic)	(-0.9110)	(-0.6447)	(0.0847)	(18.5147)	(1.0861)
Φ_{i5}	-0.0001	0.0014	-0.0062	0.0335	**0.5847**
(t-statistic)	(0.2670)	(-0.9323)	(0.5360)	(1.2234)	(12.5807)

Notes: PC1,…,PC5 represent the first through fifth principal components of the government zero-coupon bonds. t-statistics are in parentheses. Each row denotes an explanatory variable; thus Φ_{i1} represents the estimated coefficients of the first pricing factor and so on. Bold coefficients represent significance of the estimated parameters using for a two tailed t test at the 5% level.

The t-statistics for μ and Φ are shown in parentheses. The results show that there are few relationships between the principal components. A VAR(1) process is stable if all eigenvalues have modulus less than 1. I test for stability and the results show that the roots are outside the unit circle, hence, my VAR(1) process is stable.

Table 6 summarizes the estimates of the market price of risk parameters λ_0 and λ_1. The t-statistics are reported in parentheses. Following Adrian et al. (2013), I calculate the standard errors under the assumption that μ is unknown, meaning that I have used demeaned yields to accommodate the sampling uncertainty. The estimated price of risk coefficients are economically important. The first principal component is a measure of the level of the term structure (also known as “level risk”). As shown in the top element of λ_0 in table 6, the price of level risk has a negative significant negative constant component. This result is similar to Adrian et al.’s (2013) result as shown in Table 7. The second principal component also has a significant negative coefficient at the 10% level. Since the loading of level risk on the slope factor of my model is negative, therefore, the negative coefficient of PC2 in the equation of PC1 implies that a higher slope leads to a higher expected return. However, unlike the finding of Adrian et al. (2013) , my fifth principal component shows no significant effect on level risk. Like Adrian et al. (2013), I find that slope risk is not priced, as there are no significant coefficients in the row for PC2 in Table 6.

Table 6. Market Prices of risk

Factor	λ_0	$\lambda_{1.1}$	$\lambda_{1.2}$	$\lambda_{1.3}$	$\lambda_{1.4}$	$\lambda_{1.5}$
PC1	**-0.0550**	0.0029	**-0.0276**	0.0930	-0.1007	0.0255
(t-statistic)	(-2.8872)	(-0.4069)	(-1.9202)	(-1.5654)	(-0.5916)	(-0.0994)
PC2	-0.0081	-0.0033	-0.0048	-0.0837	-0.2858	-0.0720
(t-statistic)	(-0.3807)	(-0.4084)	(-0.2980)	(-1.2571)	(-1.5004)	(-0.2511)
PC3	0.0039	0.0016	-0.0070	**-0.0851**	0.1612	-0.0558
(t-statistic)	(-0.3683)	(-0.3888)	(-0.8260)	(-2.5788)	(-1.7009)	(-0.3905)
PC4	0.0015	-0.0005	0.0028	**-0.0407**	**-0.2731**	-0.0220
(t-statistic)	(-0.3307)	(-0.2762)	(-0.3321)	(-2.8764)	(-6.7169)	(-0.3529)
PC5	0.0046	0.0006	-0.0009	-0.0018	0.0466	**-0.4278**
(t-statistic)	(-1.3005)	(-0.4259)	(-1.5654)	(-0.1640)	(-1.4704)	(-8.9532)

Note: Bolded coefficients represent significance at the 10% level.

Table 7. Adrian et al. (2013) Market Prices of risk

Factor	λ_0	$\lambda_{1.1}$	$\lambda_{1.2}$	$\lambda_{1.3}$	$\lambda_{1.4}$	$\lambda_{1.5}$
PC1	**-0.019**	−0.003	**−0.016**	−0.005	0.012	**0.030**
(t-statistic)	(-2.566)	(−0.443)	(−2.160)	(−0.648)	(1.605)	(3.987)
PC2	-0.013	0.027	−0.011	−0.003	−0.011	0.015
(t-statistic)	(0.951)	(1.914)	(−0.818)	(−0.213)	(−0.792)	(1.077)
PC3	−0.030	**−0.077**	−0.001	**−0.093**	**−0.132**	−0.056
(t-statistic)	(−0.951)	(−2.466)	(−0.029)	(−2.987)	(−4.244)	(−1.783)
PC4	0.042	0.064	−0.007	0.015	−0.058	**−0.086**
(t-statistic)	(1.062)	(1.594)	(−0.189)	(0.367)	(−1.461)	(−2.147)
PC5	0.005	**−0.105**	0.012	−0.004	−0.073	**−0.324**
(t-statistic)	(0.097)	(−2.028)	(0.243)	(−0.070)	(−1.431)	(−6.287)

Source: Table 3 of Adrian, Crump, and Moench (2013)

6. Conclusion

In this paper, I follow Adrian, Cump and Moench's (2013) approach to estimate the affine term structure model. I first use the principal components approach to get the initial five principal components that are used as the state variables of the model. After deriving all the equations and programming techniques, I then obtain the implied (theoretical) government zero-coupon yield curve using only linear regression methods.

This model can be applied for solvency stress-testing. Given the value of a state vector X_t, the model can be used to derive the ZC yield for any maturity using the equation $\hat{y}_t^{(n)} = -\frac{1}{n}(A_n + B_n' X_t)$. When used as part of a multi-level stress-testing model, the analyst will obtain a "path" of the state vector X_t from the macro-scenario, and using the above bond pricing equation, will derive the implied ZC curve at every point *t* of the stress-testing horizon. Then, the ZC-curve $y_t^{(n)}$, for n = 1, 2, 3, …, 10 years (etc.) can be applied to derive the implied government zero-coupon curve.

An advantage of this approach is that it does not rely on numerical optimization to estimate the parameters, which means that unique solutions exist for the parameters. In the results section, one can see that the model gives small fitting errors. Since the method can be used with observable factors, an implied zero-coupon yield curve can also be generated with typical components of macro-stress test scenario variables such as the interest rate, GDP growth, CPI inflation and the unemployment rate, etc. However, one limitation of the model is that the raw data set from the Bank of Canada is quarterly based, and I converted the dataset to monthly based. It may be more accurate and desirable if a monthly dataset was available. Further work could be done to explore whether pricing errors for n > 10 years can be improved further.

In order to compare my results with that by Adrian et al. (2013), I have chosen the exact same sample period as theirs. However, the model could also be applied to a longer sample period. I also chose to extract five principal components to use as state variables. This is because Adrian et al. (2013) compared the results of three factor, four factor, and five factor models and found that the five factor model worked best. Further work could be done for Canada to see if different choices for the state variables would work better for deriving the zero-coupon yields.

7. References

Adrian, Tobias, Richard K. Crump, and Emmanuel Moench (2008) "Pricing the term structure with linear regressions," *Journal of Financial Economics* 110, 110-38

Amadeo, Kimberly (2020) *How US treasury yields affect the economy,* March 12, https://www.thebalance.com/treasury-yields-3305741

Ang, Andrew, and Monika Piazzesi (2003) "A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables," *Journal of Monetary Economics* 50, 745-87, doi:10.1016/S0304-3932(03)00032-1

Bolder, David (2001) "Affine term-structure models: theory and implementation," *Bank of Canada Web site,* October, https://www.bankofcanada.ca/2001/10/working-paper-2001-15/

Chen, Ren-Raw, and Louis Scott (2003) "Multi-factor Cox-Ingersoll-Ross models of the term structure: estimates and tests from a Kalman filter model," *Journal of Real Estate Finance and Economics* 27 (2): 143-72

Collins Jr., J. Patrick (2013) *Why interest rates are so important,* June 17, https://greenspringadvisors.com/blog-article/why-interest-rates-are-so-important/

Cox, J.C., J.E. Ingersoll, and S.A. Ross (1985) "A theory of the term structure of interest rates," *Econometrica* 385-407, doi:10.2307/1911242

Duffie, Darrell, and Rui Kan (1996) "A yield-factor model of interest rates," *Mathematical Finance* 6: 379-406, doi:10.1111/j/1467-9965.1996.tb00123.x

Fidelity Investment (n.d.) *Bond prices, rates, and yields,* https://www.fidelity.com/learning-center/investment-products/fixed-income-bonds/bond-prices-rates-yields

Gordon, Jack (n.d.) *Do interest rates help a budget deficit?* Hearst Newspapers, LLC, https://smallbusiness.chron.com/interest-rates-budget-deficit-81589.html

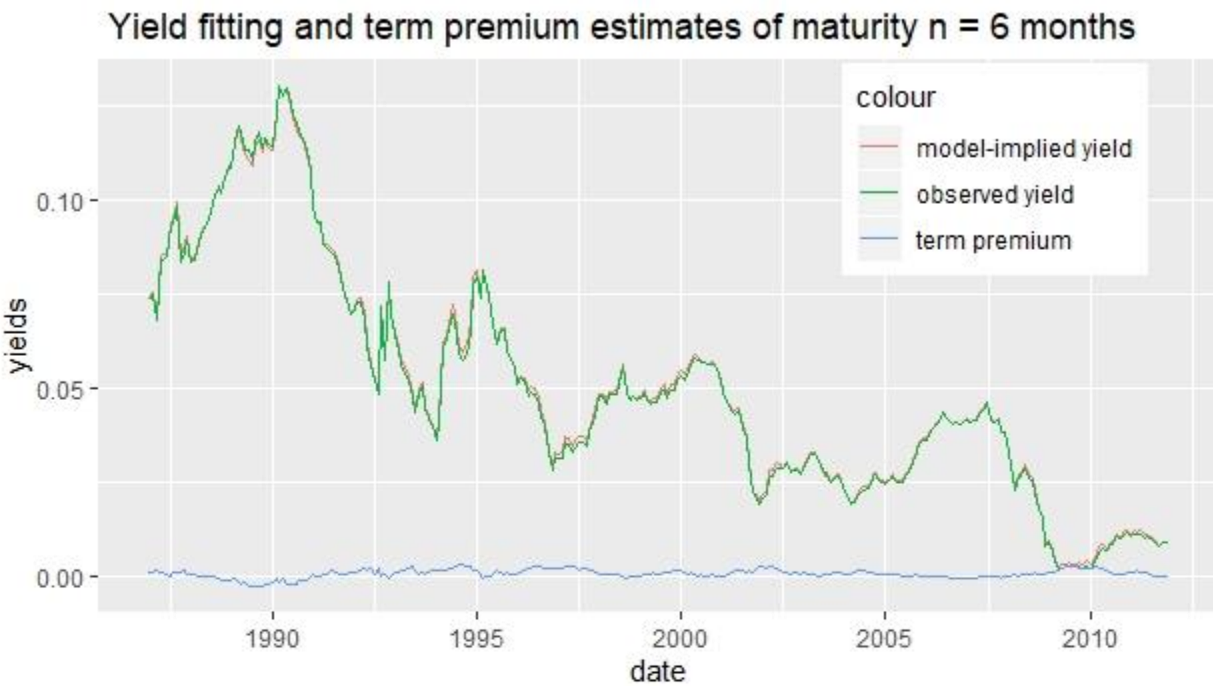
Hamilton, James D., and Jing Cynthia Wu (2012) "Identification and estimation of Gaussian affine term structure models," *Journal of Econometrics* 168, 315-31

