Direct tests of the expectations theory of the term structure: Survey expectations, the term premium and coefficient biases

Smant, David / D.J.C.

Erasmus University Rotterdam

6 January 2010

Direct tests of the expectations theory of the term structure: Survey expectations, the term premium and coefficient biases

David J.C. Smant

Erasmus School of Economics
Erasmus University Rotterdam.
P.O. Box 1738
3000 DR Rotterdam
The Netherlands

January 2010
Revised February 2011

Email: smant@ese.eur.nl
Phone: +31 10 4081408

Acknowledgements: I would like to thank Job Swank and Casper de Vries for their constructive comments. Usual disclaimer applies.

Direct tests of the expectations theory of the term structure: Survey expectations, the term premium and coefficient biases

Abstract
A common perception in the literature seems to be that the expectations theory of the term structure of interest rates (ET) is rejected by the empirical tests. However, the overall evidence across countries and time periods is actually mixed between frequent support and occasional rejection of the ET and requires careful interpretation. This paper builds on two premises. First, the general version of the expectations theory does not require a constant term premium. Second, contrary to the classic perfect-foresight-with-error hypothesis real world rational expectations do not necessarily satisfy unbiasedness and ex post efficiency conditions. Direct tests that take into account the weaknesses of the restrictive auxiliary assumptions show that the expectations theory fits the term structure data very well.

JEL: E43, G10, G12
Keywords: Expectations theory, term structure of interest rates, survey expectations, term premium

1 INTRODUCTION

A large literature has reported on various empirical tests of the expectations theory of the term structure of interest rates (ET). Common empirical tests include the forward rate and excess holding return tests, roll-over spread tests, cointegration tests, VAR model cross-equation restrictions, and the variance bound test. It appears to be the view of many economists that the ET is generally rejected in the empirical tests. They point out that estimated coefficients do not conform to their hypothesized theoretical value, particularly when using data from the U.S., and auxiliary studies tend to find problems with the assumptions of rational expectations and efficient financial markets. Without reviewing the empirical literature in great detail, a comparison of the various and frequently contradictory results shows that the empirical evidence is in fact far from conclusive, particularly when viewed across countries, across markets, across sample periods, and across various testing methodologies.[1] It is important to note that the United States Treasury bill market consistently provides results that reject the usual ET hypotheses, whereas interest rates from many other markets and other countries tend to be much more supportive of the ET. The ET is sometimes rejected for some long (bond market) or for some short (money market) maturity ranges, but supported when using the opposite maturity range. Certain statistical tests seem to result in more frequent rejections of the ET than other tests. We know that as a result of the very nature of economic theories - being simplified and abstract representations of reality, - it is always possible to empirically reject a theoretical model.[2] The essential question that should follow from the various empirical results is why the evidence is in fact so mixed.

In this paper I use survey expectations on interest rates as an essential component of a direct test of the ET. There are surprisingly few studies that test the ET using survey expectations.[3] There may be several reasons for this. First, most studies using survey data focus on the quality of the forecasts, presumably partly on the assumption that 'rational' forecasts are required before they can be used in empirical studies. Second, the survey data available are frequently not immediately suitable for direct tests of the ET due to the use of period averages and forecast horizons that do not match the available term structure data. Finally, survey data are viewed with some suspicion, because the survey responses do not necessarily represent people's true expectations and may not represent the expectations of the relevant investor(s).

As far as I have been able to determine, the earliest investigation of the ET using survey expectations was reported by Kane and Malkiel (1967). Subsequently, Friedman (1979) used quarterly data from the U.S. Goldsmith-Nagan Bond and Money Market Letter on one-quarter-ahead and two-quarters-ahead expectations for the 3-month Treasury bill. His results showed that the term premium is not zero and in fact time-varying. Froot (1989) used the same source on survey expectations for several short and long interest rates and he found that the forward premium is a biased estimate of expected future interest rate changes. For short-maturity interest rates the results suggested a time-varying term premium. For short-term changes in long-maturity interest rates a time-varying term premium appeared to be relatively unimportant, but the survey expectations exhibited systematic forecast errors.[4] Cook and Hahn (1990) also used the Goldsmith-Nagan survey data and their results showed that the survey data improve the performance of the forward rate

---

[1] Shiller (1990), Cook and Hahn (1990), Campbell (1995) and Rudebusch (1995) provide most of the typical results and associated discussions. Key empirical studies using a multi-country framework and various types of tests are Jorion and Mishkin (1991), Hardouvelis (1994), Siklos and Wohar (1996), Gerlach and Smets (1997). Recent studies include Jondeau and Ricart (1999), Dominguez and Novales (2000), Drakos (2002). Empirical results have been shown to depend on the choice of sample period (Mankiw and Miron, 1986; Hardouvelis, 1988; Simon, 1990; Hsu and Kugler, 1997), country (Kugler, 1988; Gerlach and Smets, 1997; Jondeau and Ricart, 1999), market segment (Cook and Hahn, 1990; Longstaff, 2000; Downing and Oliner, 2007), maturity horizon (Cook and Hahn, 1990; Jorion and Mishkin, 1991), data frequency (Choi and Wohar, 1991; Drakos, 2002), and estimation method (Driffill, Psaradakis and Sola, 1997, 1998).

[2] A point stressed by Fama (1991) in his survey of the efficient markets literature. With respect to the ET, Dotsey and Otrok (1995, p.65) even go so far as to argue that "the rejections are so striking that the large amount of irrationality implied by the data is too implausible".

[3] MacDonald (2000) provides a recent survey of the literature using survey data for foreign exchange, interest rates and stock market prices. A broader review of the use of survey data in economics is Pesaran and Weale (2006).

[4] Survey expectations on interest rates exhibited underreaction, probably related to a similar phenomenon in inflation expectations during the same sample period.

equations, although the forecast R^2 remained relatively low and strong rejection of the ET restriction persisted in the Treasury bill market (but not in the Eurodollar and Commercial Paper market). Batchelor (1995) employed the Blue Chip Financial Forecasts which provides monthly observations on one to three-quarter-ahead expectations of 3-month U.S. Treasury bills. He concluded that the time-varying term premium is as important as the time-variation in expected interest rate changes for explaining movements in the forward spread. MacDonald and Macmillan (1994) examined the Consensus Economics Forecasts data on one-quarter ahead U.K. 3-month interbank rates. They found that 60% of changes in the forward spread can be explained by a time-varying term premium, although this result depends crucially on the inclusion of observations related to the U.K.'s ERM crisis in the early 1990s. Jongen, Verschoor and Wolff (2005) examined an international dataset for 20 countries. They reported that the null-hypothesis for the ET coefficient is rejected for 16 of the 20 countries. Important to note for future reference is that all these studies except Friedman (1979) tested the ET by regressing the survey expectation on the forward rate rather than the reverse. This setup generates empirical tests that are probably the most sensitive to coefficient bias from a time-varying term premium in the general form of the ET.

As mentioned before, there have been many more studies interested in examining the quality of survey forecasts of interest rates and testing the restrictions of the traditional rational expectations hypothesis. Spiwoks, Bedke and Hein (2008, Table 1) provide an extensive list of studies on the accuracy of interest rate forecasts. Most of these studies are for U.S. interest rates, using one of several surveys available for that country. More recently, Jongen and Verschoor (2008) report a large set of international results on forecasts of 3-month rates. For forecast horizons of 3 and 12 months, unbiasedness of survey forecasts is rejected for 7 and 15 of the 20 countries examined. The empirical results suggest a strong tendency for forecasts to follow adaptive or extrapolative mechanisms. For reasons that I discuss below, these studies are not directly relevant for this paper or direct tests of the expectations theory of the term structure in general. Arguably, the nature of the expectations mechanism is not directly relevant to the ET, and, more importantly, tests for unbiasedness and efficiency depend on one particular version of the rational expectations hypothesis that is probably not appropriate for real world empirical tests.

The structure of the remainder of this paper is as follows. Section 2 discusses two important issues in the theoretical foundations for empirical tests of the expectations theory. These two issues separate this paper from most of the previous studies. Section 3 presents the empirical model for the direct tests of the ET in this paper and discusses the problem of coefficient biases. Section 4 presents the data and empirical results. Section 5 concludes the paper. This paper finds that the ET is not in any material way rejected in a direct empirical test that 1) uses survey data to correct for the failing classic perfect-foresight-with-error rational expectations assumption, and that 2) estimates the version of the test equation that minimizes the bias from the unobserved stochastic term premium or an expectations measurement error.

2 THEORETICAL CONSIDERATIONS

Two major auxiliary assumptions are commonly employed in theoretical expositions and empirical tests of the ET. These relate respectively to the treatment of term premiums in interest rates and the concept of rational expectations. Both auxiliary assumptions can cause serious problems in the empirical tests of the ET when not treated carefully. In fact, this paper argues that the mainstream literature frequently presents an unnecessarily restrictive view of both the term premium hypothesis and the rational expectations hypothesis, and consequently the empirical results are biased towards a rejection of the ET. The arguments presented below are in fact not new, but remain rarely acknowledged in the mainstream literature.

The term premium hypothesis

Many studies have defined the ET as a theory that supposedly presumes that risk, liquidity or term premiums in interest rates are constant or even zero. This, however, is an unnecessarily narrow view of the ET. The constant term premium hypothesis should be regarded as an auxiliary assumption that suits an easy exposition of the theory and has historically facilitated easy preliminary empirical tests when actual term premiums are unobserved variables. Historically, in recognition of the original contributors to the ET, we must observe that authors were well aware of the risk or term premium and its influence. As early as Fisher

(1896, p.91) it has been noted that the interest rate relationships are subject to variations due to uncertainties of various kinds. Similarly, Lutz (1940, p.36), frequently credited for the pure ET model with a zero term premium, states as one of his initial assumptions that in his simple model everybody concerned knows what the future short-term rates will be; and with perfect foresight there is no risk. He discusses the effect of risk and liquidity premiums later, and actually seems to prefer a preferred habitat model (Lutz, p.48). Lutz even discusses the empirical bounds on arbitrage due to trading costs.

