Journal of Empirical Finance 53 (2019) 272–290

Contents lists available at ScienceDirect

Journal of Empirical Finance

journal homepage: www.elsevier.com/locate/jempfin

Bond and option prices with permanent shocks

Haitham A. Al-Zoubi *,1

The College of Business, Alfaisal University, Riyadh, Saudi Arabia

ARTICLE INFO

JEL classification:
C15
C32
G12.

Keywords:
Short-term interest rate
Stochastic volatility
Continuous-time estimation
Jackknife

ABSTRACT

I develop and estimate an affine short-rate model that incorporates a *nonstationary* stochastic mean. In my model, the time-varying stochastic mean is subject to a sequence of permanent shocks that can better capture the source of nonlinearity in the drift than existing models. I find that the proposed model provides a better in-sample and out-of-sample fit to observed interest rates and bond prices relative to extant models. More specifically, my model outperforms constant elasticity of volatility models. It follows that the nonstationary stochastic mean model offers new insights to the implied bond option valuation and accounts for the downward bias in bond option prices generally documented in the literature.

1. Introduction

In the finance literature, the diffusion process has commonly been used to model the stochastic dynamics of interest rates following the work of Vasicek (1977) and Cox et al. (1985). A prominent empirical finding in this literature is that models that allow the volatility of interest rates to be heteroskedastic and highly sensitive to interest rate changes are superior to mean reverting models (Chan et al., 1992, hereafter CKLS). In addition to the mean reverting models being rejected, it is often found that the drift term of the stochastic process of interest rates is nonlinear, (e.g., Aït-Sahalia, 1996a).

In this paper, I consider the view of interest rates proposed by Fama (2006) who asserts the following in the context of the forecasting power of forward rates for spot rates: "The inference that this forecast power is due to mean reversion of the spot rate toward a constant expected value no longer seems valid. Instead, the predictability of the spot rate captured by forward rates seems to be due to mean reversion toward a time-varying expected value that is subject to a sequence of apparently permanent shocks that are on balance positive to mid-1981 and on balance negative thereafter" (p. 1). Cochrane and Piazzesi (2005) find that time-varying bond risk premiums and excess returns are largely unexplained by the information contained in the yield curve; a study by Cieslak and Povala (2015) demonstrates that expected inflation contains substantial predictive power for excess bond returns beyond the level, slope, and curvature of the yield curve. Duffee (2018) documents that expected inflation shocks accounts for up to 20% of variances of bond yields. A paper by Cooper and Priestley (2008) documents how the output gap contains substantial predictive power for excess bond returns, while Greenwood and Vayanos (2014) observe that measures of Treasury bond supply also help predict excess returns. Moreover, Joslin et al. (2014) argue that macroeconomic fundamentals like economic growth and inflation are important factors that can predict excess returns.[2] In line with this perspective, I modify the Vasicek model of interest rates by

* Correspondence to: The College of Business, Alfaisal University, P. O. Box 50927, Riyadh 1153, Saudi Arabia.
E-mail address: halzoubi@alfaisal.edu.

[1] I would like to thank anonymous referees, Pietro Veronesi, Robert de Jong and Jun Yu for their helpful comments and suggestions. I would also like to thank my father, my mother, my wife and my brothers and sisters for their unconditional and endless support. The project is partially supported by Alfaisal University research, Saudi Arabia grant N. 18102.

[2] Liu (2018) provides the out-of-sample evidence of economic growth and inflation on predicting excess returns. However, Bauer and Hamilton (2017) find that the statistical tests that have been used to support these findings are subject to size distortions resulting from persistent regressors and the lack of strict exogeneity of the independent variables.

https://doi.org/10.1016/j.jempfin.2019.07.010
Received 24 December 2017; Received in revised form 30 July 2019; Accepted 31 July 2019
Available online 2 August 2019

incorporating a *nonstationary* stochastic mean that allows for a time-varying expected value and possible nonlinearity in the drift function.[3] My proposed model differs from existing models such as Balduzzi et al. (1998) and Cai and Swanson (2011), in which they assume that interest rates revert to a *stationary* moving mean.[4] I find that my model provides a better fit to observed interest rates relative to a large number of alternative models. I derive the implied bond option price from my model, and show that it mitigates the downward bias in existing bond option valuation models.

The evolution of the literature on estimating spot interest rates falls into two broad classes: (1) the estimation bias and (2) the specification bias. The first class, estimation bias, comprises econometric measurement problems related to discretization bias and possible finite sample bias. Discretization bias arises when one estimates continuous time models with data that are in discrete-time. Influential contributions include the Generalized Method of Moments (Hansen and Scheinkman, 1995), Efficient Method of Moments (Gallant and Tauchen, 1996), Approximate Maximum Likelihood (Aït-Sahalia, 2002), and GMM with random parameters (Aït-Sahalia and Mykland, 2003). On the other hand, finite sample bias arises when the asymptotic distribution of the estimator tends to be a poor approximation of the finite sample distribution. Important contributions include the Jackknife Method (Phillips and Yu, 2005), the Parametric Bootstrap Method (Tang and Chen, 2009), and Least Square Bias Approximation (Yu, 2012).

The second class, specification bias, arises from the findings of CKLS that the mean reversion parameter is insignificant and the volatility parameter is the cornerstone in model specification. Using a nonparametric test based on Hermite polynomials, Aït-Sahalia (1996a) provides evidence that the drift function is nonlinear and concludes that the drift is more critical than the volatility in model specification. The same conclusion is also documented by Stanton (1997), Jiang (1998), and Al-Zoubi (2009). However, Chapman and Pearson (2000) show that even when the true model has a linear drift, the nonparametric test used in Aït-Sahalia (1996b) tends to generate nonlinearities in the drift function. Hong and Li (2005) correct a boundary bias in the Aït-Sahalia (1996b) test and strongly reject all mean-reverting models. They conclude that nonlinear drift specification does not improve the goodness of fit. However, Arapis and Gao (2006) propose a specification test for linearity in the drift and conclude that nonlinear drift plays a prominent role in derivatives pricing.

My estimation using a nonstationary mean for the spot rate involves comparing the proposed model with five affine term structure models for interest rates. My main finding is that the nonstationary time-varying mean model is better able to capture the interest rate, both in stable and unstable periods. In fact, the fits of existing models are sample dependent and the volatility function seems to be a critical component only in unstable periods. However, when the nonstationary mean is included in the model, with or without a correction for small-sample bias, the particular specification of the volatility function is less crucial. This result differs from Durham (2003) and Chan et al. (1992), who find that the behavior of interest rate is better captured by the constant elasticity of volatility (CEV) models. Basically, the advantage of CEV models is that they better capture the nonstationary mean during unstable periods. Interestingly, this benefit disappears against my proposed nonstationary mean model.

Within the same CEV specification, Aït-Sahalia (1996a) introduces a nonlinear mean-reverting model of interest rates with constant elasticity of volatility. His model allows for slow mean reversion while interest rates are in the middle of a band and strong nonlinear mean reversion while they are out of the band. I find significant evidence of nonlinearity in the drift. In fact, I find that the nonlinearity in the drift relies crucially on the direction in which interest rates move. In the recent period of low interest rates, I find largely modest evidence that the interest rate tends to revert strongly to its mean. Specifically, for periods of stable interest rates, when interest rate values are very low, a non-linear, mean-reverting model generates significant pricing errors.

My results are in line with the opportunistic approach to monetary policy developed by Orphanides and Wilcox (2002) and Aksoy et al. (2006), in which out-of-the-band inflation stimulates opportunistic policy action that re-enforces nonlinear mean reversion in the interest rate. However, when inflation is modest but still above the long-run target, the Fed does not take pro-active anti-inflation action, but waits for external forces such as beneficial supply shocks and unanticipated recessions to achieve the preferred level of inflation. Such policy is consistent with "a policy-inaction band" in which the interest rate exhibits weak or no mean reversion.[5] In contrast, out-of-the-band inflation stimulates opportunistic policy action that produces strong nonlinear mean reversion in the interest rate. In short, an additional benefit of models with nonstationary stochastic mean is that they better capture nonlinearities in the interest rate due to opportunistic monetary policies.

My analytic formulas for bond and bond option prices under my stochastic mean framework significantly attenuate the downward bias in bond option prices that is documented, for example, by Phillips and Yu (2005) and Tang and Chen (2009). In fact, my model reduces the bias in bond prices by 62%–64% and the bias in option prices by 18.5%–65% compared to the standard Vasicek (1977) model. My analysis indicates that non-stationarity in the short rate's mean is the primary reason for the downward bias inherent in existing bond option pricing models.

The remainder of the paper is as follows. Section 2 develops the dynamics of interest rates with nonstationary mean and introduces bond prices under my proposed framework. Section 3 reviews continuous time models to be considered. Section 4 outlines the data, details the specification tests used to select the best model, analyzes interest rate dynamics, and reports parameter estimates. Section 5 estimates bond and option prices using spot rates with nonstationary mean. Section 6 concludes.

[3] Bandi (2002) provides evidence that the interest rate displays martingale behavior over a range between 3% and 15%. Yu (2012) documents slow mean reversion in interest rates and shows that the nonlinear term suggested by Aït-Sahalia (1996a) is important in the near unit root situation. The slow mean reversion process is also studied by Bandi (2002) and Yu (2012).

[4] Duffee (2002) documents that the random walk forecasts outperform those from the Dai and Singleton (2000) affine model. Duffee (2011) finds that a term structure model that imposes a random walk component is superior to existing term structure models.

[5] Kozicki and Tinsley (2001) find that persistent interest rates results from historical shifts in market perceptions of the policy targeting inflation.

2. A continuous time interest rate model with *I(2)* component

In this section, I develop analytic expressions for the spot rates and for the bond price. Both expressions are articulated in terms of an continuous-time analogue of the integrated Gaussian unit root, I(2), process. I refer to continuous time interest rate model with *I(2)* component as the HZ model.

2.1. Interest rate dynamics under nonstationary mean

In my model, the spot rate, r_t, has a permanent component that is an integrated random walk process. Following Ravn and Uhlig (2002), the general representation of the nonstationary stochastic mean, μ_t, can be written as

$$d\mu_t = \vartheta_t dt, d\vartheta_t = \sigma_\mu dW_t \tag{1}$$

My model, as of most of the models in this area, is within the affine framework characterized by Duffie and Kan (1996). Uncertainty is generated by n independent Brownian motions. The instantaneous interest rate, denoted $\dot{r}_t$, is affine in the state:

$$\dot{r}_t = \delta_0 + \delta r_t$$

The equivalent martingale measure of r_t determines the bond price. The risk-neutral dynamics of the interest rate are

$$dr_t = \beta\left(r_t - \mu_t\right)dt + \sigma dZ_t \tag{2}$$

$$d\mu_t = \vartheta_t dt, \quad d\vartheta_t = \sigma_\mu dW_t$$

where dZ_t and dW_t are two independent Brownian motions. Note that under the physical measure the dynamics of μ_t are also Gaussian. I follow Ravn and Uhlig (2002) and McElroy and Trimbur (2007) in approximating the nonstationary mean in (2) by solving

$$\underset{\mu_\omega}{argmin}\left\{\sum_{t=1}^{T}\left[r_t - \mu_t\right]^2 + (1/q)\int_{\varphi}^{\omega}\left[D^2\mu_u\right]^2 du\right\}, q > 0 \tag{3}$$

where $D^2\mu_u$ denotes the second derivative of the function D over the space of all functions on the interval $[\varphi, \omega]$. The irregular component $\left(r_t - \mu_t\right)$ of the objective function mimics the mean reversion of the expected return on bonds which is due to mean reversion in expected rate of inflation as in Fama (2006). The second term $(1/q)\int_{\varphi}^{\omega}\left[D^2\mu_u\right]^2 du$ mimics the permanent shocks in the mean which is due to occasional Federal Reserve actions to control the interest rate. The parameter, $q \geq 0$, determines the smoothness of the nonstationary stochastic mean.[6]

Following McElroy and Trimbur (2007), I may write the differentiated stochastic mean and spot rate as:

$$d^2\mu_t = \eta_t,$$
$$w_t = d^2 r_t = \eta_t + d^2\left(r_{t-}\mu_t\right) = \eta_t + d^2\left(c_t\right),$$

where η_t is a white noise uncorrelated with the stationary component of interest rate $c_t = r_t - \mu_t$.

