Erasmus university Rotterdam

Erasmus school of economics

Bachelor Thesis Quantitative finance

Predicting bond returns using the output gap in expansions and recessions

Author:

Martijn Eertman

Studentnumber:

431267

Supervisor:

Dr. X. Xiao

Second assessor :

Dr. X. Xiao

July 22, 2018

ERASMUS UNIVERSITEIT ROTTERDAM

ERASMUS SCHOOL OF ECONOMICS

Abstract

In this paper new evidence will be provided about the consensus that the output gap can predict excess bond returns. Also I investigate how the predictability of the output gap changes across economic expansions and recessions. Because of the persistence of the variables and the serial correlation of the error term, I find no evidence for the statistical significance of the output gap, even when the regression is conditioned on the business cycle. In general, the business cycle gives different coefficients for the predictors with a different statistical significance. Because of the lack of observations in recessions, it is hard to make conclusions about the significance of the business cycle on the predictability of excess bond returns.

Introduction

How does the output gap affect the excess return on bonds? Can this variable add to the predictive power of the yield curve towards bond excess returns, and does this depend on the business cycle? Given that the current yield of the 10-year U.S. treasury bond is very low, it is very important to find the reasons and different factors behind the level of the yield curve, and whether this differs across economic expansions and recessions. Namely, this way strategies for investing and monetary policy can be improved. Also the impact of the financial crisis in 2008 can be better understood.

For a very long time it has been recognized that most information that is useful for predicting interest rates or bond returns is contained in the yield curve. Also from previous research there is evidence that the information contained in the yield curve can be summarized into the first three principal components, which are commonly labelled the level, slope, and curvature. For example, Heidari and Wu (2001) have shown that three factors capture 99.5% of the variation in the yield curve. When regressing the yields on the first three principal components, I indeed find a value for R^2 that is very close to 1.

The consensus of the yield factors having a large impact on interest rates and bond returns is usually put forward as the spanning hypothesis, which states that the first three principal components of the yield curve contain all information that is useful for forecasting future interest rates and bond returns. From a finance perspective this would be very convenient, because this would mean that someone that is interested in the behavior of bond returns would only have to look at the yield curve. However, even though it is clear that the yield curve factors explain most of the cross-sectional variance of yields, it is not clear whether they capture all the information that is relevant for forecasting yields and bond risk premia.

There are several predicting variables that can potentially add to the predictive power, such as for example the output gap. The output gap is an indicator of the difference between the output of the economy and the maximum potential output of the economy expressed as a fraction of GDP. Therefore a negative output gap is an indication of an underperforming economy, while a positive output gap indicates an economy that outperforms expectations. As the output gap is a good indicator of how the economy is doing, it can tell us something about bond excess returns. According to Cooper and Priestley

(2008) (CPR) the output gap can predict next year's excess returns on U.S. government bonds. This would imply that there is information that is useful for predictions that is not captured by the level, slope, and the curvature of the current yield curve, and thereby it would reject the spanning hypothesis. Other studies like Joslin et al. (2014), Ludvigson and Ng (2010) and Cochrane and Piazzesi (2005) have found that measures of economic growth and inflation, factors inferred from a large set of macro variables, and the fourth and fifth principal component of yield, respectively, help predict excess bond returns. Nevertheless, in this paper I will focus on the output gap as additional explanatory variable.

To improve the forecasts of bond returns I want to know if the output gap variable has additional value to the first three principal components, and whether or not this depends on the business cycle. In all the studies that find evidence against the spanning hypothesis the key evidence comes from future yields or excess returns that have predictive variables that are very persistent and give a good summary of the information in the yield curve. Therefore in general these predictors are correlated with the lagged forecast errors. I will investigate what influence these characteristics can have on the conclusion that the output gap has additional predictive power. To do this I analyse the regression model that shows the relationship between the bond returns and the set of explanatory variables that contains the first three principal components and the output gap. I will look at the properties of this regression model that might explain some of the results that can be obtained from testing the spanning theory, both by looking at previous research and by carrying out a simulation study.

Then, the spanning hypothesis will be tested by generating bootstrapped samples of the returns and comparing the t-statistic with the t-statistic of the actual data. To determine if the spanning hypothesis holds with regard to the variable output gap, a number of regressions is carried out all containing a combination of the variable output gap and the principal components. This way a conclusion can be made on whether the output gap is a relevant factor in predicting the bond returns.

This also brings us back to the business cycle and the influence it has on bond returns. One might expect different outcomes for regressions on economic variables in economic expansions as opposed to recessions. This can have an influence on both the first three principal components of the yield curve and the output gap. To see if there's a difference in bond risk premia across expansions and recessions, a model is estimated that differentiates between expansions and recessions using indicator functions. This way we get different coefficients for expansions and recessions. I will use a modified Campbell-Shiller regression to show the influence of the business cycle on the predictability of the excess returns. Several of these regressions are done that can contain the output gap variable and the principal components.

