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

Adam Goliński, Peter Spencer

University of York

January 3, 2017

Abstract

Linear estimators of the risk-neutral dynamics of the affine term structure model are inconsistent in the sense that they cannot reproduce the factors used in estimation. This is a serious handicap empirically, giving a worse fit than a conventional maximum likelihood estimator that ensures consistency. We show that a simple self-consistent estimator of the risk-neutral dynamics can be constructed using the eigenvalue decomposition of a regression estimator and that the remaining parameters of the model follow analytically given regression estimates of the real-world factor dynamics. The fit of this model is virtually indistinguishable from that of the maximum likelihood estimator. We apply the method to estimate the model of the U.S. Treasury yield curve and a joint model of the U.S. and German yield curves.

JEL classification: G12, C13.

Keywords: term structure, linear regression estimators, self-consistent model, estimation methods, two-country model.

Disclosure statement

I have nothing to disclose.

Adam Goliński

Disclosure statement

I am a consultant at the Bank of England.

Peter Spencer

This paper proposes a simple regression-based method of 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 construct this 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 term structure model represents a challenging numerical problem, but in recent years significant progress has been made in tackling this. Both Joslin, Singleton, and Zhu (2011) (henceforth JSZ) and Hamilton and Wu (2012) suggest ways of reducing the dimensionality of the non-linear parameter space that requires numerical search. Adrian, Crump, and Moench (2013) (henceforth ACM), Abrahams, Adrian, Crump, Moench, and Yu (2016) (henceforth $AACMY$) and Diez de Los Rios (2015) (henceforth DLR) propose linear estimators that completely eliminate the numerical search.

These new methods greatly facilitate the estimation of term structure models but have their own limitations. For instance, the JSZ method maximizes the likelihood of the cross-section of yields, but is much 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 regression-based models of ACM and $AACMY$ are very quick, but as Joslin, Le, and Singleton (2013) point out, they are over-parameterized and internally inconsistent. As we show, regression methods that do not impose the self-consistency restrictions fit the cross-section notably worse than methods that do, such as JSZ, which means that in practice, the fit needs to be improved by increasing the number of factors. The DLR estimator uses an iterative method to impose internal consistency on a regression estimate, but it is applicable only when a range of equally-spaced maturities and time observations is used in estimation. This constraint forces the use of a relatively large number of short yields, which are subject to serious measurement error and other problems. Reflecting this, we find that if we follow

ACM and $AACMY$ and omit short rates from the model used to estimate the risk-neutral dynamics, this gives a better fit for the full cross-section of yields.

This paper shows that any regression-based estimator of the risk-neutral dynamics 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 the other two are novel. First, JSZ showed that the ML estimates of the physical factor dynamics are provided by a vector autoregression (VAR) 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 from the eigenvalues of the risk-neutral response matrix of any linear regression model. Finally, we show that given these estimates, the remaining parameter, which determines the level of the yield structure follows analytically. This completes the specification of the risk-neutral dynamics, allowing the coefficients of the full cross-section of yields to be estimated using the well-known affine recursion relationships.

Although the roots of the model can be obtained from any linear estimate of the risk-neutral dynamics (such as $AACMY$), we develop a novel estimator based on the no-arbitrage relationship between spot and forward prices, which simplifies the estimation procedure and avoids various parameter biases. The eigenvalues of the risk-neutral response matrix from the spot-forward (SF) regressions can be used as starting values for the root parameters in the JSZ ML algorithm, combining the best features of the linear and JSZ approaches. However, we find that in the case of a standard data set for the U.S. Treasury market, hardly any iteration is required to get the ML values, since these are very close (see Table II). It follows that the cross-sectional fit of this simple but consistent regression-based estimator $SC(SF)$ is virtually indistinguishable from that obtained by the JSZ ML estimator.

We check the robustness of these findings using bootstrap simulations as well as other data sets. This exercise takes the ML estimates of the JSZ model in Table II as the 'true' values and uses these to generate $5,000$ artificial yield data samples of the same length and character as the original Treasury yield data set. We then present a hypothetical researcher with each data sample. Importantly they do not know what the true factors are, but follow the common procedure of backing them out of the cross-section of yields as principal components, assuming that these are measured without error. They then use these estimated factors to estimate the JSZ model using the $SC(SF)$ and ML methods. Several different starting values for the latter are tested and the best result in terms of fit is used as the ML estimate. This exercise reveals that the $SC(SF)$ and ML methods both display negligible bias and that the former fits the data almost as well as the latter. Remarkably, initiating the ML procedure using the SF estimates is as effective as initiating it from the 'true' values used to generate the artificial data.

Finally, we show how the $SC(SF)$ 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. We represent the common factors as principal components from the joint yield covariance matrix and thus avoid the usual division into global and local factors, letting the data speak without imposing any such division. We find that six factors are sufficient to give a plausible fit to the term structure for both countries, in line with typical bid-ask spreads. We find that the root-mean-square error of the $AACMY$ method is on average about double of that given by the $SC(SF)$ procedure, which is very close to the OLS fit.[1]

Our analysis raises doubts about the presence of pure global factors since we find that the fit of the joint six factor model is similar to that of two separate three factor country models. Indeed, the joint model factors simply mimic those of the single country models. We conclude that there is very little contemporaneous interaction between the U.S. and German markets that is not picked up by single country models. Nevertheless, a related paper using these data by Meldrum, Raczko, and Spencer (2016) finds strong evidence of unspanned (i.e. delayed) spillovers between these markets. Reflecting this, we find that there are significant differences between the single country and joint approaches in terms

[1]Naturally, under the null that the underlying term structure model is correctly specified, all estimation methods are *consistent* in the econometric sense that they are asymptotically efficient. Understanding this, with some abuse of nomenclature, we will refer to consistency of a model or method as a feature that allows from the estimated parameters to reproduce the factors used in estimation.

of the physical dynamics, which largely determine the way these models decompose yields into components representing interest rate expectations and the risk premium. Although, pooling the data for the U.S. and Germany seems to make little qualitative difference to the U.S. decomposition, it does give a more plausible decomposition of the German curve.

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 shows how it can be made operational in the presence of pricing errors using standard assumptions. Section II shows how the model of the risk-neutral dynamics can be estimated directly through spot-forward regression relationships assuming no-arbitrage. Section III shows how the roots underpinning the risk-neutral dynamics and the associated level parameter can be found using any such linear estimator. Section IV compares the performance of these regression-based estimators with that of the ML approach. Section V concludes with some observations on the the implications of these results for research on the term structure.

I. The affine term structure model

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

$$-\mathbf{p}_t = \mathbf{a} + \mathbf{B}'\mathbf{q}_t, \tag{1}$$

where $\mathbf{p}$ is a vector of log prices with J maturities that can be collected in a vector $\mathbf{m} = [m_1, m_2, \ldots, m_J]'$. Assume that the factors follow a $VAR(1)$ process under the risk-neutral measure:

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

with $\mathbf{u}_t^{\mathbb{Q}} \sim i.i.d.N^{\mathbb{Q}}(\mathbf{0}, \Sigma)$. The affine coefficients $\mathbf{a} = [a_{m_1}, \ldots, a_{m_J}]'$ and $\mathbf{B} = [\mathbf{b}_{m_1}, \ldots, \mathbf{b}_{m_J}]$ are given by the well-known recursions:

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

Naturally, yields do not exhibit an exact factor structure, but instead it is assumed that they are measured with error. Following Duffee (2011) and others we denote observed yields by $\mathbf{y}_t^o$ and assume additive pricing errors:

$$\mathbf{y}_t^o = \mathbf{a}_\mathbf{y} + \mathbf{B}_\mathbf{y}'\mathbf{q}_t + \mathbf{v}_t, \tag{5}$$

where $\mathbf{a}_\mathbf{y} \equiv \mathbf{M}^{-1}\mathbf{a}$, $\mathbf{B}_\mathbf{y}' \equiv \mathbf{M}^{-1}\mathbf{B}'$ and $\mathbf{v}_t \sim N(\mathbf{0}, \mathbf{\Sigma}_\mathbf{v})$, 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:

$$r_t^o = \delta_0 + \delta_1'\mathbf{q}_t + v_{1,t}. \tag{6}$$

Since the measurement errors are not priced, they have the same distribution both under the physical and under the risk-neutral measure. Moreover, if the K factors are linear combinations of the J yields, the matrix $\mathbf{\Sigma}_\mathbf{v}$ has rank $J - K$. This means that K linear combinations of the errors are zero and the corresponding combinations of yields are observed without error: i.e. $\mathbf{W}'\mathbf{v} = \mathbf{0}$for some constant $J \times K$ matrix $\mathbf{W}$.

Following JSZ we make this theoretical model operational by assuming that the factors are the first K principal components (PCs) of the cross-section of yields (or equivalently, linear transforms of these PCs):

$$\begin{aligned} \mathbf{q}_t &\equiv \mathbf{W}'\mathbf{y}_t \\ &= -\mathbf{W}'\mathbf{M}^{-1}\mathbf{p}_t. & (7) \end{aligned}$$

We follow JSZ and assume these factors are noiseless, i.e. $\mathbf{W}'\mathbf{v} = \mathbf{0}$. If this assumption is invalid, estimates of the the model parameters will exhibit measurement error bias as we show in Section III.A.

An OLS regression of the form (5) provides an unrestricted benchmark for affine models. However, when the factors are PCs, the weights $\mathbf{W}$ used in the construction of the factors (7) 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 this regression as:

$$\mathbf{y}_t^o = \mathbf{c} + \mathbf{W}\mathbf{q}_t + \widehat{\mathbf{v}}_t, \tag{8}$$

where $\mathbf{c}$ is a vector of constants and $\widehat{\mathbf{v}}_t$ is a vector of PC or equivalently OLS residuals.

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

$$\mathbf{q}_{t+1} = \mu^{\mathcal{P}} + \mathbf{\Phi}^{\mathcal{P}}\mathbf{q}_t + \mathbf{u}_{t+1}^{\mathcal{P}} \tag{9}$$

with $\mathbf{u}_t^{\mathcal{P}} \sim i.i.d.N^{\mathcal{P}}(\mathbf{0}, \Sigma)$, but it is straightforward to make an extension so that they can include other, unspanned variables (Joslin, Priebsch, and Singleton, 2014) or include multiple lags (Joslin, Le, and Singleton, 2013).

II. The Spot-Forward (SF) regression estimator

As noted in the introduction, there are several linear estimators of the risk-neutral dynamics in the recent literature, notably the multi-step procedure of $AACMY$, based on excess returns. In this section we propose a single-step estimator of the risk-neutral dynamics, based on spot-forward (henceforth SF) regressions. Appendix A compares the two estimators.