Therefore,

$$\sigma^2 = \sigma_\mu^2 + \sigma_c^2, \tag{4}$$

$$\sigma_c^2 = \frac{\sigma^2_{D^2\mu}}{q}.$$

Following McElroy and Trimbur (2011), I derive the continuous time version of the filter on the basis of signal extraction theory. Let L be the continuous-time lag operator defined by $L^x r(t) = r(t-x)$ for any x and for all times. Let $e^{-\lambda i}$ be a frequency response function, observed at point i. The associated spectral densities of the differentiated stochastic mean and interest rate are

$$f_u(\lambda) = q\sigma_c^2, \qquad f_w(\lambda) = f_u(\lambda) + \lambda^4\sigma_c^2 = \left(q + \lambda^4\right)\sigma_c^2.$$

The frequency response of the stationary component is given by the ratio $\left(1 + q/\lambda^4\right)^{-1}$ and the stationary spectrum is given by $\sigma_c^2\left(1 + q/\lambda^4\right)^{-1}$. Taking the inverse Fourier transformation of the spectrum gives the weighting kernel

$$\psi(x) = \frac{q^{1/4}\exp\left\{-|x|\,q^{1/4}/\sqrt{2}\right\}}{2\sqrt{2}}\left(\cos\left(|x|\,q^{1/4}/\sqrt{2}\right) + \sin\left(|x|\,q^{1/4}/\sqrt{2}\right)\right).$$

I move on to state the limiting distribution of the spot rate

---

[6] Note that the model corresponds to the smoothing polynomial splines of Wecker and Ansley (1983) where the first derivative is continuous and the second derivative square is integrable.

Proposition 2.1. *Let r_t be given by (2), the expected value and variance of the spot rate are given by*

$$E\left[r_t\right] = e^{\beta t}\left[r_0 + E\left[\mu_t\right]\left(e^{-\beta t} - 1\right)\right]$$

and

$$Var\left(r_t\right) = \frac{1}{2}\sigma_m^2 e^{2\beta} t^2 + \sigma^2\left(\frac{1 - e^{2\beta t}}{-2\beta}\right)$$

Proof. Applying Ito's lemma I have that

$$r_t = e^{\beta t}\left[r_0 - \int_0^t \beta\mu_t e^{-\beta u} du + \sigma\int_0^t e^{-\beta u} dZ_u\right].$$

Using integration by parts for the second term it follows that

$$r_t = e^{\beta t}\left[r_0 - \beta\mu_t\int_0^t e^{-\beta u} du + \beta\int_0^t d\mu_t\left(\int_0^t e^{-\beta t} dt\right) du + \sigma\int_0^t e^{-\beta u} dZ_u\right]$$

$$= e^{\beta t}\left[r_0 + \mu_t\left(e^{-\beta t} - 1\right) - \int_0^t \vartheta_t du\left(e^{-\beta t} - 1\right) du + \sigma\int_0^t e^{-\beta u} dZ_u\right]$$

$$= e^{\beta t}\left[r_0 + \mu_t\left(e^{-\beta t} - 1\right) + \sigma\int_0^t e^{-\beta u} dZ_u\right].$$

$$E\left[r_t\right] = e^{\beta t}\left[r_0 + E\left[\mu_t\right]\left(e^{-\beta t} - 1\right) + E\left[\sigma\int_0^t e^{-\beta u} dZ_u\right]\right]$$

By Ito's isometry I have that,

$$E\left[r_t\right] = e^{\beta t}\left[r_0 + E\left[\mu_t\right]\left(e^{-\beta t} - 1\right)\right].$$

The variances of the permanent and transitory components can be written as,

$$Cov\left[\mu_t, \mu_u\right] = \sigma_m^2 E\left[\int_0^t\int_0^t dW_s \int_0^u\int_0^u dW_s\right]. \text{ Thus,}$$

$$Var\left(\mu_t\right) = \sigma_m^2\left[\int_0^t\int_0^t ds\right] = \sigma_m^2\int_0^t (t\Lambda u) = \sigma_m^2\frac{t^2}{2}, \text{ and}$$

$$Var\left(\mu_t\left(e^{-\beta t} - 1\right)\right) = \sigma_m^2\left[\int_0^t\int_0^t ds\right] = \sigma_m^2\int_0^t (t\Lambda u) = \frac{1}{2}\sigma_m^2 t^2 e^{\beta t},$$

$$Var\left(\sigma\int_0^t e^{-\beta u} dZ_u\right) = \sigma^2\left(\frac{1 - e^{2\beta t}}{-2\beta}\right).$$

2.2. Bond price with a nonstationary I(2) mean

Proposition 2.2. *Let$P(t,T)$ denotes the price of a zero-coupon bond and let T be the maturity of that bond. The risk neutral bond price is*

$$P\left(t,T,r_t\right) = Exp\left(\left[-\int_t^T r_u\left(r_t\right) du\right] + \frac{1}{2}Var\left[-\int_t^T r_u\left(r_t\right) du\right]\right)$$
$$= Exp\left(A(t,T) r_t - \mu_t\left(A(t,T) + (T-t)\right) + B(t,T) + D(t,T)\right),$$

where

$$A(t,T) \equiv \left(\frac{1 - e^{\beta(T-t)}}{\beta}\right),$$

$$B(t,T) = \frac{\sigma^2}{2\beta^2(1+q)}\left(A(t,T) + (T-t)\right) + \frac{\sigma^2 A(t,T)^2}{4\beta(1+q)},$$

and

$$D(t,T) = \frac{q\sigma^2(T-t)^4}{48(1+q)}$$
$$= Exp\left(\left(\frac{1 - e^{\beta(T-t)}}{\beta}\right) r_t + \left(\frac{1 - e^{\beta(T-t)}}{\beta} + (T-t)\right)\mu_t + \frac{\sigma^2}{2\beta^2(1+q)}\left(\frac{1 - e^{\beta(T-t)}}{\beta}\right)\right.$$
$$\left. + \frac{\sigma^2}{2\beta^2(1+q)}(T-t) + \frac{\sigma^2}{4\beta(1+q)}\left(\frac{1 - 2e^{\beta(T-t)} + e^{2\beta(T-t)}}{2\beta^2}\right) + \frac{q\sigma^2(T-t)^4}{48(1+q)}\right)$$

Proof.

$$P(t,T) = E\left[\exp\left(-\int_0^T r_u du\right)|F_t\right].$$

Let the transitory complement be

$$c_u = r_u - \mu_u \tag{5}$$

where c_u follows the Ornstein–Uhlenbeck equation given by

$$c_u = e^{\beta u}\left(c_0 + \int_0^u \sigma_c e^{-\beta s} dZ_s\right). \tag{6}$$

with risk-neutral dynamics given by

$$dc_t = \beta c_t + \sigma_c dZ_t$$

The autocovariance of the transitory component is given by

$$\begin{aligned} Cov\left[c_t, c_u\right] &= \sigma_c^2 e^{\beta(u+t)} E\left[\int_0^t e^{-\beta s} dZ_s \left[\int_0^u e^{-\beta s} dZ_s\right]\right] \\ &= \sigma_c^2 e^{\beta(u+t)} \int_0^{u\Lambda t} e^{-2\beta s} ds = \frac{\sigma_c^2}{-2\beta} e^{\beta(u+t)}\left(e^{2\beta(u\Lambda t)} - 1\right). \end{aligned}$$

Therefore,

$$\begin{aligned} Var\left[\int_0^t c_u du\right] &= Cov\left(\int_0^t c_u du, \int_0^t c_u ds\right) \\ &= \int_0^t \int_0^t Cov\left(c_u, c_s\right) du ds = \int_0^t \int_0^t \frac{\sigma_c^2}{-2\beta} e^{\beta(u+s)}\left(e^{-2a(u\Lambda s)} - 1\right) du ds = \frac{\sigma_c^2}{2\beta^3}\left(2\beta t + 3 - 4e^{\beta t} + e^{2\beta t}\right). \end{aligned}$$

Likewise, the autocovariance of the permanent component is given by,

$$\begin{aligned} Cov\left[\mu_t, \mu_u\right] &= \sigma_m^2 E\left[\int_0^t \int_0^t dW_s \int_0^u \int_0^u dW_s\right] \\ &= \sigma_m^2\left[\int_0^{t\Lambda u} \int_0^{t\Lambda u} ds\right] = \sigma_m^2 \int_0^{t\Lambda u} (t\Lambda u) = \sigma_m^2 \frac{(t\Lambda u)^2}{2}. \end{aligned}$$

Therefore,

$$Var\left[\int_0^t \mu(u) du\right] = \sigma_m^2 \int_0^t \int_0^t \frac{(s\Lambda u)^2}{2} du ds = \sigma_m^2 \frac{t^4}{24}.$$

From (5), I have

$$E\left[-\int_0^t r_u du\right] = E\left[-\int_0^t \left(c_u + \mu_u\right) du\right]$$

Thus,

$$E\left[-\int_t^T r_u du\right] = \frac{r_t - \mu_t}{\beta}\left(1 - e^{\beta(T-t)}\right) - \mu_t (T-t) \tag{7}$$

Moreover,

$$\begin{aligned} Var\left[-\int_t^T r_u du\right] &= Var\left[\int_t^T \left(c_u + \mu_u\right) du\right] \\ &= \frac{\sigma_c^2}{2\beta^3}\left(2\beta(T-t) + 3 - 4e^{\beta(T-t)} + e^{2\beta(T-t)}\right) + \sigma_m^2 \frac{(T-t)^4}{24} \end{aligned} \tag{8}$$

I follow Harvey and Trimbur (2008) and model μ_t as

$$\mu_{t+1} = \mu_t + x_t$$
$$x_{t+1} = x_t + \eta_t$$

Because μ_t is an integrated unit root, it follows that

$$\sigma_{\nabla_m}^2 = t\sigma_\eta^2,$$

and

$$\sigma_{\nabla_\mu^2}^2 = t - (t-1)\sigma_\eta^2 = \sigma_\eta^2.$$

Therefore,

$$q = \frac{\sigma^2_{\nabla^2_\mu}}{\sigma^2_c} = \frac{\sigma^2_\eta}{\sigma^2_c} = \frac{\sigma^2_\mu}{\sigma^2_c}.$$

