
Article:
Golinski, Adam orcid.org/0000-0001-8603-1171 and Spencer, Peter orcid.org/0000-0002-5595-5360 (2021) Estimating the term structure with linear regressions: Getting to the roots of the problem. Journal of Financial Econometrics. pp. 960-984. ISSN 1479-8409

https://doi.org/10.1093/jjfinec/nbz031

Estimating the term structure with linear regressions: Getting to the roots of the problem*

Adam Goliński, Peter Spencer

University of York

Abstract

Linear estimators of the affine term structure model are inconsistent since they cannot reproduce the factors used in estimation. This is a serious handicap empirically, giving a worse fit than the conventional ML estimator that ensures consistency. We show that a simple self-consistent estimator can be constructed using the eigenvalue decomposition of a regression estimator. The remaining parameters of the model follow analytically. Estimates from this model are virtually indistinguishable from that of the ML estimator. We apply the method to estimate various models of U.S. Treasury yields. These exercises greatly extend the range of models that can be estimated.

JEL classification: C13, G12.

Keywords: term structure, linear regression estimators, self-consistent model, estimation methods, VAR(p) dynamics.

*Goliński: University of York, Department of Economics and Related Studies Heslington, York, Y010 5DD, United Kingdom; E-mail: adam.golinski@york.ac.uk.

Spencer: University of York, Department of Economics and Related Studies, Heslington, York, Y010 5DD, United Kingdom; E-mail: peter.spencer@york.ac.uk.

Introduction

This paper proposes a simple regression-based method for estimating the risk-neutral dynamics of the affine term structure model that is internally consistent in the sense that it can reproduce the factors used in estimation. We obtain this result using the eigenvalue decomposition of a linear estimator. The remaining parameters of the model follow analytically. Remarkably, the fit of this model is virtually indistinguishable from that of the maximum likelihood (ML) estimator.

The estimation of a no-arbitrage dynamic term structure model ($DTSM$) represents a challenging numerical problem, but in recent years significant progress has been made in tackling this (Joslin et al. (2011), Hamilton and Wu (2012), Adrian et al. (2013), Abrahams et al. (2015) and Diez de Los Rios (2015)). These new methods greatly facilitate the estimation of term structure models but have their own limitations. For instance, the method of Joslin et al. (2011) (henceforth JSZ) maximizes the likelihood of the cross-section of yields, but is slower than the new linear methods and requires the nature of the risk-neutral roots of the model to be specified (as real or complex and distinct or repeated). Moreover, the presence of multiple local optima means that good parameter starting values are necessary to increase the chances of finding the global optimum.

On the other hand, the linear estimators of Adrian et al. (2013) and Abrahams et al. (2015)[1] (henceforth $AACM$) yield instantaneous solutions that are unique, but as Joslin et al. (2013) point out, they contain redundant parameters. For example, with 3 factors, the risk-neutral response matrix that underpins the model of the cross section of yields, is defined by 3 characteristic roots (or eigenvalues), while the regression uses 9 parameters. Joslin et al. (2013) show that this means that the models are not self-consistent in the sense that they assume that the yield-based factors are observed without error, but the factors reconstructed from the yields fitted by the model differ from these observed values.

[1]This working paper version is not to be confused with the published version (Abrahams et al., 2016), which eschews the regression approach and instead employs the Maximum Likelihood method, which is internally consistent.

This theoretical inconsistency may or may not be important in practice, but is just one potential consequence of over-fitting. As in any econometric model, eliminating redundant parameters should improve general performance in fitting the data, as well as forecasting and simulation exercises, reducing the risk of misleading inferences being drawn.

The constrained estimator (DLR_C) proposed by Diez de Los Rios (2015) (henceforth DLR), uses an iterative method to remove these redundant parameters from initial regression estimates (DLR_U) of the risk-neutral dynamics. Instead, we do this in a way that preserves the characteristic roots of the DLR_U risk-neutral response matrix and thus its dynamic characteristics. This matrix determines the factor loadings in the model of bond yields. We then exploit the sequential nature of the arbitrage-free parameter solution to find a closed form expression for the risk-neutral level parameter, conditional upon the response matrix. This determines the intercepts in the yield model.

By eliminating redundant parameters, our estimator markedly outperforms linear estimators that are not self-consistent. In practice this means that the fit of the latter needs to be improved by increasing the number of factors. Our procedure preserves the risk-neutral dynamics that are embedded in the characteristic roots, while iterative procedures that ensure consistency, like the one used by DLR, are unlikely to do this. His estimator does not fit a standard set of U.S. Treasury yield data and simulated data sets based on this benchmark data as well as ours does.

We show that any regression-based estimator of the risk-neutral dynamics (such as $AACM$ or the unconstrained DLR estimator, DLR_U), can easily be modified to give a self-consistent term structure model. This method is based on four key observations, of which two are well-established and two are novel. First, JSZ showed that ML estimates of the physical factor dynamics can be provided by a stand-alone vector autoregression (VAR) of the principal components of yields and second, that the factor covariance matrix provided by this VAR is consistent and very close to the ML estimate. These techniques are extensively used in the new linear regression literature. We take this further by noting that the risk-neutral roots of the self-consistent model can be estimated as the characteristic roots

of the risk-neutral response matrix of any linear regression model. Moreover, although raw regression estimates (such as DLR_U) are under-identified, these eigenvalues, which characterize the $Q-$dynamics, are estimated with an extremely high degree of precision. They are strongly rooted in the data, helping to explain why our approach works so remarkably well. Finally, we show that given these values, the remaining parameters follow analytically. This completes the specification of the risk-neutral dynamics, allowing the coefficients of the full cross-section of yields to be estimated using well-known affine recursion relationships.

Using a standard data set for U.S. Treasury yields, we find that applying our simple self-consistent (SSC) procedure to the unconstrained DLR estimator gives results that are virtually indistinguishable from those obtained by the JSZ ML estimator. Similarly, we find that using the characteristic roots of the risk-neutral response matrix from the DLR_U regressions as starting values for the root parameters in the JSZ ML algorithm, hardly any iteration is required to get the ML values, since these are so close (see Table 2).

We check the robustness of these findings using bootstrap simulations. The simulation exercise takes the ML estimates of the JSZ model as the 'true' values and uses these to generate $5,000$ artificial yield data samples of the same length and character as the original U.S. Treasury yield data set. This exercise reveals that the $SSC(DLR_U)$ and ML methods both display negligible bias and that the former fits the simulated data almost as well as the latter. Remarkably, initiating the ML procedure using the DLR_U estimates is as effective as initiating it from the 'true' values used to generate the artificial data.

These exercises show how our new algorithm can estimate the $\mathcal{Q}-$dynamics in the standard $VAR(1)$ $DTSM$, quickly and consistently. However, the recent literature has been focussed on generalizing and refining the specification of the $\mathcal{P}-$dynamics. These can be used to represent market expectations that, subtracted from the corresponding yields, gives residuals that can be interpreted as term premia. These new time-series approaches are based on ML methods and can be very time-intensive, but we show that they can be speeded up enormously using our regression-based approach, greatly extending the range of problems that can be handled.

We demonstrate this using a generalization of the standard $VAR(1)$ $DTSM$ proposed by Joslin et al. (2013), who consider $VAR(p)$, $p \geq 1$, for the physical dynamics and $VAR(1)$ for the risk-neutral dynamics. For the three factor model, the Akaike (AIC) and Hannan-Quinn ($HQIC$) information criteria favour the $VAR(2)$, while the Bayesian information criterion (BIC), which is stricter, selects the $VAR(1)$. As usual, we find that many of the parameters of these time-series models are insignificant, leading to poorly-determined estimates of the term premium. So we apply the model selection procedure proposed by Joslin et al. (2014) to eliminate insignificant parameters and search over parameter combinations that is optimal in terms of these information criteria. For the $VAR(2)$ model we need to select a model from over two million (2^{21}) possible combinations. This would not be feasible using ML, but we can find this solution quickly using our regression-based framework. Imposing these restrictions reduces the variance of the term premium without significantly reducing the likelihood of the model.

The paper is set out as follows. The next section sets out the theoretical model of the risk-neutral dynamics and the term structure and discusses the alternative ML and linear estimation approaches. Section 2 shows how a self-consistent model of the risk-neutral dynamics can be found using any linear estimator. Section 3 compares the performance of these self-consistent regression-based estimators with that of the ML approach. In Section 4 we apply our estimation method in conjunction with generalized physical dynamics of the factors. Section 5 concludes with some observations on the the implications of these results for research on the term structure.

1 Estimating affine term structure models

Assume that the zero-coupon log bond prices, $p_{m,t}$, are affine in K observable factors $\mathbf{q}_t$:

$$-p_{m,t} = a_m + \mathbf{b}_m' \mathbf{q}_t. \tag{1}$$

We collect these in a vector $\mathbf{p}$ of log prices with J maturities that can be collected in a vector $\mathbf{m} = [m_1, m_2, \ldots, m_J]'$, where $-\mathbf{p}_t = \mathbf{a} + \mathbf{B}'\mathbf{q}_t$, $\mathbf{a} = [a_{m_1}, \ldots, a_{m_J}]'$ and $\mathbf{B} = [\mathbf{b}_{m_1}, \ldots, \mathbf{b}_{m_J}]$. Also assume that the factors follow a $VAR(1)$ process under the risk-neutral measure:

$$\mathbf{q}_{t+1} = \boldsymbol{\mu}^{\mathcal{Q}} + \boldsymbol{\Phi}^{\mathcal{Q}}\mathbf{q}_t + \mathbf{u}_{t+1}^{\mathcal{Q}} \tag{2}$$

with $\mathbf{u}_t^{\mathcal{Q}} \sim i.i.d.N^{\mathcal{Q}}(\mathbf{0}, \boldsymbol{\Sigma})$.

The core of the estimation problem is to find the parameters of this process: $\boldsymbol{\mu}^{\mathcal{Q}}$ and $\boldsymbol{\Phi}^{\mathcal{Q}}$. These are related to the price coefficients by the well-known recursions:

$$\begin{aligned} \mathbf{b}_m &= \mathbf{b}_1 + \boldsymbol{\Phi}^{\mathcal{Q}\prime}\mathbf{b}_{m-1}, & (3)\\ a_m &= a_1 + a_{m-1} + \mathbf{b}_{m-1}'\boldsymbol{\mu}^{\mathcal{Q}} - \frac{1}{2}\mathbf{b}_{m-1}'\boldsymbol{\Sigma}\mathbf{b}_{m-1}. & (4) \end{aligned}$$

1.1 Indirect Maximum Likelihood estimators of the risk-neutral dynamics

Until recently, the standard way of estimating the parameters of the risk-neutral dynamics was to embed them in the price coefficients using (3) and (4), substitute them back into (1) and, upon transformation to yields by $y_{m,t} = -p_{m,t}/m$, optimize the likelihood of observing the associated bond yields. Because yields do not exhibit an exact factor structure it is assumed that they are measured with error. Following Duffee (2011) and others we denote observed yields by $\mathbf{y}_t^o$ and add the vector of pricing errors $\mathbf{v}_t$:

$$\begin{aligned} \mathbf{y}_t^o &= \hat{\mathbf{y}}_t + \mathbf{v}_t \\ &= \mathbf{a}_{\mathbf{y}} + \mathbf{B}_{\mathbf{y}}'\mathbf{q}_t + \mathbf{v}_t, & (5) \end{aligned}$$

where $\mathbf{a}_{\mathbf{y}} \equiv \mathbf{M}^{-1}\mathbf{a}$, $\mathbf{B}_{\mathbf{y}}' \equiv \mathbf{M}^{-1}\mathbf{B}'$, $\mathbf{v}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{v}})$, and $\mathbf{M}$ is a diagonal price-yield transformation matrix with maturities $\mathbf{m}$ on the main diagonal. In particular, the short rate

equation ($r_t \equiv -p_{1,t}$) becomes:

$$r_t^o = \delta_0 + \boldsymbol{\delta}_1' \mathbf{q}_t + v_{1,t}. \tag{6}$$

Since the measurement errors are not priced, they have the same distribution under both the physical and the risk-neutral measure: $\mathbf{v}_t \sim i.i.d.N(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{v}})$.