1 Data

1.1 Variables and regression setup

For the research that tests the spanning hypothesis the regressions that are done are of the following form:

$$y_{t+h} = \beta_0 + \beta_1' x_{1t} + \beta_2' x_{2t} + u_{t+h}. \tag{1}$$

Here y_{t+h} is the dependent variable that represents the h-period excess return on a multi-period bond of month t. For the empirical studies that are performed the focus lies on the one-year excess returns of the five-year bond, so y_{t+h} with $h = 12$. u_{t+h} is the forecast error, and x_{1t} and x_{2t} are the vectors containing the predictors at time t. x_{1t} contains the variables that represent the information in the yield curve, which are the first three principal components of the observed yield, i.e., the level, slope and curvature, respectively:

$$x_{1t} = (PC1_t, PC2_t, PC3_t)', \tag{2}$$

while x_{2t} contains the predictor output gap of which I would like to test whether it has additional predictive power:

$$x_{2t} = gap_t, \tag{3}$$

The output gap is the deviation of the actual GDP from the potential GDP, where potential GDP is defined as the maximum sustainable output estimated based on a neoclassical production function (Cooper and Priestley (2008)). The variable output gap is lagged one month, and is measured as the deviation of the Feds Industrial Production series from a quadratic time trend. In CPR the measure of the output gap is lagged by one month to account for the publication lag of the Feds Industrial Production data. The data of the different dependent and explanatory variables run from 1952 until 2003 and consist of 618 observations.

To test whether the variable output gap adds predictive power to the regression in (1), the following null hypothesis is used

$$H_0: \quad \beta_2 = 0, \tag{4}$$

which says that the information in the yield curve that is contained by x_{1t} spans the relevant predictive information, and that the other variable x_{2t} has no additional predictive power.

1.2 Autocorrelated predictors

The serial correlation of the predictors can have a substantial influence on the results that can be obtained from the research that is done. In the case that x_{1t} and x_{2t} are very persistent and the error term is serially correlated, which is the case for overlapping bond returns, a substantial increase in R^2 is very likely when x_{2t} is added to the regression even if this variable has no significant additional predictive power.

The persistence of the different random variables can be measured as the autocorrelation of the respective variables and is shown in table 1 below along with the serial correlation of the error term. The table contains the autocorrelation of the first three principal components, the output gap and the error term at lags of 1, 3, 6 and 12 months.

Table 1: Levels of autocorrelation of the predictors and the error term

Predictor	ACF(l)			
	1	3	6	12
PC1	0.986	0.956	0.917	0.841
PC2	0.939	0.830	0.711	0.527
PC3	0.587	0.469	0.258	0.147
output gap	0.975	0.893	0.750	0.475
error term	0.866	0.678	0.425	-0.069
output gap · error term	0.881	0.638	0.376	-0.021

Table 1 shows that the variables in x_{1t} have a very high autocorrelation, especially the first two principal components, which have autocorrelations larger than 0.9 for the one month lag. This means that the level and slope of the yield curve are very persistent, which is a well known characteristic of interest rates. The first principal component is most persistent as the autocorrelation decreases much slower than the autocorrelation of the other principal components. The output gap x_{2t} is also very persistent, as the first autocorrelation is 0.975. The autocorrelation of the error term is also quite substantial, which is a well known characteristic for multi-period bond returns. These results can make it more difficult to interpret the fit of models including the output gap. It is clear that our data exhibit autocorrelation properties that are similar to those of general interest rates.

2 Implications of autocorrelated predictors

2.1 R^2-bias

In small samples the serial correlation in the residuals can increase the bias and the variance of the R^2 of the regression Carrodus and Giles (1992). To see how much the R^2 increases when the variables in x_{2t} are added we need to look at the behavior of the difference between the R^2's of the regressions with and without the variable output gap $R_2^2 - R_1^2$. Here R_2^2 and R_1^2 are the regression R^2's of the regressions with and without x_{2t}, respectively. Therefore $R_2^2 - R_1^2$ can be interpreted as the increase in the goodness of fit that is caused by adding the variable output gap.

If the null hypothesis is true ($\beta_2 = 0$), when x_{1t}, x_{2t} and u_{t+h} are stationary and satisfy standard regularity conditions, and the additional predictor is uncorrelated with each of the other predictors, then, as is explained by Bauer and Hamilton (2017), the increase in R^2 in probability converges to zero. The correlation between the output gap and the first three principal components of the yield curve is indeed very small. Only the second principal component, the slope of the yield curve, exhibits a substantial correlation with the output gap. The reason for this is that the slope of the yield curve can be a good indicator of the business cycle, which is directly linked to the output gap. When regressing the output gap on the slope of the yield curve, the R^2 is still very small. Therefore for this research we can assume the correlation between x_{1t} and x_{2t} to be negligible.