In essence, the ET is a fundamental no-arbitrage hypothesis for financial markets - a parity condition. In economic theory, adequate compensation for risk and liquidity are natural components of a no-arbitrage condition (as are arbitrage bounds due to trading costs), and there is nothing in economic theory that suggests that risk and liquidity premiums should be constant over time. Although tests of a zero or constant term premium might be called tests of the hypothetical pure form of the ET, a time-varying term premium is not a failure of the ET, but is simply an unfortunate complication for empirical research. As noted by Fama (1984, p.509), "Variation through time in these premiums can obscure the power of forward rates as predictors of future spot prices." At the same time, different premium models (i.e. liquidity: Hicks (1939), Lutz (1940), preferred habitat: Culbertson (1957), Modigliani and Sutch (1966), risk: Benninga and Protopapadakis (1983)) are not alternatives to the ET perse, but specific elaborations on a neglected but natural component of the general ET which includes a term premium.

The rational expectations hypothesis

Rational expectations are a natural complement to the hypothesis of efficient markets and zero-arbitrage conditions in financial markets. Unfortunately, most empirical studies have come to rely on one narrow view of the rational expectations hypothesis, i.e. perfect-foresight-with-error. The hoped-for advantage of the perfect-foresight hypothesis has been that in the absence of actual observations on expectations this hypothesis may allow researchers to substitute actual future interest rates for the unobserved expected interest rates of the model. Much of the empirical literature on the rational expectations hypothesis has shown that the perfect-foresight model is problematic and does not fit the real world very well. Leaving aside important empirical issues with respect to data (i.e. real-time data) and methodology (e.g. accounting for common macro forecast errors in panel studies), the fundamental reason for the weak performance of the perfect-foresight model is that this model assumes that the economic environment is stochastic stationary and that information is simply available and completely transparent. That is, we assume that all economic agents have come to know the true economic model and have sufficient information to know the future results of economic shocks. Again, in recognition of the original contributors to the rational expectations approach, the highly restrictive assumptions of the perfect-foresight-with-error model were already pointed out at an early stage (see Muth, 1961, p.316-7 and Lucas and Sargent, 1979, p.13). Specifically, Lucas and Sargent state "it has been only a matter of analytical convenience and not of necessity that equilibrium models have used the assumption of stochastically stationary shocks and the assumption that agents have already learned the probability distributions they face. Both of these assumptions can be abandoned, albeit at a cost in terms of the simplicity of the model."

Rational expectations represent a general concept of efficient and goal oriented behavior. Its precise implications depend on the characteristics of the economic environment. When real-world economies are more accurately described as stochastic non-stationary, i.e. experiencing unobserved regime shifts, structural breaks, etc., and economic agents are subject to uncertainty rather than risk,[5] and do not have full and completely transparent information, economic agents cannot have near perfect foresight. In fact, rational behavior implies that they learn rationally – possibly partly backward looking using a signal extracting approach (see for example Muth, 1960). Empirical tests of the rational expectations hypothesis that find ex post bias and inefficiency of forecasts do not necessarily signal a failure of rational expectations, but may simply reflect the need for a more realistic rational expectations model. It has even been argued that tests for bias and inefficiency are simply irrelevant.[6] Most importantly, we must conclude that when the perfect-

[5] The difference between uncertainty and risk refers to the discussion by Knight (1921, Chapter 7) and Keynes (1936, Chapter 12). See Meltzer (1982) for a comprehensive review of the issues.

[6] For a detailed critique of the common empirical rational expectations tests see Webb (1987).

foresight model of rational expectations fails, the common approach to substitute actual outcomes for expected values cannot be used to test the ET or economic theories in general.

Following the previous reconsideration of two fundamental auxiliary assumptions, we must conclude that appropriate direct tests of the expectations theory should take into account 1) the possibility of time-varying term premiums and 2) the weakness of the perfect-foresight expectations model. Discussion of the empirical tests below specifically takes into account the problem of coefficient biases and the tests will use survey expectations as a more accurate proxy for (rational) expectations.

2 SPECIFICATION OF THE EMPIRICAL TESTS

2.1 Forward rate tests

According to the ET the forward rate at time t calculated from a long maturity interest rate of n periods and a short maturity interest rate of k periods ($f_t^{(n,k)}$) should be equal to the expected future interest rate at time t+k for an interest rate of n-k maturity ($E_t r_{t+k}^{n-k}$) but corrected for a term premium (θ_t^{n-k}).

$$(1) \quad f_t^{(n,k)} = E_t r_{t+k}^{n-k} + \theta_t^{n-k}$$

There are many versions of empirical tests of the ET. However, studies using survey expectations of future interest rates generally focus on the following equation

$$(2a) \quad E_t r_{t+k}^{n-k} - r_t^{n-k} = \alpha_1 + \beta_1 (f_t^{(n,k)} - r_t^{n-k}) + \nu_{1t}^{n-k}$$

Equation (2a) includes a correction to the level of variables by subtracting the current value of the forecasted interest rate r_t^{n-k} left and right. This is done to solve econometric problems associated with non-stationary variables (i.e. spurious regression). Under the null hypothesis of the ET, coefficient β_1 is expected to equal one and the intercept α_1 represents the constant component of the unobserved term premium. The equation residual ν_t^{n-k} captures measurement errors and the stochastic component of the term premium.

It is important to note that the use of equation (2a) is a result of historical studies that have usually replaced expected value $E_t r_{t+k}^{n-k}$ with actual value r_{t+k}^{n-k} , invoking the traditional rational expectations hypothesis and therefore assuming that expectations are unbiased and efficient. Equation (2a) then allows the presumably random forecast error to be relegated to the equation residual. In principle, however, we can also write the test equation as

$$(2b) \quad f_t^{(n,k)} - r_t^{n-k} = \alpha_2 + \beta_2 (E_t r_{t+k}^{n-k} - r_t^{n-k}) + \nu_{2t}^{n-k}$$

Replacing expected value $E_t r_{t+k}^{n-k}$ with actual value r_{t+k}^{n-k} in equation (2b) would cause estimates of coefficient β_2 the be affected by an errors-in-variables bias, lowering its expected estimation value below its theoretical value of one and the bias depending on the size of the errors in forecasting interest rates. In practice and more generally, equations (2a) and (2b) both suffer from an errors-in-variables problem in the regression, but from a different origin (the imperfect measurement of interest rate expectations in equation (2b) and a stochastic but unobserved term premium in equation (2a)) and probably with a different severity.[7]

[7] The coefficient bias relating to ET regressions was introduced by Fama (1984) and Mankiw and Summers (1984). Attempts to solve the bias problem include estimation with instrumental variables (for example, Driffill, Psaradakis and Sola, 1997) and modeling the dynamics of the term premium (for example, Tzavalis and Wickens, 1997). Variants of equations (2a) and (2b) are sometimes used to examine relative contributions of expectations errors and predictable changes in the term premium or forward premium (see Froot, 1989).

Equation (2a) suffers from a bias caused by the presence of a stochastic term premium. If we assume for simplicity that the term premium is i.i.d. and if we ignore complications from non-zero correlations or serial correlations, it is easy to show that the Ordinary Least Squares estimate of equation (2a) results in an expected estimated coefficient β_1

$$(3) \quad \beta_1^{OLS} = \frac{\mathrm{var}(E_t r_{t+k}^{n-k} - r_t^{n-k})}{\mathrm{var}(E_t r_{t+k}^{n-k} - r_t^{n-k}) + \mathrm{var}(\theta_t^{n-k})} = \frac{1}{1 + \mathrm{var}(\theta_t^{n-k}) / \mathrm{var}(E_t r_{t+k}^{n-k} - r_t^{n-k})}$$

The expected value of the estimated coefficient β_1 equals its true theoretical value of one when the variance of the term premium is zero. The bias in β_1 towards zero increases when the variability of the predictable change in interest rates declines relative to the variance of the term premium. In the real world, the predictable change in short-term interest rates is heavily influenced by the type of monetary policy that exists in a particular country and a particular time period and is in practice frequently subjected to regime shifts. This insight provides a very natural interpretation of the many different results reported for various countries and sample periods (see Mankiw and Miron (1986) and Hardouvelis (1988)). Gerlach and Smets (1997) provide supporting empirical evidence in a multi-country perspective by showing how the predictability of future short-rates is related to the coefficient bias. The simulation studies of Rudebusch (1995) and Dotsey and Otrok (1995) show that statistical results are very sensitive to the time series model of the monetary policy rate. McCallum (1994) has shown that tests of the ET result in inconsistent estimates when the short-term monetary policy rate responds to changes in the long-short interest rate spread.

Following equation (2b) we assume that interest rate expectations $E_t r_{t+k}^{n-k}$ are not directly observed, but we assume the availability of survey expectations $S_t r_{t+k}^{n-k}$ as a proxy. The proxy expectations are, however, imperfect and contain an i.i.d. measurement error with mean zero, i.e. $S_t r_{t+k}^{n-k} = E_t r_{t+k}^{n-k} + \varepsilon_t^{n-k}$.

For equation (2b) we find that the OLS estimate of coefficient β_2 results in an expected value

$$(4) \quad \beta_2^{OLS} = \frac{\mathrm{var}(E_t r_{t+k}^{n-k} - r_t^{n-k})}{\mathrm{var}(E_t r_{t+k}^{n-k} - r_t^{n-k}) + \mathrm{var}(\varepsilon_t^{n-k})} = \frac{1}{1 + \mathrm{var}(\varepsilon_t^{n-k}) / \mathrm{var}(E_t r_{t+k}^{n-k} - r_t^{n-k})}$$

The estimated coefficient β_2 is expected to equal one when the variance of the measurement error in the surveyed expected rate change is zero. The bias towards zero increases when the variability of the predictable change in interest rates declines relative to the variance of the measurement error.

We can probably reduce the errors-in-variables problem in equation (2b) when we use the more direct and more accurate measurement of interest rate expectations that we obtain from surveys, rather than the noisier and empirically dubious actual rate outcomes under the classic rational expectations assumptions.[8] The choice between equation (2a) and equation (2b) as a more appropriate test of the ET depends on the relative size of the variances of the term premium and the expectations measurement error.

2.2 Roll-over spread tests

One other test of the ET is the so-called roll-over spread test. Building on equation (1) we can obtain the ET equation

[8] Survey expectations will remain imperfect measurements. Generally we do not directly observe the expectations of market participants but only of (analysts of) forecasting institutions, and even more fundamentally, we do not observe the relevant marginal buyer and seller on the market but only the average of forecasts. In addition, there are complications with the available information sets and precise timing when various survey participants prepare their responses.

$$(5) \quad R_t^n = (m/n)\sum_{k=0}^{n/m-1}[E_t r_{t+k}^m + \vartheta_t^k]$$

In the roll-over spread test we compare the average of a series of expected short-term m-period interest rates ($E_t r_{t+k}^m$) with the currently observed long-term n-period interest rate (R_t^n). In previous empirical studies this test has only been used with the auxiliary assumption of perfect-foresight rational expectations or an econometric forecasting model. When survey expectations are available it becomes possible to examine a more direct test of the ET that does not rely on a specific expectations model.