Hamilton and Wu (2012) note that an OLS regression of the form (5) provides a benchmark for these affine models, which ignores the no-arbitrage restrictions (3) and (4). They propose a minimum χ^2 estimator that minimizes the distance between the restricted and unrestricted estimates of $\mathbf{a}_{\mathbf{y}}$ and $\mathbf{B}_{\mathbf{y}}$.

1.2 Direct regression-based estimators of the risk-neutral dynamics

In contrast, the regression-based approaches attempt to estimate the parameters of the risk-neutral dynamics $\boldsymbol{\mu}^{\mathcal{Q}}$ and $\boldsymbol{\Phi}^{\mathcal{Q}}$ directly using either the OLS regression estimates of $\mathbf{a}_{\mathbf{y}}$ and $\mathbf{B}_{\mathbf{y}}$ or those of excess *ex post* return regressions. As noted, these regression estimators cannot reproduce the pricing factors. It is usual to assume that these factors are the first K principal components (PCs) of the cross-section of yields:

$$\mathbf{q}_t \equiv \mathbf{W}' \mathbf{y}_t^o. \tag{7}$$

Substituting (5) into (7), we see that internal consistency requires:

$$\mathbf{W}'\mathbf{M}^{-1}\mathbf{a} = \mathbf{0}, \tag{8}$$

$$\mathbf{W}'\mathbf{M}^{-1}\mathbf{B}' = \mathbf{I}, \tag{9}$$

and $\mathbf{W}'\mathbf{v}_t$=$\mathbf{0}$. In other words, the factors reconstructed from the yield estimates $\mathbf{q}_t = \mathbf{W}'\hat{\mathbf{y}}_t$, must coincide with those constructed from the observed yields in (7).

This problem is a consequence of the basic problem that the regression models contain redundant parameters (Hamilton and Wu, 2012, Diez de Los Rios, 2015). The indirect ML

estimators eliminate these using alternative (but equivalent) identification schemes. For example, JSZ scheme, which we use, is discussed in the next section. Conveniently, when the factors are PCs, the weights $\mathbf{W}$ used in the construction of the factors $\mathbf{q}_t$ are provided by the eigenvectors of the eigendecomposition of the yield covariance matrix and the OLS factor loadings are the transpose of this matrix, allowing us to write the Hamilton and Wu (2012) benchmark OLS regression (5) as: $\mathbf{y}_t^o = \mathbf{c} + \mathbf{W}\mathbf{q}_t + \widehat{\mathbf{v}}_t$, where $\mathbf{W} \equiv \mathbf{B}_{\mathbf{y}}'$, $\mathbf{c}$ is a vector of constants and $\widehat{\mathbf{v}}_t$ is a vector of PC or equivalently OLS residuals.

The weights $\mathbf{W}$ can thus be used to estimate $\mathbf{B}' = \mathbf{M}\mathbf{W}$ and hence the risk-neutral parameters using the Hamilton and Wu (2012) or DLR approaches. For example, the DLR cross-section regression estimator DLR_U of the risk-neutral feedback matrix $\widehat{\mathbf{\Phi}}^{\mathcal{Q}}$ in (3):

$$\widehat{\mathbf{\Phi}}^{\mathcal{Q}} = \left(\sum_{m=m_2}^{m_J} \mathbf{b}_{m-1}\mathbf{b}_{m-1}' \right)^{-1} \left(\sum_{m=m_2}^{m_J} \mathbf{b}_{m-1}(\mathbf{b}_m' - \mathbf{b}_1') \right), \tag{10}$$

(his Eq.(9)). Our self-consistent estimator focuses on this matrix.

The model is completed by specifying the physical dynamics of the factors. In this paper we first assume that they follow a $VAR(1)$ process:

$$\mathbf{q}_{t+1} = \boldsymbol{\mu} + \mathbf{\Phi}\mathbf{q}_t + \mathbf{u}_{t+1}^{\mathcal{P}} \tag{11}$$

with $\mathbf{u}_t^{\mathcal{P}} \sim i.i.d.N^{\mathcal{P}}(\mathbf{0}, \mathbf{\Sigma})$, but it is straightforward to make an extension so that they can include other, unspanned variables (Joslin et al., 2014) or include multiple lags (Joslin et al., 2013), as we do in Section 4.

2 The simple self-consistent term structure estimator

In this section, we show how the redundant parameters of any linear regression model (such as $AACM$ or the unconstrained DLR estimates outlined in the previous section) can be eliminated using the standard eigenvalue decomposition.

2.1 Structural restrictions and the JSZ identification scheme

A convenient identification scheme was proposed by JSZ. They assume that the underlying latent factors follow the risk-neutral dynamics:

$$\mathbf{x}_t = \boldsymbol{\mu}_{\mathbf{x}}^{\mathcal{Q}} + \boldsymbol{\Phi}_{\mathbf{x}}^{\mathcal{Q}} \mathbf{x}_{t-1} + \mathbf{u}_{\mathbf{x},t}^{\mathcal{Q}}, \tag{12}$$

where $\boldsymbol{\Phi}_{\mathbf{x}}^{\mathcal{Q}}$ has an ordered Jordan form determined by K parameters only and $\mathbf{u}_{\mathbf{x},t}^{\mathcal{Q}} \sim i.i.d.N^{\mathcal{Q}}(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{x}})$. The level parameter can be identified by specifying jointly $\boldsymbol{\mu}_{\mathbf{x}}^{\mathcal{Q}} = (\mu_{\infty}^{\mathcal{Q}}, 0, \ldots, 0)'$ and $\delta_{\mathbf{x},0} = 0$.[2] The theoretical log bond prices are:

$$-\mathbf{p}_t = \mathbf{a}_{\mathbf{x}} + \mathbf{B}_{\mathbf{x}}' \mathbf{x}_t, \tag{13}$$

where $\mathbf{a}_{\mathbf{x}}$ and $\mathbf{B}_{\mathbf{x}}$ follow from recursion systems analogous to (3) and (4):

$$\begin{aligned} \mathbf{b}_{\mathbf{x},m} &= \mathbf{b}_{\mathbf{x},1} + \boldsymbol{\Phi}_{\mathbf{x}}^{\mathcal{Q}\prime} \mathbf{b}_{\mathbf{x},m-1}, & (14) \\ a_{\mathbf{x},m} &= a_{\mathbf{x},1} + a_{\mathbf{x},m-1} + \mathbf{b}_{\mathbf{x},m-1}' \boldsymbol{\mu}_{\mathbf{x}}^{\mathcal{Q}} - \frac{1}{2} \mathbf{b}_{\mathbf{x},m-1}' \boldsymbol{\Sigma}_{\mathbf{x}} \mathbf{b}_{\mathbf{x},m-1}. & (15) \end{aligned}$$

The observed yields can be expressed as:

$$\mathbf{y}_t^o = \mathbf{a}_{\mathbf{y},\mathbf{x}} + \mathbf{B}_{\mathbf{y},\mathbf{x}}' \mathbf{x}_t + \mathbf{v}_t, \tag{16}$$

where $\mathbf{a}_{\mathbf{y},\mathbf{x}} = \mathbf{M}^{-1} \mathbf{a}_{\mathbf{x}}$ and $\mathbf{B}_{\mathbf{y},\mathbf{x}}' = \mathbf{M}^{-1} \mathbf{B}_{\mathbf{x}}'$, and the short rate is given by:

$$r_t = \delta_{\mathbf{x},0} + \boldsymbol{\delta}_{\mathbf{x},1}' \mathbf{x}_t + v_{1,t}, \tag{17}$$

where $\delta_{\mathbf{x},0} = a_{\mathbf{x},1}$ and $\boldsymbol{\delta}_{\mathbf{x},1} = \mathbf{b}_{\mathbf{x},1}$.

[2] Alternatively, if the factors are stationary under the risk-neutral measure (the most persistent factor has the multiplicity one with the autoregressive coefficient $|\phi_1| < 1$), the level parameter can be identified by setting $\delta_{\mathbf{x},0} = r_{\infty}^{\mathcal{Q}}$ and $\boldsymbol{\mu}_{\mathbf{x}} = \mathbf{0}$. These two identification schemes are equivalent in this case and the parameters are related by $r_{\infty}^{\mathcal{Q}} = \mu_{\infty}^{\mathcal{Q}} / (1 - \phi_1)$.

In a single-market setting (e.g. when modelling nominal government bonds in one country), $\boldsymbol{\delta}_{\mathbf{x},1}$ is not identified under the JSZ parametrization and thus can be normalized to an arbitrary vector (usually a $K-$dimensional vector of ones, $\boldsymbol{\delta}_{\mathbf{x},1} = \mathbf{1}$). In a joint model of two or more markets with common factors, however, $\boldsymbol{\delta}_{\mathbf{x},1}$ needs to be estimated, since in general the latent factor loadings for one country affect the observable factor loading of the others.[3]

Substituting (13) into (7) and assuming that $\mathbf{W}'\mathbf{v}_t = \mathbf{0}$ allows us to switch between the latent and observable factors using:

$$\mathbf{x}_t = -(\mathbf{W}'\mathbf{B}'_{\mathbf{y},\mathbf{x}})^{-1}\mathbf{W}'\mathbf{a}_{\mathbf{y},\mathbf{x}} + (\mathbf{W}'\mathbf{B}'_{\mathbf{y},\mathbf{x}})^{-1}\mathbf{q}_t. \tag{18}$$

This implies that the relationship between the coefficients in the dynamics of the observable factors (2) and latent factors (12) is:

$$\boldsymbol{\Phi}^{\mathcal{Q}} = (\mathbf{W}'\mathbf{B}'_{\mathbf{y},\mathbf{x}})\boldsymbol{\Phi}^{\mathcal{Q}}_{\mathbf{x}}(\mathbf{W}'\mathbf{B}'_{\mathbf{y},\mathbf{x}})^{-1}, \tag{19}$$

$$\boldsymbol{\mu}^{\mathcal{Q}} = (\mathbf{W}'\mathbf{B}'_{\mathbf{y},\mathbf{x}})\boldsymbol{\mu}^{\mathcal{Q}}_{\mathbf{x}} + (\mathbf{I} - \boldsymbol{\Phi}^{\mathcal{Q}})\mathbf{W}'\mathbf{a}_{\mathbf{y},\mathbf{x}}, \tag{20}$$

and

$$\boldsymbol{\Sigma}_{\mathbf{x}} = (\mathbf{W}'\mathbf{B}'_{\mathbf{y},\mathbf{x}})^{-1}\boldsymbol{\Sigma}(\mathbf{W}'\mathbf{B}'_{\mathbf{y},\mathbf{x}})^{-1\prime}. \tag{21}$$

JSZ note that the parameters $\{\boldsymbol{\mu},\boldsymbol{\Phi}\}$ describing the $\mathcal{P}-$dynamics of the observable factors (11) are unrestricted and can be estimated by OLS regression. This regression also provides a consistent (although not efficient) estimate of the $\boldsymbol{\Sigma}$ matrix, which is involved in determination of the affine coefficients a_m. Given these parameters, we are concerned in this paper with finding self-consistent estimates of the remaining parameters, which determine the risk-neutral dynamics of the factors.