Combining Eqs. (4), (7) and (8), the bond price is given by

$$\begin{aligned}
P\left(t,T,r_t\right) &= \exp\left(\left[-\int_t^T r_u\left(r_t\right)du\right] + \frac{1}{2}Var\left[-\int_t^T r_u\left(r_t\right)du\right]\right)\\
&= Exp\left(\frac{r_t-\mu_t}{\beta}\left(1-e^{\beta(T-t)}\right) - \mu_t\left(T-t\right) + \frac{\sigma_c^2}{4\beta^3}\left(2\beta\left(T-t\right)+3-4e^{\beta(T-t)}+e^{2\beta(T-t)}+\frac{\sigma_\mu^2}{\sigma_c^2}\frac{\beta^3\left(T-t\right)^4}{12}\right)\right)\\
&= Exp\left(\frac{r_t-\mu_t}{\beta}\left(1-e^{\beta(T-t)}\right) - \mu_t\left(T-t\right) + \frac{\sigma^2}{4\beta^3\left(1+q/2\right)}\left(2\beta\left(T-t\right)+3-4e^{\beta(T-t)}+e^{2\beta(T-t)}+q\frac{\beta^3\left(T-t\right)^4}{12}\right)\right)\\
&= Exp\left(\left(\left(\frac{1-e^{\beta(T-t)}}{\beta}\right)r_t + \left(\frac{1-e^{\beta(T-t)}}{\beta}+\left(T-t\right)\right)\mu_t + \frac{\sigma^2}{2\beta^2\left(1+q\right)}\left(\frac{1-e^{\beta(T-t)}}{\beta}\right)\right.\right.\\
&\quad \left.\left. + \frac{\sigma^2}{2\beta^2\left(1+q\right)}\left(T-t\right) + \frac{\sigma^2}{4\beta\left(1+q\right)}\left(\frac{1-2e^{\beta(T-t)}+e^{2\beta(T-t)}}{2\beta^2}\right) + \frac{q\sigma^2\left(T-t\right)^4}{48\left(1+q\right)}\right)\right.\\
&= Exp\left(A\left(t,T\right)r_t - \mu_t\left(A\left(t,T\right)+\left(T-t\right)\right) + B\left(t,T\right) + D\left(t,T\right)\right),
\end{aligned}$$

where

$$A\left(t,T\right) \equiv \left(\frac{1-e^{\beta(T-t)}}{\beta}\right),$$

$$B\left(t,T\right) = \frac{\sigma^2}{2\beta^2\left(1+q\right)}\left(A\left(t,T\right)+\left(T-t\right)\right) + \frac{\sigma^2 A\left(t,T\right)^2}{4\beta\left(1+q\right)},$$

and

$$D\left(t,T\right) = \frac{q\sigma^2\left(T-t\right)^4}{48\left(1+q\right)}.$$

3. Continuous time interest rate models

Although much of the focus of this paper is on single-factor models with constant elasticity of volatility, I consider the Vasicek (1977) model as my standard framework. First, I briefly review five competing single factor-models. My single factor models include the Vasicek model, the CIR (1985) model, the CKLS (1992) model, the Ahn and Gao (1999) model, and the Chapman and Pearson (2000) model. Following this introduction, I develop my nonstationary time-varying mean model and describe the econometric approach used in estimating the model.

3.1. Single-factor interest rate models

(1) The Chan, Karolyi, Longstaff and Sanders (CKLS) Model: I consider the CKLS (1992) parameterization of the interest rate diffusion process $\{r_t : t \geq 0\}$ that reads:

$$dr_t = (\alpha + \beta r_t)dt + \sigma r_t^\gamma dZ_t \tag{9}$$

where Z_t is a standard Brownian motion. α and β are the structural parameters of the conditional expected rate of change of the process (drift), and σ and γ are structural parameters of the conditional rate of change of volatility (diffusion). In this model, the interest rate reverts toward the expected value $-\alpha/\beta, -\beta$ measures the speed of mean reversion, and γ controls the sensitivity of the variance to the level of interest rates. When $\gamma > 1$, there is a so-called inverse leverage effect, in which the volatility of the interest rate increases as its level increases.

(2) The Vasicek (1977) Model: I consider a stochastic differential equation of the form:

$$dr_t = (\alpha + \beta r_t)dt + \sigma dZ_t$$

where α, β and σ are constants. The SDE is an Ornstein–Uhlenbeck process composed of a Brownian motion and a restoring drift that pushes it downwards when the process is above the mean $-\alpha/\beta$ and upwards when it is below. Thus, the distribution of the process is mean reverting and converges in equilibrium to a normally distributed mean $-\alpha/\beta$ and variance $\sigma^2/2\beta$. When $\alpha > 0, \beta < 0$, and $\gamma > 1/2$, the interest rate has a positive distribution.

(3) The Cox et al. (1985) (CIR) Model: In this model, the interest rate stochastic differential equation takes the form

$$dr_t = (\alpha + \beta r_t)dt + \sigma r_t^{1/2} dZ_t.$$

Different from the Vasicek model, the volatility term is decreasing with the level of the interest rate, allowing α to prevent the interest rate from going below zero. This condition holds as long as $\alpha \geq 1/2\sigma^2$.

(4) The Inverse CIR Model (AG): I follow Ahn and Gao (1999) and hypothesize that:

$$dr_t = (\alpha + \beta r_t + \theta r_t^2)dt + \sigma r_t^{1.5} dZ_t.$$

When $\alpha > 0$ and $\beta < 0$, the interest rate has a positive distribution. Different from the CIR model, the model generates a concave relationship between the interest rate and the yields.

(5) The Chapman and Pearson (CP) Model: I consider Chapman and Pearson (2000) representation for the following diffusion model:

$$dr_t = (\alpha + \beta r_t + \theta_1 r_t^2 + \theta_2 r_t^{-1})dt + \sigma r_t^{\gamma} dZ_t.$$

The model has a polynomial drift that enforces slow mean reversion while interest rates are in a band, and stronger nonlinear mean reversion while they are out of the band. This drift function has been investigated using a semiparametric procedure in Aït-Sahalia (1996b), using maximum likelihood with Hermite expansion in Aït-Sahalia (1999), Aït-Sahalia (2002), using Efficient Method of Moments in Gallant and Tauchen (1996), and using GMM in Chapman and Pearson (2000) and Czellar et al. (2007).

3.2. Econometrics

I follow Fama (2006) by assuming that the level of short-term interest rate is mean reverting to its long-term conditional mean which is subject to permanent shocks. The interest rate process, r_t, can be decomposed into two parts, (i) the long-term nonstationary stochastic mean, μ_t, which is unpredictable and (ii) a zero mean transitory component, c_t:

$$r_t = c_t + \mu_t, \qquad t = 1, \dots, T \tag{10}$$

I follow Hodrick and Prescott (1997) and Ravn and Uhlig (2002) and specify the permanent component, μ_t, by minimizing the loss function:

$$\sum_{t=1}^{T} \left[r_t - \mu_t\right]^2 + (1/q) \sum_{t=2}^{T} \left[\nabla^2 \mu_t\right]^2, \tag{11}$$

where ∇^2 denotes the second difference lag operator and T is the sample size. The first term in (11) penalizes the poorness of fit suggested by Fama (2006) that is due to the mean reversion of the expected return on bonds which in turn is due to mean reversion in expected rate of inflation, while the second term penalizes the variability in the mean which is in part due to the Federal Reserve actions to control money supply.

In this framework, it can be shown that there exists a unique solution to the minimization problem in (11). Denote F_T to be a known positive definite $(T \times T)$ matrix and denote I_T to be $(T \times T)$ identity matrix,

$$r = \left(\frac{F_T}{q} + I_T\right) \mu_T.$$

Danthine and Girardin (1989) show that $\mu_T = \left(\frac{F_T}{q} + I_T\right)^{-1} r_T$ is the solution for (11). Denote L to be the lag operator and L^{-1} to be the forward operator. Then, the first order condition for μ_t is

$$-2(r_t - \mu_t) - 4\frac{(\mu_{t+1} - 2\mu_t + \mu_{t-1})}{q} + 2\frac{(\mu_t - 2\mu_{t-1} + \mu_{t-2})}{q} + 2\frac{(\mu_{t+2} - 2\mu_{t+1} + \mu_t)}{q} = 0$$

which can be written as

$$r_t = \left(\frac{L^{-2}}{q} - \frac{4L^{-1}}{q} + \left(1 + \frac{6}{q} - 4\frac{L}{q} + \frac{L^2}{q}\right)\right) \mu_t,$$

$$r_t = \left(\frac{1}{q}\left(1 - L^{-1}\right)^2 (1 - L)^2 + 1\right) \mu_t.$$

Following Hodrick and Prescott (1997) and Ravn and Uhlig (2002) the solution for (11) is optimal for data generating process of the form

$$\Delta^2 \mu_t = \eta_t \sim NID\left(0, \sigma_\eta^2\right),$$

$$r_{t-}\mu_t = c_t = c_t \sim NID\left(0, \sigma_c^2\right),$$

where $NID\left(0, \sigma^2\right)$ denotes normally and independently distributed with mean zero and variance σ^2, and c_t and η_{t+1} are mutually independent, hence

$$E\left(c_t \eta_s\right) = 0; \ \forall\ t, s.$$

3.3. Interest rate models with a nonstationary stochastic mean

Following Chapman and Pearson (2000) and CKLS (1992) the parameters of the continuous-time stochastic process in (1) can be estimated using the following discrete-time econometric specification

$$r_{t+1} - r_t = \alpha + \beta r_t + \varepsilon_{t+1}, \tag{12}$$

$$E[\varepsilon_{t+1}] = 0, \qquad E[\varepsilon^2{}_{t+1}] = \sigma^2 r_t^{2\gamma}.$$

In this setting, Yu (2012) shows that if the regressor r_t in the right hand side of (12) has a permanent component (the unit root and the slow mean reversion situations), then the CKLS parameterization leads to inconsistent estimation for the mean reversion parameter.[7] In accordance, I generalize the Vasicek model by allowing for a nonstationary stochastic mean that is subject to permanent shocks, μ_t. I model the permanent component using two approaches that have strong assumptions and allow for considerable flexibility for the interest rate dynamics. In the first approach, I follow Ravn and Uhlig's (RU) (2002) decomposition procedure discussed in Section 3.2 to locate the nonstationary mean. I refer to my model parameterization as the HZ model. The econometric specification of the nonstationary stochastic mean model (HZ) reads:

(1) The Stochastic Mean Model (HZ): $r_{t+1} - r_t = \beta\left(r_t - \mu_t\right) + \varepsilon_{t+1}$ and $E[\varepsilon^2_{t+1}] = \sigma^2$.[8]

For comparison, I also follow Fama (2006) by estimating the permanent component as a moving average of the latest five years of the interest rate:

(2) The Moving Average Vasicek Model (MAVAS): $r_{t+1} - r_t = \beta\left(r_t - \mu^*_t\right) + \varepsilon^*_{t+1}$ and

$$E[\varepsilon^{*2}{}_{t+1}] = \sigma^2.$$

where μ^*_t is the moving average mentioned above. Note that the MAVAS model estimation may be inconsistent because the transitory component $c_t = r_t - \mu_t$ could be correlated with error term ε^*_{t+1}.

Since the GMM and Jackknife estimation of Phillips and Yu (2005) are feasible for these models, the estimation of nonstationary models can be based on these methods. The following specific steps are implemented in the procedure:

1. Decompose the interest rate into permanent and transitory components using the RU procedure and, in turn, the Fama (2006) approach described above.
2. Estimate the system parameters for fixed and nonstationary mean models by GMM, over the entire sample period. To estimate stationary mean models, I use standard GMM, and for nonstationary mean models, I use the estimates of μ_t and c_t, together with the corresponding diffusion process.
3. Also estimate the models' parameters by GMM for each sub-sample I consider.
4. Calculate the jackknife estimators for the system parameters using the Phillips and Yu (2005) procedure.
5. Use the parameter estimates from the above procedures to calculate bond and option prices for the HZ model.

