FEDERAL RESERVE BANK OF SAN FRANCISCO

WORKING PAPER SERIES

Unbiased Estimate of Dynamic Term Structure Models

Michael D. Bauer
Federal Reserve Bank of San Francisco

Glenn D. Rudebusch
Federal Reserve Bank of San Francisco

Jing (Cynthia) Wu
Department of Economics, University of California, San Diego

April 2011

Working Paper 2011-12
http://www.frbsf.org/publications/economics/papers/2011/wp11-12bk.pdf

Unbiased Estimation of Dynamic Term Structure Models

Michael D. Bauer*, Glenn D. Rudebusch†, Jing (Cynthia) Wu‡

April 4, 2011

Abstract

Affine dynamic term structure models (DTSMs) are the standard finance representation of the yield curve. However, the literature on DTSMs has ignored the coefficient bias that plagues estimated autoregressive models of persistent time series. We introduce new simulation-based methods for reducing or even eliminating small-sample bias in empirical affine Gaussian DTSMs. With these methods, we show that conventional estimates of DTSM coefficients are severely biased, which results in misleading estimates of expected future short-term interest rates and long-maturity term premia. Our unbiased DTSM estimates imply risk-neutral rates and term premia that are more plausible from a macro-finance perspective.

Keywords: small-sample bias correction, vector autoregression, dynamic term structure models, term premium

JEL Classifications: C53, E43, E47

*Federal Reserve Bank of San Francisco, michael.bauer@sf.frb.org
†Federal Reserve Bank of San Francisco, glenn.rudebusch@sf.frb.org
‡Department of Economics, University of California, San Diego, jingwu@ucsd.edu

1 Introduction

In affine Gaussian dynamic term structure models (DTSMs)—the standard finance representation of the yield curve—the factors underlying yields follow vector-autoregressive (VAR) dynamics with normally distributed errors. Such DTSMs have been widely estimated by maximum likelihood (ML) methods; however, as shown by Joslin et al. (2010a), ML estimates of an unrestricted Gaussian DTSM exactly recover the ordinary least squares (OLS) estimates for the dynamic system. This equivalence result highlights the issue of estimation bias, as it is well known that OLS estimates are often misleading indicators of the true values of the parameters of a dynamic system in finite samples. In particular, OLS estimates will generally be biased toward a dynamic system that displays much less persistence than the true process. Similarly, conventional estimates of DTSMs likely suffer from substantial small-sample bias that significantly distorts economic inference.[1] For example, if the degree of interest rate persistence is underestimated, expected future short rates would revert too quickly to the mean, resulting in spuriously stable estimates of risk-neutral rates. Moreover, the estimation bias would also contaminate estimates of long-maturity term premia. Therefore, in this paper, we introduce new simulation-based methods for reducing or even eliminating the small-sample bias in empirical affine Gaussian DTSMs. Our results from applying these methods show that conventional estimates of DTSM coefficients are severely biased and that the associated empirical decompositions of forward rates into the risk-neutral forward rates and term premia are incorrect. Our unbiased DTSM estimates imply risk-neutral rates and term premia that are more plausible from a both a statistical and macro-finance perspective.

Despite worries about small-sample DTSM bias noted by some researchers, there has been no systematic examination of the likely bias in DTSM estimates in the literature because of computational limitations. Standard ML estimation methods have proved to be intensive, “hands-on” procedures because these models exhibit relatively flat likelihood surfaces with many local optima. The ML computational burden has effectively precluded simulation studies of the DTSM finite-sample estimation properties.[2] However, recent work has shown that OLS can be used to estimate some or all of the parameters of a DTSM. Specifically, for a canonical affine Gaussian DTSM in which some linear combinations of yields are priced

[1] The OLS autoregressive bias is particularly severe when the estimation sample is short and the dynamic process is very persistent. Empirical DTSMs are invariably estimated under just such conditions, with data samples that contain only a limited number of highly persistent interest rate cycles. Importantly, the effective sample size relevant for assessing bias is not simply based on the number of observations. For a given temporal range of data, using a higher sampling frequency, say, daily instead of weekly, will not significantly reduce the OLS estimation bias (e.g., Pierse and Snell, 1995).

[2] Duffee and Stanton (2008) provide the most closely related analysis.

without error, Joslin et al. (2010a) show that simple OLS recovers the ML estimates of the parameters of the factor VAR. Furthermore, Hamilton and Wu (2010) show that *any* such DTSM, even if overidentifying restrictions are imposed, can be estimated by first obtaining reduced-form parameters using OLS, and then calculating all structural parameters via minimum-chi-squared estimation. We exploit these results to devise a procedure to estimate affine Gaussian DTSMs without bias and quantify the severe finite-sample bias in conventional DTSM estimates.

Our bias correction procedure corrects the small-sample bias in estimates of the parameters in a VAR system, which is the relevant dynamic system underlying most DTSMs.[3] In particular, we propose an inverse bootstrap bias correction that finds the data-generating process that leads to a mean/median of the OLS estimator equal to the original OLS estimates. In this way, our correction can completely eliminate estimation bias.[4] We provide simulation results that illustrate the advantages of our approach in comparison to standard estimates.[5] Our approach to bias correction builds on earlier contributions in the econometrics literature, which invert the mapping from the parameters for a data-generating process (DGP) to the central tendency (mean or median) of the OLS estimator. Notably, Rudebusch (1992) developed a median-unbiased estimation method for the parameters in a univariate AR(p) model in which one simulates repeatedly from a DGP, trying different parameter values, until the element-wise median of the OLS estimator is close to the original estimates.[6] In this paper, we extend this approach to multivariate (VAR) models and introduce a new algorithm which efficiently and reliably solves the numerical problem of inverting the relevant mapping.

We provide two case studies that demonstrate how our VAR results generalize to the unbiased estimation of DTSMs. In these case studies, we describe the extent of small-sample bias in the conventional estimates, provide unbiased estimates, and assess the implications for expected short rates and term premia. The first case study examines the canonical DTSM

[3]The statistical literature has developed analytical approximations for the mean bias in univariate autoregressions (e.g., Marriott and Pope (1954) and Stine and Shaman (1989)) and VARs (e.g., Nicholls and Pope (1988) and Yamamoto and Kunitomo (1984)). Examples of analytical applications include Rudebusch (1993), Amihud and Hurvich (2004), and Engsted and Pedersen (2008). Nowadays, for generality and flexibility, researchers usually employ simulation-based methods for multivariate bias correction as we do in this paper.

[4]Our method can completely eliminate *median bias*, which is what we focus on in our DTSM estimation.

[5]We also compare our results to the conventional bootstrap bias correction, which estimates the bias by using the OLS estimates found from the data as the data-generating parameters. This procedure reduces the bias to order T^{-2}.

[6]Other closely related econometric work includes Andrews (1993), who proposed an exactly median-unbiased estimation method for inferring the autoregressive parameter in an AR(1) model, and Andrews and Chen (1994), who extended this method to obtain approximately median-unbiased estimates of the sum of the AR coefficients in an AR(p) model. Tanizaki (2000) used a similar simulation-based procedure to obtain mean- and median-unbiased estimates of regression models with lagged dependent variables.

of Joslin et al. (2010a). Using the authors' same model specifications and data samples, we quantify the bias in the reported parameter estimates and describe the differences in the empirical results when the parameters governing the factor dynamics are replaced with unbiased estimates. The second case study presents a DTSM with overidentifying restrictions, which precludes estimation of the factor VAR by OLS. Unbiased estimation is achieved by adapting the consistent two-stage estimation procedure of Hamilton and Wu (2010). We first obtain both OLS and unbiased reduced-form parameter estimates and then map these into the model's structural parameters, thus uncovering the bias in a DTSM that is *not* canonical. Our results demonstrate that in each case, unbiased estimates of expected policy rates (risk-neutral rates) and term premia differ in statistically and economically significant ways from conventional estimates. The highly persistent, near-integrated nature of interest rates leads to a dramatic overestimation of the speed of short rate mean reversion in these DTSMs. Our results show that researchers and policy makers who analyze movements in interest rates are well advised to use bias-corrected estimators when trying to infer policy expectations and risk premia.

There are a number of papers in the literature that are related to our attempt to reduce the bias in DTSM parameter estimates. Notably, Jardet et al. (2009) use a "near-cointegrated" specification of the dynamic system, which amounts to obtaining estimates by averaging a stationary and a cointegrated specification. Another approach is to use the information in the cross section of interest rates to pin down the parameters of the dynamic system, by restricting the risk pricing (Bauer, 2011; Joslin et al., 2010b). Still another alternative is to use survey forecasts as additional information to incorporate in the model (Kim and Orphanides, 2005; Kim and Wright, 2005). While all of these studies can potentially reduce the mean-reversion bias to some extent, they neither quantify the bias nor provide evidence as to how much it is reduced. To our knowledge no existing study has quantified the bias in estimates of a DTSM, or even in a simple interest rate VAR.

Our paper is structured as follows: Section 2 defines the affine term structure model and discusses how recent advances in DTSM estimation make it possible to perform unbiased estimation of the model. In Section 3, we discuss the small-sample bias in OLS estimates of VAR parameters and ways to reduce this bias, including our new method, inverse bootstrap bias correction. Then we apply our method in two case studies. Section 4 demonstrates how bias correction changes the estimates of the model of Joslin et al. (2010a) and the implied risk-neutral rates and term premia. In Section 5 we perform a simulation study tailored to the JSZ case study to systematically assess the value of bias correction. Finally, in Section 6, we use the framework of Hamilton and Wu (2010) to compare least-squares and unbiased estimation of a model that is not canonical. Section 7 concludes.

2 Estimation of affine Gaussian DTSMs

In this section, we describe the specification and estimation of affine DTSMs with a focus on recent advances in estimation methodology which make unbiased estimation possible. We also discuss the assumption of stationarity that underlies essentially all implementations of affine models.

2.1 Model specification

The discrete-time affine Gaussian DTSM, the workhorse model in the term structure literature since Ang and Piazzesi (2003), has three key elements. First, a vector of N pricing factors, X_t, follows a first-order Gaussian VAR:

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

where $\varepsilon_t \overset{iid}{\sim} N(0, I_N)$ and Σ is lower triangular. Second, the short rate, r_t, is an affine function of the pricing factors:

$$r_t = \delta_0 + \delta_1' X_t. \tag{2}$$

Third, the stochastic discount factor (SDF) that prices all assets under the absence of arbitrage is of the essentially affine form (Duffee, 2002):

$$-\log(M_{t+1}) = r_t + \frac{1}{2}\lambda_t'\lambda_t + \lambda_t'\varepsilon_{t+1},$$

where the N-dimensional vector of risk prices is affine in the pricing factors,

$$\lambda_t = \lambda_0 + \lambda_1 X_t,$$

for N-vector λ_0 and $N \times N$ matrix λ_1.

With these three elements, it is well-known that a risk-neutral probability measure $\mathbb{Q}$ exists such that the price of an m-period default-free zero coupon bond is $P_t^m = E_t^{\mathbb{Q}}(e^{-\sum_{h=0}^{m-1} r_{t+h}})$. Furthermore, under $\mathbb{Q}$, the pricing factors follow a first-order Gaussian VAR,

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

and the prices of risk determine how the change of measure affects the VAR parameters:

$$\mu^{\mathbb{Q}} = \mu - \Sigma\lambda_0 \quad \Phi^{\mathbb{Q}} = \Phi - \Sigma\lambda_1. \tag{4}$$

Bond prices are exponentially affine functions of the pricing factors:

$$P_t^m = e^{\mathcal{A}_m + \mathcal{B}_m X_t},$$

and the loadings $\mathcal{A}_m = \mathcal{A}_m(\mu^{\mathbb{Q}}, \Phi^{\mathbb{Q}}, \delta_0, \delta_1, \Sigma)$ and $\mathcal{B}_m = \mathcal{B}_m(\Phi^{\mathbb{Q}}, \delta_1)$ follow the recursions

$$\begin{aligned} \mathcal{A}_{m+1} &= \mathcal{A}_m + (\mu^{\mathbb{Q}})'\mathcal{B}_m + \frac{1}{2}\mathcal{B}_m'\Sigma\Sigma'\mathcal{B}_m - \delta_0 \\ \mathcal{B}_{m+1} &= (\Phi^{\mathbb{Q}})'\mathcal{B}_m - \delta_1 \end{aligned}$$

with starting values $\mathcal{A}_0 = 0$ and $\mathcal{B}_0 = 0$. Model-implied yields are determined by $y_t^m = -m^{-1}\log P_t^m = A_m + B_m X_t$, with $A_m = -m^{-1}\mathcal{A}_m$ and $B_m = -m^{-1}\mathcal{B}_m$. Risk-neutral yields, the yields that would prevail if investors were risk-neutral, can be calculated using

$$\tilde{y}_t^m = \tilde{A}_m + \tilde{B}_m X_t, \quad \tilde{A}_m = -m^{-1}\mathcal{A}_m(\mu, \Phi, \delta_0, \delta_1, \Sigma), \quad \tilde{B}_m = -m^{-1}\mathcal{B}_m(\Phi, \delta_1).$$

Risk-neutral yields reflect policy expectations over the lifetime of the bond, $m^{-1}\sum_{h=0}^{m-1} E_t r_{t+h}$, plus a time-constant convexity term.[7] The yield term premium is defined as the difference between actual and risk-neutral yields, $ytp_t^m = y_t^m - \tilde{y}_t^m$.

