DEUTSCHE BUNDESBANK
EUROSYSTEM

# Discussion Paper

Deutsche Bundesbank
No 27/2021

## Decomposing the yield curve with linear regressions and survey information

Arne Halberstadt

Deutsche Bundesbank, Wilhelm-Epstein-Straße 14, 60431 Frankfurt am Main,
Postfach 10 06 02, 60006 Frankfurt am Main

Tel +49 69 9566-0

Please address all orders in writing to: Deutsche Bundesbank,
Press and Public Relations Division, at the above address or via fax +49 69 9566-3077

Internet http://www.bundesbank.de



ISBN 978–3–95729–835–5
ISSN 2749–2958

# Non-technical summary

## Research Question

Observed long-term bond yields can be decomposed into two hypothetical components: Average expected future short-term interest rates and premiums for the risk that short-term interest rates do not develop as expected. However, it is empirically difficult to decompose yields into these two components. In particular, it is challenging to identify the volatility of the two components well. The problem intensifies if there is only a short history of interest rate data. This paper discusses how one can obtain a robust decomposition of bond yields from a short sample of data.

## Contribution

Previous work approached the problem by imposing the constraint that model-implied interest rate expectations should be close to interest rate expectations known from surveys. I propose to restrict only the volatility of model-implied expectations to that of survey expectations. Unlike other models using survey information, other characteristics of model-implied interest rate expectations remain unrestricted. The method provides an alternative way of reconciling model-implied interest rate expectations with survey information.

## Results

The method provides a tool to analyze interest data with a short data history. It allows plausible yield decompositions in such a case. Without the restriction method, the model would imply that average expected future short-term interest rates have hardly changed over the last few years. However, my application also indicates that there is actually no need to correct the yield decomposition if it is estimated from a long sample of data. More generally, the paper thus contributes to the understanding of reasons for diverging interest rate expectations in term structure models and surveys.

# Nichttechnische Zusammenfassung

## Fragestellung

Die beobachteten langfristigen Anleiherenditen lassen sich in zwei hypothetische Komponenten zerlegen - in die durchschnittlich erwarteten kurzfristigen Zinsen und in die Prämien für das Risiko, dass sich die kurzfristigen Zinsen entgegen den Erwartungen entwickeln. Aus empirischer Sicht stellt diese Zerlegung allerdings eine große Herausforderung dar. So ist es schwierig, die Volatilität der beiden Komponenten genau zu ermitteln. Dieses Problem verschärft sich noch, wenn nur ein geringer Bestand an historischen Daten vorliegt. In diesem Beitrag wird erörtert, wie sich auch mit einer kurzen Datenreihe eine robuste Zerlegung von Anleiherenditen erreichen lässt.

## Beitrag

Die bisherigen Forschungsarbeiten näherten sich dem Problem mithilfe der Restriktion, dass die modellimpliziten Zinserwartungen nahezu den in Umfragen ermittelten Zinserwartungen entsprechen sollen. In diesem Beitrag wird nun vorgeschlagen, die Restriktion lediglich für die Volatilität der modellimpliziten und umfragebasierten Erwartungen vorzunehmen. Anders als in anderen Modellen, die Informationen aus Umfragen verwenden, bleiben die übrigen Eigenschaften der modellimpliziten Zinserwartungen unrestringiert. Mithilfe dieser methodischen Alternative lassen sich modellimplizite und umfragebasierte Zinserwartungen miteinander in Einklang bringen.

## Ergebnisse

Mit der hier vorgeschlagenen Methode können Zinssätze analysiert werden, für die keine umfangreichen historischen Daten verfügbar sind. Dadurch ist eine plausible Zerlegung der entsprechenden Renditen möglich. Ohne die Restriktionsmethode würde das Modell unterstellen, dass sich die durchschnittlich erwarteten kurzfristigen Zinsen in den letzten Jahren kaum verändert hätten. Ferner wird gezeigt, dass sich eine Korrektur der Renditezerlegung erübrigt, wenn die Schätzungen auf Grundlage einer langen Datenreihe erfolgen. Allgemeiner betrachtet leistet die vorliegende Arbeit einen Beitrag zur Klärung der Frage, weshalb sich die aus Zinsstrukturmodellen abgeleiteten Zinserwartungen anders als die Erwartungen entwickeln können, die aus Umfragen ermittelt werden.

Bundesbank Discussion Paper No 27/2021

# Decomposing the Yield Curve with Linear Regressions and Survey Information*

Arne Halberstadt

**Abstract**

The decomposition of bond yields into term premiums and average expected future short rates is impaired by the limited availability of information about the dynamics of the expectations component. Therefore, many studies require the model-implied average expected future short rates to be close to short rate expectations from surveys. In this paper, I restrict the variance of changes in model-implied average expected future short rates to match the variance of changes in short rate expectations from surveys. The variance of changes in survey expectations is relatively similar across markets and thus provides a reliable source of additional information about the expectation formation of investors. Technically, I impose a nonlinear restriction to the term structure model of Adrian, Crump, and Moench (2013). I show that typical small sample problems of term structure estimations can be mitigated if the restriction on the variance of changes is imposed. However, the analysis also makes a case for unrestricted estimations if they are based on a dataset with a typical sample length in macro finance, though.

**Keywords:** Affine Term Structure Models, Empirical Finance.

**JEL classification:** E43, E44.

*Contact address: Deutsche Bundesbank, Wilhelm-Epstein-Strasse 14, 60431 Frankfurt am Main, Germany. Phone: +49 (0)69 9566 7079. E-mail: arne.halberstadt@bundesbank.de. I would like to thank Jean-Paul Renne, Leo Krippner, Wolfgang Lemke, Emanuel Moench, Refet Gürkaynak, Christian Speck, and seminar participants at the CEF 2019 and the Bundesbank Term Structure Workshop 2019 for helpful comments and discussions. The views expressed in this paper do not necessarily reflect those of the Deutsche Bundesbank or the Eurosystem.

# 1 Introduction

Affine term structure models are widely used for pricing the cross-section of bond yields and decomposing yields into term premiums and average expected future short rates.[1] The decomposition, however, can be impaired by the limited availability of information about the dynamics under the physical probability measure. The lack of sufficient information may come from an estimation sample that is too short and/or a model framework with too many parameters that need to be identified (see, e.g., Kim and Orphanides, 2012; Wright, 2014). Consequently, the literature suggests additional restrictions or the incorporation of additional information as a way of improving the decomposition of the yield curve. Furthermore, Bauer, Rudebusch, and Wu (2012) suggest cleansing estimated model coefficients of short sample bias using sampling methods. Jardet, Monfort, and Pegoraro (2013) propose another statistical approach to obtain a reliable yield curve decomposition, namely the application of an averaging estimator.

Restrictions on the market prices of risk are of particular importance for yield curve decompositions. In order to avoid overidentification, Ang and Piazzesi (2003) propose zero restrictions for some elements of the market price of risk matrix from the outset. Furthermore, they impose zero restrictions on those elements of this matrix that are statistically not different from zero. Similarly, Joslin, Priebsch, and Singleton (2014) apply model selection criteria to reduce the number of free parameters in the market price of risk matrix. They find that imposing such zero restrictions leads to a higher persistence under the physical measure that governs the dynamics of average expected future short rates. Further, they propose an eigenvalue restriction which addresses the problem of excess stability more directly. According to this method, the persistence of the estimated autoregressive process of the state variables under the physical probability measure is made to equal that under the pricing measure.

Among others, Joslin et al. (2014) describe excessively low persistence of the risk factors under the physical measure as a key problem in term structure estimations: Especially if estimated from short samples, interest rate expectations are implausibly stable under this measure. Kim and Orphanides (2012) argue that small samples provide too little information for a precise characterization of the speed of mean reversion. This leads to an upward bias in the speed of mean reversion of interest rate expectations under the physical measure, and, consequently, to a downward bias in the model-implied persistence.

As an alternative approach to the small sample problem, Kim and Orphanides (2012) propose to incorporate survey data on interest rate expectations into the model estimation. Their model delivers plausible model-implied average expected future short rates from a relatively short data sample. Technically, the survey data are incorporated into the observation equation of a Kalman filter-based maximum likelihood estimation. Effectively, this allows the model-implied average expected future short rates to be restricted so that they are close to short rate expectations from surveys. Crump, Eusepi, and Moench (2016) avoid small sample problems completely by using survey information more extensively. They skip the estimation of a statistical model and calculate term premiums simply by subtracting survey-based expected average short rates from observed yields. They compile

[1]See, e.g., Piazzesi (2010) for a survey of affine term structure models. Rudebusch (2010), Gurkaynak and Wright (2012) and Duffee (2013) provide an overview of term structure models with a macro-financial focus.

an extensive data set of US interest rate surveys from various sources for this purpose.

In this paper, I propose an alternative method for restricting the stability of short rate expectations under the physical measure. Technically, I restrict the variance of changes in model-implied short rate expectations under the physical measure to match the variance of changes in survey expectations. Instead of binding the model-implied expectations under the physical measure to match the survey data closely, I thus only take into account how fast and how strongly survey participants change their expectations about future interest rates. The level remains unrestricted and effectively remains almost unchanged. The survey information therefore has a less comprehensive influence on model-implied average future expected short rates than it has in the model of Kim and Orphanides (2012) or Kim and Wright (2005). These models also do not take surveys too literally, though, and allow for a measurement error. Among other potential shortcomings of surveys, Kim and Orphanides (2012) warn that the average value of survey expectations may not coincide with the expectations of the marginal investor. Further, the survey participants do not have a financial incentive to reveal their true expectations.

I incorporate the variance restriction into the linear regression model of Adrian et al. (2013). All pricing factors are assumed to be observed without measurement error in this model, including in particular those which are extracted as principal components from yields. In such a model without filtered state variables, the restriction on the variance of differences provides a novel mechanism to incorporate survey information into the model estimation. Malik and Meldrum (2016) also incorporate survey information into the Adrian et al. (2013) model. They add it directly to the vector of pricing factors.

I discuss the model restriction mainly based on an application to German Bund yields in a sample from 1991 to 2019. To illustrate the importance of the sample length, I compare model results derived from this full set of data with those derived from a truncated sample (2008-2019). I also show results for an application to US data in order to compare the main results to those from an alternative data set. Three principal components are assumed as state variables under the pricing measure (spanned). Two additional (unspanned) macro factors move yields under the physical measure, thus affecting the decomposition of yields into short rate expectations and term premiums. Modelling these factors as unspanned reflects the finding that they may be relevant for forecasting future interest rates, although they are not relevant for pricing the term structure contemporaneously.[2]

The results show that the restriction method is useful for alleviating small sample problems. That is, restricted average future expected short rates from a truncated sample of 12 years are more volatile than unrestricted estimates. They feature further desired properties which are not already imposed by construction, though: First, they exhibit a sensible cyclical pattern. Second, they develop relatively similarly to those estimated from a long sample which covers the time period of the truncated sample. Third, they are relatively similar to average future expected short rates derived from alternative restricted models from the literature. In particular, I obtain similar results with the bias-corrected method of Bauer, Rudebusch, and Wu (2014) and with an eigenvalue restriction à la Joslin et al. (2014).