In the absence of arbitrage, forward prices are risk-neutral expectations of future spot prices:

$$P_{m,t}^f \equiv E_t^Q[P_{m,t+1}] = e^{r_t}P_{m+1,t}, \tag{10}$$

where r_t is the one-period interest rate. Denote the log forward price by $p^f_{m,t} \equiv \log P^f_{m,t}$, which under the log normal assumption is:

$$p^f_{m,t} = E^Q_t[p_{m,t+1}] + \frac{1}{2}Var[p_{m,t+1}]. \quad (11)$$

It follows that the corresponding forward yield, $f_{m,t} = -p^f_{m,t}/m$, is given by:

$$f_{m,t} = \frac{1}{m}a_m + \frac{1}{m}\mathbf{b}'_m\mu - \frac{1}{2m}\mathbf{b}'_m\mathbf{\Sigma}\mathbf{b}_m + \frac{1}{m}\mathbf{b}'_m\mathbf{\Phi}\mathbf{q}_t, \quad (12)$$

or in a vector notation:

$$\mathbf{f}_t = \mathbf{a_f} + \mathbf{B}'_\mathbf{f}\mathbf{q}_t, \quad (13)$$

where: $\mathbf{a_f} = \mathbf{a_y} + \mathbf{B}'_\mathbf{y}\mu - \frac{1}{2}\mathbf{M}diag\{\mathbf{B}'_\mathbf{y}\mathbf{\Sigma}\mathbf{B_y}\}$, $\mathbf{B}'_\mathbf{f} = \mathbf{B}'_\mathbf{y}\mathbf{\Phi}$ and $diag\{\mathbf{B}'_\mathbf{y}\mathbf{\Sigma}\mathbf{B_y}\}$ denotes the vector composed of the main diagonal of the matrix $\mathbf{B}'_\mathbf{y}\mathbf{\Sigma}\mathbf{B_y}$. Similarly, the risk-neutral expectation of the future factor vector in (2) can be written as:

$$\mathbf{q}^f_t = \mathbf{W}'\mathbf{f}_t = \mu_\mathbf{q} + \mathbf{\Phi}\mathbf{q}_t, \quad (14)$$

where $\mu_\mathbf{q} = \mu - \frac{1}{2}\mathbf{W}'\mathbf{M}diag\{\mathbf{B}'_\mathbf{y}\mathbf{\Sigma}\mathbf{B_y}\}$. The forward yields are generally noisy, thus the observable relationships are:

$$\mathbf{f}^o_t = \mathbf{a_f} + \mathbf{B}'_\mathbf{y}\mathbf{\Phi}\mathbf{q}_t + \mathbf{w}_t, \quad (15)$$

$$\mathbf{q}^{f,o}_t = \mu_\mathbf{q} + \mathbf{\Phi}\mathbf{q}_t + \mathbf{W}'\mathbf{w}_t, \quad (16)$$

where $\mathbf{w}_t \sim N(\mathbf{0}, \mathbf{\Sigma_w})$. Unrestricted estimates of the coefficients in these relationships can be estimated by OLS. Eq.(15) gives estimates $\hat{\mathbf{a}}_\mathbf{f}$ and $\widehat{\mathbf{B}}_\mathbf{f}$ of $\mathbf{a_f}$ and $\mathbf{B}'_\mathbf{y}\mathbf{\Phi}$. Eq.(16) gives a single-step regression estimator of $\mathbf{\Phi}$, avoiding the problems of generated regressors associated with multi-step regression procedures. The forward yields used in this estimator have the same maturities (and weights $\mathbf{W}$) as the yields used in (7) to calculate the observable factors $\mathbf{q}_t$. This is unnecessarily restrictive: one can use any other selection of forward yields as long as there are at least K of them. In this paper we follow ACM and $AACMY$ in dropping the short rate equation in (15) on the grounds that the short rate and its forward are especially noisy. We keep the other maturities used to estimate the PCs.

Dropping the first regression in (15) gives the the estimate $\widehat{\mathbf{B}}'_{\mathbf{f}(1)}$ of $\mathbf{B}'_{\mathbf{y(1)}}\mathbf{\Phi}$. Similarly, dropping the first regression in (8), the unrestricted estimate of $\mathbf{B}'_{\mathbf{y(1)}}$ is given by $\mathbf{W}_{(1)}$, which omits the first row of $\mathbf{W}$. Premultiplying $\widehat{\mathbf{B}}'_{\mathbf{f}(1)} = \mathbf{W}_{(1)}$ by its generalised inverse gives the spot-forward (SF) estimator:

$$\widehat{\mathbf{\Phi}} = (\mathbf{W}'_{(1)}\mathbf{W}_{(1)})^{-1}\mathbf{W}'_{(1)}\widehat{\mathbf{B}}'_{\mathbf{f}(1)}, \tag{17}$$

The estimate of μ can be found conditional on $\widehat{\mathbf{\Phi}}$. Specifically, we use $\widehat{\mathbf{\Phi}}$ to construct the arbitrage-free loadings $\mathbf{B}$ using the recursions (3) and then estimate μ as:

$$\widehat{\mu}= (\mathbf{B}_{(1)}\mathbf{B}'_{(1)})^{-1}\mathbf{B}_{(1)}(\widehat{\mathbf{a}}_{\mathbf{f}(1)} - \widehat{\mathbf{a}}_{\mathbf{y}(1)} + \frac{1}{2}\mathbf{M}^{-1}_{(1)}diag\{\mathbf{B}'_{(1)}\Sigma\mathbf{B}_{(1)}\}, \tag{18}$$

where $\widehat{\mathbf{a}}_{\mathbf{y}}$ and $\widehat{\mathbf{a}}_{\mathbf{f}}$ are OLS estimates of the intercept terms in (5) and (15), respectively and, as before, the subscript (1) indicates that we omit the short rate.

III. Simple self-consistent term structure estimators

Hamilton and Wu (2012) noted that the parameters of the term structure model needs to be appropriately restricted to yield a self-consistent no-arbitrage model (see also discussion in Diez de Los Rios (2015)). In this section we show how inconsistent estimates of the risk-neutral dynamics of the model obtained by linear regression (such as $AACMY$, unconstrained DLR or the SF regressions proposed in the previous section), can easily be transformed to get estimates of the structural parameters of the underlying self-consistent model.

A. Structural restrictions and the JSZ identification scheme

A convenient identification scheme was proposed by JSZ, where the underlying latent factors follow the risk-neutral dynamics:

$$\mathbf{x}_t = \mu_{\mathbf{x}} + \mathbf{\Phi}_{\mathbf{x}}\mathbf{x}_{t-1} + \mathbf{u}_{\mathbf{x},t}, \tag{19}$$

where $\mathbf{\Phi}_{\mathbf{x}}$ has an ordered Jordan form determined by K parameters only and $\mathbf{u}_{\mathbf{x},t} \sim i.i.d.N(\mathbf{0}, \Sigma_{\mathbf{x}})$. The theoretical log bond prices are:

$$-\mathbf{p}_t = \mathbf{a}_{\mathbf{x}} + \mathbf{B}'_{\mathbf{x}}\mathbf{x}_t, \tag{20}$$

where $\mathbf{a}_{\mathbf{x}}$ and $\mathbf{B}_{\mathbf{x}}$ follow from recursion systems analogous to (3) and (4). The observed yields can be expressed as:

$$\mathbf{y}^o_t = \mathbf{a}_{\mathbf{y},\mathbf{x}} + \mathbf{B}'_{\mathbf{y},\mathbf{x}}\mathbf{x}_t + \mathbf{v}_t, \tag{21}$$

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

$$r_t = \delta_{\mathbf{x},0} + \delta'_{\mathbf{x},1}\mathbf{x}_t + v_{1,t}. \tag{22}$$

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

Substituting (20) into (7) and assuming that $\mathbf{W}'\mathbf{v}_t = \mathbf{0}$ allows us to switch between the latent and observable factors using:

$$\mathbf{x}_t = -(\mathbf{W}'\mathbf{B}'_{\mathbf{y},\mathbf{x}})^{-1}\mathbf{W}'\mathbf{a}_{\mathbf{y},\mathbf{x}} + (\mathbf{W}'\mathbf{B}'_{\mathbf{y},\mathbf{x}})^{-1}\mathbf{q}_t. \tag{23}$$

This implies that the relationship between the coefficients in (2) and (19) is:

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

$$\mu = (\mathbf{W}'\mathbf{B}'_{\mathbf{y},\mathbf{x}})\mu_{\mathbf{x}} + (\mathbf{I} - \mathbf{\Phi})\mathbf{W}'\mathbf{a}_{\mathbf{y},\mathbf{x}}, \tag{25}$$

and

$$\Sigma_{\mathbf{x}} = (\mathbf{W}'\mathbf{B}'_{\mathbf{y},\mathbf{x}})^{-1}\Sigma(\mathbf{W}'\mathbf{B}'_{\mathbf{y},\mathbf{x}})^{-1\prime}. \tag{26}$$

Note that if $\mathbf{W}'\mathbf{v}_t \neq \mathbf{0}$, then this introduces a measurement error into the estimate $\mathbf{x}_t^o$ given by (23):

$$\mathbf{x}_t^o = \mathbf{x}_t - (\mathbf{W}'\mathbf{B}'_{\mathbf{y},\mathbf{x}})^{-1}\mathbf{W}'\mathbf{v}_t. \tag{27}$$

This is correlated with the measurement errors in (21), potentially biasing the parameter estimates. However, the bootstrap simulations reported in Section IV.A.4 suggest that in practice this assumption is acceptable, since the measurement error bias is negligible.

The JSZ parametrization has the very convenient feature that the parameters $\{\mu^{\mathcal{P}},\mathbf{\Phi}^{\mathcal{P}}\}$ describing the physical dynamics of the observable factors (9) are unrestricted and can be estimated by OLS regression. This regression also provides a consistent (although not efficient) estimate of the Σ 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 (19).

B. The simple consistent (SC) estimator

Any approach that returns unconstrained estimates of the $\mathbf{\Phi}$ matrix uses $K \times K$ free parameters instead of the K parameters in $\mathbf{\Phi}_{\mathbf{x}}$, resulting in a model that is overparameterised and internally inconsistent. However, the factor structure (24) of the $\mathbf{\Phi}$ matrix suggests that the underlying roots of the model can be found from the eigenvalue decomposition of a regression-based $\hat{\mathbf{\Phi}}$. 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). Having found the roots of the model in this way, we can put them in the real Jordan form $\mathbf{\Phi_x}$ and reproduce consistent slope loadings in (1) from

$$\mathbf{B}' = \mathbf{B}'_{\mathbf{x}}(\mathbf{W}'\mathbf{M}^{-1}\mathbf{B}'_{\mathbf{x}})^{-1}, \tag{28}$$

where the columns in $\mathbf{B_x}$ given by recursions:

$$\mathbf{b}_{\mathbf{x},m} = \mathbf{b}_{\mathbf{x},1} + \hat{\mathbf{\Phi}}'_{\mathbf{x}}\mathbf{b}_{\mathbf{x},m-1} \tag{29}$$

with $\mathbf{b}_{\mathbf{x},1} = \delta_{\mathbf{x},1}$.

The level parameter can be identified by specifying jointly $\mu_{\mathbf{x}} = (\mu_\infty^{\mathcal{Q}}, 0, \ldots, 0)'$ and $\delta_{\mathbf{x},0} = 0$. Alternatively, if the factors are stationary under the risk-neutral measure (the most persistent factor has the multiplicity one with autoregressive coefficient $|\phi_1| < 1$), the level parameter can be identified by setting $\delta_{\mathbf{x},0} = r_\infty^{\mathcal{Q}}$ and $\mu_{\mathbf{x}} = \mathbf{0}$. These two identification schemes are equivalent in this case and the parameters are related by $r_\infty^{\mathcal{Q}} = \mu_\infty^{\mathcal{Q}}/(1-\phi_1)$. However, if the factors contain a unit root under the risk-neutral measure, $r_\infty^{\mathcal{Q}}$ is not identified, which is the reason JSZ prefer identification of the level parameter in terms of $\mu_\infty^{\mathcal{Q}}$. Hamilton and Wu (2012) point out another advantage of this parametrization choice: when the model is parametrized in terms of $r_\infty^{\mathcal{Q}}$, the numerical ML routine tends to get stuck with one of the factors close to the unit root. The reason for this numerical problem is that at this point the likelihood surface becomes flat along the dimension of the parameter $r_\infty^{\mathcal{Q}}$.

We estimate the level parameter consistently by conditioning it on $\hat{\mathbf{\Phi}}_{\mathbf{x}}$ (the roots of $\mathbf{\Phi}$) and the associated arbitrage-free yield loading coefficients $\mathbf{B}_{\mathbf{y},\mathbf{x}}$, and hence (using (28)) $\mathbf{B}$. Substituting (23) into (21) gives:

$$\begin{aligned} \mathbf{a}_{\mathbf{y}} &= \mathbf{H}\mathbf{a}_{\mathbf{y},\mathbf{x}} \\ &= \mathbf{H}(\mathbf{c}_0\mu_\infty - \mathbf{c}_1), \end{aligned} \tag{30}$$

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

$$\mathbf{H} = \mathbf{I} - \mathbf{B}'_{\mathbf{y},\mathbf{x}}(\mathbf{W}'\mathbf{B}'_{\mathbf{y},\mathbf{x}})^{-1}\mathbf{W}' \tag{31}$$

and $\mathbf{I}$ is the identity matrix. The terms $\mathbf{c}_0$ and $\mathbf{c}_1$ are defined in Appendix B. If we adopt the parametrization of the level parameter with $\mu_{\mathbf{x}} = (\mu_{\infty}^{\mathcal{Q}}, 0, \ldots, 0)'$ and $\delta_{\mathbf{x},0} = 0$, the consistent estimate of the level parameter is given by:

$$\widehat{\mu}_{\infty}^{Q} = \left(\mathbf{c}_0'\mathbf{H}'\mathbf{H}\mathbf{c}_0\right)^{-1}\left(\widetilde{\mathbf{a}}_{\mathbf{y}} + \mathbf{H}\mathbf{c}_1\right), \tag{32}$$

where $\widetilde{\mathbf{a}}_{\mathbf{y}}$ is a vector of unrestricted yield intercepts given the arbitrage-free slope coefficients $\mathbf{B}'_y$, i.e. $\widetilde{\mathbf{a}}_{\mathbf{y}} = \overline{\mathbf{y}}^o - \mathbf{B}'_{\mathbf{y}}\overline{\mathbf{q}}$, where $\overline{\mathbf{y}}^o$ and $\overline{\mathbf{q}}$ denote vectors of sample means of yields and the observable factors, respectively. For the sake of notational simplicity, it is understood that the terms on the right hand side of (32) are evaluated conditional on the estimated roots.