It is often convenient to consider forward rates instead of yields, in particular when considering long-maturity properties of interest rates. Model-implied forward rates for loans starting at $t+n$ and maturing at $t+m$ are given by $f_t^{n,m} = (m-n)^{-1}(\log P_t^n - \log P_t^m) = (m-n)^{-1}(my_t^m - ny_t^n)$. Risk-neutral forward rates $\tilde{f}_t^{n,m}$ are calculated in analogous fashion from risk-neutral yields. The forward term premium is defined as $ftp_t^{n,m} = f_t^{n,m} - \tilde{f}_t^{n,m}$.

With regard to the choice of pricing factors, X_t, one can use unobserved factors (which are then filtered from observed variables), observables such as yields or macroeconomic variables, or any combination of latent and observable factors. In this paper, X_t will always contain the first three principal components of yields, that is, we will focus on "yields-only" models with observable factors. Our methodology, however, is applicable to cases with observable and/or unobservable yield curve factors, as well as to macro-finance DTSMs.[8] Typically, the number

[7]The convexity term equals $-m^{-1}(\mathcal{A}(\mu, \Phi, \delta_0, \delta_1, \Sigma) - \mathcal{A}(\mu, \Phi, \delta_0, \delta_1, 0))$.

[8]Specifically, our estimation method could be applied to such macro-finance models as were used in Ang and Piazzesi (2003), Joslin et al. (2010b), and Jardet et al. (2009). The only assumption that is necessary for our method to be applicable is that N linear combinations of pricing factors are priced without error.

of yields used in estimation of the model will be much larger than the number of pricing factors—after all one purpose of DTSMs is to reduce the dimensionality of the yield curve. To prevent stochastic singularity it is common to introduce measurement error. Denoting by M the number of yields in the data, we will generally assume that N linear combinations of yields are priced without error, and that there are $M - N$ independent measurement errors.

One possible parameterization of the model is in terms of $\theta = (\mu, \Phi, \mu^{\mathbb{Q}}, \Phi^{\mathbb{Q}}, \delta_0, \delta_1, \Sigma)$, leaving aside the parameters determining the measurement error distribution. Given θ, the risk sensitivity parameters λ_0 and λ_1 follow from equation (4). Model identification requires normalizing restrictions (Dai and Singleton, 2000; Hamilton and Wu, 2010). For example, θ has 34 free elements in a three-factor model, but only 22 parameters are identified, so at least 12 normalizing restrictions are necessary. If the model is exactly identified, one speaks of a "canonical" model, as opposed to an overidentified model, where additional restrictions are imposed.

2.2 Estimation with OLS

Maximum likelihood estimation of the model parameters has been found to be very difficult. Specifically, the likelihood function usually has local optima with very different economic implications, and finding a global optimum is challenging.[9] This situation largely reflects the very high persistence of the VAR system in equation (1), resulting from the near-unit-root behavior of interest rates. Consequently, μ and Φ are very hard to estimate in the sense that statistical uncertainty is high and the estimated persistence is biased downwards. Because the likelihood function is very flat in certain parameters it is difficult to find the ML estimates by numerical optimization.

Recent theoretical advances in the DTSM literature, specifically the papers by Joslin, Singleton and Zhu (2010a, henceforth JSZ) and Hamilton and Wu (2010, henceforth HW), show that OLS can be used to significantly simplify the estimation problem. JSZ prove that in a canonical model the ML estimates of μ and Φ can be obtained by OLS. HW estimate a reduced form of the model using OLS and then find the structural parameters by minimizing a chi-squared statistic. Given these methodological innovations, there is hardly any need to maximize a high-dimensional, badly behaved likelihood function in order to estimate a DTSM.[10] Conventional DTSM estimation can in most cases be performed by a consistent

[9]The list of studies that have documented such problems is long and includes Ang and Piazzesi (2003); Duffee and Stanton (2004); Kim and Orphanides (2005); Duffee (2009) and Hamilton and Wu (2010). Also see Christensen et al. (2009) and Christensen et al. (2011), who introduce an arbitrage-free Nelson-Siegel DTSM that can be readily estimated.

[10]In a Kalman filter setting, where all observations are contaminated by measurement error and the assump-

and efficient two-stage estimation procedure, where the first stage consists of simple OLS, and the second stage involves finding the remaining (JSZ) or structural (HW) parameters, without any computational or numerical difficulties. Importantly, the use of linear regressions not only simplifies conventional estimation but also allows for correction of the small-sample bias in the parameter estimates of the underlying dynamic system.[11] In this paper, we show how to incorporate bias correction into the estimation frameworks of JSZ and HW. We now discuss each framework in more detail and describe unbiased estimation of both canonical and overidentified DTSMs.

2.3 Canonical models: the JSZ separation

JSZ suggest a new normalization for canonical DTSMs and use it to derive a separation result that on one hand simplifies ML estimation and on the other hand enables us to perform unbiased estimation. The two key assumptions are that (i) there are no overidentifying restrictions, and (ii) N linear combinations of yields are exactly priced by the model. In that case, any affine Gaussian DTSM is equivalent to one where the pricing factors X_t are taken to be those linear combinations of yields. Furthermore, the distribution under $\mathbb{Q}$ of these pricing factors is uniquely determined by $r^{\mathbb{Q}}_{\infty}$, the risk-neutral unconditional mean of the short rate, $\lambda^{\mathbb{Q}}$, the eigenvalues of $\Phi^{\mathbb{Q}}$, and Σ. That is, any canonical model can be parameterized in terms of $(\mu, \Phi, \Sigma, r^{\mathbb{Q}}_{\infty}, \lambda^{\mathbb{Q}})$, which has come to be called the "JSZ normalization." The intuition is that (i) we parameterize the model in terms of physical dynamics and risk-neutral dynamics, and (ii) the normalizing restrictions are imposed on the $\mathbb{Q}$-dynamics.

This normalization is particularly useful because of the separation result that follows: the joint likelihood function of observed yields can be written as the product of (i) the "$\mathbb{P}$-likelihood," the conditional likelihood of X_t, which depends only on (μ, Φ, Σ), and (ii) the "$\mathbb{Q}$-likelihood," the conditional likelihood of the yields, which depends only on $(r^{\mathbb{Q}}_{\infty}, \lambda^{\mathbb{Q}}, \Sigma)$ and the parameters for the measurement errors.[12] Because of this separation the values of (μ, Φ) that maximize the joint likelihood function are the same as the ones that maximize the $\mathbb{P}$-likelihood. This gives rise to a simple two-step estimation procedure. In the first step, OLS is used to estimate the parameters governing the VAR for X_t. Denote the OLS

tions of JSZ and HW do not hold, these methods will generally be helpful in finding excellent starting values for ML estimation.

[11] Simulation-based bias correction methods become feasible because obtaining regression estimates on each simulated data set only takes a fraction of a second. In contrast, repeated ML estimation of a DTSM on simulated data has a prohibitive computational cost.

[12] There are $M - N$ independent measurement errors, which JSZ assume to have equal variance. This error variance is not estimated but concentrated out of the likelihood function.

estimates by $(\hat{\mu}, \hat{\Phi})$. In the second step, the remaining parameters, $(r^{\mathbb{Q}}_{\infty}, \lambda^{\mathbb{Q}}, \Sigma)$, are estimated by maximizing the joint likelihood function, fixing the values of μ and Φ at $\hat{\mu}$ and $\hat{\Phi}$. This two-step procedure exactly recovers the ML estimate but is numerically much simpler than finding the optimum of the joint likelihood function of all model parameters.[13]

The ML estimates of μ and Φ suffer from the small-sample bias that plagues all least squares estimates of autoregressive systems. In Section 3, we will elaborate on this bias and describe procedures to obtain reduced-bias or even unbiased estimates of μ and Φ. Assume for now that in addition to the OLS estimates we have available some estimates ($\tilde{\mu}$, $\tilde{\Phi}$) that we deem superior in terms of small-sample bias properties. Because of the JSZ separation result, our first-step estimates are independent of the parameter values that in the second step maximize the joint likelihood function. Differently put, we do not have to worry in the first step about cross-sectional fit. Based on this insight, we suggest the following estimation procedure: In the first step, obtain bias-corrected estimates of the VAR parameters. In the second step, maximize the joint likelihood function over $(r^{\mathbb{Q}}_{\infty}, \lambda^{\mathbb{Q}}, \Sigma)$, fixing the values of μ and Φ at $\tilde{\mu}$ and $\tilde{\Phi}$. This procedure will take care of the small-sample estimation bias for the parameters governing the dynamic system, while achieving the same cross-sectional fit as ML estimation.

In Section 4 we will demonstrate the practical consequences of bias correction for JSZ's model specification and data set, comparing results obtained using their original estimation method and this alternative bias-adjusted procedure. Moreover, independent of which estimation procedure a researcher uses, the JSZ normalization is convenient and intuitive, thus we adopt it throughout this paper.

2.4 Overidentified models: the HW approach

For models that impose overidentifying restrictions the JSZ separation result does not hold. However, many DTSM studies have imposed zero restrictions or other kinds of restrictions on model parameters.[14] How can we perform unbiased estimation in such models?

HW show that for *any* affine Gaussian DTSM that exactly prices N linear combinations of yields, all the information in the data can be summarized by the parameters of a reduced-form system, which takes the form of two sets of equations: a VAR for the exactly priced

[13]Note that Σ not only affects the $\mathbb{P}$-likelihood but also the $\mathbb{Q}$-likelihood, since it enters the affine loadings. Thus we cannot use the estimate of Σ from the VAR estimation, $\hat{\Sigma}$, but have to include it in the optimization in the second step. However, $\hat{\Sigma}$ will usually be an excellent starting value.

[14]In fact, almost every study that uses a DTSM imposes some restrictions. Among the most prominent examples are Ang and Piazzesi (2003), Kim and Orphanides (2005), Kim and Wright (2005), Cochrane and Piazzesi (2008), Rudebusch and Wu (2008) and Joslin et al. (2010b).

linear combinations of yields Y_t^1 and a contemporaneous regression equation for the linear combinations of yields Y_t^2 that are priced with error,

$$\begin{aligned} Y_t^1 &= \mu_1 + \Phi_1 Y_{t-1}^1 + u_t^1, & (5) \\ Y_t^2 &= \mu_2 + \Phi_2 Y_t^1 + u_t^2. & (6) \end{aligned}$$

We denote $Var(u_t^1) = \Omega_1$ and $Var(u_t^2) = \Omega_2$. Here, as in most DTSM models, measurement errors are taken to be uncorrelated, so Ω_2 is diagonal. In this paper, we focus on the case where the DTSM pricing factors are those linear combinations of yields that are exactly priced. Thus $Y_t^1 = X_t$, $\mu_1 = \mu$, $\Phi_1 = \Phi$, and $\Omega_1 = \Sigma\Sigma'$.

HW suggest a consistent two-step procedure for DTSM estimation. In the first step, one obtains estimates of the reduced-form parameters by OLS. In the second step, the structural model parameters are found via minimum-chi-squared estimation: a chi-squared statistic measures the distance between the estimates of the reduced-form parameters and the values implied by the structural parameters, and it is minimized via numerical optimization. This two-step procedure is asymptotically equivalent to ML estimation but greatly reduces the computational problems. The numerical optimization in the second step is generally faster and more reliable than for ML estimation: The objective function is a simple distance metric, and calculating it does not require iterating through the data as is necessary for calculating a likelihood function. This speeds up the optimization by an order of magnitude and essentially eliminates the possibility of local optima. For some specifications estimation is even simpler: if (i) the model is canonical and (ii) only one linear combination of yields is included in Y_t^2 then the number of structural parameters is equal to the number of reduced-form parameters and the second step becomes almost trivial. In this case minimum-chi-squared estimation and ML estimation are numerically identical. We will use such a specification, which HW call exact identification, for a simulation study in Section 5.