[2]See Wright (2011) for references to papers rationalizing this assumption. Bauer and Rudebusch (2017) find strong statistical evidence for spanned models, though. Duffee (2013) also reviews the spanning property.

However, my model results are hardly affected by the restriction when it is implemented with the full sample of data covering 29 years. This follows directly from the calibration of the restriction according to interest rate expectations from surveys. There is only limited information available on long-term survey expectations. The surveys on hand suggest that interest rate expectations are similarly stable to those from the unrestricted model, if it is estimated with fairly long time series. This observation thus makes a case for unrestricted estimations if they are based on a data set with a typical sample length in macro finance. Similarly, the eigenvalue restriction affects the decomposition only modestly. The bias correction, however, implies pronounced adjustments of unrestricted estimates also in the full sample estimation.

The following section describes the modelling framework. It contains a short overview of the unrestricted model (Adrian et al., 2013). The section further describes the restriction method and the data used for the empirical implementation. Section 3 is focused on the main empirical results, namely on how the restriction affects the persistence of average future short rate expectations. Section 4 adds details on the risk pricing of the model. An application of the model to US data in Section 5 reviews the implications of the restriction in another data set. Section 6 concludes.

# 2 Empirical implementation and model approach

## 2.1 The unrestricted model: Adrian et al. (2013)

I apply the model of Adrian et al. (2013, ACM) with unspanned macroeconomic factors for German Bund yields on a monthly frequency.[3] Overall, there are five factors. The first three principal components of bond yields enter the model as spanned factors ($X_t^s$, $k_s = 3$). Two macroeconomic indicators for the Euro Area, namely harmonized consumer price inflation and deviation of industrial production from trend, are assumed to be unspanned factors ($X_t^u$, $k_u = 2$). Yields load only under the physical measure on these macroeconomic factors.

I consider two different samples, one rather short sample for the period from January 2008 to December 2019 and another sample from January 1991 to December 2019. The full sample estimation serves as a benchmark against which I compare the results from the truncated sample. The full sample contains 348 observations and is thus still small in an econometric sense. Therefore, estimates from this sample may also be biased (see Appendix A for a simulation exercise). Nevertheless, the full sample has a typical length for a term structure analysis with macroeconomic factors: Joslin et al. (2014) apply their model to a sample of 276 monthly observations. They also add an estimation on an extended sample including 434 months. Kim and Orphanides (2012) discuss their approach to handling the small sample bias on a sample of 168 monthly observations. The length of their sample is thus similar to my truncated sample (144 observations). Duffee (2013) uses a sample of only 235 observations. However, his data sample of quarterly data covers a longer period of time, and hence also includes more business cycles than the other papers.

[3] German Bund yields and parameters by the method of Svensson (1994) are available on the Bundesbank's website: https://www.bundesbank.de/dynamic/action/en/statistics/time-series-databases/time-series-databases/759784/759784?listId=www_skms_it03c.

I will now briefly outline the modelling framework of Adrian et al. (2013). For details on their model and the estimation method, I refer to the original paper. The ACM model is not a shadow/lower bound model. That means that it neither states a level below which interest rates cannot fall, nor does it allow for volatility compression of interest rates approximating a lower bound.[4] Such model characteristics would generally fit well to an analysis of German Bund yields, which have hovered at very low levels in recent years. The discussion of my approach in the unchanged ACM model allows me to concentrate on the effects of the Var(Δ)-restriction, though, because the underlying model is thoroughly documented in the original paper. However, one could also incorporate a Var(Δ)-restriction into shadow/lower bound models.

In the ACM model, the factors $X_t = [X_t^s, X_t^u]$ are assumed to follow an autoregressive process:

$$X_t = \mu + \Phi \cdot X_{t-1} + \nu_t \tag{1}$$

The residuals $\nu_t$ are Gaussian distributed, with variance-covariance matrix $\Sigma$ and mean zero. The term structure of interest rates is calculated from the state variables and the factor loadings,

$$\begin{aligned} y_t^{(n)} &= -\frac{1}{n} ln(P_t^{(n)}) \\ &= -\frac{1}{n}(A_n + B_n' \cdot X_t + u_t^{(n)}), \end{aligned} \tag{2}$$

where $P_t^{(n)}$ is the price of a bond with maturity $n$ at time $t$, and $u_t^{(n)}$ is the corresponding log yield pricing error. The factor loadings are recursively calculated from the model parameters for maturities $n = 1, \ldots, N$, starting from $A_0 = 0$ and $B_0' = 0$:

$$A_n = A_{n-1} + B_{n-1}'(\mu - \lambda_0) + \frac{1}{2}(B_{n-1}'\Sigma B_{n-1} + \sigma^2) - \delta_0 \tag{3}$$

$$B_n' = B_{n-1}'(\Phi - \lambda_1) - \delta_1', \tag{4}$$

where $\sigma^2$ is the variance of the residuals from the excess return regressions, and $\lambda = [\lambda_0, \lambda_1]$ the market price of risk. Equations (2) to (4) describe bond pricing under the historical pricing measure ($\mathbb{Q}$). Under the physical measure ($\mathbb{P}$), the market price of risk does not enter recursions (3) and (4). In other words, $\Phi^{\mathbb{Q}} = (\Phi - \lambda_1)$ and $\mu^{\mathbb{Q}} = (\mu - \lambda_0)$.

Defining $A_1 = -\delta_0$ and $B_1 = -\delta_1$, one can write the affine equation for the short rate also as

$$r_t = \delta_0 + \delta_1' X_t + u_t^{(1)}. \tag{5}$$

[4] For a detailed discussion of shadow/lower bound models and a review of the literature, see Krippner (2015). For an analysis of a shadow rate model for the euro area, see, for example, Lemke and Vladu (2016) or Geiger and Schupp (2018).

Average future short rates over the horizon $h = 1, \dots, H$ are given by

$$r_{t,h} = \frac{1}{h}(r_t, \dots, r_{t+h-1}), \tag{6}$$

and model-implied expected average future short rates at time $t$ are then denoted as:

$$\begin{aligned} E_t(r_{t,h}) &= \delta_0 + \frac{1}{h}\delta_1' E_t(X_t + \cdots + X_{t+h-1}) \\ &= \delta_0 + \delta_1' \sum_{i=1}^{h-1} \Phi^i X_t. \end{aligned} \tag{7}$$

Rewriting the geometric finite series generated by $\Phi$, this equation becomes

$$E_t(r_{t,h}) = \delta_0 + \frac{1}{h}\delta_1'(I - \Phi^h)(I - \Phi)^{-1} \cdot X_t. \tag{8}$$

## 2.2 Restricted model

In the following, I show how I restrict the ACM model by incorporating information from interest rate surveys. The Var($\Delta$)-restriction aims at binding the persistence of model-implied short rate expectations over long horizons to the persistence of those from surveys.

For most countries, interest rate survey data are only available for a limited history. Consensus Economics offers a relatively long history of a survey of six-to-ten year ahead expectations for US short rates. They are available as of 1998 on a semi-annual frequency.[5] Surveys for long-term European short rate expectations start later. For the three-month euro interbank rate, Consensus Economics only provides information as of 2016. While the European data history is far too short, the modelling approach in this paper generally leaves some room for maneuver with respect to limited survey data availability: This is because I only use a specific piece of information from the survey data and do not incorporate an entire time series of observations directly. In contrast to other contributions to the literature, the modelling approach thus does not incorporate the mean of survey expectations, for example, nor does it require filtering of survey data.

For restricting the model to semi-annual survey information, expressions have to be derived for the first two moments of the model-implied short rate process which match this frequency. Analogously to Equation (8), the first moment of the short rate process is given by:[6]

$$E_{t-6}(r_{t-6,h}) = \delta_0 + \frac{1}{h}\delta_1'(I - \Phi^h)(I - \Phi)^{-1} \cdot X_{t-6} \tag{9}$$

[5]Since 2014, long-term survey expectations have been published on a quarterly frequency. For the sake of parsimony, I discard those additional observations and take only semi-annual information into account.

[6]Kim and Orphanides (2012) apply a Kalman filter in which missing observations obtain zero weight in the observation vector. Their approach can therefore also cope more easily with a change in the frequency of survey observations.

The six-month change in this model-implied expected value is then:

$$\begin{aligned}\Delta^{(6)} E_t(r_{t,h}) &= E_t(r_{t,h}) - E_t(r_{t-6,h}) \\ &= \frac{1}{h}\delta_1'(I-\Phi^h)(I-\Phi)^{-1}\cdot(X_t - X_{t-6})\end{aligned} \tag{10}$$

Furthermore, for restricting model-implied forward rates to long-term survey information, expectations of average short rates that will prevail between six years and ten years have to be derived:

$$\begin{aligned}E_t(r_{t+61,60}) &= E_t(E_{t+61}(r_{t+61,60})) \\ &= E_t(\delta_0 + \frac{1}{60}\delta_1'(I-\Phi^{60})(I-\Phi)^{-1}\cdot X_{t+61}) \\ &= \delta_0 + \frac{1}{60}\delta_1'(I-\Phi^{60})(I-\Phi)^{-1}\cdot E_t(X_{t+61}) \\ &= \delta_0 + \frac{1}{60}\delta_1'(I-\Phi^{60})(I-\Phi)^{-1}\cdot \Phi^{60} X_t\end{aligned} \tag{11}$$

The six-month change in this model-implied forecast is then:

$$\begin{aligned}\Delta^{(6)} E_t(r_{t+61,60}) &= E_t(r_{t+61,60}) - E_{t-6}(r_{t+61-6,60}) \\ &= \frac{1}{60}\delta_1'(I-\Phi^{60})(I-\Phi)^{-1}\Phi^{60}(X_t - X_{t-6})\end{aligned} \tag{12}$$

The variance of differences in forward short rate expectations can then be calculated directly from that:

$$\begin{aligned}Var_t(\Delta^{(6)} r_{t+61,60}) &= \Delta^{(6)} E_t(r_{t+61,60}) \cdot \Delta^{(6)} E_t(r_{t+61,60})' \\ &= \frac{1}{60^2}\delta_1'(I-\Phi^{60})(I-\Phi)^{-1}\Phi^{60} Var(X_t - X_{t-6}) \cdot \ldots \\ &\quad ((I-\Phi^{60})(I-\Phi)^{-1})'\delta_1\end{aligned} \tag{13}$$

For ease of notation, I will refer to this forward rate variance as $Var(\Delta fwd_t^{\mathbb{P}})$ in the figures and tables of this paper. I incorporate the survey information into the term structure model by restricting the vector autoregression under the physical measure in the following way: $\Phi$ is estimated under the restriction that the model-implied variance of differences in forward short rate expectations is equal to the variance of differences in long-term short rate expectations from surveys, $\sigma^2_{\Delta restr}$. Specifically, the constraint is implemented by setting the variance expression in Equation (13) equal to $\sigma^2_{\Delta restr}$.