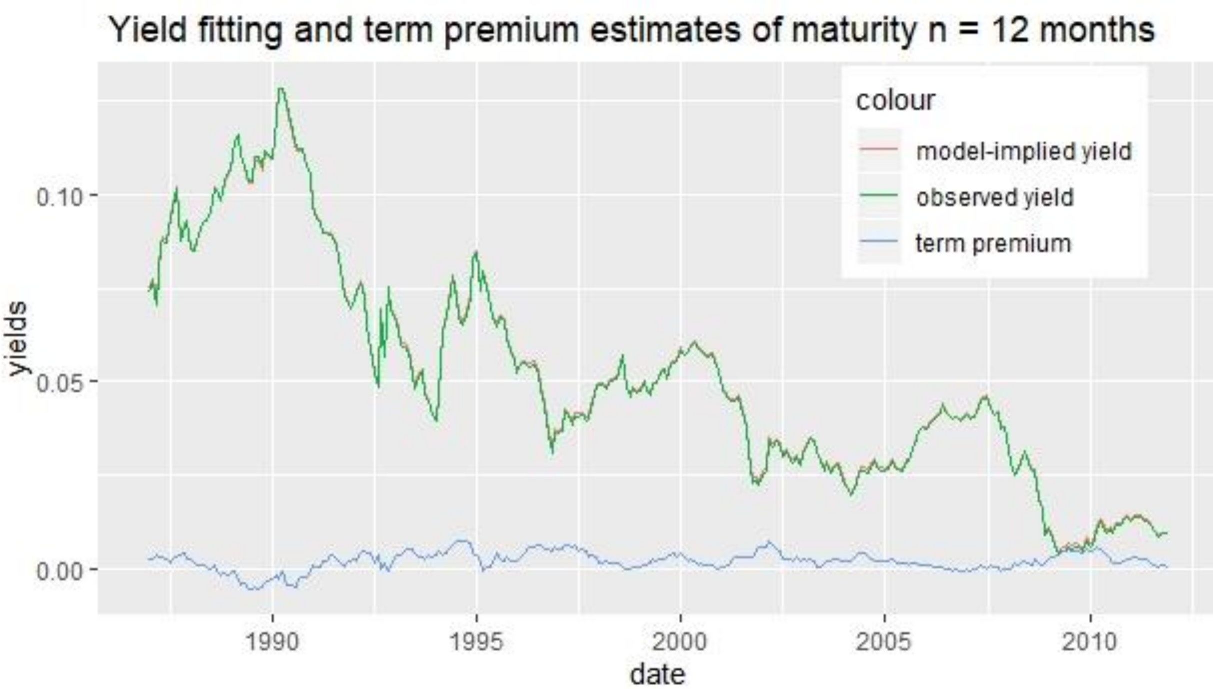
Irturk, Ali U. "Term structure of interest rates." Master's thesis, University of California, Santa Barbara, 2006.

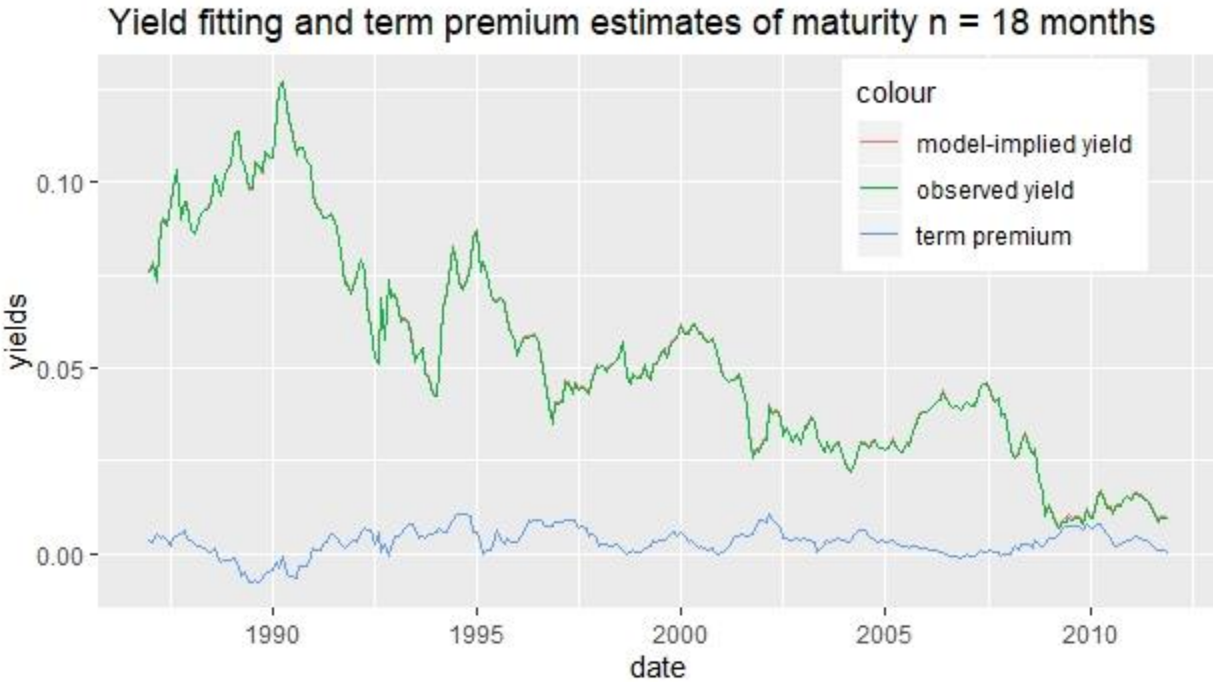
Pericoli, Marcello, and Marco Taboga (2008) "Canonical term-structure models with observable factors and the dynamics of bond risk premia," *Journal of Money, Credit, and Banking* 40 (7): 1471-88, doi:10.1111/j.1538-4616.2008.00167.x

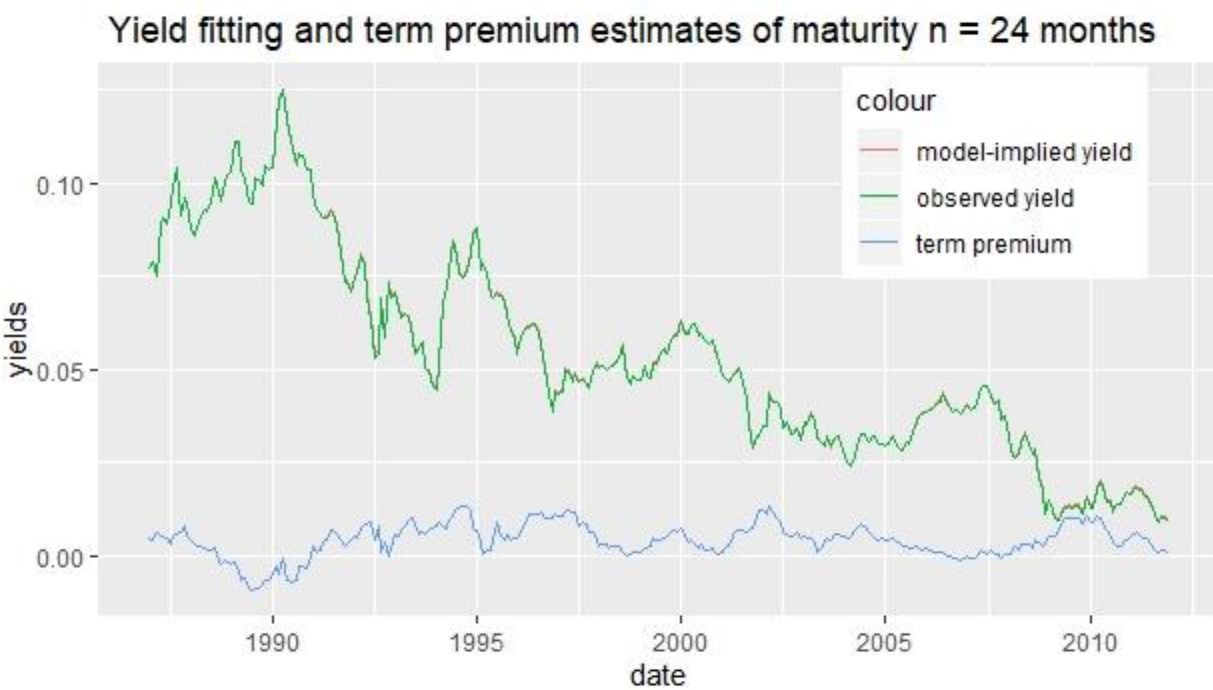
Piazzesi, Monika (2010) "Affine term structure models," Edited by Yacine Ait-Sahalia and Lars Hansen, *Handbook of Financial Econometrics* 1 (12): 691-766

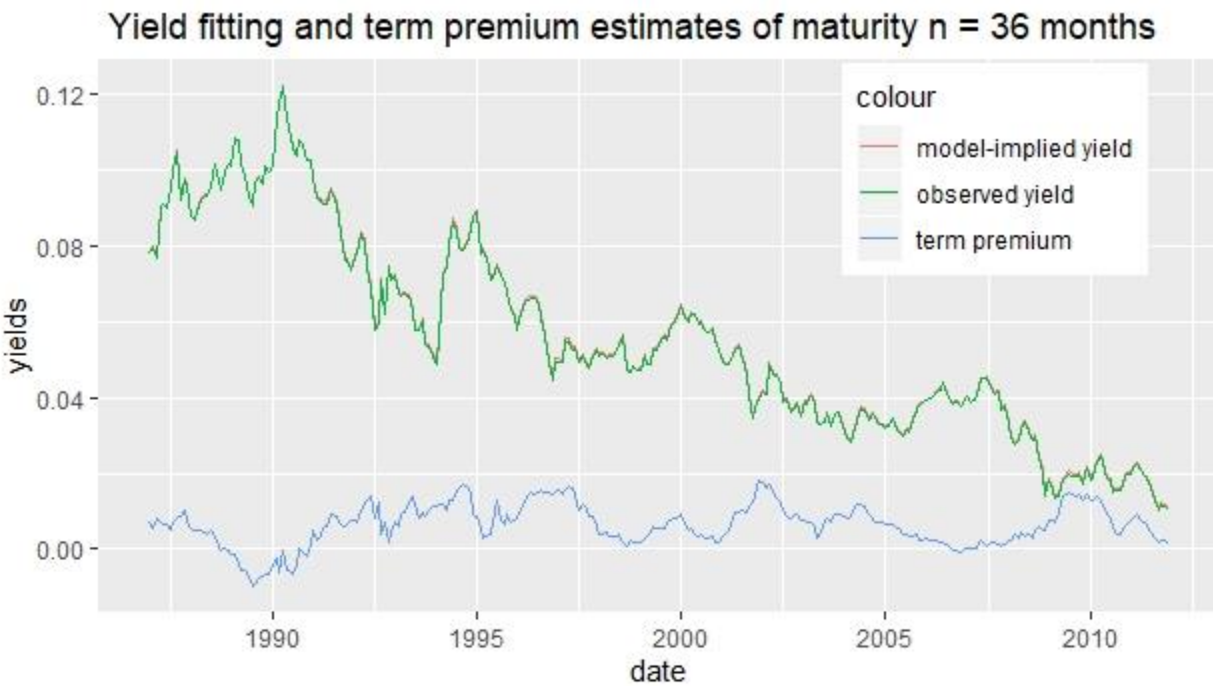
Appendix A

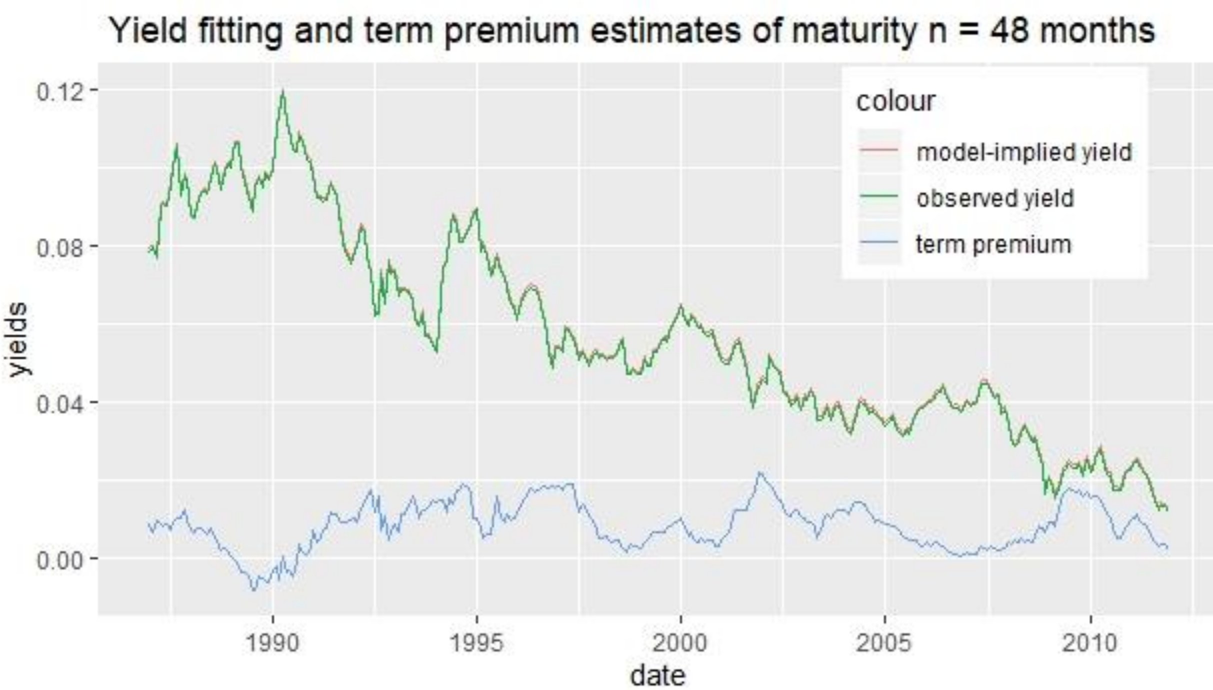

Yield fitting and term premium estimates of maturity n = 6 months
colour
model-implied yield
observed yield
term premium
0.10
0.05
0.00
yields
1990
1995
2000
2005
2010
date