Importantly, the HW procedure is applicable to both canonical and overidentified DTSMs. The procedure relies on regressions to summarize the information in the data, so it permits unbiased estimation of models that impose overidentifying restrictions. Our procedure for bias correction replaces the OLS estimates of the VAR system in equation (5) with bias-corrected parameter estimates. For the contemporaneous regression in equation (6), which has no lagged dependent variables, we continue to employ OLS since it will deliver unbiased estimates of (μ_2, Φ_2). Having obtained bias-corrected estimates of the reduced-form parameters, we perform the second stage of the estimation as before, minimizing the chi-squared distance statistic. This estimation method will deliver estimates of the structural model parameters that are free of small-sample bias, and it is applicable to a large variety of DTSMs in the literature.

We will apply this approach in the simulation study in Section 5, where we assess the implications of bias correction for the accuracy of term premium estimation in a Monte Carlo setting, as well as in the case study in Section 6, where we estimate a model with risk price restrictions on real-world data.

2.5 Stationarity of interest rates

The stationarity of yields in estimated affine DTSMs is an important issue. The persistence of interest rates is sufficiently high so that unit root tests typically cannot reject the null of a stochastic trend. However, most economists would be uncomfortable with an integrated specification since nominal interest rates generally do not turn negative and remain within a limited range. Furthermore, a largest root that exceeded one would be even more problematic, with unreasonable explosive short rate forecasts. For these reasons, empirical DTSM assumption almost invariably assume stationarity by implicitly or explicitly imposing the constraint that all roots of the factor VAR are less than one in absolute value.[15] Similarly, we assume that interest rates are mean-reverting—albeit, a reversion that is very slow and hard to detect statistically.

Conventional DTSM estimates (ML or OLS) typically imply a largest root for the factor VAR that is comfortably below one. This stationarity result is not surprising given that the small-sample distribution of the OLS estimator implies a downward bias for the largest root. Thus, even though the true value of the largest root might be very close to one, nonstationary estimates almost never occur, and the stationarity restriction is typically not binding. Bias correction on the other hand, by recentering the distribution of the estimator around the true values, leads to a more frequent occurrence of nonstationary roots. A simulation study in Section 3.4 will make this evident. The stationarity restriction imposed on DTSM estimates thus becomes more relevant: it will be binding much more often than for OLS/ML estimation. We discuss ways to deal with this in Section 3.3.4.

3 Unbiased Estimation of VAR Models

As a first step to unbiased DTSM estimation, this section describes the small-sample bias in OLS estimates of VAR systems and ways to reduce or even eliminate that bias. Specifically,

[15]Prominent examples are Dai and Singleton (2002); Duffee (2002) and Ang and Piazzesi (2003). The stationarity prior is made explicit in the Bayesian frameworks of Ang et al. (2009) and Bauer (2011).

we review the conventional bootstrap bias correction and introduce a new method that we call inverse bootstrap.

3.1 VAR models and finite-sample bias

Consider the VAR system in equation (1). We focus our exposition on a first-order VAR since the extension to higher order models is straightforward. For future reference, denote the innovations by $u_t = \Sigma\varepsilon_t \overset{iid}{\sim} N(0, \Sigma\Sigma')$. We assume that the VAR is stationary, i.e., that all the eigenvalues of Φ are less than one in modulus. The parameters of interest are $\theta = \text{vec}(\Phi)$. Denote the true values by θ_0. Consistent estimates of θ_0 can be obtained by applying OLS to each equation of the system, which is numerically equivalent to performing ML estimation (Hamilton, 1994, chap. 11.1).[16] Let $\hat{\theta}_T$ denote these OLS estimates. Because of the presence of lagged endogenous variables, the assumption of strict exogeneity is violated and the OLS estimator is biased in finite samples, i.e., $E(\hat{\theta}_T) \neq \theta_0$. The *bias function* $b_T(\theta) = E(\hat{\theta}_T) - \theta$ relates the bias of the OLS estimator to the value of θ.[17]

The small-sample bias in $\hat{\theta}_T$ is more severe the shorter the available sample and the more persistent the process is (see, for example, Nicholls and Pope, 1988). The data samples used in estimation of DTSMs are relatively short given the high persistence of the interest rate series, so the bias is potentially quite sizeable.[18] Any such bias would tend to underestimate the persistence, as measured for example by the largest eigenvalue of Φ. In this case, forecasts of X_t or of linear combinations of X_t (such as the short rate in DTSMs) would revert to their unconditional mean too quickly.

An alternative for defining bias is to consider the median instead of the mean as the relevant central tendency of the OLS estimator. Some authors, including Andrews (1993) and Rudebusch (1992), have argued that median-unbiased estimators have useful impartiality properties, given that the distribution of the OLS estimator can be highly skewed in autoregressive models for persistent processes. We denote the median bias by $B_T(\theta) = Med(\hat{\theta}_T) - \theta$. Here $Med(Y)$ is the median of random vector Y, which we take to be the element-by-element median as in Rudebusch (1992).[19]

[16] There are two asymptotically equivalent ways to estimate θ: one can either include an intercept in the regressions or one can work with demeaned values. For reasons that will become clear later, we prefer to work with demeaned values. To recover an estimate of the intercept μ, one simply uses the relation $\hat{\mu} = (I_k - \hat{\Phi})\bar{X}$.

[17] Because $\hat{\theta}_T$ is distributionally invariant with respect to μ and Σ, the bias function depends only on θ and not on μ or Σ. The proof of distributional invariance for the univariate case in Andrews (1993) naturally extends to VAR models.

[18] Sampling at higher frequency does not help: intuitively this increases the sample length but also the persistence (Pierse and Snell, 1995).

[19] There are alternative definitions of the median of a random vector since orderings of multivariate observa-

3.2 Bootstrap bias correction

The bootstrap has become a common method for correcting small-sample mean bias.[20] Denote the demeaned observations by $\tilde{X}_t$, and let B denote the number of bootstrap samples. The algorithm for mean bias correction using the bootstrap is as follows:

1. Estimate the model by OLS and save the OLS estimates $\hat{\theta} = \text{vec}(\hat{\Phi})$ and the residuals. Set $b = 1$.

2. Generate bootstrap sample b using the residual bootstrap: Resample the OLS residuals, denoting the bootstrap residuals by u_t^*. Randomly choose a starting value among the T observations. For $t > 1$, construct the bootstrap sample using $\tilde{X}_t^* = \hat{\Phi}\tilde{X}_{t-1}^* + u_t^*$.

3. Calculate the OLS estimates on bootstrap sample b and denote it by $\hat{\theta}_b^*$.

4. If $b < B$ then increase b by one and return to step two.

5. Calculate the average over all samples as $\bar{\theta}^* = B^{-1}\sum_{b=1}^{B}\hat{\theta}_b^*$.

6. Calculate the bootstrap bias-corrected estimate as

$$\tilde{\theta}^B = \hat{\theta} - \left[\bar{\theta}^* - \hat{\theta}\right] = 2\hat{\theta} - \bar{\theta}^*.$$

For large B, the estimated bias $\bar{\theta}^* - \hat{\theta}$ will be close to $b_T(\hat{\theta})$. The motivation for this approach comes from the fact that $E(b_T(\hat{\theta})) = b_T(\theta_0) + O(T^{-2})$, thus we can reduce the bias to order T^{-2} by using this bias correction (Horowitz, 2001).

The bootstrap can also be applied to correct for median bias in a straightforward fashion (although it appears not to have been employed for this purpose before). Denote by m^* the vector stacking the element-wise sample medians of $\{\hat{\theta}_b^*\}_{b=1}^{B}$. The estimate of the median bias $m^* - \hat{\theta}$ will be close to $B_T(\hat{\theta})$ for sufficiently large B.

Thus, we have at our disposal a method to obtain estimates of the parameters of a VAR system with either reduced mean bias or reduced median bias. We can apply it to perform reduced-bias estimation of DTSMs using the approach of either JSZ or HW described in Section 2.2. Just applying this bootstrap to reduce the small-sample bias in estimates of affine Gaussian DTSMs is already a big step forward. However, the bootstrap estimates can

tions are not unique. The one we use here is intuitive, has a straightforward sample analog (the element-wise sample median), and has been used elsewhere (e.g., Meerschaert and Scheffler, 2001, p. 53).

[20] For a detailed exposition see Hall (1992) or Efron and Tibshirani (1993, Chapter 10); for a review of the bootstrap including its application to bias correction, refer to Horowitz (2001). Applications in the time series literature include Kilian (1998) and Tang and Chen (2009).

potentially be improved even further. A likely shortcoming of the bootstrap bias correction is that the bias function is estimated at $\hat{\theta}_T$, whereas the true bias is determined by the bias function at θ_0. One might speak of a plug-in principle being applied here, since the bias is estimated by plugging in the OLS estimate for the unknown parameter value. If the bias were constant in a neighborhood around $\hat{\theta}_T$ this procedure would eliminate the bias (up to simulation error), which prompted MacKinnon and Smith (1998) to call this a "constant-bias-correcting" (CBC) estimator. In general, however, the bias function is not constant, thus the bootstrap will systematically get the bias estimate wrong.[21] This is illustrated by the median of the bootstrap-bias-corrected estimator, which (for large B and under the assumption that $B_T(\cdot)$ is monotone) is $\theta_0 + B_T(\theta_0) - B_T(\theta_0 + B_T(\theta_0)) \neq \theta_0$. This insight motivates the refined bias correction procedure that we introduce in the following subsection.

3.3 Inverse bootstrap bias correction

The method we propose here to obtain bias-corrected estimates of θ_0 is to choose that parameter value which yields a distribution of the OLS estimator with a central tendency equal to the OLS estimate in the actual data, using the residual bootstrap to estimate this central tendency. We call our method "inverse bootstrap" bias correction because it is based on what one might call an inversion principle, instead of the plug-in principle of conventional bootstrap bias correction. While the idea is not new, we extend it to the context of VAR estimation and, more importantly, suggest a reliable and efficient algorithm to implement it with low computational cost even if $dim(\theta)$ is large. We first discuss separately the case of mean bias and median bias correction and then present our algorithm.

3.3.1 Mean bias correction

Define $g_T(\theta) = E_\theta(\hat{\theta}_T)$, the mean of the OLS estimator if the data are generated under θ. The mean-bias-corrected estimator of θ_0 is the value of θ that solves

$$g_T(\theta) = \hat{\theta}_T.$$

We denote this estimator by $\tilde{\theta}_T$. For identifiability, we assume that $g_T(\cdot)$ is uniformly continuous and one-to-one in a neighborhood around θ_0, which ensures that there is always exactly one solution to the above equation. The residual bootstrap can be used to obtain estimates of $g_T(\theta)$ for any value of θ, with the precision depending on the number of bootstrap samples.

[21] One can iterate on the bootstrap bias correction to obtain higher accuracy (Hall, 1992), but the computational burden increases exponentially, and the bias estimates are still only approximate.

We can write $\tilde{\theta}_T = g_T^{-1}(\hat{\theta}_T)$, which makes it clear why we speak of an inversion principle. Since the motivation for this method is valid for general bias functions, $\tilde{\theta}_T$ is termed a "nonlinear-bias-correcting" (NBC) estimator by MacKinnon and Smith (1998).

This method of correcting for mean bias is closely related to indirect inference.[22] In fact, if the true model and the instrumental model are the same, so that we have a consistent estimator for the true model, the indirect inference estimator for an infinite number of simulations is exactly $\tilde{\theta}_T$ (Gourieroux et al., 2000).

Correcting for mean bias does not lead to an unbiased estimator. We have

$$E_{\theta_0}(\tilde{\theta}_T) = E_{\theta_0}(g_T^{-1}(\hat{\theta}_T)) \neq g_T^{-1}(E_{\theta_0}(\hat{\theta}_T)) = g_T^{-1}(g_T(\theta_0)) = \theta_0,$$

except for the unlikely special case that the bias function is linear, since the expectation operator does not go through nonlinear functions.

3.3.2 Median bias correction

Let $G_T(\theta) = Med_\theta(\hat{\theta}_T)$ denote the element-wise median of the OLS estimator if the DGP is governed by θ. The median bias-corrected estimator of θ_0 is the value of θ that solves

$$G_T(\theta) = \hat{\theta}_T.$$

For this estimator we will write $\breve{\theta}_T$. We make the same assumptions about $G_T(\cdot)$ as about $g_T(\cdot)$ for identifiability. Thus $\breve{\theta}_T = G_T^{-1}(\hat{\theta}_T)$.

