BANK OF ENGLAND

Working Paper No. 489

Expectations, risk premia and information spanning in dynamic term structure model estimation

Rodrigo Guimarães

March 2014

BANK OF ENGLAND

Working Paper No. 489

Expectations, risk premia and information spanning in dynamic term structure model estimation

Rodrigo Guimarães[(1)]

Abstract

This article examines the nature of the empirical instability in dynamic term structure models. I show that using survey forecasts is an effective solution because it directly addresses the information imbalance at the heart of the instability: it increases the (cross-section) information on actual dynamics, bridging the gap with the large (cross-section) information on the risk-adjusted dynamics. I relate this to other information spanning problems, particularly spanning of macro factors, and discuss the desirability of anchoring models to surveys. I also show that restricting prices of risk is not effective in ensuring stable and sensible implied expectations.

Key words: Interest rates, expectations, risk premium, dynamic term structure, robust, estimation.

JEL classification: G12, E43, C58.

(1) Bank of England. Email: rodrigo.guimaraes@bankofengland.co.uk

The views expressed in this paper are those of the author, and not necessarily those of the Bank of England or the Monetary Policy Committee. I am grateful to an anonymous referee, Michael Bauer, Mark Britten-Jones, Mike Chernov, Marcelo Ferman, Don Kim, Canlin Li, Andrew Meldrum, Fulvio Pegoraro, Jean-Paul Renne, Paul Whelan and Gabriele Zinna for helpful discussions and suggestions as well as comments from seminar participants at the Bank of England, BlackRock, 2013 World Finance Conference, the Banque de France and at the Federal Reserve Board. Ali Schomberg provided excellent research assistance with the survey data. All remaining errors are my sole responsibility. This paper was finalised on 25 February 2014.

The Bank of England's working paper series is externally refereed.

Information on the Bank's working paper series can be found at www.bankofengland.co.uk/research/Pages/workingpapers/default.aspx

Publications Group, Bank of England, Threadneedle Street, London, EC2R 8AH
Telephone +44 (0)20 7601 4030 Fax +44 (0)20 7601 3298 email publications@bankofengland.co.uk

ISSN 1749-9135 (on-line)

Summary

Market interest rates are of great interest to policymakers, not least because they play a crucial role in the monetary transmission mechanism. Moreover, financial market measures of future interest rates and inflation rates can also provide useful and timely information when making policy decisions.

This information complements and extends other sources monitored by policymakers, such as surveys of private forecasters and macroeconomic forecasting models. Market rates are available at a much higher frequency and for longer forward horizons than other data, as well as being available in a long time series. This can prove crucial in answering questions that involve the reaction to policy (such as announcements), comparisons over long periods (the effect of institutional changes, such as independence of the central bank), or effects that are expected to have distinct effects over different horizons (such as forward guidance).

In order to extract policy-relevant information from yields, it is important to understand what has driven these rates lower. Decompositions can be carried out along a number of dimensions to shed light on the drivers. First, movements in interest rates can be split into movements at different forward horizons to assess whether the changes are mainly at shorter or longer horizons. Second, movements in nominal rates can be decomposed into changes in real interest rates and changes in implied inflation rates.

And third, movements in market rates can be decomposed into two parts; one that reflects changes in expectations of future short-term rates, and another associated with changes in their required compensation for risk ('risk premia'). Disentangling both is important for policymakers because influencing the expected path of the policy rate plays an important role in the transmission mechanism of monetary policy. And estimating risk premia can give policymakers an indication of market participants' assessments of the perceived risks. In addition, some measures are designed to reduce the compensation for risk (such as quantitative easing).

While the first two decompositions – time horizon and the real versus inflation split – can be done using available data, the distinction between expectation and risk compensation components is more complicated. Extracting this information requires complicated theoretical models and statistical techniques, which raises the question of reliable decompositions. Unfortunately, the most popular class of models, both within academia and with major policy institutions, are known to be subject to instability problems that would hamper their use for policy. This paper focuses on how to obtain robust estimates from these models for the quantities of interest for policymakers: the expected path of future interest and inflation rates as well as real and inflation risk premia.

We analyse the robustness of the decomposition obtained from the workhorse model in previous work, the family of Gaussian Affine Dynamic Term Structure Models. The great advantage of this type of model, which assumes linearity and a relatively straightforward probability distribution of shocks to returns, is its tractability. At the same time their flexibility is a great asset, necessary to accommodate the rich behaviour of bond yields observed over time and

across maturities. But without enough restrictions or information to pin down the model parameters, this flexibility can become a liability, resulting in instability in the implied decomposition into risk and expectations.

Exploring recent advances in yield curve modelling this paper compares alternative methods proposed in previous work to ensure sensible decompositions. These include using survey forecasts from professional forecasters, restricting the way risk premia are allowed to vary or purely statistical techniques. This paper finds that using surveys of private professional forecasters to help anchor the model dynamics is the most reliable way to obtain robust decompositions.

In addition, the use of surveys automatically delivers ‘sensible’ decompositions because these survey forecasts (i) have been shown to provide good proxies for expected future rates (good forecasting properties); (ii) are true real-time measures (not subject to look-ahead or overfitting biases); (iii) can incorporate information that is readily available to practitioners (political events, changes in policy or policy frameworks) hard to obtain from past data.

The outputs of the models with surveys have been used to analyse the evolution of UK government bond yields in a 2012 Q3 *Quarterly Bulletin* article. The model decomposition of nominal, real and market inflation rates provided valuable insight about the behaviour of yields. It proved particularly useful in understanding the recent period of the financial crises and how it impacted market rates.

In a more technical contribution, the paper also links the ability of surveys to stabilise the decomposition of yields to new developments in term structure modelling related to spanning of information. A Monte Carlo study (based on random simulations of a theoretical model) confirms the importance of having additional information about future dynamics to reliably estimate these models. It suggests that the introduction of surveys delivers gains in precision equivalent to observing at least twice as long a sample – in other words we would need double the amount of information available (wait another 40 years) to obtain measures as reliable as those we can obtain by adding surveys.

1 Introduction

The evolution of risk-free yield curves is of fundamental importance to researchers and practitioners in finance and macroeconomics. Separately identifying expected risk-free rates and compensation for risk is important in understanding bond returns, discounting cash flows for asset prices and investment valuation, and monetary policy analysis. Theoretical dynamic term structure models (DTSM) allow the extraction of expected future interest rates and risk premia from observed yield curves. Unfortunately the most popular class of DTSM, the class of Affine models, is plagued by instability problems believed to be a combination of over-parametrization and small-sample biases due to the high persistence in yields. In this paper I investigate the nature of the instability of the decomposition of yields into risk premia and expectations components in Gaussian Affine Dynamic Term Structure Models (GADTSM), the main workhorse and benchmark in the macro-finance literature since the seminal work of Ang & Piazzesi (2003).

The standard estimation strategy for DTSM has a very large amount of information on the risk-adjusted dynamics of interest rates and very little on the actual dynamics. As a result, the expectations component, and hence the decomposition, is poorly identified. The evidence in this paper confirms that the inclusion of additional information on the actual dynamics with survey forecasts is a natural and effective way of dealing with this imbalance.[1] The stability is achieved by anchoring the long run mean of interest rates rather than the estimated persistence of the factors. In contrast, restricting risk premia is not effective in stabilizing the decomposition.

The evidence in this paper suggests that the lack of robust estimates of GADTSM is an *informational spanning problem* that is similar to other problems with unspanned risks identified in recent literature. Information about risk-adjusted dynamics is not enough to identify the actual dynamics (as in the *hidden factor* framework of Duffee (2011a)) nor is information on actual dynamics enough to identify the risk-adjusted dynamics (as in the *unspanned macro risk* framework of Joslin, Priebisch and Singleton (2012, henceforth JPS)).

[1] Given that the decompositions are the main use of these models for policy analysis and macroeconomics, it is paramount that we have reliable decompositions (see Kim & Orphanides (2007) and Rudebusch, Sack & Swanson (2007)). Ideally, we would like a strategy that guarantees sensible and robust decompositions that are not sensitive to the exact sample used in estimation. Models anchored with surveys produce similar decompositions even in very small samples.

I formalize this intuition by comparing the risk-adjusted and actual dynamics implied by GADTSM estimated with US and UK yield curve data. I use well known global distance measures between distributions applied to model-implied expectations to test the hypothesis of same implied dynamics. I apply these tests to model estimates based on different amounts of information on the risk-adjusted and actual dynamics. I show that the success of surveys in helping to identify the actual dynamics is similar to the effect of including longer-term yields to anchor the risk-adjusted dynamics.

Inspired by the empirical evidence from models with both US and UK data, I conduct a Monte Carlo experiment to quantify the effect of the 'cross-section' in providing robust estimates. The results show that the addition of a cross-section resembling available survey forecasts in US and UK, is equivalent to more than a 100 years of time series data. Even when the cross-section is very noisy, or is available for fewer horizons, it is equivalent to at least doubling the sample size used in estimation. The Monte Carlo results also confirm the information imbalance interpretation. A cross-section resembling observed yield curves is equivalent to observing a sample of more than 1000 years for the risk-adjusted dynamics, compared to the time series sample of 40 years typically available.

I also show that the model decomposition is *not* sensitive to the inclusion of surveys in an important way: *conditional on the estimated parameter, the decomposition is mostly unaffected by the inclusion or not of surveys.* While surveys help pin down the parameters in the estimation of the model, filtering the state variables is not materially affected by the exclusion of the observation equations with survey forecasts. This confirms the flexibility of these models and is suggestive that there is not much tension in matching yields and surveys at the same time.[2] This is consistent with the conclusions of Chernov & Mueller (2012), who explicitly allow for the possibility that the expectations captured by surveys might differ from the actual expectations embodied in yields but fail to reject the two are the same.

Besides reducing concerns on the appropriateness of using surveys, this means that

[2]This fact per se does not prove that there is no tension in matching yields and surveys at the same time, as it could simply be a consequence of a higher measurement error associated with surveys. Indeed, we suggest (see Section 5.1) that one way researchers can ensure that the decomposition is insensitive to the surveys is by imposing large measurement errors (here we estimate these freely). However, the fact that the inclusion of surveys does stabilize the decompositions, and therefore have a material effect on model estimates while not affecting the fit of yields, combined with the filtering insensitivity result, is indicative of them being broadly consistent, in line with Chernov & Mueller (2012).

the model can be used to decompose yields at a much higher frequency than that for which survey forecasts are available. This is particularly important for practical use of these models, such as real time monitoring for policy purposes, event studies or real time forecasting, given that surveys are available at lower frequencies.

Most of the analysis in this paper is focused on nominal yield curves without macro variables. But I also consider the estimation of joint real and nominal yield curve models for the UK, for which a long time series of liquid inflation-linked bonds is available. This allows me to illustrate the link with spanning of macro risks. The unspanned macro risks framework of JPS uses information in macro variables to identify *only* the actual dynamics of the factors, while avoiding assigning prices of risk to variables for which we have no observations for their risk-adjusted dynamics.[3] Our results confirm the difficulty in estimating price of risk for macro variables without observable asset prices and highlight that this is an identification issue rather than an issue of economic spanning of macro risks.

By not altering the risk-adjusted dynamics of yields, the JPS unspanned macro framework can be seen as an alternative to surveys (or any other extra information on actual dynamics of yields) as it increases the information on the actual dynamics without the undesirable implications (for both actual and risk-adjusted dynamics) from assuming spanned macro risks. From this perspective, it can be thought of as an econometric strategy to better identify the actual dynamics of pricing factors, with no implication about the price of risk of macro variables (we simply cannot identify them). Hence, we do not need to find a general equilibrium model that implies unspanned macro risks to rationalize the use of the JPS framework, though this might in itself be of interest.[4] It has the additional benefit relative to use of surveys of identifying the economic sources of variation in factors, but the drawback that it is subject to model and estimation risk.

Beyond instability issues, recent research has cast serious doubt on the usefulness of GADTSM (e.g. Duffee (2012 a,b) and Joslin, Le & Singleton (2011)), suggesting that the no-arbitrage restrictions have no implications for forecasting or joint distribution of yields with macro variables. Our results offer a rather more optimistic view about the

[3] With respect to actual dynamics, JPS also point out the counter-factual implication that the cross-section of bond yields should contain all information about the actual dynamics of any variable included in the pricing equation of GADTSM.

[4] JPS briefly speculate what might be the conditions under which a structural model might deliver unspanned macro risks.

usefulness of GADTSM. In a similar vein to JPS, who explore the fact that only within a no-arbitrage model it is possible to meaningfully discuss the economics of risk premia in yields, our results show that the ability of (GA)DTSM to incorporate useful information from surveys, or macro variables, while fitting the data well is a valuable property. This is not possible in purely statistical models, which do not distinguish between prices and expectations.[5]

2 Dynamic term structure models

Dynamic term structure models (DTSM) have a special place in finance theory because they capture a substantial amount of the dynamics of the intertemporal marginal rates of substitution, commonly referred to as the stochastic discount factor (SDF), or Arrow-Debreu state-price densities (see Duffie (2001)). In equilibrium, continuously-compounded zero-coupon yields are the expectation under the risk-adjusted probability measure ($\mathbb{Q}$) of future short rates, or the expectation under the actual probability measure ($\mathbb{P}$) of the SDF. The nominal (real) yield curve is the expected path of the nominal (real) SDF. If we denote by $M_{t,T} = \Pi_{i=1}^{T} M_{t+i}$ the SDF and $y_{t,T-t}$ the zero-coupon yields at date t with time to maturity $T-t$, and r_t the short-term risk-free interest rate then, in equilibrium, we have:

$$y_{t,T-t} = -\frac{1}{T} \ln E_t^P \left[M_{t,T} \right] = -\frac{1}{T} \ln E_t^Q \left[\exp \left(-\sum_{s=t}^{T-1} r_s \right) \right], \text{ for } T > t \tag{1}$$

We can use (1) to understand the nature of the identification problem of extracting expectations and term premia from yield curves, as well as the intuition for the proposed remedies. Every time we observe a snapshot of the yield curve we have a complete time series of many years (up to 50 years for some developed countries) of the $\mathbb{Q}$-dynamics of interest rates. That is, every day or even higher frequency, we have a full time series of

[5] This property separates no-arbitrage models (including GADTSM) from unrestricted VAR or dynamic factor models such as the Dynamic Nelson-Siegel (DNS) model of Diebold & Li (2006). Altavilla, Giacomini & Ragusa (2013) have shown that DNS forecasting performance broke down after 2000, while surveys continued to perform well. But the DNS model cannot match surveys and prices at the same time. As a result, the authors propose twisting forecasts of the DNS model by combining its predictions with those of surveys in an optimal way. GADTSM allow the model itself to be anchored to survey data without any loss of fit precision because of the distinction between risk-adjusted (prices) and actual (expectations) probability measures.

the risk-adjusted dynamics from the cross-section of yields. This means we can identify the $\mathbb{Q}$-dynamics very well. On the other hand, we only observe one realization (for many countries of no more than 40 years) of the $\mathbb{P}$-dynamics from the time-series of yields. What links both dynamics is the risk premium. Therefore, the less restricted risk premia are, the less information about the $\mathbb{Q}$-dynamics (cross-section) of yields is used in determining the $\mathbb{P}$-dynamics (time series).

This implies that a model that is flexible enough to fit yields should have little problem in identifying the risk-adjusted dynamics. However, we are less confident about identifying the parameters of $\mathbb{P}$-dynamics, particularly given the high persistence of yields and relatively short samples available (see Duffee & Stanton (2012) and Bauer, Rudebusch & Wu (2012, henceforth BRW)).[6] Since the decomposition into expected and risk premia components will depend on the difference between the risk adjusted and actual probabilities, the decomposition will only be as robust as the identification of the $\mathbb{P}$-dynamics.

While most of the literature focuses on Affine DTSM, the problem of identifying the risk-adjusted and actual dynamics, and the intuition for the possible remedies, apply more generally. Since the problem is that we have more information about the risk-adjusted dynamics, a natural solution is to add more information on the actual dynamics of interest rates. This can be done by including survey forecast of interest rates. In an early paper discussing instability of GADTSM, Kim & Orphanides (2012) suggested using forecasts of future interest rates from surveys which has since become a popular strategy.

Because some of the problems identified also reflect unrealistic implied risk premia, particularly in higher dimensional models (e.g. Duffee (2010) and Joslin, Singleton & Zhu (2011, henceforth JSZ)), and because restricting risk premia links the two dynamics more closely, it is not surprising that restrictions on risk premia have been proposed as an alternative to improve estimation of the $\mathbb{P}$-dynamics. When we restrict the prices of risk, we use information from the risk-adjusted dynamics to infer the actual dynamics.[7]

[6] The imbalance is perhaps worse than just the imbalance in number of samples. We have observations of the *expected path* under the risk-adjusted dynamics from the cross-section (we observe $E_t[r_{t+k}]$, plus some measurement error), whereas the time series is a *realization* of the dynamics (for an autoregressive process the observation at time $t+k$ is a function of past expected path and the sum of all shocks since, $E_t[r_{t+k}] + \sum \rho^{k-i} e_{t+i}$). For very persistent processes, the time series will provide a noisy measure of the expected dynamics. We make this more precise in Section 5 after introducing the model.

[7] This, however, raises the risk of spreading misspecification if the restrictions imposed are not warranted (see JPS). There is no guarantee that either dynamics will be well estimated by forcing them to be closer.

There are different ways of restricting risk premia. Some of the alternatives suggested include: restricting implied Sharpe ratios (Duffee (2010)), or similarly the variability of prices of risk (Bibkov & Chernov (2009) and Chernov & Mueller (2012)), restricting the dimension of risk premia by limiting the rank of the price of risk (JSZ) or by imposing many zero restrictions (Cochrane & Piazzesi (2008)). These differences also highlight a weakness of this approach: there is no clear way to restrict risk premia and restricting them too much might do more harm than good (i.e., it could distort $\mathbb{Q}$-dynamics without correctly identifying $\mathbb{P}$-dynamics).