To summarize, our self-consistent spot-forward estimator, $SC(SF)$, is obtained as follows. In a single market model, the vector of parameters to estimate is $\Theta = \{r_{\infty}^{\mathcal{Q}}, \mathbf{\Phi}_{\mathbf{x}}, \mu^{\mathcal{P}}, \mathbf{\Phi}^{\mathcal{P}}, \Sigma, \Sigma_{\mathbf{v}}\}$. We follow JSZ, $AACMY$ and others in estimating the physical dynamics of the observable factors $(\mu^{\mathcal{P}}, \mathbf{\Phi}^{\mathcal{P}}, \Sigma)$ in (9) by OLS. We obtain $\hat{\mathbf{\Phi}}$ using OLS estimates of the spot-forward relations (15) and (17). The roots of the model are the eigenvalues of $\hat{\mathbf{\Phi}}$, which form $\hat{\mathbf{\Phi}}_{\mathbf{x}}$. Given the estimates of Σ and $\hat{\mathbf{\Phi}}_{\mathbf{x}}$, we compute the loadings $\mathbf{B}$ and find the level parameter using (32). Similarly, our simple consistent $AACMY$ estimator, $SC(AACMY)$, is obtained by finding $\hat{\mathbf{\Phi}}$ from excess return regressions as discussed in Appendix A.

IV. Empirical applications

A. The U.S. Treasury market

A.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 (forthcoming). This is similar to the data set used by ACM 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, Sack, and Wright (2007) using the Svensson (1994) parametric method, which are published by the Federal Reserve Board.[2] 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 (forthcoming).

A.2. The performance of different estimators

Table I shows the performance of different estimators, measured in terms of the root-mean-square error for each yield and the average across yields for 3, 4 and 5 factors. The OLS and ML estimate of the JSZ model (reported as JSZ for short) constitute the benchmark for other models. The OLS model gives the best possible linear fit of yields to the observable factors. The MLE routine searches numerically for the parameters (including Σ) that maximize the likelihood function subject to the no-arbitrage constraints, providing the upper bound for the no-arbitrage approaches. It is the only one that optimizes the likelihood function for these eight base yields. The other approaches that we consider optimize the fit of forward rates or ex post returns for different sets of maturities. The methods that relax the consistency condition involve extra parameters that could, in principle, allow them more

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

freedom to fit their specific objective criterion, but which could lead to extrapolation errors when fitting the eight base yields. Finally, these other approaches satisfy the restriction that the Σ parameters appearing in the cross-section (2) and the time series dynamics (9) are the same, by using the parameters of the latter in the former, while the MLE routine uses these parameters to optimize the fit of both jointly. However, all methods considered in the table use the same observable factors, which are the first K PCs of our eight base yields.

In this comparison, we also consider the estimator proposed by Diez de Los Rios (2015) . This method is based on a linear regression which fits all consecutive maturities to obtain the unrestricted loadings and intercepts. For the short maturities, $1-3$ months we use the yields from the Fama Treasury bills files, and for other maturities we use the Gurkaynak, Sack, and Wright (2007) data. As noted in the introduction, the virtue of the DLR method is that it can be constrained using iterative methods to satisfy the self-consistency requirement.[3]

[Insert Table I near here]

Table I indicates that three factors are sufficient to give a plausible fit to the term structure, in line with typical bid-ask spreads in the 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 performance of the constrained linear estimator proposed by DLR is always visibly worse than the JSZ benchmark. This should not be surprising, since the DLR estimator is based on all consecutive maturities including a relatively large number of noisy short and long term yields, rather than the eight base yields fitted by MLE.

The next two results are for the $AACMY$ and SF regressions, which do not impose the internal self-consistency condition. The $AACMY$ 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 this is noticeably worse than the OLS and MLE benchmarks. The average $RMSE$ for the SF method for the 3 factor model is about 8.5 basis points, which is much lower than for the

[3]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.

DLR and $AACMY$ routines. The fit obtained by the SF method for the 5 factor model is, however, the worst of all approaches. This is mainly due to the poor fit apparent for the $1-$year yield.

The last two entries in the table, $SC(AACMY)$ and $SC(SF)$, compare the fit of the two self-consistent estimation methods based on the roots of the estimates of the $\mathbf{\Phi}$ matrix obtained by the $AACMY$ and SF regressions,. Imposing the self-consistency restrictions always improves the fit dramatically. For instance, $SC(AACMY)$ 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 $AACMY$ and SF estimators are reduced from about $1.6-2.0$ basis points to under 1 basis point. Remarkably, the $SC(SF)$ procedure achieves a fit very close to that obtained by the MLE. This is impressive given the fact that the key parameters are estimated by simple linear regression, using data for a selection of spot-forward 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 $\Phi_{\mathbf{x}}$ provided by the basic SF regressions are very close to those of the ML estimates. The rest of this section tests the robustness of this finding.

A.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 II reports the ML estimates of the roots of $\Phi_{\mathbf{x}}$ 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 II also reports (as $SC(AACMY)$ and $SC(SF)$, respectively) the eigenvalues of the Φ matrix estimated by the $AACMY$ and SF

regressions. It is striking how close the roots of the SF estimator always are to the MLE estimates. The roots obtained from the $AACMY$ estimator are generally some distance away from the MLE values. Most problematic for the $AACMY$ 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.

[**Insert Table II near here**]

A.4. Robustness

To examine the robustness of these results, we have compared the performance of the SC and MLE methods across several other data sets, including those reported in the next section. We also conducted a bootstrap simulation exercise that takes the ML estimates of the JSZ model in Table II 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. The rest of this section reports the results.

In the first stage we use the estimates of the physical dynamics in (9) to simulate the time series of the PCs[4]. 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) then adding the measurement errors. These are 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 PCs and the JSZ assumption that weighting up the 'true' residuals $\mathbf{v}_t^*$ using the sample-dependent weights $\mathbf{W}$ gives $\mathbf{W}'\mathbf{v}_t^* = 0$. Setting

[4]Specifically, starting from the value $\mathbf{q}_1$ in the first period of the original sample, we use (9) 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 were 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.

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 III.A. Finally, our researcher uses these estimated factors to estimate the JSZ model using the SC and MLE methods. The starting values for the parameters used to initiate the ML routine are discussed in the next subsection. Table II reports the bias and root-mean-square error ($RMSE$) of each method. It shows that the $SC(SF)$ and ML estimates display negligible bias and that their $RMSE's$ are close. However, the $SC(AACMY)$ estimator always has a much larger $RMSE$ than the $SC(SF)$. It also exhibits large negative bias in the least persistent real root, consistent with our observation in the previous paragraph regarding the point estimate of this parameter.

A.5. Linear regression estimates as starting values in ML estimation

So far we have considered the SF and $SC(SF)$ 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 SF method as potentially useful starting values. To help throw light on this option, we used these estimates as one of the sets of parameter starting values for the ML routine in the bootstrap simulations. We used the 'true' values which were used to simulate the artificial data as another set. Finally, to represent the usual practice, we use 5 sets of random starting values.[5] The parameters reported in Table II 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

[5]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 Σ again come from the VAR model of the physical dynamics.

to distinguish the global from other local optima.

As Hamilton and Wu (2012) observed, we found that parametrization in terms of the level parameter $r^{\mathbb{Q}}_{\infty}$ frequently led the numerical routine to stall. They showed that this problem is alleviated by parametrization in terms of $\mu^{\mathbb{Q}}_{\infty}$, but we go a step further with our MLE algorithm by concentrating the level parameter out from the likelihood function, as shown in Section III.B. Thus, we only search numerically for the K roots of $\mathbf{\Phi}$ matrix and the $K(K+1)/2$ parameters of Σ. We find that this algorithm is very robust for the model with 3 factors, always converging to the same maximum (with tolerance of +1), irrespective of the choice of starting values. The ability of different specifications to find the maximum maximorum for the 4 and 5 factor models is reported in Figure 1. This shows that with 4 and 5 factors, initiating the search from the roots of the SF 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 are used, this is found in only 91% of simulations (see Figure 7(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 7(b)).

[**Insert Figure 1 near here**]

B. The U.S. Treasury and German Bund markets

Encouraged by the results for the U.S. Treasury market, we then examined the performance of the SC approach for other countries like Germany, and then went on to look at a joint two-country model. $AACMY$ had already used their regression approach to model the U.S. Treasury and $TIPS$ markets jointly. In a similar spirit we use the $SC(SF)$ estimator to estimate a two-country German-U.S. model, which should offer an interesting test. Specifically, we look at U.S. Treasury bonds and German bunds from the perspective of a

U.S. dollar based investor and use the dollar risk-neutral probability measure for pricing. The forward rate algebra and the common factor model are set out in Appendix C.

For this two-country application we used data for the period January 1987 to December 2015, which was dictated by the availability of data for the German market. For the U.S., the annual maturities again come from Gurkaynak, Sack, and Wright (2007). For Germany, we use estimates published by the Bundesbank, also estimated using the Svensson method.[6] We use maturities 1 month, 1, 2, 3, 4, 5, 7, and 10 years for both countries. In the absence of the equivalent Fama-Bliss estimates of German short rates we use the 1−month Euribor rate and, for consistency, we also use the 1−month Libor to complete the U.S. data set. These data are shown in Figure 2.

[**Insert Figure 2 near here**]

The first principal component for each country is depicted in Figure 3. These are similar, exhibiting a strong downward tend, but there is some suggestion that the U.S. leads Germany. Figure 4 shows the factor loadings for the two countries, which are remarkably close, indicating that a common factor model might be appropriate.

[**Insert Figure 3 near here**]

[**Insert Figure 4 near here**]

We begin by estimating separate models for the two countries. Table III shows the results of this exercise in terms of model fit. Once again, the improvement in fit of the $SC(SF)$ estimator over $AACMY$ is impressive. To conserve space, we do not report other methods, such as SF or SC estimator based on $AACMY$ estimate of the $\mathbf{\Phi}$ matrix, but the ranking of the results is consistent with those reported in Section IV.A: the $SC(SF)$ method greatly improves upon the estimators that do not impose the self-consistency restrictions and gives

[6] Available at: http://www.bundesbank.de/Navigation/EN/Statistics/Money_and_capital_markets/Interest_rates_and_yields/Term_structure_of_interest_rates/term_structure_of_interest_rates.html.

the best fit among the methods based on linear regression. Generally, as reported in Table III the $SC(SF)$ estimator gives a close fit to the unrestricted OLS regressions for any number of factors. The fit for Germany is better than for the U.S. For the models with 3 factors the average $RMSE$ for the U.S. obtained by the $SC(SF)$ method is 1 basis point larger for the U.S. yields and 0.2 basis points for Germany, relative to the OLS fit. For the 5 factor model the fit of the $SC(SF)$ is virtually the same as the OLS benchmark. The $RMSE$ results for the $AACMY$ method are generally much higher, often more than double that of the OLS.

[Insert Table III near here]

Table IV reports the cumulative percentage of the variance in the U.S. and German yields explained by the PCs extracted (a) separately from the single country yield covariance matrices and (b) from the joint covariance matrix of yields. This provides a *prima facie* indication of the performance of an unrestricted OLS common factor model. These results do not suggest that common factors are present since to achieve a similar fit, the common factor model needs as many factors as the total number used in the two separate country models. For instance, if we assume that the yields of each country are appropriately fitted by the 3 first PCs extracted from each country's data (which explain about 99.97% of their variance), then we need to use 6 PCs extracted from the joint data set to achieve the same performance.