4. Data and empirical results

4.1. Data

I use three-month secondary market rates that consist of end-of-month observations of the annualized yield on the U.S. Treasury bills from January 1934 through November 2013, collected from the Federal Reserve Economic Database (FRED) of the Federal Reserve Bank of St. Louis. This series embodies the lengthiest monthly set of observations on the U.S. T-bill rate that I recognized. The Treasury bill rates comprised in the daily H.15 release. The release is posted daily Monday through Friday at 4:15pm. The total number of observations is 959. Similar source of data is previously used in the literature by Anderson and Lund (1997) and Stanton (1997).

To compare my results with Aït-Sahalia (1999), I also consider the period of his sample from January 1963 to December 1998. Table 1 shows the means, standard deviations, and the first five lagged autocorrelations of the three-month yield and the three-month changes in the yield. The mean of the short-term riskless rate is 3.64% with a standard deviation of 3.16%. For the period considered in Aït-Sahalia (1999), the mean is 6.3% with a standard deviation of 2.51%.

Panels A and B of Table 2 report the results of the Augmented Dickey-Fuller (ADF) test and the Augmented Weighted Symmetric (WS) test in Pantula et al. (1994) for the entire sample period and for the sub-sample in Aït-Sahalia (1999).[9] As shown, I cannot

[7] Yu (2012) shows that the bias resulting from persistent interest rate is related to bad approximation of Cesaro sums and cannot be removed by jackknifing or bootstrapping.

[8] Consider the original Vasicek model $dr_t = \kappa(b - r_t)dt + \sigma dZ_t$, where b is the long term mean and k is the speed of adjustment. [This model can also be written in its Eq. (12) form as $dr_t = \alpha + \beta r_t + \sigma dZ_t$, with $\alpha = \beta$b and $k = -\beta$.] The discretized form reads $r_{t+1} - r_t = \kappa(b - r_t) + \varepsilon_{t+1}$, where $\varepsilon_{t+1} \sim \sigma N(0,1)$. This is equivalent to HZ with b replaced by μ_t and k *by* $-\beta$.

[9] The ADF and WS asymptotic p-values are computed using MacKinnon (1994) approximation (robust to size distortion). To save space, I only report the results of ADF and WS tests under the null of driftless unit root; the same results are obtained under the null of unit root with drift.

Table 1

Summary statistics. Panel A presents summary statistics for the three-month Treasury-bill rate, using the entire sample period (959 annualized monthly observations from January 1934 through November, 2013). Panel B presents the three-month Treasury-bill rate for the period studied in Aït-Sahalia (1999) (432 annualized monthly observations from January 1963 through December 1998). I report the mean, the standard deviation, and a set of monthly autocorrelations (ρ_j denotes the autocorrelation at lag *j*) for the interest rate in levels and first differences.

Variables	N	Mean	Standard deviation	ρ_1	ρ_2	ρ_3	ρ_4	ρ_5
Panel A: January 1934 through November 2013								
r_t	959	0.0366	0.0316	0.9932	0.9810	0.9713	−0.9611	0.9522
$r_{t+1} - r_t$	958	0.0000	0.0037	0.336	−0.072	0.0757	−0.0176	0.0368
Panel B: January 1963 through December 1998								
r_t	432	0.0630	0.0259	0.9794	0.9461	0.9160	0.8925	0.8691
$r_{t+1} - r_t$	431	0.0000	0.0052	0.3233	−0.1048	0.1193	−0.0431	0.0365

Table 2

Nonstationarity tests with a constant in the Fitted Regressions. The table presents the results of two nonstationarity tests applied to the three-month Treasury-bill rate examined in this study (959 annualized monthly observations from January 1934 through November 2013). I implement the Augmented Dickey-Fuller test (ADF) and the Augmented Weighted Symmetric test (WS) in Pantula et al. (1994). The ADF test statistic is the ratio $\hat{\alpha}/S_{\hat{\alpha}}$ from the estimated model $\delta r_{t+1} = a + \alpha r_t + \sum_{i=1}^{n} \phi_i \delta r_t + u_{t+1,}$ where $S_{\hat{\alpha}}$ is the standard error of the parameter estimate $\hat{\alpha}$. The optimal lag length n is chosen on the basis of the Akaike information criterion (AIC). The WS estimator of test statistic is given by $\hat{\rho}_{WS} = \frac{\sum_{t=2}^{n} r_t r_{t+1}}{\sum_{t=2}^{n-1} r_{t+1}^2 + n^{-1} \sum_{t=1}^{n} r_{t+1}^2}$,
where n is the length of the sequence r. I reject the null hypothesis of a unit root if the *p*-value is below a relevant significance level. ***P-values*** for the relevant statistics are computed using the approximation of MacKinnon (1994) on the basis of a regression surface. I report simulated *p*-values based on 1000 replications drawn from the normal distribution with zero mean and OLS squared residuals variances (the Wild bootstrap). Panel A reports test statistics for the entire sample period (959 annualized monthly observations from January 1934 through November 2013). Panel B reports test statistics for the period studied in Aït-Sahalia (1999) (432 annualized monthly observations from January 1963 through December 1998). Panel C reports test statistics for the "pre" period in Aït-Sahalia (1999) (348 annualized monthly observations from January 1934 through December 1962). Panel D reports test statistics for the "post-Aït-Sahalia (1999)" period (179 annualized monthly observations from January 1999 through November 2013).

	WS (*p*-value)	ADF (*p*-value)
Panel A: 1934 through 2013 period	−2.0337 (0.6215)	−1.8181 (0.6959)
Panel B: 1963 through 1998 period	−2.2833 (0.4416)	−2.3711 (0.3951)
Panel C: 1934 through 1963 period	−3.2869 (0.0379)	−4.0848 (0.0066)
Panel D: 1963 through 1999 period	−3.0903 (0.0662)	−3.1794 (0.0885)

reject the null of unit root across all tests.[10] The *p*-values of all of the tests are high, which provides evidence that the short-term interest rate is persistent.[11]

Since the series displays long-time intervals of stable behavior and low mean and volatility at either ends (1934–1963 and 1999–2013), I test for a unit root in those subperiods to evaluate the impact of low mean and volatility on the test results. As shown in Panels C and D, the null of unit root is rejected at a 5% significance level for the subperiod 1934 through 1963 and at a 10% significance level for the subperiod 1999 through 2013. I document stationary interest rates only when interest rates and volatilities are low.

A study by Conley et al. (1997) finds that the increase in volatility is a mechanism of inducing stationarity when interest rates are high. However, I document stationary interest rates only when interest rates and volatilities are low. A pattern seems to emerge: higher volatility periods do not necessary inject stationarity in the data. Consistent with Orphanides and Wilcox (2002), I infer that volatility induces stationarity only when the interest rate is high enough to force the Federal Reserve to intervene to curb inflation.

In order to gain some insight into the nature of the interest rate, Fig. 1 plots the short-term interest rate and the estimated nonstationary mean. As mentioned in the previous section, I approximate the nonstationary mean by choosing two smoothing constants: $1/q = 129{,}600$ (Panel A) and $1/q = 43{,}200$ (Panel B). These are equivalent to the Ravn and Uhlig's (2002) adjustments of the fourth and third power of the frequency of observations, respectively.[12] The fluctuation process of the stationary component of short-term interest rate is also illustrated in the Figure. As shown, the stationary component reverts systematically about its nonstationary mean except for the period spanning from September 1979 to July 1982, where substantial changes occur in the

[10] Perron (1989) concludes that the nominal interest rate is a unit root process; the same conclusion is also documented by Aït-Sahalia (1996b) and Bandi (2002). On the other hand, according to Bierens (1997), Al-Zoubi (2008, 2017) and Al-Zoubi et al. (2018), this cannot be realistic for two reasons: First, if the interest rate is a driftless random walk then it must allow for negative values, which is unrealistic; second, if it is a random walk with positive drift then it would converge to infinity, which is improbable.

[11] See Perron (1989) for further discussion.

[12] For models estimation, I follow Mehra (2004) in calculating the nonstationary mean at time t. In particular, I run the standard RU mean serially through time using only the data available at time t for calculating the mean value for that point in time.

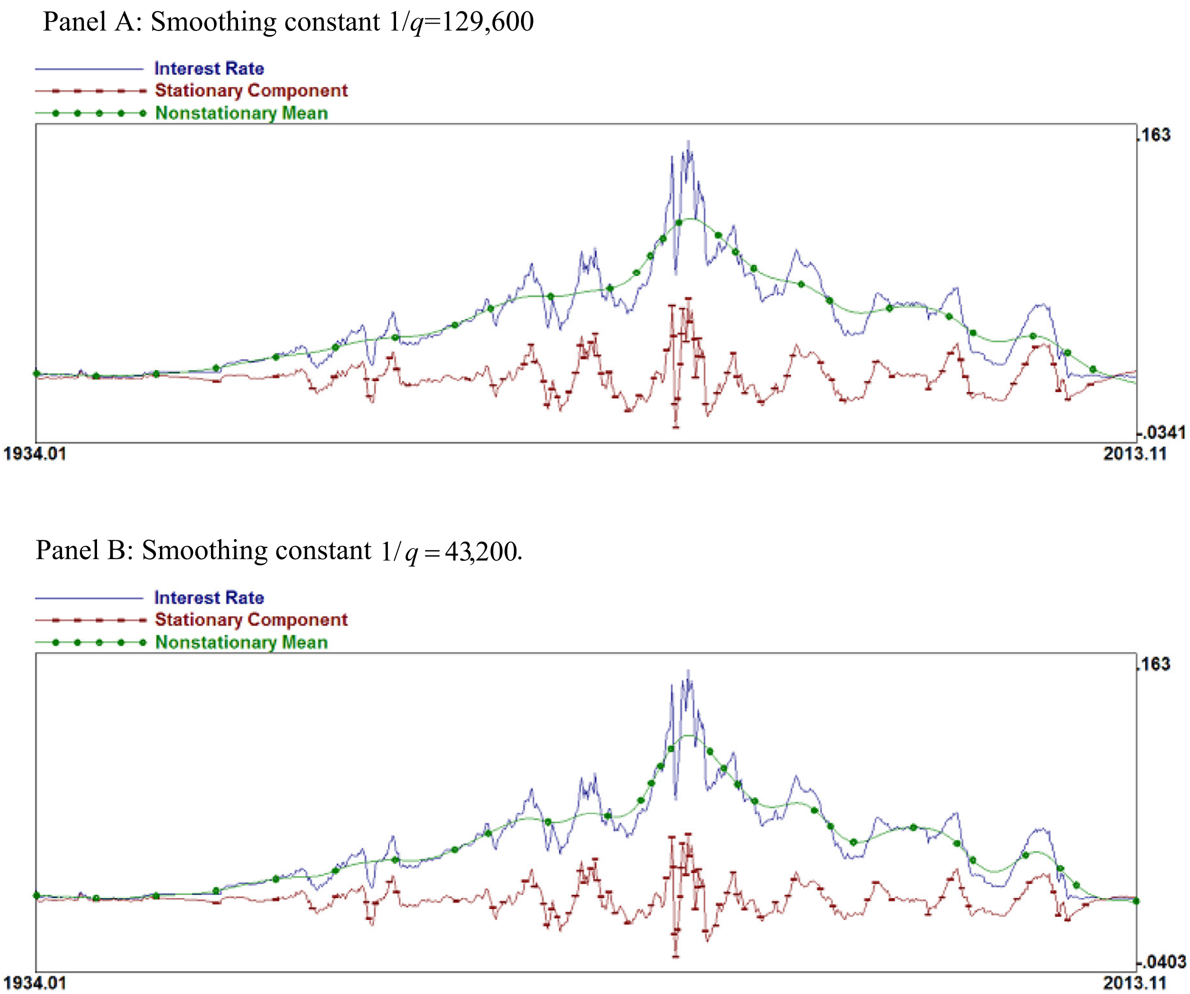

Fig. 1. Interest Rate from January 1934 through November 2013. The figure plots three series. The first is the Federal Reserve Bank of St. Louis three-month secondary market Treasury bill rate. The second series is the approximated nonstationary mean of the interest rate. The third series is the stationary component of the interest rate. In Panels A and B, I use smoothing constants 1/q of 129,600 and 43,200, respectively. These are equivalent to the Ravn and Uhlig's (2002) adjustment of the fourth and third power of the frequency of observations, respectively. The sample is from January 1934 through November 2013 (959 annualized monthly observations).

variability of the interest rate around the trend. This may suggest a structural break during that period, in that the Federal Reserve emphasized targets for monetary aggregates to curb inflation after September 1979.