[3]More technical details and the application of the estimator to a joint two-country model can be found in the working paper version of this article, Goliński and Spencer (2019).

2.2 The simple self-consistent (SSC) estimator

Any approach that returns unconstrained estimates of the risk-neutral dynamics uses $K \times K$ free parameters in the response $\mathbf{\Phi}^{\mathcal{Q}}$ matrix instead of the K parameters in $\mathbf{\Phi}^{\mathcal{Q}}_{\mathbf{x}}$ and K parameters in $\boldsymbol{\mu}^{\mathcal{Q}}$ instead of the single level parameter in $\boldsymbol{\mu}^{\mathcal{Q}}_{\mathbf{x}}$, and thus results in a model that is over-parametrized and therefore internally inconsistent. However, the factor structure of the $\mathbf{\Phi}^{\mathcal{Q}}$ matrix (19) suggests that the underlying roots of the model can be found from the eigenvalue decomposition of a regression-based $\hat{\mathbf{\Phi}}^{\mathcal{Q}}$.[4] We can put these roots in the real Jordan form $\hat{\mathbf{\Phi}}^{\mathcal{Q}}_{\mathbf{x}}$ and replace the regression-based response $\mathbf{\Phi}^{\mathcal{Q}}$ matrix by its self-consistent counterpart corresponding to (19): $\tilde{\mathbf{\Phi}}^{\mathcal{Q}} = (\mathbf{W}'\mathbf{B}'_{\mathbf{y},\mathbf{x}})\hat{\mathbf{\Phi}}^{\mathcal{Q}}_{\mathbf{x}}(\mathbf{W}'\mathbf{B}'_{\mathbf{y},\mathbf{x}})^{-1}$, where the loadings $\mathbf{B}_{\mathbf{y},\mathbf{x}}$ follow from (14) and (16). Substituting the latent factors (18) into (16), which represents yields in terms of the latent factors, and equating this with (5), which expresses them in terms of the observed factors, allows us to find its loadings and intercepts:

$$\mathbf{B}'_{\mathbf{y}} = \mathbf{B}'_{\mathbf{y},\mathbf{x}}(\mathbf{W}'\mathbf{B}'_{\mathbf{y},\mathbf{x}})^{-1} \tag{22}$$

$$\mathbf{a}_{\mathbf{y}} = \mathbf{H}\mathbf{a}_{\mathbf{y},\mathbf{x}}, \tag{23}$$

where $\mathbf{H}$ is the idempotent matrix:

$$\mathbf{H} = \mathbf{I} - \mathbf{B}'_{\mathbf{y},\mathbf{x}}(\mathbf{W}'\mathbf{B}'_{\mathbf{y},\mathbf{x}})^{-1}\mathbf{W}' \tag{24}$$

and $\mathbf{I}$ is the identity matrix. Also, integrating the intercept recursion in (15) and substituting $\boldsymbol{\mu}^{\mathcal{Q}}_{\mathbf{x}} = (\mu^{\mathcal{Q}}_{\infty}, 0, \ldots, 0)'$ gives the closed form:

$$\mathbf{a}_{\mathbf{y},\mathbf{x}} = \mu^{\mathcal{Q}}_{\infty}\mathbf{c}_0 - \mathbf{c}_1, \tag{25}$$

[4] One of the advantages of this approach is that we do not need to test different root configurations (i.e. whether they are real/complex, distinct/repeated).

where the elements of $\mathbf{c}_0$ and $\mathbf{c}_1$ are:

$$c_{0,m} = \frac{1}{m}\sum_{j=1}^{m} b_{\mathbf{x},1,j-1}, \qquad c_{1,m} = \frac{1}{2m}\sum_{j=1}^{m} \mathbf{b}'_{\mathbf{x},j-1}\boldsymbol{\Sigma}_{\mathbf{x}}\mathbf{b}_{\mathbf{x},j-1}. \tag{26}$$

The consistent estimate of the level parameter is given by substituting (25) into (23) and equating this vector with the unrestricted yield intercepts: $\widetilde{\mathbf{a}}_{\mathbf{y}} = \overline{\mathbf{y}}^o - \mathbf{B}'_{\mathbf{y}}\overline{\mathbf{q}}$, where $\overline{\mathbf{y}}^o$ and $\overline{\mathbf{q}}$ denote vectors of sample means of yields and the observable factors, respectively, and $\mathbf{B}_{\mathbf{y}}$ are the arbitrage-free slope coefficients (22). We can then solve for the level parameter as:[5]

$$\widehat{\mu}_{\infty}^{\mathcal{Q}} = \left(\mathbf{c}_0'\mathbf{H}'\mathbf{H}\mathbf{c}_0\right)^{-1}\mathbf{c}_0'\mathbf{H}'\left(\widetilde{\mathbf{a}}_{\mathbf{y}} + \mathbf{H}\mathbf{c}_1\right). \tag{27}$$

The terms on the right hand side of (27) are evaluated conditional on the estimated roots.

To summarize, our simple self-consistent estimator, $SSC(DLR_U)$, is obtained as follows. The vector of parameters to estimate is $\boldsymbol{\Theta} = \{\mu_{\infty}^{\mathcal{Q}}, \boldsymbol{\Phi}_{\mathbf{x}}^{\mathcal{Q}}, \boldsymbol{\mu}, \boldsymbol{\Phi}, \boldsymbol{\Sigma}, \boldsymbol{\Sigma}_{\mathbf{v}}\}$. We follow JSZ, $AACM$ and others in estimating the $\mathcal{P}-$dynamics of the observable factors ($\boldsymbol{\mu}$,$\boldsymbol{\Phi}$,$\boldsymbol{\Sigma}$) in (11) by OLS. We obtain $\boldsymbol{\Phi}^{\mathcal{Q}}$ using (10). The roots of the model are the eigenvalues of $\boldsymbol{\Phi}^{\mathcal{Q}}$, which form $\boldsymbol{\Phi}_{\mathbf{x}}^{\mathcal{Q}}$. Given the estimates of $\boldsymbol{\Sigma}$ and $\boldsymbol{\Phi}_{\mathbf{x}}^{\mathcal{Q}}$, we compute the loadings $\mathbf{B}$ using (22) and find the level parameter using (27). Similarly, our simple self-consistent $AACM$ estimator, $SSC(AACM)$, is obtained by finding $\boldsymbol{\Phi}^{\mathcal{Q}}$ from excess return regressions.

3 Empirical evaluation of the SSC estimator

This section reports the results of using the basic $VAR(1)$ $DTSM$ to represent a standard data set for the U.S. Treasury market. Later sections look at extensions of the model stemming from the recent literature on the physical dynamics.

[5]The formula for $r_{\infty}^{\mathcal{Q}}$ is identical, with $\mathbf{c}_0$ equal to a vector of ones, $\mathbf{c}_0 = \mathbf{1}$.

3.1 Data

To compare the performance of this new approach with others, we first employ these on a standard monthly data set for the U.S. Treasury market, one used for example in Goliński and Spencer (2017). This is similar to the data set used by Adrian et al. (2013) and comprises eight base yields with maturities 1 month, 1, 2, 3, 5, 7, 10 and 15 years for the period January 1983 to December 2015, which followed the Volker experiment. The annual maturities come from the well-known data set constructed by Gurkaynak et al. (2007) using the Svensson (1994) parametric method, which are published by the Federal Reserve Board.[6] The short rate data comes from the Fama Treasury bills files available from the Center of Research in Security Prices. For more detailed description of the data, see Goliński and Spencer (2017).

3.2 The performance of different estimators

Table 1 shows the performance of different estimators, measured in terms of the root-mean-square error ($RMSE$) for each yield and the average across yields for 3, 4 and 5 factors. This is a natural measure of performance used for example by Adrian et al. (2013), even though their estimates are based on return regressions, perhaps making $RMSEs$ for returns more relevant. The Hamilton and Wu (2012) regression model (5) and ML estimate of the JSZ model (reported respectively as OLS and JSZ for short) constitute benchmarks for other models. The OLS regression is the best linear unbiased estimator. The ML routine searches numerically for the parameters (including $\mathbf{\Sigma}$) that maximize the likelihood function subject to the no-arbitrage constraints, providing the upper bound for the no-arbitrage approaches. These two routines optimize the likelihood function for these eight base yields directly, while the other methods that we consider estimate the price coefficients indirectly using estimates of the parameters of the risk-neutral dynamics for different sets of maturities. The information set used in these estimators is larger than in JSZ, since it includes both the eight base maturities and eight adjacent maturities. Finally, these other approaches satisfy

[6] Available at: http://www.federalreserve.gov/pubs/feds/2006/200628/200628abs.html.

the restriction that the $\mathbf{\Sigma}$ parameters appearing in the cross-section (2) and the time series dynamics (11) are the same by using the estimates from the the latter, while the ML routine uses these parameters to optimize the fit of both dimensions jointly. However, all methods considered in the table use the same observable factors, which are the first K principal components of our eight base yields.

In our comparison we include the linear estimators that do not impose self-consistency restrictions, namely the $AACM$ and DLR_U, as well as estimators that do, like the estimator proposed by Diez de Los Rios (2015), denoted DLR_C, adapted to our selection of yield maturities.[7] As noted in the introduction, the DLR routine imposes the self-consistency constraints using iterative methods. However, as opposed to the original DLR estimator, we use the OLS version of the DLR method, since we found that his GLS estimator diverges when the number of factors is larger than three. For three factors, the DLR routine results obtained by constrained OLS and constrained GLS are similar. As input for our SSC method we then use the estimates of $\mathbf{\Phi}^{\mathbb{Q}}$ from the $AACM$ and DLR_U estimators to get $SSC(AACM)$ and $SSC(DLR_U)$, respectively.

Finally, in the comparison we use the constrained estimator DLR_C as input to get $SSC(DLR_C)$. Since the DLR_C estimator is already self-consistent, this estimator returns the same estimate of the $\mathbf{\Phi}^{\mathbb{Q}}$ matrix, and thus the $\mathbf{B_y}$ factor loadings. The difference between these two methods comes solely from the method of estimation of $\boldsymbol{\mu}$; while the DLR method simultaneously uses all risk-neutral parameters to fit the model to its OLS counterpart, the SSC estimator obtains the level parameter using (27) *conditional* on the $\mathbf{\Phi}^{\mathbb{Q}}$ matrix.