[Insert Table IV near here]

The results for the common factor models are shown in Table V. The technical details extending the $SC(SF)$ estimator to multi-market framework are relegated to Appendix C. For Germany, the fit of the 6 factor joint model is similar to that for the 3 factor single-country model shown in Table III, although the fit of the 6 factor joint model for the U.S. is slightly better than that obtained for the single country model. Importantly, consistent with the single-country results, the $SC(SF)$ estimator performs very well, giving $RMSEs$ generally very close to those of the OLS benchmark. On the other hand, the $AACMY$

estimator (and SF regressions in unreported results), which do not impose self-consistency of the model, perform much worse with $RMSEs$ often more than double those given by the OLS estimator

[Insert Table V near here]

These results support the argument that the $SC(SF)$ estimator offers a computationally efficient alternative to the full ML method used by JSZ. It provides a similar fit to the data, which is not achieved by the routines that do not impose the self-consistency restrictions. In the remainder of this section we look at the relationship between the single and two-country models and the view that the U.S. and German markets are spanned by common factors.

The factors in the common and single country models exhibit an interesting pattern. The first common factor is depicted in Figure 3 and is clearly a compromise between the level factors in the two single country models. We find that the second common factor is highly correlated with the difference between the level factors in the two single country models. This relationship is shown in the top panel of Figure 5. Furthermore, we find that regressing the level factors for the single country models on the first two common factors gives an almost perfect fit (not reported). We find similar pair-wise factor-rotation relationships between the slope and curvature factors. Specifically, the third common factor is a slope factor and is a compromise between the second factors in the single country models, while the fourth common factor reflects the difference between them, as shown by the middle panel in Figure 5. Similarly, the fifth common factor is a compromise between the third (curvature) factors in the single country models and the sixth common factor reflects the difference between them, as shown by the final panel in Figure 5.

[Insert Figure 5 near here]

These pair-wise factor-rotation relationships are reflected in the factor loadings for the two countries in the 6 common factor model shown in Figure 6. The continuous lines show the loadings on the odd-numbered factors, which are interpreted as common level (black), slope (blue) and curvature (red) factor loadings respectively. These clearly resemble the respective loadings in the single country models shown in Figure 4. The loadings on the associated even numbered common factors are shown with dashed lines. Consistent with the view that these allow for relative effects missed by the associated common factors, the signs are reversed for the U.S. and Germany.

[**Insert Figure 6 near here**]

Thus, we find that if we do not restrict the loadings but just allow the data to speak, these data tell us that the joint model really just mimics the U.S. and German single country models, giving a similar fit. Figure 3 shows that the level factor in the joint model picks up the down-trend in the single country factors, but misses out important relative effects - indeed there is a suggestion that the U.S. level factor leads the German level factor. Clearly, the second joint factor is needed to allow for these differences, with similar effects in the slope and curvature factors. We conclude that most of the contemporaneous interaction between the U.S. and Germany can be spanned equally well by single-country models.

Nevertheless, a related paper using these data finds strong evidence of *unspanned* (i.e. delayed) spillovers between the U.S. and Germany (Meldrum, Raczko, and Spencer (2016)). This finds that overseas unspanned factors - constructed from the components of overseas yields that are uncorrelated with domestic yields - have significant explanatory power for subsequent domestic bond returns. Moreover, reflecting this, we find that there *are* significant differences between the single country and joint approaches in terms of the physical dynamics and hence the risk premia (which reflect differences between the risk-neutral and physical dynamics). Figure 7 shows how the model decomposes the $10-$year yield into components representing $10-$year interest rate expectations and a risk premium. This decomposition largely depends upon the physical dynamics represented by the time series VAR in (9), and in particular its maximum eigenvalue which governs their persistence. The VAR is used to forecast the interest rates and hence compute the yields that would be observed if the expectations hypothesis held and investors were risk-neutral. Subtracting this estimate from the model-fitted yield for any maturity then gives the standard estimate of the risk premium. This is shown in Figure 7 for both countries and for both single and common factor models. Pooling the data for the two countries seems to make little qualitative difference to the U.S. decomposition. However, it does give a more plausible decomposition of the German curve. This is explained by the difference in the persistence of the physical dynamics and hence the risk-neutral expectations.

[**Insert Figure 7 near here**]

Table VI shows that the maximum eigenvalue for the common factor VAR is straddled by those of the two single country $VARs$. The maximum eigenvalue in our German-only VAR model is 0.9992, which is very close to a unit root. This makes interest rates highly persistent, suggesting that the negative rates seen recently will last, making the risk-neutral expectations component of the 10−year yield negative. Thus, the German-only model attributes most of the fall in the 10−year yield since 2012 to falling interest rate expectations, leaving the risk premium relatively stable. In contrast, the joint model suggests that German rates will normalize more quickly, attributing the recent fall in the 10−year yield largely to the fall in the risk premium, with falling interest rate expectations playing a secondary role.

[**Insert Table VI near here**]

V. Conclusion

This paper proposes a simple regression-based estimator of the risk-neutral dynamics of the bond market that is consistent in the sense that it reproduces the factors used in estimation. It is reassuring to find that our parameter estimates are very close to ML estimates of the risk-neutral parameters embedded in bond yields, and that the bias revealed by the bootstrap 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, more of a surprise, to fit the term structure of bond yields by extrapolation. However, the speed of the regression-based approach gives it a clear edge in simulation exercises involving a large number of replications. Reflecting this, regression methods have been extensively employed by central banks and other policymakers, despite their lack of consistency and efficiency. Our new estimator avoids this trade-off and allows both academic researchers and practitioners to obtain results with speed, consistency and efficiency.

References

Abrahams, M., T. Adrian, R. Crump, E. Moench, and R. Yu (2016): "Decomposing real and nominal yield curves," *Journal of Monetary Economics*, (84), 182–200.

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

Diez de Los Rios, A. (2015): "A new linear estimator for Gaussian dynamic term structure models," *Journal of Business & Economic Statistics*, 33(2), 282–295.

Duffee, G. R. (2011): "Information in (and not in) the term structure," *Review of Financial Studies*, 24(9), 2895–2934.

Goliński, A., and P. D. Spencer (forthcoming): "The advantages of using excess returns to model the term structure," *Journal of Financial Economics*.

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.

Hamilton, J. D., and J. C. Wu (2012): "Identification and estimation of Gaussian affine term structure models," *Journal of Econometrics*, 168(2), 315–331.

Joslin, S., A. Le, and K. J. Singleton (2013): "Gaussian macro-finance term structure models with lags," *Journal of Financial Econometrics*, 11(4), 581–609.

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.

Joslin, S., K. J. Singleton, and H. Zhu (2011): "A new perspective on Gaussian dynamic term structure models," *Review of Financial Studies*, 24(3), 926–970.

Meldrum, A., M. Raczko, and P. Spencer (2016): "Overseas unspanned factors and domestic bond returns," Bank of England working paper 618, Bank of England.

Politis, D. N., and J. P. Romano (1994): "The stationary bootstrap," *Journal of the American Statistical Association*, 89(428), 1303–1313.

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

Appendix A Regression models of the term structure

AACMY use *ex post* excess return rather than forward yield regressions. The *ex post* excess return $rx_{m,t}$ on the bond with maturity m in period t is found by subtracting its t period forward price from the actual price at $t+1$: $rx_{m,t} = p_{m,t+1} - p^f_{m,t}$. Stacking these returns, substituting (1) and (11) and then adding a vector of Gaussian error terms $\mathbf{e}_t$ gives their regression system:

$$\mathbf{rx}_{t+1} = \mathbf{p}_{t+1} - \mathbf{p}^f_t = \mathbf{a_{rx}} + \mathbf{C}'\mathbf{q}_t - \mathbf{B}'\mathbf{q}_{t+1} + \mathbf{e}_t, \tag{A-1}$$

where: $\mathbf{a_{rx}} = \mathbf{B}'\mu - \frac{1}{2}diag\{\mathbf{B}'\mathbf{\Sigma}\mathbf{B}\}$ and $\mathbf{C}' = \mathbf{B}'\mathbf{\Phi}$. The *AACMY* estimate of $\mathbf{\Phi}$ is:

$$\hat{\mathbf{\Phi}}_{GLS} = (\hat{\mathbf{B}}'\hat{\mathbf{\Sigma}}_e^{-1}\hat{\mathbf{B}})^{-1}\hat{\mathbf{B}}'\hat{\mathbf{\Sigma}}_e^{-1}\hat{\mathbf{C}}' \tag{A-2}$$

where $\hat{\mathbf{\Sigma}}_e$ is the estimate of the covariance matrix of $\mathbf{e}_t$ obtained from the *OLS* residuals in (A-1). Finally, this estimator is used to obtain more efficient generalised least squares

(GLS) estimates of $\mathbf{a}$ and $\mathbf{B}$ using the regression:

$$\mathbf{rx}_{t+1} = \mathbf{a}_{GLS} + \mathbf{B}'_{GLS}(\hat{\mathbf{\Phi}}_{GLS}\mathbf{q}_t - \mathbf{q}_{t+1}) + \mathbf{e}_{GLS,t}, \tag{A-3}$$

This estimator differs from our spot-forward estimator in several ways. First, $AACMY$'s excess returns use log prices and their forwards while we use regressions for yields and their forwards. Second, $AACMY$ obtain $\hat{\mathbf{B}}$ and $\hat{\mathbf{C}}'$ from a single regression (A-1) while we obtain the analogues $\mathbf{W}_{(1)}$ and $\widehat{\mathbf{B}}'_{\mathbf{f}(1)}$ from separate regression equations for yields (the principal component analysis) and their forwards (15). Their regression (A-1) effectively combines separate projections for log prices and their forwards onto $\mathbf{q}_t$ and $\mathbf{q}_{t+1}$ which we may write as:

$$\mathbf{p}_{t+1} = \mathbf{a}_{\mathbf{p}} + \mathbf{B}'_{\mathbf{1}}\mathbf{q}_t - \mathbf{B}'_2\mathbf{q}_{t+1} + \mathbf{e}_{\mathbf{p},t+1}, \tag{A-4}$$

$$-\mathbf{p}^f_t = \mathbf{a}_{\mathbf{fp}} + \mathbf{C}'_{\mathbf{1}}\mathbf{q}_t - \mathbf{C}'_2\mathbf{q}_{t+1} + \mathbf{e}_{\mathbf{f},t}, \tag{A-5}$$

where their projection coefficients are related to this decomposition by: $\mathbf{B}' = \mathbf{B}'_2 + \mathbf{C}'_2$ and $\mathbf{C}' = \mathbf{C}'_{\mathbf{1}} + \mathbf{B}'_{\mathbf{1}}$. Since the factors $\mathbf{q}_t$ are autocorrelated this means that $\mathbf{B}$ and $\mathbf{C}$ differ from the estimates that would be obtained by regressing $\mathbf{p}_{t+1}$ (or in our system (8) $\mathbf{y}_{t+1}$) on $\mathbf{q}_{t+1}$ and $\mathbf{p}^f_t$ (or in (15) $\mathbf{f}_{t+1}$) on $\mathbf{q}_t$ separately. Finally, we find that in spot-forward regressions OLS estimators such as $\widehat{\mathbf{\Phi}}$ are generally more robust than GLS estimators such as $\hat{\mathbf{\Phi}}_{GLS}$.