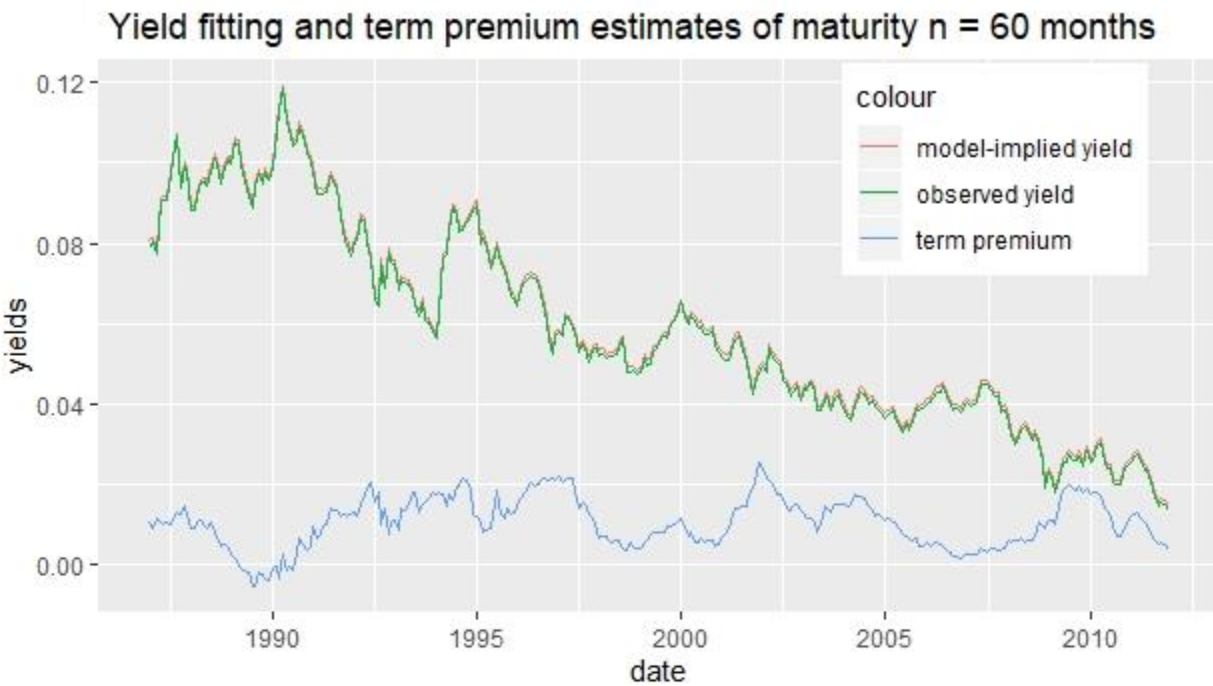

Yield fitting and term premium estimates of maturity n = 12 months
colour
model-implied yield
observed yield
term premium
0.10
0.05
0.00
yields
1990
1995
2000
2005
2010
date

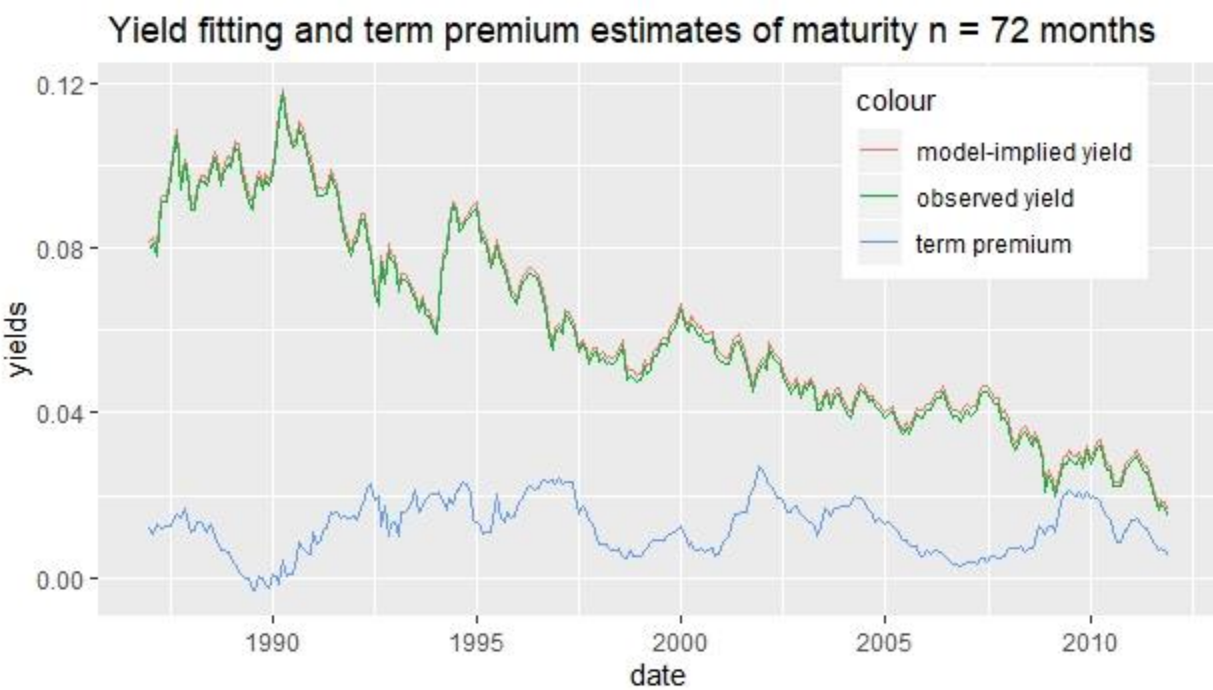

Yield fitting and term premium estimates of maturity n = 18 months
colour
model-implied yield
observed yield
term premium
yields
0.10
0.05
0.00
1990
1995
2000
2005
2010
date

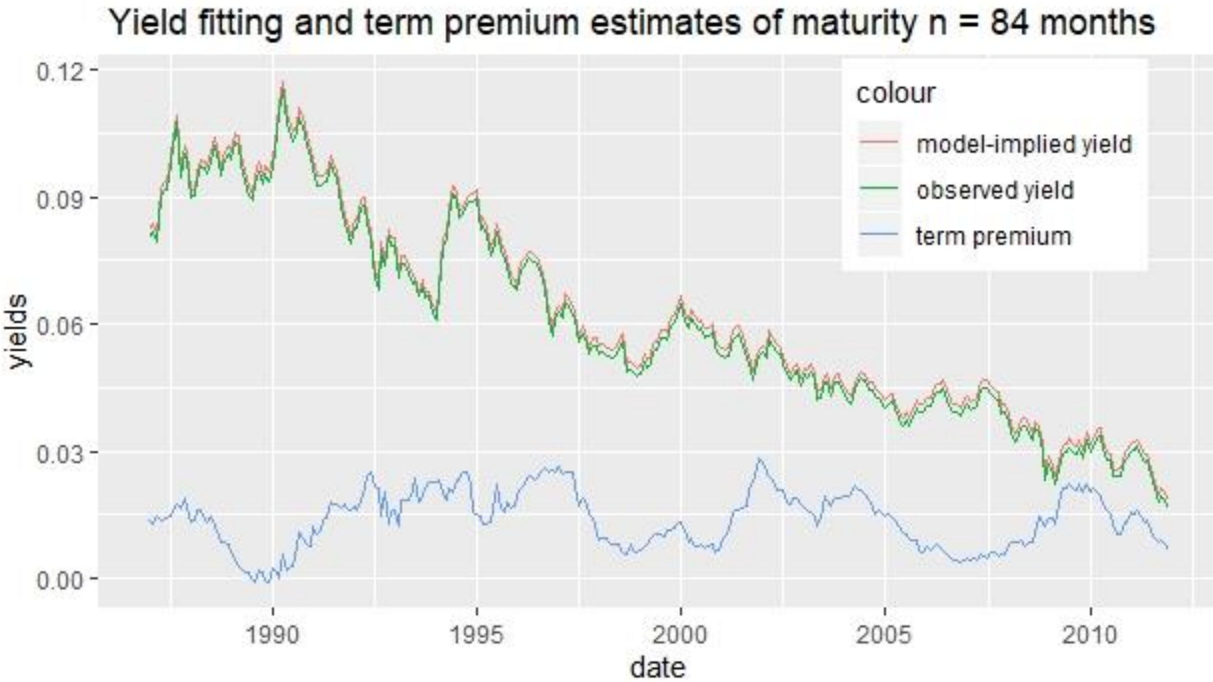

Yield fitting and term premium estimates of maturity n = 24 months
colour
model-implied yield
observed yield
term premium
yields
0.10
0.05
0.00
1990
1995
2000
2005
2010
date

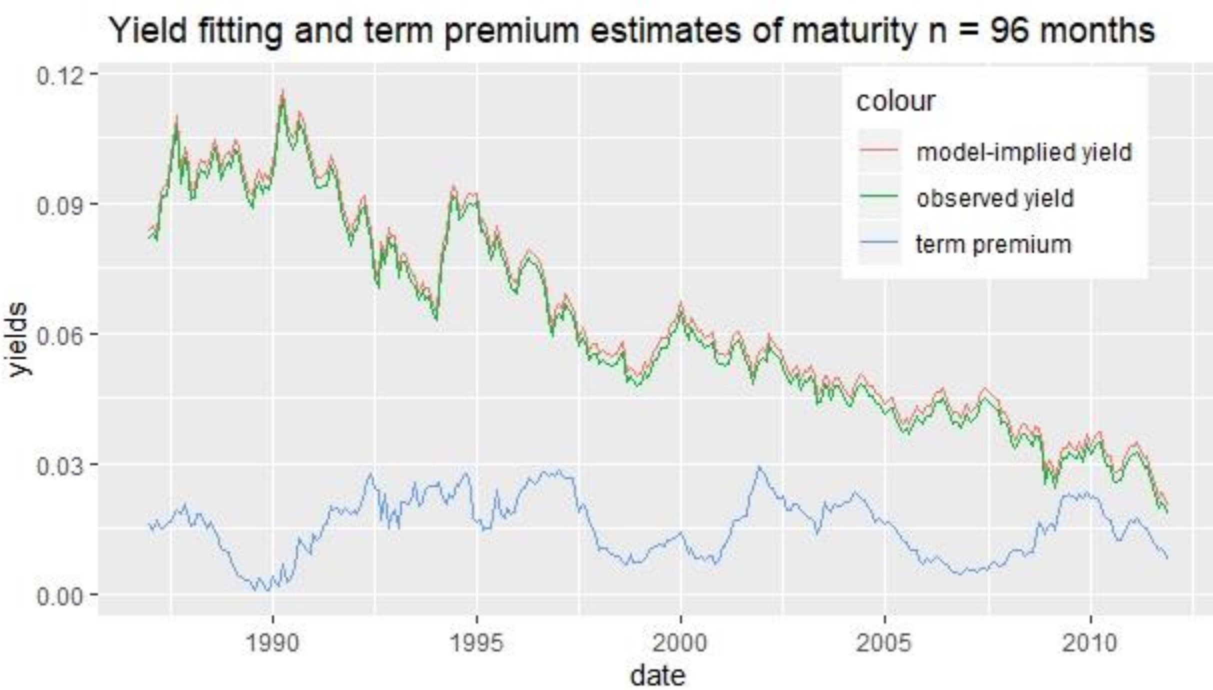

Yield fitting and term premium estimates of maturity n = 36 months
yields
date
colour
model-implied yield
observed yield
term premium
0.12
0.08
0.04
0.00
1990
1995
2000
2005
2010