However, due to the serial correlation, adding the output gap to the regression can still lead to an increase in R^2 even if it does not belong in the regression. If $x_{2t}u_{t+h}$ has a positive serial correlation, $R_2^2 - R_1^2$ has a higher expected value and a higher variance compared to situations when it is not (Bauer and Hamilton (2017)). This serial correlation in $x_{2t}u_{t+h}$, that can be seen in table 1, contributes to larger values for $R_2^2 - R_1^2$ on average and to a larger variability in $R_2^2 - R_1^2$ across different samples. Therefore including x_{2t} can substantially increase the R^2 even if the spanning hypothesis (equation (4)) holds true.

2.2 small-sample distortions of the t-test

Alongside with the R^2-bias the persistence of the explanatory variables can lead to size distortion of the tests that are used to assess the spanning hypothesis with respect to the output gap. The OLS t-test of $\beta_2 = 0$ asymptotically has a standard normal distribution if the standard first-order asymptotic approximation is used. The fact that the standard errors of the estimates of β_2 are inaccurate leads to the distortion of standard inference. The cause of the problem can be shown using the following setting where the OLS estimation of (1) in the case of h=1 is considered:

$$y_{t+1} = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + u_{t+1}, \text{ with} \tag{5}$$

$$x_{1,t+1} = \rho_1 x_{1t} + \epsilon_{1,t+1}. \tag{6}$$

$$x_{2,t+1} = \rho_2 x_{2t} + \epsilon_{2,t+1}. \tag{7}$$

In this setting x_{1t} and x_{2t} are very persistent scalars with serial correlation coefficients ρ_1 and ρ_2 close to 1. The error terms ϵ_{1t}, ϵ_{2t}, u_t follow martingale difference sequences. A variable is a martingale difference sequence if its expectation conditional on the past is zero, which is something that naturally applies to error terms. The errors terms jointly have the following variance matrix:

$$V = E \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ u_t \end{bmatrix} \begin{bmatrix} \epsilon_{1t} & \epsilon_{2t} & u_t \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & 0 & \delta\sigma_1\sigma_u \\ 0 & \sigma_2^2 & 0 \\ \delta\sigma_1\sigma_u & 0 & \sigma_u^2 \end{bmatrix}. \tag{8}$$

The error terms ϵ_{1t}, ϵ_{2t}, u_t have variances σ_1^2, σ_2^2, and σ_u^2, respectively. δ is the correlation coefficient. The matrix V shows that the coefficient δ needs to be zero for the variables in x_{1t} to be strictly exogenous. This means that in general the correlation between the dependent variable and the predictors δ leads to endogeneity of the variable x_{1t}. The product between the output gap and the error term for the excess returns $x_{2t}u_{t+1}$ is serially uncorrelated no matter what the value of δ is.

If the error term u_t is correlated with x_{2t}, and the predictors x_{1t} and x_{2t} are very persistent, the null hypothesis $\beta_2 = 0$ will be rejected too often when a standard t-test is applied in small sampled data (Bauer and Hamilton (2017)). The OLS estimation of the parameter of the output gap β_2 can give some intuition to this result. The output gap is regressed on the variable x_{1t} containing the principal components, while the bond excess returns are regressed on the principal components. The residuals of the first regression are regressed on the residuals of the second regression. When x_{1t} and x_{2t} are very persistent, the small-sample properties of the first regression will be different. As x_{1t} is not strictly exogenous this can make a big difference for the small-sample distribution of the OLS estimate of β_2.

3 Simulation studies

3.1 Theoretical simulation

The effect of the persistence of the explanatory variables can be shown using a simulation study, that does not involve any economic data. It can be seen that the persistence of the predictors leads to size distortions of the t-tests. These t-tests are done to examine the relevance of an artificial variable x_{2t} for predicting an artificial dependent variable y_t. The simulation study generates 50,000 artificial data samples as replacements for the excess bond returns and explanatory variables that will be used later to test the predicting power of the output gap, and is also used in Bauer and Hamilton (2017). Each sample is created and is used for a regression using the following steps. The variables x_{1t} and x_{2t} are generated recursively using the equations (6) and (7), starting with the initial values $x_{1,0} = x_{2,0} = 0$. Here the error terms ϵ_{1t} and ϵ_{2t} are serially independent Gaussian random variables with variance equal to one. Then to create values for the dependent variable y_t the just generated values of the explanatory variables are used. This can be seen in the following equation, that is an adjusted version of equation (5) with parameters