Appendix B Estimating the level parameter for a single market

The affine coefficients $\mathbf{a}_{\mathbf{x}}$ follow a recursion similar to (4): .

$$a_{\mathbf{x},m} = a_{\mathbf{x},1} + a_{\mathbf{x},m-1} + \mathbf{b}'_{\mathbf{x},m-1}\mu_{\mathbf{x}} - \frac{1}{2}\mathbf{b}'_{\mathbf{x},m-1}\mathbf{\Sigma}_{\mathbf{x}}\mathbf{b}_{\mathbf{x},m-1}, \tag{B-1}$$

which implies the closed form:

$$a_{\mathbf{x},m} = m\delta_{\mathbf{x},0} + \left(\sum_{j=1}^{m} \mathbf{b}'_{\mathbf{x},j-1}\right)\mu_{\mathbf{x}} - \frac{1}{2}\sum_{j=1}^{m} \mathbf{b}'_{\mathbf{x},j-1}\mathbf{\Sigma_x}\mathbf{b}_{\mathbf{x},j-1}. \tag{B-2}$$

If we adopt the parametrization of the level parameter with $\mu_{\mathbf{x}} = (\mu_\infty^Q, 0, \ldots, 0)'$ and $\delta_{\mathbf{x},0} = 0$, we collect the terms in (B-2) in a vector and substitute into (30) to get;

$$\mathbf{a_y} = \mathbf{H}(\mathbf{c}_0\mu_\infty - \mathbf{c}_1), \tag{B-3}$$

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

$$c_{0,m} = \frac{1}{m}\mathbf{e}'_1\sum_{j=1}^{m}\mathbf{b}_{x,j-1}, \tag{B-4}$$

$$c_{1,m} = \frac{1}{2m}\sum_{j=1}^{m}\mathbf{b}'_{\mathbf{x},j-1}\mathbf{\Sigma_x}\mathbf{b}_{\mathbf{x},j-1}, \tag{B-5}$$

where $\mathbf{e}_1$ is a $K-$dimensional indicator vector with 1 in the first row and zeros elsewhere. Note that $\mathbf{c}_1$ is the Jensen convexity term. The parametrization of the model in terms of r_∞^Q yields the same formula (32) with the only difference that $\mathbf{c}_0$ being the $J-$dimensional vector of ones:

$$\widehat{r}_\infty^Q = \left(\mathbf{1}'\mathbf{H}'\mathbf{H}\mathbf{1}\right)^{-1}\left(\widetilde{\mathbf{a}}_\mathbf{y} + \mathbf{H}\mathbf{c}_1\right). \tag{B-6}$$

This linear estimator of the level parameter can also be used to concentrate it out of the ML routine used to estimate the JSZ model, further reducing the parameter space and improving the numerical convergence of the routine. It follows that the estimate of the level parameter conditional on the ML estimates of the roots yields the ML estimator of the level parameter.

Appendix C The two-country model

Let $P^{€}_{m,t}$ represent the price in euro of an $m-$maturity German zero coupon bond and S_t denote the spot euro-U.S. dollar exchange rate, expressed as the number of dollars exchanged for 1 euro. The dollar cost of this bond is thus $P^{€}_{m,t}S_t$. Similarly, the one period forward price of the German bond in dollars ($P^{f,€\to\$}_{m,t}$) is equal to the product of forward price in euro ($P^{f,€}_{m,t}$) and the forward exchange rate (S^f_t) and is the the dollar risk-neutral expectation of the dollar price in the next period:

$$\begin{aligned} P^{f,€\to\$}_{m,t} &= P^{f,€}_{m,t}S^f_t, \\ &= E^Q_t[P^{€}_{m,t+1}S_{t+1}], \end{aligned} \tag{C-1}$$

where S^f_t denotes the one period forward exchange rate. Evaluating this expectation assuming that bond prices and exchange rates are Gaussian; taking logs and rearranging gives an expression for the logarithm of the forward price in €:

$$\begin{aligned} p^{f,€}_{m,t} &= E^Q_t[p^{€}_{m,t+1} + s_{t+1}] - s^f_t + \frac{1}{2}Var_t\left[p^{€}_{m,t+1} + s_{t+1}\right] \\ &= E^Q_t[p^{€}_{m,t+1}] + E^Q_t[s_{t+1}] - s^f_t + \frac{1}{2}Var_t[p^{€}_{m,t+1}] + \frac{1}{2}Var_t[s_{t+1}] + Cov_t[p_{m,t+1}, s_{t+1}] \end{aligned}$$

where the lower case letters denote the logarithm of the upper case counterparts. We can replace the log forward exchange rate using a similar representation: $s^f_t \; = \log(E^Q_t[S_{t+1}])$ $= E^Q_t[s_{t+1}] + \frac{1}{2}Var_t[s_{t+1}]$. Cancelling common terms and rearranging allows us to represent the dollar risk-neutral expectation in terms of the log forward rate plus a constant (given homoscedasticity) specific to each maturity:

$$E^Q_t[p^{€}_{m,t+1}] = p^{f,€}_{m,t} - \frac{1}{2}Var_t[p^{€}_{m,t+1}] - Cov_t[p^{€}_{m,t+1}, s_{t+1}],$$

The risk-neutral expectations of the log Treasury bond prices follow directly from (11).

We collect bond prices for the two countries in one ($J = J^{\$} + J^{€}$) vector and assume that they are driven by a common set of K factors:

$$-\mathbf{p}_t = -\begin{bmatrix} \mathbf{p}_t^{\$} \\ \mathbf{p}_t^{€} \end{bmatrix} = \begin{bmatrix} \mathbf{a}^{\$} \\ \mathbf{a}^{€} \end{bmatrix} + \begin{bmatrix} \mathbf{B}^{\$\prime} \\ \mathbf{B}^{€\prime} \end{bmatrix} \mathbf{q}_t. \tag{C-2}$$

As before we assume that portfolios of bond yields represented by common principal components:

$$\mathbf{q}_t = -\mathbf{W}'\mathbf{M}^{-1}\mathbf{p}_t,$$

are noiseless, where $\mathbf{M}$ is a diagonal matrix composed of the respective maturities of the U.S. and German bonds, $\mathbf{m} = [\mathbf{m}^{\$\prime}, \mathbf{m}^{€\prime}]'$. Thus, we can estimate the roots of the risk-neutral dynamics from the eigenvalues of the $\hat{\mathbf{\Phi}}$ matrix estimated from (17). The estimation of the level parameters for the two countries is described in Appendix C.

In a multi-market setting we need to estimate the common factor parameters $\{\mathbf{\Phi_x}, \mu^{\mathcal{P}}, \mathbf{\Phi}^{\mathcal{P}}, \Sigma\}$ and $\{r_\infty^{\mathcal{Q}}, \delta_{\mathbf{x},1}, \Sigma_{\mathbf{v}}\}$ specifically for each market. In a multi-market setting with common factors, the loading recursions (29) are common to all markets, as Appendix C explains, but $\mathbf{b}_{\mathbf{x},1}$ needs to be estimated for each one. To estimate these, we propose the following strategy. First, from (28) we can write the relation between $\delta_{\mathbf{x},1}$ and δ_1 for each market:

$$\begin{aligned} \delta'_{\mathbf{x},1} &= \delta'_1(\mathbf{W}'\mathbf{B}'_{\mathbf{y},\mathbf{x}}) \\ &= \delta'_1(\mathbf{W}'\mathbf{B}'_{\mathbf{y},\mathbf{x}})\hat{\mathbf{\Phi}}_{\mathbf{x}}(\mathbf{W}'\mathbf{B}'_{\mathbf{y},\mathbf{x}})^{-1}(\mathbf{W}'\mathbf{B}'_{\mathbf{y},\mathbf{x}})\hat{\mathbf{\Phi}}_{\mathbf{x}}^{-1} \\ &= \delta'_1\hat{\mathbf{\Phi}}(\mathbf{W}'\mathbf{B}'_{\mathbf{y},\mathbf{x}})\hat{\mathbf{\Phi}}_{\mathbf{x}}^{-1}. \end{aligned} \tag{C-3}$$

Although the matrix $\mathbf{\Phi}$ and the coefficients $\mathbf{B_x}$ in a multi-country setting depend on the short rate coefficients $\delta_{\mathbf{x},1}$, we can substitute the eigendecomposition of the empirical estimate of $(\widehat{\mathbf{\Phi}} = \widehat{\mathbf{L}}\hat{\mathbf{\Phi}}_{\mathbf{x}}\widehat{\mathbf{L}}^{-1})$ and use the fact that $\widehat{\mathbf{L}} \simeq (\mathbf{W}'\mathbf{B}'_{\mathbf{y},\mathbf{x}})$ to obtain the estimated short rate loadings as:

$$\delta_{\mathbf{x},1} = \hat{\mathbf{L}}'\delta_1 \tag{C-4}$$

using the OLS estimate of δ_1 for each market as in (6).

The level parameters for a multi-market model follow in a similar manner as for the one-market model. Since there are two countries, the level of the term structures might be different and thus a more natural and intuitive parametrization of the level is in terms of r_{∞}^{Q} rather than μ_{∞}^{Q}; since e.g. level of the term structure in each country is generally different, we need multiple level parameters, and is it more natural to interpret multiple r_{∞}^{Q} rather than multiple μ_{∞}^{Q}. Recall that in this parametrization scheme we need to assume that there is no unit root under the risk-neutral measure. Although this assumption might seem restrictive, in the model estimation it is hardly an issue, since the probability of estimating the one of the risk-neutral roots of the model of exactly 1 in a linear regression is virtually zero. Having calculated the loading parameters $\mathbf{B_x}$= $[\mathbf{B}_{\mathbf{x}}^{\$}, \mathbf{B}_{\mathbf{x}}^{€}]$, we can find $\mathbf{H}$ and $\mathbf{c}_1 = [\mathbf{c}_1^{\$\prime}, \mathbf{c}_1^{€\prime}]'$ from (31) and (B-5), respectively. The vector of level parameters, $\mathbf{r}_{\infty}^{Q} = [r_{\infty}^{Q\$}, r_{\infty}^{Q€}]'$, can be estimated by:

$$\widehat{r}_{\infty}^{Q} = \left(\mathbf{C}_0'\mathbf{H}'\mathbf{H}\mathbf{C}_0\right)^{-1}\left(\widetilde{\mathbf{a}}_{\mathbf{y}} + \mathbf{H}\mathbf{c}_1\right), \tag{C-5}$$

where $\mathbf{C}_0$ is a $J \times 2$ block diagonal matrix:

$$\mathbf{C}_0 = \begin{bmatrix} \iota_{J^{\$}} & \mathbf{0} \\ \mathbf{0} & \iota_{J^{€}} \end{bmatrix}$$

and $\widetilde{\mathbf{a}}_{\mathbf{y}} = [\widetilde{\mathbf{a}}_{\mathbf{y}}^{\$\prime}, \widetilde{\mathbf{a}}_{\mathbf{y}}^{€\prime}]'$ is a vector of unrestricted yield intercepts given the arbitrage-free slope coefficients given by $\widetilde{\mathbf{a}}_{\mathbf{y}} = \overline{\mathbf{y}}^{o} - \mathbf{B}_{\mathbf{y}}'\overline{\mathbf{q}}$.

Tables

yields:	1m	1y	2y	3y	5y	7y	10y	15y	Av.RMSE
					Root-mean-square error (bp)				
					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
DLR	4.3	17.8	3.6	8.9	13.9	11.4	5.4	20.5	10.72
$AACMY$	2.5	20.4	6.4	7.9	14.4	15.2	10.6	20.2	12.20
SF	2.5	17.6	7.0	7.3	8.9	7.5	7.6	10.0	8.55
$SC(AACMY)$	2.7	13.1	4.3	8.3	8.8	6.7	6.4	11.0	7.65
$SC(SF)$	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
DLR	0.3	3.7	3.3	3.8	1.5	4.1	5.6	5.3	3.32
$AACMY$	0.3	5.0	3.8	3.5	1.6	3.8	4.0	4.2	3.28
SF	0.3	4.5	3.3	3.2	1.5	3.4	3.6	4.1	2.99
$SC(AACMY)$	0.3	3.4	3.3	3.2	1.7	3.5	3.6	4.2	2.89
$SC(SF)$	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
DLR	0.1	0.9	1.5	0.8	1.7	0.9	2.4	1.0	1.17
$AACMY$	0.1	3.2	2.6	1.4	1.9	0.8	2.2	1.0	1.64
SF	0.1	6.5	2.8	1.6	1.5	0.5	1.7	0.9	1.96
$SC(AACMY)$	0.1	0.9	1.6	0.4	1.5	0.5	1.9	0.9	0.97
$SC(SF)$	0.1	0.8	1.5	0.6	1.4	0.6	1.7	0.8	0.92

Table I. Performance in terms of fit for the Treasury yields, as measured by the root-mean-square error of the measurement noise. 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.