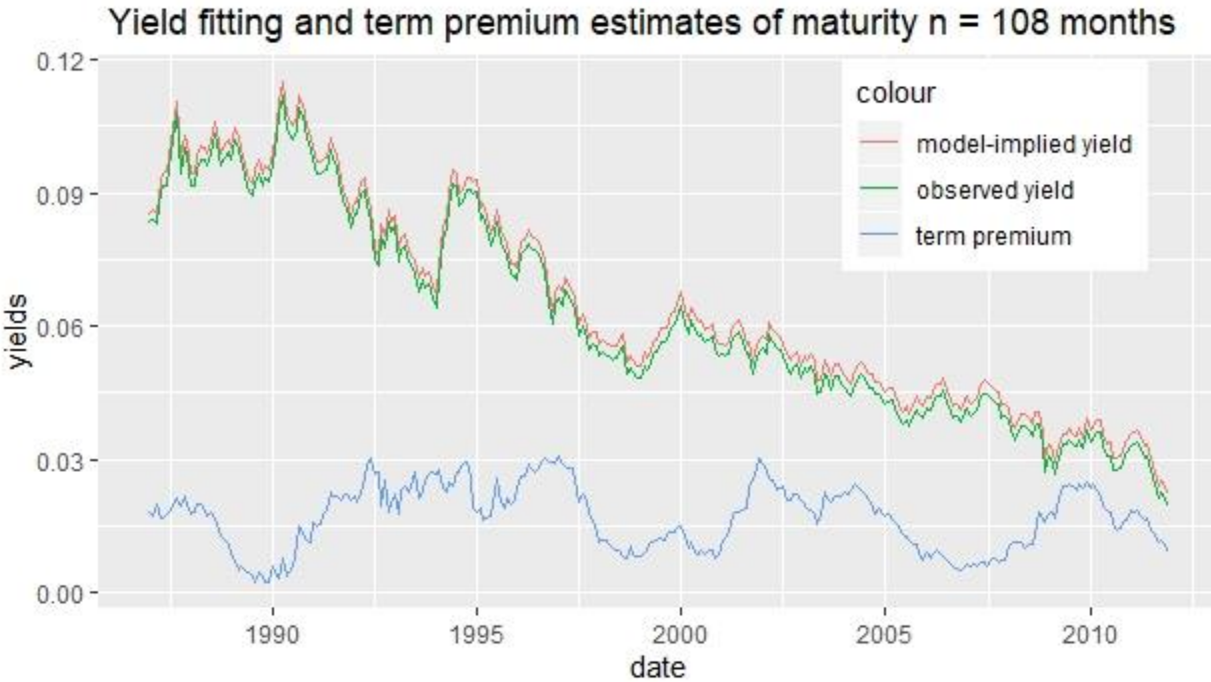

Yield fitting and term premium estimates of maturity n = 48 months
yields
date
colour
model-implied yield
observed yield
term premium
0.12
0.08
0.04
0.00
1990
1995
2000
2005
2010

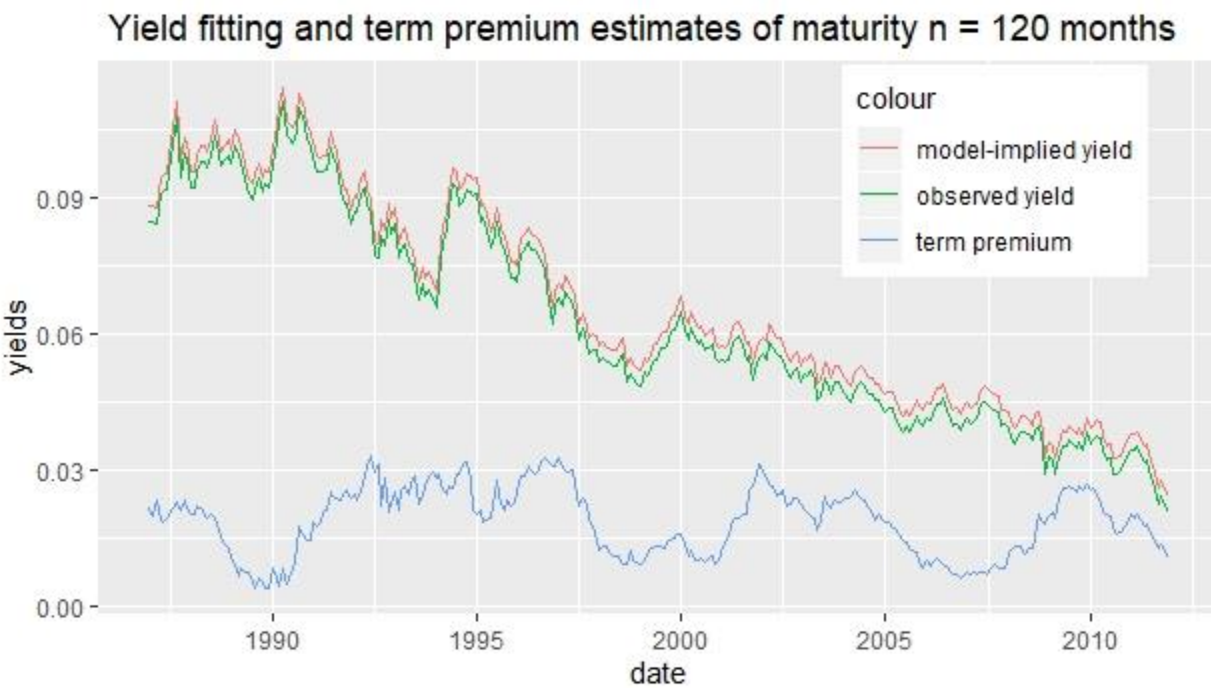

Yield fitting and term premium estimates of maturity n = 60 months
colour
model-implied yield
observed yield
term premium
yields
0.12
0.08
0.04
0.00
1990
1995
2000
2005
2010
date


Yield fitting and term premium estimates of maturity n = 72 months
colour
model-implied yield
observed yield
term premium
yields
0.12
0.08
0.04
0.00
1990
1995
2000
2005
2010
date


Yield fitting and term premium estimates of maturity n = 84 months
0.12
0.09
0.06
0.03
0.00
yields
1990
1995
2000
2005
2010
date
colour
model-implied yield
observed yield
term premium


Yield fitting and term premium estimates of maturity n = 96 months
0.12
0.09
0.06
0.03
0.00
yields
1990
1995
2000
2005
2010
date
colour
model-implied yield
observed yield
term premium


Yield fitting and term premium estimates of maturity n = 108 months
colour
model-implied yield
observed yield
term premium
yields
0.12
0.09
0.06
0.03
0.00
1990
1995
2000
2005
2010
date


Yield fitting and term premium estimates of maturity n = 120 months
colour
model-implied yield
observed yield
term premium
yields
0.09
0.06
0.03
0.00
1990
1995
2000
2005
2010
date

Appendix B: Derivation of equations of Adrian, Crump and Moech (2013) model

Tsun Yan Chan

The Adrian, Crump and Moech (2013) model[1]

Assume the vector of $K \times 1$ state variables X_t follows the vector autoregression VAR(1) process:

$$\underbrace{X_{t+1}}_{(k\times 1)} = \underbrace{\mu}_{(k\times 1)} + \underbrace{\Phi}_{(k\times k)} \underbrace{X_t}_{(k\times 1)} + \underbrace{v_{t+1}}_{(k\times 1)} \tag{1}$$

We also assume that the shocks v_{t+1} conditionally follow a Gaussian distribution with variance-covariance matrix Σ:

$$v_{t+1} | \{X_s\}_{s=0}^{t} \sim N(0, \Sigma)$$

The assumption of no-arbitrage implies that there exists a pricing kernel M_t such that

$$P_t^n = E_t[M_{t+1} P_{t+1}^{n-1}] => 1 = E_t[M_{t+1} \frac{P_{t+1}^{t-1}}{P_t^n}] \tag{2}$$

We also assume that the pricing kernel M_{t+1} is exponentially affine:

$$M_{t+1} = \mathrm{e}^{-r_t - \frac{1}{2}\lambda_t' \lambda_t - \lambda_t' \Sigma^{-\frac{1}{2}} v_{t+1}} \tag{3}$$

where r_t=$lnP_t^{(1)}$ is the continuously compounded risk-free rate.

We further assume market prices of risks

$$\underbrace{\lambda_t}_{(KX1)} = \underbrace{\underbrace{\Sigma^{-\frac{1}{2}}}_{(KXK)} (\underbrace{\lambda_0}_{(KX1)} + \underbrace{\lambda_1}_{(KXK)} \underbrace{x_t}_{(KX1)})}_{(KX1)} \tag{4}$$

We denote $rx_{t+1}^{(n-1)}$ as the log excess holding return of a bond maturing in n periods:

$$rx_{t+1}^{(n-1)} = r_{t+1}^{(n-1)} - r_t = lnP_{t+1}^{(n-1)} - lnP_t^{(n)} - r_t$$
$$rx_{t+1}^{(n-1)} + r_t = lnP_{t+1}^{(n-1)} - lnP_t^{(n)}$$
$$rx_{t+1}^{(n-1)} + r_t = ln(\frac{P_{t+1}^{(n-1)}}{P_t^{(n)}})$$

Take logs on both sides of the equation:

[1]See https://www.sciencedirect.com/science/article/abs/pii/S0304405X13001335

$$e^{rx_{t+1}^{(n-1)}+r_t} = \frac{P_{t+1}^{(n-1)}}{P_t^{(n)}} \tag{5}$$

Now we have,

$$\begin{aligned}
1 &= E_t[e^{-r_t-\frac{1}{2}\lambda_t'\lambda_t-\lambda_t'\Sigma^{-\frac{1}{2}}v_{t+1}}(e^{rx_{t+1}^{(n-1)}+r_t})] \\
1 &= E_t[e^{rx_{t+1}^{(n-1)}-\frac{1}{2}\lambda_t'\lambda_t-\lambda_t'\Sigma^{-\frac{1}{2}}v_{t+1}}] \\
1 &= e^{-\frac{1}{2}\lambda_t'\lambda_t}E_t[e^{rx_{t+1}^{(n-1)}-\lambda_t'\Sigma^{-\frac{1}{2}}v_{t+1}}]
\end{aligned} \tag{6}$$

Assume $\{rx_{t+1}^{(n-1)}$,$v_{t+1}\}$ are jointly normally distributed

$$\begin{bmatrix} rx_{t+1}^{(n-1)} \\ v_{t+1} \end{bmatrix} \sim N[\begin{bmatrix} E[rx_{t+1}^{(n-1)}] \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_t^2(rx_{t+1}^{(n-1)}) & cov(rx_{t+1}^{(n-1)}, v_{t+1}') \\ cov(rx_{t+1}^{(n-1)}, v_{t+1}) & \Sigma \end{bmatrix}]$$