Lastly, to the extent that part of the problem is the capacity to identify the actual dynamics from persistent time series in small samples, statistical methods designed to improve estimation of persistent time series can be used. A recent example of this alternative is the bias-correction method of BRW.

The solutions can hence be divided into two groups: the yield data itself is not sufficient to identify both dynamics (so augment the data with additional information from surveys) or it is just a problem of identification that can be solved by adding restrictions to the model specification (limiting the variation or dimension of prices of risk) or altering the estimation method (statistical bias correction).To date, however, there has been no study establishing the capacity of each in isolation in achieving robust estimation, or comparing them. In this paper we do just that.

The instability concerns have became stronger as the literature increasingly focused on higher dimensional models following results suggesting that important information for bond excess returns (Cochrane & Piazzesi (2005)) and interest rate volatility (Collin-Dufresne, Goldstein & Jones (2009)) were not spanned by the traditional level, slope and curvature factors. Cochrane & Piazzesi (2005) showed that a factor unspanned by the first 3 Principal Components (PC) could forecast a substantial fraction of bond excess returns. This suggested that the typical 3 factor model that had dominated the literature since the results of Litterman & Scheinkman (1991) would not be capable of capturing basic excess return properties. Since then, a large number of models have adopted either 4 or 5 factors (e.g. Cochrane & Piazzesi (2008), Duffee (2011a,b), and Chernov & Mueller (2012)). As the number of parameters more than doubles from the 3 to 5 factor model, over-parametrization and small sample identification issues are heightened.

A second reason for recent concern is related to the financial crisis and the ensuing effective zero lower bound (ZLB) period in many of the advanced economies, including the US and UK, which raises the possibility of a regime change in bond yields dynamics. Potential regime changes are the common reason for using samples much shorter than the available yield data (e.g. Kim & Wright (2005), Joyce, Lildholt & Sorensen (2010) and JSZ). Given that real time forecasts are conditional on the regime they are in, anchoring expectations to surveys might help alleviate or solve problems with regime changes.[8] We therefore either need a reliable way of estimating sufficiently flexible models that will have no trouble in matching the whole time series or a method that has the potential to deliver reliable estimates for very short-term samples. This leads us to include very short subsamples in our empirical comparison.[9]

2.1 Why Gaussian homoscedastic term structure models?

This family of models owes its popularity in the literature to the ease with which macro variables can be jointly modelled with yields, since it closely resembles commonly used macro-econometric techniques. Our choice of GADTSM is driven not only by the ease with which these models can be linked to conventional theoretical and empirical macroeconomic models, but also because of feasibility concerns. Because factors in affine term structure models are latent, we need to impose identifying restrictions to estimate the models. An identification strategy recently proposed by JSZ has drastically reduced the estimation times for the GADTSM family, which ultimately makes the analysis in this paper feasible. Estimation of these models has been reduced from weeks or days, with complicated numerical searches with ill-behaved likelihoods, to just a few minutes or hours.[10]

These models are clearly misspecified as yields display significant heteroscedasticity. However, this will always be true about some aspects of the data.[11] The question is

[8] Many 'regime changes' in parameters in one model can be accommodated by an enlarged model with fixed parameters. For example, if the regime change is the inflation target or inflation volatility, then if these are a variable instead of a parameter the model would not suffer from regime changes. It then becomes an empirical question whether the latent factors in DTSM capture the relevant variables that might be viewed as having different regimes in subsamples.

[9] This would allow estimation of models for new instruments, such as the (nominal) OIS curve or a joint model with the inflation swaps curve. Data on these instruments span less than a decade.

[10] Because in our baseline exercise we estimate models for 2 countries, with 3 different dimensions, 8 samples and 5 alternative estimation strategies for nominal yield curves alone, we have to estimate more than 240 different models.

[11] See Kim (2007) for a discussion of the different properties term structure models need to have to match different frequencies of the data.

whether these models serve as a good approximation for the purposes at hand. Feldhüter (2008) examines the capacity of ADTSM to match different yield moments with different parametrization of risk premia (which only differ when models have at least one factor driving the stochastic volatility, all are equivalent to the essentially affine risk premia used here when models are Gaussian). He finds that none of the models with stochastic volatility can match the Gaussian model in replicating the Campbell and Shiller (1991) regressions. This extends the results of Duffee (2002), who compared the completely affine specification of prices of risk with his proposed essentially affine model, to the more flexible price of risk specifications of Duarte (2004) and Cheridito, Filipovic & Kimmel (2007).

Bibkov & Chernov (2011) also conclude that stochastic volatility is not essential when only matching yields. Heidari & Wu (2009) suggest that including stochastic volatility in ADTSM is not the key to match option-implied volatility, what is important is to add factors to capture the rich factor structure present in yield residuals beyond what is explained by 3 factors. In other words, while stochastic volatility models have very different implications for higher moments that better match those moments in the data and derivatives, regarding the first moment of yields, which is what we are concerned with, the Gaussian model seems to perform just as well.

3 Methodology

To document the instability of unrestricted GADTSM, and to evaluate the success of including survey information or price of risk restrictions to deliver the desired robust and sensible decompositions, I estimate models that vary in number of factors and sample periods used for the different estimation strategies. Specifically, I consider (i) unrestricted models, (ii) models that include survey forecast information, (iii) models that restrict the implied maximum Sharpe ratio (which effectively is a restriction on variability of prices of risk) (iv) and models that restrict the number of time-varying risk premia. I consider different limits on average Sharpe ratio and also combine surveys and Sharpe ratio restrictions.

For each of these estimation strategies, and for both US and UK, I estimate a total of 24 models: 3 different dimensions for the factors driving yields and 8 different estimation samples. I estimate models with 3 to 5 factors, which encompasses most of the models used in the latent factor literature. For a given number of factors, I estimate the model with

8 different samples each, with the start dates ranging from January 1972 every 5 years to January 2007. These dates will capture most of the time periods commonly suggested as candidates for regime break dates for both US and UK.

3.1 Data

I use continuously compounded zero-coupon end-of-month yields for the US and UK from January 1972 through November 2010[12]. For the US we follow Duffee (2011a) and use monthly CRSP data for spot yields with maturities of 0.25, 1, 2, 3, 4 and 5 years augmented with 7 and 10 year maturity yields from Gurkaynak, Sack & Wright (2007). For UK yields I use end-of-month zero-coupon spot yields estimated by the Bank of England with maturities of 0.5, 1, 2, 3, 4, 5, 7 and 10 years.

Survey forecasts on short-term nominal rates for the US are taken from the Survey of Professional Forecasters of the Philadelphia Fed. I use the forecasts for 3 month T-bill rates 1, 2, 3 and 4 quarters ahead, available on a quarterly basis since September 1981, and the forecast for the average over the next 10 years, available annually since 1992. For UK, I use forecasts from the Bank of England's Survey of External Forecasters for Bank Rate 1,2 and 3 years ahead, available quarterly since 1999. Figure 1 shows the time series of yields and of survey forecasts for both countries.

For the joint model of real and nominal UK yields, I use continuously compounded zero-coupon real spot yields with 1, 2, 3, 4, 5, 7 and 10 years maturity and survey forecasts of retail price index (RPI) inflation from Consensus Forecasts for 1, 2, 3, 4, 5 years ahead and the average between 6 and 10 years ahead.

Table 1 shows some basic properties of the data used in this paper. It clearly shows the high persistence of yields of all maturities, as well as a positive average slope for the yield curves.

3.2 Models

Discrete time GADTSM are characterized by three assumptions. The first assumption is that the risk-free short-term interest rate is a linear function of a vector of state variables,

[12] We start our sample in January 1972 because of data availability. UK yields are only available since 1971 and the 10 year spot yield from Gurkaynak, Sack and Wright (2007) is only available since August 1971.

or factors:

$$i_t = \delta_0 + \delta_1 X_t \tag{2}$$

In the macro-finance literature the factors include observed macroeconomic variables and latent yield factors. In this paper I will only consider latent factor GADTSM.

The second assumption is that the dynamics of the state variables are described by a VAR under the risk-adjusted pricing measure ($\mathbb{Q}$):

$$X_{t+1} = \mu^Q + \Phi^Q X_t + \Sigma \varepsilon_{t+1}^Q \tag{3}$$

where $\varepsilon_{t+1}^Q \mid X_t \sim N(0, I)$. These two assumptions are sufficient to imply that bond prices are exponential-affine functions of the state variables:

$$P_{t,n} = \exp(A_n + B_n X_t) \tag{4}$$

where $P_{t,n}$ is the price of a bond with n years to maturity, and the loadings $A_n = \mathcal{A}\left(\delta_0, \delta_1, \mu^Q, \Phi^Q, \Sigma, n\right), B_n = \mathcal{B}\left(\delta_1, \Phi^Q, n\right)$ satisfy recursive equations derived in Appendix A.

These models are referred to as Gaussian Affine DTSM because continuously compounded yields are affine function of the factors:

$$y_{t,n} = A_n^Q + B_n^Q X_t \tag{5}$$

with $A_n^Q = -\frac{A_n}{n}, B_n^Q = -\frac{B_n}{n}$ and because of the homoscedastic assumption yields will have a Gaussian distribution.[13] The final element, required to estimate the model and to get the decomposition of yields into expected and risk premia components, is the price of risk, which transforms risk-adjusted probabilities into actual probabilities and vice versa. Following Duffee (2002), the vector of prices of risk is also specified as a linear function of the state variables :

$$\Lambda_t = \lambda_0 + \lambda_1 X_t \tag{6}$$

Given this choice of prices of risk, the dynamics under the actual probability measure

[13] For a discussion of non-Gaussian ADTSM, including stochastic volatility and jumps see Dai & Singleton (2003) or Piazzesi (2010).

($\mathbb{P}$) of the factors are also described by a VAR:

$$X_{t+1} = \mu + \Phi X_t + \Sigma \varepsilon_{t+1} \tag{7}$$

where $\varepsilon_{t+1} \mid X_t \sim N(0, I)$, $\mu^Q = \mu - \Sigma\lambda_0$ and $\Phi^Q = \Phi - \Sigma\lambda_1$, and the log-SDF is given by:

$$\ln M_{t+1} = -i_t - \frac{\Lambda_t' \Lambda_t}{2} - \Lambda_t' \varepsilon_{t+1} \tag{8}$$

The parameter set can be defined by $\Theta = \left\{\delta_0, \delta_1, \mu, \Phi, \mu^Q, \Phi^Q, \Sigma\right\}$.

3.2.1 Joint real-nominal models

The GADTSM can equally be applied to nominal or inflation-linked bonds. The only difference is that for the former the relevant short-term interest rate (Equation 2) is the nominal rate while for the latter it is the real rate. For the UK I also consider a joint nominal-real GADTSM[14] were both real and nominal rates are linear functions of a common set of factors, and so inflation is also allowed to be a function of all factors. If we denote the one period nominal risk-free rate by i_t, the real risk-free rate by r_t and expected one period inflation by π_t, and assume:[15]

$$r_t = \rho_0 + \rho_1 X_t \tag{9a}$$
$$i_t = \delta_0 + \delta_1 X_t \tag{9b}$$
$$\pi_t = i_t - r_t \tag{9c}$$

then spot nominal $(y_{t,\cdot})$, real $\left(y_{t,\cdot}^r\right)$ and inflation $(\pi_{t,\cdot})$ rates will be affine functions of the factors (Equation (5)) with the loadings given by the same recursions with the corresponding short rate parameters (see Appendix A).

Some models have assumed that inflation and real rates are driven by separate factors

[14] The UK has a long history of very liquid inflation-linked bonds whereas US inflation-linked bonds have well documented liquidity issues (see Chernov & Mueller (2012) and D'Amico, Kim & Wei (2010)).

[15] For the sake of simplicity and parsimony, we consider the case where inflation is a deterministic process with stochastic expected inflation, which is driven by all priced factors. This reduces the number of parameters to be estimated by N relative to a more general specification with (homoscedastic) stochastic inflation, and implies that the nominal and real prices of risk are the same. The only difference between real and nominal payoffs is the exposure to different risks, not the compensation for each risk. For the more general homoscedastic inflation case, only the level of risk premia would be different. See Appendix A.1.

(e.g. Chernov & Mueller (2012) and Joyce, Lildholt & Sorensen (2010)). There are two reasons I prefer to allow nominal and real (and hence inflation) rates to depend on all factors. First, this partition is at odds with most of the structural models with inflation, where, in equilibrium, every endogenous quantity and price are typically a function of all state variables. Since the model is reduced form, its strength is allowing the factors to be whatever is required to empirically explain the historical behaviour of yields. This includes the possibility that the factors that are important for real rates have little impact on inflation. Arbitrarily restricting the model reduces this strength while still not being a structural model.

Second, and perhaps more importantly, introducing an arbitrary partition on the factors increases the risk that the model will have difficulties in dealing with different regimes. The low frequency changes in inflation and real rates observed over the last 40 years might be captured by a common factor, or a combination of factors, but if inflation and real rates are not allowed to respond to common factors then the model might have trouble fitting the entire sample. More importantly, since I use the same assumption with the different estimation strategies, this should not affect our comparison between them.

3.3 Estimation

Since the factors are latent, we need to impose normalization restrictions on the parameters to achieve identification. I follow the new identification scheme proposed by JSZ , which is what has made this study feasible. Before JSZ, the standard identification scheme for ADTSM followed Dai & Singleton (2000), which normalized the volatility matrix. Unfortunately, because the volatility is measure independent, this effectively created a strong numerical link between the $\mathbb{P}$ and $\mathbb{Q}$ parameters. JSZ propose to leave the volatility matrix unrestricted and instead normalize the short rate and the drift of the $\mathbb{Q}$-dynamics (see Appendix B for details). This not only breaks down the strong dependence between the $\mathbb{P}$ and $\mathbb{Q}$ parameters, but it also allows for much more efficient estimation of each.

Under the assumption that a subset or a portfolio of yields are priced without error, then $\mathbb{P}$ parameters can be estimated by OLS, or at least very good starting points can be obtained, and it is easy to find good starting values for the $\mathbb{Q}$ parameters. Even if we do not want to assume that a subset of yields or portfolios of yields are perfectly priced, we can still use these as good starting points to obtain global maximum, as I do here. The JSZ normalization means the models take at most hours to estimate (for the more highly parameterized and constrained) without the expensive numerical searches that meant potentially weeks or months to have some confidence of a global maximum.[16]

I estimate models by Maximum Likelihood using the Kalman Filter. The Kalman Filter is naturally able to deal with missing data and inclusion of surveys. It has also been shown to provide better estimates than assuming that a subset or a portfolio of yields are priced without error (see Bibkov & Chernov (2009) and Duffee & Stanton (2012)), and is able to accommodate factors *unspanned* by the cross-section (see Duffee (2011a), Chernov & Mueller (2012)). I use the estimation strategy with observed PCs suggested by JSZ as a first step to find good starting values for the Kalman filter. The reduced rank restrictions on price of risk of JSZ and the bias correction of BRW are not feasible with the Kalman Filter estimation. I discuss these separately.

The transition equation is always given by the VAR of the latent factors in Equation (7). The observation equations will depend on the number of bond yields and survey forecasts used. All yields are assumed to be observed with error ($y_{t,m}^{obs} = y_{t,m} + \sigma_n u_m$). The observation equations includes 7 equations for nominal spot yields for bonds with maturities of 6 months, 1, 2, 3, 4, 5, 7 and 10 years:

$$\begin{aligned} y_{t,m}^{obs} &= y_{t,m} + \sigma_n u_m \\ &= A_m^Q + B_m^Q X_t + \sigma_n u_m^n \end{aligned} \tag{10}$$

Restrictions on risk premia will imply restrictions on the $\mathbb{Q}-$parameters that enter the recursive pricing coefficients A_m^Q, B_m^Q (see Appendix B).

When I include surveys forecasts of nominal rates in the estimation, additional observation equations for each survey maturity will be added:

$$y_{t,n-m}^{sur} = A_{n-m}^e + B_{n-m}^e X_t + \sigma_{ns} u_m^{ns} \tag{11}$$

where $A_n^e = -\frac{1}{n}\mathcal{A}\left(\delta_0, \delta_1, \mu, \Phi, 0, n\right)$ and $B_n^e = -\frac{1}{n}\mathcal{B}\left(\delta_0, \delta_1, \mu, \Phi, 0, n\right)$ (see Appendix C).

[16] See Kim (2007), Duffee & Stanton (2008), and a previous version of Duffee (2011b), for a discussion of the difficulties in estimating these models with previous identification methods. Chernov & Mueller (2012) report evaluating the likelihood in 2 billion Sobol points and then optimize the likelihood using the best 20,000 points as starting values.

When I consider jointly the real and nominal yield curves, additional observation equations for each maturity of real yields used in estimation will be of the form

$$y_{t,m}^{r,obs} = A_m^r + B_m^r X_t + \sigma_r u_m^r \tag{12}$$

Finally, when I include surveys on inflation in the estimation of the joint model for UK, for each survey forecast we also have

$$\pi_{t,n-m}^{sur} = A_{n-m}^{\pi e} + B_{n-m}^{\pi e} X_t + \sigma_{is} u_m^{is} \tag{13}$$

For the sake of parsimony, I use a common variance for the observation equation noise for each group (nominal yields, real yields, nominal interest rate surveys and inflation rate surveys) instead of a separate parameter for each maturity in each group.[17]

For models with Sharpe ratio constraints I follow Duffee (2010) and add the restriction on the maximal log returns implied by the model

$$\sqrt{(\Sigma^{-1}\Lambda)_t' \, \Sigma^{-1}\Lambda_t} \leq c \tag{14}$$

which is similar to the restriction on variability of risk premia used by Bibkov & Chernov (2009) and Chernov & Mueller (2012), who add a penalty to the likelihood function proportional to $\Lambda_t' \Lambda_t$. Based on the results in Duffee (2010) I consider constraints of 0.3 and 0.5 for the average maximal Sharpe Ratio.