4.2. Empirical results

In this section, I first estimate and compare the five affine models with the nonstationary mean models of the short-term interest rate process. To reiterate, I model the permanent component using two approaches that make strong assumptions and allow for considerable flexibility for the interest rate dynamics. In the first approach, I use Ravn and Uhlig's (2002) method described in Section 3.2 to locate the nonstationary mean. For comparison purposes I also follow Fama (2006) by estimating the permanent component as a moving average of the latest five years of the riskless rate.

Before addressing estimation bias, I employ the naïve GMM technique. This allows us to obtain preliminary estimates and demonstrate the role of specification error. I compare the performance of the five models with each other and with the modified models. Later, I correct for bias using the jackknife technique developed by Phillips and Yu (2005).

I follow Chapman and Pearson (2000) in defining the orthogonality conditions for the CP model, Ahn and Gao (1999) for the AG model, and Chan et al. (1992) for the Vasicek, CIR, and CKLS models. Since GMM is now common in the literature, I report the relevant moment conditions in Appendix. The general approach is well explained in detail in Chan et al. (1992).

Table 3

GMM estimate of five single-factor models and three modified models for short-term interest rate dynamics for the period from January 1963 through December 1998. Panel A presents GMM estimate of five single-factor interest rate models, namely Chapman and Pearson (2000), Ahn and Gao (1999), Chan et al. (1992), Cox et al. (1985), and Vasicek (1977) for the period studied in Aït-Sahalia (1999) (432 annualized monthly observations from January 1963 through December 1998). The models I consider can be nested within the Chapman and Pearson (2000) model: $dr_t = (\alpha + \beta r_t + \theta_1 r_t^2 + \theta_2 r_t^{-1})dt + \sigma r_t^{\gamma} dZ_t$, where Z_t is the standard Brownian motion. I examine the following discrete-time econometric specification

$$r_{t+1} - r_t = \alpha + \beta r_t + \theta_1 r_t^2 + \theta_2 r_t^{-1} + \varepsilon_{t+1}$$
$$E[\varepsilon_{t+1}] = 0, \qquad E[\varepsilon_{t+1}^2] = \sigma^2 r_t^{2\gamma}.$$

I follow Chapman and Pearson (2000) to estimate the models by using GMM. Define λ as the entire parameter vector of $\alpha, \beta, \theta_1, \theta_2, \sigma,$ *and* γ. The relevant orthogonality conditions are as follows:

$$h\left(r_{t+1}, \lambda\right) = \left[\varepsilon_{t+1}, \varepsilon_{t+1} r_t, \varepsilon_{t+1} r_t^2, \varepsilon_{t+1} r_t^{-1} - \sigma^2 r_t^{2\gamma}, \left(\varepsilon_{t+1}^2 - \sigma^2 r_t^{2\gamma}\right) r_t\right], \text{ where } \varepsilon_{t+1} = [r_{t+1} - r_t - \alpha - \beta r_t - \theta_1 r_t^2 - \theta_2 r_t^{-1}].$$

Panel B presents GMM estimate of the modified mean reverting interest rates models, HZ and MAVAS models for the period studied in Aït-Sahalia (1999) (432 annualized monthly observations from January 1963 through December 1998). I follow Ravn and Uhlig (RU) (2002) and de Jong and Sakarya (2016) in approximating the nonstationary mean, μ_t. I choose a smoothing constant $1/q = 129,600$. This is equivalent to the Ravn and Uhlig's (2002) adjustment of the fourth power of observations frequency. For comparison, I also follow Fama (2006) and approximate μ_t as a moving average of the latest five years of interest rate data. I impose various restrictions on the parameters (see Appendix C) to obtain standard parametric models for the interest rate dynamics. GMM estimate of the coefficients, overidentifying tests of Hansen (1982), their significance levels, and goodness of fit tests (RAMSE, MAE, ME, and MAPE). The corresponding *p*-values are in parentheses. I follow Inoue and Shintani (2006) by using the Parzen kernel of Gallant (1987) with two lags to compute the moment-weighting matrix. The covariance matrix is robust to heteroskedasticity and autocorrelation.

Model	α (*p*-value)	β (*p*-value)	θ_1 (*p*-value)	θ_2 (*p*-value)	σ^2 (*p*-value)	γ (*p*-value)	Overidentifying Test (*p*-value)	RMSE	MAE	ME	MAPE
Panel A: Single-factor models											
Chapman &Pearson	−0.03366 (0.0442)	0.52762 (0.0264)	−2.48081 (0.0091)	0.00065 (0.0549)	0.14583 (0.2797)	1.66439 (0.0000)	EXACT	0.00515	0.00298	0.00007	4.49
AG	−0.00328 (0.1384)	0.09767 (0.1044)	−0.58304 (0.1069)	0	0.04503 (0.0000)	1.5	3.65610 (0.16073)	0.00522	0.00300	−0.00013	4.791
CKLS	0.00225 (0.1653)	−0.03331 (0.2352)	0	0	0.14758 (0.2789)	1.71750 (0.0000)	EXACT	0.00523	0.00302	−0.00012	4.496
CIR	0.00112 (0.4681)	−0.01492 (0.5771)			0.00019 (0.0000)	0.5	6.61641 (0.0101)	0.00523	0.00296	−0.00015	4.333
VAS	0.00010 (0.4938)	−0.01383 (0.6029)	0	0	0.00001 (0.0000)	0	7.47546 (0.0063)	0.00529	0.00307	0.00080	4.475
Panel B: Modified models											
HZ	0	−0.04747 (0.0285)	0	0	0.00001 (0.0299)	0	6.09671 (0.0474)	0.00517	0.00290	0.00003	4.251
MAVAS	0	−0.02879 (0.0068)	0	0	0.00001 (0.0049)	0	7.45312 (0.0240)	0.00525	0.00298	0.00007	4.368

To evaluate the forecast performance of each model I define the following goodness-of-fit measures based on the Bergstrom (1986, 1989) generalized discrete form:

$$RMSE = \sqrt{\frac{1}{K}\sum_{t=T+1}^{T+K}(\hat{r} - r)^2}$$

$$MAE = \frac{1}{K}\sum_{t=T+1}^{T+K}|\hat{r} - r|$$

$$ME = \frac{1}{K}\sum_{t=T+1}^{T+K}(\hat{r} - r)$$

$$MAPE = \frac{1}{K}\sum_{t=T+1}^{T+K}\left|\frac{\hat{r} - r}{r}\right|$$

where $\hat{r}$ is the forecast value of r and K is the forecast period. *RMSE* measures root-mean-square error; *MAE* and *ME* measure mean-absolute and mean error, respectively; and *MAPE* measures mean-absolute-percentage error.

4.2.1. *Aït-Sahalia (1999) Period: 1963–1998*

Panel A of Table 3 reports the parameter estimates, the asymptotic *p*-values of the individual parameters, the GMM overidentifying (χ^2) test, and the goodness of fit tests, for the unmodified models over the period spanning the years 1963 to 1998 (the period considered by Aït-Sahalia, 1999). As shown, for CIR and Vasicek models, the *p*-values of the GMM criterion test are below 0.05, which suggests that these models are misspecified. My results are consistent with Chapman and Pearson (2000) and Bandi (2002), who argue that none of the models are correctly specified. The AG model has the lowest goodness of fit statistic with a *p*-value well above 0.05 (0.16), but none of the parameters of the drift term are significant.

Table 4
GMM estimate of five single-factor models and three modified models for short-term interest rate dynamics for the period from January 1934 through November 2013. Panel A presents GMM estimate of five single-factor interest rate models, namely Chapman and Pearson (2000), Ahn and Gao (1999), Chan et al. (1992), Cox et al. (1985), and Vasicek (1977) for the entire sample period (959 annualized monthly observations from January 1973 through November 2013). The models I consider can be nested within the Chapman and Pearson (2000) model: $dr_t = (\alpha + \beta r_t + \theta_1 r_t^2 + \theta_2 r_t^{-1})dt + \sigma r_t^{\gamma} dZ_t$, where Z_t is the standard Brownian motion. I examine the following discrete-time econometric specification

$$r_{t+1} - r_t = \alpha + \beta r_t + \theta_1 r_t^2 + \theta_2 r_t^{-1} + \varepsilon_{t+1}$$
$$E[\varepsilon_{t+1}] = 0, \qquad E[\varepsilon_{t+1}^2] = \sigma^2 r_t^{2\gamma}.$$

I follow Chapman and Pearson (2000) to estimate the models by using GMM. Define λ as the entire parameter vector of $\alpha, \beta, \theta_1, \theta_2, \sigma,$ *and* γ. The relevant orthogonality conditions are as follows:

$$h\left(r_{t+1}, \lambda\right) = \left[\varepsilon_{t+1}, \varepsilon_{t+1} r_t, \varepsilon_{t+1} r_t^2, \varepsilon_{t+1} r_t^{-1} - \sigma^2 r_t^{2\gamma}, \left(\varepsilon_{t+1}^2 - \sigma^2 r_t^{2\gamma}\right) r_t\right], \text{ where } \varepsilon_{t+1} = [r_{t+1} - r_t - \alpha - \beta r_t - \theta_1 r_t^2 - \theta_2 r_t^{-1}].$$

Panel B presents GMM estimate of the modified mean reverting interest rates models, HZ and MAVAS models for the entire sample period (959 annualized monthly observations from January 1973 through November 2013). To evaluate the effect of the nonstationary mean on the diffusion, I also modify the CKLS model by incorporating nonstationary means in the drift. I follow Ravn and Uhlig (RU) (2002) and de Jong and Sakarya (2016) in approximating the nonstationary mean, μ_t. I choose a smoothing constant $1/q = 129{,}600$. This is equivalent to the Ravn and Uhlig's (2002) adjustment of the fourth power of observations frequency. For comparison, I also follow Fama (2006) and approximate μ_t as a moving average of the latest five years of interest rate data. I impose various restrictions on the parameters (see Appendix C) to obtain standard parametric models for the interest rate dynamics. GMM estimate of the coefficients, overidentifying tests of Hansen (1982), their significance levels, and goodness of fit tests (RAMSE, MAE, ME, and MAPE). The corresponding *p*-values are in parentheses. I follow Inoue and Shintani (2006) by using the Parzen kernel of Gallant (1987) with two lags to compute the moment-weighting matrix. The covariance matrix is robust to heteroskedasticity and autocorrelation..