For the case of an AR(1) model, this estimator exactly corresponds to the one proposed by Andrews (1993). Rudebusch (1992) used this estimator to obtain parameter estimates in an AR(p) model. We extend it to VAR models.

As opposed to the case of mean bias correction, this estimator is in fact exactly median-unbiased, under the assumptions we made about $G_T(\cdot)$:

$$Med_{\theta_0}(\breve{\theta}_T) = Med_{\theta_0}(G_T^{-1}(\hat{\theta}_T)) = G_T^{-1}(Med_{\theta_0}(\hat{\theta}_T)) = G_T^{-1}(G_T(\theta_0)) = \theta_0.$$

The crucial difference is that the median operator does go through monotone functions.

[22] For a review of indirect inference, see Gourieroux and Monfort (1996, Chapter 4).

3.3.3 Algorithmic implementation

We now present an algorithm that can be used to reliably and rather quickly find the bias-corrected estimates, focusing on $\tilde{\theta}_T$ for this exposition. Define $R(\theta) = \hat{\theta}_T - g_T(\theta)$. Since the bootstrap gives us noisy measurements of $g_T(\theta)$, the problem of finding a solution to $R(\theta) = 0$ is different from classical root-finding.[23] Instead it is a problem in the area of "stochastic approximation," pioneered by Robbins and Monro (1951). Their crucial insight was that for each attempted value of θ we do not need a very precise measurement of $R(\theta)$, because it is only used to lead us in the right direction. Thus a small number of bootstrap replications is sufficient in each iteration, which greatly lowers our computational cost. The basic stochastic approximation algorithm is to construct a sequence according to

$$\theta^{(j+1)} = \theta^{(j)} + \alpha^{(j)} Y^{(j)}, \tag{7}$$

where $\alpha^{(j)}$ is a deterministic scalar sequence and $Y^{(j)}$ is a noisy measurement of $R(\theta^{(j)})$. Under some specific conditions about $\alpha^{(j)}$, the sequence will converge to $\tilde{\theta}_T$. However, the sequence of averages, $\bar{\theta}^{(j)} = j^{-1} \sum_{i=1}^{j} \theta^{(i)}$, converges even if $\alpha^{(j)}$ is taken to be a constant (between zero and one) and it does so at an optimal rate (Polyak and Juditsky, 1992). Under some rather weak conditions on $R(\cdot)$, $\alpha^{(j)}$ and the measurement error, we have $\bar{\theta}^{(j)} \to \tilde{\theta}_T$ almost surely, $\sqrt{j}$-asymptotic normality, as well as optimality in the sense of a maximum rate of convergence.[24] Motivated by these results, we use the following algorithm:

1. Choose as a starting value $\theta^{(1)} = \hat{\theta}_T$. Set $j = 1$.

2. Using $\theta^{(j)}$, obtain a measurement $Y^{(j)}$: estimate $g_T(\theta^{(j)})$ using a residual bootstrap with B replications (for details, see below) and set $Y^{(j)}$ equal to the difference between $\hat{\theta}_T$ and this estimate.

3. Calculate $\theta^{(j+1)}$ using equation (7).

4. Calculate $\bar{\theta}^{(j+1)}$, the average of the $\theta^{(j)}$'s for $j = 1, \ldots, j+1$.

5. If an exit condition is fulfilled, take $\bar{\theta}^{(j+1)}$ as the bias-corrected estimate of θ_0. Otherwise, increase j and return to step 2.

---

[23] Our assumptions about $g_T(\cdot)$ imply that $R(\theta) = 0$ has a unique solution.

[24] The only assumption that needs mentioning here is that the Jacobian at the solution point needs to be a Hurwitz matrix, i.e., the real parts of the eigenvalues of $R'(\tilde{\theta}_T)$ need to be strictly negative. Only if $R(\cdot)$ is decreasing in this sense does it make sense to increase the value of $\theta^{(j)}$ when we have positive measurements (equation 7). We check this condition by estimating the Jacobian at $\hat{\theta}_T$, verifying that it is Hurwitz, and relying on the assumption that this does not change between $\hat{\theta}_T$ and $\tilde{\theta}_T$. Details on how we estimate the Jacobian in this particular setting are available upon request.

In step two the approximate mean of the OLS estimator for a given $\theta^{(j)}$, i.e., an estimate of $g_T(\theta^{(j)})$, is obtained using a residual bootstrap with B replications. We randomly choose the starting values among the T observations. For $t > 1$ the bootstrapped series is obtained using $\tilde{X}_t^* = \Phi^{(j)}\tilde{X}_{t-1}^* + u_t^*$, where u_t^* are the bootstrap residuals, and $\Phi^{(j)}$ denotes the $N \times N$ matrix containing the elements of of $\theta^{(j)}$. Importantly, the bootstrap residuals have to be obtained for a given $\theta^{(j)}$: One cannot resample the original VAR residuals since these do not, together with $\theta^{(j)}$, generate the original data. Instead one has to first obtain a series of residuals $\hat{u}_t = \tilde{X}_t - \Phi^{(j)}\tilde{X}_{t-1}$, for $t > 1$, which then can be resampled in the usual way to create the bootstrap residuals u_t^*.[25]

We choose $\alpha^{(j)} = 0.5$ and $B = 50$, unless otherwise specified. Instead of devising a specific exit condition which might be computationally costly to check, we simply run the algorithm for a fixed number of iterations. We do not calculate $\bar{\theta}^{(j)}$ using all iterations but instead discard the first part of the sample, corresponding to the idea of a burn-in sample in the Markov chain Monte Carlo literature. Unless otherwise specified, we use 1000 iterations as a burn-in sample and then take as our estimate the average of the next 5000 iterations.

To verify the convergence of the algorithm, we then check how close $\bar{\theta}^{(j)}$ is to $\tilde{\theta}_T$. This is feasible despite $\tilde{\theta}_T$ being unknown, since we can obtain a measurement of $R(\bar{\theta}^{(j)})$ with arbitrarily small noise by using a large B, and check how close it is to zero. As the distance measure, we take root mean-square distance, that is $d(a, b) = (l^{-1}(a - b)'(a - b))^{1/2}$ for two vectors a, b of equal length l. We use this distance metric because it is invariant to the dimensionality of θ. We calculate $d(Y^{(j)}, 0)$, using a precision for the measurement of $B = 100,000$, and verify that this distance is less than 10^{-3}.

While the structure of our algorithm has solid theoretical foundations, our specific configuration ($\alpha^{(j)}$, B, number of burn-in/actual iterations) is admittedly arbitrary. We chose it based on our own experience with the algorithm. The specifics of the problem likely would allow us to reduce the computational cost further by choosing the configuration in some optimal way. We leave this for future research. With our configuration, the computational costs are very manageable: For a VAR(1) with 3 variables and about 300 observations, using a Dell Laptop with Intel Core i5 CPU (2.53 GHz, 3.42 GB RAM), it takes about five minutes to run the algorithm.

3.3.4 Explosive VAR dynamics

A problematic aspect of bias correction in VAR models is the possibility of explosive estimates, which contradicts our prior that interest rate dynamics should be stationary. There are two

[25] This notation suppresses the dependence on the bootstrap replication b and on the iteration j.

alternative ways to ensure stationarity of the bias-corrected estimates. First, the restriction could be systematically incorporated into the bias correction procedure, so in each iteration in the stochastic approximation algorithm, the proposed update, $\theta^{(j)}+\alpha^{(j)}Y^{(j)}$, is projected into a constraint set. Alternatively, as in Kilian (1998), a stationarity adjustment can be applied after obtaining unrestricted bias-corrected estimates. There are three advantages to this second approach. First, it will often be the case that bias-corrected estimates are stationary so no adjustment is needed. Second, when needed, a two-step procedure is straightforward and easy to implement. Third, it can be instructive to identify cases in which unrestricted bias correction would lead to nonstationary estimates. In our experience, while nonstationary estimates are certainly possible, even if the DGP is stationary and the model is correctly specified (as shown below), such estimates seem most prevalent when the VAR is misspecified. Hence, nonstationary roots after bias correction may raise a useful warning sign.

In our two empirical case studies in Sections 4 and 6, our bias-corrected estimates turn out to be stationary, so explosive eigenvalues are not an issue. In our VAR simulation study below (Section 3.4), we focus on parameter bias—we do not ensure stationarity and simply report the frequency of nonstationary roots. However, in the DTSM simulation study in Section 5, stationarity of the estimated system is crucial, therefore we will apply Kilian's stationarity adjustment to those bias-corrected estimates which have nonstationary roots.

3.4 AR/VAR Monte Carlo study

To assess the performance of our proposed bias correction method, we present the results of a simulation study, which considers both a univariate AR model as well as a two-variable VAR model. To create a setting that is comparable to the reality faced by researchers analyzing interest rates, we first estimate AR and VAR models on quarterly interest rate data. Specifically we use zero coupon rates at 120 maturities spanning one month to 10 years, starting in 1985:Q1 and ending in 2009:Q4.[26] The sample length is thus $T = 100$. We extract the first two principal components from the cross section of yields; for the AR model, we use just the first principal component. For this sample, the Schwartz-Bayes information criterion suggests a lag order of one for both the AR and VAR. We take the OLS estimates as our DGP, setting

[26]The zero coupon rates are constructed from observed bond prices using unsmoothed Fama-Bliss forward rates. We thank Anh Le from the University of North Carolina for providing us with this data set. It has been used, for example, in Joslin et al. (2010b).

the lower-left element of Φ to zero since it is insignificant:

$$X_{t+1} = \begin{pmatrix} .95 & .14 \\ 0 & .87 \end{pmatrix} X_t + \Sigma\varepsilon_{t+1}, \quad \Sigma = \begin{pmatrix} .066 & 0 \\ -.006 & .019 \end{pmatrix}, \quad \varepsilon_t \overset{iid}{\sim} N(0, I_2).$$

For the AR(1) model, the autoregressive coefficient is estimated by OLS as 0.95. We also consider the bias-corrected estimate of the autoregressive coefficient, obtained using the inverse bootstrap, which turns out to be 0.99.

We generate $M = 10,000$ samples and calculate for each replication the OLS estimates and mean-bias-corrected as well as median bias-corrected estimates using both conventional bootstrap (CB) and inverse bootstrap (IB). For the conventional bootstrap bias correction, we use 1,000 replications. For the inverse bootstrap, we use 10 replications in each step and iterate the algorithm 1,100 times, discarding the first 100 values as the burn-in sample.

Table 1 shows the results of our simulations. For each parameter, the first column shows the true value, the next three columns show the mean bias for OLS and the mean-bias-corrected estimates, and columns five through seven show median bias for OLS and the median bias-corrected estimates. In addition, the last row of each panel shows the relative frequency of nonstationary roots. For the VAR, the bottom panel also reports the "root-mean-squared bias," which is the square root of the mean-squared bias across the four parameters, the "total absolute bias," which is the sum of the absolute values of the bias in each parameter, and the mean/median bias for the largest eigenvalue of Φ.

The results for the AR(1) models demonstrate how the well-known Kendall bias is reduced by an order of magnitude by means of the conventional bootstrap and inverse bootstrap. Theory predicts that for mean bias either method could do better (MacKinnon and Smith, 1998) but that for median bias the inverse bootstrap should be superior (Andrews, 1993), and this is exactly what we find. Turning to the VAR model, the results also show how bias correction significantly reduces the bias. The total absolute bias is reduced by 80-90% using the bootstrap, and the dramatic downward bias in the largest eigenvalue disappears. Relative to the standard bootstrap, the inverse bootstrap estimates most elements of Φ with less bias; indeed, the measures of total bias are roughly cut in half. These observations hold both for mean and median bias. One caveat is that the largest root is estimated with slightly larger bias than with the bootstrap, but the improvement over OLS is still sizeable.