4 Results

4.1 Nominal decomposition

For a given number of factors, the mean absolute errors for each maturity are very similar across models estimated with different strategies and samples (for sake of space

[17] This simplifying assumption is commonly employed (e.g. Duffee (2011a)), and does not lead to materially different estimates to the case where each maturity is allowed to have different variances (JSZ). There is evidence that errors are cross-sectionally correlated and display significant autocorrelation (Dempster & Tang (2011)), but since this assumption is applied to all models, it should not affect their comparison. Furthermore, Kim & Orphanides (2012) show that allowing for correlated errors in surveys does not affect their results, while being more cumbersome to estimate.

these are included in the online appendix). This confirms the flexibility of these models to fit the data, which also gives rise to their small-sample instability. From the perspective of fit, there would be no strong reason to go beyond 4 factors, and 5 factor models can have an average precision that is higher than the precision with which these yield curves are estimated from bond prices. Because these models fit the data so well, and hence predicted yields are nearly identical across models, comparing term premia estimates across models implies essentially the same degree of variability of the expectations component.

Figure 3 shows the time series of US 10 year spot nominal term premia estimates for models varying by number of factors and sample period and both for unrestricted models (right column) and models estimated using survey forecasts as additional observation equations (left column). The charts only show the estimates for the sample period used in estimating each model. From these charts it is already evident that unrestricted models can provide very different estimates of term premia when different samples are used for the same number of factors. Further, the variability for the same sample with different number of factors is more pronounced when comparing 3 and 5 factor models, particularly for the US. Figure 4 show the equivalent estimates for UK data, with similar resutls.

In contrast, the models estimated using surveys to anchor expectations provide very similar estimates of term premia across models and samples. This is not surprising, and is indeed the desired outcome, and in part is just a reflection that the model is flexible enough to match observed yields with implied expectations that are consistent with survey forecasts. However, this result is also not automatic given that the survey forecasts used are available for much shorter maturities or considerably shorter time series, or both for the case of the UK. I will discuss the role of surveys and the role they play in anchoring estimates further in Sections 5 and 6.

The results for models limiting the number of time-varying risk parameters are particularly poor, raising concerns that they might not be a global solution in the maximum likelihood optimization. The difficulty in estimating the models is a further drawback in this strategy, beyond the arbitrary nature of restrictions. The likelihood concentration result of JSZ for unconstrained models no longer applies (see Appendix B) and the PC might not be good initial proxies. While the models limiting the rank of prices of risk (as in JSZ) are not as problematic, they are not as successful in stabilizing term premia estimates. These restrictions also cannot be estimated with Kalman Filter, which limits the comparison with other strategies.

To directly compare the success of the different strategies, I consider a common sample comparison that highlights the difference in estimated dynamics. I compare the estimated term premia used with different estimation subsamples when filtered for the entire sample using the Kalman Filter. Figure 5 shows the time series of the full sample estimates of US term premia for the unrestricted model, the model with surveys and models with Sharpe ratio constraints of 0.3 and 0.5 (c in Equation (14)). Figure 6 contains the estimates for UK. For the US, the unrestricted model estimates for the sample starting in January 2007 are particularly poor so I do not include them in the charts.[18] It is apparent from these charts that the estimates with surveys differ by less than the unrestricted or those with Sharpe ratio constraints, across models of different size and different estimation samples.[19] In Section 5 I formalize these comparisons using two well known measures of divergence.

4.1.1 Persistence and mean

An interesting question is to investigate the source of gains in stability. Previous literature has focused primarily on the persistence of the factors as a source of the instability (see BRW), with less attention devoted to the level they are expected to converge to. All the estimates of the unconditional mean and half-life are reported in the online appendix.

I find no systematic difference between the largest eigenvalue of the estimated $\mathbb{P}$-dynamics across the models. For all the estimation strategies the sample length seems to dominate, with estimates using shorter samples typically implying lower persistence. The highest estimated half-life tends to be for samples starting in 1982, a sample in which interest rates have a clear negative trend.

There is a clear difference, however, with respect to the unconditional mean interest rate $(\delta_0 + \delta_1 (I_N - \Phi)^{-1})$. The estimates from unrestricted models and models with Sharpe

[18] This might be due to the lack of variability in the US data for this sample, as yields in the US fell earlier than in UK and have since been relatively stable. We also find a tendency of 5 factor models to imply unrealistically high average Sharpe Ratios for US yields, documented by both Duffee (2010) and JSZ. This is not an issue with UK data.

[19] In the online appendix we compare the range of estimates normalized by yield level. This is the relevant measure for the question of how much of observed yields reflect compensation for risk. These charts suggest that anchoring the model with surveys is the more reliable strategy, one that delivers robustness across countries, model dimensions and subsamples, despite differences in the dynamics of yields in the US and UK. This conclusion is not dependent on the choice of 10 year maturity. In the US, instability is more pronounced in 5 factor models (Duffee (2010) and JSZ also report problems with estimates of 5 factor models for US yields) and there are no clear gains in stability of the decomposition from tightening the Sharpe ratio constraint.

ratio restrictions are highly unstable, particularly for the US. In contrast, the unconditional mean interest rate estimate from models with surveys are within a narrow range. I discuss this further in light of the Monte Carlo experiment in Section 5.2.

4.2 Joint nominal-real decomposition

Figure 7 shows the estimates of inflation risk premia with and without surveys (a similar figure for model estimates of real term premia is in the online appendix). These confirm the stabilizing role of survey forecasts in estimating GADTSM for nominal curves shown above. Three points are worth highlighting relative to the results of the nominal models.

First, there is greater uncertainty over the decomposition for the period for which there is neither real yield data (before 1985) nor inflation survey forecasts (before 1990), but this uncertainty does not seem to affect the decomposition for the period for which I have both. This is in contrast with the sensitivity due to lack of survey data for the nominal case. For the UK nominal case I have bond data for the entire sample, but only a short sample of survey forecast data and with short horizons. But even short samples of survey forecasts for relatively short-term horizons is sufficient to stabilize the decomposition of observed yields. The survey forecasts for inflation used here are available for nearly double the sample size I have nominal interest rate survey forecasts and for longer maturities, but bond data is missing for roughly a third of the entire sample. This is related to spanning of $\mathbb{Q}$-dynamics, which I explore further in Section 5. In the absence of observed yields there is an additional identification problem which is worse than just decomposing an observed price.

Second, the joint model estimates anchored with surveys reveal a real term premia that is typically negative and more stable than the inflation risk premia, particularly for the sample period for which I have surveys. This is not apparent from the unrestricted model estimates. These estimates are consistent with the economic intuition that real term premia should be negative (see Campbell, Shiller & Viceira (2009)) and that inflation risk premia accounts for a large fraction of nominal term premia variation (Le & Singleton (2013)).

Third, there was a negative inflation risk premia at the height of the crisis, which is consistent with inflation options market prices and probabilistic survey forecasts of inflation (see Guimarães (2012) and Smith (2012)). This is in line with the intuition of Campbell, Sundaram & Viceira (2010), who argue that when bad times for the economy are associated with deflation, the term premia on nominal bonds can reflect a negative inflation risk premia. This suggests that the argument that nominal term premia should be counter-cyclical is true only for periods when the covariance of inflation risks and the state of the economy is not positive.

5 The importance of the 'cross-section'

To make the stability comparison more formal, I propose to use standard tests of discrepancy of empirical distributions applied to the model implied dynamics from each measure. Specifically, I propose to use the two-sample Cramér-von Mises test of global discrepancy. I apply this test to the empirical distributions of model-implied expectations under $\mathbb{P}$ and $\mathbb{Q}$ to measure the discrepancy in dynamics under both measures for the different models. I do this for several maturities to identify how well anchored each of these two measures are at different horizons.

Let the empirical distribution of the $\mathbb{T}$-expectation of yields with maturity n from model i be denoted as $\mathbb{F}^{\mathbb{T}}_{n,i}$, for $\mathbb{T} = \{\mathbb{P}, \mathbb{Q}\}$. For example, $\mathbb{F}^{\mathbb{P}}_{10,i}$ is the empirical distribution of the time series of the model i estimate of the 10 year $\mathbb{P}$-expected average interest rate $\left\{y^{\mathbb{P},i}_{s,10}\right\}_{s=t:T}$. The Cramér-von Mises (CvM) distances between Models i, j for maturity n are given by:

$$CvM^{\mathbb{T}}_{i,j}(n) = T \int \left(\mathbb{F}^{\mathbb{T}}_{n,,j} - \mathbb{F}^{\mathbb{T}}_{n,i}\right)^2 d\mathbb{F}^{\mathbb{T}}_{n,i+j} \tag{15}$$

where $\mathbb{F}^{\mathbb{T}}_{n,i+j} = \frac{1}{2}\left(\mathbb{F}^{\mathbb{T}}_{n,,j} + \mathbb{F}^{\mathbb{T}}_{n,i}\right)$.[20]

Table 2 shows the median p-values for the null hypothesis of identical $\mathbb{Q}$-distributions across sample periods for the different models and number of factors for US and UK. As in Figures 5 and 6, I use the full sample model implications (which implies I am using backward out-of-sample predictions for all the models except the one estimated with the full sample). We can see from the table that the $\mathbb{Q}$-dynamics are well anchored across all models, at least up to the maturity of yields used in estimation. We cannot reject the model's average $\mathbb{Q}$-expectations up to 10 ahead, and sometimes up to 30 years, come from the same distribution.

[20] I have also implemented the Kolmogorov-Smirnov test, given by $KS^{\mathbb{T}}_{i,j}(n) = \sqrt{T}\left\|\mathbb{F}^{\mathbb{T}}_{n,,j} - \mathbb{F}^{\mathbb{T}}_{n,i}\right\|_{\infty}$, with nearly identical resutls.

The results for $\mathbb{P}$-dynamics, shown in Table 3, confirms the instability and impact of surveys. While for the unrestricted model there is clear evidence against the hypothesis that the estimates of the $\mathbb{P}$-expectations come from the same distribution beyond 2 year maturity (and often even for the 1 year), Table 3 shows that we cannot reject the null for maturities of up to 5 years for the model estimates using surveys. Models with restricted risk premia improve a little relative to the unrestricted model, mainly for the larger models. The somewhat stronger results for anchoring with the US are likely due to the longer availability of surveys (quarterly from 1981 for US and 1999 for UK), while the weaker stability relative to $\mathbb{Q}$-expectations is in line with the smaller 'cross-section' under $\mathbb{P}$ (both in maturities and in number of observations for each maturity).[21] Our Monte Carlo results are consistent with this interpretation.

Table 4 illustrates the intuition for the importance of the amount of cross-sectional information more clearly. The table shows the same statistics as in the previous tables for three unrestricted model specifications with alternative amounts of cross-section information for the $\mathbb{Q}$-dynamics. Besides the benchmark unrestricted model, two additional test models where two bonds are removed from estimation are considered. The first, labeled "without intermediate", removes the 4 year and 7 year US bond yields, while the second removes the longest maturities (the 7 year and 10 year), labeled "without long". I used the same samples and number of factors for each alternative. While the model without the intermediate maturities shows similar stability in $\mathbb{Q}$-dynamics as the benchmark model, the model without the longer maturities is clearly de-anchored at the maturities removed from estimation.

Creal & Wu (2013) claim that the reason the $\mathbb{Q}$-dynamics are precisely estimated is because of the "high powered polynomial functions of Φ^Q in the bond loadings" and not because of the large amount of information in the cross-section. However, in their exercise, Creal & Wu (2013) keep the longer-maturity bonds and only vary the amount of intermediate maturity bonds used in estimation to conclude that the amount of cross-sectional information is not important. Our results show that the length of the cross-section matters, confirming the intuition provided in Section 2. The longest maturity will determine the length of the sample from the $\mathbb{Q}$-dynamics observed on each date, which

[21] For the US we have a total of 117 observations of survey forecasts for short horizons, and only 19 for the 10 year average, whereas we have 467 observations of bond prices of each maturity. For the UK the number of survey forecast observations is only 45.

matters more for estimating the dynamics than additional intermediate observations. This is also confirmed by the Monte Carlo results in Section 5.2. The difference between the cross-section and time series is not so much about 'high powered polynomials' as one of error structure.[22]

I have chosen the data to use in this study to match the standard choice of maturities used in empirical applications of DTSM for comparison. Typically, the longest maturity used in estimation is the 10 year yield. An exception is Christensen, Diebold & Rudebusch (2011), who use 15, 20 and 30 year yields in estimating their affine Nelson-Siegel models. In fact, it is more common to have models estimated with maturities of 5 years or less (e.g. Bauer (2011), Duffee (2011b)). Our results, however, suggest that it is important to include longer maturities in estimation. Our estimates of longer-maturity bond yields differ by several percentage points for maturities of 30 years. Both for the UK and US, bond with maturities of 30 years are available for long periods (though not since 1972) and could be included to further help identify the model.

5.1 Relation to unspanned macro risks

The UK joint model offers an interesting opportunity to test if the cross-section intuition extends to macro variables and illustrate some of the identification difficulties raised by JPS. I re-estimate the joint term structure model with less information on the real yield curve, but maintaining the surveys on nominal interest rates and inflation rates as before. Furthermore, I start the optimization from the maximum likelihood parameter estimate using all of the real yield curve data and use the 10 year inflation-linked bond, both advantages an econometrician would not have when asset prices contingent on a macro variable are unavailable. Table 5 compares the model fit of the model with all of the data as described in Section 3 with the fit of a model with only the 10 year spot real yield from the real yield curve. The table shows that the fit deteriorates significantly, with average absolute errors 3 to 6 times larger for shorter maturities, and maximum errors of around 300 basis points or higher.

[22] The time series realisation of interest rate can be written as the conditional expectation relative to an initial state, which is a polynomial function of the intial state ($E_t\left[i_{t+k}\right] = \delta_0 + \mu \sum_{i=0}^{k-1} \Phi^i + \delta_1 \Phi^k X_t$), and weighted sum of all of the previous shocks ($i_{t+k} = E_t\left[i_{t+k}\right] + \delta_1 \sum_{i=1}^{k} \Phi^{k-i} \Sigma \varepsilon_{t+i}$) whereas the forecasts of future interest rates will have the same 'high powered polynomials' ($f_{t,k} = E_t\left[i_{t+k}\right] + \Omega u_{t,k}$) , but only have a measurement error.

For the UK joint model, we know that RPI $\mathbb{Q}$-dynamics should be spanned in the sense that inflation-linked bonds in the UK are a function of RPI. However, without the information from inflation-linked bonds I have shown I cannot pin down the real yield from nominal yields even with the help of surveys on RPI inflation. This is true despite using plenty of information on the RPI $\mathbb{P}$-dynamics and even allowing information that the econometrician without observed macro asset prices would not have: the estimation starts from the ML point estimate using the full real bond term structure and the longer maturity (10 year) real bond yield. For this exercise, the fact I am using a parsimonious joint model (see Section 3.2.1) should enhance our chances of identifying the real term structure by reducing the number of parameters that are only identifiable from real bond prices.

Although the RPI $\mathbb{Q}$-dynamics information is economically spanned by the bonds, and I am using survey information on expected RPI inflation[23], unless I use the inflation-linked bonds price data I cannot identify the $\mathbb{Q}$-dynamics properly. This is another example of missing information on the $\mathbb{Q}$ dynamics rather than $\mathbb{P}$. It is not specific about macro risks.

5.2 Monte Carlo evidence

The Monte Carlo design is meant to measure the benefit of adding a cross-section of interest rate forecasts for model estimation. I therefore compare standard VAR estimates from different sample sizes with estimates enhanced by different amount of cross-section forecasts. Because we have seen that the $\mathbb{Q}$-dynamics are estimated very precisely, and in view of the near concentration of the likelihood result of JSZ, I focus only on the drift parameters of the VAR (μ, Φ). The details of the Monte Carlo are relegated to Appendix D. The amount of cross-section forecasts added, and degree of noise of forecasts, are also varied to replicate the amount of information available in bond prices ('Like Q') and availability of survey forecasts in the US ('Data US') and UK ('Data UK').

Robustness tests

Table 6 shows the median p-values of the CvM test applied to 10 different forecast horizons, for the different time series and cross-section estimates of the baseline specification ('MC1' in Appendix D).[24] Panel A shows the results for estimates using only the time series

[23] We are using survey forecasts for the RPI measure of inflation, which is the same measure the UK inflation-linked bonds are indexed to, so there is no 'correct inflation' measurement issue.

[24] The equivalent tables for the other Monte Carlo designs (MC2, MC3, and MC4) are shown in the online appendix. They confirm the results shown here for more persistent and/or higher dimensional models.

for 11 different sample lengths, ranging from 20 to 1000 years. Panel B shows the median p-values for 3 different degrees of cross-section information as well as 3 levels of cross-section precision each, for the same forecast maturities as in Panel A.

The p-values in Table 6 are not directly comparable to those in Table 2 and 3 because the latter were calculated over estimates with shrinking sample size. Nevertheless, we can see that the degree of stability of estimates replicates several of the results seen earlier.