In the roll-over spread tests of the ET we expect to find coefficient β_3=1 in the regression equation

$$(6a) \quad (m/n)\sum_{k=0}^{n/m-1} E_t r_{t+k*m}^m - r_t^m = \alpha_3 + \beta_3 (R_t^n - r_t^m) + v_{3t}^n$$

As we did before with equations (2a) and (2b), we can reverse the left and right hand side variables and estimate a roll-over spread equation (6b) – not shown here – with the ET hypothesis for its slope coefficient β_4=1. In a similar way as before, it can be shown that the two equations (6a) and (6b) differ in their sensitivity to the stochastic term premium and the expectations measurement error. It is also important to note that equation (6a) for the roll-over spread and equation (2a) for the forward spread differ in their sensitivity to the stochastic term premium. The aggregate term premium included in the long-term rate is an average of a series of individual short-term forward rate term premiums. If the stochastic component in the forward term premiums averages to zero, the coefficient bias in equation (6a) may be less than the bias in equation (2a). Similar reasoning may be applied to a comparison between equations (2b) and (6b), where the measurement error in a series of survey interest rate expectations may average towards zero.

The discussion in this section has shown that empirical tests of the ET are generally subject to coefficient biases with different origins and with different severity. Empirical research must therefore examine these biases and interpret the empirical results carefully before drawing conclusions on the validity of the ET.

3 EMPIRICAL RESULTS

3.1 Data sources and issues

The main source of interest rate expectations used in this paper is the monthly survey by Consensus Economics for the G-7 countries and Western Europe. Every month a number of private sector forecasting institutions, normally from each of the selected countries, are surveyed and the individual forecasts and the arithmetic average, mean or 'consensus' forecast are reported. The interest rate forecasts are part of a range of macroeconomic forecasts, including inflation, real GDP growth and foreign exchange rates. The official survey date is the second Monday of every calendar month, although most survey participants send their responses on the Friday before the publication day. For the empirical analysis, Jongen et al. (2005) use the Friday before the official survey date, whereas MacDonald and Macmillan (1994) decided to use the 2nd Monday date. I follow the latter, assuming that this is the day that forecasts will have been worked into market prices. The forecast horizon for the monthly forecasts is the end of the month 3-months and 12-months ahead (i.e. the January survey forecasts are for the end of April and the end of January next year). Given the survey date, the true forecast horizon is actually slightly longer than 3 and 12 months. I use the monthly survey data for the 3-month interest rate for eight countries: the United States, the United Kingdom, Canada, Germany, Japan, Sweden, Norway and Switzerland.[9] The sample period is January 1996 to December 2007. The start is determined by the constraints of a subscription to the survey. The sample ends with 2007 in order to avoid more serious complications from the recent turmoil in financial markets.[10] Data

[9] Monthly survey data on 10-year government bond yields is also available. Semi-annual surveys in April and October include long-term forecasts for the 10-year government bond yield for 2 to 10 years ahead.

[10] Problems related to the U.S. subprime mortgage market started to emerge in the first half of 2007. In January 2008 the Federal Reserve responded to the mounting problems in the financial system with unusually large interest rate cuts of 75bp and 50bp.

for Norway and Switzerland were added to the Consensus Economics survey in June 1998. For Canada a longer data set is available that starts in October 1989.

The money market term structure data are in most cases obtained from Thompson Reuters Datastream. Selection of the appropriate time series variables involved some careful consideration of competing data sources and data definitions, as well as interpreting the moment of some changes in the forecasted rate (see the data appendix). Use of the money market rates is relatively straightforward, taking into account money market conventions of simple interest rate calculations. The only special adjustment made to the data is the conversion of U.S. Treasury bill discount rates into investment yields using the formula $r_y = r_d\ [100 / (100 - r_d\ m/360)]$ where r_d and m are the discount rate and the maturity of the T-bill (m=91 and 182). Canadian T-bills are already reported on an investment yield basis, although they are annualized using 365 days (as are U.K. rates). In the case of Japan, data availability necessitates that we change the interest rate definition from the CD rate indicated in the survey to the Euro-yen and interbank rate, but these rates match the survey data very well. The 3-month forward rate 3-months ahead is calculated from the 6-month (long) and 3-month (short) interest rates using the formula $f^{6,3} = 400*((1+R^6/200) / (1+r^3/400)-1)$. As mentioned above, there is a slight mismatch between the forecast horizon (3+ months) and the available term structure data (generally 1, 3, 6 and 12 months). Unfortunately, this data problem, as well as some other refinements such as compounding and day-counting will remain in the realm of approximation and measurement error.

In addition to the monthly forecasts, the Consensus Economics Forecasts survey also provides quarterly (March, June, September, December) forecasts for 3-month interest rates with quarterly intervals up to seven quarters ahead. As far as I know, these data have not been used before to test the ET. In the roll-over spread tests, I use the current 3-month rate and the expected 3-month rates for Q+1 to Q+3 to match survey expectations with the actual 12-month interest rate from the money market term structure. Taking into account money market conventions, the 12-month rate implied by the survey forecasts is $(1+R^{12}/100) = (1+r^3/400)\ (1+s^3_{+3}/400)\ (1+s^3_{+6}/400)\ (1+s^3_{+9}/400)$, where R, r, and s refer to the current long rate and short rate and the survey forecasts. Calculations for the U.S. Treasury bill data are slightly different because they are provided on a discount rate basis and because the series for the 12-month Treasury bill secondary market rate has been discontinued. The U.S. term structure data are in this case obtained from the Treasury constant maturity yield curve. The constant maturity data refer to bond yields with semi-annual compounding. After the conversion from discount rate to investment yield as described earlier, the 3-month T-bill rates imply a 12-month constant maturity bond yield (y^{12}) of $(1+y^{12}/200)^2 = (1+r^3/400)\ (1+s^3_{+3}/400)\ (1+s^3_{+6}/400)\ (1+s^3_{+9}/400)$.

To examine the robustness of the results to changes in the sample period and survey, additional forecasts for the 3-month U.S. Treasury bill rate are also taken from the U.S. Survey of Professional Forecasters. Quarterly interest forecasts in the SPF start with the survey of the third quarter of 1982. At least two assumptions must be made when using these survey data. First, there are no explicit survey dates mentioned in the SPF source for before 1992Q3 when the Federal Reserve Bank of Philadelphia took over responsibility for the survey. I maintain the assumption that the survey date for 1982-1992 has been the 20th of the middle month of every quarter. If the 20th falls on Saturday or Sunday the survey date is moved forward to Monday in order to obtain the relevant term structure data. The second problem is that the SPF forecasts refer to the quarterly average rate, rather than the rate on a specific date. One solution employed in previous studies is to assume that the quarterly average can be used for the middle of the mid-quarter month (i.e. 15th). In matching the 3-month forward rate 3-months ahead and the survey expectation the assumption is that the one-quarter ahead quarterly forecast can be treated as the expected rate exactly 3-months ahead from the survey date.

Recent studies have argued that consensus forecasts may be subject to an aggregation bias and that it is worthwhile to exploit cross-sectional information from surveys, particularly when time series data is limited. Although forecasts from individual forecast institutions are available in the survey, I will only use the average (= mean) consensus forecast. Two comments are relevant in this respect. First, the median of forecasts may be preferable in order to eliminate possible sensational outlier forecasts, but this measure is not available without a substantial data collection effort. Also, outliers that would affect the mean are much more likely to be present in longer horizon forecasts whereas this study is limited to relatively short horizon forecasts. Second, in addition to practical data collection considerations, the main reason for not using a pooled estimation strategy is that it incorrectly assumes that market efficiency requires all individual forecasts to be rational and efficient. Market efficiency, however, relates to average market behavior or more specifically the influence of the all-important marginal investor. Observing individual irrationality in survey participants simply cannot disprove market efficiency and/or the ET.

3.2 Results on the forward rate test

Table 1 provides some simple descriptive statistics on the three variables of importance to the empirical analysis: the forward spread $f_t^{6,3} - r_t^3$ obtained from the observed term structure data, the expected rate change $S_t r_{t+3}^3 - r_t^3$ obtained from the survey data, and the implied term premium $f_t^{6,3} - S_t r_{t+3}^3$.[11] The average term premium is significant and positive for the U.S., Germany and Canada. For the U.K. the average term premium is negative but not significant. More important for this study are the measures of variability, because the earlier discussion suggested a very important role for the relative variability of expected rate changes and the term premium in the bias of estimated coefficients. The normal range of ratios between variances of the term premium and the expected rate change seems to be between 0.6 (Norway) and 0.9 (UK, Canada and Germany). Japan (2.2) and Switzerland (1.4) are outliers in this group of countries. These ratios indicate a high potential for substantial coefficient bias in estimates of equation (2a). Estimates based on equation (2b) depend on the variability of the measurement error in the survey expectations, but the measurement error remains unobserved. In fact, in Table 1 the expectations measurement error is a part of the calculated term premium $f_t^{6,3} - S_t r_{t+3}^3$. At this stage the maintained hypothesis will be that the term premium dominates the measurement error. Finally, the Jarque-Bera test statistic suggests that regression estimation and tests may suffer from deviations from the normal distribution. The underlying statistics indicate that excess kurtosis is the main problem.

Results for the estimation of equations (2a) and (2b) are presented in Table 2. Equations (2a) and (2b) are estimated using the ordinary least squares method.[12] The monthly data frequency together with the multi-month forecast horizon introduces an overlapping observations problem, in which case residual serial correlation is expected. The estimation results therefore include Newey-West HAC standard errors that correct for serial correlation as well as heteroskedasticity. Consistent with previous studies we find that in equation (2a) the hypothesis β_1=1 is strongly rejected for all eight countries. The forward rate in the money market term structure is found to be a biased indicator of market expectations. However, the results on the hypothesis β_2=1 in equation (2b) provide a very different message. The ET hypothesis β_2=1 is not rejected, except in the case of Switzerland. On a secondary level, the weakest results for the ET are found for the U.S. Treasury bill market (p-value 5.3%) and Japan (p-value 9.1%). The overall results in Table 2 thus suggest a strong reversal in the performance of the ET by taking into account the econometric problem of coefficient bias.