I obtain estimates for the restricted coefficients $\Phi^r$ by minimizing the sum of squared residuals of the vector autoregression under the physical measure. Specifically, I optimize the criterion function

$$\min_{\Phi} Q(\Phi, X_t) = g(\Phi, X_t)\cdot W \cdot g(\Phi, X_t)' \tag{14}$$

$$s.t. \quad Var(\Delta fwd_t^{\mathbb{P}}) = \sigma^2_{\Delta restr},$$

where

$$g(\Phi, X_t) = \frac{1}{T-1} \sum_{t=2}^{T} (\Phi \cdot X_{t-1} - X_t). \tag{15}$$

The optimization method is similar to the generalized method of moments (GMM) approach, but I impose equal weighting by $W = I$. Equal weighting turned out to be most robust in this analysis compared to opimal weighting, particularly for the short sample applications discussed below. For consistency, all results discussed in this paper are based on an optimization with equal weighting. For completeness, however, I add the calculation of the GMM weighting matrix in my setting. The description is based on the detailed discussion of the procedure in Hamilton (1994). In the GMM estimation, $W$ is a positive-definite weighting matrix,

$$W = \frac{1}{T-1} \sum_{t=2}^{T} (\Phi \cdot X_{t-1} - X_t) \cdot (\Phi \cdot X_{t-1} - X_t)'. \tag{16}$$

$W$ is initialized according to this equation with the unconstrained OLS estimate of $\Phi$. During the optimization iterations, it is repeatedly updated.[7]

The constrained optimization of $\Phi$ is carried out after the unrestricted three-step estimation described by ACM (see section 2.2 in their paper). Specifically, the three-step estimation of ACM provides first estimates of $\Sigma$, $\Phi$, $\lambda_0$ and $\lambda_1$. Yields under the pricing measure are also determined according to this three-step estimation. Then, $\delta_0$ and $\delta_1$ are obtained by regressing the short rate on an intercept and the state variables. The restricted optimization described in Equations (14)-(15) is carried out afterwards to obtain the restricted $\Phi^r$ ($\Sigma$, $\Phi$, $\delta_0$, $\delta_1$). The risk parameters $\lambda_0$ and $\lambda_1$ are updated for the restricted $\Phi^r$, $\lambda_1 = \Phi^r - \Phi^{\mathbb{Q}}$ and $\lambda_0 = \mu^r - \mu^{\mathbb{Q}}$.

# 3 Results

## 3.1 Comparing unrestricted model results to survey information

Standard estimations of term structure models tend to imply long-term expected short-term rates that are too stable compared to survey information. This is particularly the case for short sample estimations (Kim and Orphanides, 2012; Wright, 2014). To better understand what surveys actually reveal about the stability of interest rate expectations, I start by considering descriptive statistics of survey data for different markets. I then compare the statistics to estimates from an unrestricted model.

At this stage, I do not estimate the persistence of surveys in a vector autoregression. Instead, I consider descriptive statistics of the data that provide information about the stability of surveys expectations, but are independent of modelling assumptions.

[7] To implement the optimization, I use the estimation routine *fmincon* in Matlab for any choice of the weighting matrix. The constrained optimization takes about seven seconds on an ordinary laptop.

Table 1 contains information about long-term expectations (six-to-ten years ahead) for different markets and two different time samples: US short rates, ten-year US Treasury yields, and ten-year German Bund yields. It shows that not only the interest rate data, but also the survey data trended downwards over the last decades. This causes an obvious commonality in the properties of the surveys: The downward trend increased the variance of the series, relative to the variance of those series in first differences. The variances are thus higher than the variances of the data in first differences.

| | 2008-2019 | | | | 1991-2019 | | | |
|---|---|---|---|---|---|---|---|---|
| | US 3m | US 10y | DE 10y | ACM | US 3m* | US 10y | DE 10y | ACM |
| Var | 0.36 | 0.50 | 0.58 | 0.02 | 0.55 | 1.05 | 2.49 | 1.57 |
| Var($\Delta$) | 0.04 | 0.04 | 0.08 | 0.00 | 0.06 | 0.09 | 0.09 | 0.08 |

Table 1: Empirical variance of six-to-ten years ahead interest rate expectations from surveys of Consensus Economics and the unrestricted model of Adrian et al. (2013) for the sample from 1991 to 2019 and from 2008 to 2019 with German Bund yields, rounded to the two nearest basis points. Survey data for US short rates are only available as of April 1998 (*).

The variance of differences is relatively similar across markets. This is particularly the case for the long sample starting in 1991. The similarity is somewhat surprising given that the surveys refer to interest rates of very different maturities and from different jurisdictions. The variance of differences thus appears to provide a relatively solid piece of information about the stability of interest rate expectations from surveys. Therefore, I will calibrate the restriction on these values.

As mentioned above, Consensus Economics has provided quarterly updates of its survey of long-term expectations only since July 2014. I do not take the additional observations into account, because this would complicate the matching of the frequencies. Quantitatively, the impact of this simplification appears to be small, because the variance of the survey data series with and without the additional observations is relatively similar. A thorough assessment of the impact is difficult, however, given that the greater frequency has made only ten additional observations more available on top of the semiannual observations after July 2014.

The table also contains information about the stability of the long-term average of future short-term rates implied by an unrestricted term structure model (Adrian et al., 2013). The six-to-ten years ahead forward short rate expectations refer to the same forecast horizon as the long-term survey expectations. However, as already discussed above, there is no survey for German or Euro Area long-term short rate expectations with which one could compare the model-implied values.

However, judging from the survey data available from the US, it is not obvious that the stability of model-implied short rate expectations diverges materially from that from surveys. Considering the full sample starting in 1991, the model-implied values actually appear to be in line with the surveys: The variance of differences of model-implied German average expected future short rates is 0.08. This is a bit higher than what surveys report for US short rate expectations (0.06), and slightly lower than what surveys report for German and US ten-year government bond yield expectations (0.09).

The picture is different when considering the truncated sample. All surveys do indicate that the interest rate expectation are more stable in recent years. In the unrestricted model estimates, however, almost no variation can be detected. The variance of the forward short rate expectations is as low as 0.02, and the variance of differences is even nil. The comparison with survey information thus clearly corroborates the results in the literature: Unrestricted average expected future short rates are excessively stable if estimated from a small sample.

In the optimization, I calibrate the model restriction with information about the variance of US short rates (i.e. $\sigma^2_{\Delta restr} = 0.06$ in the full sample estimation and $\sigma^2_{\Delta restr} = 0.04$ in the truncated sample estimation). The implementation with US data is not an indisputable choice for an estimation of the German term structure. However, apart from the obvious lack of a sufficiently long history of survey information on European short rate expectations, there are ex ante a couple of reasons for this choice: I choose this proxy primarily because it provides the longest history of long-term surveys about expectations for a short rate. Second, the actual value for the variance of differences of the US short rate survey is the lowest of this metric for the surveys under consideration. Calibrating the restrictions on this value thus provides a reasonable upper limit for illustrating the effect of the restriction. Note, further, that I discuss the sensitivity of my results to alternative choices for $\sigma^2_{\Delta restr}$ in greater detail below.

## 3.2 Results from the restricted model

I begin the discussion of results from the restricted model by considering the model-based persistence. Table 2 shows the maximal absolute eigenvalues of the VAR coefficients from various models, again for both the truncated sample and the full sample estimation. The table also provides the variances of implied forward rates in first differences. This allows the (imposed) change of the variance of differences to be compared with its effect on eigenvalues. Additionally, it serves to compare the results with alternative models below. Note, however, that these variances are actually already reported in Table 1. The variances of differences from the restricted model coincide with those from the surveys shown above, because the restriction of matching the survey value is always precisely achieved.

The table also provides half-life periods, which illustrate the variation in the persistence of all models intuitively. The half-life $H_{.5}$ describes the number of months it takes for the average expected future short rates to halve their initial value, given the model-implied speed of mean reversion and an expected exponential decay of expected short rates:[8]

$$E(r_{t+H|t}) = \phi^H r_t.$$

[8] Piazzesi (2010) discusses the persistence of yields based on half-life periods of shocks to these yields. More generally, Dias and Marques (2010) review the characteristics of this metric, its limitations as a measure for persistence, and alternative measures for persistence.

Taking the maximal eigenvalue as the decay constant, the half-life $H_{.5}$ is given by:

$$\begin{aligned} eig(\phi)^{H_{.5}} &= \frac{1}{2} \\ H_{.5} \cdot log(eig(\phi)) &= -log(2) \\ H_{.5} &= -\frac{log(2)}{log(eig(\phi))} \end{aligned} \tag{15}$$

Additionally to the values of $H_{.5}$ in the table, Appendix B also contains a figure which illustrates the functional relationship of Equation (15).

Consider the eigenvalues for the truncated sample first (2008-2019). The eigenvalue of the unrestricted coefficient matrix under the physical measure, $\Phi$, is relatively low ($eig = 0.9811$), indicating a lower persistence of short rate expectations under the physical measure than under the pricing measure (second panel of Table 2, column ACM). The persistence from the restricted estimation is materially higher ($eig = 0.9967$ in column Var($\Delta$)). It is thus very similar to the persistence under $\mathbb{Q}$ which is unaffected by the restriction. The higher variance of differences in the restricted model thus corresponds with a higher model-implied persistence as measured by the eigenvalues. Therefore, imposing the restriction delivers the desired effect: It prevents implausibly low persistence levels in short sample estimations. Effectively, it brings the implied persistence under the physical measure from the truncated sample closer to the value estimated from the full sample.

The variance restriction affects the persistence similarly to the eigenvalue restriction of Joslin et al. (2014). This can be seen in Table 2 which contains results from an implementation of an eigenvalue constraint à la Joslin et al. (2014) in the Adrian et al. (2013) model framework. Actually, their modelling approach is technically very different to my approach. They set the largest eigenvalue of the transition matrix under the physical measure equal to the one under the pricing measure straightaway in order to impose a higher persistence on $\Phi$. This reveals that their approach follows a different motive than mine: They argue that the persistence of expected short rate dynamics can be assumed to be the same under both probability measures. My model restriction is instead motivated by information from survey data. The reason for similar results in both models is that the survey data support the idea of Joslin et al. (2014) of a similar persistence under both probability measures. However, the forward rate variance $Var(\Delta fwd_t^{\mathbb{P}})$ from an eigenvalue-constrained model does not turn out to vary as much as survey data suggest.