First, the model with monthly forecasts, with same maturities and a similar precision as observed for yields (column Like Q, Yields) display the same level of stability we saw for $\mathbb{Q}$-dynamics, which is higher than we would achieve from time series estimates using 1000 years of monthly observations. Second, as the amount of cross-section information is reduced ('Like Q' to the other two), the stability of estimated model dynamics falls, with longer-horizon forecast more useful than short ones ('Data US' versus 'Data UK'), as we observed for $\mathbb{P}$-dynamics. Third, the gain in stability in estimated dynamics are sizable: even adding very noisy cross-section forecasts ('Noise' columns) is equivalent to roughly doubling the sample size. With forecasts with similar relative precision as in surveys (or yields) the stability achieved is equivalent to observing centuries (or millennia) of data. These conclusions extend to the other Monte Carlo specifications described in the appendix, which increase persistence and number of factors.

Persistence and mean

Table 7 shows the distribution of estimated unconditional mean of interest rates and the half-life of the factors for the same models as the previous table. With respect to the unconditional mean, all models are roughly unbiased, in the sense that the median of the estimates are very close to the true value (4.96%), but with significant variation in dispersion. The relative tightness of the bounds confirms the conclusions drawn from the tests of equality of dynamics: the higher the precision and availability of forecasts, the tighter the bounds, with longer-term forecasts proving more useful (US versus UK). With the exception of the estimate using noisy and short-term forecasts ('Data UK, Noise'), the gains are significant relative to the time series estimates. As above, the gains are equivalent to centuries ('Data US, Surveys') or millennia ('Like Q, Yields') of data.

Regarding the estimated persistence, the pattern of dispersion of estimates are very similar to those for the unconditional mean. The main difference is that the estimates are biased downwards, as is well known for persistent time series. Even with 1000 year samples, the median estimated half-life is still lower than the true value (7 years). As before, the precision of forecasts is a significant factor in reducing bias of the forecast-enhanced estimates. It is interesting to note that the case of only short-term forecasts remains severely biased even when surveys are as precise as yields in the data ('Data UK, Yields'), highlighting the importance of the length of the cross-section once again.

Out-of-Sample forecasts

The comparison of the out-of-sample forecasts using our artificial samples suggests these are poor metrics for evaluating these models. Unlike the noticeable differences seen in stability and bias that adding different quantities and precision of cross-section forecasts deliver, the out-of-sample root mean squared error (RMSE) of forecasts shows only marginal differences across estimates. Similar evidence was presented in Kim & Orphanides (2012), who showed that a worse model could result in better RMSE (in sample) because the look-ahead bias would dominate the improvements in RMSE from reduction in the bias of estimates. For sake of space, we do not report these results.

6 Survey forecasts

The results show that adding surveys as suggested by Kim & Orphanides (2012) is a better alternative to obtain stable decompositions of the term structure of yields. This raises two issues: the availability of survey forecasts and whether it is desirable to have model implied expectations broadly consistent, or anchored, to survey forecasts.

The results in Section 4.1 suggest that even short samples of surveys for short horizons are sufficient to significantly stabilize the decomposition. These tend to be available for most industrialized countries. Therefore, for the economies for which reasonable time series of yields are available, survey availability should not be a major issue. The issue of survey frequency is addressed below.

There is substantial evidence that survey forecasts provide sensible proxies for expected dynamics. There is evidence that surveys from professional forecasters are consistent with theory[25] and have good forecasting properties, not only for interest rates and inflation.[26] There is also growing evidence that forecasts from structural macro DSGE models also benefit by anchoring model implied expectations to survey forecasts.[27] Even if not the best on forecasting horse races, models that roughly match survey forecasts have the benefit that they are automatically sensible in different time periods and environments, not subject to data snooping and consistent with information sets in real time. This is not true about the implied expectations from models that restrict risk premia, or that employ purely statistical methods, as I discuss in Section 6.2.

More importantly, the study of Chernov & Mueller (2012) directly tackles the question of consistency between survey forecasts and bond expectations and conclude they are consistent. They compare a model that allows for differences between the (objective) expectations in bonds and the (subjective) expectations of forecasters and prefer the model where the two expectations are the same. In a separate exercise, they also conclude that including survey forecasts improves the model forecasting performance.

Because surveys are relatively infrequent (surveys used in this paper are quarterly, some semiannual or even annual), another objection to using surveys is that it limits the high frequency application of the models. Yet another objection, is that by anchoring the model to surveys we might be losing the independent information in market yields. To address these concerns I now examine the impact of including surveys when filtering the states, and the resulting decomposition, once I have estimated the parameters. I answer the question: conditional on parameter identification, how much does inclusion of survey forecasts distort the information contained in bond prices?

6.1 Sensitivity of decomposition to filtering latent factors using surveys

The goal is to ascertain how much the decomposition is being distorted by requiring the model to match surveys beyond helping to estimate the parameters. This amounts

[25] Carroll (2003) and Easaw & Golinelli (2010) show that surveys of professional forecasters are consistent with rational expectations and tend to lead surveys of households. Carvalho & Necchio (2012) show that the SPF forecasts are consistent with predictions from monetary economics theory.

[26] For evidence on the forecasting power of surveys of interest rates and inflation see Ang, Bekaert & Wei (2007), Duffee (2012a), Faust & Wright (2012), Chun (2011) and Altavilla, Giacomini & Ragusa (2013). For a discussion of forecasting of macro variables see Hess & Orbe (2013).

[27] See Fuhrer (2013), Del Negro & Schorfheide (2012) and Milani & Rajbhandari (2012).

to comparing the filtered latent factors, and resulting decompositions, with and without the observation equations for surveys in the Kalman Filter. Table 8 shows some of the percentiles of the absolute differences between the estimates of spot term premia for the models estimated using survey forecasts when the surveys are not used in filtering the states. The differences are calculated for spot term premia with horizons between 10 and 20 years, and are shown for each of the subsamples used in estimation and the different number of factors used.

The table shows that for the US (top panel), for both 3 and 4 factors there is virtually no difference between the decompositions, with absolute differences of only a few basis points. The exception is for models estimated for the 2007-2010 subsample, but the differences are still small relative to the range of estimates across models, with 95% of the differences smaller than 5 basis points. For 5 factor models the impact is larger, with median absolute difference between 6 and 13 basis points depending on subsample. Similar results apply to the UK, shown in the bottom panel.

While this relative insensitivity to surveys has come out as the result of the optimization, where the precision of survey forecasts has been chosen to maximize the log-likelihood (with estimated survey precision always much smaller than for yields), in practice this can be ensured by imposing a cap on the precision of surveys versus yields, or even choosing this precision. This would guarantee that the filtered states, and hence the model decompositions, are not very sensitive to the inclusion of surveys while avoiding implied expectations that are clearly incompatible with surveys. Given the relatively low estimation cost of these models, researchers can consider a range of estimates with varying degrees of relative precision. This can be viewed as an approximate way of imposing a prior on the informativeness of different instruments.

6.2 Sensible expectations

Two additional methods seem worth discussing here to highlight the benefits of using survey forecasts to anchor the expectations component: the bias correction method of BRW, a purely statistical method, and the Cochrane & Piazzesi (2008) model, which is the most restricted risk premia specification while allowing for time-varying risk premia. In the Cochrane & Piazzesi (2008) model, only one parameter in λ_0 and one parameter in λ_1 are allowed to be non-zero. These models are good candidates to test the hypothesis that over-parameterized risk premia and statistical biases are the main culprits of the instability of GADTSM, as opposed to the cross-section information.

The estimates from Cochrane & Piazzesi (2008) imply longer-term expectations that are more volatile than short-term expectations. This behaviour of expectations is hard to reconcile with basic predictions of traditional economic models and with observed survey forecasts. The mean reversion present in most surveys and asset prices is a feature not only of interest rates, but also for a wide variety of economic and financial variables.

Similar problems result from the bias-correction estimates of BRW. The expectations from BRW bias correction are almost as volatile at 10 years ahead as they are for 1 month ahead (see Panel A of Figure 8). The bias corrected estimate suggests that the average expected 1 month rate in 10 years time in early 2000s was just as low as the observed 1 month rate, around 1%, which is clearly inconsistent with the term structure of survey forecasts from the SPF, as shown in Panel B of Figure 8. In addition, the bias corrected estimate of BRW implies negative 1 month forward rates 20 years ahead for more than half (51.4%) of the sample between 1990 and 2007, even before rates reached the zero lower bound (ZLB). In contrast, none of the models estimated with surveys imply long-term forecasts of 1 month rates that are negative even after reaching the ZLB in early 2009.

This suggests that just aiming to increase the persistence of estimated dynamics is not the solution to achieving reasonable decompositions for policy analysis.[28] While the OLS estimates might underestimate the persistence of yields in small samples, the BRW estimate of 0.9991 (see Table 2 in BRW) seems to be too high compared to the persistence in the data (see Table 1) or the estimated persistence for the $\mathbb{Q}$-dynamics. In addition, it is not clear that the high risk-adjusted persistence is a good indicator for actual persistence.[29]

7 Conclusion

I have shown evidence for the US and UK nominal yield curves that anchoring model-implied expectations with survey forecasts of future rates seems to be the most reliable way of achieving robust and sensible model decompositions. Restricting risk premia seems neither necessary nor sufficient to achieve sensible yield curve decompositions. The same

[28] See also the discussion in Wright (2013) of the term premia estimates of BRW.

[29] Collin-Dufresne, Johannes & Lochstoer (2013) provide an example of a general equilibrium model in which parameter learning by agents leads to risk-adjusted dynamics being more persistent than actual.

can be said about correcting for the persistence of yields: while some unrestricted models might tend to severely underestimate the persistence of yields, implying expected rates at the 10 year horizon and beyond that are too stable (and hence attributing all the changes in longer-term forward yields to risk premia), increasing the persistence *per se* is not enough to guarantee sensible decompositions.

The evidence suggests that the lack of robust estimates of DTSM is an informational problem similar to other cross-sectional spanning issues identified in the literature. This does not seem to be a problem specific to GADTSM, or even ADTSM. Our results points to not only including surveys to increase the $\mathbb{P}$ cross-section, but also using yield data with longer maturities. To the extent that surveys do not seem to anchor the estimate of persistence, which tends to be lower for shorter samples, if the researcher does not desire to use a long sample, then methods to correct for this bias (e.g. Jardet, Monfort & Pegoraro (2013), BRW, JSZ, Minnesota prior) in addition to surveys should be considered. While models anchored with surveys give very similar decompositions for the period for which data is used, the implied dynamics will diverge significantly for long maturities.

An interesting empirical question is the substitutability between adding macro variables with the unspanned framework of JPS, which improves the $\mathbb{P}$-dynamics, and survey forecasts. One inherent advantage of surveys, which might more than compensate any improvement in forecasting from using macro variables, is that they are truly real-time proxies of expectations with no look-ahead bias. Any model with macro variables will benefit from hindsight (either from information not available in real time or in choice of variables and methods) and might therefore be subject to the usual out-of-sample instability. Pseudo out-of-sample tests are notoriously unreliable. As argued by Kim & Orphanides (2012), out-of-sample improvements are not necessarily a reflection of better measure of expectations, as some outcomes were not expected in real time. The current crises and aftermath are a perfect example.

A Nominal and real Affine pricing recursions

If we assume that nominal bond prices are exponential-affine:

$$P_{t,n} = \exp(A_n + B_n X_t)$$

combined with the no-arbitrage condition

$$P_{t,n} = E_t^Q \left[e^{-i_t} P_{t+1,n-1} \right]$$

we obtain the recursion:

$$\begin{aligned}
A_n + B_n X_t &= \ln P_{t,n} \\
&= \ln E_t^Q \left[\exp\left\{ -i_t + A_{n-1} + B_{n-1} X_{t+1} \right\} \right] \\
&= \ln E_t^Q \left[\exp\left\{ -\left(\delta_0 + \delta_1 X_t\right) + A_{n-1} + B_{n-1} \left(\mu^Q + \Phi^Q X_t + \Sigma \varepsilon_{t+1}^Q \right) \right\} \right] \\
&= -\left(\delta_0 + \delta_1 X_t\right) + A_{n-1} + B_{n-1} \left(\mu^Q + \Phi^Q X_t \right) + \ln E_t^Q \left[\exp\left\{ B_{n-1} \Sigma \varepsilon_{t+1}^Q \right\} \right] \\
&= -\left(\delta_0 + \delta_1 X_t\right) + A_{n-1} + B_{n-1} \left(\mu^Q + \Phi^Q X_t \right) + \frac{1}{2} B_{n-1} \Sigma \Sigma' B_{n-1}^{r\prime}
\end{aligned}$$

$$\begin{aligned}
A_n &= -\delta_0 + A_{n-1} + \frac{1}{2} B_{n-1} \Sigma \Sigma' B_{n-1}^{r\prime} + B_{n-1} \mu^Q \equiv \mathcal{A}\left(\delta_0, \delta, \mu^Q, \Phi^Q, \Sigma, n\right) \\
B_n &= -\delta_1 + B_{n-1} \Phi^Q \equiv \mathcal{B}\left(\delta, \Phi^Q, n\right)
\end{aligned}$$

with initial conditions $A_0, B_0 = 0$, so

$$y_{t,n} = A_n^Q + B_n^Q X_t$$

where

$$\begin{aligned}
A_n^Q &= -\frac{1}{n} \mathcal{A}\left(\delta_0, \delta_1, \mu^Q, \Phi^Q, \Sigma, n\right) \\
B_n^Q &= -\frac{1}{n} \mathcal{B}\left(\delta_1, \Phi^Q, n\right)
\end{aligned}$$

To calculate the yields that would hold under expectation hypothesis, that is, the yields that would hold if agents utility functions were linear[30]

$$\begin{aligned}
y_{t,n}^P &= -\frac{1}{n} \ln E_t \left[\exp\left\{ -i_t + A_{n-1} + B_{n-1} X_{t+1} \right\} \right] \\
&= A_n^P + B_n^P X_t
\end{aligned}$$

[30] Some refer to these as risk-neutral rates (e.g. BRW), but because the risk-adjusted ($\mathbb{Q}$) measure is also commonly referred to as risk-neutral measure we avoid using risk-neutral rates to prevent any confusion.

all we need to do is use the $\mathbb{P}$-dynamics parameters in the recursions above:

$$\begin{aligned} A_n^P &= -\frac{1}{n}\mathcal{A}\left(\delta_0, \delta_1, \mu, \Phi, \Sigma, n\right) \\ B_n^P &= -\frac{1}{n}\mathcal{B}\left(\delta_1, \Phi, n\right) \end{aligned}$$

It is worth pointing out that this is not the expectation of future interest rates because of a convexity term (see Appendix C).

From these, spot term premia is easily calculated as

$$tp_{t,n} = \left(A_n^Q - A_n^P\right) + \left(B_n^Q - B_n^P\right) X_t$$

We can also easily calculate forward rates starting in n periods in the future maturing in $m > n$ periods, $y_{t,n-m} = (m-n)^{-1}\left(my_{t,m} - ny_{t,n}\right)$ which becomes

$$y_{t,n-m} = A_{n-m}^Q + B_{n-m}^Q X_t$$

with $A_{n-m}^Q = (m-n)^{-1}\left[mA_m^Q - nA_n^Q\right]$ and $B_{n-m}^Q = (m-n)^{-1}\left[mB_m^Q - nB_n^Q\right]$.

For the real curve we have the same pricing equations but with the real risk-free short rate so the same affine yield result holds

$$y_{t,n}^r = A_n^r + B_n^r X_t$$

with:

$$\begin{aligned} A_n^r &= -\frac{1}{n}\mathcal{A}\left(\rho_0, \rho_1, \mu^Q, \Phi^Q, \Sigma, n\right) \\ B_n^r &= -\frac{1}{n}\mathcal{B}\left(\rho_0, \rho_1, \mu^Q, \Phi^Q, \Sigma, n\right) \end{aligned}$$

We could have also used the SDF to compute the bond prices using prices of risk and $\mathbb{P}$-dynamics:

$$\begin{aligned} \ln P_{t,n} &= A_n + B_n X_t \\ &= \ln E_t^Q\left[\exp\left\{-i_t + A_{n-1} + B_{n-1}X_{t+1}\right\}\right] \\ &= \ln E_t\left[\exp\left\{m_{t+1} + A_{n-1} + B_{n-1}X_{t+1}\right\}\right] \\ &= \ln E_t\left[\exp\left\{\begin{array}{c} -\left(\delta_0 + \delta_1 X_t\right) - \frac{1}{2}\left(\lambda_0 + \lambda_1 X_t\right)'\left(\lambda_0 + \lambda_1 X_t\right) - \left(\lambda_0 + \lambda_1 X_t\right)'\varepsilon_{t+1} \\ +A_{n-1} + B_{n-1}\left(\mu + \Phi X_t + \Sigma\varepsilon_{t+1}\right) \end{array}\right\}\right] \\ &= -\left(\delta_0 + \delta_1 X_t\right) + A_{n-1} - B_{n-1}\Sigma\left(\lambda_0 + \lambda_1 X_t\right) + \frac{1}{2}B_{n-1}\Sigma\Sigma' B_{n-1}^{r\prime} + B_{n-1}\left(\mu + \Phi X_t\right) \end{aligned}$$

$$\begin{aligned} A_n &= -\delta_0 + A_{n-1} + \frac{1}{2}B_{n-1}\Sigma\Sigma' B_{n-1}^{r\prime} + B_{n-1}\left(\mu - \Sigma\lambda_0\right) \equiv \mathcal{A}\left(\delta_0, \delta, \mu^Q, \Phi^Q, \Sigma, n\right) \\ B_n &= -\delta_1 + B_{n-1}\left(\Phi - \Sigma\lambda_1\right) \equiv \mathcal{B}\left(\delta_1, \Phi^Q, n\right) \end{aligned}$$

using $\mu^Q = \mu - \Sigma\lambda_0$ and $\Phi^Q = \Phi - \Sigma\lambda_1$.