Model	α (*p*-value)	β (*p*-value)	θ_1 (*p*-value)	θ_2 (*p*-value)	σ^2 (*p*-value)	γ (*p*-value)	Overidentifying Test (*p*-value)	RMSE	MAE	ME	MAPE
Panel A: Single-factor models											
Chapman & Pearson	−0.00209	0.0839	−0.6735	0.00000	0.50681	1.6167	EXACT	0.00403	0.00228	−0.00005	128.173
	(0.0174)	(0.0136)	(0.0268)	(0.0219)	(0.2251)	(0.0000)					
AG	−0.00142	0.0552	−0.3917	0	0.05367	1.5	2.2211	0.00378	0.00214	0.00031	42.709
	(0.0669)	(0.0621)	(0.1187)		(0.0000)		(0.3294)				
CKLS	0.001256	−0.02262	0	0	0.45265	1.6612	EXACT	0.00379	0.00212	−0.00044	42.473
	(0.2273)	(0.2986)			(0.2370)	(0.0000)					
CIR	0.000333	−0.0044			0.000189	0.5	6.5263	0.00374	0.00187	−0.00018	16.687
	(0.7307)	(0.8254)			(0.0000)		0.01061				
VAS	0.000613	−0.0102	0	0	0.000009	0	8.235	0.00374	0.00193	−0.00025	24.186
	(0.045)	(0.5488)			(0.6187)		0.0041				
Panel B: Modified models											
HZ	0	−0.0385	0	0	0.00001	0	5.9988	0.00369	0.00182	−0.00000	10.598
		(0.0158)			(0.0345)		(0.0498)				
MAVAS	0	−0.0296	0	0	0.00009	0	7.9748	0.00386	0.00198	0.00000	13.307
		(0.0005)			(0.0051)		(0.0185)				

The result emerging from Table 3, Panel A is that CP model provides a very good estimate of the drift term. Unlike for the AG and CKLS models, the estimated parameters of the drift are significantly different from zero. Consistent with the literature, CEV models like CP model outperform the rest of the models in terms of forecast power. On the basis of *RMSE* and *ME* measures, CP model performs the best. The other parametric models, namely CIR and Vasicek, perform relatively less well. Based on *RMSE*, the AG model outperforms the CKLS, CIR, and Vasicek models. The CKLS model, on the other hand, outperforms the AG, CIR, and Vasicek models on the basis of *ME*. Across all four measures, the Vasicek model performs the least well; however, no robust conclusion can be gleaned about the linear mean-reverting behavior in this case because the GMM overidentifying test fails to accept the null hypothesis for both CIR and Vasicek models, which results in an inconsistent estimate of the parameters (see, Aït-Sahalia, 1996b, for further discussion). Ignoring the inconsistency in the entire parameter vector, one could, however, reject the linearity of the drift term.[13]

Panel B of Table 3 reports parameter estimates for the nonstationary mean models. The results support my prediction, in striking fashion. As shown, the overidentifying tests suggest that the modified Vasicek model is correctly specified using the RU nonstationary mean. The model has a *p*-value in excess of 0.05 and cannot be rejected at a 95% confidence level. The goodness of fit measures also support my prediction as the modified Vasicek model with the RU nonstationary mean setting performs the best on the basis of the *RMSE, MAE*, *ME*, and *MAPE* measures.

An important feature of this ranking is that it can be classified by the persistence of the data generating process, that is, when the interest rate has a permanent component, CEV models perform well and CIR and Vasicek models perform poorly. Correspondingly, when Vasicek model is modified to account for permanent shocksCP, CKLS and AG models no longer outperform mean-reverting models.

My estimates of the parameter of mean reversion β are as expected. For the modified Vasicek model, the estimates of β are negative and significantly different from zero at 5% level, which implies reversion about a nonlinear, time-varying, mean. This in

[13] Inconsistent with CKLS, Bandi (2002) finds that the parameters in the drift term are significant and concludes that the drift term has a linear structure.

turn suggests that the model misspecification induced by not including enough time-varying components in the drift term may lead to substantially biased estimates of the remaining parameters in the model.

4.2.2. Full sample period: 1934-2013

Panel A of Table 4 summarizes estimation results for the entire sample period. The estimates are not particularly supportive of CEV models in the sense that the goodness-of-fit measures sharply increase relative to Panel A of Table 3. Moreover, the CIR model performs the best relative to the CP, CKLS, and AG models. These results are in line with Cai and Swanson (2011) for the period of stable interest rates. Consistent with Bali and Wu (2006) and Cai and Swanson (2011), I conclude that model performance is highly sample dependent. Stationary mean models like the CIR and Vasicek models outperform CEV models. Of further note is that the CP model performs modestly, and the Vasicek model outperforms the CP, CKLS, and AG models.

In Panel B of Table 4, I also estimate the modified version of my model. Again, estimation results support the nonstationary mean model, in the sense that the parameter estimates are correctly identified and forecasting power substantially increased for both HZ and MAVAS models. As shown, for HZ model, the *p*-values for the goodness-of fit-statistic are well above 5% and the models cannot be rejected. However, for MAVAS, the model is still rejected at a 5% significance level.

Based on forecasting power measures compared with models in Panel A of Table 4, the modified model with the RU mean performs the best across all measures. The modified models with MA mean also exhibit good performance. Furthermore, for each nonstationary mean model, the mean reversion parameter β becomes negative and significant with an increase in the speed of mean reversion. Therefore, the finding of mean reversion about the nonlinear mean is robust.

4.3. Out-of-sample forecast

To check models performance out-of-sample, I run the Diebold and Mariano (1995) forecast accuracy test for the HZ model against the five affine models. Let $\hat{r}_{1t}$ and $\hat{r}_{2t}$ be the estimate of the interest rate r_t in two competing models. The Diebold–Mariano test statistic (DM) is the ratio $\bar{d}/\sqrt{AVAR}$ where $\bar{d}$ is the sample mean of the loss differential for the two competing models $\left(\bar{d} = K^{-1}\sum_{t=T+1}^{T+K}\left(\left(\hat{r}_{1t}-r_t\right)^2-\left(\hat{r}_{2t}-r_t\right)^2\right)\right)$ and $AVAR$ is the estimate of the asymptotic variance of the loss differential described in Diebold and Mariano (1995):

$$AVAR\left(\bar{d}\right) = \frac{2\pi \hat{f}_d(0)}{K},$$

where K is the forecast horizon and $\hat{f}_d(0)$ is a consistent estimate of the spectral density function of d at frequency zero defined by:

$$\hat{f}_d(0) = \frac{1}{2\pi}\sum_{K=-(T-1)}^{T-1} I\left(\frac{h}{K-1}\right)\hat{\gamma}_d(h),$$

where

$$\hat{\gamma}_d(h) = \frac{1}{T}\sum_{t+|h|+1}^{T}\left(d_t - d\right)\left(d_{t-|-h|} - \bar{d}\right),$$

and

$$I\left(\frac{h}{K-1}\right) = \begin{cases} 1 & for\left|\frac{h}{K-1}\right| \le 1 \\ 0 & otherwise \end{cases};$$

thus

$$\hat{f}_d(0) = \frac{1}{2\pi}\left(\hat{\gamma}_d(0) + 2\sum_{h=1}^{K-1}\hat{\gamma}_d(0)\right).$$

To overcome the size distortion of the test in a finite sample, I follow Harvey et al. (1997) and correct the bias of the DM test by comparing it to a Student-t distribution with *(T-1)* degrees of freedom. The corrected statistic takes the form

$$DM^* = \sqrt{\frac{T+1-2h+h(h-1)}{T}}DM.$$

For each test, I use three forecast horizons: 1, 6, and 24 months and use data over the period January 1934–November 2003 for estimation and reserve data over the period December 2003–November 2013 for out-of-sample forecasting. The model is estimated using the first data points and the one-period-ahead forecast is generated, for a total of 120 observations. The null hypothesis is that the models have the same forecast accuracy against the alternative hypothesis that the modified model is more accurate than the competing model. The results are reported in Table 5. They are broadly consistent with the previous findings. Specifically, for one- and six-month forecast horizons the null that model forecasts are equally accurate is mostly rejected at the 5% significance level in favor of HZ. For the 24-month forecast horizon, the AG and CKLS models are rejected at the 1% significance level and the CP and Vasicek models are rejected at the 10% significance level in favor of HZ.

Table 5
Diebold–Mariano forecast accuracy test of the modified model versus five single-factor models for short-term interest rate dynamics for the period from January 1963 through November 2013. The table presents the results of the Diebold and Mariano (1995) forecast accuracy test applied to forecast errors of five single-factor models against HZ model for the short-term interest rate. The sample period is from January 1934 through November 2013. I use data over the period January 1934–November 2003 for estimation and reserve data over the period December 2003–November 2013 for out-of-sample forecasting. The model is estimated using the first data points and the one-period-ahead forecast is generated, for a total of 120 observations. The Diebold–Mariano test statistic is the ratio $\overline{d}/\sqrt{AVAR}$ where $\overline{d}$ is the sample mean of the loss differential for the two competing models $\left(\overline{d} = K^{-1}\sum_{t=T+1}^{T+K}\left(\left(\hat{r}_{1t} - r_t\right)^2 - \left(\hat{r}_{2t} - r_t\right)^2\right)\right)$ and $AVAR$ is the estimate of the asymptotic variance of the loss differential described in Diebold and Mariano (1995): $AVAR\left(\overline{d}\right) = \frac{2\pi \hat{f}_d(0)}{K}$, where $\hat{f}_d(0)$ is a consistent estimate of the spectral density function of d at frequency zero and K is the forecast horizon. To overcome the size distortion of the test in a finite sample, I follow Harvey et al. (1997) and correct the bias of the DM test by comparing it to a Student-t distribution with $(T-1)$ degrees of freedom For each test, I use three forecast horizons (1, 6, and 24 months). The null hypothesis is that the two models have the same forecast accuracy against the alternative hypothesis that the modified model is more accurate than the competing model.

	1-month forecast horizon	6-month forecast horizon	24-month forecast horizon
Model 2	HZ	HZ	HZ
Model 1			
CP	3.4859	1.9969	1.7329
	(0.0005)	(0.0461)	(0.08385)
AG	5.4620	4.1521	4.7069
	(0.0000)	(0.0000)	(0.0000)
CKLS	4.2272	3.3218	3.3133
	(0.0000)	(0.0009)	(0.0009)
CIR	2.0549	1.9218	1.5542
	(0.0402)	(0.05492)	(0.1205)
Vasicek	3.3859	2.4596	1.8821
	(0.0007)	(0.0141)	(0.06012)

4.4. The jackknife estimate of the drift and the diffusion function

In this Section 1 implement the method of Phillips and Yu (2005) for correcting the bias of consistent estimators for the short-rate parametric models. They find that their jackknife procedure provides a very considerable bias reduction over the Maximum likelihood Estimator (MLE).[14] The performance of the jackknife is also examined in Phillips and Yu (2009) and compared with approximation based methods of ML, like the Euler method studied in Merton (1980) and Lo (1988), the Milstein scheme proposed in Elerian (1998), the Nowman (1997) expansion, and the Hermite expansion proposed in Aït-Sahalia (1999) and developed in Aït-Sahalia (2002). Phillips and Yu (2009) conclude that the jackknife method significantly outperforms all existing approximations in reducing bias. Tang and Chen (2009) compare the jackknife with a parametric bootstrap, and the estimation proposed in Gourieroux et al. (1993); they find that the jackknife delivers satisfactory results in bias reduction even though it inflates the variance. As noted by Phillips and Yu (2005, 2009)) and Bao et al. (2014), a carefully designed jackknife procedure substantially reduces bias, leading to a decrease in the *RMSE*.