Bias correction often leads to nonstationary roots. Although the DGP is stationary and the estimated model is correctly specified, there is a non-zero probability that the realized value of the OLS estimator is such that bias correction leads to explosive estimates. Consequently, in practice some type of stationarity adjustment will be necessary to ensure that estimated

Table 1: Simulation results – parameter bias

	true value	Mean bias			Median bias		
		OLS	CB	IB	OLS	CB	IB
AR(1), $\rho = .95$							
ρ	0.95	-0.0474	-0.0066	-0.0029	-0.0374	-0.0057	-0.0003
freq. expl.		0.10%	12.33%	16.17%	0.10%	6.91%	9.84%
AR(1), $\rho = .99$							
ρ	0.99	-0.0533	-0.0113	-0.0120	-0.0437	-0.0102	-0.0024
freq. expl.		1.84%	39.74%	42.57%	1.84%	29.76%	32.27%
VAR model							
Φ_{11}	0.95	-0.0499	-0.0062	-0.0031	-0.0408	-0.0052	0.0005
Φ_{12}	0.14	0.0078	-0.0010	-0.0032	0.0072	-0.0019	-0.0047
Φ_{21}	0	-0.0011	-0.0003	-0.0002	-0.0012	-0.0003	-0.0002
Φ_{22}	0.87	-0.0544	-0.0079	-0.0022	-0.0450	-0.0064	-0.0008
root-mean-sq. bias		0.0371	0.0051	0.0025	0.0306	0.0042	0.0024
total abs. bias		0.1132	0.0154	0.0086	0.0942	0.0137	0.0061
max. eig.	0.95	-0.0419	0.0074	0.0113	-0.0368	0.0047	0.0122
freq. expl.		0.02%	15.79 %	19.40%	0.02%	9.66%	12.83%

Notes: True values, mean bias, median bias, and summary statistics for different estimators in a Monte Carlo study with M=10,000 replications and T=100 observations, using an AR(1)/VAR(1) as a DGP (for details see text). The first row in the first two panels and the first four rows in the third panel show for parameter the true value, the mean bias of the OLS estimates, and the mean-bias-corrected estimates using the conventional bootstrap (CB) and the inverse bootstrap (IB), as well as the median bias of the OLS estimates and the median bias-corrected estimates using the bootstrap and the inverse bootstrap. The last row in each panel shows the frequency with which explosive eigenvalues occur. The remaining rows in the third panel show the root-mean-squared bias across parameters, the total absolute bias across parameters, and the true value and mean/median bias for the largest eigenvalue of Φ.

dynamics are not explosive.

We draw two conclusions from this simulation study: First, both the bootstrap and inverse bootstrap are useful and reliable methods to reduce the bias in OLS estimates of VAR parameters. Second, the inverse bootstrap is a superior bias correction method compared to the standard bootstrap correction—it can further reduce mean bias and essentially eliminate median bias in the parameter estimates.

Table 2: JSZ case study – OLS and unbiased parameter estimates

	OLS			MU		
$\mu \times 1200$	-0.5440	-0.1263	0.0700	-0.3402	-0.2192	0.0770
Φ	0.9788	0.0133	0.4362	0.9962	0.0005	0.4399
	0.0027	0.9737	0.3532	-0.0013	0.9901	0.3517
	-0.0025	-0.0023	0.8537	-0.0002	0.0014	0.8732
$\lvert eig(\Phi)\rvert$	0.9678	0.9678	0.8706	0.9949	0.9949	0.8696
$r^{\mathbb{Q}}_{\infty} \times 1200$	8.6055			8.6478		
$\lambda^{\mathbb{Q}}$	0.9976	0.9519	0.9287	0.9976	0.9519	0.9287
$\Sigma \times 1200$	0.6365	0	0	0.6430	0	0
	-0.1453	0.2097	0	-0.1464	0.2105	0
	0.0630	-0.0117	0.0867	0.0633	-0.0115	0.0872

Notes: Parameter estimates of the DTSM in Joslin et al. (2010a). The left panel shows point estimates for the case that the dynamic system (factor VAR) was estimated by OLS; the right panel shows the corresponding results for the case of median-unbiased (MU) estimation.

4 JSZ Case Study

In an important contribution to the term structure literature, JSZ develop a new normalization that allows separate estimation of the physical and risk-neutral dynamics. In this section, we replicate the model estimates in JSZ, provide bias-corrected estimates of that model, and compare the associated risk-neutral forward rates and term premia.

4.1 Unbiased JSZ model estimation

We replicate JSZ's "RPC" canonical DTSM. The pricing factors X_t are the first three principal components of yields, where $\Phi^{\mathbb{Q}}$ has distinct real eigenvalues. That is, the pricing factors are $X_t = WY_t$, where W consists of the eigenvectors corresponding to the three largest eigenvalues of the covariance matrix of Y_t. There are no overidentifying restrictions, thus there are 22 free parameters, not counting measurement error variances. The free parameters are μ (3), Φ (9), $r^{\mathbb{Q}}_{\infty}$, $\lambda^{\mathbb{Q}}$ (3), and Σ (6). The JSZ data consist of zero-coupon yields that are bootstrapped from end-of-month constant maturity Treasury yields from the Fed's H.15 release, from January 1990 to December 2007. The yield maturities are 6 months and 1, 2, 3, 5, 7 and 10 years.[27] The $M = 7$ observed yields at time t are stacked in the vector $Y_t = [y_t^{m_1}, y_t^{m_2}, ..., y_t^{m_M}]'$.

We perform ML estimation and median-unbiased (MU) estimation of the model. For the

[27] JSZ have made their data and the MATLAB code used for estimation available online at http://www.stanford.edu/~kenneths/jsz.zip.

Table 3: JSZ case study – summary statistics

	OLS	MU
max. eig.	0.9678	0.9949
half-life	24.7	136.7
CIR	34.1	162.7
IRF at 5y	0.16	0.76
$\sigma(f_t^{47,48})$	1.392	1.392
$\sigma(\tilde{f}_t^{47,48})$	0.388	1.356
$\sigma(ftp_t^{47,48})$	1.301	1.440

Notes: Summary statistics for OLS and median-unbiased (MU) estimates of the DTSM in Joslin et al. (2010a). First row: maximum eigenvalue of the estimated Φ. Second to fourth row: half-life, cumulative impulse response, and value of the impulse response function at the five-year horizon for the response of the level factor to a level shock. Rows six to eight show standard deviations of the fitted 47-to-48-month forward rates and of the corresponding risk-neutral forward rates and forward term premia.

ML estimation, we exactly follow JSZ and are able to replicate their estimates, shown in the left panel of Table 2.[28] To obtain bias-corrected estimates of the model parameters, we follow the procedure descripted in Section 2.3. In the first stage, we obtain MU estimates of (μ, Φ) using the inverse bootstrap. These are reported in the top right section of Table 2. Note how the bias correction procedure significantly increases the absolute value of the largest eigenvalue of Φ, which now is 0.9949 instead of 0.9678. In the second stage, we maximize the likelihood function over $\left(r^{\mathbb{Q}}, \lambda^{\mathbb{Q}}, \Sigma\right)$, taking as given the unbiased estimates of the VAR parameters. The estimates we obtain are reported in the bottom right of Table 2. Because of the JSZ separation result, the estimated risk-neutral dynamics and the estimated Σ are very similar, whether we use OLS or MU estimation for the physical dynamics.[29]

4.2 Economic implications of bias correction

To highlight the economic implications of bias-corrected DTSM estimates, we examine measures of persistence and decompose forward rates into expectations and risk premium components.

[28]Compare the top left panel of our Table 2 with the top row of JSZ's Table 3, noting that $(I - \Phi)^{-1} \mu = \theta^{\mathbb{P}}/12$ and $(\Phi - I) = K_1^{\mathbb{P}}/12$. Compare the middle left panel of our Table 2 with the top row of JSZ's Table 2, noting that our risk-neutral eigenvalues are one plus JSZ's risk-neutral eigenvalues.

[29]The differences in the $\mathbb{Q}$-parameters between the left and the right panel stem from the fact that Σ enters both the $\mathbb{P}$-likelihood and the $\mathbb{Q}$-likelihood. Therefore, different values of μ and Φ will lead to different optimal values of $\left(r^{\mathbb{Q}}, \lambda^{\mathbb{Q}}, \Sigma\right)$ in the second stage.

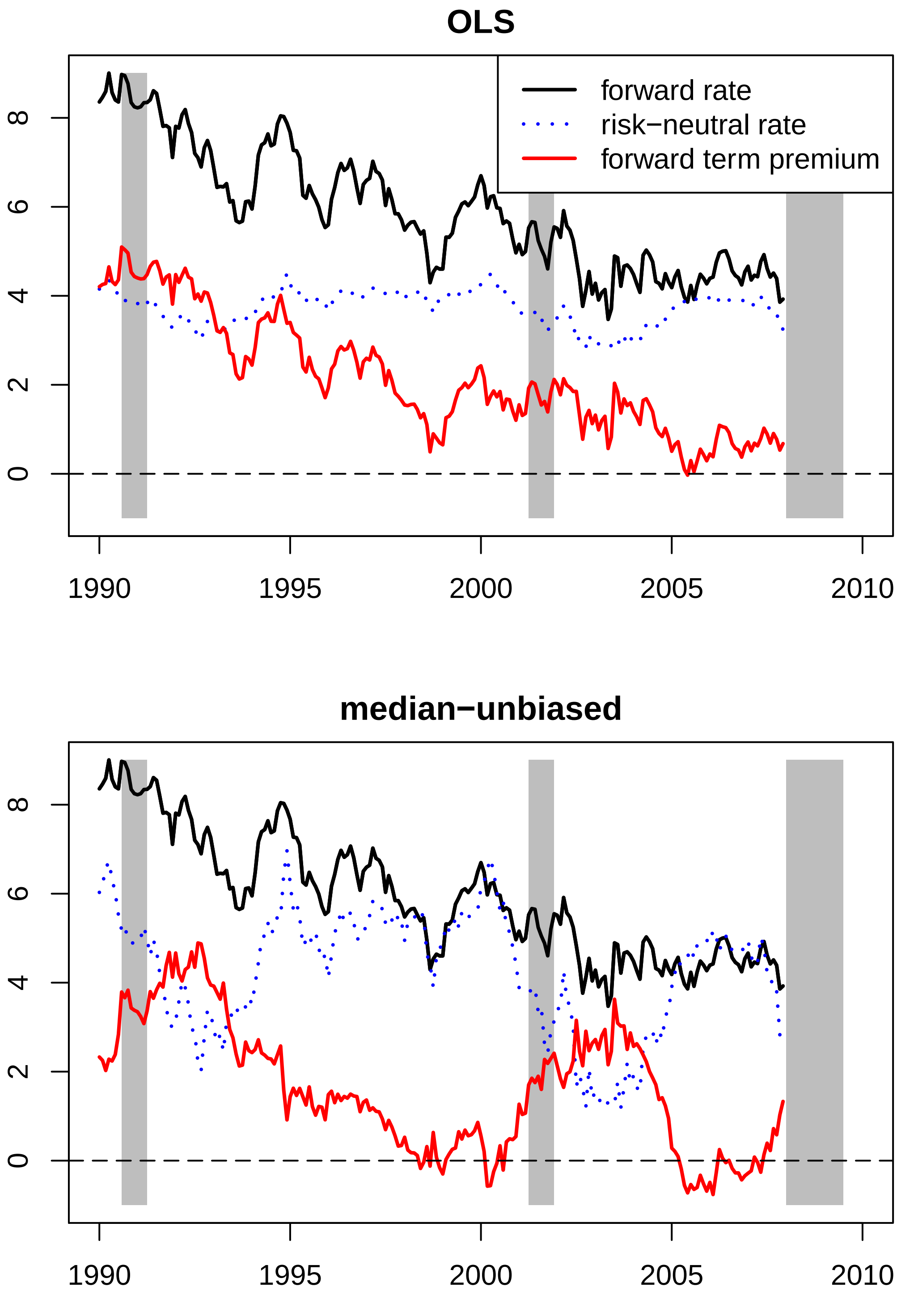
The top panel of Table 3 provides four measures of the persistence of the estimated dynamics. The first row reports the maximum absolute eigenvalue of the estimated mean-reversion matrix Φ, which is considerably closer to one after bias correction. The second to fourth rows are based on the impulse response function (IRF) of the level factor to a level shock. The second row shows the half-life, which is around two years for OLS and around eleven years for MU. The third row reports the cumulative impulse response (CIR), which is increased by the bias correction by a factor of about 4.5. The fourth row reports the value of the IRF at the five-year horizon, which shows a similar increase after bias correction. The unescapable conclusion is that the OLS JSZ model estimates greatly understate the persistence of the dynamic system.