The moment generating function of a jointly normal distribution is:
If X~(μ, Σ) then $E[e^{s'x}] = e^{s'\mu+\frac{1}{2}s'\Sigma s}$ From equation (5), we have

$$s' = \underbrace{\begin{bmatrix} \underbrace{1}_{(1X1)} & \underbrace{-\lambda_t'}_{(1XK)} & \underbrace{\Sigma^{-\frac{1}{2}}}_{(KXK)} \end{bmatrix}}_{1X(1XK)}$$

$$s = \begin{bmatrix} 1 \\ -\Sigma^{-\frac{1}{2}}\lambda_t \end{bmatrix}$$

Therefore,

$$1 = exp(-\frac{1}{2}\lambda_t'\lambda_t)exp(\begin{bmatrix} 1 & -\lambda_t'\Sigma^{-\frac{1}{2}} \end{bmatrix} \begin{bmatrix} E[rx_{t+1}^{(n-1)}] \\ 0 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} \sigma_t^2(rx_{t+1}^{(n-1)}) & cov(rx_{t+1}^{(n-1)}, v_{t+1}') \\ cov(rx_{t+1}^{(n-1)}, v_{t+1}) & \Sigma \end{bmatrix} t')$$

$$1 = exp(-\frac{1}{2}\lambda_t'\lambda_t)exp(E_t[rx_{t+1}^{(n-1)}])exp(\frac{1}{2}\begin{bmatrix} 1 & -\lambda_t'\Sigma^{-\frac{1}{2}} \end{bmatrix} \begin{bmatrix} \sigma_t^2(rx_{t+1}^{(n-1)}) & cov(rx_{t+1}^{(n-1)}, v_{t+1}') \\ cov(rx_{t+1}^{(n-1)}, v_{t+1}) & \Sigma \end{bmatrix} \begin{bmatrix} 1 \\ -\Sigma^{-\frac{1}{2}}\lambda_t \end{bmatrix})$$

$$1 = exp(-\frac{1}{2}\lambda_t'\lambda_t)+E_t[rx_{t+1}^{(n-1)}])exp(\frac{1}{2}\begin{bmatrix} \sigma_t^2 rx_{t+1}^{(n-1)} - \lambda_t'\Sigma^{-\frac{1}{2}}cov(rx_{t+1}^{(n-1)}, v_{t+1}') & cov(rx_{t+1}^{(n-1)}, V_{t+1}) - \lambda_t'\Sigma^{-\frac{1}{2}} \end{bmatrix} \begin{bmatrix} 1 \\ -\Sigma^{-\frac{1}{2}}\lambda_t \end{bmatrix})$$

$$1 = exp(-\frac{1}{2}\lambda_t'\lambda_t)+E_t[rx_{t+1}^{(n-1)}])exp(\frac{1}{2}(\sigma_t^2 rx_{t+1}^{(n-1)}-\lambda_t'\Sigma^{-\frac{1}{2}}cov(rx_{t+1}^{(n-1)}, v_{t+1}')-cov(rx_{t+1}^{(n-1)}, V_{t+1})\Sigma^{-\frac{1}{2}}\lambda_t+\lambda_t'\Sigma^{\frac{1}{2}}\Sigma^{-\frac{1}{2}}\lambda_t)$$

$$1 = exp(-\frac{1}{2}\lambda_t'\lambda_t + E_t[rx_{t+1}^{(n-1)}] + \frac{1}{2}\sigma_t^2 rx_{t+1}^{(n-1)} - cov(rx_{t+1}^{(n-1)}, v_{t+1}')\Sigma^{-\frac{1}{2}}\lambda_t + \frac{1}{2}\lambda_t'\lambda_t)$$

$$1 = exp(E_t[rx_{t+1}^{(n-1)}] + \frac{1}{2}var_t[rx_{t+1}^{(n-1)}] - cov[rx_{t+1}^{(n-1)}, v_{t+1}']\Sigma^{-\frac{1}{2}}\lambda_t)$$

Take logs on both sides,

$$0 = E_t[rx_{t+1}^{(n-1)}] + \frac{1}{2}var_t[rx_{t+1}^{(n-1)}] - cov[rx_{t+1}^{(n-1)}, v'_{t+1}]\Sigma^{-\frac{1}{2}}\lambda_t$$

$$\therefore E_t[rx_{t+1}^{(n-1)}] = Cov[rx_{t+1}^{(n-1)}, v'_{t+1}\Sigma^{-\frac{1}{2}}\lambda_t] - \frac{1}{2}var_t[rx_{t+1}^{(n-1)}] \tag{7}$$

Now, we denote

$$\beta_t^{(n-1)'} = Cov_t[rx_{t+1}^{(n-1)}, v'_{t+1}]\Sigma^{-1} \tag{8}$$

By using equation (4), we have

$$E_t[rx_{t+1}^{(n-1)}] = \beta_t^{(n-1)'}(\lambda_0 + \lambda_1 X_t) - \frac{1}{2}var_t[rx_{t+1}^{(n-1)}] \tag{9}$$

FACT: (From Adrian et al. (2015) equation 11)

$$rx_{t+1}^{(n-1)} - E_t[rx_{t+1}^{(n-1)}] = \gamma_t^{(n-1)'}v_{t+1} + e_{t+1}^{(n-1)} \tag{10}$$

where $e_{t+1}^{(n-1)}$ is the return pricing errors.
We want to show that $\gamma_t^{(n-1)} = \beta_t^{(n-1)}$:
Since

$$Cov(x,y) = E[x-\mu_x][y-\mu_y]$$

From equation (8):

$$\begin{aligned} \beta_t^{(n-1)'} &= cov_t[rx_{t+1}^{(n-1)}, v'_{t+1}]\Sigma^{-1} \\ &= E_t[rx_{t+1}^{(n-1)} - E_t[rx_{t+1}^{(n-1)}]][v'_{t+1} - 0]\Sigma^{-1} \end{aligned} \tag{11}$$

Substitute (10) into (11),

$$\begin{aligned} \beta_t^{(n-1)'} &= E_t[\gamma_t^{(n-1)'}v_{t+1} + e_{t+1}^{(n-1)}][v'_{t+1}]\Sigma^{-1} \\ &= (\gamma_t^{(n-1)'}\underbrace{E_t[v_{t+1}v'_{t+1}]}_{\Sigma} + \underbrace{E_t[e_{t+1}^{(n-1)}v'_{t+1}])}_{0}\Sigma^{-1} \\ &= \gamma_t^{(n-1)'}\Sigma\Sigma^{-1} \\ \therefore \beta_t^{(n-1)'} &= \gamma_t^{(n-1)'} \end{aligned}$$

Next, we find that

$$var[rx_{t+1}^{(n-1)}]$$

$$var[rx_{t+1}^{(n-1)}] = E[(rx_{t+1}^{(n-1)} - E[rx_{t+1}^{(n-1)}])^2] \tag{12}$$

Substitute equation (11) into (12)

$$
\begin{aligned}
var[rx_{t+1}^{(n-1)}] &= E_t[(\gamma_t^{(n-1)'} v_{t+1} + e_{t+1}^{(n-1)})^2] \\
&= E_t[\gamma_t^{(n-1)'} v_{t+1} v_{t+1}' \gamma_t^(n-1) + 2\gamma_t^{(n-1)'} v_{t+1} e_{t+1}^{(n-1)} + (e_{t+1}^{(n-1)})^2] \\
&= \gamma_t^{(n-1)'} \underbrace{E_t[v_{t+1} v_{t+1}']}_{\sigma} \gamma_t^{(n-1)} + 2\gamma_t^{(n-1)'} \underbrace{E_t[v_{t+1} e_{t+1}^{(n-1)}]}_{0} + \underbrace{E_t[(e_{t+1}^{(n-1)})^2]}_{\sigma}{}^{2} \\
&= \gamma_t^{(n-1)'} \Sigma \gamma_t^{(n-1)} + \sigma^2
\end{aligned}
$$

$$
\therefore var[rx_{t+1}^{(n-1)}] = \gamma_t^{(n-1)'} \Sigma \gamma_t^{(n-1)} + \sigma^2 \tag{13}
$$

Substitute equation (12) into equation (9)

$$
\begin{aligned}
E_t[rx_{t+1}^{(n-1)}] &= \beta_t^{(n-1)'} (\lambda_0 + \lambda_1 x_t) - \tfrac{1}{2}[\gamma_t^{(n-1)'} \Sigma \gamma_t^{(n-1)} + \sigma^2] \\
&= \beta_t^{(n-1)'} (\lambda_0 + \lambda_1 x_t) - \tfrac{1}{2}[\beta_t^{(n-1)'} \Sigma \beta_t^{(n-1)} + \sigma^2]
\end{aligned} \tag{14}
$$

Substitute equation (13) into equation (10)

$$
\begin{aligned}
rx_{t+1}^{(n-1)} &= \beta_t^{(n-1)'} (\lambda_0 + \lambda_1 x_t) - \tfrac{1}{2}[\beta_t^{(n-1)'} \Sigma \beta_t^{(n-1)} + \sigma^2] + \beta_t^{(n-1)'} v_{t+1} + e_{t+1}^{(n-1)} \\
&= \beta_t^{(n-1)'} \beta_0 + \beta_t (\lambda_1 x_t) - \tfrac{1}{2}[\beta_t^{(n-1)'} \Sigma \beta_t^{(n-1)}] - \tfrac{1}{2}\sigma^2 + \beta_t^{(n-1)'} v_{t+1} + e_{t+1}^{(n-1)}
\end{aligned}
$$

Now, we have

$$
\underbrace{rx_{t+1}^{(n-1)}}_{(1X1)} = \underbrace{\underbrace{\beta_t^{(n-1)'}}_{(1XK)} \underbrace{\lambda_0}_{(KX1)}}_{(1X1)} + \underbrace{\underbrace{\beta_t}_{(1XK)} (\underbrace{\lambda_1}_{(KXK)} \underbrace{x_t}_{(KX1)})}_{(1X1)} - \frac{1}{2} [\underbrace{\underbrace{\beta_t^{(n-1)'}}_{(1XK)} \underbrace{\Sigma}_{(KXK)} \underbrace{\beta_t^{(n-1)}}_{(1X1)}}_{(1X1)} + \underbrace{\sigma^2}_{(1X1)}] + \underbrace{\underbrace{\beta_t^{(n-1)'}}_{(1XK)} \underbrace{v_{t+1}}_{(KX1)}}_{(1X1)} + \underbrace{e_{t+1}}_{(1X1)} \tag{15}
$$

By putting all T observations of bond (n-1) together

$$
\begin{bmatrix} rx_1^{(n-1)} & rx_2^{(n-1)} & \cdots & rx_T^{(n-1)} \end{bmatrix} = (\beta^{(n-1)'} \lambda_0 - \frac{1}{2}(\beta^{(n-1)'} \Sigma \beta^{(n-1)} + \sigma^2)) \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix} +
$$
$$
\beta^{(n-1)'} \lambda_1 \begin{bmatrix} x_0 & x_1 & \cdots & x_{T-1} \end{bmatrix} + \beta^{(n-1)'} \begin{bmatrix} V_1 & V_2 & \cdots & V_T \end{bmatrix} + \begin{bmatrix} e_1^{(n-1)} & e_2^{(n-1)} & \cdots & e_T^{(n-1)} \end{bmatrix}
$$

Let

$$
X\_ = \begin{bmatrix} x_0 & x_1 & \cdots & x_{T-1} \end{bmatrix}
$$
$$
V = \begin{bmatrix} V_1 & V_2 & \cdots & V_T \end{bmatrix}
$$
$$
E = \begin{bmatrix} e_1^{(n-1)} & e_2^{(n-1)} & \cdots & e_T^{(n-1)} \end{bmatrix}
$$

By substituting $X\_$, V and E into the equation,

$$\begin{bmatrix} rx_1^{(n-1)} & rx_2^{(n-1)} & \cdots & rx_T^{(n-1)} \end{bmatrix} = (\beta^{(n-1)'}\lambda_0 - \frac{1}{2}(\beta^{(n-1)'}\Sigma\beta^{(n-1)} + \sigma^2))\iota_T' + \beta^{(n-1)'}\lambda_1 X_- + \beta^{(n-1)'}V + E \quad (16)$$