Table 1 indicates that three factors are sufficient to give a plausible fit to the term structure, in line with typical bid-ask spreads in the U.S. Treasury market. However, a four or perhaps a five factor model is needed to get the performance of the JSZ model close to that of the unrestricted OLS model. The loadings on the first three principal components are shown in Figure 1. They have the shapes to the level, slope and curvature factors, as typically found in the literature.

[7] We are grateful to Antonio Diez de los Rios for providing us with his code.

The next two results are for the $AACM$ and DLR_U regressions, which do not impose the self-consistency condition. The $AACM$ algorithm gives the worst fit of all approaches for the 3 factor model, with an average $RMSE$ exceeding 12 basis points. The table shows that it requires 4 (or perhaps 5) factors to get a plausible fit, but even then it is noticeably worse than the OLS and JSZ benchmarks. The average $RMSE$ for the DLR_U method for the 3 factor model is about 8.5 basis points, which is much lower than for the $AACM$ routines. The fit obtained by the DLR_U method for the 5 factor model is, however, slightly worse than for the $AACM$ method. This is mainly due to the poor fit apparent for the $1-$year yield.

Surprisingly, the performance of the constrained linear estimator proposed by DLR is always visibly worse than the JSZ benchmark. This is particularly glaring for the model with 4 and 5 factors, where the $RMSEs$ exceed those of the JSZ benchmark by a several basis points.

The last three entries in the table, $SSC(AACM)$, $SSC(DLR_C)$ and $SSC(DLR_U)$, compare the fit of the simple self-consistent estimation methods based on the roots of the estimates of the $\mathbf{\Phi}^{\mathcal{Q}}$ matrix obtained by different methods. Imposing the self-consistency restrictions in this way always improves the fit. For instance, $SSC(AACM)$ with 3 factors reduces the average $RMSE$ from over 12 basis points to under 8 basis points. The absolute improvement in fit for a larger number of factors is smaller but is still considerable: e.g. for the 5 factor model the $RMSE$ of the $AACM$ and DLR_U estimators are reduced from about $1.6-2.0$ basis points to under 1 basis point. Remarkably, the $SSC(DLR_U)$ estimator is very close to that obtained by the JSZ method. This is impressive given the fact that the key parameters are estimated by simple linear regression, using data for a selection of DLR_U regressions rather than by optimizing the fit for the eight base yields, using all the model parameters. This reflects the observation made in the introduction that the estimates of the roots of $\mathbf{\Phi}_{\mathbf{x}}^{\mathcal{Q}}$ provided by the basic DLR_U regressions are very close to those of the ML estimates. Finally, it is worth noting that applying our SSC method to the DLR_C gives considerably better fit to the data, than the DLR_C routine, which demonstrates the

gain from estimating the level parameter upon conditioning on the fitted loadings.

3.3 Parameter estimates

While practitioners are likely to be interested in the ability of a method to fit the cross-section of bond yields, a researcher is more likely to be interested in drawing statistical inferences about the risk-neutral parameters. In this respect, direct regression-based estimates provide a useful check upon inferences drawn indirectly from the parameters embedded in bond yields. Table 2 reports the ML estimates of $\mu_\infty^{\mathcal{Q}}$ and the roots of $\mathbf{\Phi}_{\mathbf{x}}^{\mathcal{Q}}$ in the JSZ model for 3, 4 and 5 factors. We checked different root configurations and report those with the highest likelihood value. Specifically, the roots for the 3 factor models are all real, while those for the 4 and 5 factor models include a complex pair. Table 2 also reports the structural parameters implied by the DLR_C routine, as well as $SSC(AACM)$ and $SSC(DLR_U)$ methods.

It is striking how close the roots of the DLR_U estimator always are to the ML estimates. The roots obtained from the $SSC(AACM)$ estimator are generally some distance away from the ML values. Most problematic for the $SSC(AACM)$ method is the estimate of the least persistent root, which is always much too small. This is most apparent for the 4 and 5 factor models, where the estimates of the least persistent real root is negative, while the ML estimate is about 0.6.

Interestingly, while the direct regression methods, $AACM$ and DLR_U, return the same roots as JSZ, i.e. real roots of $\mathbf{\Phi}_{\mathbf{x}}^{\mathcal{Q}}$ for the 3 factor model and complex roots for 4 and 5 factors, for the 4 factor model the roots obtained by the DLR_C method are all real. This demonstrates the likely 'distortion' caused by the iterative procedure of imposing nonlinear self-consistency constraints that can drive the roots away from those in the linear regression estimates (DLR_U). Also, for the specification with 4 and 5 factors, the 2 largest roots obtained by the DLR_C method are furthest away from the ML estimates. The next section analyses the robustness of these findings.

The standard errors reported in this table come from the asymptotic variance matrix

of the parameters found using the delta method (see e.g. Greene, 2002, Theorem D.21A). Specifically, the distribution of the risk-neutral parameters $\boldsymbol{\Psi} = (\mu_\infty^{\mathcal{Q}}, \boldsymbol{\lambda}')'$, where $\boldsymbol{\lambda}$ are the risk-neutral roots, is calculated from

$$\sqrt{T}\left(\widehat{\boldsymbol{\Psi}} - \boldsymbol{\Psi}\right) \xrightarrow{d} N\left(\mathbf{0}, \boldsymbol{\Gamma}\boldsymbol{\Omega}\boldsymbol{\Gamma}'\right), \tag{28}$$

where $\boldsymbol{\Omega}$ is the variance matrix of the reduced form parameters $\mathbf{a_y}$ and $\mathbf{B_y}$ in (5) for both the set of J yields with our basic maturities and J yields with subsequent maturities, and $\boldsymbol{\Gamma}$ is the Jacobian matrix of partial derivatives calculated numerically. Since $\boldsymbol{\Sigma}$ has no effect on the roots and only a second-order effect on the level parameter, we assume that it is known with certainty in these calculations, having checked that it does not materially affect the results.

Alternatively, the standard errors could be calculated in a bootstrap exercise. Given the rapid execution of the linear regression, its applicability to bootstrap and other simulation-based problems offers further advantage over the ML methods. Throughout this paper, however, we report asymptotic standard errors for two reasons. First, it gives a direct comparison with other results reported in the related literature (Joslin et al., 2011, Bauer et al., 2012, Adrian et al., 2013, Joslin et al., 2014, Diez de Los Rios, 2015). Second, since in application the linear regression method puts emphasis on simplicity and time efficiency, for consistency we stay with the standard errors generated by an instantaneous solution of the delta method rather than the time-consuming bootstrap procedure.

As we would expect, the efficient JSZ maximum likelihood procedure generates the smallest standard errors. The self-consistent linear regression methods based on yields, DLR_C and $SSC(DLR_U)$ have slightly larger standard errors, but they are generally of the same order of magnitude as those returned by the maximum likelihood algorithm. In particular, the very low standard errors of the latter indicate that although the raw regression estimates are under-identified, their eigenvalues are strongly rooted in the data, helping to explain why our approach works as well as it does. On the other hand, the standard errors from the $SSC(AACM)$ routine are an order of magnitude larger, making some of them

statistically insignificant. This is particularly evident for the level parameter, which is poorly identified from the excess return regressions. This provides further evidence that the model parameters obtained from excess returns should be treated with caution.

3.4 Robustness

To examine the robustness of these results, we conducted a bootstrap simulation exercise that takes the ML estimates of the JSZ model in Table 2 as the 'true' values and uses these to generate $5,000$ artificial data samples of the same length and character as the original data set.

In the first stage we use the estimates of the $\mathcal{P}-$dynamics in (11) to simulate the time series of the principal components.[8] This gives the 'true factors' for each sample. We next use each set of factors to generate a cross-section of yields using (5) with $\mathbf{a}_{\mathbf{y}}^*$ and $\mathbf{B}_{\mathbf{y}}^*$ obtained from the recursions (3-4) using the 'true' JSZ parameters. Then we add the measurement errors, obtained by randomly drawing the joint residuals $\mathbf{v}_t^*$ from the 'true' JSZ model estimates using the circular stationary bootstrap proposed by Politis and Romano (1994).

We then present a hypothetical researcher with each data sample. Importantly they only see the cross-section of yields. They do not know what the true factors are, but have to back these out of the cross-section using principal components and the JSZ assumption that weighting up the 'true' residuals $\mathbf{v}_t^*$ using the sample-dependent weights $\mathbf{W}$ gives $\mathbf{W}'\mathbf{v}_t^* = \mathbf{0}$. Since this may not be true in these samples, setting the simulations up in this way allows us to test the validity of this assumption by checking the parameter estimates for the measurement error bias discussed in Section 2.1. Finally, our researcher uses these estimated factors to estimate the term structure model using the ML, DLR_C and $SSC(DLR_U)$ methods. Table 3 reports the bias and root-mean-square error ($RMSE$) of each method. It shows that the $SSC(DLR_U)$ and ML estimates display negligible bias and that their $RMSEs$ are

[8] Specifically, starting from the value $\mathbf{q}_1$ in the first period of the original sample, we use (11) to project this forward in time over a period of 896 months given the values of the previous period and adding forecast errors $\mathbf{u}_{t+1}^{\mathcal{P}}$ that are randomly selected (with replacement) from the original sample of forecast errors. We dispose of the first 500 generated observations, leaving a set of 396 observations that match the size and character of the historical sample.

close. However, the $SSC(AACM)$ estimator always has a much larger $RMSE$ than the $SSC(DLR_U)$. It also exhibits large negative bias in the least persistent real root, consistent with our observation in Section 3.3 regarding the point estimate of this parameter.

The largest bias and $RMSEs$, however, are obtained by the DLR_C estimator, which are generally an order of magnitude bigger than for other methods. Particularly notable are the large biases for the 2 largest roots for the 4 and 5 factor models, which, again, is consistent with the comparison of the point estimates in Table 2.

As mentioned in the introduction, one of advantages of the SSC method is that it does not require the researcher to specify whether the roots of the model are real or complex, distinct or repeated. This could potentially inhibit the simulation exercise, since the estimated roots might not match the 'true' roots. We found, however, that the SSC methods estimated the same roots as 'true' roots in over 99% of simulations (when calculating the bias and $RMSE$ we rejected the remaining results when the estimated roots where 'incorrect'). Also, the DLR_C estimator returned the correct roots in over 99% cases for the model with 3 and 5 factors, but for the model with 4 factors it found the correct roots in only about 70% of simulations.

3.5 Linear regression estimates as starting values in ML estimation

So far we have considered the DLR_U and $SSC(DLR_U)$ estimators as stand-alone estimation methods. However, it is well known that the likelihood function exhibits multiple local maxima and it is standard practice to use a range of starting values to identify these, making it likely (though not certain) that they include the global optimum. As suggested in the introduction, a researcher interested in the ML estimates could view the roots from the linear DLR_U method (i.e. the SSC_{DLR_U} estimates) as potentially useful starting values. To investigate this option, we used these estimates as one of the sets of parameter starting values for the ML routine in the bootstrap simulations. As another set we used the 'true' values which were used to simulate the artificial data. Finally, to represent the usual prac-

tice, we use 5 sets of random starting values.[9] The parameters reported in Table 2 are those that gave the maximum for the optimized likelihood out of these seven sets. How likely is it that any particular starting value specification returns this best fit? Although there can be no guarantee that this represents the true global optimum, this should indicate the ability of different specifications to distinguish the global from other local optima.