A.1 Link between real and nominal SDFs

Let the real SDF $\hat{M}$ be given by

$$\ln \hat{M}_{t+1} = \hat{m}_{t+1} = -r_t - \frac{\hat{\Lambda}_t' \hat{\Lambda}_t}{2} - \Lambda_t' \varepsilon_{t+1}, \qquad (16)$$

and log inflation be given by

$$\pi_{t+1} = \mu_t - \frac{\sigma_t' \sigma_t}{2} + \sigma_t' \varepsilon_{t+1}$$

so that μ_t is the expected inflation rate. Then the log nominal SDF will be given by

$$\begin{aligned} m_{t+1} &= \hat{m}_{t+1} - \pi_{t+1} & (17) \\ &= -r_t - \frac{\hat{\Lambda}_t' \hat{\Lambda}_t}{2} - \hat{\Lambda}_t' \varepsilon_{t+1} - \left(\mu_t - \frac{\sigma_t^2}{2} + \sigma_t' \varepsilon_{t+1} \right) \\ &= -r_t - \mu_t - \left(\frac{\hat{\Lambda}_t' \hat{\Lambda}_t - \sigma_t' \sigma_t}{2} \right) - \left(\hat{\Lambda}_t + \sigma_t \right)' \varepsilon_{t+1} \\ &= -i_t - \Lambda_t' \varepsilon_{t+1} & (18) \end{aligned}$$

where

$$\begin{aligned} \Lambda_t &= \hat{\Lambda}_t + \sigma_t \\ i_t &= r_t + \mu_t + \sigma_t' \left(\sigma_t + \hat{\Lambda}_t \right) \end{aligned}$$

So as long as the price process is not a stochastic volatility process (or the stochastic volatility is uncorrelated to the real price of risk) then we can think of specifying either the real or nominal price of risk as affine and 2 of real, nominal or inflation rates as affine and the remaining one will also be affine. Campbell, Sundaram & Viceira (2010) is an example of affine real SDF with inflation following a stochastic volatility process, which results in a quadratic nominal term structure model.

B JSZ identification

JSZ propose a normalization that not only allows for identification of the model described in Section 3.2, but also simplifies the task of finding a global maximum of the likelihood function. With this normalization, there is a near separation of the likelihood in $\mathbb{P}$ and $\mathbb{Q}$ parameters and it is easier to estimate each of them.

JSZ show that any canonical GADTSM (as described in Section 3.2) is observationally equivalent to a canonical GADTSM where the parameters governing bond pricing, $\left(\delta_0, \delta, \mu^Q, \Phi^Q, \Sigma\right)$, are uniquely mapped into a smaller set of parameters $\left(k_\infty^{\mathbb{Q}}, \lambda^{\mathbb{Q}}, \Sigma\right)$, where $k_\infty^{\mathbb{Q}}$ is proportional to the long-run mean of the short rate under $\mathbb{Q}$ and $\lambda^{\mathbb{Q}}$ is the N-vector of ordered eigenvalues. In the JSZ normalization, with parameter set $\Theta^{JSZ} = \left\{k_\infty^{\mathbb{Q}}, \lambda^{\mathbb{Q}}, \mu, \Phi, \Sigma\right\}$, (i) the risk-free rate (Equation (2)) becomes $r_t = \iota X_t$, where ι is a vector of ones; (ii) Σ is lower triangular with positive diagonal entries, (iii) and the $\mathbb{Q}$-drift is given by $\mu_1^Q = k_\infty^{\mathbb{Q}}$ and $\mu_i^Q = 0$ for $i \neq 1$, and $\Phi^Q = J\left(\lambda^Q\right)$ is in real Jordan form (see JSZ Appendix C for the real Jordan form for different assumptions on eigenvalues).

B.1 OLS estimates with PCs

Under the assumption that N portfolios of bonds are priced perfectly, $\mathcal{P}_t \equiv W y_t = W y_t^{obs} \equiv \mathcal{P}_t^{obs}$, then there is full separation of the likelihood and OLS recovers the ML estimates of $\{\mu, \Phi\}$. The observed yields are allowed to differ from their model-implied counterpart through a J-vector of measurement errors $u_t \sim N\left(0, \sigma_u^2 I_J\right)$. Note that here it is assumed for simplicity that the variance of the measurement errors is the same across all long-term yields used to fit the model.

The likelihood function of the model is then given by

$$\mathcal{L}\left(y_t^{obs}, \mathcal{P}_t | \mathcal{P}_{t-1}; \Theta, \sigma_u\right) = \mathcal{L}\left(y_t^{obs} | \mathcal{P}_t; k_\infty^{\mathbb{Q}}, \lambda^{\mathbb{Q}}, \Sigma, \sigma_u\right) \times \mathcal{L}\left(\mathcal{P}_t | \mathcal{P}_{t-1}; \mu, \Phi, \Sigma\right) \tag{19}$$

where y_t^{obs} is the vector of observed yields. A convenient feature of the normalization proposed by JSZ is that the ML estimate of μ and Φ, that is $\hat{\mu}$ and $\hat{\Phi}$, are obtained by OLS. Conditional on $\left\{\hat{\mu}, \hat{\Phi}\right\}$, an optimization algorithm searches for the values of $r_\infty^{\mathbb{Q}}$, $\lambda^{\mathbb{Q}}$, Σ, and σ_u in order to find the global maximum of the likelihood function. First, good starting values for the parameters in Σ can be obtained by running an OLS regression of $\mathcal{P}_t$ on $\mathcal{P}_{t-1}$. Also, good starting values for $k_\infty^{\mathbb{Q}}$ and $\lambda^{\mathbb{Q}}$ are not difficult to obtain because these parameters are rotation-invariant and therefore carry economic interpretation.

A natural candidate for the portfolio of bonds are the PC of yields, since following Litterman & Scheinkman (1991) it has been well documented the first 3 PC typically explain well over 95% of variation in yields.

B.2 Good starting values for Kalman Filter

Even if we don't make the assumption that N portfolios of bonds are priced perfectly, we can still use PC-based optimization as good starting points for the Kalman Filter optimization. Since the first 3 PC of yields explain a very large fraction of the variation in yields and are easy to compute, and $\mathcal{P}_t \cong \mathcal{P}_t^{obs}$. By choosing the PC to obtain initial estimates we can expect the errors for individual yields will be reasonably small and so the regression of $\mathcal{P}_t^{obs}$ on $\mathcal{P}_{t-1}^{obs}$ should provide good starting points for the ML estimates of $\{\mu, \Phi\}$ to the extent that $\mathcal{P}_t \cong \mathcal{P}_t^{obs}$.

While the likelihood concentration does not hold exactly if all yields (and portfolio of bonds) are observed with error, the $\mathbb{P}$ parameters should have a weak dependency on $\mathbb{Q}$ parameters to the extent that the filtered states do not vary much when $\mathbb{Q}$ parameters change. Indeed, for 3 factors the Kalman Filter is very fast because the filtered factors end up being very similar to the first 3 PC of yields. For 5 factor models the optimization with Kalman Filter takes substantially longer as the 4th and 5th filtered factors are not as close to the respective PC of yields.

C Matching surveys of nominal interest rates and inflation

If we wish to calculate expected nominal spot rates (not just risk neutral), then we can set $\Sigma = 0$:

$$
\begin{aligned}
y^e_{t,n} &= \frac{1}{n} E_t \left[\sum_{i=0}^{n-1} i_{t+i} \right] \\
&= \frac{1}{n} E_t \left[\sum_{i=0}^{n-1} \left(\delta_0 + \delta_1 X_{t+i} \right) \right] \\
&= -\frac{1}{n} \tilde{A}^e_n - \frac{1}{n} \tilde{B}^e_n X_t \\
&= A^e_n + B^e_n X_t
\end{aligned}
$$

$$
\begin{aligned}
\tilde{A}^e_n &= -\delta_0 + \tilde{A}^e_{n-1} + \tilde{B}^e_{n-1} \mu \\
\tilde{B}^e_n &= -\delta_1 + \tilde{B}^e_{n-1} \Phi
\end{aligned}
$$

with $A^e_0, B^e_0 = 0$. We can therefore use the same recursions as before with a zero matrix in place of the variance-covariance matrix Σ:

$$
\begin{aligned}
A^e_n &= -\frac{1}{n} \mathcal{A} \left(\delta_0, \delta_1, \mu, \Phi, 0, n \right) \\
B^e_n &= -\frac{1}{n} \mathcal{B} \left(\delta_1, \Phi, n \right)
\end{aligned}
$$

The difference between risk-neutral yields and expected yields is the convexity effect which in essence is a Jensen term (bonds are the expected value of a convex function of interest rates). In homoscedastic models like those considered here these effects will be constant for each maturity.

To match expected nominal forward rates, which is what most of the survey forecasts are about, we can use

$$
y^e_{t,n-m} = A^e_{n-m} + B^e_{n-m} X_t
$$

To match forecasts of forward real rates we have the same formulas, but using the real short-rate parameters in the recursion, as shown in Appendix A. To calculate expected forward inflation rates, which are the difference between the nominal and real forward rates, we can use the same recursions to get:

$$
\pi^e_{t,n-m} = A^{\pi,e}_{n-m} + B^{\pi,e}_{n-m} X_t
$$

where $A^{\pi,e}_{n-m} = A^e_{n-m} - A^{r,e}_{n-m}$ and $B^{\pi,e}_{n-m} = B^e_{n-m} - B^{r,e}_{n-m}$.

D Monte Carlo Design

A total of 1000 samples were generated for each specification of the data generating process (DGP): 3 and 4 factor models with two different parameter sets each capturing different levels of persistence. The parameters were taken from US $\mathbb{P}$-dynamics estimates (from the 3 and 4 factor full sample estimates with surveys), and the UK $\mathbb{Q}$-dynamics estimates (the most persistent estimates). These are summarized in the table below:

A. Alternative Data Generating Processes					
	MC1	MC2	MC3	MC4	
Number of Factors	3	3	N=4	N=4	
Source estimate	US P-dyn	UK Q-dyn	US P-dyn	UK Q-dyn	
	Empirical Volatilities Estimates				
max vol factors	0.033	0.035	0.033	0.035	
vol yield errors	0.003	0.002	0.002	0.002	
vol survey errors	0.012	0.010	0.013	0.010	
	Persistence				
max Eigenvalue	0.9918	0.9997	0.9914	0.9989	
half-life (years)	7	208	7	51	
B. Design of Cross-Section Information					
	Varying Cross-Section Quantity of Information				
Name	Short	Like Q	Data US	Data UK	Cross Section
Time series length	10	40	40	40	0
Forecast time series length	10	40	30	15	40
Frequency of forecasts (months)	1	1	3	3	1
Forecast Maturities (years)	0.5,1,2,3,4,5,10	0.5,1,2,3,4,5,10	0.5, 1, 10	1,2,3	0.5,1,2,3,4,5,10
	Varying Cross-Section Precision of Information				
Name	Noise	High	Moderate	Surveys	Yields
Multiple of max vol of factors	50	10	5	1	0.1

The parameters to be estimated are μ, Φ. I do not estimate $\delta_0, \delta_1, \Sigma$ because these enter the $\mathbb{Q}$-likelihood and Ω can be chosen by the econometrician rather than estimated, and is not the focus of the experiment. For each model I take $\mu, \Phi, \delta_0, \delta_1, \Sigma$ from empirical estimates and vary Ω.

For each of the 4 DGP, standard OLS estimates for the VAR drift (μ) and autoregressive (Φ) parameters were obtained for samples of length of {20,30,40,50,100,200,300,400,500,700,1000} years with monthly data.

For models with cross-section of interest rate forecasts, forecasts of up to 10 years ahead were generated, with different degrees of volatility in errors added, as shown in panel B of the table. I also consider a pure cross-section estimate (‘Cross-Section’) where only the likelihood of forecast errors are included. A total of 25 models with forecasts were estimated for each of the 4 DGP (5 different variations of availability of time series and forecasts (length and maturities), with 5 different levels of measurement error volatilities, shown in Panel B).

References

[1] Altavilla, C., R. Giacomini & G. Ragusa (2013). "Anchoring the Yield Curve with Survey Expectations", mimeo UCL.

[2] Ang, A., G. Bekaert and M. Wei (2007). "Do Macro Variables, Asset Markets or Surveys Forecast Inflation Better?", *Journal of Monetary Economics*, vol. 50 (4).

[3] Ang, A. and M. Piazzesi (2003). "A No-Arbitrage Vector Regression of Term Structure Dynamics with Macroeconomic and Latent Variables", *Journal of Monetary Economics*, vol. 50 (4).

[4] Bauer, M. (2011). "Bayesian Estimation of Dynamic Term Structure Models under Restrictions on Risk Pricing", *Federal Reserve Bank of San Francisc*o working paper 2011-03.

[5] Bauer, M., G. Rudebusch and J. Wu (2012). "Correcting Estimation Bias in Dynamic Term Structure Models", *Journal of Business and Economic Statistics*, vol. 30 (3).

[6] Bibkov, R. and M. Chernov (2009). "Unspanned Stochastic Volatility in Affine Models: Evidence from Eurodollar Futures and Options", *Management Science*, vol. 55 (8).

[7] Bibkov, R. and M. Chernov (2011). "Yield Curve and Volatility: Lessons from Eurodollar Futures and Options", *Journal of Financial Econometrics*, vol. 9 (1).

[8] Campbell, J. and R. Shiller (1991). "Yield Spreads and Interest Rate Movements: A Bird's Eye View", *Review of Economic Studies*, vol. 58.

[9] Campbell, J., R. Shiller and L. Viceira (2009). "Understanding Inflation-Indexed Bond Markets", *Brookings Papers on Economic Activity*, Spring.

[10] Campbell, J., A. Sundaram and L. Viceira (2010). "Inflation Bets or Deflation Hedges? The Changing Risks of Nominal Bonds", mimeo Harvard University.

[11] Carroll, C. (2003). "Macroeconomic Expectations of Households and Professional Forecasters", *Quarterly Journal of Economics*, vol. 118 (1).

[12] Carvalho, C. and F. Necchio (2012). "Do People Understand Monetary Policy?", *FRBSF Working Paper* #2012-01

[13] Cheridito, P., D. Filipovic and R. Kimmel (2007). "Market Price of Risk Specifications for Affine Models: Theory and Evidence", *Journal of Financial Economics*, vol. 83.

[14] Chernov, M. and P. Mueller (2012). "The Term Structure of Inflation Expectations", *Journal of Financial Economics*, vol. 106

[15] Christensen, J. , F. Diebold, and G. Rudebusch (2011). "The Affine Arbitrage-free Class of Nelson-Siegel Term Structure Models", *Journal of Econometrics,* vol. 164 (1).

[16] Chun, A. (2011). "Expectations, bond yields and monetary policy", *Review of Financial Studies* vol. 24.

[17] Cochrane, J. and M. Piazzesi (2005). “Bond Risk Premia”, *American Economic Review*, vol. 95 (1).

[18] Cochrane, J. and M. Piazzesi (2008). "Decomposing the Yield Curve", mimeo Graduate School of Business, University of Chicago.

[19] Cole, S. and F. Milani (2013). "The Misspecification of Expectations in New Keynesian Models: A DSGE-VAR Approach", mimeo UC Irvine.

[20] Collin-Dufresne, P., B. Goldstein and C. Jones (2009). "Can Interest Rate Volatility be Extracted from the Cross Section of Bond Yields? An Investigation of Unspanned Stochastic Volatility", *Journal of Financial Economics*, vol. 94 (1).

[21] Collin-Dufresne, P., M. Johannes and L. Lochstoer (2013). "Parameter Learning in General Equilibrium: The Asset Pricing Implications", working paper Columbia Business School.

[22] Creal, D. and C. Wu (2013). "Estimation of non-Gaussian Affine Term Structure Models", mimeo Chicago Booth.

[23] Dai, Q. and K. Singleton (2000). “Specification Analysis of Affine Term Structure Models”, *Journal of Finance*, vol. 55.

[24] Dai, Q. and K. Singleton (2002). “Expectation Puzzles, Time-varying Risk Premia, and Affine Models of the Term Structure”, *Journal of Financial Economics*, vol. 63 (3).

[25] Dai, Q. and K. Singleton (2003). “Term Structure Dynamics in Theory and Reality”, *Review of Financial Studies*, vol.16.

[26] D’Amico, S., D. Kim and M. Wei (2010). "Tips from TIPS: the informational content of Treasury Inflation-Protected Security prices", Federal Reserve Board Finance and Economics Discussion Series 2010-19.

[27] Del Negro, M. and F. Schorfheide (2012). "DSGE Model-Based Forecasting", working paper UPenn.

[28] Dempster, M. and K. Tang (2011). "Estimating exponential affine models with correlated measurement errors: Applications to fixed income and commodities", *Journal of Banking and Finance*, vol. 35 (3).

[29] Diebold, F. and C. Li (2006). “Forecasting the Term Structure of Government Bond Yields”, *Journal of Econometrics*, vol 127 ().

[30] Duarte, J. (2004). “Evaluating an Alternative Risk Preference in Affine Term Structure Models”, *Review of Financial Studies*, vol. 17(2).

[31] Duffee, G. (2002). “Term Premia and Interest Rate Forecasts in Affine Models”, *Journal of Finance*, vol.57.

[32] Duffee, G. (2010). "Sharpe ratios in term structure models", working paper Johns Hopkins University.

[33] Duffee, G. (2011a). "Information In (and Not In) the Term Structure", *Review of Financial Studies*, vol.24 (9).

[34] Duffee, G. (2011b). "Forecasting with the Term Structure: The role of no-arbitrage restrictions", working paper Johns Hopkins University.