The relatively weak performance of the ET in the U.S. Treasury bill market is not entirely surprising given the previous results available from cross-country and cross-market studies. Empirical results in the literature show that using U.S. Commercial Paper or Eurodollar market rates tends to be more favorable to the ET. Unfortunately, the survey data for these rates are not available and we cannot examine this particular issue in this paper. Why the U.S. Treasury bill market is such a special case also remains a question to be answered, but many have suggested an effect from the importance of T-bills in U.S. central bank operations and banks' liquidity requirements.

Next, a closer examination of the regression results and the time series of the underlying data suggests that several countries experienced sizable shocks to the term premium that could have an effect on the regression estimates. The U.K., Japan, Germany, Sweden, Norway and Switzerland show a large increase in the observed term premium in three months centered on September or August 1999. Perhaps this can be related

[11] Each of the variables was dutifully examined for unit roots and stationarity using Ng-Perron and KPSS test statistics (results available as appendix). As is not uncommon, the results of the tests are not always in agreement and require some judgment. Based on the criterion that at least one of the two tests rejects the unit root or does not reject stationarity the combined results suggest that the three variables can be considered stationary,

[12] Alternative estimation methods such as GMM or IV depend on uncertain validity of available instrument variables.

to the year 2K problem that affected financial markets at the end of 1999. Japan had a further similar experience around October 1998 and this may be related to discussions of a bank bailout plan in Japan. This issue of specific term premium shocks may require further discussion and analysis. Further examination also shows that the strong statistical rejection of the ET with the Swiss data disappears when the sample period is limited to 2000-2007. We note that the Swiss dataset consists of only a small sample period that starts in 1998M06 and the Swiss result may therefore be particularly sensitive to time specific influences, but changing the data sample arbitrarily is of course not an adequate solution. Addition of future survey data will have to provide a more robust analysis of the Swiss evidence on the ET.

For the longer sample that is available for Canada, descriptive statistics can be found in Panel B of Table 1. The corresponding regression results are presented in Panel B of Table 2. We find that the results for the longer sample are similar to the results based on the shorter sample in Table 1.

Descriptive statistics for the U.S. Survey of Professional Forecasters can be found in Panel C of Table 1. The regression results are presented in Panel C of Table 2. The results are very similar to the results for the shorter sample used in Table 1. However, the probability value for the t-test on $\beta_2=1$ is now 3.3%, and this means a rejection of the ET on a conventional significance level of 5% rather than 10%. Further examination of the data and regression estimates shows that over longer time periods, in particular before 1995, the calculated term premium for the U.S. Treasury bill market is not on average constant. The U.S. T-bill term premium has gradually decreased over time from at least 1981. This apparent (broken) trend in the term premium may be an explanation for the coefficient bias that we find for the U.S. in the long sample estimates. In an attempt to capture this trending behaviour of the term premium I re-estimated equations (2a) and (2b) with an intercept dummy and linear trend. The estimated beta coefficients move towards one, no longer rejecting the hypothesis $\beta_1=1$ in equation (2a) but still rejecting $\beta_2=1$ in equation (2b). The results suggest that in these regressions the measurement error associated with the SPF data dominates the stochastic component of the term premium.

3.3 Results on the roll-over spread test

Results for equation (6a) and its complement (6b) are presented in Table 3. Again, as in Table 2 we find that in equation (6a) the hypothesis $\beta_3=1$ is strongly rejected in all eight countries. However, the results on the hypothesis $\beta_4=1$ in equation (6b) again provide the opposite message. The ET hypothesis $\beta_4=1$ is not rejected for any of the eight countries. Compared to the forward rate test in Table 2 the roll-over spread test also eliminates or avoids the troublesome result for Switzerland, emphasizing the importance of selecting a specific test type. The weakest result in Table 3, as usual, is for the U.S. Treasury bill rates (β_4 p-value equals 5.9%).

Panel B of Table 3 presents additional results for the roll-over spread test using the longer time series from the U.S. SPF survey. The results are similar to Table 2, panel C. The two hypotheses $\beta_3=1$ and $\beta_4=1$ are both rejected using the SPF data. After taking again into account the long-term broken trend in the T-bill term premium the hypothesis $\beta_3=1$ is not rejected on a much larger p-value than was the case for $\beta_1=1$ in Table 2. Still, the alternative test $\beta_4=1$ remains strongly rejected, and again the suggestion based on our discussion of the theoretical model is that the measurement error for market expectations introduced by using the SPF data is important.

4 SUMMARY AND CONCLUSIONS

The expectations theory of the term structure of interest rates has a tarnished reputation in the empirical literature. Although the results of empirical tests are actually mixed when viewed across sample periods, markets and countries, the strong focus on studies using U.S. data causes that many economists regard the ET as empirically rejected.

This paper has highlighted three major issues with respect to the existing empirical literature. Two issues relate to two common auxiliary theoretical assumptions. First, a time-varying term premium may be an empirical nuisance and may represent a rejection of the simplified pure ET found in finance textbooks, but it does not represent a rejection of the general ET. The general ET is appropriately defined as a no-arbitrage condition or parity relationship that naturally includes an appropriate compensation for risk. Economic theory does not suggest that risk and risk premiums are constant over time. Second, the specific variant of the rational expectations hypothesis commonly used in empirically studies, i.e. perfect-foresight-with-error, is empirically weak and frequently rejected by the data. But rejection of the perfect-foresight hypothesis is not a rejection of the general rational expectations model, a model that includes rational learning and signal extraction in stochastic non-stationary economies. These points are not particularly new but are rarely acknowledged in the mainstream literature. The third empirical issue is that both time-varying term premiums and measurement errors in interest rate expectations cause a coefficient bias in the empirical tests of the ET. However, different tests exhibit different sensitivity to the sources of the coefficient bias and should be evaluated carefully.

Direct empirical tests of the ET must thus take into account the problem of coefficient bias, from time-varying term premiums, and require a better proxy for expected interest rates, such a survey data. The empirical results in this paper show that when using survey data on interest rate expectations and when taking into account the sensitivity of tests to coefficient bias, the expectations theory fits the term structure data very well in the eight countries examined in this paper.

REFERENCES

Batchelor, R. (1995), 'On the importance of the term premium in the T-bill market,' Applied Financial Economics, vol.5 (4) August: 235-42.

Benninga, S. and A. Protopapadakis (1983), "Real and nominal interest rates under uncertainty: The Fisher theorem and the term structure," Journal of Political Economy, vol. 91(5) October: 856-67.

Campbell, J.Y. (1995), 'Some lessons from the yield curve,' Journal of Economic Perspectives, vol.9 (3) Summer: 129-52.

Choi, S. and M.E. Wohar (1991), 'New evidence concerning the expectations theory for the short end of the maturity spectrum,' Journal of Financial Research, vol.14 (1) Spring: 83-92.

Cook, T. and T. Hahn (1990), 'Interest rate expectations and the slope of the money market yield curve,' FRB Richmond Economic Review, vol.76 September/October: 3-26.

Culbertson, J.M. (1957), 'The term structure of interest rates,' Quarterly Journal of Economics, vol.71 (4) November: 485-517.

Dominguez, E. and A. Novales (2000), 'Testing the expectations hypothesis in Eurodeposits,' Journal of International Money and Finance, vol.19 (5) October: 713-36.

Dotsey, M. and C. Otrok (1995), 'The rational expectations hypothesis of the term structure, monetary policy, and time-varying term premia,' FRB Richmond Economic Quarterly, Winter: 65-81.

Downing, C. and S. Oliner (2007), 'The term structure of commercial paper rates,' Journal of Financial Economics, vol.83 (1) January: 59-86.

Drakos, K. (2002), 'A daily view of the term structure dynamics: Some international evidence,' De Economist, vol.150 (1): 41-52.

Driffill, J., Z. Psaradakis and M. Sola (1997), 'A reconciliation of paradoxical empirical results on the expectations model of the term structure,' Oxford Bulletin of Economics and Statistics, vol.59 (1) February: 29-42.

Driffill, J., Z. Psaradakis and M. Sola (1998), 'Testing the expectations hypothesis of the term structure using instrumental variables,' International Journal of Finance and Economics, vol.3 (4) October: 321-5.

Fama, E.F. (1984), 'The information in the term structure,' Journal of Financial Economics, vol.13 (4) December: 509-28.

Fama, E.F. (1991), 'Efficient capital markets: II,' Journal of Finance, vol.46 (5) December: 1575-1617.

Fisher, I. (1896), 'Appreciation and interest: a study of the influence of monetary appreciation and depreciation on the rate of interest with applications to the bimetallic controversy and the theory of interest. Publications of the American Economic Association, vol.11.

Friedman, B.M. (1979), 'Interest rate expectations versus forward rates: Evidence from an expectations survey,' Journal of Finance, vol.34 (4) September: 965-73.