		λ_1	λ_2	λ_3	λ_{imag}	$\pm$ xi
		Panel A: $K = 3$				
JSZ	Estimates	0.9971	0.9714	0.7537		
	Bias	-0.0001	0.0023	-0.0015		
	RMSE	0.0003	0.0028	0.0244		
$SC(AACMY)$	Estimates	0.9958	0.9779	0.1657		
	Bias	-0.0019	0.0034	-0.0758		
	RMSE	0.0158	0.0051	0.0876		
$SC(SF)$	Estimates	0.9976	0.9686	0.7862		
	Bias	0.0001	-0.0012	0.0133		
	RMSE	0.0004	0.0022	0.0371		
		Panel B: $K = 4$				
JSZ	Estimates	0.9998	0.5930		0.9664	0.0194
	Bias	0.0002	0.0064		0.0014	-0.0027
	RMSE	0.0003	0.1057		0.0022	0.0040
$SC(AACMY)$	Estimates	0.9991	-0.0666		0.9656	0.0221
	Bias	0.0000	-0.1068		0.0005	0.0031
	RMSE	0.0005	0.1221		0.0035	0.0311
$SC(SF)$	Estimates	0.9998	0.6103		0.9655	0.0188
	Bias	0.0000	0.0010		-0.0005	-0.0005
	RMSE	0.0002	0.0564		0.0022	0.0096
		Panel B: $K = 5$				
JSZ	Estimates	0.9998	0.9621	0.6020	0.9728	0.0184
	Bias	0.0004	-0.0011	0.0291	-0.0001	-0.0031
	RMSE	0.0007	0.0098	0.0346	0.0039	0.0051
$SC(AACMY)$	Estimates	1.0012	0.9741	-0.0656	0.9625	0.0233
	Bias	-0.0016	0.0092	-0.1058	-0.0056	0.0061
	RMSE	0.0317	0.0170	0.1241	0.0497	0.0253
$SC(SF)$	Estimates	1.0003	0.9528	0.7100	0.9789	0.0211
	Bias	0.0002	-0.0051	0.0441	0.0025	0.0014
	RMSE	0.0004	0.0099	0.0548	0.0035	0.0028

Table II. Estimates of the roots and estimator performance. The table reports estimates of the roots of the model by different methods. Bias and RMSE calculated for different estimation methods 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.

Model	Country	Root-mean-square error (bp)								
		1m	1y	2y	3y	4y	5y	7y	10y	Av.RMSE
					Panel A: $K = 2$					
OLS	U.S.	24.6	12.5	14.4	12.3	8.4	4.5	7.8	18.0	12.81
	Germany	25.2	12.4	14.9	12.4	8.0	3.6	7.4	17.4	12.66
AACMY	U.S.	24.6	21.4	18.6	19.9	19.4	16.5	10.2	33.4	20.50
	Germany	25.2	15.6	19.0	22.3	23.5	21.9	13.1	31.1	21.45
$SC(SF)$	U.S.	26.6	17.5	15.1	12.3	9.2	7.0	8.7	19.6	14.50
	Germany	26.6	15.1	15.9	12.5	8.3	5.7	8.5	17.8	13.80
					Panel B: $K = 3$					
OLS	U.S.	3.0	7.5	2.6	3.0	4.2	4.1	1.8	6.1	4.03
	Germany	2.4	7.1	2.0	2.9	3.4	3.0	1.3	4.6	3.34
AACMY	U.S.	3.0	14.3	5.7	4.6	6.6	7.3	5.3	7.4	6.77
	Germany	2.4	8.4	3.6	3.3	3.8	3.7	2.7	5.1	4.14
$SC(SF)$	U.S.	3.6	10.2	4.0	4.4	4.6	4.4	2.6	6.5	5.03
	Germany	2.5	7.4	2.4	3.1	3.4	3.1	1.6	4.6	3.51
					Panel C: $K = 4$					
OLS	U.S.	0.3	2.0	2.1	1.4	0.5	1.4	1.8	1.5	1.38
	Germany	0.1	1.4	2.0	1.0	0.6	1.4	1.3	1.5	1.18
AACMY	U.S.	0.3	6.4	6.5	6.6	4.8	2.8	2.6	1.7	3.97
	Germany	0.1	5.1	2.8	1.7	1.3	1.7	1.5	1.6	1.98
$SC(SF)$	U.S.	0.4	2.6	3.2	1.5	1.2	1.8	2.1	1.7	1.81
	Germany	0.1	1.5	2.2	1.1	0.7	1.5	1.3	1.6	1.26
					Panel D: $K = 5$					
OLS	U.S.	0.0	0.2	0.6	0.3	0.4	0.1	0.7	0.3	0.33
	Germany	0.0	0.1	0.4	0.3	0.3	0.1	0.5	0.2	0.25
AACMY	U.S.	0.0	6.4	4.2	2.6	1.7	1.3	1.6	1.2	2.37
	Germany	0.0	4.0	1.7	1.1	0.9	0.7	0.6	0.3	1.16
$SC(SF)$	U.S.	0.0	0.3	0.7	0.4	0.5	0.1	0.7	0.3	0.38
	Germany	0.0	0.1	0.5	0.3	0.3	0.1	0.5	0.3	0.26

Table III. Estimator performance in terms of fit for the two separate U.S. and German markets, as measured by the root-mean-square error of the measurement noise. The model is estimated separately for the U.S. and for Germany with 2, 3, 4 and 5 principal components of yields. The reported values are in basis points. The sample period is January 1987 to December 2015.

	Separate models		Joint model	
#	U.S.	Germany	U.S.	Germany
1	97.1893	97.2250	89.9414	89.7814
2	99.6709	99.6657	97.8825	97.8953
3	99.9679	99.9766	99.2972	99.2636
4	99.9962	99.9972	99.7138	99.7115
5	99.9998	99.9998	99.8848	99.8983
6	100.0000	100.0000	99.9722	99.9784
7	100.0000	100.0000	99.9928	99.9857
8	100.0000	100.0000	99.9967	99.9974

Table IV. The explanatory power of single country and global factors. The table shows the percentage of the variance of the U.S. and German yields explained by single-country and two-country principal components. The sample period is June 1986 to December 2015.

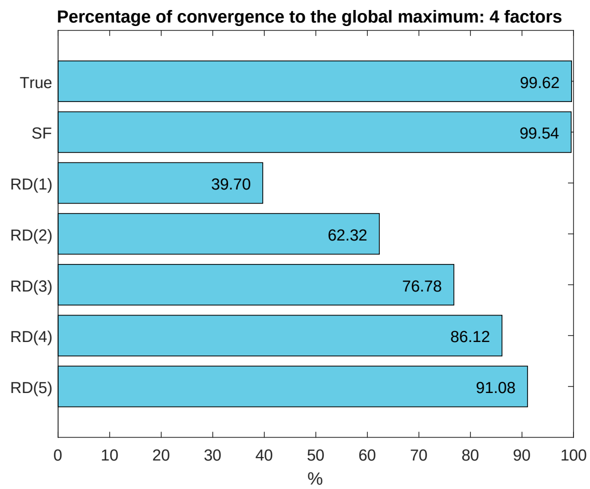
		Root-mean-square error (bp)								
Model	Country	1m	1y	2y	3y	4y	5y	7y	10y	Av.RMSE
		Panel A: $K = 4$								
OLS	U.S.	22.4	11.9	12.9	10.7	7.0	3.7	8.6	18.1	11.90
	Germany	23.4	11.0	14.1	12.3	8.3	4.1	6.1	15.5	11.83
$AACMY$	U.S.	22.4	24.8	22.7	22.4	20.1	15.9	13.3	35.9	22.16
	Germany	23.4	19.2	25.4	32.3	35.3	33.4	19.3	46.1	29.30
$SC(SF)$	U.S.	24.7	17.1	13.5	10.9	8.1	6.3	9.2	19.6	13.68
	Germany	24.9	15.5	15.4	12.5	9.1	6.8	7.5	16.6	13.55
		Panel B: $K = 5$								
OLS	U.S.	11.7	9.8	8.4	6.8	4.8	3.6	6.1	11.6	7.85
	Germany	13.2	7.8	7.6	6.8	4.9	3.1	4.1	9.7	7.18
$AACMY$	U.S.	11.7	14.4	11.8	12.4	12.1	9.6	7.2	17.7	12.09
	Germany	13.2	14.8	12.3	9.2	6.9	6.2	6.8	12.0	10.16
$SC(SF)$	U.S.	12.2	11.8	8.9	7.1	5.3	4.3	6.2	12.3	8.50
	Germany	13.4	9.3	7.8	7.4	5.8	4.0	4.5	10.9	7.89
		Panel C: $K = 6$								
OLS	U.S.	2.9	7.1	2.3	2.9	3.8	3.5	1.6	5.7	3.73
	Germany	2.6	6.8	2.1	2.6	3.2	2.9	1.3	4.3	3.23
$AACMY$	U.S.	2.9	14.6	7.6	4.3	8.2	9.9	6.3	11.4	8.16
	Germany	2.6	13.5	7.8	3.9	4.0	5.0	4.0	5.9	5.83
$SC(SF)$	U.S.	3.6	10.0	3.9	4.5	4.3	3.9	2.8	6.1	4.88
	Germany	2.7	7.6	2.6	3.1	3.3	3.0	1.7	4.3	3.54
		Panel D: $K = 7$								
OLS	U.S.	1.2	3.2	2.1	1.6	1.7	1.9	1.6	2.6	2.00
	Germany	2.1	5.5	1.9	2.3	2.4	2.2	1.3	3.5	2.65
$AACMY$	U.S.	1.2	10.6	6.8	5.5	4.3	3.5	2.3	3.2	4.66
	Germany	2.1	9.3	2.3	2.5	2.7	2.8	2.6	3.8	3.51
$SC(SF)$	U.S.	1.4	4.2	3.0	1.9	1.9	2.0	1.8	2.6	2.35
	Germany	2.4	6.3	2.1	2.7	2.6	2.3	1.6	3.8	2.97
		Panel E: $K = 8$								
OLS	U.S.	0.3	1.8	2.0	1.3	0.5	1.4	1.6	1.5	1.28
	Germany	0.1	1.3	1.9	0.9	0.6	1.4	1.2	1.5	1.12
$AACMY$	U.S.	0.3	8.5	6.8	6.5	4.6	2.9	2.1	1.9	4.20
	Germany	0.1	8.3	3.5	2.1	1.4	2.0	2.1	2.0	2.68
$SC(SF)$	U.S.	0.3	2.2	2.9	1.4	1.0	1.6	1.8	1.7	1.61
	Germany	0.1	1.5	2.3	1.0	0.9	1.5	1.3	1.6	1.26

Table V. Estimator performance in terms of fit for the joint two-country model, as measured by the root-mean-square error of the measurement noise. The model is estimated with 4, 5, 6, 7 and 8 principal components of yields. The reported RMSE values are in basis points. The sample period is January 1987 to December 2015.

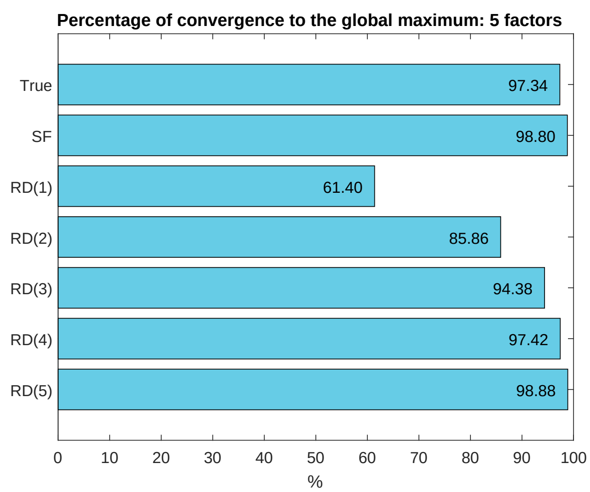
	joint	U.S. only	German only
Max(eig$\{\Phi^P\}$)	0.9945	0.9918	0.9992
ρ_1	0.9925	0.9915	0.9923
ρ_{12}	0.8746	0.8497	0.8672
ρ_{24}	0.7403	0.6676	0.7195
ADF	-0.0026	-0.0034	-0.0023
$KPSS$	15.5124	14.6489	14.0887

Table VI. The table shows time series (physical) properties of the principal components of the joint model and U.S. and German only yields. The first row shows the maximum eigenvalue of the feedback matrix under the physical measure for the joint and individual models. In the following three rows the table reports the first, 12th and 24th order autocorrelation of the first principal component. The last two rows present the ADF test and $KPSS$ test for unit root and stationarity, respectively with one lag. The 10% critical value for the ADF test with no intercept and no time trend amounts to -2.5657. The 1% critical value for the $KPSS$ test amounts to 0.739.