Next, we stack all N assets

$$\begin{bmatrix} rx^{(1)} \\ rx^{(2)} \\ \vdots \\ rx^{(n-1)} \\ \vdots \\ rx^{(N)} \end{bmatrix} = \begin{bmatrix} \beta^{(1)'}\lambda_0 - \frac{1}{2}(\beta^{(1)'}\Sigma\beta^{(1)} + \sigma^2) \\ \beta^{(2)'}\lambda_0 - \frac{1}{2}(\beta^{(2)'}\Sigma\beta^{(2)} + \sigma^2) \\ \vdots \\ \beta^{(n-1)'}\lambda_0 - \frac{1}{2}(\beta^{(n-1)'}\Sigma\beta^{(n-1)} + \sigma^2) \\ \vdots \\ \beta^{(N)'}\lambda_0 - \frac{1}{2}(\beta^{(N)'}\Sigma\beta^{(N)} + \sigma^2) \end{bmatrix} \iota_T' + \begin{bmatrix} \beta^{(1)'}\lambda_1 \\ \beta^{(2)'}\lambda_1 \\ \vdots \\ \beta^{(n-1)'}\lambda_1 \\ \vdots \\ \beta^{(N)'}\lambda_1 \end{bmatrix} X_- + \begin{bmatrix} \beta^{(1)'} \\ \beta^{(2)'} \\ \vdots \\ \beta^{(n-1)'} \\ \vdots \\ \beta^{(N)'} \end{bmatrix} V + E \quad (17)$$

We define

$$a = \begin{bmatrix} \beta^{(1)'}\lambda_0 - \frac{1}{2}(\beta^{(1)'}\Sigma\beta^{(1)} + \sigma^2) \\ \beta^{(2)'}\lambda_0 - \frac{1}{2}(\beta^{(2)'}\Sigma\beta^{(2)} + \sigma^2) \\ \vdots \\ \beta^{(n-1)'}\lambda_0 - \frac{1}{2}(\beta^{(n-1)'}\Sigma\beta^{(n-1)} + \sigma^2) \\ \vdots \\ \beta^{(N)'}\lambda_0 - \frac{1}{2}(\beta^{(N)'}\Sigma\beta^{(N)} + \sigma^2) \end{bmatrix}$$

$$c = \begin{bmatrix} \beta^{(1)'}\lambda_1 \\ \beta^{(2)'}\lambda_1 \\ \vdots \\ \beta^{(n-1)'}\lambda_1 \\ \vdots \\ \beta^{(N)'}\lambda_1 \end{bmatrix}$$

We have

$$a = \begin{bmatrix} a^{(1)} \\ a^{(2)} \\ \vdots \\ a^{(n-1)} \\ \vdots \\ a^{(N)} \end{bmatrix} = \begin{bmatrix} \beta^{(1)'}\lambda_0 - \frac{1}{2}(\beta^{(1)'}\Sigma\beta^{(1)} + \sigma^2) \\ \beta^{(2)'}\lambda_0 - \frac{1}{2}(\beta^{(2)'}\Sigma\beta^{(2)} + \sigma^2) \\ \vdots \\ \beta^{(n-1)'}\lambda_0 - \frac{1}{2}(\beta^{(n-1)'}\Sigma\beta^{(n-1)} + \sigma^2) \\ \vdots \\ \beta^{(N)'}\lambda_0 - \frac{1}{2}(\beta^{(N)'}\Sigma\beta^{(N)} + \sigma^2) \end{bmatrix} = \begin{bmatrix} \beta^{(1)'}\lambda_0 \\ \beta^{(2)'}\lambda_0 \\ \vdots \\ \beta^{(n-1)'}\lambda_0 \\ \vdots \\ \beta^{(N)'}\lambda_0 \end{bmatrix} - \frac{1}{2}(\begin{bmatrix} \beta^{(1)'}\Sigma\beta^{(1)} \\ \beta^{(2)'}\Sigma\beta^{(2)} \\ \vdots \\ \beta^{(n-1)'}\Sigma\beta^{(n-1)} \\ \vdots \\ \beta^{(N)'}\Sigma\beta^{(N)} \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ \vdots \\ 1 \end{bmatrix})\sigma^2 \quad (18)$$

Next, we need to show

$$\begin{bmatrix} \beta^{(1)'}\Sigma\beta^{(1)} \\ \beta^{(2)'}\Sigma\beta^{(2)} \\ \vdots \\ \beta^{(n-1)'}\Sigma\beta^{(n-1)} \\ \vdots \\ \beta^{(N)'}\Sigma\beta^{(N)} \end{bmatrix} = B^* vec(\Sigma)$$

where

$$\beta^* = \begin{bmatrix} vec(\beta^{(1)}\beta^{(1)'}) & \cdots & vec(\beta^{(N)}\beta^{(N)'}) \end{bmatrix}'$$

Since $\beta^{(n)'}\Sigma\beta^{(n)}$ is a (1X1) matrix, therefore $\beta^{(n)'}\Sigma\beta^{(n)} = vec(\beta^{(n)'}\Sigma\beta^{(n)})$
Property 1:

$$vec(ABC) = (C' \otimes A)vec(B)$$

By property 1, we thus have

$$\beta^{(n)'}\Sigma\beta^{(n)} = vec(\beta^{(n)'}\Sigma\beta^{(n)}) = \underbrace{(\underbrace{\beta^{(n)'}}_{(1XK)} \otimes \underbrace{\beta^{(n)'}}_{(1XK)})\underbrace{vec(\Sigma)}_{(K^2X1)}}_{(1X1)}$$

And so,

$$\begin{bmatrix} \beta^{(1)'}\Sigma\beta^{(1)} \\ \beta^{(2)'}\Sigma\beta^{(2)} \\ \vdots \\ \beta^{(n-1)'}\Sigma\beta^{(n-1)} \\ \vdots \\ \beta^{(N)'}\Sigma\beta^{(N)} \end{bmatrix} = \begin{bmatrix} (\beta^{(1)'} \otimes \beta^{(1)'}) \\ (\beta^{(2)'} \otimes \beta^{(2)'}) \\ \vdots \\ (\beta^{(n-1)'} \otimes \beta^{(n-1)'}) \\ \vdots \\ (\beta^{(N)'} \otimes \beta^{(N)'}) \end{bmatrix} vec(\Sigma) \tag{19}$$

Property 2:

$$(A \otimes B)' = A' \otimes B'$$

$$\therefore \begin{bmatrix} (\beta^{(1)'} \otimes \beta^{(1)'}) \\ (\beta^{(2)'} \otimes \beta^{(2)'}) \\ \vdots \\ (\beta^{(n-1)'} \otimes \beta^{(n-1)'}) \\ \vdots \\ (\beta^{(N)'} \otimes \beta^{(N)'}) \end{bmatrix} = \begin{bmatrix} (\beta^{(1)} \otimes \beta^{(1)})' \\ (\beta^{(2)} \otimes \beta^{(2)})' \\ \vdots \\ (\beta^{(n-1)} \otimes \beta^{(n-1)})' \\ \vdots \\ (\beta^{(N)} \otimes \beta^{(N)})' \end{bmatrix}$$

Property 3:

$$vec(cd') = d \otimes c$$

By property 3, we thus have:

$$\begin{bmatrix} (\beta^{(1)} \otimes \beta^{(1)})' \\ (\beta^{(2)} \otimes \beta^{(2)})' \\ \vdots \\ (\beta^{(n-1)} \otimes \beta^{(n-1)})' \\ \vdots \\ (\beta^{(N)} \otimes \beta^{(N)})' \end{bmatrix} = \begin{bmatrix} vec(\beta^{(1)}\beta^{(1)'})' \\ vec(\beta^{(2)}\beta^{(2)'})' \\ \vdots \\ vec(\beta^{(n-1)}\beta^{(n-1)'})' \\ \vdots \\ vec(\beta^{(N)}\beta^{(N)'})' \end{bmatrix}$$

Therefore, substitute it back to equation (19)
LHS:

$$\begin{bmatrix} \beta^{(1)'}\Sigma\beta^{(1)} \\ \beta^{(2)'}\Sigma\beta^{(2)} \\ \vdots \\ \beta^{(n-1)'}\Sigma\beta^{(n-1)} \\ \vdots \\ \beta^{(N)'}\Sigma\beta^{(N)} \end{bmatrix} = \begin{bmatrix} (\beta^{(1)} \otimes \beta^{(1)})' \\ (\beta^{(2)} \otimes \beta^{(2)})' \\ \vdots \\ (\beta^{(n-1)} \otimes \beta^{(n-1)})' \\ \vdots \\ (\beta^{(N)} \otimes \beta^{(N)})' \end{bmatrix} vec(\Sigma) = \begin{bmatrix} vec(\beta^{(1)}\beta^{(1)'})' \\ vec(\beta^{(2)}\beta^{(2)'})' \\ \vdots \\ vec(\beta^{(n-1)}\beta^{(n-1)'})' \\ \vdots \\ vec(\beta^{(N)}\beta^{(N)'})' \end{bmatrix} vec(\Sigma) \tag{20}$$

RHS:

$$B^*vec(\Sigma) = \left[vec(\beta^{(1)}\beta^{(1)'}) \quad \cdots \quad vec(\beta^{(N)}\beta^{(N)'})\right]' vec(\Sigma) = \begin{bmatrix} vec(\beta^{(1)}\beta^{(1)'})' \\ vec(\beta^{(2)}\beta^{(2)'})' \\ \vdots \\ vec(\beta^{(n-1)}\beta^{(n-1)'})' \\ \vdots \\ vec(\beta^{(N)}\beta^{(N)'})' \end{bmatrix} vec(\Sigma)$$

$\therefore$ LHS = RHS

$$\therefore \begin{bmatrix} \beta^{(1)'}\Sigma\beta^{(1)} \\ \beta^{(2)'}\Sigma\beta^{(2)} \\ \vdots \\ \beta^{(n-1)'}\Sigma\beta^{(n-1)} \\ \vdots \\ \beta^{(N)'}\Sigma\beta^{(N)} \end{bmatrix} = B^*vec(\Sigma)$$

$$\therefore a = \beta'\lambda_0 - \frac{1}{2}[B^*vec(\Sigma) + \iota_N\sigma] \tag{21}$$

Substitute a into equation (17)

$$\begin{bmatrix} rx^{(1)} \\ rx^{(2)} \\ \vdots \\ rx^{(n-1)} \\ \vdots \\ rx^{(N)} \end{bmatrix} = a\iota_t' + \begin{bmatrix} \beta^{(1)'}\lambda_1 \\ \beta^{(2)'}\lambda_1 \\ \vdots \\ \beta^{(n-1)'}\lambda_1 \\ \vdots \\ \beta^{(N)'}\lambda_1 \end{bmatrix} X_{-} + \begin{bmatrix} \beta^{(1)'} \\ \beta^{(2)'} \\ \vdots \\ \beta^{(n-1)'} \\ \vdots \\ \beta^{(N)'} \end{bmatrix} V + E$$

Let

$$\beta'\lambda_1 = a\iota_t' + \begin{bmatrix} \beta^{(1)'}\lambda_1 \\ \beta^{(2)'}\lambda_1 \\ \vdots \\ \beta^{(n-1)'}\lambda_1 \\ \vdots \\ \beta^{(N)'}\lambda_1 \end{bmatrix}, \ \beta' = \begin{bmatrix} \beta^{(1)'} \\ \beta^{(2)'} \\ \vdots \\ \beta^{(n-1)'} \\ \vdots \\ \beta^{(N)'} \end{bmatrix} \ and \ rx = \begin{bmatrix} rx^{(1)} \\ rx^{(2)} \\ \vdots \\ rx^{(n-1)} \\ \vdots \\ rx^{(N)} \end{bmatrix}$$

$$rx = (\beta'\lambda_0 - \frac{1}{2}(B^*vec(\Sigma) + \iota_N\sigma^2))\iota_T' + \beta'\lambda_1 X\_ + \beta'V + E$$

$$rx = \beta'(\lambda_0\iota_T' + \lambda_1 X\_) - \frac{1}{2}(\beta^*vec(\Sigma) + \iota_N\sigma^2)\iota_T' + \beta'V + E \tag{22}$$

$$rx = (\beta'\lambda_0 - \frac{1}{2}(\beta^*vec(\Sigma) + \iota_N\sigma^2))\iota_T' + \beta'\lambda_1 X\_ + \beta'V + E \tag{23}$$

$$\therefore rx = a\iota_T' + \beta'V + cX\_ + E \tag{24}$$

$$\therefore rx = \begin{bmatrix} a & \beta' & c \end{bmatrix} \begin{bmatrix} \iota_T' \\ V \\ X\_ \end{bmatrix} + E$$