I turn now to the full sample estimation (1991-2019). The variance of differences of the unrestricted estimates is similar to that from survey expectations (see upper panel of Table 2). The restriction criterion therefore requires only a marginal adjustment in the variance of first differences, and this translates to a small impact on the model-implied persistence: The eigenvalues are 0.9939 and 0.9930 for the unrestricted and the restricted model, respectively. The model-implied persistence from the full sample is thus in line with survey information: This suggests that also the unrestricted model estimates are not biased. The eigenvalue restriction as in Joslin et al. (2014) causes a stronger, but still moderate adjustment of the persistence under the physical measure, because this is not very different from the persistence under the pricing measure ($eig = 0.9896$). Consequently, it follows that the forward rate variance $Var(\Delta fwd_t^{\mathbb{P}})$ from the eigenvalue-restricted model is lower than the unrestricted one in the full sample estimation.

| | | ACM | Var(Δ) | JPS-eig | BRW |
|---|---|---|---|---|---|
| 1991-2019 | $\text{eig}(\Phi^{\mathbb{Q}})$ | 0.9896 | 0.9896 | 0.9896 | 0.9947 |
| | $H_{.5}$ $\text{eig}(\Phi^{\mathbb{Q}})$ | 5.5 | 5.5 | 5.5 | 10.9 |
| | $\text{eig}(\Phi)$ | 0.9939 | 0.9930 | 0.9896 | 1.0011 |
| | $H_{.5}(\text{eig}(\Phi))$ | 9.4 | 8.3 | 5.5 | - |
| | $\text{Var}(\Delta fwd_t\text{-}TP)$ | 0.2046 | 0.2046 | 0.2569 | 0.3602 |
| | $\text{Mean}(\Delta fwd_t\text{-}TP)$ | 2.04 | 2.03 | 2.03 | 2.16 |
| | $\text{Var}(\Delta fwd_t^{\mathbb{P}})$ | 0.0785 | 0.0643 | 0.0479 | 0.4677 |
| | $\text{Mean}(\Delta fwd_t^{\mathbb{P}})$ | 2.49 | 2.50 | 2.50 | 2.39 |
| | $\min(\Delta fwd_t^{\mathbb{P}})$ | 0.40 | 0.58 | 1.07 | -2.10 |
| | $\max(\Delta fwd_t^{\mathbb{P}})$ | 5.27 | 5.03 | 4.60 | 8.59 |
| 2008-2019 | $\text{eig}(\Phi^{\mathbb{Q}})$ | 0.9956 | 0.9956 | 0.9956 | 0.9981 |
| | $H_{.5}$ $\text{eig}(\Phi^{\mathbb{Q}})$ | 13.1 | 13.1 | 13.1 | 30.4 |
| | $\text{eig}(\Phi)$ | 0.9811 | 0.9967 | 0.9956 | 1.0030 |
| | $H_{.5}(\text{eig}(\Phi))$ | 3.0 | 17.5 | 13.1 | - |
| | $\text{Var}(\Delta fwd_t\text{-}TP)$ | 0.2676 | 0.1769 | 0.1842 | 0.1690 |
| | $\text{Mean}(\Delta fwd_t\text{-}TP)$ | 2.21 | 2.22 | 2.22 | 2.28 |
| | $\text{Var}(\Delta fwd_t^{\mathbb{P}})$ | 0.0023 | 0.0403 | 0.0287 | 0.0520 |
| | $\text{Mean}(\Delta fwd_t^{\mathbb{P}})$ | 0.15 | 0.13 | 0.13 | 0.14 |
| | $\min(\Delta fwd_t^{\mathbb{P}})$ | -0.06 | -0.74 | -0.63 | -0.88 |
| | $\max(\Delta fwd_t^{\mathbb{P}})$ | 0.52 | 1.55 | 1.38 | 1.79 |
| EAPP (91-19) | $\text{Var}(\Delta fwd_t^{\mathbb{P}})$ | 0.0126 | 0.0111 | 0.0113 | 0.0503 |
| | $\text{Mean}(\Delta fwd_t^{\mathbb{P}})$ | 0.69 | 0.83 | 1.31 | -1.47 |
| | $\min(\Delta fwd_t^{\mathbb{P}})$ | 0.40 | 0.58 | 1.07 | -2.10 |
| | $\max(\Delta fwd_t^{\mathbb{P}})$ | 0.99 | 1.12 | 1.57 | -1.01 |
| EAPP (08-19) | $\text{Var}(\Delta fwd_t^{\mathbb{P}})$ | 0.0006 | 0.0157 | 0.0111 | 0.0201 |
| | $\text{Mean}(\Delta fwd_t^{\mathbb{P}})$ | 0.01 | -0.39 | -0.34 | -0.48 |
| | $\min(\Delta fwd_t^{\mathbb{P}})$ | -0.06 | -0.74 | -0.63 | -0.88 |
| | $\max(\Delta fwd_t^{\mathbb{P}})$ | 0.08 | -0.16 | -0.14 | -0.21 |
| Data | | Var(Δfwd) | Mean(fwd) | Min(fwd) | Max(fwd) |
| 1991-2019 | | 0.2728 | 4.54 | -0.48 | 8.34 |
| 2008-2019 | | 0.2967 | 2.41 | -0.48 | 5.14 |

Table 2: The table shows the absolute values of the largest eigenvalues of the transition matrices, corresponding half-life periods $H_{.5}$ in years, and time series properties of six-to-ten year forward rate expectations under the physical measure and six-to-ten year forward rate term premiums for the following models: The unrestricted model of Adrian et al. (2013, ACM column), the model with the Var(Δ)-restriction, the ACM model with an eigenvalue constraint as in Joslin et al. (2014), and the bias-corrected model of Bauer et al. (2012). The two upper panels describe results from the full sample and the truncated sample. The two panels below provide time series properties of forward rate expectations for the low interest rate environment since the start of the EAPP in March 2015, calculated from the results of each of the two estimation samples. The bottom panel shows descriptive statistics of forward rate data.

The Var(Δ)-restriction and the eigenvalue restriction imply, in line with survey information, less persistent average expected future short rates than the unrestricted model. The literature tends to discuss the opposite case. Besides of the aforementioned reason of excess stability as a prominent symptom of short sample problems in unrestricted estimations, there is another obvious explanation for that: Joslin et al. (2014) discuss their model based on an application to US data. For US data, I also find that unrestricted average expected future short rates are less persistent than survey data, even considering a long sample of data (see Section 5 below). For UK data, however, Malik and Meldrum (2016) find that the inclusion of survey data makes the physical dynamics less persistent. My results for German Bund yields thus complement the findings from applications to US and UK data: For an application with a fairly long sample of data, they describe an example of an unrestricted estimation in which the model dynamics are more persistent under the physical measure than under the pricing measure.

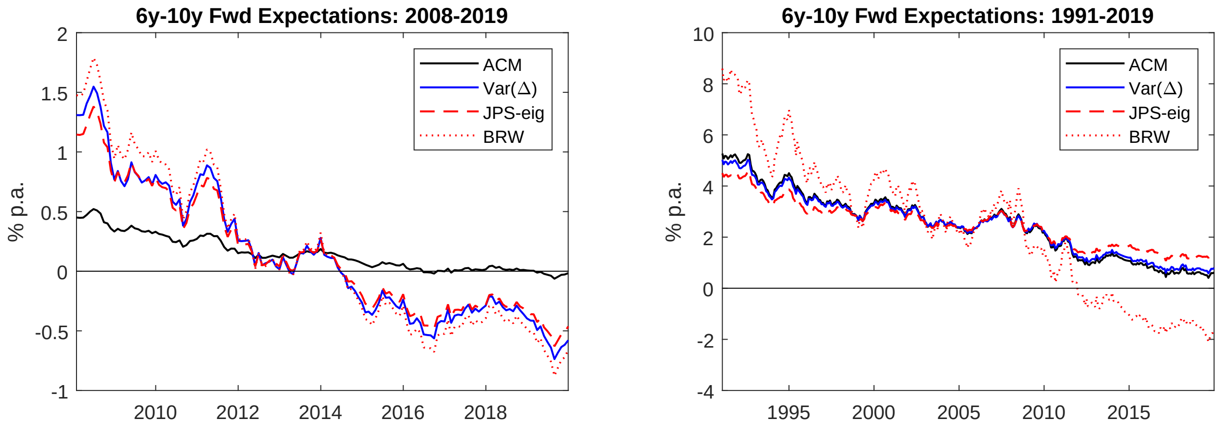


Figure 1: Six-to-ten year forward short rate expectations under the physical measure from different models, once estimated from a truncated sample (left), and once from the full sample (right) for German Bund yields. For model references to the mnemonics in the legend, see Table 2.

I also compare the results with Bauer et al. (2012). The last column of Table 2 contains the results of an application of their bias-corrected regression to Bund yield data. Their idea of correcting biased estimates using sampling-based methods is technically unrelated to my approach. The aim of correcting the persistence under the physical measure, however, is the same. The eigenvalues are overall higher and even slightly above one. Therefore, I consider these results only with caution. A thorough discussion of the Bauer et al. (2012) application to these data samples would require implementing the model with a restriction that prevents such explosive dynamics. The bias correction appears to have a similar effect on the physical dynamics in both the full sample and the truncated sample implementation. The variance of forward rate expectations is closer to the level of survey expectations in the case of the small sample estimation. The equivalent value from the full sample estimation is remarkably high and may indicate an overcorrection. This observation may not come as a surprise because of the explosive dynamics under the physical measure in my application. However, by comparing bias-corrected average expected future short rates of Bauer et al. (2012) to surveys, Wright (2014) also presumes an overcorrection of the bias by the method of Bauer et al. (2012), particularly in the application to German data. In their application, the largest eigenvalue is lower than one, though.

Figure 1 shows model-implied forward short rate expectations under the physical measure over six to ten years from all considered models. It illustrates a high co-movement of all restricted estimates from the truncated sample (left-hand panel). The unrestricted indicator is visibly less volatile than the restricted indicators and hardly varies at all in the last years of the observation period. Specifically, since the Eurosystem launched the Expanded Asset Purchase Programme (EAPP) in March 2015, the unrestricted indicator varied only between a minimum of -6 basis points and a maximum of +8 basis points (bottom panel of Table 2). The Var(Δ)-restricted estimates, instead, varied within a range of 58 basis points in this time period. This is comparable to the widths of the unrestricted and Var(Δ)-restricted indicators from the full sample estimation.