$\beta_0 = \beta_2 = 0$ and $\beta_1 = \rho_1$:

$$y_{t+1} = \rho_1 x_{1t} + u_{t+1}, \text{ with} \tag{9}$$

$$u_t = \delta\epsilon_{1t} + \sqrt{1-\delta^2}v_t \tag{10}$$

with $v_t \sim N(0, 1)$ an *i.i.d.* random variable, and the parameter δ now representing the correlation between ϵ_{1t} and u_t. The variances of the error terms σ_1, σ_2, and σ_u are equal to one as the covariance between ϵ_{1t} and ϵ_{2t} is considered to be zero. This simulation is done for different values for the degree of endogeneity δ, the persistence of the predictors ρ with $\rho_1 = \rho_2 = \rho$, and the sample size T. This is done to investigate whether the theoretical influence of these parameters can be seen in this simulations study. Now that an artificial sample with values for y_t, x_{1t}, and x_{2t} has been generated, the regression in equation (5) is done with OLS, to get the t-statistic of the parameter β using standard errors of the OLS estimation. This means that for each sample i out of the total number of 50,000 a t-statistic is obtained that is denoted as t_i. This value will be compared with t the two-sided 5% critical value of a t-distribution with $T - 3$ degrees of freedom. The test sizes are then defined as the fraction of samples in which the t-statistic of sample i is larger in absolute value than the 5% critical value t, so the fraction in which $|t_i| > t$ holds. The size distortion is the difference between the just calculated (true) size and the nominal size of 0.05. This means that if the size distortion is very small, the true size should be very close to 0.05. The sizes for the simulations using different values for the parameters δ, ρ, and T are shown in table 2.

The first thing that becomes clear from table 2 is that the values for the different combinations of T, δ, and ρ are all not too far distant from the nominal size of 0.05 and are all below 0.11. The sizes seem to increase when the persistence of the regressors ρ increases, and we can speak of size distortions in cases when $\rho > 0.9$, and the sizes increase when ρ increases. δ does not have a significant effect on the size distortions. The size distortions are decreasing when the sample size T increases and ρ is smaller than 1. In the case of $\rho = 1$ the sample size seems to have no effect on the true size of the test. The largest value in table 2 is 0.1051, which means that the t-test would reject the null hypothesis more than twice as often as it should. The reason behind the size distortions is the fact that the OLS standard errors substantially underestimate the sampling variability of the OLS estimates of β_1 and β_2 (Bauer and Hamilton (2017)).

3.2 Empirical simulation

Empirical research will be used in the form of a bootstrap procedure as an alternative of conventional tests. The inference of the research is based on the small sample distributions of the test statistics using a parametric bootstrap under the null hypothesis $H_0 : \beta_2 = 0$. According to Horowitz (2001) and Hall and Wilson (1991) using bootstraps under the null hypothesis makes it possible to find out the small-sample size of the conventional tests, and is therefore preferred to bootstrapping under an alternative hypothesis. Furthermore using bootstraps under the null hypothesis leads to more accuracy and more powerful tests compared to making use of bootstraps under alternative hypotheses.

The spanning hypothesis $H_0 : \beta_2 = 0$ will be tested by generating bootstrapped samples under this null hypothesis. First the bond excess returns are regressed on the first three principal components of the yield curve to obtain the weighting vector $\hat{w}$ for the bond excess returns, that gives weight to each of the principal components for explaining the variation in the bond excess returns:

$$y_{t+h} = \hat{w}x_{1t} + \hat{u}_t, \tag{11}$$

Also a VAR(1)-model for x_{1t} is estimated with OLS:

$$x_{1t} = \hat{\phi}_0 + \hat{\phi}_1 x_{1,t-1} + e_{1t} \quad t = 1, ..., T. \tag{12}$$

The VAR(1)-model captures the linear interdependence, the dynamic development, and the cross-sectional dependence of the yields even though it does not impose absence of arbitrage. A model in this simple form therefore fits and forecasts the behavior of the yield curve very well (Duffee (2011) and Hamilton and Wu (2011)).

Using the just calculated estimates of the parameters w, ϕ_0, and ϕ_1, 5000 artificial yield data samples with length $T = 618$, the size of the original sample, are generated. The starting values for the recursion are drawn from the unconditional distribution implied by the estimated VAR for x_{1t}. The iterations starting with the just calculated starting values are done with equation (12) using parameter estimates $\hat{\phi}_0$ and $\hat{\phi}_1$:

$$x^*_{1t} = \hat{\phi}_0 + \hat{\phi}_1 x^*_{1,t-1} + e^*_{1t}. \tag{13}$$

The bootstrap residuals for the bootstrap variables e^*_{1t} are drawn from the empirical distribution of e_{1t}. In a similar way a bootstrap for the other predictor x^*_{2t} is made.