As Hamilton and Wu (2012) observed, we find that parametrization in terms of the level parameter $r_\infty^{\mathbb{Q}}$ frequently leads the numerical routine to stall. They showed that this problem is alleviated by parametrization in terms of $\mu_\infty^{\mathbb{Q}}$, but we go a step further with our ML algorithm by concentrating the level parameter out from the likelihood function, as shown in Section 2.2. Thus, we only estimate the K roots of the $\mathbf{\Phi}^{\mathbb{Q}}$ matrix and the $K(K+1)/2$ parameters of $\mathbf{\Sigma}$. We find that this algorithm is very robust for the model with 3 factors, always converging very close to the same maximum, irrespective of the choice of starting values. The ability of different specifications to find the largest local maximum for the 4 and 5 factor models is reported in Figure 2. This shows that with 4 and 5 factors, initiating the search from the roots of the DLR_U estimator as the starting values gives the best fit about as often as initiating the search from the 'true' values of the parameters. In both cases the convergence rate is about 99%. This contrasts starkly with the performance of random values. When using a single set of random starting values, the algorithm converges to the best fit in about 40% and 61% of simulations for the 4 and 5 factor models, respectively. Even if 5 sets of different starting values are used, this is found in only 91% of simulations (see Figure 2(a)). Interestingly, the performance of random starting values for the 5 factor model appears to be better than for the 4 factor model (see Figure 2(b)).

[9]Specifically, we randomly draw the starting values for the real roots (and the real part of the complex roots) from the range $(1-0.1K, 1)$, where K is the number of factors, and the imaginary part of the complex roots is drawn from the uniform distribution in the range $(0, 0.1)$. The starting values for $\mathbf{\Sigma}$ again come from the VAR model of the physical dynamics.

3.6 Discussion

It is perhaps not surprising to find that imposing self-consistency on regression estimators by removing redundant parameters improves their performance. But it is a surprise to find that in practice the constrained DLR estimator does not perform as well as our SSC approach. This is reflected by the $RMSEs$ of DLR_C, $SSC(DLR_U)$ and $SSC(DLR_C)$ in Table 1 and the parameter biases and $RMSEs$ shown in Table 3.

The DLR_C method is based on a linear approximation of the highly nonlinear self-consistency constraints. Indeed, several iterations are typically necessary to reach a self-consistent solution, which indicates that the nonlinearities involved are important. In other words, the convergence gradient is not optimal in the ML sense and in the iteration process the roots tend to move away from their OLS and ML counterparts. On the other hand, the roots recovered by the SSC method preserve the characteristics of the OLS risk-neutral dynamics. As originally noted by Hamilton and Wu (2014), the OLS slope coefficients are numerically very close to the no-arbitrage values estimated by their χ^2 minimization procedure (and by implication ML), so there is little efficiency loss in using our SSC approach instead. In contrast, Hamilton and Wu (2014) noted that the no-arbitrage constraints on the yield intercepts were strongly rejected in their data set.

Importantly, the SSC method estimates the level parameter conditional on the risk-neutral roots, which is similar to concentrating the parameters out of the likelihood in ML procedures. On the other hand, the DLR_C estimator constrains all parameters simultaneously, which means that, due to the iteration process, the risk-neutral level parameter $\mu_\infty^{\mathcal{Q}}$ may not be estimated efficiently. The importance of this point can be seen by comparing $RMSEs$ of DLR_C and $SSC(DLR_C)$ in Table 1. Since both estimators use the same roots, they have the same factor loadings $\mathbf{B_y}$, and therefore the difference in performance between them is solely due to the superior SSC estimate of the level parameter.

4 Modelling the physical dynamics and the term premium[10]

Thus far, we have focussed on our new algorithms for estimating the $\mathcal{Q}-$dynamics quickly and consistently. However, the recent literature has been focussed on the $\mathcal{P}-$dynamics, which are used to represent market expectations. These are especially important from a policy perspective because subtracting these expectations from the corresponding yields gives residuals that can be interpreted as term premiums. Unfortunately, because the physical dynamics are estimated from the time-series dynamics they are subject to prediction errors that are an order of magnitude larger than the measurement errors found in the cross-section, as noted by Dai and Singleton (2000), Cochrane and Piazzesi (2008) and others. They usually contain many insignificant parameters and may also be subject to small sample bias. The new time series methods have been focussed on ways of dealing with these problems. However, they can be very time-intensive when used in conjugation with ML estimates of the risk-neutral dynamics. Because our algorithms solve almost instantly for the $\mathcal{Q}-$dynamics, they complement these innovations nicely.

4.1 Generalizing the dynamic structure

Thus far, we have assumed that the dynamics are $VAR(1)$ under both the risk-neutral and the physical measures, which is the most common $DTSM$ specification in this literature. However, recently some papers, have considered generalizing the factor dynamics. For instance, Abbritti et al. (2016) proposed $VARFIMA$ real world dynamics, while Goliński and Spencer (2017) considered $ARFIMA$ risk-neutral factor dynamics. Both of these papers assumed a linear specification of the price of risk, which implies non-standard dynamics for the corresponding equivalent measure through the no-arbitrage condition. Le et al. (2010) considered a model with $VAR(1)$ dynamics under the risk-neutral measure but with a non-

[10]We are grateful to two referees of this journal who urged us to investigate more general dynamic models and recently developed methods for improving the performance of the physical dynamics.

linear specification of the price of risk, which also implies non-standard $\mathcal{P}-$dynamics. There is also an immense literature that examines the time series properties of interest rates outside the no-arbitrage framework.

These models are computationally difficult and lie well beyond the linear-in-variables framework considered here. However, given the evidence suggesting that the $VAR(1)$ dynamics are too simple, we investigate a linear generalization proposed by Joslin, Le and Singleton (2014), who consider $VAR(p)$, $p > 1$, for the physical dynamics and $VAR(1)$ for the risk-neutral dynamics. The advantage of this framework is that the dynamics are explicit under both probability measures, while the price of risk specification remains implicit and linear-in-variables. This means that the decomposition of a yield into an expectation and risk premium remains just as straightforward as it is in the basic $VAR(1)$ model. In this section we employ standard statistical selection criteria to find the best specification of the time series dynamics. We show that our linear estimation methods allow restrictions within this generalized setting to be tested much more quickly than ML does.

Thus, we now consider a model defined by (5) and (6), but replace the specification of the $\mathcal{P}-$dynamics (11) by:

$$\mathbf{q}_{t+1} = \boldsymbol{\mu} + \sum_{j=1}^{p} \boldsymbol{\Phi}_i \mathbf{q}_{t-j+1} + \mathbf{u}_{t+1}^{\mathcal{P}}. \tag{29}$$

We restrict our attention to the three factor model, $K = 3$, since a higher number of factors poses significant risk of over-fitting the model. We consider the $VAR(p)$ model with up to $p = 12$ lags. Table 4 reports (the negative of) the information criteria for the first six lags. The Akaike (AIC) and Hannan-Quinn ($HQIC$) information criteria favour the $VAR(2)$, while the Bayesian information criterion (BIC), which is stricter, selects the $VAR(1)$. The unrestricted parameters of the $VAR(1)$ process for the factors are reported in Panel A of Table 5, while Panel A of Table 6 reports the estimates for the $VAR(2)$ model.

4.2 Selecting the best model of the physical dynamics

As we have noted, many of the parameters of the price of risk and the physical dynamics are insignificant, making estimates of the term premium poorly determined. Moreover, the physical dynamics are less persistent, possibly due to the well-known small sample bias. Researchers have recently been trying to improve these estimates by imposing various parameter restrictions. We can exploit our rapid estimation procedure to implement these for models that would otherwise be too time consuming to consider.

We first apply the model selection procedure proposed by Joslin et al. (2014). Following Bauer (2018) we write the parameters of the $\mathcal{P}-$dynamics in (11) in terms of those of the $\mathcal{Q}-$dynamics in the $VAR(1)$ model (2) using:

$$\boldsymbol{\mu} = \boldsymbol{\mu}^{\mathcal{Q}} + \mathbf{l}_0, \quad \boldsymbol{\Phi} = \boldsymbol{\Phi}^{\mathcal{Q}} + \mathbf{L}_1, \tag{30}$$

and testing zero restrictions on the risk-adjustment parameters $\mathbf{l}_0$ and $\mathbf{L}_1$. Similarly, in the $VAR(2)$ model, we test the significance of $\mathbf{l}_0$, $\mathbf{L}_1$ and the extra lag parameters ($\boldsymbol{\Phi}_2 = \mathbf{L}_2$). Since the zero price-of-risk restrictions effectively push the parameters of the $\mathcal{P}-$dynamics towards those of the $\mathcal{Q}-$dynamics, which are more strongly determined, one would expect that the restricted models would be more persistent (Cochrane and Piazzesi (2008)).

Thus, for the optimal $VAR(p)$ model of the $\mathcal{P}-$dynamics selected by the information criteria, we consider all possible combinations of the zero restrictions on the price of risk parameters. Specifically, we estimate 2^{12} different model specifications for $VAR(1)$ selected by the BIC and 2^{21} specifications for $VAR(2)$ using AIC and $HQIC$.

Importantly, the restricted $\mathcal{P}-$dynamics can be estimated by the method of restricted least squares, which fits neatly into our linear-in-variables framework. Moreover the roots of the risk-neutral response matrix $\boldsymbol{\Phi}^{\mathcal{Q}}$ do not change as this model of the $\mathcal{P}-$dynamics shifts (i.e. they remain as reported in Table 2).

However, as in the basic $VAR(1)$ model, the level parameter introduces a slight complication: our solution for $\boldsymbol{\Sigma}$ and hence $\boldsymbol{\mu}^{\mathcal{Q}}$ is no longer sequential as it is with the uncon-

strained OLS model of the $\mathcal{P}-$dynamics in earlier sections. We need to find the constrained $\mathcal{P}-$dynamics and $\boldsymbol{\mu}^{\mathcal{Q}}$ simultaneously, which we do using a recursive estimation strategy. Specifically, we start with the estimate of $\boldsymbol{\Sigma}$ from the unconstrained OLS model, use this to find $\boldsymbol{\mu}^{\mathcal{Q}}$ and hence the constrained $\mathcal{P}-$dynamics. We repeat this procedure until the norm of the $\boldsymbol{\Sigma}$ matrix is smaller than 10^{-6}. This procedure is similar in spirit to that proposed by Diez de Los Rios (2015). However, in practice because $\boldsymbol{\Sigma}$ has only a second-order effect on the $\mathbf{a}$ coefficients, we find that the cross-sectional fit is virtually unaffected by the specification of the $\mathcal{P}-$dynamics and typically convergence is achieved in a single iteration.

In marked contrast, the selection procedure for the full ML estimates requires the whole model of the cross-section to be re-estimated for each combination of the restrictions. That is because as Joslin et al. (2014) note, when there are restrictions across the model of the $\mathcal{Q}-$ and $\mathcal{P}-$dynamics as in (30), the latter can no longer be estimated separately. We find that this puts up the time required for ML estimation from around a second for the unrestricted model to between 8 and 20 seconds for the restricted model. Using this procedure on a similar macro-finance model, which generalizes the $\mathcal{P}-$dynamics by introducing growth and inflation variables, Joslin et al. (2014) manage to select a model from 2^{19} possible combinations. This is unlikely to be feasible for our 3 factor $VAR(2)$ model, for which there are over two million combinations, four times as many as in their problem. However, using our linear method we are able to analyse this number of combinations in just $7-8$ hours using a single desktop computer.