The right-hand panel of Figure 1 shows the results from the full sample. The results from the estimation with the restriction on the variance of differences almost coincides with the unrestricted estimates. This is not surprising, given that the unrestricted variance of differences is very similar to the value on which it is restricted: The restriction simply does not affect the model results if the information it adds is almost redundant.

The full sample estimates in the right-hand panel are more heterogeneous overall. The eigenvalue-restricted forward rate expectations are still relatively similar to the unrestricted estimates and only deviate moderately from them at the beginning and the end of the observation period. They are, for example, on average 64 basis points higher in the recent years in which the Eurosystem has been running the EAPP. Bias-corrected estimates are already reported to be materially more volatile also in this observation period, partly due to a unit root in the physical dynamics (see Table 1 above). The figure shows that this volatility is consistent with high deviations from the estimates of the other models. The bias-correction suggests very low levels of forward rate expectations six to ten years ahead particularly for the last years. It falls to $-2\%$ in 2019 and thus spans a range of more than 10 percentage points from this minimum to its maximum. This is more than double the amplitude than in the other models in the full sample estimation.

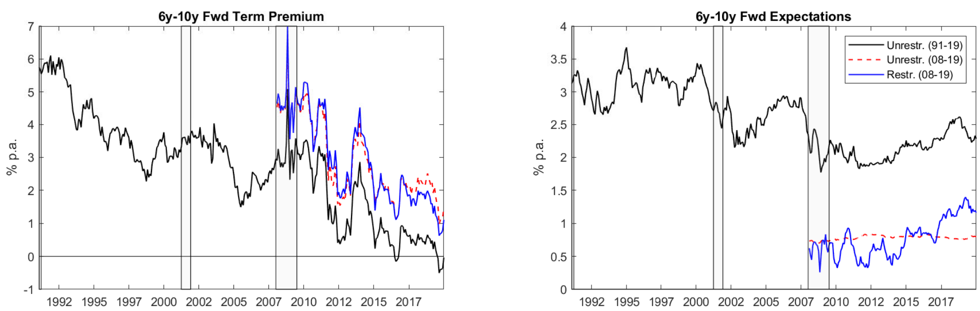


Figure 2: Comparison of forward rate components estimated from different data samples: Unrestricted results are plotted from both sample estimations. To achieve a clearer presentation, the restricted estimates from the full sample are omitted. They are very similar to the unrestricted full sample estimates (see Figure 1, right-hand panel).

In the following, I focus on the unrestricted model and the model with the restriction on the variance of differences. Figure 2 illustrates the key similarities and differences of these two approaches. It contains the six-to-ten year ahead forward rate components of the two sample estimations. The left-hand panel shows the forward term premium components. Restricted and unrestricted forward term premiums estimated from the truncated sample do not appear to differ materially. This is a consequence of the low level of the complementary forward yield component, i.e. the forward short rate expectations. Those are shown in the right-hand panel of Figure 2: This panel highlights the strong co-movement of the unrestricted full sample results and the restricted truncated sample results. However, forward expectations (forward term premiums) from the truncated sample are shifted downwards (upwards). Empirically, this mirrors a lower sample mean of forward rate expectations in the truncated sample and ultimately follows from downward trending interest rates over the last decades. The sample mean is assumed to be invariant over time in the model.

The marked difference in the level of forward rate expectations from the two samples, which is visible in Figure 2, gives rise to the question of whether the level should also be restricted. In principle, such a second condition could be added to the optimization approach in Equation (14). One could impose, for example, that the level complies with that from very long historical samples, as they are provided by the Macrohistory Database of Jorda, Schularick, and Taylor.[9] Alternatively, the level of model-implied forward rate expectations could also be restricted to that from survey expectations.

To evaluate the potential impact of such a restriction on the level of survey expectations, I briefly digress from the discussion of long-term forward rates. I check whether the sample means of model-implied rates are in line with survey information for shorter forecast horizons. The digression to surveys for shorter forecasting horizons is necessary because very little survey information on long-term short rate expectations is available for Germany, as already mentioned earlier. The average of one year ahead three-month interest rate expectations is 2.86% for 1991 until 2019, according to Consensus Economics. For 2008 to 2019, the mean of that survey is 0.73%. Survey expectations show a decline in short rate expectations of about 2 percentage points for these observation periods. Model-implied means of forward short rate expectations 12 to 15 months ahead are almost identical for the unrestricted model and the $\text{Var}(\Delta)$-restricted model. These measures are roughly in line with the aforementioned values from surveys (2.56% for the full sample, 0.15% for the truncated sample).

I now return to the long-term forward rate expectations. Note that the restricted model overestimates the decrease in forward short rate expectations over the sample period, at least compared to the restricted full sample estimation. Forward short rate expectations are 189 basis points lower in December 2019 than in January 2008. According to the restricted (unrestricted) full sample estimation, they decreased only by 166 basis points (186 basis points) over the same period.

Three recession periods are marked as shaded areas in the figure. The recession chronology is taken from the German Council of Economic Experts.[10] Methodologically, the recession identification follows the definition applied by NBER for the US and CEPR for the Euro Area. The full sample estimates reveal business cycle properties of yield

[9]See https://www.macrohistory.net/database and Jorda, Schularick, and Taylor (2017).

[10]See Box 7 in the third chapter of the Annual Report of the German Council of Economic Experts, https://www.sachverstaendigenrat-wirtschaft.de/en/publications/annual-reports/previous-annual-reports/annual-report-201718.html

components, albeit dampened due to the long-term horizon of the forward rates. Forward term premiums increase during recessions and fall during expansions: The path of future nominal short rate developments may appear more uncertain in recession periods, and market participants demand a higher premium for bearing this interest rate exposure affiliated with long-term bonds in times of recession. The cyclical pattern of forward rate expectations is less prominent. They decrease during recessions, but also during most of the expansions, contributing to the downward trend of yields over the last decades.

There are both theoretical and empirical contributions to the literature about the cyclicality of risk premiums, and I refer in the following only to a few of these studies which are of particular relevance for my application. Wachter (2006) develops a theoretical model in which term premiums are countercyclical. The counter-cyclicality originates from a preference structure featuring habit persistence and the assumption of risky returns of long-term bonds. The empirical evidence on the cyclicality of term premiums is rather mixed in the literature. Jardet et al. (2013) find countercyclical term premiums by applying an averaging estimator. Their averaging estimator helps to mitigate the problem of highly persistent model factors which I approach with the $Var(\Delta)$-restriction in this paper. Bauer et al. (2014) also report pronounced countercyclical forward term premiums for German bond yields from 1990 to 2009. But in contrast to the unrestricted ACM estimates and the Var($\Delta$)-restricted estimates considered in this paper, their forward term premium estimates remain ultimately unchanged over expansion periods, or even increase slightly. Consequently, the forward term premiums increase over their whole estimation period from 1990 to 2009. Wright (2011), by contrast, finds a relatively flat development of term premiums for Germany from 1990 to 2009.

Considering business cycle properties in the truncated sample yields few insights. This is not only because the truncated sample contains just one recession period in this particular application. Containing too few ups and downs of a cycle is in fact a determining element of a small sample bias that drives a wedge between small and long sample estimates: Kim and Orphanides (2012) describe a lack of information about the speed of mean reversion of short rate expectations as responsible for excessively stable short rate expectations in small sample applications.

The variance of differences is only restricted at one horizon in my model. Its value is marked as a dash-dotted line in Figure 3. It is calibrated to match the variance of differences from surveys, which is indicated by the red thick line at the right of each panel. However, model-implied yield expectations of all horizons are affected by the restriction, because the restricted transition matrix $\Phi$ enters the recursion in Equation (4) for every maturity. The figure illustrates this effect on the variance of differences of the entire term structure. Restricted and unrestricted variances of differences exhibit a similar cross-sectional mapping under the physical measure in both samples.

In the full sample estimation (right-hand panel), these concur surprisingly well with survey values for shorter forecast horizons which are not incorporated in the model estimation. This confirms the finding from above that the unrestricted model results for the full sample implementation are also well in line with survey information.

In the truncated sample (left-hand panel), survey values for forecast horizons of 48 months and less are not matched by the restricted nor by the unrestricted model. For horizons up to four years, the restriction at most allows the stability of average expected future short rates to move in the direction of survey expectations.

In applications which solely consider small data samples, it could therefore be worth to restricting a medium-term horizon instead of - or additionally to - a long-term horizon. I leave this question for future research. Note, however, that the variances of differences for shorter forecast horizons differ more across markets. A plot of the variances of differences of survey expectations for different markets is provided in the Appendix C (Figure 7). This observation supports the case for applying the restriction to a long-term forward rate in this application.

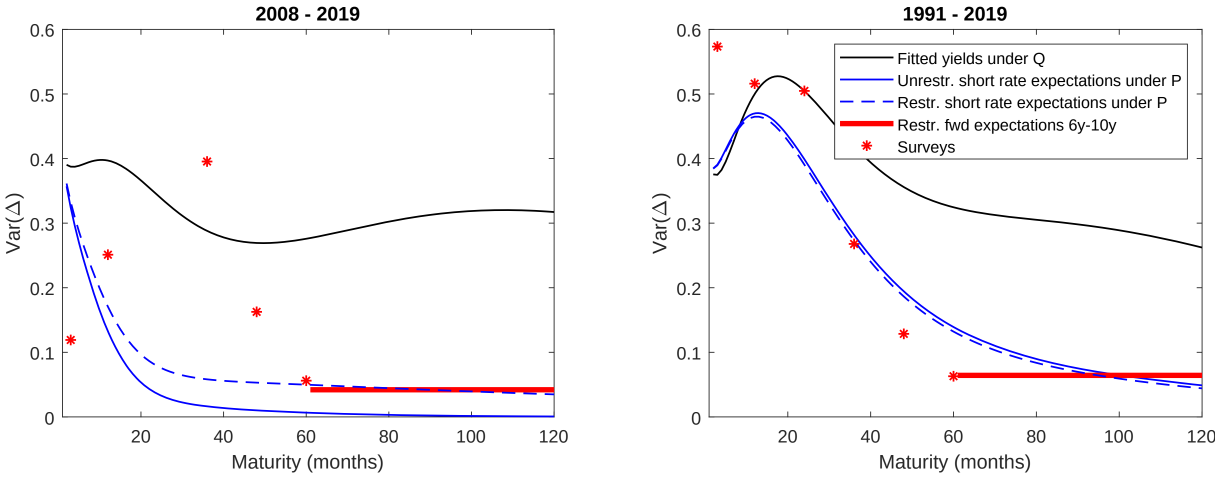


Figure 3: Variance of differences of model-implied yields under the pricing measure and the physical measure, the latter from the restricted and the unrestricted model. In the restricted model, the variance of differences in six-to-ten year forward short rate expectations under the physical measure is required to match the value known from a survey. The thick red line marks this value calculated from the survey for the six-to-ten year horizon. Red asterisks denote the values calculated from the survey for other horizons. For consistency, all survey information in this figure refers to US short rate expectations from Consensus Economics. The cross-sectional mappings of survey expectations for other markets are of a similar shape, though (Figure 7).