Figures



(a)



(b)

Figure 1. 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 Φ matrix estimated by the spot-forward regressions ('SF') 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.

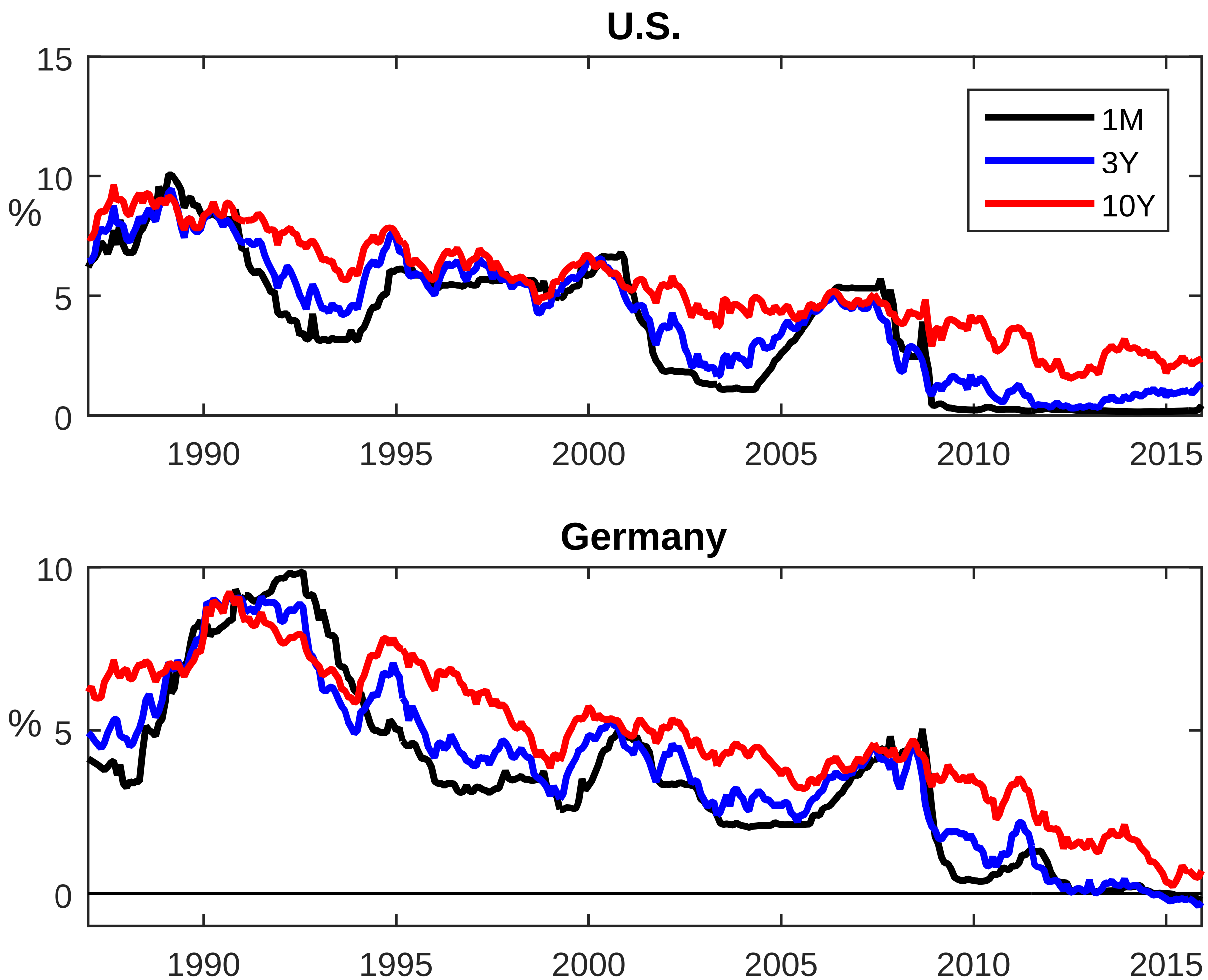


Figure 2. U.S. and German Treasury markets. This figure shows yield data for three of the maturities used in the two country model. For the U.S. the annual maturities come from Gurkaynak, Sack, and Wright (2007) estimated using the Svensson (1994) method. For Germany, we use estimates published by the Bundesbank, also estimated using the Svensson method.

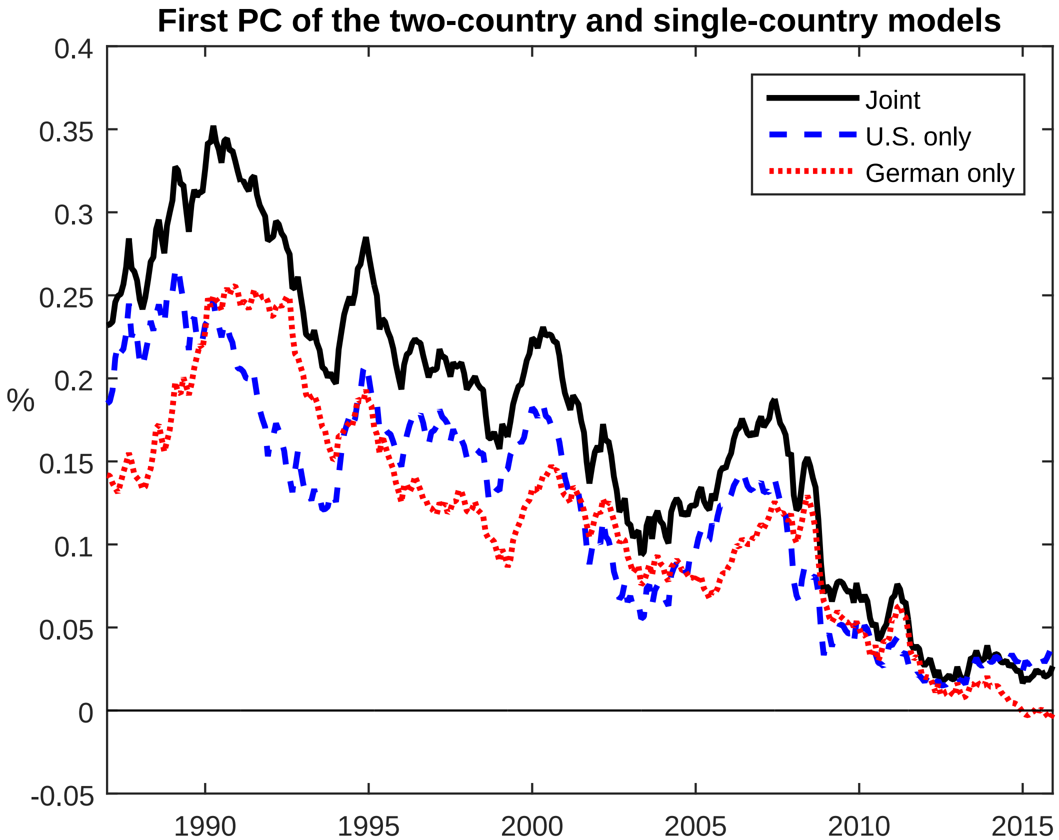


Figure 3. The levels of yields in the U.S. and German markets. This figure shows the first principal components for the two 3−factor U.S. and German models and for the joint 6−factor model. These are similar, exhibiting the strong downward tend in the level of interest rates in the two countries, but there is a tendency for the U.S. to lead the German market. The continuous black line shows the first common factor in the joint model and is a compromise between the level factors in the two single country models. Their differences are reflected in the second common factor, shown in top panel of Figure 5.

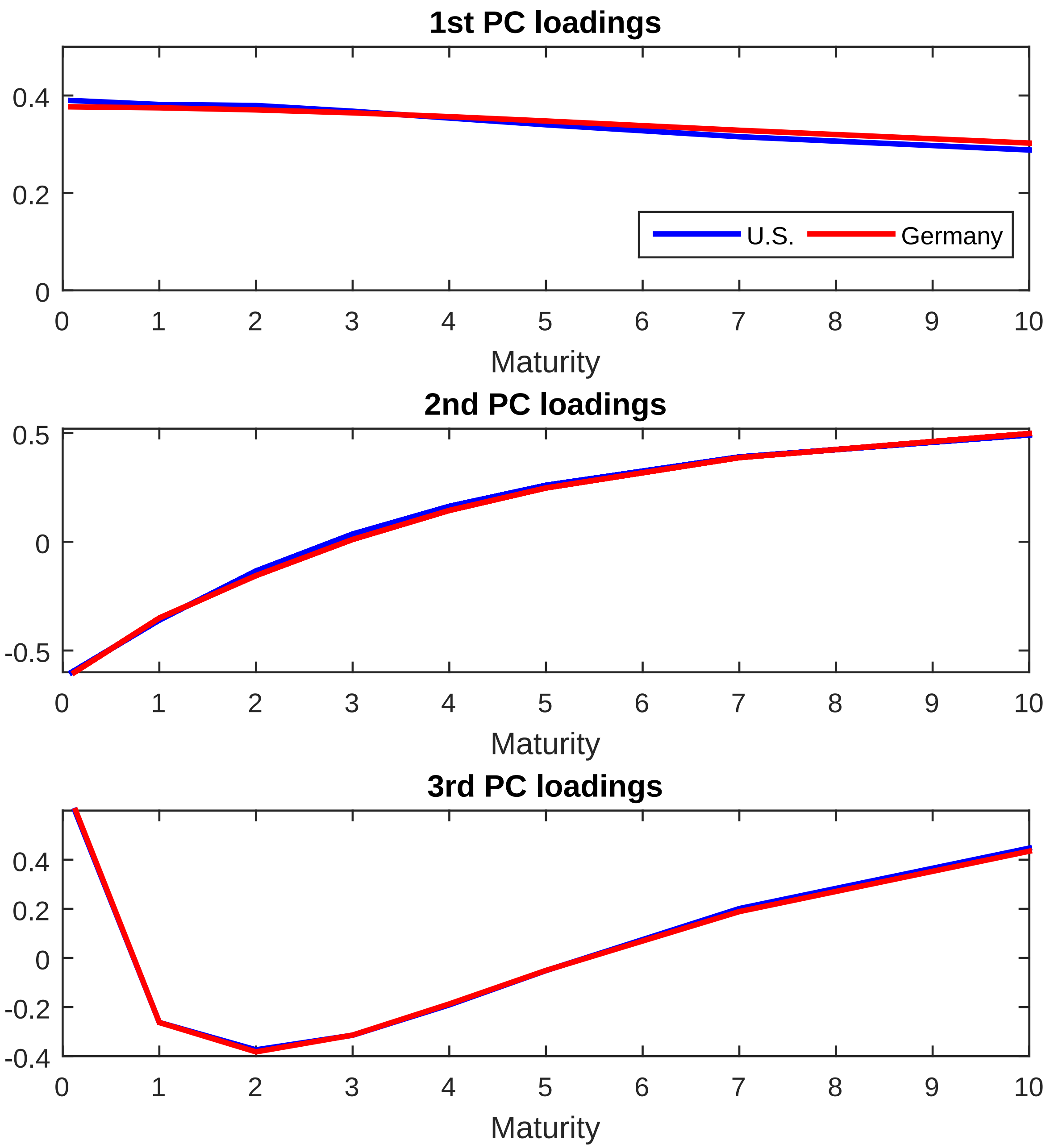


Figure 4. Factor loadings in the separate U.S. and German market models. This figure shows the factor loadings for the two separate 3−factor country models. These are remarkably close, indicating that a common factor model is appropriate.