[35] Duffee, G. (2012a). "Forecasting Interest Rates", prepared for the *Handbook of Economic Forecasting*, Vol. 2.

[36] Duffee, G. (2012b). "Bond Pricing and the Macroeconomy," prepared for the Handbook of the Economics of Finance

[37] Duffee, G. and R. Stanton (2008). "Evidence on Simulation Inference for Near Unit-Root Processes with Implications for Term Structure Estimation," *Journal of Financial Econometrics*, vol. 6 (1).

[38] Duffee, G. and R. Stanton (2012). "Estimation of Dynamic Term Structure Models", *Quarterly Journal of Finance*, vol. 2 (2).

[39] Duffie, D. (2001). *Dynamic Asset Pricing Theory*, Princeton University Press, 3rd Ed.

[40] Easaw, J. and R. Golinelli (2010). "Households Forming Inflation Expectations: Active and Passive Absorption Rates," *The B.E. Journal of Macroeconomics* (Contributions), vol. 10 (1).

[41] Faust, J. and J. Wright (2012). "Forecasting Inflation", prepared for the *Handbook of Forecasting.*

[42] Feldhüter, P. (2008). "Can Affine Models Match the Moments in Bond Yields?", mimeo London Business School.

[43] Fuhrer, J. (2013). "Real Expectations: Replacing Rational Expectations with Survey Expectations in Dynamic Macro Models", working paper Federal Reserve Bank of Boston.

[44] Guimarães, R. (2012). "What accounts for the fall in UK ten-year government bond yields?", *Bank of England Quarterly Bulletin*, vol. 52 (3).

[45] Gurkaynak, R., B. Sack & J. Wright (2007). "The U.S. Treasury yield curve: 1961 to the present", *Journal of Monetary Economics*, vol. 54.

[46] Heidari, M. and L. Wu (2009). "A Joint Framework for Consistently Pricing Interest Rates and Interest Rate Derivatives", *Journal of Financial and Quantitative Analysis*, vol. 44 (3).

[47] Hess, D. and S. Orbe (2013). "Irrationality or Efficiency of Macroeconomic Survey Forecasts? Implications from the Anchoring Bias Test", *Review of Finance*, vol. 17 (6).

[48] Jardet, C., A. Monfort and F. Pegoraro (2013). "No-arbitrage Near-Cointegrated VAR(p) term structure models, term premia and GDP growth", *Journal of Banking and Finance*, vol. 37.

[49] Joslin, S., A. Le and K. Singleton (2011). "Why Gaussian Macro-Finance Term Structure Models Are (Nearly) Unconstrained Factor-VARs", *Journal of Financial Economics*, vol. 109.

[50] Joslin, S., M. Priebisch and K. Singleton (2012). "Risk Premiums in Dynamic Term Structure Models with Unspanned Macro Risks", *Journal of Finance*, forthcoming.

[51] Joslin, S., K. Singleton and H. Zhu (2011). "A New Perspective on Gaussian Dynamic Term Structure Models", *Review of Financial Studies*, vol.24 (3).

[52] Joyce, M., P. Lildholdt & S. Sorensen (2010). "Extracting inflation expectations and inflation risk premia from the term structure: A joint model of the UK nominal and real yield curves", *Journal of Banking and Finance*, vol. 34

[53] Kim, D. (2007). "Challenges in macro-finance modeling", BIS working paper 240.

[54] Kim, D. and A. Orphanides (2012). "Term structure estimation with survey data on interest rate forecasts", *Journal of Financial and Quantitative Analysis*, vol. 47 (1).

[55] Kim, D. and A. Orphanides (2007). "The bond market term premium: what is it, and how can we measure it?", BIS Quarterly Review, June.

[56] Kim, D. and J. Wright (2005). "An arbitrage-free three-factor term structure model and the recent behavior of long-term yields and distant-horizon forward rates", Finance and Economics Discussion Series 2005-33, Federal Reserve Board.

[57] Le, A. and K. Singleton (2013). "A Robust Analysis of the Risk-Structure of Equilibrium Term Structures of Bond Yields", working paper Stanford.

[58] Litterman, R. and J. Scheinkman (1991). "Common Factors Affecting Bond Returns", *Journal of Fixed Income,* June.

[59] Milani, F. & A. Rajbhandari (2012). "Expectation formation and monetary DSGE models: Beyond rational expectations paradigm", working paper Univ. California Irvine.

[60] Piazzesi, M. (2010). “Affine Term Structure Models”, in *Handbook of Financial Econometrics*, Eds. Yacine Ait-Sahalia and Lars Peter Hansen.

[61] Rudebusch, G., B. Sack and E. Swanson (2007). "Macroeconomic Implications of Changes in the Term Premium", *Federal Reserve Bank of St. Louis Review*, vol. 89 (4).

[62] Smith, T (2012), ‘Option-implied probability distributions for future inflation’, Bank of England Quarterly Bulletin, Vol. 52, No. 3.

[63] Wright, J. (2011). "Term Premia and Inflation Uncertainty: Empirical Evidence from an International Panel Dataset", *American Economic Review*, vol. 101 (4).

[64] Wright, J. (2013). "Term Premia and Inflation Uncertainty: Empirical Evidence from an International Panel Dataset: Reply", *American Economic Review*, forthcoming.

Table 1: Descriptive statistics of yield curve data - US and UK

	A. US Nominal Yields							
	3m	1y	2y	3y	4y	5y	7y	10y
Average	5.7	6.0	6.3	6.5	6.7	6.8	7.0	7.3
Std. Dev.	3.2	3.2	3.1	3.0	2.9	2.8	2.7	2.5
Autocorrelation	0.981	0.982	0.984	0.985	0.985	0.986	0.987	0.987
Observations	467	467	467	467	467	467	467	467
	B. UK Nominal Yields							
	6m	1y	2y	3y	4y	5y	7y	10y
Average	7.4	7.8	8.0	8.1	8.2	8.3	8.5	8.6
Std. Dev.	3.6	3.4	3.3	3.3	3.2	3.2	3.3	3.3
Autocorrelation		0.982	0.983	0.984	0.985	0.986	0.988	0.989
Observations	396	465	467	467	467	467	467	467
	C. UK Real Yields							
		1y	2y	3y	4y	5y	7y	10y
Average		2.2	2.6	2.6	2.7	2.7	2.7	2.8
Std. Dev.		1.7	1.2	1.2	1.1	1.0	1.0	1.1
Autocorrelation						0.950	0.963	0.975
Observations		48	191	292	310	311	311	311

Notes: The table shows the average, standard deviation, autocorrelation and number of observations for each of the yields used in this paper. Note that there are missing observations for UK nominal and real yields for the short maturities. All figures are expressed in annualized percentage points.

Table 2: P-values of C-vM Global Discrepancy Measure of Q-dynamics

	US											
	Unrestricted			With Surveys			Sharpe ratio 0.3			Sharpe ratio 0.5		
	3 factors	4 factors	5 factors	3 factors	4 factors	5 factors	3 factors	4 factors	5 factors	3 factors	4 factors	5 factors
1y	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
5y	99%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
10y	96%	100%	100%	87%	100%	100%	95%	100%	100%	93%	100%	100%
15y	4%	87%	99%	3%	93%	94%	2%	13%	45%	3%	97%	63%
20y	0%	33%	78%	0%	40%	45%	0%	0%	1%	0%	39%	8%
30y	0%	1%	12%	0%	1%	1%	0%	0%	0%	0%	1%	0%
50y	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%

	UK											
	Unrestricted			With Surveys			Sharpe ratio 0.3			Sharpe ratio 0.5		
	3 factors	4 factors	5 factors	3 factors	4 factors	5 factors	3 factors	4 factors	5 factors	3 factors	4 factors	5 factors
1y	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
5y	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
10y	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
15y	35%	41%	41%	40%	40%	41%	43%	41%	27%	35%	39%	43%
20y	2%	0%	1%	4%	0%	1%	3%	0%	0%	3%	0%	0%
30y	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
50y	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%

Notes: The table shows the median of the p-values from the Cramér-von Mises test for common $\mathbb{Q}$-dynamics. For each estimation strategy and number of factors, the p-values shown are the median p-values for the null hypothesis of same distribution from all 28 (8x7/2) pairwise comparisons for the different estimation samples. This is shown for several different maturities. Though models are estimated for different subsamples of the data, I use the full sample model predictions to compute the statistics. The top panel shows the results for the US and the bottom panel shows the same results for the UK. The median p-values for longer maturities are zero.

Table 3: P-values of C-vM Global Discrepancy Measure of P-dynamics

	Unrestricted			With Surveys			Sharpe ratio 0.3			Sharpe ratio 0.5		
US												
	3 factors	4 factors	5 factors	3 factors	4 factors	5 factors	3 factors	4 factors	5 factors	3 factors	4 factors	5 factors
1y	18%	36%	0%	99%	99%	37%	41%	75%	87%	61%	60%	58%
2y	1%	2%	0%	99%	92%	20%	7%	64%	48%	9%	36%	20%
3y	0%	0%	0%	97%	43%	13%	1%	22%	11%	1%	5%	2%
4y	0%	0%	0%	80%	24%	8%	0%	4%	2%	0%	1%	0%
5y	0%	0%	0%	56%	17%	5%	0%	1%	0%	0%	0%	0%
6y	0%	0%	0%	43%	8%	3%	0%	0%	0%	0%	0%	0%
7y	0%	0%	0%	28%	5%	1%	0%	0%	0%	0%	0%	0%
UK												
	3 factors	4 factors	5 factors	3 factors	4 factors	5 factors	3 factors	4 factors	5 factors	3 factors	4 factors	5 factors
1y	4%	10%	56%	86%	47%	75%	19%	83%	81%	4%	27%	78%
2y	0%	0%	1%	49%	45%	58%	4%	33%	79%	0%	0%	17%
3y	0%	0%	0%	35%	45%	44%	5%	32%	36%	0%	0%	2%
4y	1%	0%	0%	39%	35%	15%	3%	9%	4%	1%	0%	1%
5y	1%	0%	0%	39%	34%	4%	1%	0%	1%	1%	0%	1%
6y	0%	0%	0%	22%	13%	1%	0%	0%	0%	0%	0%	0%
7y	0%	0%	0%	2%	3%	2%	0%	0%	0%	0%	0%	0%

Notes: The table shows the median of the p-values from the Cramér-von Mises test for common $\mathbb{P}$-dynamics. For each estimation strategy and number of factors, the p-values shown are the median p-values for the null hypothesis of same distribution from the 15 (6x5/2) pairwise comparisons for the different estimation samples excluding the 2002 and 2007 estimation samples. The online appendix has the pairwise correlations with these two additional samples. This is shown for several different maturities. Though models are estimated for different subsamples of the data, I use the full sample model predictions to compute the statistics. The top panel shows the results for the US and the bottom panel shows the same results for the UK.

Table 4: P-values of C-vM Global Discrepancy Measure of Q-dynamics for Alternative Cross-section of Bonds

	Unrestricted			Without Intermediate			Without Long		
	3 factors	4 factors	5 factors	3 factors	4 factors	5 factors	3 factors	4 factors	5 factors
1y	100%	100%	100%	100%	100%	100%	100%	100%	100%
5y	99%	100%	100%	99%	100%	100%	99%	100%	100%
10y	96%	100%	100%	97%	100%	100%	1%	10%	0%
15y	4%	87%	99%	4%	37%	0%	0%	0%	0%
20y	0%	33%	78%	0%	2%	0%	0%	0%	0%
30y	0%	1%	12%	0%	0%	0%	0%	0%	0%
50y	0%	0%	0%	0%	0%	0%	0%	0%	0%

Notes: The table shows the median of the p-values from the Cramér-von Mises test for common $\mathbb{Q}$-dynamics for US with different bond maturities used in estimation. The "unrestricted" model uses all maturities as described in the data section. The "without intermediate" model removes the bonds with maturities of 4 and 7 years from estimation, and model "without long" removes the bonds with 7 and 10 years. For each estimation strategy and number of factors, the p-values shown are the median p-values for the null hypothesis of same distribution from all 28 (8x7/2) pairwise comparisons for the different estimation samples. This is shown for several different maturities. Though models are estimated for different subsamples of the data, I use the full sample model predictions to compute the statistics. The median p-values for longer maturities are zero.

Table 5: Real yield curve fit with reduced real bond yield data in estimation

	Mean pricing errors					
	3 Factors		4 Factors		5 Factors	
	baseline	only 10yr	baseline	only 10yr	baseline	only 10yr
1y	24.9	81.4	24.0	99.4	6.9	137.0
2y	19.3	64.3	20.4	68.2	4.9	90.7
3y	8.4	50.3	8.7	56.3	3.6	62.8
4y	7.8	43.1	7.3	46.3	3.3	45.3
5y	10.6	38.5	10.0	40.3	3.3	36.9
7y	16.0	33.6	15.5	34.6	2.1	29.8
10y	21.4	31.5	21.4	32.6	6.0	27.3
	Max pricing errors					
	3 Factors		4 Factors		5 Factors	
	baseline	only 10yr	baseline	only 10yr	baseline	only 10yr
1y	90.1	262.3	70.7	324.4	16.3	339.0
2y	66.7	319.5	71.4	227.5	20.3	529.3
3y	54.6	424.9	43.9	403.3	24.3	440.6
4y	56.7	358.1	36.8	340.6	14.9	349.7
5y	57.0	303.5	45.6	290.4	13.0	269.1
7y	66.4	217.2	53.2	212.1	11.0	169.1
10y	89.8	122.3	87.7	143.4	22.2	124.8

Notes: The table shows the mean (top panel) and maximum (bottom panel) absolute fitting errors for each maturity of the real yield curve and number of factors for two different estimates. The column labeled "baseline" show the real yield curve fit for the joint model estimated with the full sample of real yield data from 1972 and surveys described in Section 3.1. The column labeled "only 10yr" shows the fit for the same model estimated with the same data for nominal yields and surveys as the "baseline" models, but only the 10 year maturity real yield using as starting parameter values the estimates using all of the real yields (parameter values for column "baseline" are used as starting values for estimation of model shown in "only 10yr"). All figures are expressed in annualized basis points.

Table 6: Median P-values of Cramer-von Mises Tests for Alternative Time Series and Cross-Section Monte Carlo Estimates

	A. Estimates from Time Series								
	40	50	70	100	200	300	500	700	1000
Horizon									
0.5	3%	7%	38%	75%	89%	95%	97%	99%	99%
1	0%	0%	7%	34%	55%	70%	79%	89%	96%
2	0%	0%	0%	4%	12%	23%	32%	48%	64%
3	0%	0%	0%	0%	2%	5%	9%	19%	33%
5	0%	0%	0%	0%	0%	0%	1%	2%	5%
7	0%	0%	0%	0%	0%	0%	0%	0%	1%
10	0%	0%	0%	0%	0%	0%	0%	0%	0%
15	0%	0%	0%	0%	0%	0%	0%	0%	0%
20	0%	0%	0%	0%	0%	0%	0%	0%	0%
30	0%	0%	0%	0%	0%	0%	0%	0%	0%
50	0%	0%	0%	0%	0%	0%	0%	0%	0%

	B. Estimates with Cross-Section Forecasts								
	Like Q			Data US			Data UK		
	Noise	Survey	Yields	Noise	Survey	Yields	Noise	Survey	Yields
Horizon									
0.5	69%	100%	100%	25%	95%	100%	19%	73%	99%
1	30%	99%	100%	2%	74%	99%	1%	34%	98%
2	3%	99%	100%	0%	37%	91%	0%	4%	95%
3	0%	99%	100%	0%	20%	78%	0%	0%	59%
5	0%	99%	100%	0%	4%	69%	0%	0%	2%
7	0%	96%	100%	0%	1%	80%	0%	0%	0%
10	0%	29%	100%	0%	0%	82%	0%	0%	0%
15	0%	0%	72%	0%	0%	3%	0%	0%	0%
20	0%	0%	12%	0%	0%	0%	0%	0%	0%
30	0%	0%	0%	0%	0%	0%	0%	0%	0%
50	0%	0%	0%	0%	0%	0%	0%	0%	0%

Notes: The table shows the median of the p-values from the pairwise Cramér-von Mises test for common estimated dynamics from the Monte Carlo experiment (model 'MC1', described in page 32). The test is applied to all pairwise combinations of forecasts from the 1000 estimates. The forecasts are generated using actual estimated factors from US data (the same from which the true parameters were taken) for different forecast horizons (expressed in years) along the rows.