We use our regression approach[11] to select the best $VAR(1)$ and $VAR(2)$ models using information criteria and report the ML estimates of the resulting models in Tables 5 and 6. For the $VAR(1)$ dynamics, the BIC selects the model with nine restrictions, leaving only the first column of the $\boldsymbol{\Phi}_1^{\mathcal{Q}}$ matrix unconstrained. For the $VAR(2)$ dynamics, the AIC

[11]The information criteria select the optimal model based on the value of the likelihood function at the maximum. As explained in Joslin et al. (2011), the model likelihood can be decomposed into a part coming from the cross-sectional fit and another from the physical dynamics of the factors. In order to avoid the risk of distorting the likelihood value by using our sequential linear estimation method for finding the cross-sectional part of the likelihood function, the cross-sectional part of the likelihood is obtained for each model from the benchmark OLS regression (5).

criterion selects a model with six restrictions. The $HQIC$, which is stricter, selects the same six restrictions as the AIC plus three additional restrictions. As expected, the restricted models are more persistent than the unrestricted ones, as indicated by the largest eigenvalue of the $\mathcal{P}-$dynamics, reported in the last column of Tables 5 and 6.

4.3 Directly restricted degree of persistence

To deal directly with the problem of small sample bias in the $\mathcal{P}-$dynamics Bauer et al. (2012) propose a simulation-based correction. We apply this technique to the $VAR(2)$ process, to compare with the results of the price of risk restrictions shown in Panels A to C of Table 6. In our data, the constraint in the adjustment procedure that keeps the process stationary is binding and the shrinkage method is necessary to keep the eigenvalues of the system smaller than unity. The estimates are reported in Panel D of Table 6. The highest eigenvalue that results is 0.9999. Since the eigenvalues are within the unit circle, the stationarity assumption is satisfied and the standard asymptotics hold.

Finally, we apply the restriction proposed by Joslin et al. (2011), that the highest eigenvalue under the physical measure is equal to the value (of 0.9976) under the risk-neutral measure, shown in Panel A of Table 2. We implement this adjustment using linear reduced rank regression. Since the process under the risk-neutral measure is more persistent than under the physical measure, this method is also designed to reduce the small sample autoregressive bias. We call this model $Eig(\mathcal{Q})$ and report the estimates in Panel E of Table 6.

4.4 Evaluating the term premium

For each specification of the $\mathcal{P}-$dynamics, we calculate the term premium, defined as the difference between the fitted yield and the average forecast of the $1-$month rate for a given maturity. We focus on the $10-$year term premium, which is the common benchmark in the literature. The diagonal entries in Table 7 show the standard deviation of each estimate of the term premium. The largest variation (1.07%) is displayed by the term premium from the

model with unrestricted $VAR(2)$ $\mathcal{P}-$dynamics, while the $VAR(1)$ BIC model, which has the most heavily restricted dynamics, exhibits the smallest variation (0.19%). The cross-correlations between these six estimates are reported in the off-diagonal entries of Table 7.

Figure 3 plots the estimates of the different term premia. The grey areas in the plot denote the NBER recession periods. Clearly, the differences between different term premium estimates can be very large at times. Consistent with the results of Table 7, the term premium from the unrestricted $VAR(2)$ model takes the most extreme values, while the BIC term premium is the smoothest in our sample.

The decomposition of the yield into a market expectation and a term premium can be used to analyse the behaviour of the U.S. Treasury market during key monetary policy episodes. For example, this has been used extensively to analyse the well-known 'conundrum' identified originally by Alan Greenspan, when in 2004 and 2005 short term policy rates rose but the 10 year forward rate continued to fall (see e.g. Backus and Wright, 2007). The term premia in these models exhibit counter-cyclical behaviour, increasing during recessions and declining between recessions. So they suggest that the premium would decline as short rates increase. However, the response in the basic $VAR(1)$ BIC model is hard to discern from the chart, as is the response in the $VAR(2)$ $HQIC$. These models suggest that the conundrum occurred because long term rate expectations fell or were flat. However, the premium in the more flexible $VAR(2)$ AIC, BRW and $Eig(\mathcal{Q})$ models fell back markedly over this period. Like the recent macro-finance term structure literature (Joslin et al., 2011), this suggests more plausibly, that the conundrum occurred because the premium fell, offsetting an increase in long term rate expectations. These results underline the importance of developing richer models of dynamic adjustment in term structure models carefully.

5 Conclusion

Until the advent of the recent regression-based estimators, estimates of the parameters $\boldsymbol{\mu}^{\mathcal{Q}}$ and $\boldsymbol{\Phi}^{\mathcal{Q}}$ of the risk-neutral dynamics were obtained by embedding them in the bond pricing coefficients and then using numerical methods to find values that optimized the fit of these equations. However, the new linear regression methods allow us to estimate this structure by OLS. Although these estimators are not internally consistent, we can estimate the response matrix $\boldsymbol{\Phi}^{\mathcal{Q}}$ by OLS and use its characteristic roots to obtain a consistent response matrix for the JSZ canonical form. The level parameter then follows directly from the affine bond pricing recursions and the remaining risk-neutral parameters are determined by the internal consistency conditions. Finally, the affine recursions determine the bond yields.

It is reassuring to find that our parameter estimates are very close to global ML estimates of the risk-neutral parameters embedded in bond yields, and that the bias revealed by our bootstrap simulations is negligibly small. It is clear that the risk-neutral dynamics are strongly rooted in the data, allowing both approaches to be used to draw statistical inferences about the risk-neutral parameters and, in the case of our estimator, to fit the term structure of bond yields by extrapolation.

Our algorithms ensure consistency by eliminating redundant parameters from the risk-neutral dynamics embedded in the cross-section of yields. Recent innovations in handling the time-series part of the $DTSM$ work by eliminating insignificant parameters from the $\mathcal{P}-$dynamics and using other constraints to improve precision. However, they are used in conjunction with ML methods for estimating the risk-neutral dynamics, which are very time intensive when used in this way. Because our algorithms solve almost instantly for the $\mathcal{Q}-$dynamics, they complement these innovations nicely.

Our linear estimator offers the researcher a useful addition to the term structure toolkit. As the exercises reported here demonstrate, its speed gives it a clear edge in model selection and simulation exercises involving a large number of models or replications. However, this estimator is likely to prove very useful in other areas as well. For example, in the working

paper version of this article (Goliński and Spencer, 2019), we show how the SSC procedure can be extended to handle multi-market term structures with common factors and use this to model the U.S. and German government bond markets jointly, representing the common factors as principal components from the joint yield covariance matrix. We find that six factors are sufficient to give a plausible fit to the term structure for both countries and that the $RMSEs$ of the $SSC(DLR_U)$ procedure is very close to the fit of an OLS benchmark. Although this fit is similar to that obtained using two separate-country three-factor models, we find that there are significant differences between the single country and joint approaches in terms of the physical dynamics. Looking forward, we intend to use the methods of Section 4 to refine the model of the physical dynamics and thus the term premia.

Regression methods have been extensively employed by central banks and other policy-makers (Bauer and Rudebusch, 2014), despite their lack of consistency and efficiency, but our estimator avoids this trade-off and allows both academic researchers and practitioners to obtain results with speed, consistency and efficiency.

References

Abbritti, M., Gil-Alana, L. A., Lovcha, Y., Moreno, A., 2016. Term structure persistence. Journal of Financial Econometrics 14, 331–352.

Abrahams, M., Adrian, T., Crump, R. K., Moench, E., 2015. Decomposing real and nominal yield curves. Staff Reports 570, Federal Reserve Bank of New York.

Abrahams, M., Adrian, T., Crump, R. K., Moench, E., Yu, R., 2016. Decomposing real and nominal yield curves. Journal of Monetary Economics pp. 182–200.

Adrian, T., Crump, R. K., Moench, E., 2013. Pricing the term structure with linear regressions. Journal of Financial Economics 10, 110–138.

Backus, D. K., Wright, J. H., 2007. Cracking the conundrum. Brookings Papers on Economic Activity 38, 293–329.

Bauer, M. D., 2018. Restrictions on risk prices in dynamic term structure models. Journal of Business & Economic Statistics 36, 196–211.

Bauer, M. D., Rudebusch, G. D., 2014. The signaling channel for Federal Reserve bond purchases. International Journal of Central Banking 10, 233–289.

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

Cochrane, J. H., Piazzesi, M., 2008. Decomposing the yield curve. AFA 2010 Atlanta Meetings Paper .

Dai, Q., Singleton, K., 2000. Specification analysis of affine term structure models. Journal of Finance 55, 415–41.

Diez de Los Rios, A., 2015. A new linear estimator for Gaussian dynamic term structure models. Journal of Business & Economic Statistics 33, 282–295.

Duffee, G. R., 2011. Information in (and not in) the term structure. Review of Financial Studies 24, 2895–2934.

Goliński, A., Spencer, P., 2017. The advantages of using excess returns to model the term structure. Journal of Financial Economics 125, 163–181.

Goliński, A., Spencer, P., 2019. Estimating the term structure with linear regressions: Getting to the roots of the problem. Discussion Paper in Economics 19/05, University of York.

Greene, W. H., 2002. Econometric Analysis. 5th Edition, Prentice Hall, London.

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

Hamilton, J. D., Wu, J. C., 2012. Identification and estimation of Gaussian affine term structure models. Journal of Econometrics 168, 315–331.

Hamilton, J. D., Wu, J. C., 2014. Testable implications of affine term structure models. Journal of Econometrics 178, 231–242.

Joslin, S., Le, A., Singleton, K. J., 2013. Gaussian macro-finance term structure models with lags. Journal of Financial Econometrics 11, 581–609.

Joslin, S., Priebsch, M., Singleton, K. J., 2014. Risk premiums in dynamic term structure models with unspanned macro risks. Journal of Finance 69, 1197–1233.

Joslin, S., Singleton, K. J., Zhu, H., 2011. A new perspective on Gaussian dynamic term structure models. Review of Financial Studies 24, 926–970.

Le, A., Singleton, K. J., Dai, Q., 2010. Discrete-time affine$^{\mathbb{Q}}$ term structure models with generalized market prices of risk. Review of Financial Studies 23, 2184–2227.

Politis, D. N., Romano, J. P., 1994. The stationary bootstrap. Journal of the American Statistical Association 89, 1303–1313.

Svensson, L. E. O., 1994. Estimating and interpreting forward interest rates; Sweden 1992-1994. IMF working papers 94/114, International Monetary Fund.