Then, the artificial returns are obtained with equation(11) using the artificial values for the variable x_{1t} and the parameter estimate $\hat{w}$:

$$y^*_{t+h} = \hat{w}x^*_{1t} + \hat{u}^*_t, \tag{14}$$

The bootstrap residuals $\hat{u}^*_t$ are drawn from a normal distribution with mean zero and variance σ^2_u. The standard deviation of the errors, σ_u, is equal to the sample standard deviation of the fitting errors $\hat{u}_t$. This simulation creates artificial samples of returns y^*_{t+h} that have dynamics that are very similar to those of the observed returns. For example the recursion gives values for y_{t+h} for which only the first three principal components have predictive power, which is also the case for actual bond returns for which the spanning hypothesis holds. This means that for this bootstrap the yield curve contains all the information necessary to forecast future returns.

With the help of a t-test I can test the spanning hypothesis using our just created bootstraps. The t-test tests whether the parameter for the output gap is significantly different from zero. The t-statistic in the data is denoted by t and the t-statistic in bootstrap i by t^*_i. The true size of a conventional t-test is estimated as the fraction of samples in which $|t^*_i|$ exceeds the usual asymptotic critical value. This value is equal to 0.0406, which is quite close to 0.05. Unfortunately, we cannot really draw a conclusion from this results as the method that is used is not valid. Namely the distribution of the test statistic depends on ρ_1 and ρ_2 through local-to-unity parameters that cannot be estimated consistently Bauer and Hamilton (2017). The parameters that are used in the VAR-equations are only estimates, while only the true values for these parameters would give a right sized test. Therefore, the bootstrap gives a test of H_0 that does not have a size of five percent. A solution to this problem is to use the Monte Carlo simulation, that is mentioned in section 3.1.

The Monte Carlo simulation applied to the empirical framework will generate 5000 samples using the following steps. For each sample i that is simulated , I calculate the t-statistic ($\tilde{t}_i$) for testing the null hypothesis ($\beta_2 = 0$), estimate the autoregressive models for the predictors using OLS, generate a bootstrap simulation using the estimated autoregressive coefficients and estimate a regression on the simulated bootstrap:

$$y^*_{t+h} = \hat{\phi}_1 x^*_{1.t-1} + v^*_t. \tag{15}$$

Here the values for v^*_t are drawn from the distribution of the residuals in a regression of

y_t on $x_{1,t-1}$, and is jointly drawn with the errors ϵ^*_{1t} and ϵ^*_{2t} to keep the same correlation as in the data. After the regression of the bootstrap I calculate the t-test of $H_0 : \beta_2 = 0$ (t^*_i). With these t-values the value c is calculated such that $|t^*_i| > c$ holds in five percent of the samples. This value is reported in table 3. For each sample the null hypothesis is rejected if the original t-value is larger than this number c, so if $\tilde{t}_i > c$. The true size of the bootstrap procedure applied to the model is the fraction of samples for which this is the case. The calculated size is equal to 0.047, the value from table 3, which is slightly below five percent, and therefore gives a small size distortion. This means that if the bootstrap procedure applied to this model does not reject the spanning hypothesis, it is save to say that the evidence against the spanning hypothesis is not persuasive.

4 The effect of the output gap in expansions and recessions

4.1 The effect of the output gap independent of the business cycle

To further elaborate on the simulation inference in determining the extent to which the output gap is useful for predicting excess bond returns, I make a number of predictive regressions with the dependent variable excess bond return on the five-year bond y_{t+h} and the independent variables output gap and the first three principal components. All regressions that are done contain a combination of the variable gap and the first three principal components, and they all contain a constant. From these regressions the coefficients of the output gap variable are obtained from which we can for example see whether it has a positive or negative influence on the excess bond returns. I have also calculated the OLS t-statistics and HAC t-statistics and performed the bootstrap method described in section 3.2 to see if the coefficients for the output gap variable are significant. For the HAC t-statistics Newey-West is ued with 22 lags as described by CPR.

The results are shown in table 3.The first regression only contains the lagged output gap x_{2t}:

$$y_{t+h} = \beta_0 + \beta_2' x_{2t} + u_{t+h}. \tag{16}$$

It becomes clear that the variable hardly has an impact on the regression results. In this regression the variable output gap has a negative coefficient that is very close to zero, that is not significant. Both the OLS and HAC t-statistic are very close to zero, which results in a very high p-value. This means that the variable output gap is insignificant at the 5% level.

The second regression only contains the first three principal components in x_{1t} as explanatory variables.

$$y_{t+h} = \beta_0 + \beta_1' x_{1t} + u_{t+h}. \tag{17}$$

This regression makes clear that the principal components of the yield curve are very important for predicting the excess returns on bond returns. The coefficients are highly significant as the OLS and HAC t-statistics are very large, and the HAC p-values are practically zero. The one thing that can be concluded from this is that the principal components of the yield curve are indeed very important for the prediction of excess returns on bonds.