Froot, K.A. (1989), 'New hope for the expectations hypothesis of the term structure of interest rates,' Journal of Finance, vol.44 (2) June: 283-305.
Gerlach, S. and F. Smets (1997), 'The term structure of Euro-rates: Some evidence in support of the expectations hypothesis,' Journal of International Money and Finance, vol.16 (2) April: 305-21.
Hardouvelis, G.A. (1988), 'The predictive power of the term structure during recent monetary regimes,' Journal of Finance, vol.43 (2) June: 339-56.
Hardouvelis, G.A. (1994), 'The term structure spread and future changes in long and short rates in the G7 countries: Is there a puzzle?' Journal of Monetary Economics, vol.33 (2) March: 255-83.
Hicks, J. (1939), Value and Capital. Clarendon Press.
Hsu, C. and P. Kugler (1997), 'The revival of the expectations hypothesis of the US term structure of interest rates,' Economics Letters, vol.55 (1) August: 115-20.
Jondeau, E. and R. Ricart (1999), 'The expectations hypothesis of the term structure: Tests on US, German, French, and UK Euro-rates,' Journal of International Money and Finance, vol.18 (5) October: 725-50.
Jongen, R. and W.F.C. Verschoor (2008), 'Further evidence on the rationality of interest rate expectations,' Journal of International Financial Markets, Institutions and Money, vol.18 (5) December: 438-48.
Jongen, R., W.F.C. Verschoor and C.C.P. Wolff (2005), 'Time variation in term premia: International evidence,' CEPR Discussion paper no.4959.
Jorion, P. and F.S. Mishkin (1991), 'A multicountry comparison of term-structure forecasts at long horizons,' Journal of Financial Economics, vol.29 (1) March: 59-80.
Kane, E.J. and B.G. Malkiel (1967), 'The term structure of interest rates: An analysis of a survey of interest-rate expectations, Review of Economics and Statistics, vol.49 (3) August: 343-355.
Keynes, J.M. (1936), The General Theory of Employment, Interest, Money. Harcourt Brace.
Knight, F.H. (1921), Risk, Uncertainty and Profit. Houghton Mifflin.
Kugler, P. (1988), 'An empirical note on the term structure and interest rate stabilization policies,' Quarterly Journal of Economics, vol.103 (4) November: 789-92.
Longstaff, F.A. (2000), 'The term structure of very short-term rates: New evidence for the expectations hypothesis,' Journal of Financial Economics, vol.58 (3) December: 397-415.
Lucas, Jr., R.E. and T.J. Sargent (1979), 'After Keynesian macroeconomics,' FRB Minneapolis Quarterly Review, vol.3 (2) Spring: 1-6.
Lutz, F.A. (1940), 'The structure of interest rates,' Quarterly Journal of Economics, vol.55 (1) November: 36-63.
MacDonald, R. and P. Macmillan (1994), 'On the expectations view of the term structure, term premia and survey-based expectations,' Economic Journal, vol.104 (426) September: 1070-86.
MacDonald, R., 'Expectations formation and risk in three financial markets: Surveying what the surveys say,' Journal of Economic Surveys, vol.14 (1) February: 69-100.
Mankiw, N.G. (1986), 'The term structure of interest rates revisited," Brookings Papers on Economic Activity, no.1: 61-110.
Mankiw, N.G. and J.A. Miron (1986), 'The changing behavior of the term structure of interest rates,' Quarterly Journal of Economics, vol.101 (2) May: 211-28.
Mankiw, N.G. and L.H. Summers (1984), 'Do long-term interest rates overreact to short-term interest rates?' Brookings Papers on Economic Activity, no.1: 223-42.
McCallum, B.T. (1994), 'Monetary policy and the term structure of interest rates,' NBER Working paper no.4938.
Meltzer, A.H., 'Rational expectations, risk, uncertainty, and market responses,' in P. Wachtel (ed.) Crises in the Economic and Financial Structure. Lexington, Mass.: Lexington Books, 1982: 3-22.
Modigliani, F. and R. Sutch (1966), 'Innovations in interest rate policy,' American Economic Review, vol.56 (1/2) March-May: 178-97.
Muth, J.F. (1960), 'Optimal properties of exponentially weighted forecasts,' Journal of the American Statistical Association, vol.55 (290) June 1960: 299-306.
Muth, J.F. (1961), 'Rational expectations and the theory of price movements,' Econometrica, vol.29 (3) July: 315-35.
Pesaran, M.H. and M. Weale (2006), 'Survey expectations,' in G. Elliott, C. Granger and A. Timmermann (eds.) Handbook of Economic Forecasting. North-Holland.

Rudebusch, G.D. (1995), 'Federal Reserve interest rate targeting, rational expectations, and the term structure.' Journal of Monetary Economics, vol.35 (2) April: 245-274.
Shiller, R.J. (1990), 'The term structure of interest rates' in B.M. Friedman and F.H. Hahn (eds.) Handbook of Monetary Economics, vol.1. Amsterdam: North-Holland: 627-722.
Siklos, P.L. and M.E. Wohar (1996), 'Cointegration and the term structure: A multicountry comparison,' International Review of Economics and Finance, vol.5 (1) : 21-34.
Simon, D.P. (1990), 'Expectations and the Treasury bill – Federal funds rate spread over recent monetary policy regimes,' Journal of Finance, vol.45 (2) June: 567-77.
Spiwoks, M., N. Bedke and O. Hein (2008), 'Forecasting the past: the case of US interest rate forecasts,' Financial Markets and Portfolio Management, vol.22 (4) December: 357-79.
Tzavalis, E. and M.R. Wickens (1997), 'Explaining the failures of the term spread models of the rational expectations hypothesis of the term structure,' Journal of Money, Credit, and Banking, vol.29 (3) August: 364-80.
Webb, R.H. (1987), 'On the irrelevance of tests for bias in series of macroeconomic forecasts,' FRB Richmond Economic Review, vol.73 (6) November/December: 3-9.

TABLE 1. Simple descriptive statistics

		Forward spread $f_t^{6,3} - r_t^3$	Expected change $S_t r_{t+3}^3 - r_t^3$	Term premium $f_t^{6,3} - S_t r_{t+3}^3$	Forecast error $r_{t+3}^3 - S_t r_{t+3}^3$
Panel A. Sample period 1996M01-2007M12, 144 monthly obs.					
US	Mean	0.252	0.225	0.027	-0.291
	St.dev.	0.237	0.192	0.172	0.547
	JB	[0.480]	[0.095]	[0.000]	[0.000]
UK	Mean	0.028	0.018	0.010	-0.010
	St.dev.	0.250	0.190	0.179	0.403
	JB	[0.001]	[0.000]	[0.000]	[0.020]
CN	Mean	0.260	0.151	0.109	-0.190
	St.dev.	0.266	0.189	0.175	0.529
	JB	[0.000]	[0.110]	[0.007]	[0.000]
JP	Mean	0.077	0.052	0.025	-0.026
	St.dev.	0.165	0.105	0.157	0.160
	JB	[0.000]	[0.001]	[0.000]	[0.000]
GE	Mean	0.070	0.010	0.061	0.027
	St.dev.	0.204	0.149	0.142	0.304
	JB	[0.124]	[0.005]	[0.000]	[0.723]
SWE	Mean	0.119	0.070	0.048	-0.147
	St.dev.	0.253	0.201	0.166	0.410
	JB	[0.000]	[0.018]	[0.000]	[0.001]
NO*	Mean	-0.153	-0.202	0.049	0.217
	St.dev.	0.414	0.327	0.250	0.715
	JB	[0.000]	[0.011]	[0.046]	[0.000]
SWI*	Mean	0.163	0.115	0.049	-0.089
	St.dev.	0.214	0.164	0.191	0.395
	JB	[0.000]	[0.000]	[0.000]	[0.001]
Panel B. Sample period 1989M10-2007M12, 219 monthly obs.					
CN	Mean	0.183	0.018	0.165	-0.153
	St.dev.	0.467	0.357	0.360	0.838
	JB	[0.000]	[0.000]	[0.000]	[0.000]
Panel C. Sample period 1981Q3 – 2007Q4, 106 quarterly obs.					
US	Mean	0.363	0.102	0.261	-0.174
	St.dev.	0.428	0.386	0.375	0.799
	JB	[0.000]	[0.000]	[0.000]	[0.000]

Notes: * Sample for Norway and Switzerland starts in 1998M06 with 115 obs. Panel A uses the monthly Consensus Economics survey data. Panel B for Canada uses an extended sample from the same survey. Panel C for the U.S. uses the quarterly Survey of Professional Forecasters.
JB is the probability value of the Jarque-Bera test of normality.

TABLE 2. Regression estimates of the forward rate test

$$(2a)\quad S_t r_{t+3}^3 - r_t^3 = \alpha_1 + \beta_1 (f_t^{(6,3)} - r_t^3) + \nu_{1t}^3$$

$$(2b)\quad f_t^{(6,3)} - r_t^3 = \alpha_2 + \beta_2 (S_t r_{t+3}^3 - r_t^3) + \nu_{2t}^3$$

					Probability values of t-tests		
		α	β	R^2	α=0 p-value	β=0 p-value	β=1 p-value
Panel A. Sample period 1996M01-2007M12, 144 monthly obs.							
US	Eq2a	0.082	0.566	0.486			
		(0.020)	(0.073)		0.000	0.000	0.000
	Eq2b	0.059	0.859	0.486			
		(0.030)	(0.072)		0.053	0.000	0.053
UK	Eq2a	0.003	0.534	0.491			
		(0.020)	(0.095)		0.894	0.000	0.000
	Eq2b	0.012	0.918	0.491			
		(0.025)	(0.104)		0.637	0.000	0.432
CN	Eq2a	0.011	0.537	0.569			
		(0.023)	(0.047)		0.625	0.000	0.000
	Eq2b	0.100	1.060	0.569			
		(0.021)	(0.141)		0.000	0.000	0.671
JP	Eq2a	0.033	0.251	0.155			
		(0.016)	(0.059)		0.037	0.000	0.000
	Eq2b	0.045	0.617	0.155			
		(0.013)	(0.225)		0.001	0.007	0.091
GE	Eq2a	-0.027	0.523	0.514			
		(0.014)	(0.073)		0.063	0.000	0.000
	Eq2b	0.061	0.983	0.514			
		(0.019)	(0.126)		0.002	0.000	0.891
SWE	Eq2a	-0.001	0.599	0.573			
		(0.021)	(0.107)		0.964	0.000	0.000
	Eq2b	0.052	0.956	0.573			
		(0.022)	(0.081)		0.022	0.000	0.591
NO*	Eq2a	-0.106	0.628	0.634			
		(0.035)	(0.049)		0.003	0.000	0.000
	Eq2b	0.051	1.009	0.634			
		(0.028)	(0.181)		0.075	0.000	0.960
SWI*	Eq2a	0.050	0.397	0.269			
		(0.024)	(0.108)		0.038	0.000	0.000
	Eq2b	0.086	0.677	0.269			
		(0.039)	(0.129)		0.030	0.000	0.014
Panel B. Sample period 1989M10-2007M12, 219 monthly obs.							
CN	Eq2a	-0.073	0.495	0.419			
		(0.034)	(0.069)		0.033	0.000	0.000
	Eq2b	0.168	0.847	0.419			
		(0.043)	(0.191)		0.000	0.000	0.425
Panel C. Sample period 1981Q3 – 2007Q4, 106 quarterly obs.							
US	Eq2a	-0.088	0.523	0.337			
		(0.065)	(0.169)		0.179	0.003	0.006
	Eq2b	0.297	0.644	0.337			
		(0.059)	(0.165)		0.000	0.000	0.033
US	Eq2a (lt)	0.003	0.697	0.477			
		(0.047)	(0.179)		0.946	0.000	0.093
	Eq2b (lt)	0.113	0.675	0.588			
		(0.038)	(0.108)		0.004	0.000	0.003

Notes: * Sample for Norway and Switzerland starts 1998M06. Newey-West HAC standard errors in parentheses (lag truncation=4).
Panel A and B use the Consensus Economics survey. Panel C uses the quarterly U.S. Survey of Professional Forecasters. Equations 2a and 2b (lt) include a linear trend and intercept dummy for the period 1981Q3-1997Q4 to capture the decline in the U.S. T-bill term premium.