Table 7: Monte Carlo Percentiles of the Bias for Estimates of Unconditional Mean and Half-life

	A. Estimates from Time Series								
	40	50	100	200	300	400	500	700	1000
Percentiles									
	Unconditional Mean in percentage points (DGP = 4.96)								
0.01	-5.2	-3.8	-2.8	-2.2	-1.7	-1.3	-1.2	-1.0	-0.8
0.05	-3.1	-2.8	-2.0	-1.5	-1.1	-0.9	-0.8	-0.7	-0.6
0.10	-2.3	-2.2	-1.6	-1.1	-0.9	-0.7	-0.6	-0.5	-0.4
0.50	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.90	2.5	2.2	1.7	1.1	0.9	0.7	0.7	0.6	0.5
0.95	3.3	2.8	2.0	1.4	1.2	1.0	0.8	0.7	0.6
0.99	5.0	4.1	2.9	2.1	1.8	1.4	1.1	1.0	0.8
	Half-life in years (DGP=7)								
0.01	-6.0	-5.9	-5.1	-4.2	-3.6	-3.3	-3.0	-2.6	-2.2
0.05	-5.7	-5.6	-4.6	-3.4	-2.9	-2.7	-2.5	-2.1	-1.7
0.10	-5.4	-5.2	-4.2	-3.1	-2.6	-2.3	-2.1	-1.7	-1.4
0.50	-3.6	-3.2	-2.0	-1.2	-0.7	-0.6	-0.5	-0.3	-0.2
0.90	0.8	1.5	2.1	2.0	1.8	1.6	1.5	1.3	1.2
0.95	4.1	4.4	3.8	3.0	2.8	2.3	2.2	2.1	1.7
0.99	23.0	12.4	7.3	5.2	4.3	3.9	4.0	3.4	2.7

	B. Estimates with Cross-Section Forecasts								
	Like Q			Data US			Data UK		
	Noise	Survey	Yields	Noise	Survey	Yields	Noise	Survey	Yields
Percentiles									
	Unconditional Mean in percentage points (DGP = 4.96)								
0.01	-1.9	-0.5	-0.1	-3.7	-0.9	-0.4	-5.7	-2.3	-1.8
0.05	-1.2	-0.3	-0.1	-2.5	-0.6	-0.3	-3.0	-1.6	-1.1
0.10	-0.9	-0.2	-0.1	-2.0	-0.4	-0.2	-2.3	-1.2	-0.8
0.50	0.0	0.0	0.0	0.1	0.0	0.0	0.1	0.1	0.0
0.90	1.0	0.2	0.1	2.1	0.4	0.2	2.4	1.3	0.9
0.95	1.3	0.3	0.1	2.7	0.6	0.2	3.1	1.6	1.1
0.99	1.8	0.4	0.1	3.5	0.9	0.4	4.3	2.5	2.0
	Half-life in years (DGP=7)								
0.01	-5.5	-2.3	-0.3	-5.6	-4.7	-2.3	-5.6	-5.2	-4.6
0.05	-4.7	-1.4	-0.1	-5.0	-3.9	-1.4	-5.1	-4.5	-3.8
0.10	-4.2	-1.1	-0.1	-4.5	-3.4	-0.9	-4.5	-4.1	-3.4
0.50	-2.1	0.0	0.0	-2.5	-1.2	0.0	-2.5	-2.2	-1.3
0.90	1.0	0.8	0.1	0.9	1.0	0.8	1.2	1.5	2.4
0.95	2.3	1.5	0.1	3.3	1.9	1.4	4.8	3.2	4.3
0.99	5.4	3.4	0.2	12.4	3.8	3.0	25.2	9.2	11.4

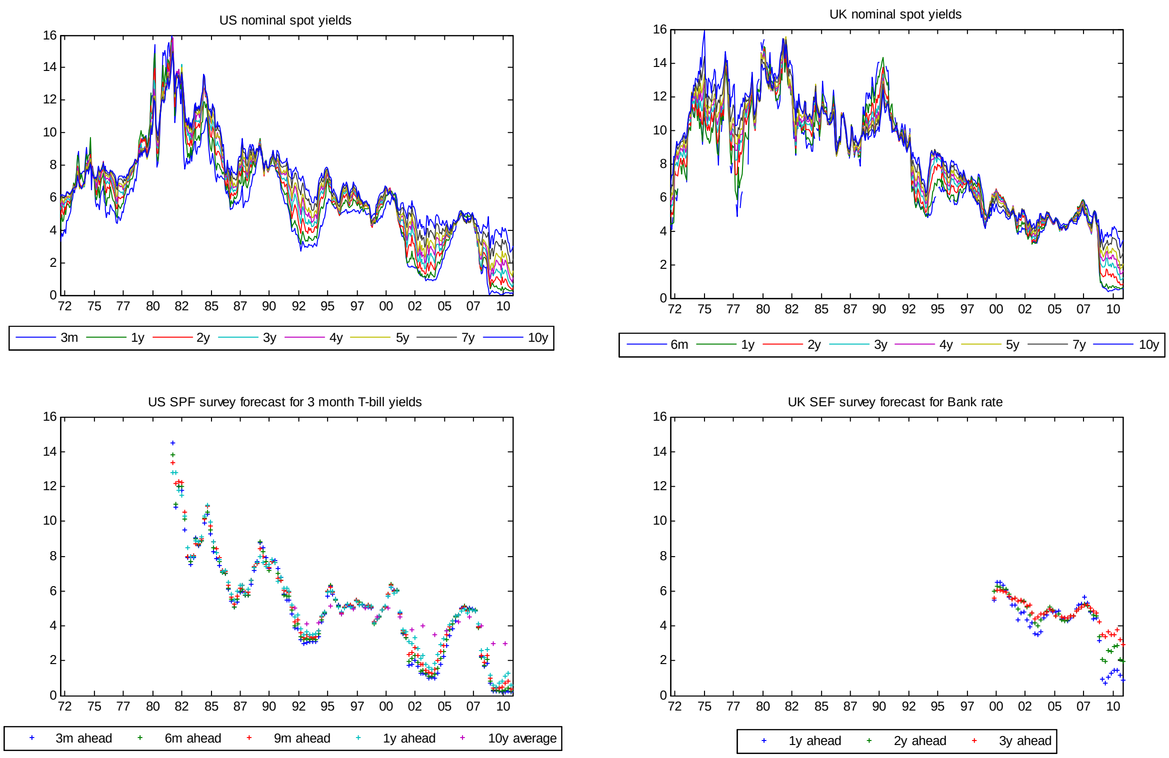
Notes: The table shows the percentiles of the bias in estimated unconditional mean of intereste rates $(\delta_0 + \delta_1 (I_N - \Phi)^{-1} \mu)$ and half-life of the largest eigenvalue of Φ $(\frac{1}{12}\frac{\ln(0.5)}{\ln(\max(eig(\Phi)))})$ from the Monte Carlo experiment (model 'MC1', described in page 32).

Table 8: Sensitivity of term premia estimates to the inclusion of surveys in filtering

Percentile	Estimation sample (starting year))							
	1972	1977	1982	1987	1992	1997	2002	2007
	A. US Nominal Yields							
	3 factors							
0.5	0.0	0.0	0.0	0.1	0.1	0.1	0.1	1.1
0.9	0.8	0.9	0.8	0.7	0.5	0.5	1.4	6.0
	4 factors							
0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.5
0.9	0.2	0.3	0.2	0.3	0.1	0.4	0.4	5.4
	5 factors							
0.5	13.7	7.1	9.8	7.7	6.3	7.0	5.6	0.1
0.9	49.5	20.7	26.0	28.6	22.2	25.6	25.5	1.7
	B. UK Nominal Yields							
	3 factors							
0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.9	0.0	0.1	0.0	0.1	0.1	0.1	0.1	0.2
	4 factors							
0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.9	2.5	3.1	3.8	0.6	1.1	0.4	0.0	0.1
	5 factors							
0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1
0.9	1.7	2.3	2.2	0.7	0.4	0.3	0.3	0.7

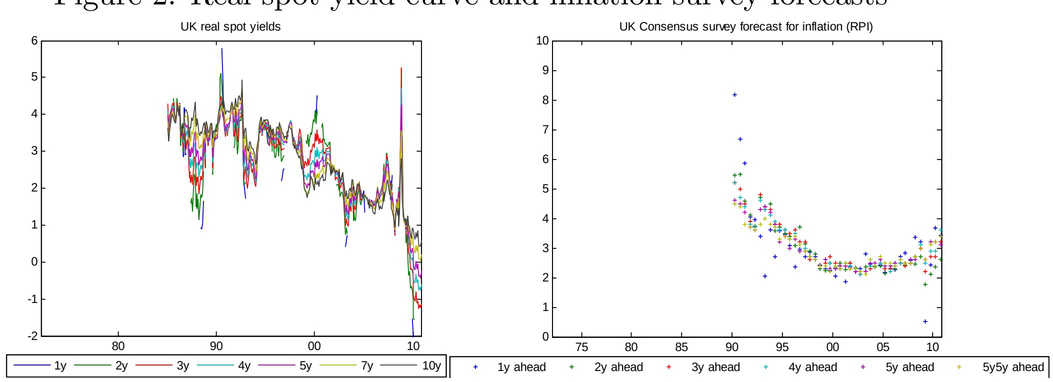
Notes: The table shows the percentiles of the absolute differences between the estimates of spot term premia for the models estimated using survey forecasts when the surveys are not used in filtering the states. For each number of factors (blocks of rows), and each sample estimation period (columns), the percentiles of the absolute difference between the term premia estimates for maturities from 10 years to 20 years, with and without surveys used in filtering, are shown along the rows for each block. All figures are expressed in annualized basis points.

Figure 1: Nominal spot yield curves and survey forecasts



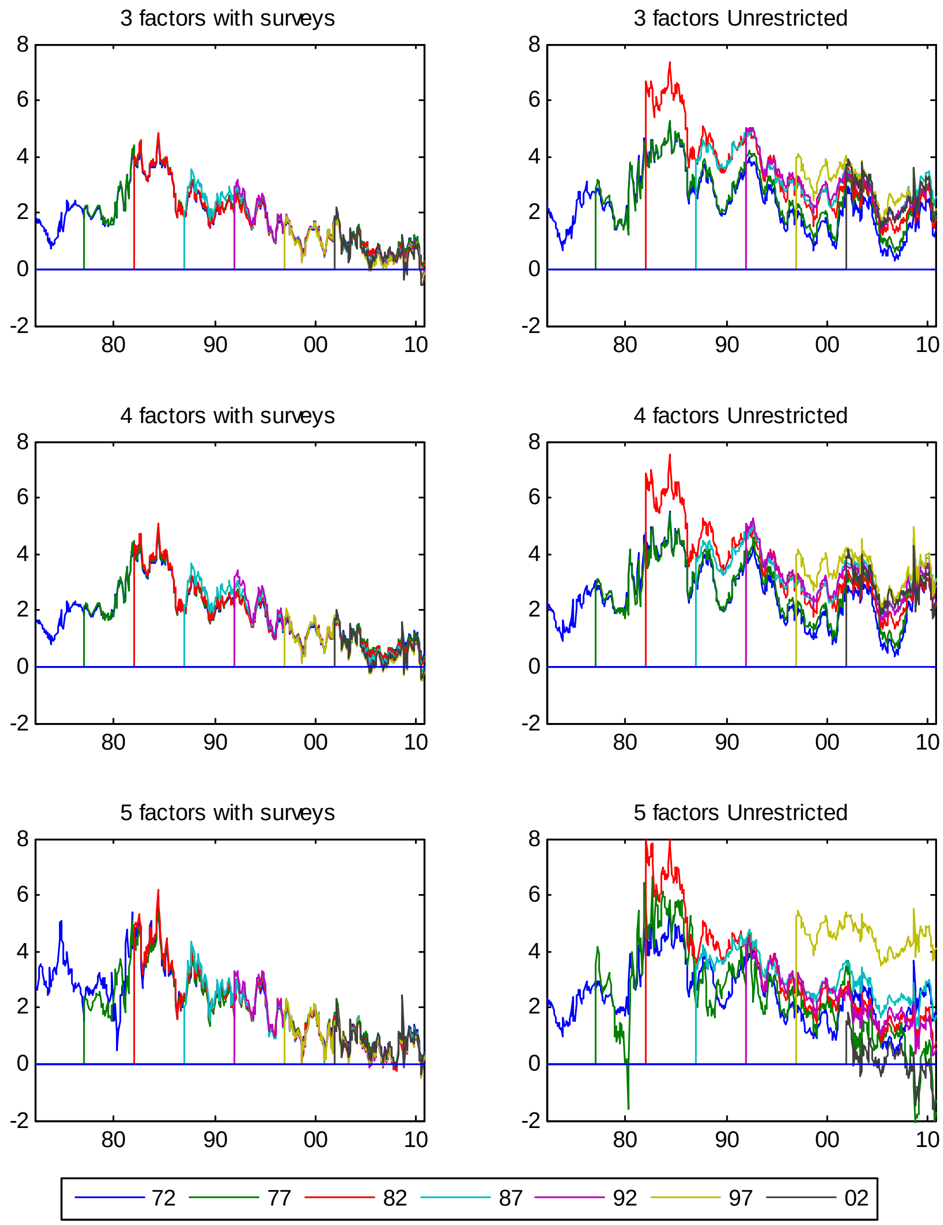
Notes: The figure shows the term structure of nominal spot yields on government bonds and survey forecasts of nominal interest rates used in this paper. The left column shows the data for the US and right column the data for the UK. All figures are in per cent per annum. For the US the forecasts are for 3 month T-bill yields from the Survey of Professional Forecasters of the Philadelphia Federal Reserve Bank. For the UK the forecasts are for Bank Rate from the Survey of External Forecasters of the Bank of England.

Figure 2: Real spot yield curve and inflation survey forecasts



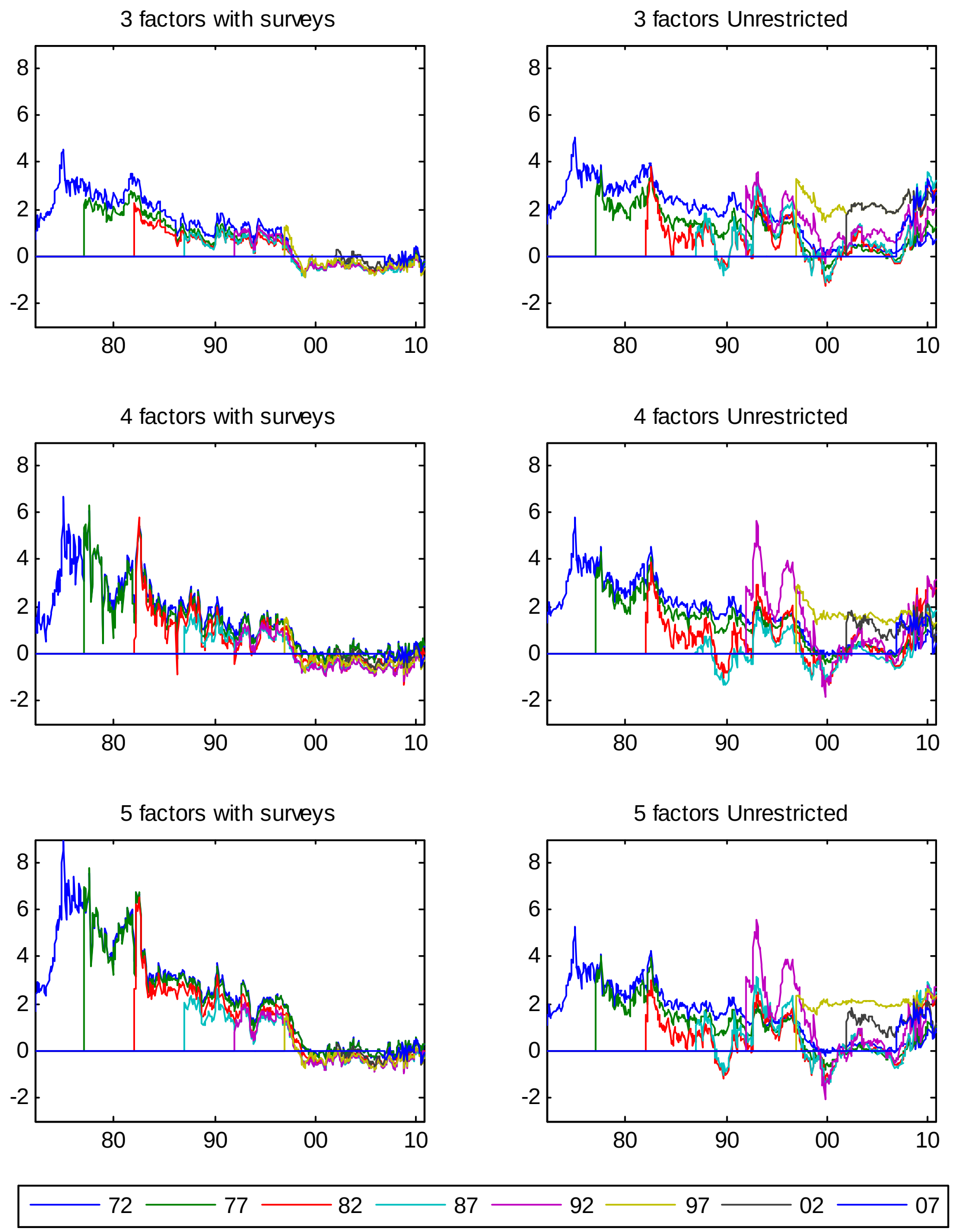
Notes: The figure shows the term structure of real spot yields on government bonds and survey forecasts of inflation rates used in this paper. All figures are in per cent per annum. The survey forecasts of RPI inflation are from Consensus Forecasts for 1, 2, 3, 4, 5 years ahead and the average between 6 and 10 years ahead.