The third specification includes the first three principal components of the yield curve, and is therefore most interesting as it shows how the output gap relates to the principal components with respect to their predictive value:

$$y_{t+h} = \beta_0 + \beta_1' x_{1t} + \beta_2' x_{2t} + u_{t+h}. \tag{18}$$

The bootstrap method is applied for this specification to see how the result of this test differs from the HAC t-statistic. Namely, the size of the HAC t-test is substantially distorted, because the true size is more than 21 %, as can be seen in table 3. However, the size of the bootstrap is very close to 5% as was discussed in 3.2. The regression shows results that are very similar to the first two specifications. The output gap variable is still statistically very insignificant, while the principal components are statistically very significant in the regression. The bootstrap critical value is significantly larger than the HAC t-statistic, which again leads to the conclusion that the HAC t-test is substantially oversized. The bootstrap p-value is very close to 50 %, which means that we can conclude that the output gap is insignificant for predicting the excess bond returns, as the size of the test is not seriously distorted. After carrying out these regression I have not found any evidence that the output gap can help predict the excess bond returns, both from the HAC t-test and the bootstrap test, and so there is no reason to reject the spanning hypothesis with respect to the output gap.

4.2 The business cycle dependent effect of the output gap

There is reasonable amount of evidence that the business cycle has an influence on the behavior of interest rates. Studies like Hamilton (1988) and Gray (1996) have shown that interest rates are more persistent in expansions than in recessions and that two-state models describe interest rate dynamics much better than single-state models. I would like to see how the predictability of the excess bond returns compares between expansions and recessions. This can be done by using a regression that is conditional on the business cycle, so variables get separate coefficients for expansions and recessions.

Firstly, a modified regression that is based on the following regression from Campbell and Shiller (1991) is used:

$$y_{t+m,k-m} - y_{t,k} = \alpha_k + \beta_k \frac{m}{k-m}(y_{t,k} - y_{t,m}) + u_{t+m,k}, \tag{19}$$

where $y_{t,k}$ refers to the k-period bond yields in period t. The modified Campbell-Shiller regression conditions on the state of the business cycle, and therefore has dummies to check for expansions and recessions:

$$\begin{aligned} y_{t+m,k-m} - y_{t,k} = \alpha_k^{EXP} 1_{\{z_t \geq c\}} + \beta_k^{EXP} \frac{m}{k-m} 1_{\{z_t \geq c\}}(y_{t,k} - y_{t,m}) + \alpha_k^{REC}(1 - 1_{\{z_t \geq c\}}) \\ + \beta_k^{REC} \frac{m}{k-m}(1 - 1_{\{z_t \geq c\}})(y_{t,k} - y_{t,m}) + \tilde{u}_{t+m,k}. \end{aligned} \tag{20}$$

Here $1_{\{z_t \geq c\}}$ is an indicator function that has a value of one for expansions if $z_t \geq c$ and a value of zero for recessions if $z_t < c$. To identify recessions the variable z_t is used. This variable refers to the Purchasing Managers' Index (PMI), where $c = 44.5$ is the threshold value from Berge and Jorda (2011) that is used to differentiate between recessions and expansions. The PMI is a well known indicator of business cycle activity. The advantage of using PMI is that it is available in real time without any publication lags or subsequent data revision. Christiansen et al. (2014) make clear that the PMI is the best recession indicator among a large number of economic variables. When applying this indicator it turns out that 544 of the 618 observations are from expansions, while the other 74 observations are from recessions.

The regression is run for $m = 3$ and $k = 60$, because we are dealing with five-year bonds. The results of this regression are shown in table 4 below.
The first thing of interest is if the conditional Campbell-Shiller regression has a better goodness-of-fit than the standard regression, by checking if the R^2 has increased. The R^2

of the standard regression is equal to 0.027 and the R^2 of the model that is conditioned on the business cycle is equal to 0.105. This means that the R^2 increases substantially when we condition on the business cycle. When comparing the coefficients across the regressions it becomes clear that coefficients that are specific for recessions or expansions are not statistically more significant than their general counterparts in the standard regression. In the standard regression the coefficient for the constant is significant as the p-value is much smaller than 0.05. In the conditional regression the coefficient for the constant during expansions and recessions is still statistically significant. The coefficient of the slope in expansions is statistically significant, but its counterpart in recessions is statistically very insignificant at the 5% level.

Secondly, I can test whether the coefficients are significantly different across expansions and recessions. In other words, I test whether $\alpha_k^{\Delta} = \alpha_k^{REC} - \alpha_k^{EXP}$ and $\beta_k^{\Delta} = \beta_k^{REC} - \beta_k^{EXP}$ are large. This should give a good indication of the importance of the business cycle for predicting bond returns with the yield curve. I find that the coefficients are statistically significantly different at the 5% level. Therefore it becomes clear that the business cycle is very important for predicting bond yields.