To motivate the jackknife method of Phillips and Yu (2005), let N be the number of observations in the entire sample and divide the sample into m successive blocks each with l equal observations, that is $N = l \times m$. Let $\hat{\theta}_N$ be a consistent estimator of a certain parameter θ utilizing the entire sample. Furthermore, let $\hat{\theta}_{li}$ be a consistent estimator of θ utilizing the i'th block. The jackknife bias-corrected estimator is

$$\hat{\theta}_{Jack} = \frac{m}{m-1}\theta_N - \frac{\sum_i^m \hat{\theta}_{li}}{m^2 - m}.$$

Phillips and Yu (2005) prove that under quite broad conditions which insure that the bias of the estimates $\hat{\theta}$ is correctly asymptotically sized with power of $\left(N^{-1}\right)$, the bias in jackknife estimate $\hat{\theta}_{Jack}$ is of order $O\left(N^{-2}\right)$ rather than $O\left(N^{-1}\right)$. Of note, the jackknife is practically simple and is relevant to a wide range of models, including models with unknown asymptotic bias expansions such as CEV models.

Table 6 reports the jackknife estimates and goodness of fit tests for the diffusion models for the entire sample period. Recall that my GMM estimations in Tables 3 and 4 experienced some identification problems in estimating the MAVAS model. The p-value of the GMM overidentifying $\left(\chi^2\right)$ test was well below 5%, which led to inconsistent parameter estimation. However, when I used the RU mean, the p-values of the $\left(\chi^2\right)$ statistic increased above 5%, indicating consistent GMM estimation. Therefore, below, I use the RU mean specification as my general nonlinear drift specification. I let HZ1 and HZ2 denote the HZ model with the Ravn and Uhlig's (2002) adjustments using the fourth and third power of the frequency of observations, respectively.

As shown, the jackknife provides improvements across all models in terms of *RMSE*, *MAE*, and *ME* measures. Moreover, the CKLS model performs the best amongst the CP, AG, CKLS, CIR, and Vasicek models. This is followed by CP and AG models, respectively.

[14] Duffee and Stanton (2012) find that the Efficient Method of Moments (EMM) exhibits high finite sample bias even in the simplest term structure models. Bauer et al. (2012) also use simulation based bias correction method for finite-sample bias of the affine dynamic term structure model that helps in measuring policy. implications or the nominal term premia.

Table 6

Jackknife estimate of five single-factor models and two modified models for short-term interest rate dynamics for the period from January 1934 through November 2013. Panel A presents jackknife estimates of five single-factor interest rate models, namely Chapman and Pearson (2000), Ahn and Gao (1999), Chan et al. (1992), Cox et al. (1985), and Vasicek (1977), for the entire sample period (959 annualized monthly observations from January 1934 through November 2013). The models I consider can be nested within the Chapman and Pearson (2000) model: $dr_t = (\alpha + \beta r_t + \theta_1 r_t^2 + \theta_2 r_t^{-1})dt + \sigma r_t^{\gamma} dZ_t$, where Z_t is the standard Brownian motion. I examine the following discrete-time econometric specification

$r_{t+1} - r_t = \alpha + \beta r_t + \theta_1 r_t^2 + \theta_2 r_t^{-1} + \varepsilon_{t+1}$

$E[\varepsilon_{t+1}] = 0, \qquad E[\varepsilon_{t+1}^2] = \sigma^2 r_t^{2\gamma}.$

I follow Phillips and Yu (2005) to estimate the models. Let N be the number of observations in the entire sample and divide the sample into $m = 2$ successive blocks each with l equal observations, that is $N = l \times m$. Let $\hat{\theta}_N$ be a consistent estimator of a certain parameter θ utilizing the entire sample. Furthermore, let $\hat{\theta}_{li}$ be a consistent estimator of θ utilizing the i'th block. The jackknife bias-corrected estimator is

$\hat{\theta}_{Jack} = \frac{m}{m-1}\theta_N - \frac{\sum_i^m \hat{\theta}_{li}}{m^2-m}.$

Panel B presents jackknife estimate of the modified mean reverting interest rates models, HZ model for the entire sample period (959 annualized monthly observations from January 1973 through November 2013). I consider a stochastic differential equation of the form: $dr_t = \beta\left(r_t - \mu_t\right)dt + \sigma dZ_t$ for the HZ model. I follow Ravn and Uhlig (RU) (2002) and de Jong and Sakarya (2016) in approximating the nonstationary mean, μ_t. In particular, I choose two smoothing constants $1/q = 129,600$ *and* $43,200$. These are equivalent to the Ravn and Uhlig's (2002) adjustments of the fourth and third power of the frequency of observations, respectively. I impose various restrictions on the parameters (see appendix C) to obtain standard parametric models for the interest rate dynamics. HZ1 and HZ2 stand for the HZ model with the Ravn and Uhlig's (2002) adjustments to the fourth and third power of the frequency of observations, respectively. I provide jackknife estimates of the coefficients and goodness of fit tests (RAMSE, MAE, and ME).

Panel A: Single-factor models									
Model	α	β	θ_1	θ_2	σ^2	γ	RAMSE	MAE	ME
CP	−0.00259	0.10093	−0.89064	0.000001	0.04095	1.27214	0.00388	0.00232	0.00066
AG	−0.00086	0.03294	−0.12421	0	0.05347	1.5	0.00381	0.00202	−0.00006
CKLS	0.00101	−0.01566	0	0	1.10866	1.31889	0.00377	0.00205	−0.00044
CIR	0.00033	−0.00535			0.00019	0.5	0.00373	0.00187	−0.00014
VAS	0.00074	−0.01470	0	0	0.00001	0	0.00374	0.00197	−0.00021
Panel B: Modified models									
HZ1	0	−0.03305	0	0	0.00007	0	0.003692	0.00182	−0.0000075
HZ2	0	−0.05869	0	0	0.00001	0	0.003667	0.00181	−0.0000076

Again, the HZ model outperforms all competing models. On the basis of *RMSE*, *MAE* and *ME*, the HZ2 model performs the best, followed by the HZ1 model. Consistent with the results from Table 4, CEV models perform modestly, and underperform relative to the CIR and Vasicek models.

The parameter estimates for the drift functions in Table 6 provide supporting evidence for my earlier findings from the stochastic mean models. For the entire sample period, it is apparent that the estimates of parameter β are negative across all stochastic mean models. This is consistent with the intuition that the interest rate should converge more quickly to a central tendency that is nonstationary than a constant mean.

5. Estimating bond and option prices using spot rates with nonstationary means

In earlier literature, the problem of biased estimation turns out to be significant in pricing bonds and options. For example, Phillips and Yu (2005) document that estimation bias leads to a 1% downward bias in the bond price and a 24.4% downward bias in the option price. Tang and Chen (2009) also document large underestimation in option values under both the Vasicek and the CIR models. While both jackknife and bootstrap methods have been shown to be efficient in reducing the bias for bond and option prices even if the model is misspecified, the bias in the estimation is still large. For instant, Phillips and Yu (2005) show that the jackknife method is able to reduce the bias in option price from −21.04% to −5.89%, −40.02% to −23.02%, and −57.65% to −47.72% for in-the money, at-the-money, and out-of-the-money options, respectively. Unfortunately, misspecification error still accounts for the majority of bias and becomes more severe when fewer observations are available or for options with relatively longer maturities.

To address the problem of biased estimation using nonstationary mean models for spot rates, I use the analytic solutions for option prices for the Vasicek, HZ1, and HZ2 models. I denote $C_{t,T,S,K}(\theta)$ to be the theoretical price of a European call option at time t with strike price K and maturity S on the bond with price $P_{t,T}(\theta)$ maturing at $T > S$. Section 2.2 provides the analytic formula for the bond price in the case of HZ model, as functions of the estimated parameters. For the Vasicek model, I use the analytic formula given in Vasicek (1977). I first compute the implied bond prices by plugging in the parameter estimates of β and σ^2 from Vasicek, HZ1, and HZ2 models with $t = 0$ and $T = 3$. I evaluate the performance of the models using *RMSE* and compare implied bond prices with observed (true) bond prices. The true zero-coupon bond price is given by

$$P\left(t, T, r_t\right) = Exp\left(-r_t T\right).$$

The implied option prices for the three models are calculated using the analytic formula provided in Jamshidian (1989) for a European call option with a face value of a bond = \$100, K = \$90 and $S = 1$. I used bond prices generated from both the Vasicek and HZ models to calculate option prices. The call option price that matures at time S on a zero-coupon bond maturing at time T is

$$LP(0,T)\,N(h) - KP(0,S)\,N\left(h - \sigma_P\right)$$

Table 7
Zero-coupon bond prices and European call option prices under the nonstationary mean interest rate model. The table presents bond and option prices for the nonstationary mean (HZ) and Vasicek (1977) models. $\hat{P}$ and $\hat{C}$ are the estimated zero-coupon bond and European call option prices respectively, P is the observed-zero coupon bond price, Avg, Std Dev, and RMSE are the estimated average, standard deviation, and root-mean-square error respectively. I use GMM and jackknife parameter estimates for the entire sample to estimate $\hat{P}$ and $\hat{C}$. To test model performance, I compute $\hat{P}$ and $\hat{C}$ on a monthly basis using the observed three-month Treasury-bill rate for the entire sample period (959 annualized monthly observations from January 1934 through November 2013). HZ1 and HZ2 stand for the HZ model with the Ravn and Uhlig's (2002) adjustments to the fourth and third power of the frequency of observations, respectively. For the HZ model, the bond price is given by

$$\hat{P} = Exp\left(A(t,T)r_t - \mu_t(A(t,T)+1) + B(t,T) + D(t,T)\right),$$

where

$$A(t,T) = \left(\frac{1-e^{\beta(T-t)}}{\beta}\right),\ B(t,T) = \frac{\sigma^2}{2\beta^2(1+q)}(A(t,T)+(T-t)) + \frac{\sigma^2 A(t,T)^2}{4\beta(1+q)},\ \text{and}\ D(t,T) = \left(\frac{q\sigma^2\beta^3(T-t)^4}{48(1+q)}\right).$$

The column "*True P*" is the average of the true zero-coupon bond price given by

$$P(t,T,r_t) = Exp(-r_t T).$$

The column "$\hat{C}$" reports the average of the numerical call option prices calculated using Jamshidian (1989) formula for bond prices implied from the HZ and Vasicek models.

	α	β	σ^2	$\hat{P}$	$\hat{C}$	*True P*
Under the Vasicek process						
GMM estimates	0.00061	−0.01020	0.000009			
Avg				0.89877	4.02295	0.89983
St Dev				0.08055	3.78326	0.08186
RMSE				0.00863		
Jackknife estimates	0.00074	−0.014703	0.00001			
Avg				0.89887	4.11465	
St Dev				0.08005	3.75971	
RMSE				0.00868		
Under the HZ1 Process						
GMM estimates	0	−0.03851	0.00001			
Avg				0.89925	4.16609	
St Dev				0.08130	3.85562	
RMSE				0.00157		
Jackknife estimates	0	−0.03305	0.00007			
Avg				0.89927	4.16832	
St Dev				0.08137	3.86371	
RMSE				0.00138		
Under the HZ2 Process						
GMM estimates	0	−0.06014	0.00001			
Avg				0.89945	4.17549	
St Dev				0.08122	3.86158	
RMSE				0.00192		
Jackknife estimates	0	−0.05869	0.00001			
Avg				0.89940	4.17192	
St Dev				0.08123	3.86000	
RMSE				0.00189		

where L is the principal of the zero coupon bond, K is the strike price, $P(0,T)$ is the bond price derived in Proposition 2.2, and

$$h = \frac{1}{\sigma_p} ln \frac{LP(0,T)}{KP(0,S)} + \frac{\sigma_P}{2}$$

$$\sigma_P = \frac{\sigma}{\beta}\left[e^{\beta(T-s)} - 1\right]\sqrt{\frac{e^{2\beta S}-1}{2\beta}}.$$

The following steps are taken to implement the technique.