Figure 3: US 10 year nominal term premium with and without surveys



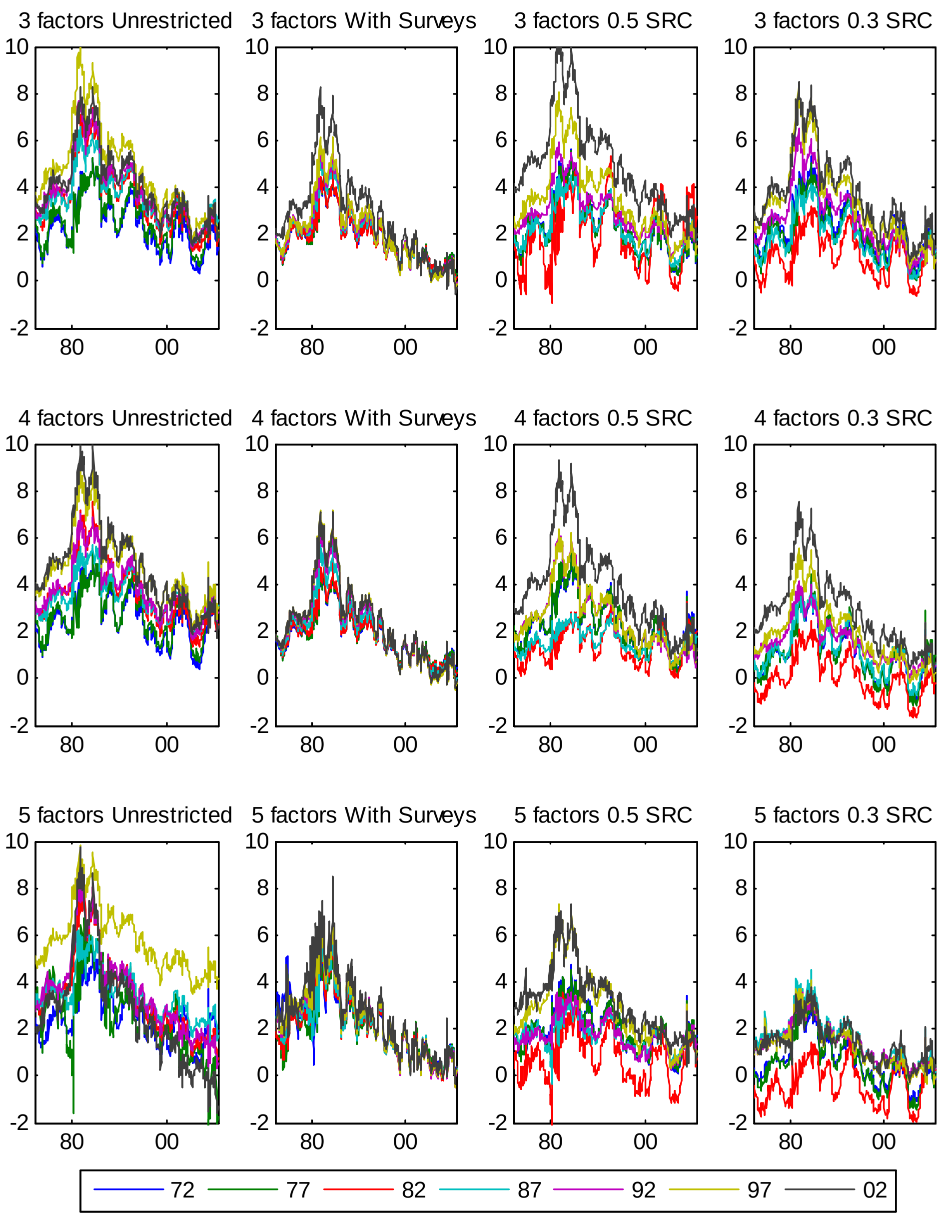
Notes: The figure shows the 10 year spot term premium estimates for the US nominal government bond yields for a total of 45 estimated models. All figures are in percentage points per annum. The models vary by sample, with 7 different samples shown in each chart. The samples vary by start date, starting every 5 years from 1972 to 2002, with all samples ending in Dec 2008. The 3 models varying by number of factors (3 to 5) are displayed along the rows. The models using surveys are displayed in the left column and the unrestricted models in the right column.

Figure 4: UK 10 year nominal term premium with and without surveys



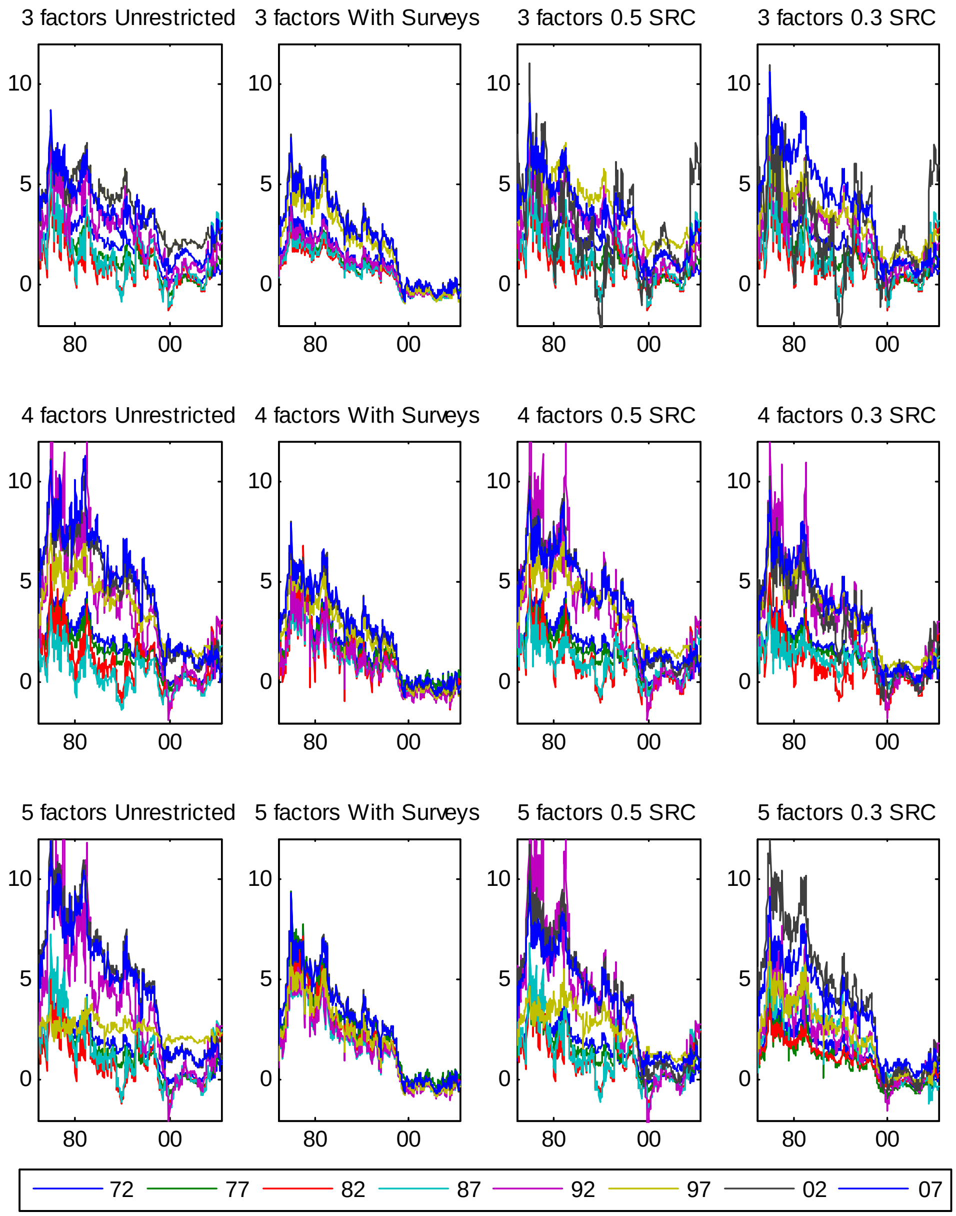
Notes: The figure shows the 10 year spot term premium estimates for the UK nominal government bond yields for a total of 48 estimated models. All figures are in percentage points per annum. The models vary by sample, with 8 different samples shown in each chart. The samples vary by start date, starting every 5 years from 1972 to 2007, with all samples ending in Dec 2010. The 3 models varying by number of factors (3 to 5) are displayed along the rows. The models using surveys are displayed in the left column and the unrestricted models in the right column. The forecasts for 1, 2 and 3 years ahead Bank Rate from the Bank of England's Survey of External Forecasters were used for estimation of the models with surveys.

Figure 5: Full sample comparison of 10 year nominal spot term premia estimates for US



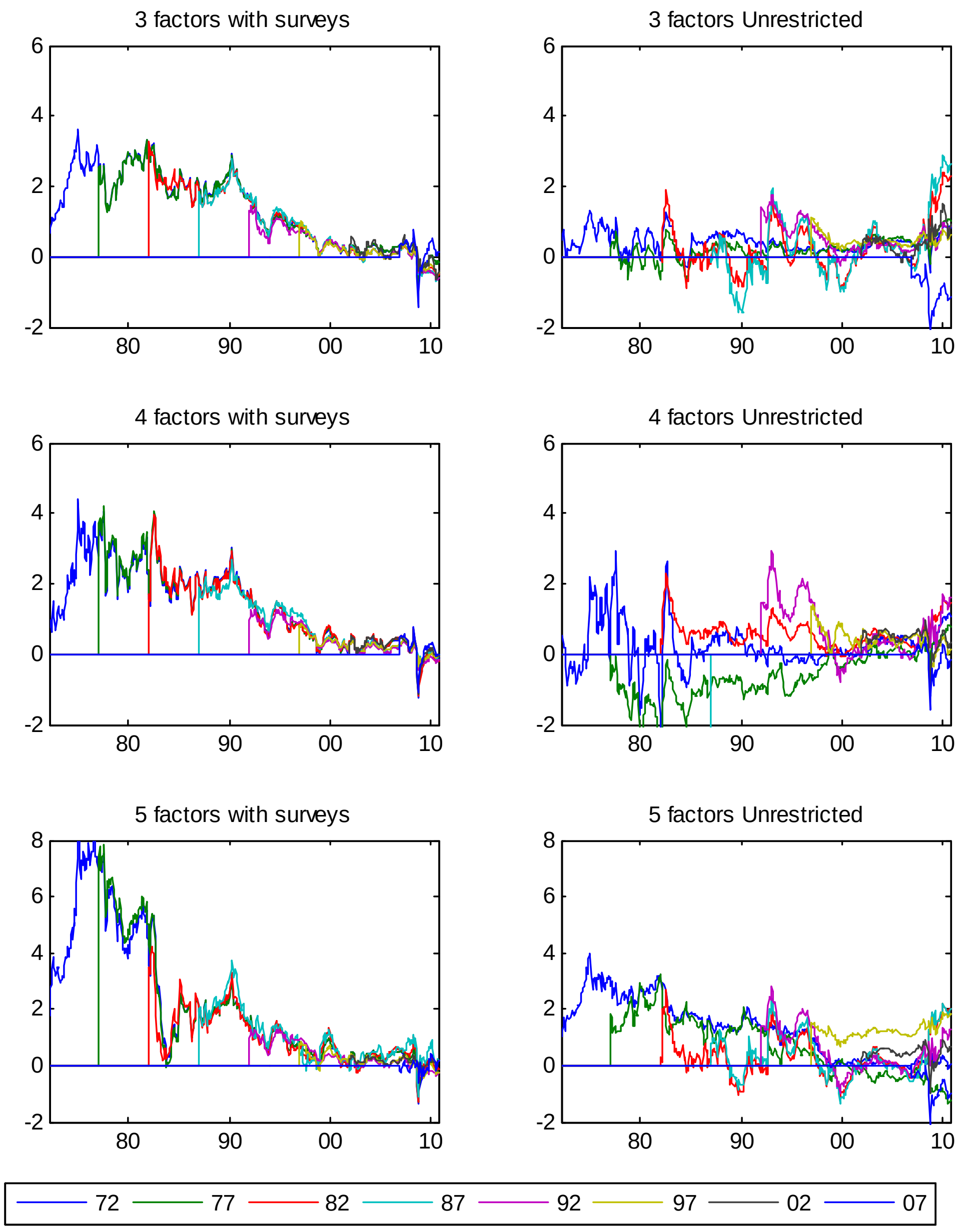
Notes: The figure shows the different estimates of the 10 year spot term premium for US nominal government bond yields filtered for the entire sample using the parameters estimated for the different subsamples. Each chart shows the estimates for the 7 different estimation samples (with starting dates 1972:5:2002). Each row shows models with the same number of factors (3 to 5), while each columns is a different strategy (unrestricted, using surveys and with a 0.5 and 0.3 average constraints on the maximum Sharpe ratio).

Figure 6: Full sample comparison of 10 year nominal spot term premia estimates for UK



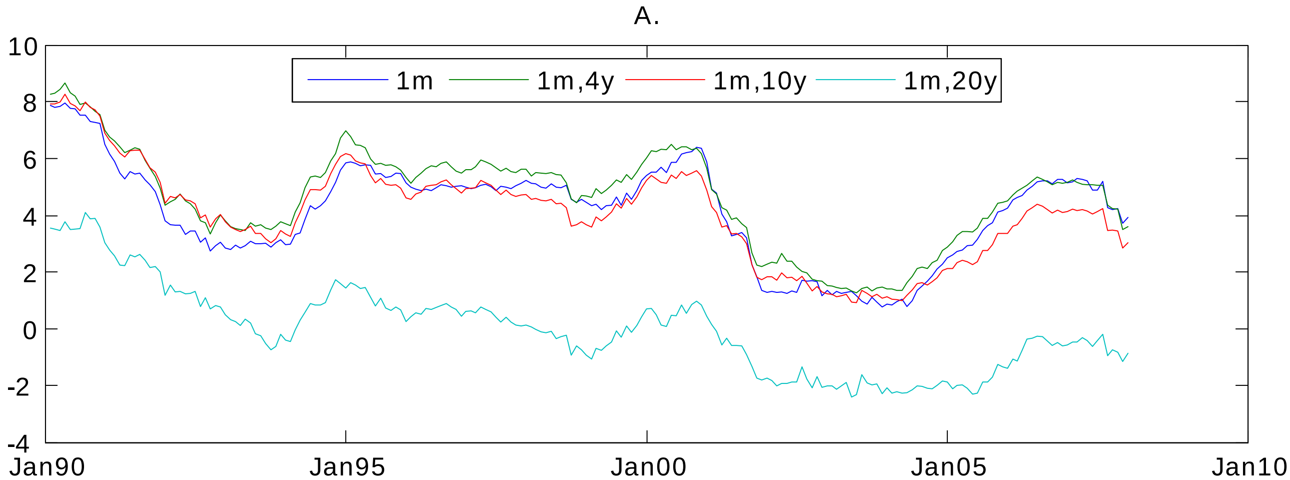
Notes: The figure shows the different estimates of the 10 year spot term premium for US nominal government bond yields filtered for the entire sample using the parameters estimated for the different subsamples. Each chart shows the estimates for the 8 different estimation samples (with starting dates 1972:5:2007). Each row shows models with the same number of factors (3 to 5), while each columns is a different strategy (unrestricted, using surveys and with a 0.5 and 0.3 average constraints on the maximum Sharpe ratio).

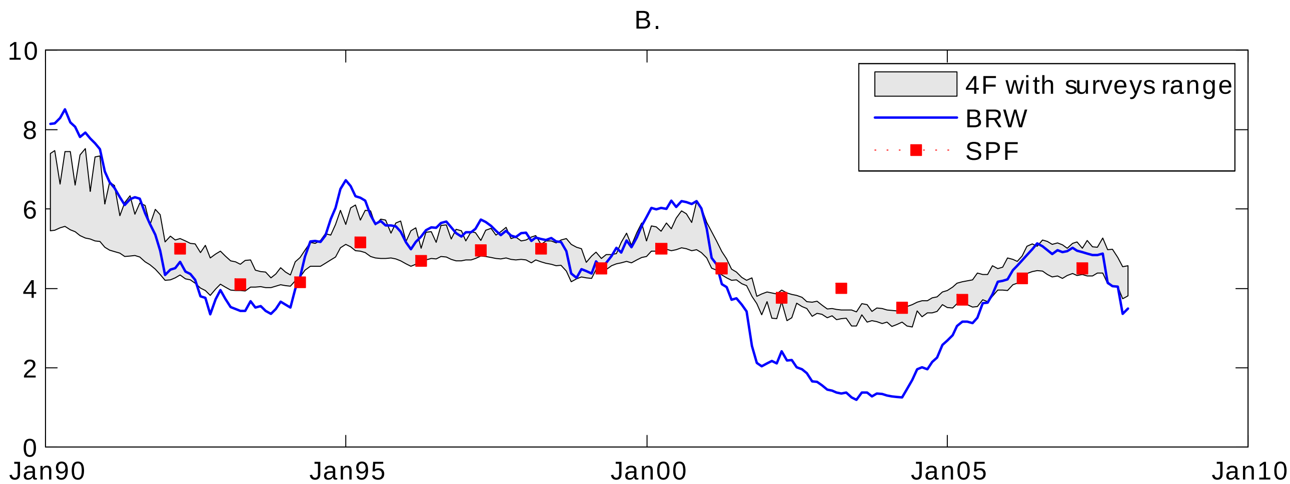
Figure 7: UK 10 year inflation risk premium with and without surveys



Notes: The figure shows the 10 year spot inflation risk premium estimates for the UK government bond yields for a total of 48 estimated models. All figures are in percentage points per annum. The models vary by sample, with 8 different samples shown in each chart. The samples vary by start date, starting every 5 years from 1972 to 2007, with all samples ending in Dec 2010. The 3 models varying by number of factors (3 to 5) are displayed along the rows. The models using surveys are displayed in the left column and the unrestricted models in the right column. The forecasts for 1, 2 and 3 years ahead Bank Rate from the Bank of England's Survey of External Forecasters and the forecasts for inflation from Consensus Forecasts for 1, 2, 3, 4, 5 years ahead and the average between 6 and 10 years ahead were used for estimation of the models with surveys.

Figure 8: Comparisson of 'Bias Corrected' Interest Rate Forecasts and Surveys





Notes: The top panel (A) of the figure shows the implied expected 1 month rates using the data and 'Bias Corrected' parameter estimates from Bauer, Rudebusch & Wu (2012, BRW), for selected forward horizons. Calculations are my own. The bottom panel (B) show the expected average range over a horizon of 10 years from BRW 'Bias Corrected' estimates, the Survey of Professional Forecasters (SPF) and the range of estimates from the 4 factor models estimated with surveys from this paper. All figures are in percentage points per annum. The sample period is Jan 1990 through Dec 2007.