On practical effect of the JSZ underestimation of persistence can be illustrated with a decomposition of forward rates. We focus on the one-month forward rate for a loan maturing in four years, i.e., $f_t^{47,48}$. The last three rows of Table 3 show standard deviations of this forward rate and of the component risk-neutral forward rate and forward term premium. Not surprisingly, the slower mean reversion of the unbiased DTSM estimate produces more volatile future short rate expectations and hence risk-neutral rates.[30]

This greater volatility following the correction for estimation bias is evident in Figure 1, which shows the fitted forward rate, risk-neutral forward rate, and forward term premium for our sample based on OLS estimates (top panel) and MU estimates (bottom panel). The risk-neutral forward rate resulting from OLS estimates displays little variation, and the associated term premium closely mirrors the movements of the forward rate. The secular decline in the forward rate is attributed to the risk premium component. In contrast, the risk-neutral forward rates obtained from the unbiased DTSM estimates vary over time and account for a considerable portion of the secular decline in the forward rate. From a macro-finance perspective, the later decomposition seems more plausible. The MU future short rate expectations are low during recessions (denoted by the shaded bars) and in the early stages of expansions, which is consistent with a sustained countercyclical monetary policy response. In particular, such time variation is evident in survey-based interest rate forecasts (Kim and Orphanides, 2005) and far-ahead inflation expectations (Kozicki and Tinsley, 2001). Similarly, the MU estimated term premium varies significantly over time and—fittingly for a compensation for risk—rises notably during recessions. Most macroeconomists believe that risk premia behave in such a countercyclical fashion, given theoretical work such as Campbell and Cochrane (1999) and Wachter (2006) as well as empirical evidence from Harvey (1989) to Lustig et al. (2010). In

[30]The variability of the forward term premium is also higher for the unbiased case, which is due to the covariation of the term premium and risk-neutral forward rates. This result appears specific to the particular data set used, while the greater variability of risk-neutral forward rates seems more general.

Figure 1: JSZ case study – decomposition of forward rates



One-month forward rates with four years maturity, decomposed into risk-neutral forward rates and forward term premia using the affine Gaussian DTSM of Joslin et al. (2010a). Sample: Monthly observations from January 1990 to December 2007. Top panel: factor VAR estimated by OLS. Bottom panel: factor VAR estimated by MU.

contrast, the OLS estimated term premium is very stable and, if anything, appears to decline a bit during economic recessions. The pronounced countercyclical pattern in the unbiased term premium makes it a more plausible measure of bond risk premia than conventional term premia that suffer from small-sample bias.

In sum, this case study shows that the small-sample bias inherent in the DTSM estimates of Joslin et al. (2010a) is significant and that taking account of this bias leads to term premium estimates that are economically different from the term premium estimates implied by the results reported in JSZ.

5 DTSM Monte Carlo study

The simulation study in Section 3.4 has demonstrated how bias correction methods can successfully reduce the small-sample bias in the VAR parameter estimates. However, as illustrated in the preceding case study, the objects of interest in a DTSM are often nonlinear functions of the parameters, such as expected risk-neutral rates and term premia, which may not show as much of a difference from bias correction. Thus, to follow up on the JSZ case study, it is useful to systematically investigate whether bias correction actually improves inference about expected short rates and term premia. For this purpose we perform a Monte Carlo study that repeatedly simulates interest rate data from the estimated JSZ DTSM and then re-estimates that specification using OLS and MU methods. Since we know the population properties as well as the exact time series in each sample of risk-neutral rates and term premia, we can assess the accuracy of the conventional and bias-corrected methods.

5.1 Simulation and estimation

The DGP is the three-factor JSZ specification estimated in Section 4. We assume the first three principal components of yields are pricing factors are exactly priced by the model. We simulate 1000 data sets for the same seven maturities as in JSZ. As DGP parameters, we take the MU estimates that were reported in Table 2. For the pricing factors X_t, we simulate time series with T=300 observations from the VAR, drawing the starting values from their stationary distribution.[31] Model-implied yields are given by $Y_t = A + BX_t$, where $A = (A_{m_1}, \ldots, A_{m_7})'$ and $B = (B_{m_1}, \ldots, B_{m_7})'$. We obtain the yield-loadings for given $(r^{\mathbb{Q}}_{\infty}, \lambda^{\mathbb{Q}}, \Omega, W)$, with the 3×7 matrix W containing the loadings for the first three principal components in the original

[31] The unconditional mean is $E(X_t) = (I - \Phi)^{-1}\mu$. The unconditional variance $V_0 = Var(X_t)$ is related to the innovation variance $\Sigma\Sigma'$ by the equation $vec(V_0) = (I_{N^2} - \Phi \otimes \Phi)^{-1} vec(\Sigma\Sigma')$ (Hamilton, 1994, p. 265).

yield data. Note that by construction $X_t = WY_t$. The pricing factors, which are three linear combinations of yields, are assumed to be measured without error, i.e., $Y_t^1 = X_t$.

We use the HW framework for estimation, which is computationally very simple for the exactly identified case where the number of measured series is equal to $N+1$. In particular, we only generate one series measured with error, $Y_t^2 = W_2 Y_t + e_t$, where $W_2 = (0, 0, 0, 0, 0, -7, 10)$ and $e_t \overset{iid}{\sim} N(0, \sigma_e^2)$, with $\sigma_e = .0001$. This choice of W_2 makes Y_t^2 proportional to the seven-to-ten-year forward rate.[32] The specific advantage of the HW estimation method for the case of exact identification is that the second-stage estimation is much simpler and more reliable than for the overidentified case. Since the number of structural parameters is equal to the number of reduced-form parameters, the chi-squared statistic achieves a value of exactly zero (up to numerical accuracy of the nonlinear equation solver). Thus, we know for sure whether we have found the global minimum in the second stage. Furthermore, the task of finding the structural parameters can be divided into a sequence of smaller steps—first find the $\mathbb{P}$-parameters, second find $\lambda^{\mathbb{Q}}$, third find $r_\infty^{\mathbb{Q}}$—which accelerates and simplifies finding the minimum of the chi-squared statistic. The estimation works as follows: In the first stage we estimate the VAR for Y_t^1, which provides us with estimates of (μ, Φ, Σ), and we regress Y_t^2 on Y_t^1, resulting in reduced-form loadings $\hat{\mu}_2$ (a scalar) and $\hat{\Phi}_2$ (a 1×3 vector). For the VAR, we perform both OLS and MU estimation.[33] Bias correction is not necessary for the cross-sectional regression. In the second stage, we minimize the chi-squared statistic, which is very straightforward because of exact identification. From the VAR estimates, we already know the $\mathbb{P}$-dynamics, and we only need to find $\lambda^{\mathbb{Q}}$ and $r_\infty^{\mathbb{Q}}$. We choose $\lambda^{\mathbb{Q}}$ to make the model-implied loadings of Y_t^2 on X_t, given by $W_2 B$, match the reduced-form loading $\hat{\Phi}_2$. Then we choose $r_\infty^{\mathbb{Q}}$ such that the intercept of Y_t^2, given by $W_2 A$, matches $\hat{\mu}_2$. This estimation procedure is very fast and very reliable.[34]

Since our inference focuses on objects that are highly sensitive to persistence of the estimated VAR dynamics, we need to ensure stationarity of the estimates. For MU estimates that have nonstationary roots we apply the stationarity adjustment proposed in Kilian (1998), which shrinks the bias-corrected estimates to the (stationary) OLS estimates just enough to make them stationary. In those rare cases where OLS estimates are explosive, we shrink the

[32] The specific choice of W_2 is of minor importance, the only requirement being that it not be an exact linear combination of the rows of W.

[33] To find MU estimates using the inverse bootstrap, we calculate the average over 1000 iterations, after 100 burn-in iterations. In each iteration, we calculate the median of the OLS estimates across 10 replications and update using $\alpha^{(j)} = .5$.

[34] Note that we use the same W and W_2 in the data simulation and in estimation, which ignores that in reality a researcher would have to decide which linear combinations are taken to be measured with and without error, and that loadings of principal components would be different in each data set. We do this for simplicity and to focus attention on the inference about parameters.

estimates towards zero in a similar fashion, and only then proceed to obtain bias-corrected estimates. Our procedure ensures stationarity of the estimated dynamic system in every Monte Carlo replication.

We calculate fitted four-to-five-year forward rates and the corresponding risk-neutral rates and term premia. Based on the realized sample path of X_t, these rates and premia can be calculated (i) for the true DGP parameters, (ii) for the OLS parameters, and (iii) for the MU parameters. Interest lies in the question of how close the estimated series correspond to the true series, and we assess this by focusing on two types of summary statistics: volatilities and root-mean-squared errors (RMSEs). We report these statistics based on annualized interest rates in percentage points. The motivation for considering volatilities is that the small-sample bias in conventional DTSM estimates supposedly makes risk-neutral forward rates too stable. We look at both sample and population volatilities: For each simulation, we calculate in-sample volatilities of rates and premia for the DGP, OLS, and MU.[35] We expect volatilities of risk-neutral rates implied by MU to be larger and closer to the true DGP volatilities than those for OLS. The RMSEs that we calculate measure the accuracy of estimated rates and premia in relation to the true series implied by the DGP parameters.

5.2 Results

In Table 4 we show the DGP parameters and the median bias in the OLS and MU estimates. The bias in the estimates of the $\mathbb{P}$-dynamics is quite significant for OLS, while the MU estimates generally show much smaller bias, as was to be expected in light of the results of Section 3.4. With regard to the $\mathbb{Q}$-dynamics, $r^{\mathbb{Q}}_{\infty}$ is estimated with a slight downward bias both by OLS and MU. The intuitive reason is that the largest root under $\mathbb{Q}$ is very close to one, which makes inference about the long-run mean under $\mathbb{Q}$ more difficult.[36] Finally, $\lambda^{\mathbb{Q}}$ and Σ are estimated without any systematic bias by both OLS and MU. In sum, the physical dynamics of the DGP are more accurately recovered if one uses MU estimation as opposed to OLS, while the remaining parameters are estimated with similar accuracy in either case.

Table 5 presents summary statistics about the persistence of the estimated dynamic system and about how well forward rates and risk premia are estimated. In the first four rows, we show the true values (DGP) and the medians across replications of the estimated values

[35] We do not report population volatilities, which in principle can be calculated based on the estimated long-run variance of the pricing factors and the loadings of forward rates and term premia. These are highly sensitive to the VAR parameters, and bias in parameter estimates implies that the parameter-implied population variances are not comparable.

[36] This has motivated some researchers to restrict the largest root under $\mathbb{Q}$ to one and $r^{\mathbb{Q}}_{\infty}$ to zero, in which case the long end of the yield curve is determined by a level factor (Bauer, 2011; Christensen et al., 2011).

Table 4: DTSM Monte Carlo study – true parameter value and bias

	DGP			Median bias OLS			Median bias MU		
$\mu \times 1200$	-0.340	-0.219	0.077	-0.189	0.067	-0.008	-0.064	0.023	-0.008
Φ	0.996	0.000	0.440	-0.013	0.015	0.005	-0.004	0.005	0.012
	-0.001	0.990	0.352	0.003	-0.012	0.000	0.001	-0.003	-0.003
	0.000	0.001	0.873	-0.002	-0.001	-0.012	-0.001	0.000	0.000
$\lvert eig(\Phi)\rvert$	0.995	0.995	0.870	-0.008	-0.030	0.000	0.004	-0.016	0.005
$r_\infty^{\mathbb{Q}} \times 1200$	8.648			-0.057			-0.040		
$\lambda^{\mathbb{Q}}$	0.998	0.952	0.929	0.000	0.000	0.000	0.000	0.000	0.000
$\Sigma \times 1200$	0.643	0.000	0.000	-0.007	0.000	0.000	-0.005	0.000	0.000
	-0.146	0.210	0.000	0.002	-0.002	0.000	0.002	-0.002	0.000
	0.063	-0.012	0.087	-0.001	0.000	-0.001	-0.001	0.000	-0.001

Notes: DGP parameter values (columns 1-3), median bias of OLS estimates (columns 4-6), and median bias of MU estimates (columns 7-9). For details on the Monte Carlo set-up, refer to the main text.

(OLS/MU) for selected persistence measures, namely the largest root, half-life, cumulative impulse response (CIR), and the value of the impulse response function (IRF) at the five-year horizon (for the response of the level factor to own shocks). Not surprisingly, the persistence of interest rates is significantly underestimated by OLS, with median values being far away from the values corresponding to the true DGP. Unbiased estimation generally implies much higher estimated persistence.