TABLE 3. Roll-over spread regressions

$$(6a)\quad (1/4)\sum_{k=0}^{3} S_t r_{t+k*3}^{3} - r_t^{3} = \alpha_3 + \beta_3\,(R_t^{12} - r_t^{3}) + \nu_{3t}^{12}$$

$$(6b)\quad R_t^{12} - r_t^{3} = \alpha_4 + \beta_4\,((1/4)\sum_{k=0}^{3} S_t r_{t+k}^{3} - r_t^{3}) + \nu_{4t}^{12}$$

		α	β	R^2	Probability values of t-tests α=0 p-value	β=0 p-value	β=1 p-value
Panel A. Quarterly sample period 1996Q1-2007Q4							
US	Eq6a	0.102	0.403	0.254			
		(0.028)	(0.096)		0.001	0.000	0.000
	Eq6b	0.145	0.629	0.254			
		(0.074)	(0.192)		0.056	0.002	0.059
UK	Eq6a	0.040	0.632	0.568			
		(0.045)	(0.121)		0.384	0.000	0.004
	Eq6b	0.036	0.899	0.568			
		(0.034)	(0.102)		0.297	0.000	0.326
CN	Eq6a	0.113	0.431	0.496			
		(0.037)	(0.050)		0.003	0.000	0.000
	Eq6b	0.067	1.150	0.496			
		(0.054)	(0.166)		0.206	0.000	0.371
JP	Eq6a	0.032	0.449	0.364			
		(0.020)	(0.151)		0.115	0.005	0.001
	Eq6b	0.035	0.810	0.364			
		(0.011)	(0.144)		0.002	0.000	0.194
GE	Eq6a	0.014	0.557	0.636			
		(0.022)	(0.087)		0.535	0.000	0.000
	Eq6b	0.039	1.141	0.636			
		(0.041)	(0.251)		0.353	0.000	0.576
SWE	Eq6a	0.079	0.555	0.581			
		(0.032)	(0.076)		0.016	0.000	0.000
	Eq6b	0.030	1.047	0.581			
		(0.032)	(0.121)		0.346	0.000	0.698
NO*	Eq6a	0.002	0.710	0.796			
		(0.040)	(0.060)		0.956	0.000	0.000
	Eq6b	-0.014	1.120	0.796			
		(0.044)	(0.115)		0.758	0.000	0.303
SWI*	Eq6a	0.147	0.286	0.265			
		(0.029)	(0.103)		0.000	0.009	0.000
	Eq6b	0.046	0.926	0.265			
		(0.106)	(0.354)		0.667	0.013	0.836
Panel B. Quarterly sample period 1981Q3 – 2007Q4, 106 quarterly obs.							
US	Eq6a	-0.117	0.533	0.333			
		(0.076)	(0.158)		0.128	0.001	0.004
	Eq6b	0.347	0.624	0.333			
		(0.060)	(0.139)		0.000	0.000	0.008
US	Eq6a (lt)	0.000	0.796	0.553			
		(0.051)	(0.183)		0.944	0.000	0.267
	Eq6b (lt)	0.093	0.685	0.671			
		(0.034)	(0.081)		0.007	0.000	0.000

Notes: * Sample for Norway and Switzerland starts 1998Q2. Newey-West HAC standard errors in parentheses (lag truncation=3).

Panel A uses Consensus Economics survey data. The September 2006 data were not available and are treated as a missing obs. Panel B uses U.S. Survey of Professional Forecasters. Equations 6a and 6b (lt) include a linear trend and intercept dummy for the period 1981Q3-1997Q4 to capture the decline in the US T-bill term premium.

APPENDIX. Data sources and definitions

Consensus Economics Forecasts
Although the Consensus Economics forecast survey (CF) states the specific interest rate that is forecasted, creating the term structure dataset has at least three complications. First, sometimes alternative sources or definitions exist. Second, term structure data may not be available. Third, the survey date data sometimes suggest that a shift in source or definition has occurred at a date different from the one indicated in the publication. The following series have been selected by matching the published survey date observation to the available data series.
United States: 3-month Treasury bill rate (secondary market, discount rate)
Data: Although CF data seem to be sourced from the Financial Times, US forecasting institutions are more likely to relate to the Federal Reserve H.15 time series. Datastream FRTBS3M, FRTBS6M
U.S. Treasury bill discount rates are converted into coupon equivalent or investment yield interest rates using the formula $r_y = r_d\ [100 / (100 - r_d\ m/360)]$ where r_d and m are the discount rate and maturity of the T-bill (m=91 for 3 months and 182 for 6 months)
For the quarterly analysis using the 12-month roll-over approach the Treasury constant maturity data are used because the FRTBS1Y is unavailable after August 2001. The constant maturity data are semi-annual compounded bond yields. Datastream FRTCM3M, FRTCM1Y.
United Kingdom: 3-month Interbank rate
Source Tradition. Data: Datastream LDNIB3M, LDNIB6M, LDNIB1Y. This series fits the CF data slightly better than the Libid/Libor rate available from the Bank of England.
Canada: 3-month Treasury bill rate (yield basis)
Data: Before 31Dec1993 only the weekly auction rate is available, but two series exist: one for the rate measured on Wednesdays, the other for the most recent Thursday auction. The Wednesday rate is used as the representative series. Source Cansim/Bank of Canada using Datastream CN13884, CN13885, CN13886. From 31Dec1993 the secondary market series. Source Bank of Canada using Datastream CNTBB3M, CNTBB6M, CNTBB1Y.
Japan: 3-month CD rate
Data: There is no term structure data available for the CD rate. Best match with the published CF data are: until Nov1998 Euro-Yen rate London. Source: FT/ICAP/TR using Datastream ECJAP3M, ECJAP6M, ECJAP1Y. From Dec1998 Tokyo Interbank Offered Rate (TIBOR). Source Japan Bankers Association through Datastream JPIBK3M, JPIBK6M, JPIBK1Y.
Germany: 3-month Euro-DM rate, from Jan1999 3-month Euribor
Data: Until Dec1998 Euro-Mark rate London. Source: FT/ICAP/TR using Datastream ECWGM3M, ECWGM6M, ECWGM1Y. From Jan1999 Euribor. Datastream EIBOR3M, EIBOR6M, EIBOR1Y.
Sweden: 3-month Euro-Krona, from Mar2004 3-month Deposit rate, from Aug2005 3-month Interbank rate
Data: Euro-Krona - until Mar1997 Source Goldman-Sachs Datastream GSSEK3M, GSSEK6M, GSSEK1Y and from Apr1997 source FT/ICAP/TR Datastream ECSWE3M, ECSWE6M, ECSWE1Y. The GSSEK series do not match during Nov1998-Dec1999. Deposit/Interbank rate: from Mar2004 Stockholm Interbank Offer Rate (STIBOR). Data: Source Bank of Sweden using Datastream SIBOR3M, SIBOR6M, SIBOR1Y. The Deposit rate data seems to suffer from infrequent trading or updating from approx. 2001.
Norway: 3-month Euro-Krone, from Mar2004 3-month Interbank rate
Data: Euro-Krone, until Jun1999. Source FT/ICAP/TR using Datastream ECNOR3M, ECNOR6M, ECNOR1Y. Interbank: Interbank (effective) rate fits the CF data best from Jul1999. Source Norges Bank using Datastream NWIBE3M, NWIBE1Y; corresponding 6-month (effective) rate obtained from the Norges Bank website.
Switzerland: 3-month Euro-SwissFranc
Data: Euro-SwissFranc rate London. Source FT/ICAP/TR through Datastream ECSWF3M, ECSWF6M, ECSWF1Y. It remains uncertain if Swiss forecasting institutions are more likely to relate to the Zurich Euromarket rate.

Survey of Professional Forecasters
The quarterly SPF asks participants to forecast the quarterly average 3-month Treasury bill rate (secondary market, discount rate). Source: Federal Reserve H.15 time series using Datastream FRTBS3M, FRTBS6M.
U.S. Treasury bill discount rates are converted into coupon equivalent or investment yield interest rates using the formula $r_y = r_d\ [100 / (100 - r_d\ m/360)]$ where r_d and m are the discount rate and maturity of the T-bill (m=91 for 3 months and 182 for 6 months).