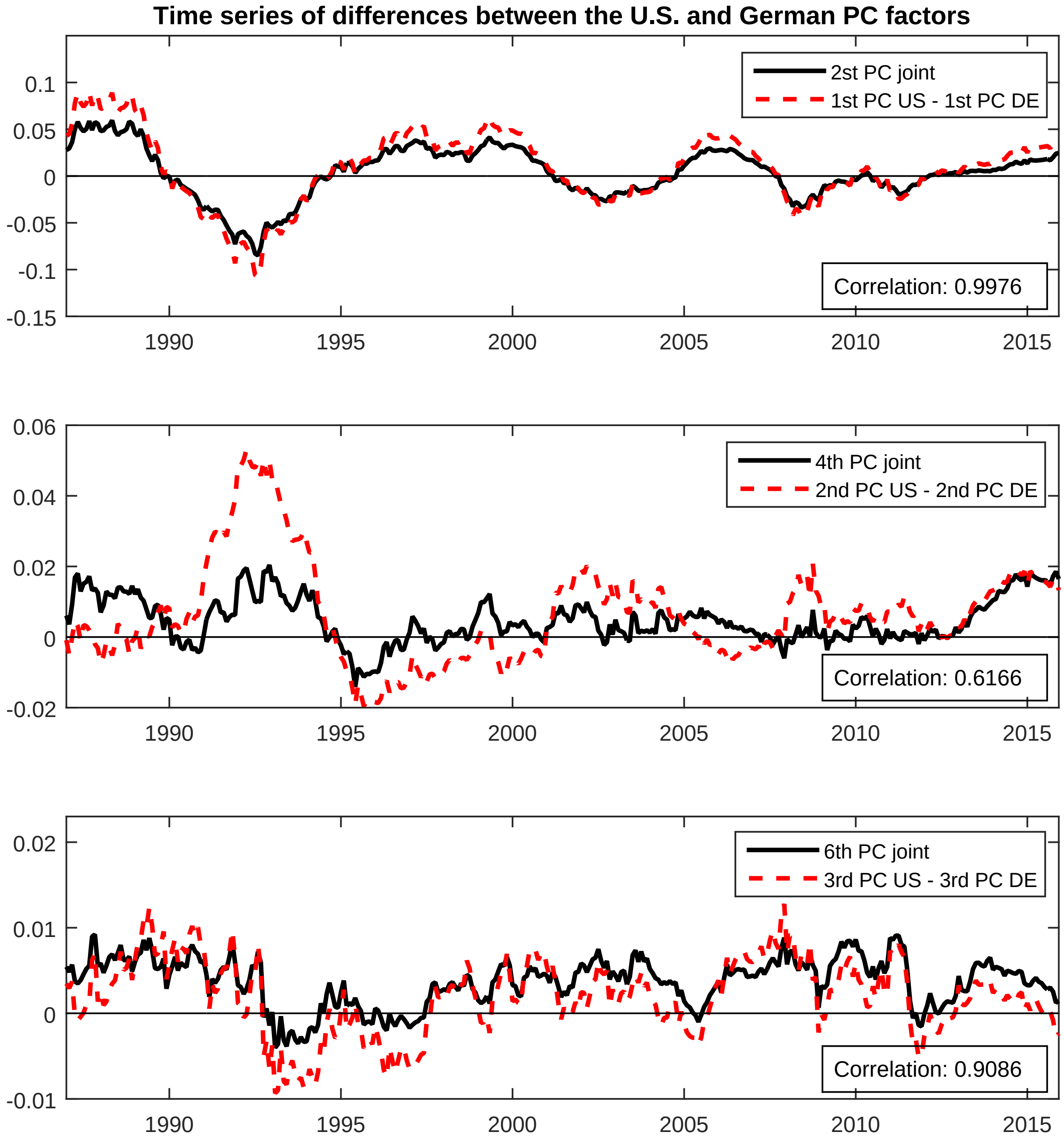


Figure 5. The relationships between the factors in the separate and joint U.S.-German models. This figure shows how the even-numbered common factors allow for relative effects missed by the associated odd-numbered common factors. So for example, the top panel shows how the second common factor is aligned with the difference between the level factors in the two single country models and thus complements the first common factor shown in Figure 3. Thus, we find that regressing the level factors in the single country models on the first two common factors gives an almost perfect fit. Similarly, the third and fourth (fifth and sixth, respectively) factors in the joint model explain most of the variance in the slope (curvature) factors in the single country models. The loadings of the yields on these factors are shown in Figure 6.

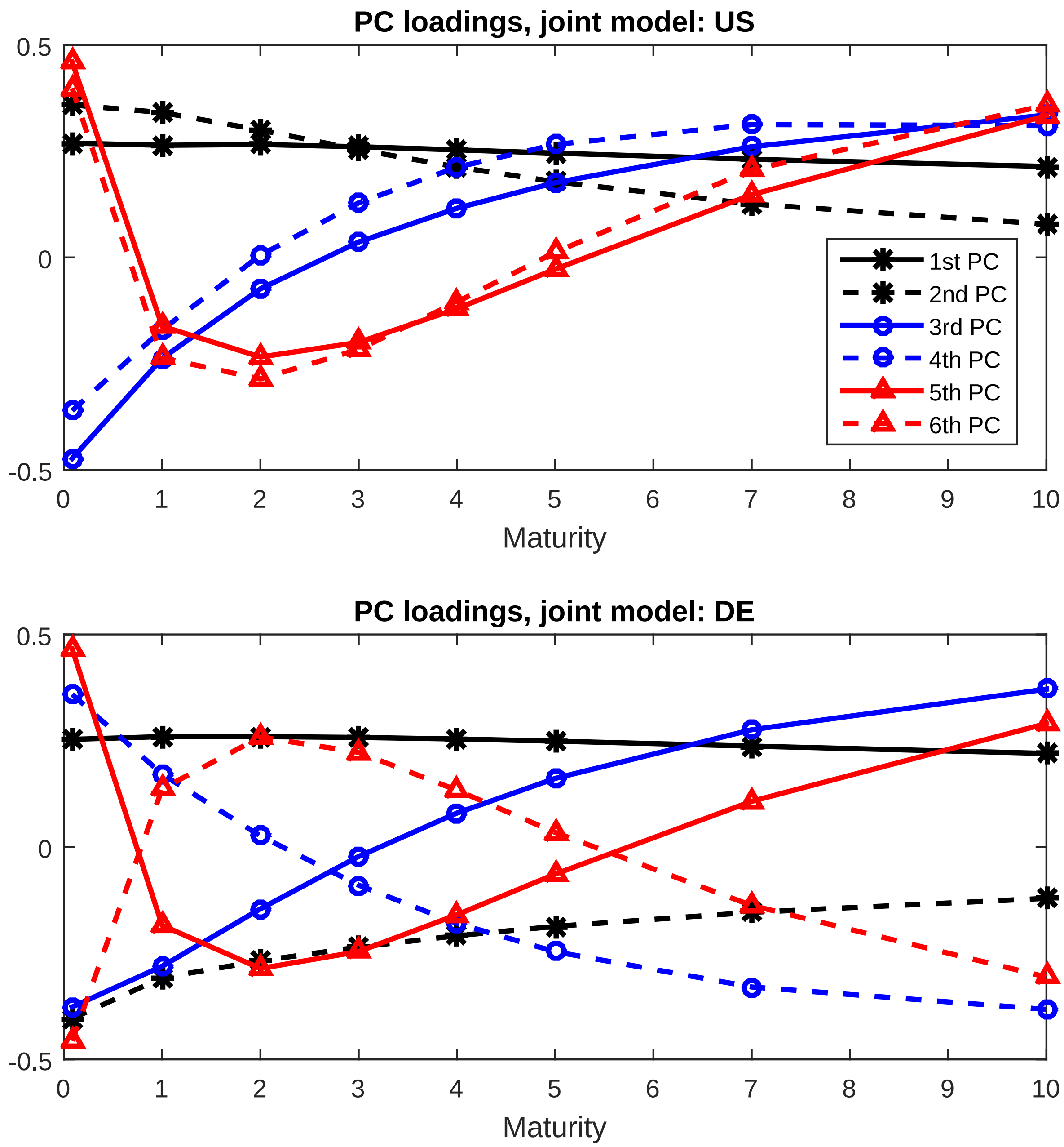
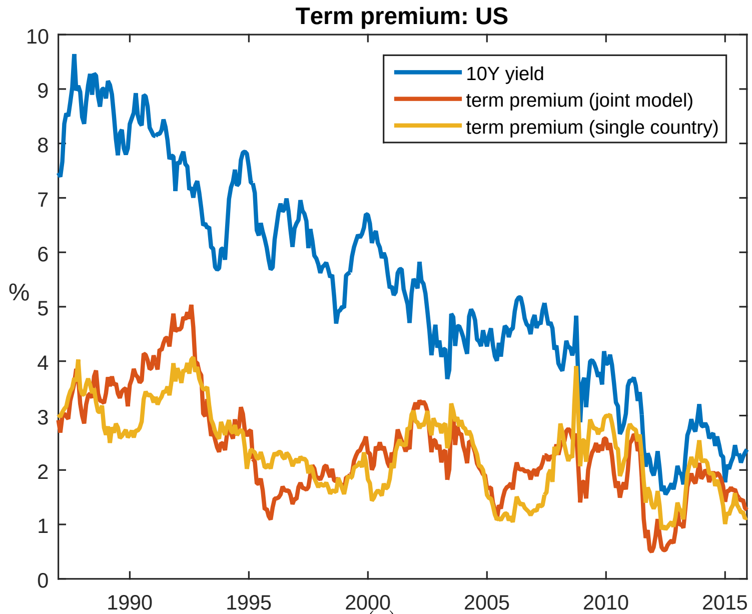
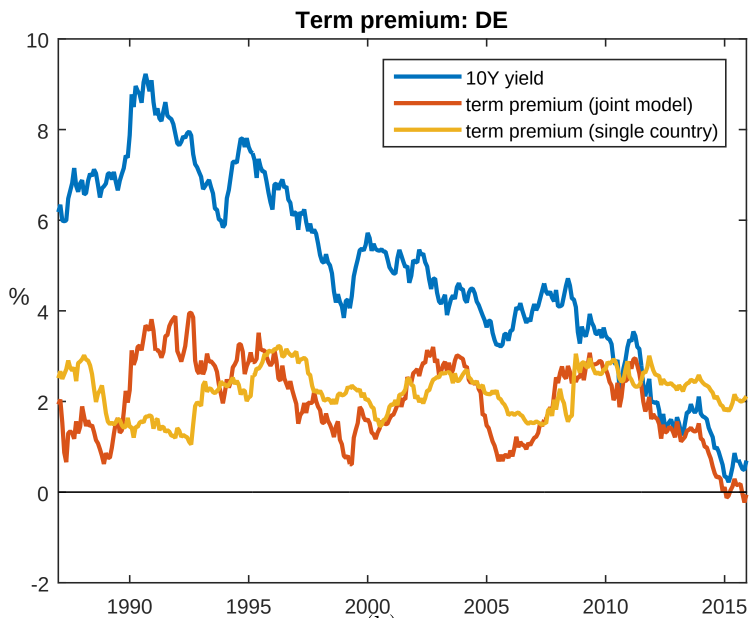


Figure 6. The relationships between the factor loadings in the separate and joint U.S.-German models. The pair-wise factor-rotation relationships shown in Figure 5 are reflected in the factor loadings for the two countries in the 6 common factor model shown in this figure. The continuous lines show the loadings on the odd-numbered factors, which are interpreted as common level (black), slope (blue) and curvature (red) factor loadings respectively. These clearly resemble the respective loadings in the single country models shown in Figure 4. The loadings on the associated even-numbered common factors are shown with dashed lines. Since these allow for relative level, slope and curvature effects, the signs are reversed for the U.S. and Germany.



(a)



(b)

Figure 7. 10-year yield decomposition. This figure shows shows how these models decompose the 10−year yield into components representing 10−year interest rate expectations and a risk premium for the U.S. in Panel (a) and Germany in Panel (b). This decomposition largely depends upon the physical dynamics represented by the time series VAR, Eq.(9) which is used to forecast the interest rates and hence the yields that would be observed if investors were risk-neutral. Subtracting this estimate from the model-fitted yield for any maturity then gives the estimate of the risk premium.