Let $\hat{z} = \begin{bmatrix} \iota_T & \hat{V}' & X\_' \end{bmatrix}' = \begin{bmatrix} \iota_T' \\ \hat{V} \\ X\_ \end{bmatrix}$

Substitute $\hat{z}$ into equation (23)

$$rx = \begin{bmatrix} a & \beta' & c \end{bmatrix} \hat{Z} + E$$

Let $\delta = \begin{bmatrix} a & \beta' & c \end{bmatrix}$

$$\therefore rx = \delta\hat{z} + E$$
$$E = rx - \delta\hat{Z} \tag{25}$$

Criterion: For single OLS estimation, we have to find $\hat{\beta}$ that minimize the sum of squared residuals. Therefore we take the derivate with espect to $\hat{\beta}$ For multiple OLS estimation, we have to find the parameters that minimize the trace(EE')

$$\begin{aligned} tr(EE') = & \quad tr((rx - \delta\hat{Z})(rx - \delta\hat{Z})) \\ = & \quad tr(rx(rx' - \hat{Z}'\delta') - \delta\hat{Z}(rx' - \hat{Z}'\delta')) \\ = & \quad tr(rxrx' - rx\hat{Z}'\delta' - \delta\hat{Z}rx' + \delta\hat{Z}\hat{Z}'\delta') \end{aligned}$$

Property 1

$$tr(A + B + C + D) = tr(A) + tr(B) + tr(C) + tr(D)$$

$$\therefore tr(EE') = \quad tr(rxrx') - tr(rx\hat{Z}'\delta') - tr(\delta\hat{Z}'rx') + tr(\delta\hat{Z}\hat{Z}'\delta')$$

Property 2

$$tr(A) = tr(A') \;\; \therefore tr(rx\hat{Z}'\delta') = tr(\delta\hat{Z}rx')$$

$$\begin{aligned}
\therefore tr(EE') = & \quad tr(rxrx') - tr((rx\hat{Z}'\delta')') - tr(\delta\hat{Z}'rx') + tr(\delta\hat{Z}\hat{Z}'\delta') \\
= & \quad tr(rxrx') - tr(\delta\hat{Z}rx') - tr(\delta\hat{Z}rx') + tr(\delta\hat{Z}\hat{Z}'\delta') \\
= & \quad tr(rxrx') - 2tr(\delta\hat{Z}rx') + tr(\delta\hat{Z}\hat{Z}'\delta') \\
= & \quad \underbrace{tr(rxrx')}_{a1} - \underbrace{2tr(\delta\hat{Z}rx')}_{a2} + \underbrace{tr(\delta(\hat{Z}\hat{Z}')\delta')}_{a3}
\end{aligned} \tag{26}$$

We can find δ^* by taking the first derivative (i.e. $\frac{\partial tr(EE')}{\partial \delta} = 0$)

First, we take FOC of part a1:

$$\frac{\partial tr(rxrx')}{\partial \delta} = 0$$

$\therefore$ FOC of a1 is 0.

Next,we need to take Foc of part a2: By using property 3: $\frac{\partial tr(AB)}{\partial A} = B'$ We have

$$\frac{\partial tr(\delta\hat{Z}rx')}{\partial \delta} = rx\hat{Z}'$$

$\therefore$ FOC of part a2 is $-2rx\hat{Z}$

Lastly, we have to take FOC of part a3: By using property 4: $\frac{\partial tr(CDC')}{\partial C} = CD' + CD$

$$\therefore \frac{\partial tr(\delta(\hat{Z}\hat{Z})'\delta')}{\partial \delta} = \delta(\hat{Z}\hat{Z}') + \delta(\hat{Z}\hat{Z}') = 2\delta(\hat{Z}\hat{Z}')$$

We have,

$$\frac{\partial tr(EE')}{\partial \delta} = -2rx\hat{Z} + 2\delta(\hat{Z}\hat{Z}') = 0rx\hat{Z}' = \delta(\hat{Z}\hat{Z}')$$

Premultiplying both sides by $(\hat{Z}\hat{Z}')^{-1}$

$$rx\hat{Z}'(\hat{Z}\hat{Z}')^{-1} = \delta(\hat{Z}\hat{Z}')(\hat{Z}\hat{Z}')^{-1}\delta = rx\hat{Z}'(\hat{Z}\hat{Z}')^{-1}$$

$$\therefore \begin{bmatrix} a & \beta' & c \end{bmatrix} = rx\hat{Z}'(\hat{Z}\hat{Z}')^{-1} \tag{27}$$

From equation (21), we have

$$a = \beta'\lambda_0 - \frac{1}{2}[B^* vec(\Sigma) + \iota_N \sigma^2]$$

$$\therefore \underbrace{\hat{a}}_{(NX1)} = \underbrace{\underbrace{\hat{\beta}'}_{(NXK)} \underbrace{\lambda_0}_{(KX1)}}_{(NX1)} - \frac{1}{2}(\underbrace{\underbrace{\hat{B}^*}_{(NXK^2)} \underbrace{vec(\Sigma)}_{(K^2X1)}}_{(NX1)} + \underbrace{\iota_N \sigma^2}_{(NX1)})$$

$$\hat{a} + \frac{1}{2}(B^* \hat{vec}(\Sigma) + \iota_N \sigma^2) = \hat{\beta}'\lambda_0$$

We multiply both sides by $\hat{\beta}$, we get:

$$\hat{\beta}[\hat{a} + \frac{1}{2}(\hat{B}^* vec(\Sigma) + \iota_N \sigma^2)] = \hat{\beta}\hat{\beta}'\lambda_0$$

We pre-multiply both sides by $(\hat{\beta}\hat{\beta}')^{-1}$, we get:

$$(\hat{\beta}\hat{\beta}')^{-1}\hat{\beta}[\hat{a}+\frac{1}{2}(\hat{B^{*}}vec(\Sigma)+\iota_N\sigma^2)]=(\hat{\beta}\hat{\beta}')^{-1}(\hat{\beta}\hat{\beta}')\lambda_0$$

$$(\hat{\beta}\hat{\beta}')^{-1}\hat{\beta}(\hat{a}+\frac{1}{2}(B^{*}vec(\hat{\Sigma})+\hat{\sigma^2}\iota_N))=I_k\hat{\lambda_0}$$

$$\therefore \hat{\lambda_0}=(\hat{\beta}\hat{\beta}')^{-1}\hat{\beta}(\hat{a}+\frac{1}{2}(\hat{B^{*}}vec(\hat{\Sigma})+\hat{\sigma^2}\iota_N)) \tag{28}$$

Since we have $\hat{c}=\hat{\beta}'\lambda_1$

$$\therefore \hat{c}=\hat{\beta}'\hat{\lambda_1}$$

We multiply both sides by $\hat{\beta}$

$$\hat{\beta}\hat{c}=\hat{\beta}\hat{\beta}'\hat{\lambda_1}$$

We pre-multiply both sides by $(\hat{\beta}\hat{\beta}')^{-1}$

$$\begin{aligned}(\hat{\beta}\hat{\beta}')^{-1}\hat{\beta}\hat{c}= &\quad (\hat{\beta}\hat{\beta}')^{-1}(\hat{\beta}\hat{\beta}')\hat{\lambda_1}\\ \hat{\lambda_1}= &\quad (\hat{\beta}\hat{\beta}')^{-1}\hat{\beta}\hat{c}\end{aligned} \tag{29}$$

Inference

$$\hat{\lambda_0}=(\hat{\beta}\hat{\beta}')^{-1}\hat{\beta}(\hat{a}+\frac{1}{2}(\hat{B^{*}}vec(\hat{\Sigma})+\hat{\sigma^2}\iota_N))$$

$$\hat{\lambda_1}=(\hat{\beta}\hat{\beta}')^{-1}\hat{\beta}\hat{c}$$

Let us denote

$$\Lambda=\begin{bmatrix}\lambda_0 & \lambda_1\end{bmatrix}$$

$$\begin{aligned}\Lambda= &\quad \begin{bmatrix}\lambda_0 & \lambda_1\end{bmatrix}\\ = &\quad \begin{bmatrix}(\hat{\beta}\hat{\beta}')^{-1}\hat{\beta}(\hat{a}+\tfrac{1}{2}(\hat{B^{*}}vec(\hat{\Sigma})+\hat{\sigma^2}\iota_N)) & (\hat{\beta}\hat{\beta}')^{-1}\hat{\beta}\hat{c}\end{bmatrix}\\ = &\quad (\hat{\beta}\hat{\beta}')^{-1}\hat{\beta}\begin{bmatrix}\hat{a}+\tfrac{1}{2}(\hat{B^{*}}vec(\hat{\Sigma})+\hat{\sigma^2}\iota_N) & \hat{c}\end{bmatrix}\end{aligned} \tag{30}$$

From equation (14), we have $rx=a\iota_T'+\beta'\hat{V}+cX\_+E$

$$rx=a\iota_T'+cX\_+\beta'\hat{V}+E$$

$$rx=\underbrace{\begin{bmatrix}a & c & \beta'\end{bmatrix}}_{\alpha}\underbrace{\begin{bmatrix}\iota_T'\\ X\_\\ \hat{V}\end{bmatrix}}_{W}+E=\alpha W+E$$

Let

$$\alpha=\begin{bmatrix}a & c & \beta'\end{bmatrix},W=\begin{bmatrix}\iota_T'\\ X\_\\ \hat{V}\end{bmatrix}$$

Then, the OLS estimator for α is $\hat{\alpha} = \begin{bmatrix}\hat{a} & \hat{c} & \hat{\beta}'\end{bmatrix} = rxW'(WW')^{-1}$

Let

$$Z\_ = \begin{bmatrix}\iota_T & X\_'\end{bmatrix}' = \begin{bmatrix}\iota_T \\ X\_'\end{bmatrix}$$

then,

$$W = \begin{bmatrix}Z\_ \\ \hat{V}\end{bmatrix}$$

Since $W = \begin{bmatrix}\iota_T' \\ X\_ \\ \hat{V}\end{bmatrix}$. Therefore $W = \begin{bmatrix}\iota_T' \\ X\_ \\ \hat{V}\end{bmatrix} = \begin{bmatrix}Z\_ \\ \hat{V}\end{bmatrix}$

Then

$$WW' = \begin{bmatrix}Z\_ \\ \hat{V}\end{bmatrix}\begin{bmatrix}Z\_' & \hat{V}'\end{bmatrix} = \begin{bmatrix}Z\_Z\_' & Z\_\hat{V}' \\ \hat{V}Z\_' & \hat{V}\hat{V}'\end{bmatrix}$$

By partitioned matrix we need $(WW')^{-1}$. By the ? of the partitioned matrix inverse, we have:

$$A = \begin{bmatrix}A_{11} & A_{12} \\ A_{21} & A_{22}\end{bmatrix}$$

$$A^{-1} = \begin{bmatrix}D & -DA_{12}A_{22}^{-1} \\ -A_{22}^{-1}A_{21}D & A_{22}^{-1} + A_{22}^{-1}A_{21}DA_{12}A_{22}^{-1}\end{bmatrix}$$

where

$$D = (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1}$$

Then, we have that

$$WW' = \begin{bmatrix}Z\_Z\_' & Z\_\hat{V}' \\ \hat{V}Z\_' & \hat{V}\hat{V}'\end{bmatrix}$$

$$D = (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1} = (Z\_Z\_' - Z\_\hat{V}'(\hat{V}\hat{V}')^{-1}\hat{V}Z\_')^{-1} = (Z\_(I_T - \hat{V}'(\hat{V}\hat{V}')^{-1}\hat{V})Z\_')^{-1}$$