## 3.3 Robustness of the results

In the following, I review the reliability of the results of the restricted model. First, I will analyze the sensitivity of results with respect to the restriction level. Second, I will confront model-implied interest rate expectations with ex post realized short rates. In Appendix D, I also consider whether the impact of the restriction on the model-implied persistence depends on the length of the time series.

How sensitive is the model-implied persistence to changes in the Var(Δ)-restriction level? To answer this first question, I run the model for different restriction levels. Specifically, I consider a range from $\sigma^2_{\Delta restr} = 0.03$ to $\sigma^2_{\Delta restr} = 0.09$. All values observed from surveys fall into this range, as it is reported in Table 1. The result for the model-implied persistence hardly comes as a surprise: The lower the Var(Δ)-restriction level, the lower is the eigenvalue. The calibration on the highest variance level implies slightly nonstationary dynamics, because the largest eigenvalue is above one. What is remarkable, though, is the low amplitude of the implied eigenvalues (Table 3). This observation holds for

| | Restricted: 1991-2019 | | | Unrestricted 91-19 | |
|---|---|---|---|---|---|
| | $Var(\Delta) = 0.03$ | $Var(\Delta) = 0.05$ | $Var(\Delta) = 0.09$ | $eig(\Phi)$ | $eig(\Phi^{\mathbb{Q}})$ |
| $eig(\Phi)$ | 0.9897 | 0.9920 | 0.9945 | 0.9939 | 0.9896 |
| | Restricted: 2008-2019 | | | Unrestricted: 08-19 | |
| | $Var(\Delta) = 0.03$ | $Var(\Delta) = 0.05$ | $Var(\Delta) = 0.09$ | $eig(\Phi)$ | $eig(\Phi^{\mathbb{Q}})$ |
| $eig(\Phi)$ | 0.9951 | 0.9978 | 1.0001 | 0.9811 | 0.9956 |

Table 3: Absolute values of the largest eigenvalues of the transition matrix under the physical measure, $\Phi$, for a range of restriction levels and two different samples. For comparison, the eigenvalues of the unrestricted $\Phi$ and $\Phi^{\mathbb{Q}}$ are included in the right-hand panel. The unrestricted model-implied value of the variance of difference is $Var(\Delta) = 0.08$ in the full sample and $Var(\Delta) = 0.00$ in the truncated sample.

both the full sample and the truncated sample. If one assumes that the true level of persistence lays anywhere near to the survey information considered above, then the risk of mis-calibrating the restriction appears limited.

Second, manipulating $\Phi$ by calibrating the implied volatility on exogenously chosen levels may have consequences for the predictive ability of the model. Therefore, I now perform a quick forecasting exercise to make sure that the restriction has no unintended consequences. Such an investigation is particularly warranted in an overidentified model like the Adrian et al. (2013) model on which my model builds.

Further, and more generally, Bauer et al. (2012) argue that biased parameter estimates under the physical measure do not provide a correct picture of the speed of mean reversion of short rate expectations. Given the lack of data on the actual speed of mean reversion, I test whether investors would eventually be better off by forming their expectations according to the restricted or the unrestricted model. To answer this question, I will compare the predictive ability of the restricted and unrestricted model under the physical measure. I carry out a prediction exercise over a subsample of 17 years (01:1991-01:2008). Over this period, I calculate a series of realized average short rates, $\bar{y}^1_{t:t+H|t+H}$. This series contains ex post observed realizations to which I compare ex ante expectations from the restricted or the unrestricted model:

$$\bar{y}^1_{t:t+H|t+H} = \frac{1}{H}\sum_{h=0}^{H} y^1_{t+h}$$

In line with the longest horizon of the surveys, I set $H = 120$. This limits the latest point $t$ for a prediction to be ten years before the end of the main estimation sample.

The difference in the prediction accuracy, as measured by root mean squared errors, is very small overall (see Table 4, left-hand panel). The aim of an in-sample alignment of the volatility of the physical dynamics to survey expectations thus does not come at the cost of deteriorating predictions.

As a cross-check of the forecasting performance of the model in general, I also evaluate the out-of-sample forecasting performance under the pricing measure in the subsample from January 2001 to December 2019. For that, I compare model-implied yields of all maturities with their ex post realizations. The model has an acceptable forecasting performance, as one can see from the root mean squared errors in the right-hand panel of Table 4.

Furthermore, the residuals of the estimated restricted vector autoregressive process are generally not equal to the unrestricted estimates. Consequently, the implied variance-covariance matrix of the restricted vector autoregression (VAR), $\Sigma$, differs from the unrestricted variance-covariance matrix in this case. However, the in-sample fit of the state variable VAR deteriorates only negligibly under the restriction. The root mean squared errors are the same for the restricted and the unrestricted estimations up to several decimal places. This is also true for the estimation in the truncated sample.

| | $\mathbb{P}$ | | $\mathbb{Q}$ | | | |
|---|---|---|---|---|---|---|
| | $\bar{y}^1$ unrestr. | $\bar{y}^1$ restr. | 1m | 12m | 60m | 120m |
| h=3 | 0.47 | 0.47 | 0.29 | 0.19 | 0.18 | 0.14 |
| h=6 | 0.44 | 0.44 | 0.55 | 0.54 | 0.47 | 0.38 |
| h=12 | 0.58 | 0.59 | 1.21 | 1.31 | 1.18 | 1.01 |

Table 4: H-step ahead root mean squared errors of average short rate expectations under the physical measure (left-hand panel) and for out-of-sample forecasts under the pricing measure (right-hand panel).

# 4 Model inference and risk pricing

Obtaining reasonably persistent estimates of average future short rate expectations is the main objective of this paper. However, this should not come at the cost of implausible results elsewhere in the model. I will consider the model fit and the relevance of the state variables first. The number of state variables is of great relevance for the model fit and the size of the model. Three latent factors were found to be sufficient for determining the historical yield dynamics (see, e.g., Litterman and Scheinkman, 1991; Piazzesi, 2010). A Wald test clearly confirms that all latent factors are indeed spanned, i.e. none of the three spanned latent factors is redundant. Also with respect to the model fit, three principal components as spanned factors are enough for pricing the term structure accurately in the full sample estimation. Table 5 compares the root mean squared errors from the model containing three spanned factors and two unspanned factors with an alternative approach using five principal components as spanned factors. Adrian et al. (2013) favor such a setup with five spanned factors. The better model fit impedes the risk of serially correlated return pricing errors and consequently excess return predictability. In my application, however, the root mean squared errors are relatively small, and adding further spanned factors only marginally improves the model fit in the full sample estimation.

This result does not hold for the implementation with the truncated sample. The model fit does improve if I replace the two unspanned macro factors with two spanned principal components, namely by almost 5 basis points. As suggested by Adrian et al. (2013), I therefore consider the autocorrelation in the return pricing errors, $e_{t+1}^{(n-1)}$:

$$e_{t+1}^{(n-1)} = u_{t+1}^{(n-1)} - u_t^{(n)} + u_t^{(1)}. \tag{16}$$

| | $k_s = 5$, $k_u = 0$ | | $k_s = 3$, $k_u = 2$ | |
|---|---|---|---|---|
| | 08-19 | 91-19 | 08-19 | 91-19 |
| RMSE (unrestr.) | 0.0227 | 0.0640 | 0.0703 | 0.0659 |
| RMSE (restr.) | 0.0233 | 0.0641 | 0.0709 | 0.0660 |
| *Autocorrelations* | | | | |
| 12m | 0.3293 | 0.3221 | 0.1405 | 0.2764 |
| 120m | 0.3082 | 0.3271 | 0.1835 | 0.2871 |

Table 5: In-sample fit and autocorrelations of return pricing errors for different samples and factor sets. Autocorrelations of order one are calculated for pricing errors of bond yields with 12 and 120 months to maturity.

The autocorrelation of the return pricing errors remains relatively small in the truncated sample estimation (see the last two rows of Table 5). The risk of excess return predictability thus remains limited in both the truncated and the full sample estimation.

| | Unrestricted | | | | | |
|---|---|---|---|---|---|---|
| | $\lambda_0$ | PC1 | PC2 | PC3 | HICP | IP-Gap |
| PC1 | **-0.0092** | -0.0014 | **-0.0077** | 0.0052 | -0.0068 | 0.0024 |
| t-stat | -2.3090 | -0.2477 | -1.7595 | 1.1935 | -1.0981 | 0.4916 |
| PC2 | -0.0080 | 0.0138 | -0.0150 | -0.0163 | 0.0006 | -0.0190 |
| t-stat | -0.7230 | 0.9047 | -1.2426 | -1.3590 | 0.0377 | -1.4909 |
| PC3 | 0.0521 | 0.0106 | 0.0140 | -0.0566 | 0.0739 | -0.0281 |
| t-stat | 1.6436 | 0.2519 | 0.4138 | -1-6802 | 1.7121 | -0.8363 |
| | Restricted | | | | | |
| | $\lambda_0$ | PC1 | PC2 | PC3 | HICP | IP-Gap |
| PC1 | **-0.0092** | -0.0020 | -0.0076 | 0.0052 | -0.0072 | 0.0022 |
| t-stat | -2.3088 | -0.3544 | -1.7354 | 1.2036 | -1.1658 | 0.4474 |
| PC2 | -0.0080 | 0.0140 | -0.0150 | -0.0163 | 0.0007 | -0.0189 |
| t-stat | -0.7230 | 0.9168 | -1.2440 | -1.3590 | 0.0456 | -1.4869 |
| PC3 | 0.0521 | 0.0107 | 0.0140 | -0.0566 | 0.0740 | -0.0281 |
| t-stat | 1.6436 | 0.2548 | 0.4131 | -1.6805 | 1.7141 | -0.8350 |

Table 6: Risk premium parameter matrix $[\lambda_0, \lambda_1]$ and t-statistics from the unrestricted and the restricted estimation. In both applications, two of the coefficients are significant at a 5% level (bold). Asymptotic properties of the market prices of risk are derived by Adrian et al. (2013).