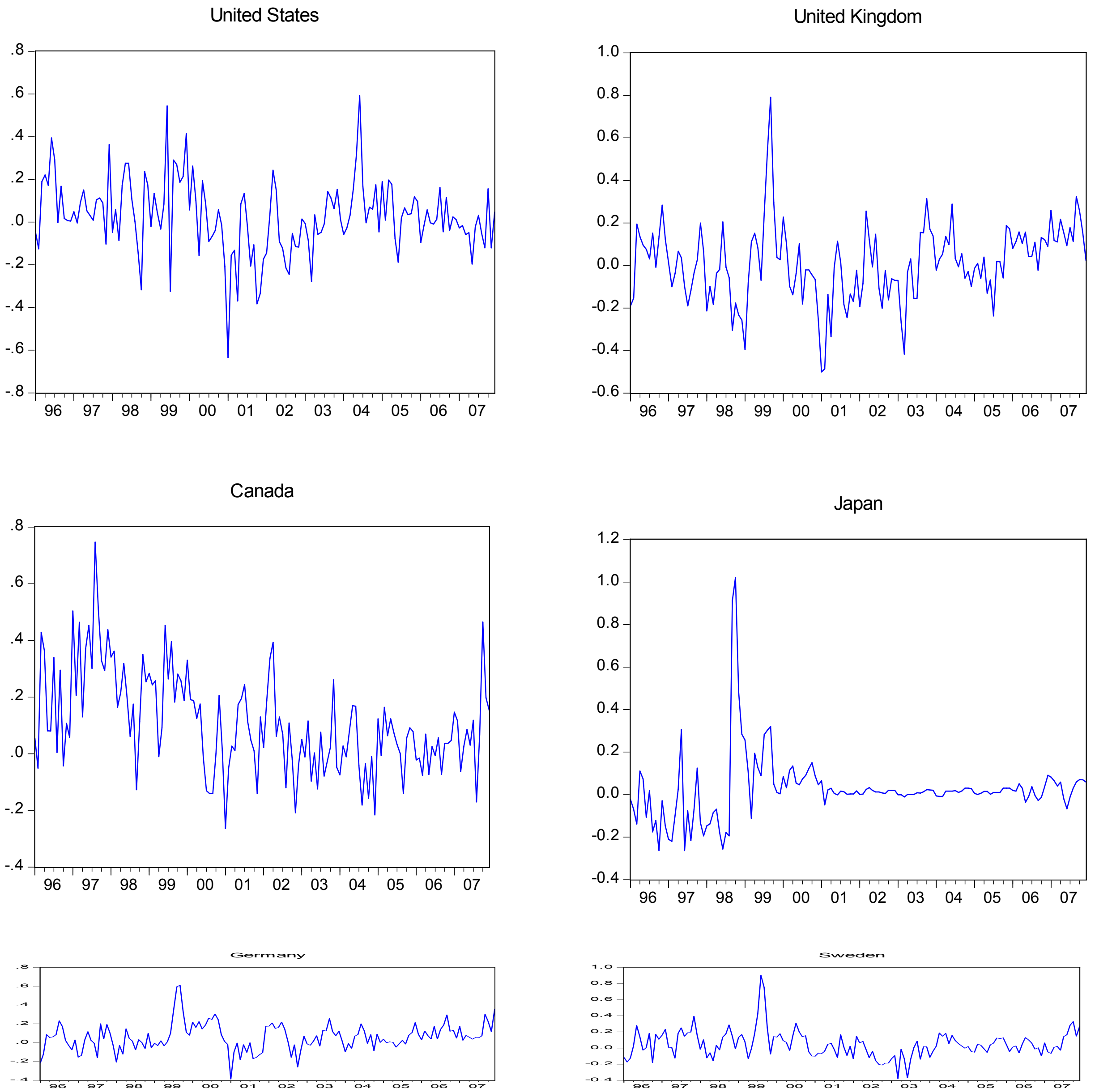
APPENDIX NOT FOR PUBLICATION
Unit root and stationarity tests

		Forward spread $f_t^{6,3} - r_t^3$		Expected change $S_t r_{t+3}^3 - r_t^3$		Term premium $f_t^{6,3} - S_t r_{t+3}^3$	
A. Sample period 1996M01-2007M12							
US	MZa	-10.21	**	-7.80	*	-60.35	***
	KPSS	0.14		0.31		0.25	
UK	MZa	-6.91	*	-3.41		-29.15	***
	KPSS	0.04		0.18		0.25	
CN	MZa	-11.32	**	-13.82	***	-29.79	***
	KPSS	0.45	*	0.08		0.90	***
JP	MZa	-77.68	***	-27.74	***	-44.57	***
	KPSS	0.31		0.43	*	0.13	
GE	MZa	-9.69	**	-24.23	***	-10.32	**
	KPSS	0.12		0.11		0.14	
SWE	MZa	-4.56		-3.28		-32.51	***
	KPSS	0.22		0.47		0.25	
NO*	MZa	-9.35	**	-10.31	**	-22.81	***
	KPSS	0.69	**	0.84	***	0.13	
SWI*	MZa	-22.88	***	-14.30	***	-26.69	***
	KPSS	0.14		0.49	**	0.25	
B. Sample period 1989M10-2006M05							
CN	MZa	-15.21	***	-7.69	*	-37.67	***
	KPSS	0.35	*	0.96	***	0.41	*

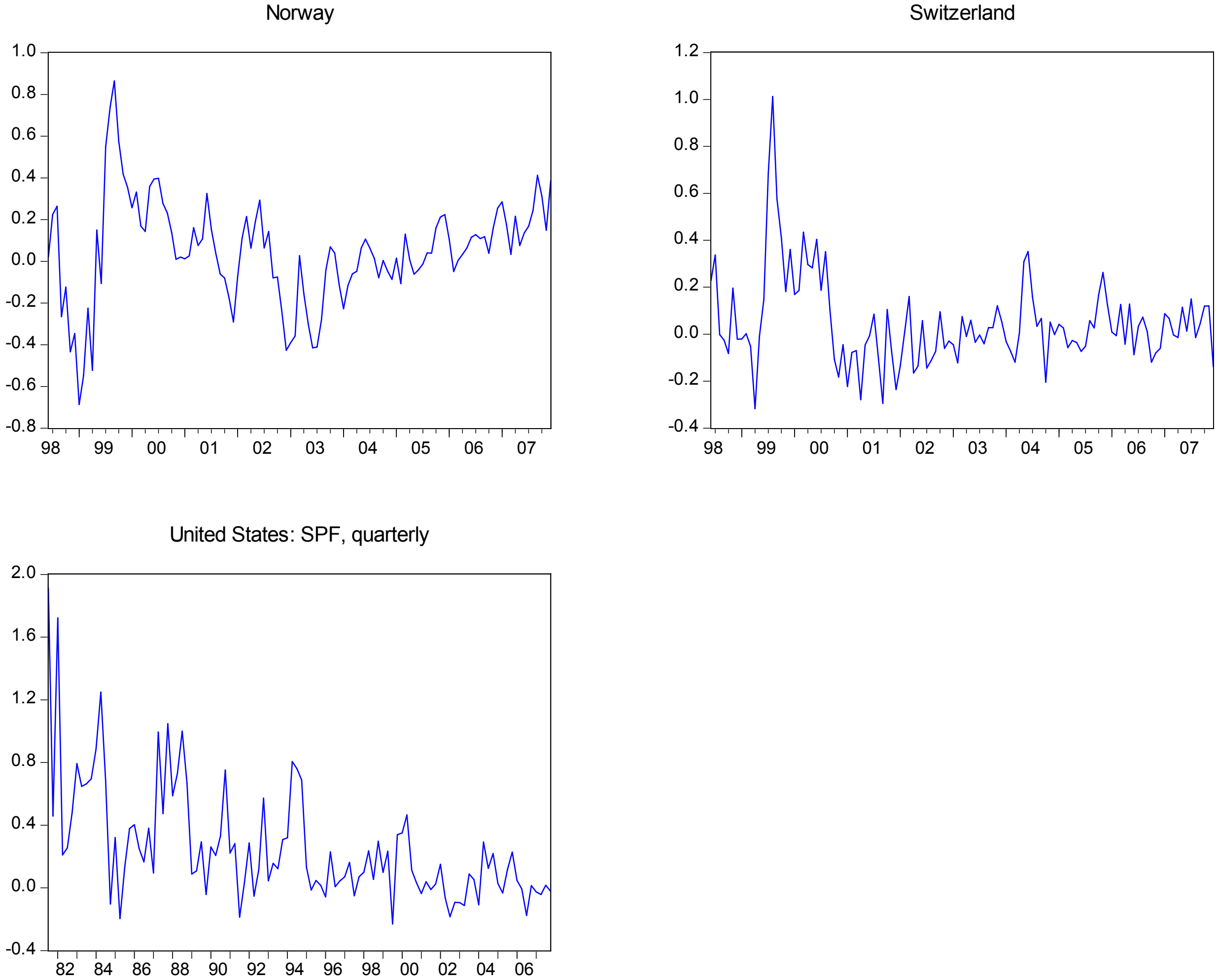
Notes: * Sample for Norway and Switzerland starts 1998M06.
MZa refers to the Ng-Perron modified unit root test with 1%, 5% and 10% critical values of -13.8, -8.1 and -5.7. KPSS test for stationarity, with critical values 0.739, 0.463, 0.347. ***, **, * symbols indicate significance at 1, 5 and 10 percent levels.

APPENDIX NOT FOR PUBLICATION

Term premiums in the 3-month forward rate: $f_t^{6,3} - S_t r_{t+3}^3$

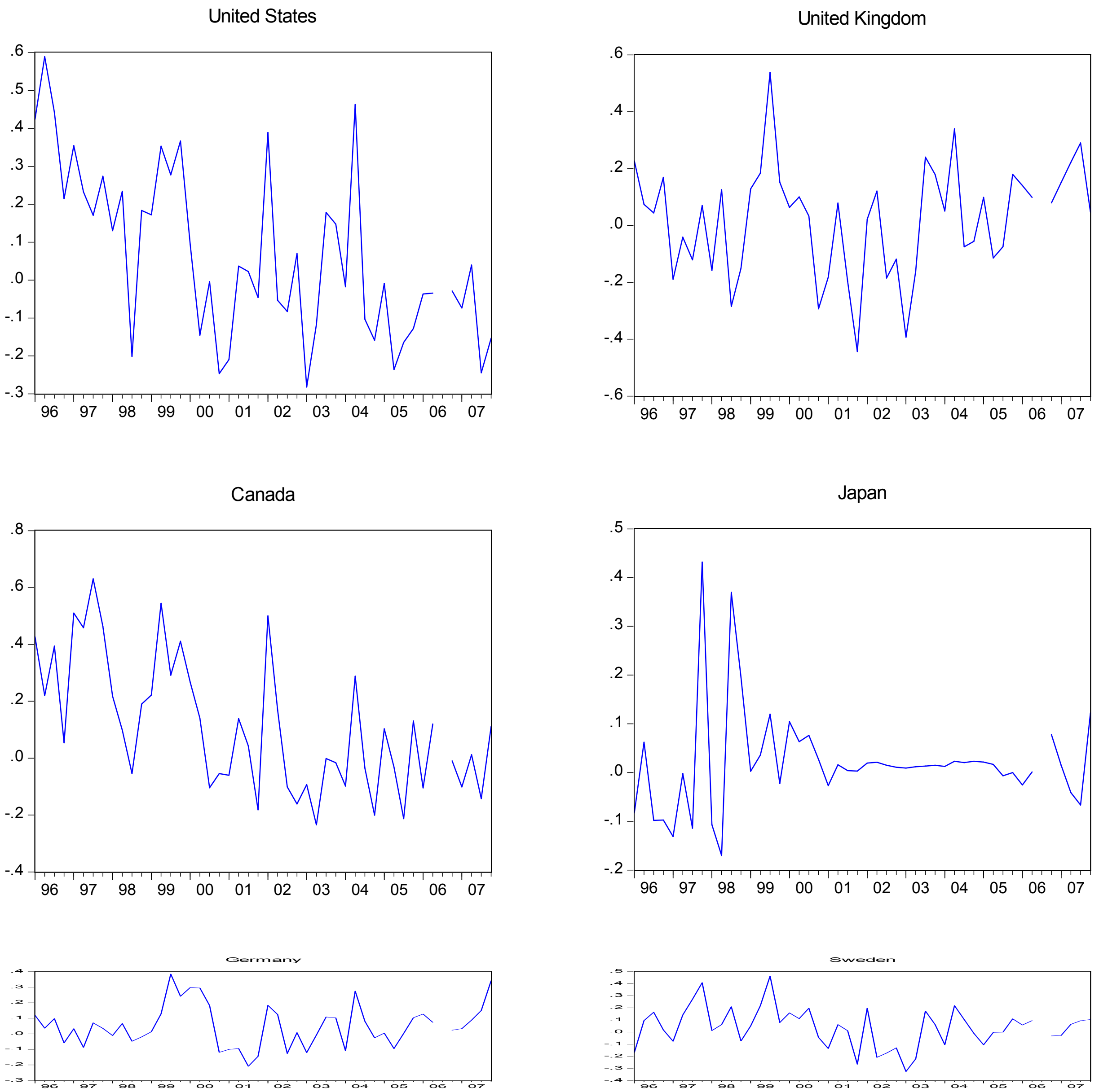


Term premiums in the 3-month forward rate (cont)