We show medians of in-sample volatilities in rows five to seven of Table 5. Volatilities of forward rates naturally are similar for the DGP and estimated series because the model is fitted to the cross section of interest rates. Importantly, for risk-neutral forward rates, OLS implies volatilities that tend to be much below those of the true risk-neutral rates. On the other hand, MU estimation leads to risk-neutral rates that are equally volatile as the true rates. This strongly confirms the intuition that expectations of future short rates and risk-neutral interest rates should be calculated based on unbiased estimates instead of conventional estimates of the DTSM parameters.

In the last three rows of the table, we present RMSEs. Forward rates are naturally fit very accurately, with about three basis points of error on average. Risk-neutral forward rates and forward premia are estimated much more imprecisely, with RMSEs around 1.5 percentage points. Interestingly, the MU estimates generally imply lower errors than OLS. This is another encouraging result, demonstrating the higher accuracy of unbiased DTSM estimation for inference about short rate expectations and risk premia.

Table 5: DTSM Monte Carlo study – summary statistics

	DGP	OLS	MU
$max(eig(\Phi))$	0.9949	0.9867	0.9993
half-life	136.68	30.67	168.84
CIR	162.73	39.11	124.53
IRF at 5y	0.76	0.23	0.58
$\sigma(f_t)$	1.67	1.67	1.67
$\sigma(\tilde{f}_t)$	1.75	1.08	1.89
$\sigma(ftp_t)$	2.05	1.84	2.20
RMSE(f_t)		< 0.01	< 0.01
RMSE($\tilde{f}_t$)		1.31	1.18
RMSE(ftp_t)		1.31	1.18

Notes: Persistence and accuracy of estimated rates and premia in DTSM Monte Carlo study. First four rows show true values (DGP) and medians of estimated values for largest root of Φ, half-life, cumulative impulse response (CIR), and impulse response function (IRF) at the five-year horizon for response of first pricing factor to own shocks. Rows five to seven show medians of in-sample standard deviations of forward rates, risk-neutral forward rates, and forward premia. Last three rows show medians of root-mean-squared errors (RMSE) for estimated rates and premia. For details on Monte Carlo set-up, refer to main text.

6 HW Case Study

Almost all studies in the DTSM literature impose some restrictions on model parameters. For example, Ang and Piazzesi (2003) and Kim and Orphanides (2005) impose zero restrictions on Φ and (λ_0, λ_1). Recently, the value of restrictions on the risk pricing in DTSMs, i.e., on the elements of λ_0 and λ_1, for reducing small-sample bias has been stressed in the literature (Joslin et al., 2010b; Bauer, 2011). The purpose of this case study is to implement the procedure suggested in Section 2.4, which adapts the HW estimation framework to correct for small-sample bias, for unbiased estimation of DTSMs with overidentifying restrictions.

6.1 Data, model specification, and estimation

We use the monthly unsmoothed Fama-Bliss zero-coupon rate data set described above. Our sample starts in January 1990 and extends through December 2009, which amounts to $T = 239$ observations. We include yields with monthly maturities of $m_i \in \{3, 6, 12, 24, 36, 60, 84, 120\}$. The measured yields are denoted by $\hat{Y}_t$ and the model-implied yields by Y_t.

As before we assume that the first three principal components of yields, Y_t^1, are priced exactly by the model, and take these as our pricing factors, i.e., $X_t = Y_t^1 = WY_t$, where

W contains the eigenvectors corresponding to the largest three eigenvalues of the covariance matrix of $\hat{Y}_t$. The remaining five principal components, Y_t^2, are assumed to be measured with error: $Y_t^2 = W_2 Y_t + u_t^2 = W_2 A + W_2 B X_t + u_t^2$, where W_2 contains the other five eigenvectors of $Var(\hat{Y}_t)$.

For the identifying restrictions, we again stick with the JSZ normalization. Thus the model parameters are $(\mu, \Phi, \Sigma, r_\infty^{\mathbb{Q}}, \lambda^{\mathbb{Q}})$ or alternatively $(\lambda_0, \lambda_1, \Sigma, r_\infty^{\mathbb{Q}}, \lambda^{\mathbb{Q}})$, plus the measurement error variance Ω_2. The second parameterization will be useful when we impose restrictions on the risk prices.

In order to decide which restrictions to impose, we first estimate a canonical model without bias correction. The first step is to estimate the VAR for Y_t^1 and the five cross-sectional regressions for Y_t^2 by OLS. In the second step we find the structural parameters that minimize the chi-squared distance.[37] We focus on $(\Sigma\lambda_0, \Sigma\lambda_1)$ instead of on (λ_0, λ_1), since we do not want our inference to depend on the arbitrary factorization of the covariance matrix of the VAR innovations (Joslin et al., 2010b). We report parameter estimates and standard errors for $(\Sigma\lambda_0, \Sigma\lambda_1, r_\infty^{\mathbb{Q}}, \lambda^{\mathbb{Q}}, \Sigma)$ in Table 6; for details on how to calculate standard errors, refer to HW.

Based on the t-statistics, we decide to specify our restricted model with the $(3, 2)$ and $(2, 3)$ elements of $\Sigma\lambda_1$ set to zero.[38] The restricted model is estimated using both OLS estimates as well as MU estimates of the reduced-form system. The minimum-chi-squared step is performed in exactly similar fashion for both cases. We report the resulting parameter estimates in the middle and right panel of Table 6. Naturally the $\mathbb{Q}$-parameters are very similar across all three sets of estimates. Notably, the estimates of the unrestricted risk price parameters change quite significantly between OLS (middle panel) and MU (right panel). Clearly, bias correction has a sizeable impact on the magnitudes of the estimated risk sensitivities. The economic consequences of these differences will become evident below.

6.2 Economic implications of bias correction

We decompose five-to-ten year forward rates, as for example in Wright (2011), since these accurately represent the long-maturity properties of short rate expectations and risk premia.

[37] The reduced-form estimates are $(\hat{\mu}_1, \hat{\Phi}_1, \hat{\Omega}_1, \hat{\mu}_2, \hat{\Phi}_2, \hat{\Omega}_2)$. Because there are no restrictions on risk prices or $\mathbb{Q}$-dynamics, $\hat{\mu} = \hat{\mu}_1$ and $\hat{\Phi} = \hat{\Phi}_1$. And, as always, the measurement error variance is both a structural and a reduced-form parameter. Thus the numerical minimization of the chi-squared statistic only involves finding $r_\infty^{\mathbb{Q}}$, $\lambda^{\mathbb{Q}}$ and Σ.

[38] Such an ad hoc choice of restrictions is common practice in the DTSM literature, although there is significant model uncertainty, given the very different economic implications of different restrictions. This has been noticed, among others, by Kim and Orphanides (2005). Bauer (2011) deals with the problems of model selection and model uncertainty more systematically.

Table 6: HW case study – parameter estimation

	OLS, unrestricted			OLS, restricted			MU, restricted		
$\Sigma\lambda_0 \times 1200$	-0.4163	-0.0278	-0.0318	-0.4431	-0.0025	-0.0293	-0.3675	-0.0601	-0.0575
	(0.2201)	(0.1286)	(0.0545)						
$\Sigma\lambda_1$	-0.0076	0.0910	-0.3290	-0.0077	0.0922	-0.3573	0.0019	0.0770	-0.3719
	(0.0090)	(0.0358)	(0.1621)						
	-0.0079	-0.0228	-0.0309	-0.0078	-0.0231	0	-0.0084	-0.0046	0
	(0.0053)	(0.0209)	(0.0948)						
	0.0010	0.0007	-0.0812	0.0010	0	-0.0812	-0.0016	0	-0.0662
	(0.0022)	(0.0089)	(0.0404)						
$r^{\mathbb{Q}} \times 1200$	8.9846			9.9810			9.0033		
	(0.3609)								
$\lambda^{\mathbb{Q}}$	0.9931	0.9673	0.9179	0.9931	0.9674	0.9178	0.9931	0.9674	0.9178
	(0.0008)	(0.0034)	(0.0082)						
$\Sigma \times 1200$	0.6694	0	0	0.6694	0	0	0.6770	0	0
	(0.0304)								
	-0.2216	0.3358	0	-0.2216	0.3358	0	-0.2214	0.3369	0
	(0.0229)	(0.0146)							
	-0.0530	-0.0448	0.1485	-0.0530	-0.0448	0.1485	-0.0526	-0.0446	0.1494
	(0.0103)	(0.0095)	(0.0068)						

Notes: Parameter estimates for canonical model specification (left panel), as well as conventional (OLS) and median-unbiased (MU) estimates of the model with zero restrictions on risk price parameters.

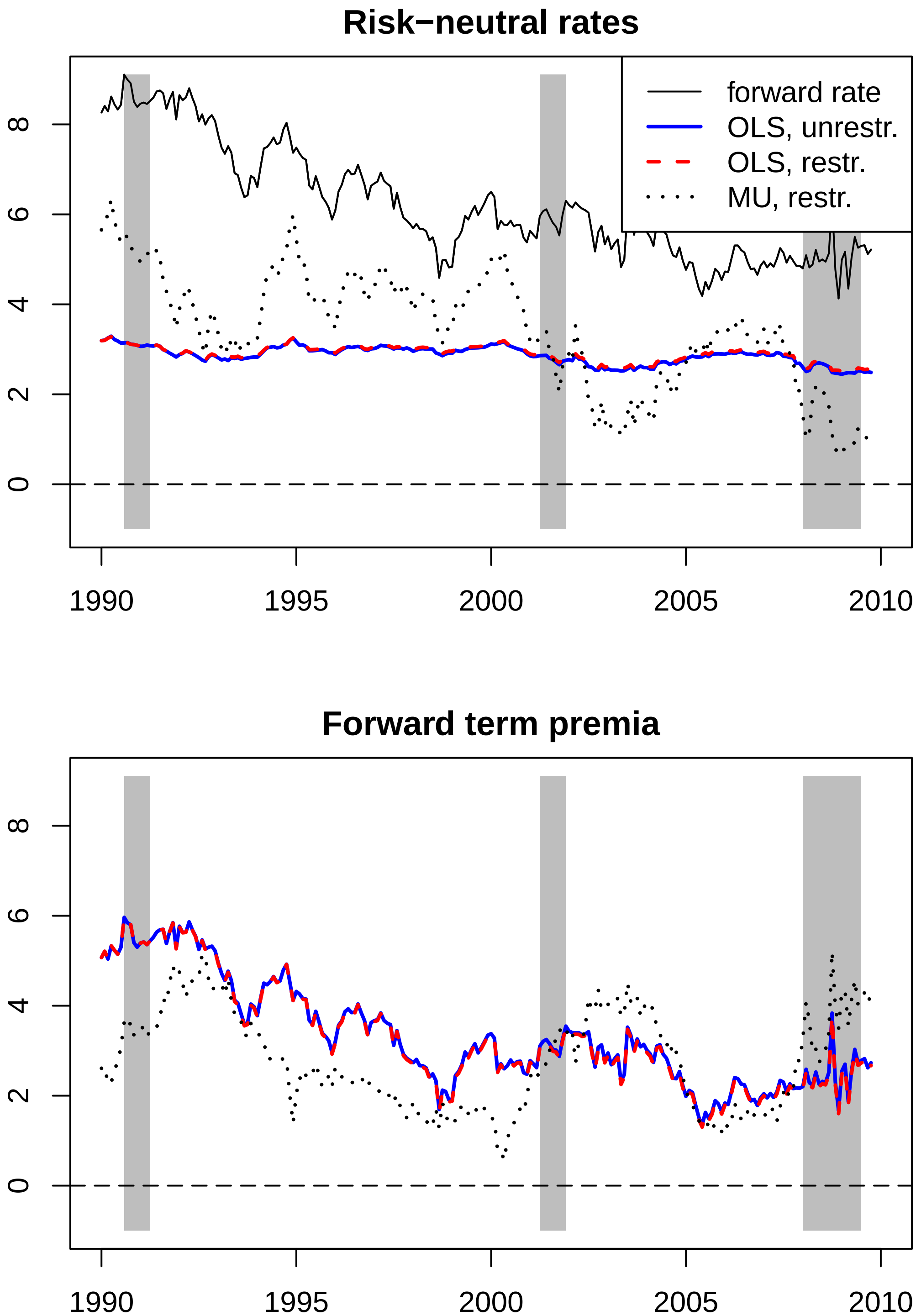
The top panel of Figure 2 shows the time series of the fitted forward rate (which is essentially identical across estimates) and the three alternative estimates of the risk-neutral forward rate, and the bottom panel shows the alternative estimates of the forward term premium. In Table 7 we show the sample standard deviations for the estimated rates and premia.