To now further expand on this notion a regression for the bond excess returns is done that contains the first three principal components and the output gap as explanatory variables, and is conditioned on the business cycle. This regression is the transformation of (18) to make it fully dependent on the business cycle. For each parameter in this specification there are now two parameters for expansions and recessions, respectively.

$$\begin{aligned} y_t = {} & \beta_0^{EXP} 1_{\{z_t \geq c\}} + \beta_0^{REC}(1 - 1_{\{z_t \geq c\}}) + \beta_1^{EXP} 1_{\{z_t \geq c\}} x_{1t} \\ & + \beta_1^{REC}(1 - 1_{\{z_t \geq c\}}) x_{1t} + \beta_2^{EXP} 1_{\{z_t \geq c\}} x_{2t} + \beta_2^{REC}(1 - 1_{\{z_t \geq c\}}) x_{2t} + u_t, \end{aligned} \tag{21}$$

The results of the regression in (21) are shown in table 5.

The R^2 of this specification is equal to 0.275, which is substantially higher than the R^2 of the unconditional regression of which the results are shown in table 3, which is 0.232. Most of the coefficients are statistically more significant in expansions compared to recessions, which might be related to the difference in the number of observations between expansions and recessions. The HAC and bootstrap tests give similar results for this specification compared to the regression that is not dependent on the business cycle. Again the size of the HAC t-test is substantially distorted, while this is not the case for the bootstrap test. Both test indicate statistical insignificance for the output gap in both expansions and recessions. The business cycle does have a substantial influence on the bond excess returns, as the coefficients of the principal components are substantially different across expansions and recessions. The coefficients of the output gap both for expansions and recessions are statistically insignificant at the 5% level. So even if we condition the regression on the business cycle we still find no evidence that the output gap has significant predictive power with respect to the excess bond returns.

5 Conclusion

With the different techniques used in this paper the work of previous research has been challenged and extended. The spanning hypothesis, that states that all the information that is required to predict interest rates or risk premia is contained by the yield curve, is applied to the additional variable output gap. Because of the persistence of the explanatory variables and the serial correlation of the error terms, an increase in R^2 after adding the variable output gap does not have to mean that this variable has additional predictive power. Also because of the persistence of the variables the size of the t-test for testing the spanning hypothesis are distorted, as is shown by the simulation studies.

Using the Campbell-Shiller regression I find that the predictability of the bond yields increases when the regression is made conditional on the business cycle, as I find different coefficients for expansions and recessions that are statistically more significant than the coefficients for the unconditional regression. This influence of the business cycle also becomes clear when I regress the excess bond returns on the principal components of the yield curve and the output gap. For expansions and recessions I find different coefficients, but the coefficients for the output gap are statistically insignificant. In general we find somewhat higher values for R^2 when the regression is conditioned on the business cycle, which can be explained by the persistence in the explanatory variables and the serial correlation of the error terms.

In both the simulation study and the regressions the output gap showed no significant predictive power, even when I differentiate between expansions and recessions. Therefore we cannot reject the spanning hypothesis. The business cycle has a positive influence on the predictability of the excess returns, as it makes the predictors more explanatory and in general provides a better fit. If we want to predict bond excess returns the output gap does not have to be used as an additional predictor to the yield curve, and it is better to condition the predictions on expansion and recessions.

References

Bauer, M. and Hamilton, J. (2017). Robust bond risk premia. *The Review of Financial Studies*, 31(2):399–448.

Berge, T. and Jorda, O. (2011). Evaluating the classification of economic activity into recessions and expansions. *American Economic Journal: Macroeconomics*, 3(2):246–277.

Campbell, J. and Shiller, R. (1991). Yield spread and interest rate movements: A birds eye view. *The Review of Economic Studies*, 58(3):495–514.

Carrodus, M. and Giles, D. (1992). The exact distribution of r 2 when the regression disturbances are autocorrelated. *Economics Letters*, 38(4):375–380.

Christiansen, C., Eriksen, J., and Moller, S. (2014). Forecasting recessions: The role of sentiment. *Journal of Banking and Finance*, 49:459–468.

Cochrane, J. and Piazzesi, M. (2005). Bond risk premia. *American Economic Review*, 95(1):138–160.

Cooper, I. and Priestley, R. (2008). Time-varying risk premiums and the output gap. *The Review of Financial Studies*, 22(7):2801–2833.

Duffee, G. (2011). Forecasting with the term structure: the role of no-arbitrage.

Gray, S. (1996). Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics*, 42(1):27–62.

Hall, P. and Wilson, S. (1991). Two guidelines for bootstrap hypothesis testing. *Biometrics*, 47(2):757–762.

Hamilton, J. (1988). Rational-expectations econometric analysis of changes in regime: An investigation of the term structure of interest rates. *Journal of Economic Dynamic and Control*, 12(2–3):385– 423.

Hamilton, J. and Wu, J. (2011). Testable implications of affine term structure models. *Journal of econometrics*, 178(2):231–242.

Heidari, M. and Wu, L. (2001). Are interest rate derivatives spanned by the term structure of interest rates?