I will now consider further aspects of model inference and model-implied risk pricing. For the sake of brevity, I will only report results from a full sample estimation regarding the risk pricing. The macroeconomic factors are modelled as unspanned, which keeps the number of parameters in the pricing equation low. A limited set of pricing factors avoids the problem of overfitting (Joslin et al., 2014). Implausibly high Sharpe ratios would hint at an overfitted model (Duffee, 2010), but in this application they turn out to be reasonably low on average (maximally 0.18 in the full sample estimation). In contrast to Joslin et al. (2014), I do not impose any zero restrictions on the risk premium parameters. However, I find a similar influence of the risk prices on expected excess returns as Joslin et al. (2014) for US data: Specifically, the real activity factor affects level risk pro-cyclically and slope risk counter-cyclically in an unrestricted estimation (Table 6). One important difference to Joslin et al. (2014) is, however, that inflation risk affects level risk premiums counter-cyclically. Imposing the variance restriction affects the risk premium parameters only quantitatively, but not their direction of influence. The diverging cyclicality of yield components from the two estimation approaches illustrated in Figure 2 thus does not follow from qualitatively different risk premium estimates. It is instead caused by quantitative differences in implied risk premiums.

Only a few coefficients related to the first principal component (i.e., level risk) are significantly different from zero at a 5% level. Nevertheless, slope and curvature risk are also relevant at hardly less strict significance levels. There are risk premium coefficients related to these factors that are significant at a 10% level, in both of the two estimation approaches. The real activity factor also affects slope risk significantly at a 10% level.

# 5 Application to US data

I also apply the restricted estimation to US data. In this way, I check whether the implications of the restriction discussed above also appear in an application with another data set. I consider the US data over the same time periods as the German / Euro Area data (1991-2019 and 2008-2019). As yield data, I use the US Treasury yield curve data of Gurkaynak, Sack, and Wright (2007) which can be downloaded from the website of the Federal Reserve Board.[11] I follow a working paper version of Joslin et al. (2014) in their choice of the macro data, because all these data are freely available for download: The historical (real-time) data of the Chicago Fed National Activity Index serves as a measure of real economic activity.[12] The measure of price inflation is the first principal component of the CPI (all items) and the personal consumption deflator, both in log differences. The CPI data are published by the U.S. Bureau of Labor Statistics, and the personal consumption deflator by the U.S. Bureau of Economic Analysis. Both can be downloaded from the FRED economic database of the St. Louis Fed.[13] I calibrate the Var($\Delta$)-restriction on the same level as in the German / Euro Area application (i.e., 0.04 for the truncated sample and 0.06 for the full sample).

That makes the restriction weightier in the US application, because the variance of differences of unrestricted US forward rates under the physical measure is lower than that for the German ones, both in the truncated sample and in the full sample. Consequently, restricted and unrestricted estimates for the US do differ in the full sample implementation (Figure 4). Also the eigenvalues indicate a stronger effect of the restriction on the persistence, even in the full sample estimation.

In the short sample, the observed variance of differences is practically zero. The largest eigenvalues under the two pricing measures from the unrestricted estimation differ markedly in the truncated sample. Figure 4 illustrates again excessively stable short rate expectations in a small sample estimation. By contrast, the restricted short rate from the

[11]See https://www.federalreserve.gov/pubs/feds/2006/200628/200628abs.html.

[12]The data can be downloaded from https://www.chicagofed.org/research/data/cfnai/historical-data.

[13]See https://fred.stlouisfed.org/.

truncated sample displays a similar development to the full sample estimate over the last years, albeit more volatile.

The restriction thus implies similar effects for the decomposition of both US Treasury yields and German Bund yields. It is rather the size of the effects which differs between these two applications.

| | 2008-2019 | | 1991-2019 | |
|---|---|---|---|---|
| | ACM | Var(Δ) | ACM | Var(Δ) |
| $\text{eig}(\Phi^{\mathbb{Q}})$ | 1.0051 | 1.0051 | 1.0020 | 1.0020 |
| $\text{eig}(\Phi)$ | 0.9589 | 0.9924 | 0.9839 | 0.9856 |
| $Var(\Delta fwd_t^{\mathbb{P}})$ | 0.0005 | 0.0403 | 0.0475 | 0.0643 |

Table 7: The table reports the absolute values of the largest eigenvalues of the transition matrices for the unrestricted model of Adrian et al. (2013, ACM column) and the model with the Var(Δ)-restriction for an application with US Treasury yields.

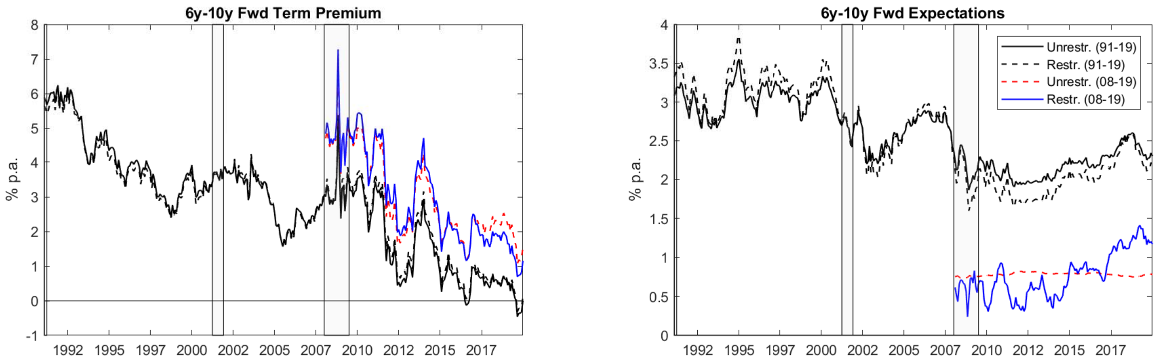


Figure 4: This figure compares US forward rate components estimated from different data samples (analogously to the illustration of German forward rate components in Figure 2): Unrestricted results are plotted from both sample estimations.

# 6 Conclusion

A bias in term structure model estimations from small samples has been repeatedly described in the literature: Model-implied average future expected short rates have been found to be excessively stable. I have also illustrated this effect for German Bund data in this paper.

Taking into account interest rate expectations from surveys has proven to be useful for avoiding small sample problems (Kim and Orphanides, 2012). In this paper, I provide an alternative method for incorporating survey expectations into a term structure model. The method allows for the inclusion of survey information in term structure models which are estimated with linear regressions and observed state variables. The method aligns the persistence of average future expected short rates with that observed in interest rate surveys and thus avoids the small sample bias.

Therefore, the restriction method may prove to be particularly useful in applications in which the data samples are rather short. For example, the data history on inflation-linked bond yields is rather short in many countries.

In the case of a reasonably long data sample, however, the persistence of unrestricted model-implied average future expected short rates does not differ much from survey evidence. My results thus also support the application of unrestricted models to term structure analyses if there is a sufficiently long data history available. From a broader perspective, the paper thus contributes to the understanding of the actual benefits of incorporating survey information into term structure model estimations.

# References


Adrian, T., R. K. Crump, and E. Moench (2013). Pricing the term structure with linear regressions. *Journal of Financial Economics 110*(1), 110–138.

Ang, A. and M. Piazzesi (2003). A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. *Journal of Monetary Economics 50*(4), 745–787.

Bauer, M. D. and G. D. Rudebusch (2017). Resolving the Spanning Puzzle in Macro-Finance Term Structure Models. *Review of Finance 21*(2), 511–553.

Bauer, M. D., G. D. Rudebusch, and J. C. Wu (2012). Correcting estimation bias in dynamic term structure models. *Journal of Business & Economic Statistics 30*(3), 454–467.

Bauer, M. D., G. D. Rudebusch, and J. C. Wu (2014). Term Premia and Inflation Uncertainty: Empirical Evidence from an International Panel Dataset: Comment. *American Economic Review 104*(1), 323–337.

Crump, R. K., S. Eusepi, and E. Moench (2016). The term structure of expectations and bond yields. Staff Reports 775, Federal Reserve Bank of New York.

Dias, D. A. and C. R. Marques (2010). Using mean reversion as a measure of persistence. *Economic Modelling 27*(1), 262–273.

Duffee, G. R. (2010). Sharpe ratios in term structure models. Economics Working Paper Archive 575, The Johns Hopkins University, Department of Economics.

Duffee, G. R. (2013). Bond Pricing and the Macroeconomy. *Handbook of the Economics of Finance 2*(Part B), 907–967.

Geiger, F. and F. Schupp (2018). With a little help from my friends: Survey-based derivation of euro area short rate expectations at the effective lower bound. Discussion Papers 27/2018, Deutsche Bundesbank.

Gurkaynak, R. S., B. Sack, and J. H. Wright (2007). The U.S. treasury yield curve: 1961 to the present. *Journal of Monetary Economics 54*(8), 2291–2304.

Gurkaynak, R. S. and J. H. Wright (2012). Macroeconomics and the term structure. *Journal of Economic Literature 50*(2), 331–367.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press.

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 & Finance 37*(2), 389–402.

Jorda, O., M. Schularick, and A. M. Taylor (2017). Macrofinancial History and the New Business Cycle Facts. *NBER Macroeconomics Annual 31*(1), 213–263.

Joslin, S., M. Priebsch, and K. J. Singleton (2014). Risk Premiums in Dynamic Term Structure Models with Unspanned Macro Risks. *Journal of Finance 69*(3), 1197–1233.

Kim, D. H. and A. Orphanides (2012). Term Structure Estimation with Survey Data on Interest Rate Forecasts. *Journal of Financial and Quantitative Analysis 47*(01), 241–272.

Kim, D. H. and J. H. 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, Board of Governors of the Federal Reserve System (U.S.).

Krippner, L. (2015). *Zero Lower Bound Term Structure Modeling: A Practitioner's Guide*. New York: Palgrave-Macmillan.

Lemke, W. and A. Vladu (2016). Below the zero lower bound - a shadow-rate term structure model for the euro area. *Bundesbank Discussion Paper 32*.

Litterman, R. B. and J. Scheinkman (1991). Common factors affecting bond returns. *Journal of Fixed Income 1*, 54–61.

Luetkepohl, H. (2005). *New Introduction To Multiple Time Series Analysis*. Berlin, Heidelberg: Springer-Verlag.

Malik, S. and A. Meldrum (2016). Evaluating the robustness of UK term structure decompositions using linear regression methods. *Journal of Banking & Finance 67*, 85 – 102.

Piazzesi, M. (2010). *Affine term structure models*, Chapter 12, pp. 691–766. Handbook of Financial Econometrics, ed. Y. Ait-Sahalia and L. Hansen.

Rudebusch, G. D. (2010). Macro-finance models of interest rates and the economy. *The Manchester School 78*(s1), 25–52.

Svensson, L. E. O. (1994). Estimating and interpreting forward interest rates: Sweden 1992-1994. NBER Working Papers 4871, National Bureau of Economic Research, Inc.

Wachter, J. A. (2006). A consumption-based model of the term structure of interest rates. *Journal of Financial Economics 79*(2), 365–399.

Wright, J. H. (2011). Term premia and inflation uncertainty: Empirical evidence from an international panel dataset. *American Economic Review 101*(4), 1514–34.

Wright, J. H. (2014). Term premia and inflation uncertainty: Empirical evidence from an international panel dataset: Reply. *American Economic Review 104*(1), 338–41.