Tables

	Root-mean-square error (bp)								
yields:	1m	1y	2y	3y	5y	7y	10y	15y	Av.RMSE
				Panel A: $K = 3$					
OLS	2.5	11.2	3.0	6.6	8.4	5.8	4.0	9.9	6.43
JSZ	2.6	12.0	3.7	7.4	8.5	6.1	5.6	10.2	7.01
$AACM$	2.5	20.4	6.4	7.9	14.4	15.2	10.6	20.2	12.20
DLR_U	2.5	17.6	7.0	7.3	8.9	7.5	7.6	10.0	8.55
DLR_C	2.6	11.7	4.9	7.3	9.4	8.3	6.5	13.6	8.05
$SSC(AACM)$	2.7	13.1	4.3	8.3	8.8	6.7	6.4	11.0	7.65
$SSC(DLR_C)$	2.6	12.4	3.9	7.7	8.7	6.1	5.9	10.1	7.17
$SSC(DLR_U)$	2.6	12.4	4.0	7.7	8.7	6.1	5.8	10.1	7.16
				Panel B: $K = 4$					
OLS	0.3	3.1	3.0	3.1	1.4	3.4	3.4	4.0	2.71
JSZ	0.3	3.1	3.0	3.1	1.5	3.4	3.4	4.1	2.72
$AACM$	0.3	5.0	3.8	3.5	1.6	3.8	4.0	4.2	3.28
DLR_U	0.3	4.5	3.3	3.2	1.5	3.4	3.6	4.1	2.99
DLR_C	0.5	6.6	5.1	7.5	1.6	7.9	9.9	11.1	6.29
$SSC(AACM)$	0.3	3.4	3.3	3.2	1.7	3.5	3.6	4.2	2.89
$SSC(DLR_C)$	0.6	4.3	3.7	4.5	1.5	4.8	5.7	6.4	3.90
$SSC(DLR_U)$	0.3	3.1	3.0	3.1	1.5	3.4	3.4	4.1	2.72
				Panel C: $K = 5$					
OLS	0.1	0.8	1.3	0.3	1.3	0.4	1.7	0.8	0.82
JSZ	0.1	0.8	1.4	0.3	1.3	0.5	1.7	0.8	0.86
$AACM$	0.1	3.2	2.6	1.4	1.9	0.8	2.2	1.0	1.64
DLR_U	0.1	6.5	2.8	1.6	1.5	0.5	1.7	0.9	1.96
DLR_C	0.3	3.6	5.1	2.2	6.3	3.3	9.8	4.3	4.34
$SSC(AACM)$	0.1	0.9	1.6	0.4	1.5	0.5	1.9	0.9	0.97
$SSC(DLR_C)$	0.2	2.1	3.2	1.2	3.7	1.6	5.2	2.3	2.42
$SSC(DLR_U)$	0.1	0.8	1.5	0.6	1.4	0.6	1.7	0.8	0.92

Table 1. Goodness of fit.
The table reports the fit of the model for the Treasury yields, as measured by the root-mean-square error of the measurement errors. The model is estimated with 3, 4 and 5 principal components of yields. The reported values are in basis points. The sample period is January 1983 to December 2015.

	$\mu_\infty^{\mathcal{Q}} \times 100$	λ_1	λ_2	λ_3	λ_4	$(\lambda_{real}$	$\pm\, i\lambda_{imag})$
			Panel A: $K = 3$				
JSZ	0.0403 0.0005	0.9971 0.0001	0.9714 0.0006	0.7537 0.0098			
$SSC(AACM)$	0.0434 0.2871	0.9958 0.0409	0.9779 0.0654	0.1657 0.2986			
$SSC(DLR_U)$	0.0338 0.0019	0.9976 0.0003	0.9686 0.0021	0.7862 0.0322			
DLR_C	0.0391 0.0016	0.9975 0.0004	0.9677 0.0026	0.7711 0.0301			
			Panel B: $K = 4$				
JSZ	0.0143 0.0004	0.9998 0.0001	0.5930 0.0250			0.9664 0.0008	0.0194 0.0008
$SSC(AACM)$	0.0204 0.0631	0.9991 0.0089	−0.0666 0.2562			0.9656 0.0434	0.0221 0.0563
$SSC(DLR_U)$	0.0136 0.0015	0.9998 0.0002	0.6103 0.0630			0.9655 0.0020	0.0188 0.0019
DLR_C	0.0199 0.0089	0.9958 0.0040	0.9791 0.0153	0.9472 0.0217	0.6163 0.0660		
			Panel B: $K = 5$				
JSZ	0.0129 0.0003	0.9998 0.0001	0.9621 0.0052	0.6020 0.0186		0.9728 0.0019	0.0184 0.0011
$SSC(AACM)$	−0.0006 0.0241	1.0012 0.0020	0.9741 0.0932	−0.0656 0.2446		0.9625 0.0286	0.0233 0.0692
$SSC(DLR_U)$	0.0125 0.0022	1.0003 0.0003	0.9528 0.0066	0.7100 0.0395		0.9789 0.0010	0.0211 0.0018
DLR_C	0.0092 0.0010	1.0041 0.0007	0.8691 0.0181	0.7249 0.0332		0.9878 0.0011	0.0282 0.0008

Table 2. Estimates of the risk-neutral parameters.
The estimates of risk-neutral dynamics are obtained by different methods for 3, 4 and 5 factor JSZ models.

		$\mu_\infty^{\mathbb{Q}} \times 100$	λ_1	λ_2	λ_3	λ_4	λ_{imag}	$\pm$	xi
				Panel A: $K = 3$					
JSZ	Bias	0.2521	-0.0001	0.0023	-0.0015				
	RMSE	0.2634	0.0003	0.0028	0.0244				
$SSC(AACM)$	Bias	0.6978	-0.0012	0.0034	-0.0747				
	RMSE	0.7329	0.0016	0.0050	0.0822				
$SSC(DLR_U)$	Bias	0.6094	0.0001	-0.0012	0.0133				
	RMSE	0.6341	0.0004	0.0022	0.0371				
DLR_C	Bias	0.5710	0.0002	-0.0022	0.0095				
	RMSE	0.5944	0.0007	0.0036	0.0353				
				Panel B: $K = 4$					
JSZ	Bias	0.2480	0.0002	0.0064			0.0014	-0.0027	
	RMSE	0.2563	0.0003	0.1057			0.0022	0.0040	
$SSC(AACM)$	Bias	0.5469	0.0000	-0.1084			0.0005	0.0013	
	RMSE	0.5696	0.0005	0.1204			0.0034	0.0037	
$SSC(DLR_U)$	Bias	0.5561	0.0000	0.0010			-0.0005	-0.0006	
	RMSE	0.5805	0.0002	0.0562			0.0022	0.0027	
DLR_C	Bias	0.8418	-0.0077	-0.0193			0.0031	-0.0024	
	RMSE	6.1602	0.0240	0.0674			0.0082	0.0077	
				Panel B: $K = 5$					
JSZ	Bias	0.4522	0.0004	-0.0011	0.0291		-0.0001	-0.0031	
	RMSE	0.4748	0.0007	0.0098	0.0346		0.0039	0.0051	
$SSC(AACM)$	Bias	0.7212	-0.0016	0.0092	-0.1070		-0.0055	0.0049	
	RMSE	0.7701	0.0318	0.0170	0.1229		0.0498	0.0082	
$SSC(DLR_U)$	Bias	0.6570	0.0002	-0.0051	0.0441		0.0025	0.0014	
	RMSE	0.6864	0.0004	0.0099	0.0548		0.0035	0.0028	
DLR_C	Bias	0.4024	0.0033	-0.0569	0.0744		0.0127	0.0070	
	RMSE	0.4258	0.0035	0.0584	0.0800		0.0128	0.0071	

Table 3. Bias and RMSE for different estimation methods.
The statistics are calculated for 3, 4 and 5 factor JSZ model based on $5,000$ simulation bootstrap assuming the JSZ estimates are the 'true' values. The simulated model for 3 factors was based on 3 real roots, while for 4 and 5 factors included a complex root.

Lags (k):	1	2	3	4	5	6
$-BIC$	**32.7973**	32.7744	32.6677	32.5754	32.4667	32.3538
$-AIC$	32.8899	**32.9596**	32.9455	32.9458	32.9296	32.9094
$-HQIC$	32.8531	**32.8861**	32.8353	32.7988	32.7460	32.6890

Table 4. VAR lags selection for physical factor dynamics.
The table shows the negative of Bayesian, Akaike and Hannan-Quinn information criteria. The bold font denotes the highest value for each criterion. The sample period is January 1983 to December 2015.

$\boldsymbol{\mu}$	$\boldsymbol{\Phi}_1$			$\mathcal{P}-$eigenvalues
Panel A: VAR(1) (unrestricted)				
0.0023 0.0014	0.9936 0.0048	0.0008 0.0295	−0.2055 0.0880	0.9933
−0.0019 0.0007	−0.0014 0.0024	0.9483 0.0147	0.3683 0.0439	0.9678
0.0030 0.0007	0.0009 0.0021	0.0186 0.0133	0.6191 0.0397	0.6000
Panel B: Var(1) BIC (9 restrictions)				
−0.0014 0.0004	1.0043 0.0026	**0.0617** 0.0075	**−0.2112** 0.0398	0.9995
−0.0008 0.0002	−0.0075 0.0016	**0.9754** 0.0047	**0.2944** 0.0295	0.9709
0.0013 0.0001	0.0071 0.0012	**−0.0013** 0.0030	**0.7342** 0.0066	0.7436

Table 5. MLE Estimates for different $VAR(1)$ specifications of the physical dynamics. Panel A shows the unrestricted estimates of the $VAR(1)$ model. Panel B shows the estimates of the model selected by the BIC model. The bold font denotes restricted parameters. The sample period is January 1983 to December 2015.

$\boldsymbol{\mu}$	$\boldsymbol{\Phi}_1$			$\boldsymbol{\Phi}_2$			$\mathcal{P}-$eigenvalues
Panel A: VAR(2) (unrestricted)							
0.0012	1.1334	−0.1155	−0.3231	−0.1395	0.1067	0.2585	0.9938
0.0015	0.0524	0.1264	0.1358	0.0523	0.1247	0.1467	
−0.0017	−0.0295	0.9811	0.4140	0.0284	−0.0294	−0.0789	0.9658
0.0007	0.0264	0.0637	0.0684	0.0264	0.0628	0.0739	
0.0030	−0.0640	0.1607	0.6285	0.0652	−0.1460	−0.0113	0.6033
0.0007	0.0235	0.0567	0.0609	0.0234	0.0559	0.0658	
Panel B: VAR(2) AIC (6 restrictions)							
−0.0014	1.1388	0.0201	**−0.2110**	−0.1400	**0**	0.2449	0.9995
0.0004	0.0478	0.0248	0.0399	0.0477	0.0000	0.0758	
−0.0023	**−0.0170**	**0.9755**	0.4259	0.0169	−0.0192	−0.0650	0.9695
0.0007	0.0018	0.0047	0.0509	0.0029	0.0152	0.0473	
0.0031	−0.0688	0.1593	0.6153	0.0679	−0.1458	**0**	0.6015
0.0006	0.0186	0.0371	0.0395	0.0186	0.0376	0.0000	
Panel C: VAR(2) HQIC (9 restrictions)							
−0.0014	1.1312	**0.0616**	**−0.2109**	−0.1361	**0**	0.1933	0.9963
0.0004	0.0475	0.0075	0.0399	0.0476	0.0000	0.0632	
−0.0028	**−0.0170**	**0.9755**	0.3769	0.0162	**0**	**0**	0.9801
0.0006	0.0018	0.0047	0.0424	0.0030	0.0000	0.0000	
0.0033	−0.0674	0.1426	0.6157	0.0685	−0.1367	**0**	0.6098
0.0006	0.0187	0.0345	0.0393	0.0187	0.0344	0.0000	
Panel D: VAR(2) BRW							
0.0004	1.1291	−0.1098	−0.3279	−0.1291	0.1022	0.2545	0.9999
0.0015	0.0520	0.1255	0.1349	0.0519	0.1239	0.1457	
−0.0020	−0.0300	0.9909	0.4149	0.0297	−0.0301	−0.0872	0.9778
0.0007	0.0262	0.0631	0.0678	0.0261	0.0623	0.0733	
0.0029	−0.0653	0.1612	0.6363	0.0657	−0.1435	−0.0026	0.6170
0.0007	0.0233	0.0562	0.0604	0.0233	0.0554	0.0652	
Panel E: VAR(2) Eig($\mathcal{Q}$) (1 restriction)							
0.0007	1.1360	−0.1160	−0.3223	−0.1388	0.1071	0.2617	0.9971
0.0015	0.0520	0.1253	0.1347	0.0519	0.1237	0.1455	
−0.0018	−0.0292	0.9811	0.4141	0.0284	−0.0294	−0.0784	0.9658
0.0007	0.0262	0.0631	0.0678	0.0261	0.0623	0.0733	
0.0031	−0.0643	0.1608	0.6284	0.0652	−0.1461	−0.0117	0.6027
0.0007	0.0233	0.0562	0.0603	0.0232	0.0554	0.0652	