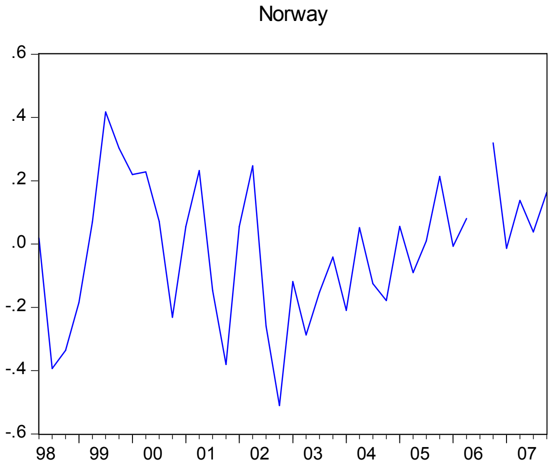


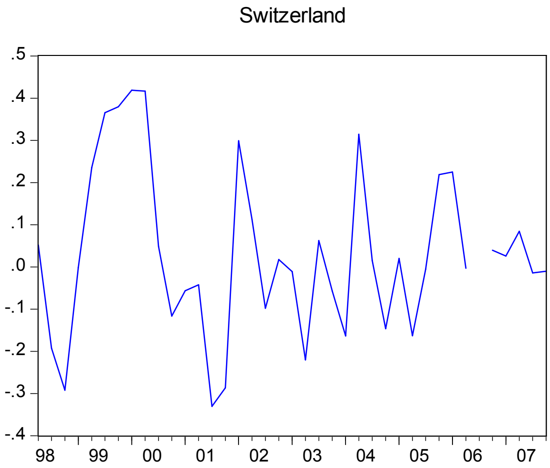
APPENDIX NOT FOR PUBLICATION

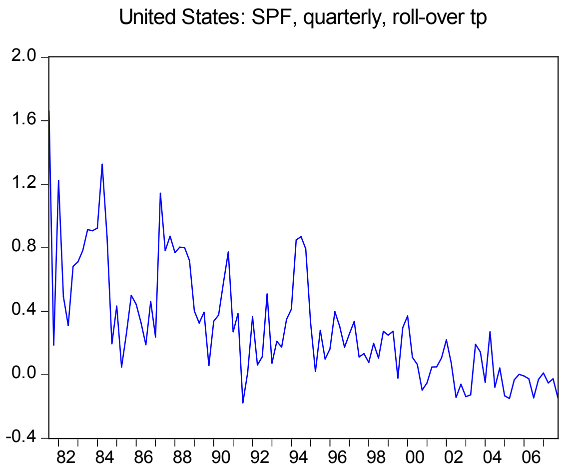
Term premiums in the roll-over spread: $R_t^{12} - (1/4)\sum_{k=0}^{3} S_t r_{t+3*k}^{3}$



Term premiums in the roll-over spread (Cont)







APPENDIX NOT FOR PUBLICATION
Derivation of coefficient bias equations

Coefficient bias derivations follow discussions in Fama (1984), Mankiw and Summers (1984) and others. Simplifying assumptions are used for some covariances and autocorrelations.

$$f_t^{(n,k)} = E_t r_{t+k}^{n-k} + \theta_t^{n-k}$$

Stochastic term premium bias
We start with the theoretical ET equation

$$(') \quad E_t r_{t+k}^{n-k} - r_t^{n-k} = -\vartheta_t^{n-k} + (f_t^{(n,k)} - r_t^{n-k})$$

The estimation and test equation is

$$('') \quad E_t r_{t+k}^{n-k} - r_t^{n-k} = \alpha_1 + \beta_1 (f_t^{(n,k)} - r_t^{n-k}) + v_{1t}^{n-k}$$

We assume that the term premium ϑ_t^{n-k} is not a constant, but is stochastic i.i.d. with a constant average and variance var(ϑ_t^{n-k}).

Ordinary Least Squares (OLS) estimation of regression equation ('') is based on the estimator of β_1

$$\beta_1^{OLS} = \frac{\text{cov}[(E_t r_{t+k}^{n-k} - r_t^{n-k}), (f_t^{(n,k)} - r_t^{n-k})]}{\text{var}[f_t^{(n,k)} - r_t^{n-k}]} = \frac{\text{var}[E_t r_{t+k}^{n-k} - r_t^{n-k}] + \text{cov}[(E_t r_{t+k}^{n-k} - r_t^{n-k}), \vartheta_t^{n-k}]}{\text{var}[E_t r_{t+k}^{n-k} - r_t^{n-k}] + \text{var}[\vartheta_t^{n-k}]}$$

Assuming that covariances with the term premium are zero results in equation (3) in the main text.

Expectations measurement error bias
We start with the theoretical ET equation

$$(*) \quad f_t^{(n,k)} - r_t^{n-k} = \vartheta_t^{n-k} + (E_t r_{t+k}^{n-k} - r_t^{n-k})$$

The estimation and test equation is

$$(**) \quad f_t^{(n,k)} - r_t^{n-k} = \alpha_2 + \beta_2 (S_t r_{t+k}^{n-k} - r_t^{n-k}) + v_{2t}^{n-k}$$

We assume that the survey expectations only proxy the true expectations and contain an i.i.d. measurement error, i.e. $S_t r_{t+k}^{n-k} = E_t r_{t+k}^{n-k} + \varepsilon_t^{n-k}$. Measurement error ε_t^{n-k} has a zero mean and constant variance var(ε_t^{n-k}).

OLS estimation of regression equation (**) provides the estimate of β_2 as

$$\beta_2^{OLS} = \frac{\text{cov}[(S_t r_{t+k}^{n-k} - r_t^{n-k}), (f_t^{(n,k)} - r_t^{n-k})]}{\text{var}[S_t r_{t+k}^{n-k} - r_t^{n-k}]} = \frac{\text{var}[E_t r_{t+k}^{n-k} - r_t^{n-k}]}{\text{var}[E_t r_{t+k}^{n-k} - r_t^{n-k}] + \text{var}[\varepsilon_t^{n-k}]} +$$

$$\frac{\text{cov}[(E_t r_{t+k}^{n-k} - r_t^{n-k}), \varepsilon_t^{n-k}] + \text{cov}[(E_t r_{t+k}^{n-k} - r_t^{n-k}), \vartheta_t^{n-k}] + \text{cov}[(\varepsilon_{t+k}^{n-k}, \vartheta_t^{n-k}]}{\text{var}[E_t r_{t+k}^{n-k} - r_t^{n-k}] + \text{var}[\varepsilon_t^{n-k}]}$$

Assuming that covariances with the term premium and measurement error are zero results in equation (4) in the main text.

Roll-over spread test
Assume $f_t^{(k+1,k)} = E_t r_{t+k}^1 + \theta_t^k$ and $\vartheta_t^0 = 0$

We start with the theoretical ET equation

$$(\#) \quad (1/n)\sum_{k=0}^{n-1} E_t r_{t+k}^1 - r_t^1 = -(1/n)\sum_{k=1}^{n-1} \vartheta_t^k + (R_t^n - r_t^1)$$

The estimation and test equation is

$$(\#\#) \quad (1/n)\sum_{k=0}^{n-1} E_t r^1_{t+k*m} - r^1_t = \alpha_3 + \beta_3\,(R^n_t - r^1_t) + v^n_{3t}$$

We assume that the term premiums ϑ^k_t are not constant, but are stochastic i.i.d. with equal variances var(ϑ_t).

Estimating regression equation (##) provides the estimate of β_3 as

$$\beta_3^{OLS} = \frac{\text{cov}[(1/n\sum_{k=0}^{n-1} E_t r^1_{t+k} - r^1_t), (R^n_t - r^1_t)]}{\text{var}[R^n_t - r^1_t]}$$

With respect to innovations in interest rates and associated variances it will be practical to use

$$(1/n)\sum_{k=0}^{n-1} E_t r^1_{t+k*m} - r^1_t = \sum_{k=1}^{n-1} (1-k/n) E_t \Delta r^1_{t+k})$$

$$\beta_3^{OLS} = \frac{\text{var}[E_t \Delta r^1_{t+k}]\sum_{k=1}^{n-1}(1-k/n)^2}{\text{var}[E_t \Delta r^1_{t+k}]\sum_{k=1}^{n-1}(1-k/n)^2 + \text{var}[\vartheta_t](n-1)/n^2} +$$

$$\frac{\text{cov}[(\sum_{k=1}^{n-1}(1-k/n)E_t\Delta r^1_{t+k}), (1/n)\sum_{k=1}^{n-1}\vartheta^k_t]}{\text{var}[E_t \Delta r^1_{t+k}]\sum_{k=1}^{n-1}(1-k/n)^2 + \text{var}[\vartheta_t](n-1)/n^2}$$

Assuming that covariances with the term premium are zero results in

$$\beta_3^{OLS} = \frac{1}{1 + \dfrac{\text{var}[\vartheta_t]\,(n-1)/n^2}{\text{var}[E_t \Delta r^1_{t+k}]\,\sum_{k=1}^{n-1}(1-k/n)^2}}$$