Imposing the restrictions we choose hardly has any effect on estimated risk-neutral forward rates and term premia—the two series corresponding to OLS-unrestricted and OLS-restricted in each panel are essentially indistinguishable. A look at the summary statistics reveals that the restrictions make the risk-neutral forward rate slighly less volatile and the forward term premium slightly more volatile. These differences however are not economically significant, as evident from the graph.

As in the case of the JSZ model and data set in Section 4, bias-correcting the DTSM estimates has important economic consequences. The observations to be made here very much parallel the ones from our previous case study. Risk-neutral forward rates, for OLS close to constant around 3%, become much more volatile and contribute more to the secular decline in forward rates. The forward term premium does not move in tandem with the actual forward rate anymore, but instead shows interesting independent time variation. It displays a distinct countercyclical pattern, in line with the conventional macro wisdom about risk premia.

Our sample includes data up to December 2009, after the end of the most recent recession.

Figure 2: HW case study – decomposition of forward rates



Five-to-ten year forward rates decomposed into risk-neutral forward rates and forward term premia using three different sets of estimates: (i) OLS, unrestricted, (ii) OLS, restricted, and (iii) MU, restricted. For details, refer to main text. Sample: Monthly observations from January 1990 to December 2009. Top panel: forward rate and alternative estimates of risk-neutral forward rates. Bottom panel: alternative estimates of the forward term premium. Gray shaded areas correspond to NBER recessions.

Table 7: HW case study – summary statistics

	OLS-UR	OLS-R	MU-R
$\sigma(f_t^{61,120})$	1.257	1.257	1.257
$\sigma(\tilde{f}_t^{61,120})$	0.195	0.179	1.286
$\sigma(ftp_t^{61,120})$	1.166	1.182	1.086

Notes: Summary statistics for OLS-unrestricted (OLS-UR), OLS-restricted (OLS-R) and MU-restricted (MU-R) estimates – standard deviations of the fitted five-to-ten-year forward rates and of the corresponding risk-neutral forward rates and forward term premia.

In contrast to the conventional estimates, MU estimation implies a term premium that increases significantly before and during the times of that economic downturn. This is a salient feature of our results, which demonstrates, along with several other observations we made in this paper, the economic plausibility of the risk premium estimates obtained using our unbiased estimation framework.

7 Conclusion

Correcting for finite-sample bias in estimates of the VAR dynamics of affine term structure models has dramatic implications for the estimated persistence of interest rates and for the inference about term premia. Risk-neutral rates, which reflect expectations of future monetary policy, show significantly more variation if one performs unbiased estimation of the underlying VAR dynamics instead of the commonly used OLS/ML estimation. Our paper shows how one can overcome the problem of artificially stable far-ahead short rate expectations that several previous studies have criticized.

In addition to showing how unbiased estimation methods can be brought to bear on the problem of estimating affine Gaussian DTSMs, we make a methodological contribution to the field of time series econometrics: We show how to obtain unbiased estimates of the parameters in VAR models, using a simulation-based inference method that we call inverse bootstrap. The method is superior to conventional bootstrap bias correction in terms of accuracy, and is novel specifically in the way our algorithm prescribes how to construct a sequence of estimates that converges in an optimal way to the solution of the problem.

Our paper is the first to quantify the bias in estimates of DTSMs and opens up several promising directions for future research. In particular, the question of how other methods that aim at improving the specification and/or estimation of the dynamic system fare in terms of bias reduction can be answered using our framework. Among the approaches that have been proposed in the literature are inclusion of survey information (Kim and Orphanides, 2005), restrictions on risk pricing (Bauer, 2011; Joslin et al., 2010b), near-cointegrated specification of the VAR dynamics (Jardet et al., 2009) and fractional integration (Gil-Alana and Moreno, 2007; Schotman et al., 2008). The key open question is how model-implied short rate forecasts that result from these different approaches compare to those that are based on unbiased estimates of a stationary VAR. In terms of extensions of our approach, generalizing it to the context of non-affine and non-Gaussian term structure models are the logical next steps, in particular allowing for unbiased estimation of DTSMs with stochastic volatility.

References

Amihud, Yakov and Clifford M. Hurvich, "Predictive Regressions: A Reduced-Biad Estimation Method," *Journal of Financial and Quantitative Analysis*, December 2004, *39* (4), 813–841.

Andrews, Donald W. K., "Exactly Median-Unbiased Estimation of First Order Autoregressive/Unit Root Models," *Econometrica*, 1993, *61* (1), 139–165.

Andrews, D.W.K. and H.Y. Chen, "Approximately median-unbiased estimation of autoregressive models," *Journal of Business & Economic Statistics*, 1994, *12* (2), 187–204.

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

_ , Jean Boivin, and Sen Dong, "Monetary Policy Shifts and the Term Structure," NBER Working Paper 15270, National Bureau of Economic Research, Inc August 2009.

Bauer, Michael D., "Term Premia and the News," Working Paper 2011-03, Federal Reserve Bank of San Francisco Jan 2011.

Campbell, John Y. and John H. Cochrane, "By force of habit: A consumption-based explanation of aggregate stock market behavior," *Journal of Political Economy*, 1999, *107* (2), 205–251.

Christensen, Jens H. E., Francis X. Diebold, and Glenn D. Rudebusch, "An Arbitrage-Free Generalized Nelson–Siegel Term Structure Model," *Econometrics Journal*, November 2009, *12* (3), C33–C64.

Christensen, Jens H.E., Francis X. Diebold, and Glenn D. Rudebusch, "The Affine Arbitrage-Free Class of Nelson-Siegel Term Structure Models," *forthcoming in Journal of Econometrics*, 2011.

Cochrane, John H. and Monika Piazzesi, "Decomposing the Yield Curve," manuscript 2008.

Dai, Qiang and Kenneth J. Singleton, "Specification Analysis of Affine Term Structure Models," *Journal of Finance*, Oct 2000, *55* (5), 1943–1978.

_ and _ , "Expectation puzzles, time-varying risk premia, and affine models of the term structure," *Journal of Financial Economics*, Mar 2002, *63* (3), 415–441.

Duffee, Gregory R., "Term Premia and Interest Rate Forecasts in Affine Models," *Journal of Finance*, 02 2002, *57* (1), 405–443.

_ , "Forecasting with the Term Structure: the Role of No-Arbitrage," Working Paper June, Haas School of Business, University of California, Berkeley 2009.

_ and Richard H. Stanton, "Estimation of Dynamic Term Structure Models," working paper, Haas School of Business Mar 2004.

_ and _ , "Evidence on Simulation Inference for Near Unit-Root Processes with Implications for Term Structure Estimation," *Journal of Financial Econometrics*, Winter 2008, *6* (1), 108–142.

Efron, Bradley and Robert J. Tibshirani, *An introduction to the bootstrap*, Chapman & Hall/CRC, 1993.

Engsted, Tom and Thomas Q. Pedersen, "Return Predictability and Intertemporal Asset Allocation: Evidence from a Bias-Adjusted VAR Model," *SSRN eLibrary*, 2008.

Gil-Alana, L. and A. Moreno, "Uncovering the US Term Premium: An Alternative Route," Working Paper 2007.

Gourieroux, Christian and Alain Monfort, *Simulation-based econometric methods*, Oxford University Press, 1996.

_ , Eric Renault, and Nizar Touzi, "Calibration by simulation for small sample bias correction," in Robert Mariano, Til Shcuermann, and Melvyn J. Weeks, eds., *Simulation-based Inference in Econometrics: Methods and Applications*, Cambridge University Press, 2000, chapter 13, pp. 328–358.

Hall, Peter, *The bootstrap and Edgeworth expansion*, Springer Verlag, 1992.

Hamilton, James D., *Time Series Analysis*, Princeton University Press, 1994.

— **and Jing (Cynthia) Wu**, "Identification and Estimation of Affine Term Structure Models," working paper, University of California, San Diego July 2010.

Harvey, Campbell R., "Time-varying conditional covariances in tests of asset pricing models," *Journal of Financial Economics*, 1989, *24* (2), 289–317.

Horowitz, Joel L., "The Bootstrap," in J.J. Heckman and E.E. Leamer, eds., *Handbook of Econometrics*, Vol. 5, Elsevier, 2001, chapter 52, pp. 3159–3228.

Jardet, Caroline, Alain Monfort, and Fulvio Pegoraro, "No-arbitrage near-cointegrated VAR(p) term structure models, term premia and GDP growth," working paper, Banque de France Jun 2009.

Joslin, Scott, Kenneth J. Singleton, and Haoxiang Zhu, "A New Perspective on Gaussian DTSMs," *forthcoming in Review of Financial Studies*, 2010.

— **, Marcel Priebsch, and Kenneth J. Singleton**, "Risk Premiums in Dynamic Term Structure Models with Unspanned Macro Risks," working paper May 2010.

Kilian, Lutz, "Small-sample confidence intervals for impulse response functions," *Review of economics and statistics*, 1998, *80* (2), 218–230.

Kim, Don H. and Athanasios Orphanides, "Term Structure Estimation with Survey Data on Interest Rate Forecasts," Computing in Economics and Finance 2005 474, Society for Computational Economics November 2005.

— **and Jonathan H. Wright**, "An arbitrage-free three-factor term structure model and the recent behavior of long-term yields and distant-horizon forward rates," Finance and Economics Discussion Series 2005-33, Board of Governors of the Federal Reserve System (U.S.) 2005.

Kozicki, S. and P.A. Tinsley, "Shifting endpoints in the term structure of interest rates," *Journal of Monetary Economics*, 2001, *47* (3), 613–652.

Lustig, Hanno N., Nikolai L. Roussanov, and Adrien Verdelhan, "Countercyclical Currency Risk Premia," NBER Working Papers 16427, National Bureau of Economic Research, Inc September 2010.

MacKinnon, James G. and Anthony A. Jr. Smith, "Approximate bias correction in econometrics," *Journal of Econometrics*, 1998, *85* (2), 205–230.

Marriott, F. H. C. and J. A. Pope, "Bias in the Estimation of Autocorrelations," *Biometrika*, 1954, *41* (3/4), 390–402.

Meerschaert, M.M. and H.P. Scheffler, *Limit distributions for sums of independent random vectors: Heavy tails in theory and practice*, Wiley-Interscience, 2001.

Nicholls, DF and AL Pope, "Bias in the estimation of multivariate autoregressions," *Australian & New Zealand Journal of Statistics*, 1988, *30* (1), 296–309.

Pierse, R.G. and A.J. Snell, "Temporal aggregation and the power of tests for a unit root," *Journal of Econometrics*, 1995, *65* (2), 333–345.

Polyak, B.T. and A.B. Juditsky, "Acceleration of stochastic approximation by averaging," *SIAM Journal on Control and Optimization*, 1992, *30*, 838.

Robbins, H. and S. Monro, "A stochastic approximation method," *The Annals of Mathematical Statistics*, 1951, *22* (3), 400–407.

Rudebusch, Glenn D., "Trends and Random Walks in Macroeconomic Time Series: A Re-Examination," *International Economic Review*, 1992, *33* (3), 661–680.

_ , "The Uncertain Unit Root in Real GNP," *International Economic Review*, March 1993, *83* (1), 264–272.

_ and Tao Wu, "A Macro-Finance Model of the Term Structure, Monetary Policy and the Economy," *Economic Journal*, 07 2008, *118* (530), 906–926.

_ , Eric T. Swanson, and Tao Wu, "The Bond Yield "Conundrum" from a Macro-Finance Perspective," *Monetary and Economic Studies*, December 2006, *24* (S1), 83–109.

Schotman, P.C., R. Tschernig, and J. Budek, "Long memory and the term structure of risk," *Journal of Financial Econometrics*, 2008, *6* (4), 459.

Stine, Robert A. and Paul Shaman, "A Fixed Point Characterization for Bias of Autoregressive Estimators," *Annals of Statistics*, 1989, *17* (3), 1275–1284.

Tang, C.Y. and S.X. Chen, "Parameter estimation and bias correction for diffusion processes," *Journal of Econometrics*, 2009, *149* (1), 65–81.

Tanizaki, Hisashi, "Bias correction of OLSE in the regression model with lagged dependent variables," *Computational Statistics & Data Analysis*, 2000, *34* (4), 495–511.

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

Wright, Jonathan H., "Term Premia and Inflation Uncertainty: Empirical Evidence from an International Panel Dataset," *American Economic Review*, 2011, *forthcoming.*

Yamamoto, T. and N. Kunitomo, "Asymptotic bias of the least squares estimator for multivariate autoregressive models," *Annals of the Institute of Statistical Mathematics*, 1984, *36* (1), 419–430.