# A Small sample bias of OLS estimates: Illustration

OLS estimates can generally be biased when they are derived from small samples. This is particularly the case if the variables are highly persistent. As a result, model-implied coefficients are prone to be downward biased (see, for example, Chapter 8.2 of Hamilton, 1994, for details on these small sample properties of OLS). Figure 5 illustrates this effect for the AR(1) process

$$x_t = \rho x_{t-1} + \epsilon_t,$$

where $\epsilon_t \sim \mathcal{N}(0,\, 1)$. I assume $\rho = .99$ and simulate the small sample distribution of this process using Monte Carlo methods as outlined in Hamilton (1994): I draw 5,000 starting values for $x_t$ from a normal distribution, $x_t \sim \mathcal{N}(0,\, 1/(1-\rho^2))$. From those starting values, I calculate $T-1$ observations for $x_t$ with the equation of the AR(1) process above. I then estimate $\hat{\rho}$ for each of the generated draws using OLS.

For a sample with $T = 144$ observations, the size of the truncated sample, the mean of the simulated AR-coefficient is smaller than its true counterpart, $\hat{\rho} = .9794$. For $T = 348$ observations, the simulated mean of the AR-coefficient $\hat{\rho} = .9849$ is closer to the true value, but still downward biased.

The downward bias of $\hat{\rho}$ leads directly to a downward bias in model-implied forecasts, $\hat{\rho}^h x_t$. In the given example, where $E(\hat{\rho}) = .9794$, the five-year ahead forecast would be

$$x_{t,h=60} = .9794^{60} \cdot x_t = .4061 \cdot x_t,$$

whereas the true forecast would imply

$$x_{t,h=60} = .99^{60} \cdot x_t = .5472 \cdot x_t.$$

In addition, the variance of changes of an h-step ahead forecast, $Var(\rho^h \Delta x_t)$, is affected by the estimation bias:

$$\begin{aligned} Var(\rho^h \Delta x_t) &= \rho^{2h}(2Var(x_t) - Cov(x_t, x_{t-1}) - Cov(x_{t-1}, x_t)) \\ &= \rho^{2h}(2Var(x_t) - 2\rho Var(x_t)) \\ &= 2\rho^{2h}(1-\rho)Var(x_t) \end{aligned}$$

Hence, in the case of the variance in changes of the five-year ahead forecast, $Var(\rho^h \Delta x_t)$ is about 23% higher than $Var(\hat{\rho}^h \Delta x_t)$.

The simulation thus illustrates the sensitivity of the model-implied variance of changes in forecasts to the small sample bias. This highlights the potential benefits of restricting the variance of changes in expectations proposed in this paper: If the variance of changes in survey expectations is a plausible target for the model-implied variance of changes, then this may improve the estimation of the factor's persistence in a small sample estimation.

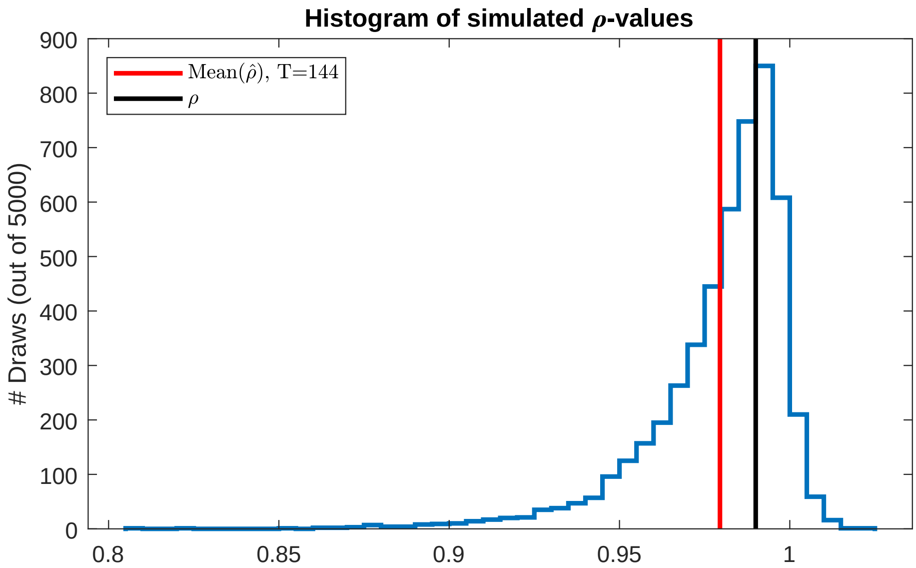

Figure 5: Histogram of the simulated small sample distribution of $\hat{\rho}$.

# B Illustration of half-life periods for different eigenvalues

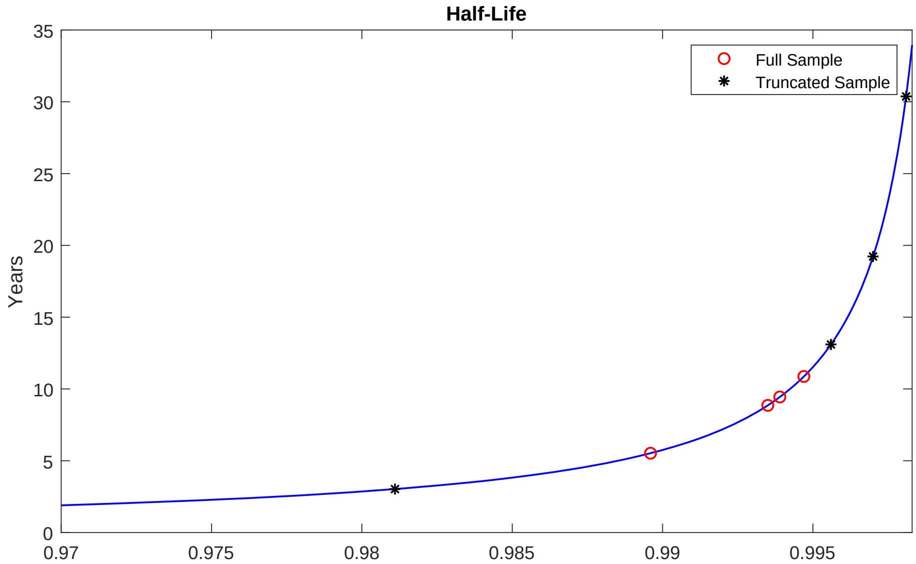


Figure 6: The figure plots the half-life for a range of eigenvalues (Equation 15). Asterisks (truncated sample estimates) and circles (full sample estimates) mark half-life values of the models reported in Table 2. Specifically, from left to right, the marks refer to the following eigenvalues: The truncated sample ACM estimation under $\mathbb{P}$, the full sample ACM estimation under $\mathbb{Q}$, the full sample Var($\Delta$)-restricted estimation under $\mathbb{P}$, the full sample ACM estimation under $\mathbb{P}$, the full sample BRW estimation under $\mathbb{Q}$, the truncated sample ACM estimation under $\mathbb{Q}$, the truncated sample Var($\Delta$)-restricted estimation under $\mathbb{P}$, and the truncated sample BRW estimation under $\mathbb{Q}$.

# C Variance of differences in survey expectations

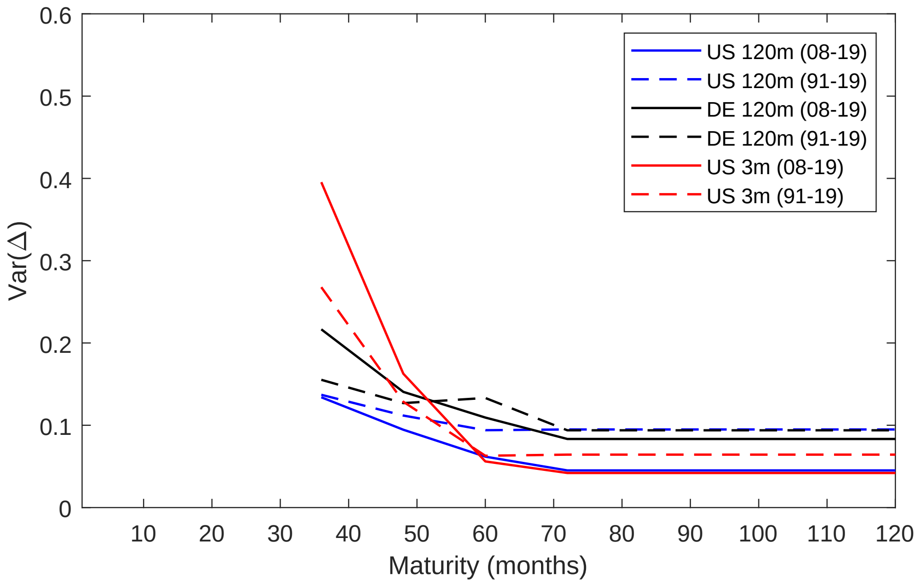


Figure 7: Variance of differences in surveys from different markets and for different forecast horizons. Source: Consensus Economics and own calculations.

# D Simulation results

A simulation indicates that the influence of the Var(Δ)-restriction on the eigenvalues would be similar in a longer sample:[14] I generate artificial state variables from the estimated parameter matrices and errors from a multivariate normal distribution, i.e. $u_t \sim \mathcal{N}(0, \Sigma)$:

$$X_t^{sim} = \tilde{\Phi} \cdot X_{t-1}^{sim} + u_t, \tag{14}$$

where $\tilde{\Phi}$ is either $\Phi^{\mathbb{Q}}$, the unrestricted $\Phi$, or the restricted $\Phi^r$. I re-calculate the regression coefficients for the VAR of the simulated states $X_t^{sim}$. The persistence of the simulated parameters is very similar to that from the original model estimates. The left-hand panel of Table 8 shows that the absolute differences of the largest eigenvalues from the simulation to the model-implied eigenvalue are very small.

| | Bund | | US Treasuries | |
|---|---|---|---|---|
| | $T = 1000$ | $T = 10000$ | $T = 1000$ | $T = 10000$ |
| $\Delta\text{eig}(\Phi^{\mathbb{Q}})$ | 0.0018 | 0.0010 | 0.0006 | 0.0000 |
| $\Delta\text{eig}(\Phi)$ | 0.0010 | 0.0010 | 0.0061 | 0.0024 |
| $\Delta\text{eig}(\Phi^r)$(restr.) | 0.0011 | 0.0011 | 0.0008 | 0.0024 |

Table 8: Differences in the eigenvalues of model-implied parameters and simulated parameters, measured by the absolute differences of the largest eigenvalues. I have simulated $T = 1000$ or $T = 10000$ observations for the state variables from model-implied coefficients and errors drawn from a multivariate normal distribution with zero mean and model-implied variances.

[14]For an overview of simulation methods, see e.g. Luetkepohl (2005).