Table 6. MLE Estimates for different $VAR(2)$ specifications of the physical dynamics. Panel A shows the unrestricted estimates of the $VAR(2)$ model. Panel B and Panel C report the estimates of the models selected by AIC and $HQIC$, respectively. Panel D reports the $VAR(2)$ estimates with the Bauer et al. (2012). Panel E shows the estimates for a model that restricts the largest eigenvalue of the physical dynamics to be the same as the one for the risk-neutral dynamics shown in Panel A of Table 2. The bold font denotes constrained parameters. The sample period is January 1983 to December 2015.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) VAR(1) unrestr.	1.0372						
(2) VAR(2) unrestr.	0.9903	1.0658					
(3) VAR(1) BIC	0.2943	0.1658	0.1905				
(4) VAR(2) AIC	0.8387	0.7701	0.7274	0.5650			
(5) VAR(2) HQIC	0.8298	0.8959	-0.2765	0.4461	0.5152		
(6) VAR(2) BRW	0.7267	0.6345	0.8626	0.9703	0.2426	0.9395	
(7) VAR(2) Eig($\mathcal{Q}$)	0.9670	0.9294	0.5147	0.9469	0.6783	0.8749	0.8590

Table 7. Standard deviation and correlation of different risk premia.
The standard deviation of the term premium estimates is on the diagonal. The lower-triangular part of the matrix shows the correlations between particular term premium estimates. The sample period is January 1983 to December 2015.

Figures

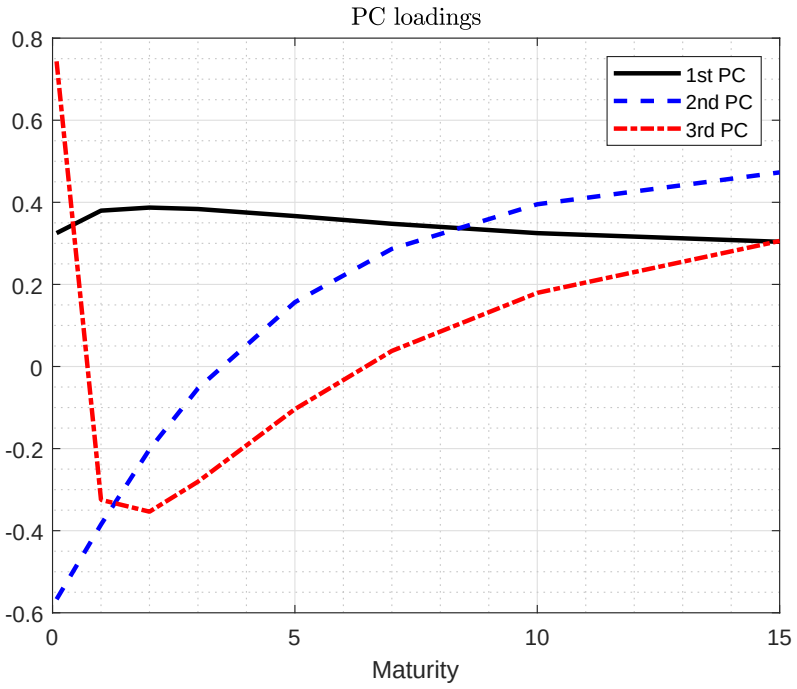
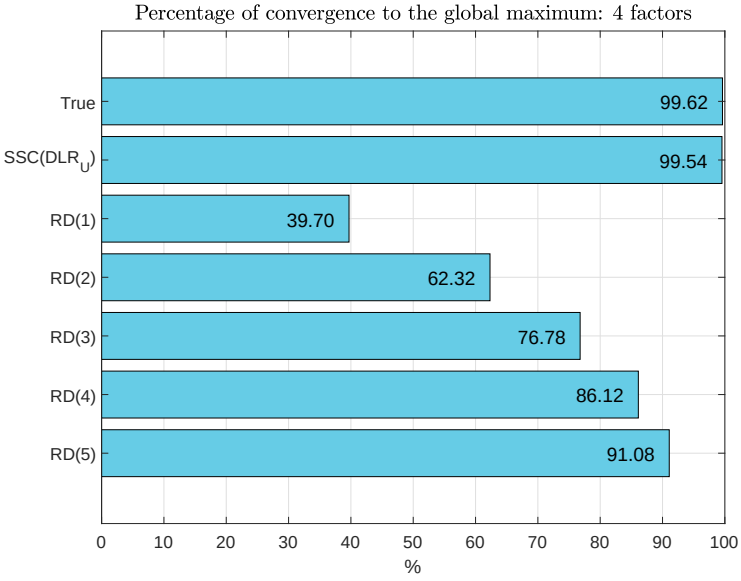
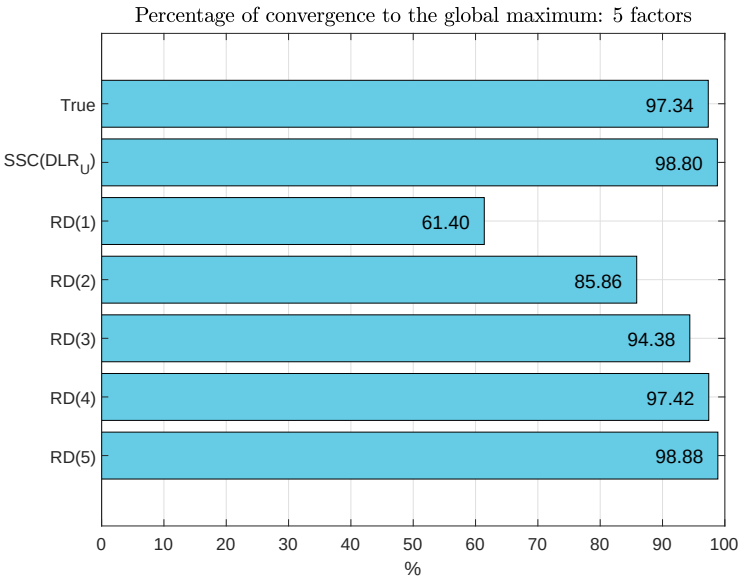


Figure 1. Principal components loadings. The figure shows the loadings on the first three principal components. The loadings are are obtained as the eigenvectors corresponding to the first three largest eigenvalues of the covariance matrix of yields.



(a)



(b)

Figure 2. Finding the global maximum likelihood. The figure shows the percentage of simulations converging to the global maximum when initiating the numerical MLE routine from different starting values for the 4 and 5 factor model in Panel (a) and Panel (b), respectively. The starting values are the values used to generate the simulated yield samples ('True'), the roots of the $\mathbf{\Phi}^{\mathcal{Q}}$ matrix estimated by the DLR_U estimator, $SSC(DLR_U)$, and from 1 to 5 different random starting values ('RD'). The global maximum is assumed to be the optimum with the highest likelihood values found from any of these starting values.

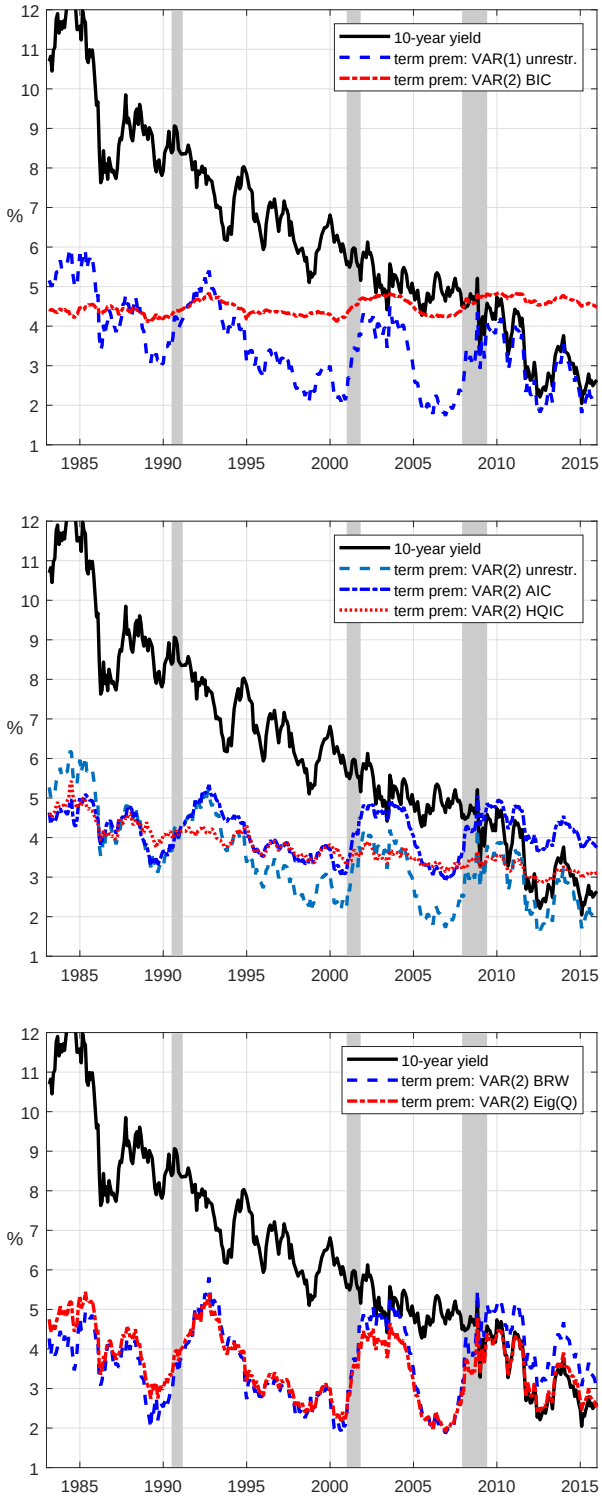


Figure 3. U.S. 10-year term premium. This figure shows 10−year yield and the 10−year term premium implied by different specifications of the $\mathcal{P}-$dynamics. The grey areas denote the NBER recession periods. The time series models are as specified in Tables 5 and 6.