Let

$$M_{\hat{V}} = I_T - \hat{V}'(\hat{V}\hat{V}')^{-1}\hat{V}$$

$$\therefore D = (Z\_M_{\hat{V}}Z\_')^{-1}$$

$$-A_{22}^{-1}A_{21}D = -(\hat{V}\hat{V}')^{-1}\hat{V}Z\_'D$$

$$\begin{aligned}
\begin{bmatrix}a & c\end{bmatrix} &= rx\begin{bmatrix}Z_' & \hat{V}'\end{bmatrix}\begin{bmatrix}D \\ -A_{22}^{-1}A_{21}D\end{bmatrix} \\
&= rx\begin{bmatrix}Z_' & \hat{V}'\end{bmatrix}\begin{bmatrix}D \\ -(\hat{V}\hat{V}')^{-1}\hat{V}Z_'D\end{bmatrix} && (31) \\
&= rxZ_'D - rx\hat{V}'(\hat{V}\hat{V}')^{-1}\hat{V}Z_'D && (32) \\
&= rx(I_T - \hat{V}'(\hat{V}\hat{V}')^{-1}\hat{V})Z_'D && (33)
\end{aligned}$$

Since

$$M_{\hat{V}} = I_T - \hat{V}'(\hat{V}\hat{V}')^{-1}\hat{V}$$

$$\begin{bmatrix}\hat{a} & \hat{c}\end{bmatrix} = rx(I_T - \hat{V}'(\hat{V}\hat{V})) = rxM_{\hat{v}}Z\_'D$$

$$\therefore \begin{bmatrix}\underbrace{\hat{a}}_{(NX1)} & \underbrace{\hat{c}}_{(NXK)}\end{bmatrix} = \underbrace{\underbrace{rx}_{(NXT)}\underbrace{M_{\hat{v}}}_{(TXT)}\underbrace{Z\_'}_{(TX(1+K))}}_{(NX(1+K))}\underbrace{(\underbrace{Z\_}_{(1+k)xt}\underbrace{M_{\hat{v}}}_{(TXT)}\underbrace{Z\_'}_{(TX(1+K))})^{-1}}_{(1+K)(1+K)} \tag{34}$$

$$\Lambda = (\hat{\beta}\hat{\beta}')^{-1}\hat{\beta}\begin{bmatrix}\hat{a} + \frac{1}{2}(\hat{B^*}vec(\hat{\Sigma}) + \hat{\sigma^2}\iota_N) & \hat{c}\end{bmatrix} = (\hat{\beta}\hat{\beta}')^{-1}\hat{\beta}[\begin{bmatrix}\hat{a} & \hat{c}\end{bmatrix} + \begin{bmatrix}\frac{1}{2}(\hat{B^*}vec(\Sigma) + \hat{\sigma^2}\iota_N) & 0\end{bmatrix}]$$

$$\because \begin{bmatrix}\frac{1}{2}(\hat{B^*}vec(\Sigma) + \hat{\sigma^2}\iota_N) & 0\end{bmatrix} = \frac{1}{2}(\hat{B}^*vec(\hat{\Sigma}) + \sigma^2\iota_N)\begin{bmatrix}1 & 0\end{bmatrix} \therefore \Lambda = (\hat{\beta}\hat{\beta}')^{-1}\hat{\beta}[\begin{bmatrix}\hat{a} & \hat{c}\end{bmatrix} + \frac{1}{2}(\hat{B}^*vec(\hat{\Sigma} + \sigma^2\iota_N)\begin{bmatrix}1 & 0\end{bmatrix}]$$

Substitute $\begin{bmatrix}\hat{a} & \hat{c}\end{bmatrix} = rxM_{\hat{V}}Z\_'(Z\_M_{\hat{V}}Z\_')^{-1}$

$$\hat{\Lambda} = (\hat{\beta}\hat{\beta}')^{-1}\hat{\beta}(rxM_{\hat{V}}Z\_' + \frac{1}{2}(\hat{B}^*vec(\hat{\Sigma}) + \sigma^2\iota_N)\begin{bmatrix}1 & 0\end{bmatrix})$$

CLAIM: $\begin{bmatrix}1 & 0\end{bmatrix} = \iota_T' M_{\hat{V}}Z\_'(Z\_M_{\hat{V}}Z\_')^{-1} = \varrho_1'$

Now, we will prove that $\iota_T' M_{\hat{V}}Z\_'(Z\_M_{\hat{V}}Z\_')^{-1} = \varrho_1'$, where ϱ_1 is a (K+1)X1 vector with first element equal to one and zero elsewhere.

$$\iota_T' M_{\hat{V}}Z\_'(Z\_M_{\hat{V}}Z\_')^{-1} = \varrho_1' = \begin{bmatrix}1\\0\end{bmatrix}' = \begin{bmatrix}1 & 0\end{bmatrix}$$

Let $A = \iota_T' M_{\hat{V}}Z\_'$ and $B^{-1} = (Z\_M_{\hat{V}}Z\_')^{-1}$

Let us derive A:

$$\begin{aligned} A = \iota_T' M_{\hat{V}}Z\_' = &\ (\iota_T'(I_T - \hat{V}'(\hat{V}\hat{V}')^{-1}\hat{V}))Z\_' \\ = &\ (\iota_T' - \iota_T'\hat{V}'(\hat{V}\hat{V}')^{-1}\hat{V})Z\_' \\ = &\ \iota_T' Z\_' \end{aligned}$$

Substitute $Z\_' = \begin{bmatrix}\iota_T & X\_'\end{bmatrix}$

$$\therefore A = \iota_T' Z\_' = \iota_T'\begin{bmatrix}\iota_T & X\_'\end{bmatrix} = \begin{bmatrix}\iota_T'\iota_T & \iota_T' X\_'\end{bmatrix}$$

Now, let us derive B^{-1}:

$$\begin{aligned} B^{-1} = &\ (Z\_M_{\hat{V}}Z\_')^{-1} \\ = &\ (\begin{bmatrix}\iota_T\\X\_\end{bmatrix}(I_T - \hat{V}'(\hat{V}\hat{V}')^{-1}\hat{V})\begin{bmatrix}\iota_T & X\_'\end{bmatrix})^{-1} \\ = &\ (\begin{bmatrix}\iota_T I_T - \iota_T'\hat{V}'(\hat{V}\hat{V}')^{-1}\hat{V}\\ X\_I_T - X\_\hat{V}'(\hat{V}\hat{V}')^{-1}\hat{V}\end{bmatrix}\begin{bmatrix}\iota_T & X\_'\end{bmatrix})^{-1} \\ = &\ (\begin{bmatrix}\iota_T'\\ X\_ - X\_\hat{V}'(\hat{V}\hat{V}')^{-1}\hat{V}\end{bmatrix}\begin{bmatrix}\iota_T & X\_'\end{bmatrix})^{-1} \\ = &\ \begin{bmatrix}\iota_T'\iota_T & \iota_T' X\_' \\ X\_\iota_T - X\_\hat{V}'(\hat{V}\hat{V}')^{-1}\hat{V}\iota_T & X\_X\_' - X\_\hat{V}'(\hat{V}\hat{V}')^{-1}\hat{V}X\_'\end{bmatrix}^{-1} \\ = &\ \begin{bmatrix}\iota_T'\iota_T & \iota_T' X\_' \\ X\_\iota_T & X\_[I_T - \hat{V}'(\hat{V}\hat{V}')^{-1}\hat{V}]X\_'\end{bmatrix}^{-1} \\ = &\ \begin{bmatrix}\iota_T'\iota_T & \iota_T' X\_' \\ X\_\iota_T & X\_M_{\hat{V}}X\_'\end{bmatrix}^{-1} \end{aligned}$$

Therefore, we have

$$AB^{-1} = \iota_T' M_{\hat{V}} Z\_'(Z\_M_{\hat{V}}Z\_')^{-1} = \begin{bmatrix} \iota_T'\iota_T & \iota_T'X\_' \end{bmatrix} \begin{bmatrix} \iota_T'\iota_T & \iota_T'X\_' \\ X\_\iota_T & X\_M_{\hat{V}}X\_' \end{bmatrix}^{-1}$$

Let

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = \begin{bmatrix} \iota_T'\iota_T & \iota_T'X\_' \\ X\_\iota_T & X\_M_{\hat{V}}X\_' \end{bmatrix}$$

Therefore, we have

$$AB^{-1} = \iota_T' M_{\hat{V}} Z\_'(Z\_M_{\hat{V}}Z\_')^{-1} = \begin{bmatrix} A_{11} & A_{12} \end{bmatrix} \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}^{-1}$$

We need $B^{-1} = (Z\_M_{\hat{V}}Z\_')^{-1}$, by the formula of the partitioned matrices. (P.660 of Lutkepohl's New Introduction to Multiple Time Series Analysis)

$$B^{-1} = \begin{bmatrix} \underbrace{A_{11}^{-1}}_{(1X1)} + \underbrace{A_{11}^{-1}}_{(1X1)} \underbrace{A_{12}}_{(1XK)} \underbrace{G}_{(KXK)} \underbrace{A_{21}}_{(KX1)} \underbrace{A_{11}^{-1}}_{(1X1)} & -\underbrace{A_{11}^{-1}}_{(1X1)} \underbrace{A_{12}}_{(1XK)} \underbrace{G}_{(KXK)} \\ -\underbrace{G}_{(KXK)} \underbrace{A_{21}}_{(KX1)} \underbrace{A_{11}^{-1}}_{(1X1)} & \underbrace{G}_{(KXK)} \end{bmatrix}$$

where

$$\underbrace{G}_{(KXK)} = (\underbrace{A_{22}}_{(KXK)} - \underbrace{A_{21}}_{(KX1)} \underbrace{A_{11}^{-1}}_{(1X1)} \underbrace{A_{12}}_{(1XK)})^{-1}$$

$$\begin{aligned}
\therefore AB^{-1} &= \begin{bmatrix} \underbrace{A_{11}}_{(1X1)} & \underbrace{A_{12}}_{(1XK)} \end{bmatrix} \begin{bmatrix} \underbrace{A_{11}^{-1} + A_{11}^{-1}A_{12}GA_{21}A_{11}^{-1}}_{(1X1)} & \underbrace{-A_{11}^{-1}A_{12}G}_{(1XK)} \\ \underbrace{-GA_{21}A_{11}^{-1}}_{(KX1)} & \underbrace{G}_{(KXK)} \end{bmatrix} \\
&= \begin{bmatrix} \underbrace{A_{11}A_{11}^{-1}}_{=1} + \underbrace{A_{11}A_{11}^{-1}}_{=1}\underbrace{A_{12}GA_{21}A_{11}^{-1}}_{(1X1)} - \underbrace{A_{12}G1_{21}A_{11}^{-1}}_{(1X1)} & \underbrace{-A_{11}A_{11}^{-1}}_{=1}\underbrace{A_{12}}_{(1XK)}\underbrace{G}_{(KXK)} + \underbrace{A_{12}}_{(1XK)}\underbrace{G}_{(KXK)} \end{bmatrix} \\
&= \begin{bmatrix} \underbrace{1}_{(1X1)} + \underbrace{A_{12}GA_{21}A_{11}^{-1}}_{(1X1)} - \underbrace{A_{12}GA_{21}A_{11}^{-1}}_{(1X1)} & \underbrace{-A_{12}G}_{(1XK)} + \underbrace{A_{12}G}_{(1XK)} \end{bmatrix} \\
&= \begin{bmatrix} \underbrace{1}_{(1X1)} & \underbrace{0}_{(1XK)} \end{bmatrix}
\end{aligned}$$

Therefore,

$$AB^{-1} = \iota_T' M_{\hat{V}} Z\_'(Z\_M_{\hat{V}}Z\_')^{-1} = \begin{bmatrix} 1 & 0 \end{bmatrix}$$

Now, we can substitute this into equation (32)

$$\hat{\Lambda} = (\hat{\beta}\hat{\beta}')^{-1}\hat{\beta}[rxM_{\hat{V}}Z\_'(Z\_M_{\hat{V}}Z\_')^{-1}]$$

$$\therefore \hat{\Lambda} = (\hat{\beta}\hat{\beta}')^{-1}(rx + \frac{1}{2}\hat{B}^{*}vec(\hat{\Sigma})\iota_T' + \frac{1}{2}\sigma^2\iota_N\iota_T')M_{\hat{V}}Z\_'(Z\_M_{\hat{V}}Z\_')^{-1} \tag{35}$$