Horowitz, J. (2001). The bootstrap. 5:3159–3228.

Joslin, S., Priebsch, M., and Singleton, K. (2014). Risk premiums in dynamic term structure models with unspanned macro risks. *Journal of Finance*, 69(3):1197–1233.

Ludvigson, S. and Ng, S. (2010). Macro factors in bond risk premia. *Review of Financial*

Studies, 22(12):5027–5067.

6 Appendix

Table 2: Simulation study: true sizes of conventional t-tests

T	$\delta = 0$			$\delta = 0.8$			$\delta = 1$		
	$\rho = 0.9$	$\rho = 0.99$	$\rho = 1$	$\rho = 0.9$	$\rho = 0.99$	$\rho = 1$	$\rho = 0.9$	$\rho = 0.99$	$\rho = 1$
50	0.0694	0.0963	0.1019	0.0685	0.0948	0.0994	0.0703	0.0952	0.0987
100	0.0627	0.0911	0.0989	0.0630	0.0880	0.0941	0.0630	0.0904	0.0973
200	0.0608	0.0934	0.1052	0.0550	0.0903	0.0972	0.0569	0.0914	0.0990
500	0.0502	0.0793	0.1051	0.0511	0.0769	0.1038	0.0512	0.0776	0.1057

The true size, as a fraction, of a conventional t-test of the null hypothesis $H_0 : \beta_2 = 0$ with nominal size equal to 5%, in simulated small samples. The correlation coefficients, that represent the persistence of the predictors, are set to $\rho_1 = \rho_2 = \rho$, which means that they are the same for x_{1t} and x_{2t}. δ determines the degree of endogeneity, which is the correlation of x_{1t} with the lagged error term u_t.

Table 3: Results of the regressions

	constant	output gap	PC1	PC2	PC3
Coefficient	0.008	-0.126			
OLS t-statistic	3.832	-3.225			
HAC t-statistic	1.226	-1.080			
HAC p-value	0.221	0.281			
Coefficient	-0.031		0.001	0.035	-0.062
OLS t-statistic	-5.445		4.564	11.323	-3.429
HAC t-statistic	-3.005		1.666	5.459	-2.057
HAC p-value	0.003		0.096	0.000	0.040
Coefficient	-0.035	0.147	0.001	0.043	-0.067
OLS t-statistic	-6.125	3.523	4.377	11.496	-3.693
HAC t-statistic	-3.120	1.235	1.595	4.881	-2.282
HAC p-value	0.002	0.217	0.111	0.000	0.023
Bootstrap 5% c.v.		1.976			
Bootstrap p-value		0.499			
HAC size		0.214			
Bootstrap size		0.047			

The table contains the results of the different regressions done with the one-year excess return on five-year bonds. In the regressions the set of explanatory variables contains either the output gap that is lagged one month, the first three principal components of the yield curve, or all four of these variables. The HAC t-statistic and the HAC p-value are calculated using the HAC standard errors, that are based on the Newey-West estimator with 22 lags. The bootstrap 5% critical value is calculated in 3.2

Table 4: Results of the Campbell-Shiller regressions

	α_5	β_5		α_5^{EXP}	α_5^{REC}	β_5^{EXP}	β_5^{REC}
Coefficient	0.001	-1.954		0.002	-0.005	-2.993	0.097
OLS t-statistic	2.203	-3.750		4.774	-5.131	-5.439	0.072
HAC t-statistic	2.478	-5.059		2.893	-5.837	-5.583	0.053
HAC p-value	0.014	0.000		0.040	0.000	0.000	0.958

The table contains the results of the Campbell-Shiller regressions with on the left the results of the unconditional Campbell-Shiller regression and on the right the results of the regression conditioned on the business cycle.

Table 5: Results of the conditional regression

	β_0^{EXP}	β_0^{REC}	β_1^{EXP}			β_1^{REC}			β_2^{EXP}	β_2^{REC}
			pc1	pc2	pc3	pc1	pc2	pc3		
Coefficient	-0.032	-0.026	0.001	0.048	-0.078	0.003	0.007	0.027	0.212	0.070
OLS t-statistic	-5.398	-1.616	2.264	12.302	-4.083	4.545	0.641	0.555	4.449	0.786
HAC t-statistic	-2.407	-1.536	0.931	4.588	-2.852	1.554	0.402	0.573	1.581	0.418
HAC p-value	0.016	0.125	0.352	0.000	0.005	0.121	0.688	0.567	0.114	0.676
Bootstrap 5% c.v.									1.952	1.938
Bootstrap p-value									0.495	0.493
HAC size									0.116	0.682
Bootstrap size									0.047	0.047

The table contains the results of the regression conditional on the business cycle with the explanatory variable x_{1t} containing the first three principal components (pc1, pc2, pc3) and x_{2t} containing the output gap.