1. Using the parameter estimates in Tables 4 and 6, I first calculate bond prices of the Vasicek and HZ models for the entire sample period. I use the analytic formula derived in Proposition 2.2 to calculate the HZ bond prices.
2. Compute the *RMSE* of the Vasicek and HZ models by comparing both models' prices with the true bond prices using the entire sample period.
3. Using the Vasicek and HZ bond prices calculated in step 1, calculate the call option prices for each model using the analytic formula provided in Jamshidian (1989).

I Compare the implied option prices of Vasicek and HZ models with the results of Phillips and Yu (2005) and Tang and Chen (2009). As a robust check, I also compare the option prices implied by the three models with their jackknife counterparts. If the nonstationary mean is the main source of misspecification bias, one would expect little or no improvements from jackknifing the HZ1 and HZ2 models.

Table 7 reports the estimation results for bond and option prices implied from Vasicek, HZ1, and HZ2. As shown, the nonstationary mean plays an important role in all cases. In fact, the *RMSE* of HZ1 and HZ2 bond prices are 18% (0.0015/0.0086) and 22% (0.0019/0.008) of that of the Vasicek model, respectively. Also, the HZ1 and HZ2 models increase the implied option prices by 3.6% (4.166/4.02–1) and 3.8% (4.175/4.02–1), respectively, in comparison to Vasicek model. The bias reduction of my proposed model is comparable to the size of the bias reported by Phillips and Yu (2005), who find that the estimation bias of the mean-reversion parameter leads to a 6%–21% downward bias in the in-the-money bond option price and a 1% downward bias in the discount bond price. Also, notice that the jackknife method increases implied option prices by 2.2% for the Vasicek model. However, no material bias reduction is reported in the case of HZ1 and HZ2 models when the jackknife method is used. These results confirm that the nonstationary mean is the main source of downward bias in implied bond option prices.

6. Conclusion

In this paper, I develop an interest rate model that embeds a nonstationary mean in the interest rate process. Relative to several well-known models, the model that provides the best fit with the data in fact is the one with a nonstationary stochastic mean, which also allows for possible nonlinearity in the drift function. My modeling accords with Fama's (2006) assertion that the interest rate exhibits strong evidence of a nonstationary mean. Empirical results show the procedure of imposing a nonstationary mean in the Vasicek model is highly effective and offers substantial improvements in terms of root-mean-square error. In fact, I find that the nonstationary mean governs a substantial majority of the dynamics of bond and bond option prices.

Appendix

This appendix provides details for GMM estimations used in the paper. The stationary mean models are based on the following discrete-time econometric specification

$$r_{t+1} - r_t = \alpha + \beta r_t + \theta_1 r_t^2 + \theta_2 r_t^{-1} + \varepsilon_{t+1} \tag{13}$$

$$E[\varepsilon_{t+1}] = 0, \qquad E[\varepsilon^2{}_{t+1}] = \sigma^2 r_t^{2\gamma}.$$

The five affine models are linked by placing the following restrictions on the parameters:

1. The Chapman and Pearson model: No restrictions.
2. The Ahn and Gao model: $\theta_2 = 0,\ \gamma = 1.5$.
3. The CKLS model: $\theta_1 = 0,\ \theta_2 = 0$.
4. The CIR Model: $\theta_1 = 0,\ \theta_2 = 0,\ \gamma = 0.5$.
5. The Vasicek model: $\theta_1 = 0,\ \theta_2 = 0,\ \gamma = 0$.

Following Hansen (1982) and Chapman and Pearson (2000), the econometric specification of (12) can fit into a GMM framework as follows. Define λ as the entire parameter vector of $\alpha, \beta, \theta_1, \theta_2, \sigma,$ *and* γ. The relevant orthogonality conditions are as follows:

1. **The Chapman and Pearson Model:**

$$E\left[h\left(r_{t+1}, \lambda\right)\right], \qquad h\left(r_{t+1}, \lambda\right) = \begin{bmatrix} \varepsilon_{t+1} \\ \varepsilon_{t+1} r_t \\ \varepsilon_{t+1} r_t^2 \\ \varepsilon_{t+1} r_t^{-1} \\ \varepsilon_{t+1}^2 - \sigma^2 r_t^{2\gamma} \\ \left(\varepsilon_{t+1}^2 - \sigma^2 r_t^{2\gamma}\right) r_t \end{bmatrix},$$

where $\varepsilon_{t+1} = [r_{t+1} - r_t - \alpha - \beta r_t - \theta_1 r_t^2 - \theta_2 r_t^{-1}]$.

2. **The CKLS model:**

$$h\left(r_{t+1}, \lambda\right) = \left[\varepsilon_{t+1},\ \varepsilon_{t+1} r_t,\ \varepsilon_{t+1}^2 - \sigma^2 r_t^{2\gamma}, \left(\varepsilon_{t+1}^2 - \sigma^2 r_t^{2\gamma}\right) r_t\right],$$

where $\varepsilon_{t+1} = [r_{t+1} - r_t - \alpha - \beta r_t]$.

3. **The AG model:**

$$h\left(r_{t+1},\ \lambda\right) = \left[\varepsilon_{t+1},\ \varepsilon_{t+1} r_t,\ \varepsilon_{t+1} r_t^2,\ \varepsilon_{t+1}^2 - \sigma^2 r_t^{2\gamma}, \left(\varepsilon_{t+1}^2 - \sigma^2 r_t^3\right) r_t, \left(\varepsilon_{t+1}^2 - \sigma^2 r_t^3\right) r_t^3\right],$$

where $\varepsilon_{t+1} = [r_{t+1} - r_t - \alpha - \beta r_t - \theta_1 r_t^2]$.

4. **The CIR Model:**

$$h\left(r_{t+1}, \lambda\right) = \left[\varepsilon_{t+1},\ \varepsilon_{t+1} r_t,\ \varepsilon_{t+1}^2 - \sigma^2 r_t,\ \left(\varepsilon_{t+1}^2 - \sigma^2 r_t\right) r_t\right]$$

where $\varepsilon_{t+1} = [r_{t+1} - r_t - \alpha - \beta r_t]$.

5. **The Vasicek Model:**

$$h\left(r_{t+1},\lambda\right)=\left[\varepsilon_{t+1},\ \varepsilon_{t+1}r_t,\ \varepsilon_{t+1}^2-\sigma^2,\ \left(\varepsilon_{t+1}^2-\sigma^2\right)r_t\right]$$

where $\varepsilon_{t+1}=[r_{t+1}-r_t-\alpha-\beta r_t]$.

I also estimate the stochastic mean models (HZ and MAVAS):

6. **The stochastic mean HZ Model:**

$$h\left(r_{t+1},\lambda\right)=\left[\varepsilon_{t+1},\ \varepsilon_{t+1}c_t,\ \varepsilon_{t+1}^2-\sigma^2,\left(\varepsilon_{t+1}^2-\sigma^2\right)c_t\right]$$

where $\varepsilon_{t+1}=[r_{t+1}-r_t-\beta c_t]$.

7. **The Moving Average Vasicek Model:**

$$h\left(r_{t+1},\lambda\right)=\left[\varepsilon^*{}_{t+1},\ \varepsilon^*{}_{t+1}c_t,\ \varepsilon^{*2}{}_{t+1}-\sigma^{*2},\ \left(\varepsilon^{*2}{}_{t+1}-\sigma^{*2}\right)c_t\right],$$

where $\varepsilon^*{}_{t+1}=[r_{t+1}-r_t-\beta c_t]$.

Define Y_t to be a K-dimensional vector of instrumental variables with finite variance that are included in the information set, and let the function f be

$$f\left(x_{t+1},Y_t,\lambda\right)=h\left(x_{t+1},\lambda\right)\otimes Y_t, \tag{14}$$

where $\otimes$ is the Kronecker product. The minimum distance estimator of λ that is defined by the unconditional expectation of (14) can be estimated by minimizing the GMM criterion of the form,

$$\frac{1}{T}\left[f\left(x_{t+1},Y_t,\lambda\right)\right]'W_T\frac{1}{T}\left[f\left(x_{t+1},Y_t,\lambda\right)\right] \tag{15}$$

where W_T is a consistent estimate of $\left(\mathrm{var}\left[(1/T)\left(f\left(x_{t+1},Y_t,\lambda\right)_t\right)\right]\right)^{-1}$.

To test for validity of restrictions among the models, the GMM test of overidentifying restrictions is used. Let L be the number of moment restrictions that are greater than the elements in the parameter vector. The minimized GMM criterion in (15) can be used to test if the remaining, $L-K$, linearly independent moments conditions are zero. Under the null hypothesis that these restrictions are valid, I have that:

$$\frac{1}{T}\left[f\left(x_{t+1},Y_t,\lambda\right)\right]'W_T\frac{1}{T}\left[f\left(x_{t+1},Y_t,\lambda\right)\right]\overset{a}{\sim}\chi^2_{L-k}. \tag{16}$$

For jackknife refinements, I follow Inoue and Shintani (2006) by using the Parzen kernel of Gallant (1987) with two lags to compute the moments weighting matrix. The covariance matrix is robust to heteroskedasticity and autocorrelation.

References

Ahn, D., Gao, B., 1999. A parametric nonlinear model of term structure dynamics. Rev. Financ. Stud. 12, 721–762.
Aït-Sahalia, Y., 1996a. Testing continuous-time models of the spot interest rate. Rev. Financ. Stud. 9, 385–426.
Aït-Sahalia, Y., 1996b. Nonparametric pricing of interest rate derivatives securities. Econometrica 64, 527–560.
Aït-Sahalia, Y., 1999. Transition densities for interest rate and other nonlinear diffusions. The journal of finance 54, 1361–1395.
Aït-Sahalia, Y., 2002. Maximum likelihood estimation of discretely sampled diffusions: A closed-approximation approach. Econometrica 70, 223–262.
Aït-Sahalia, Y., Mykland, P., 2003. The effects of random and discrete sampling when estimating continuous–time diffusions. Econometrica 71, 483–549.
Aksoy, Y., Orphanides, A., Small, D., Wieland, V., Wilcox, D., 2006. A quantitative exploration of the opportunistic approach to disinflation. J. Monetary Econ. 53, 1877–1893.
Al-Zoubi, H., 2008. The long swings in the spot exchange rates and the complex unit roots hypothesis. J. Int. Financ. Mark. Inst. Money 18, 236–244.
Al-Zoubi, H., 2009. Short-term spot rate models with nonparametric deterministic drift. Q. Rev. Econ. Finance 49, 731–747.
Al-Zoubi, H., 2017. Cyclical and persistent carry trade returns and forward premia. Q. J. Financ. 7, 1750010.
Al-Zoubi, H., O'Sullivan, J., Alwathnani, A., 2018. Business cycles, financial cycles and capital structure. Ann. Financ. 14, 105–123.
Anderson, T., Lund, J., 1997. Estimating continuous-time stochastic volatility models of the short-term interest rate. J. Econometrics 77, 343–377.
Arapis, M., Gao, J., 2006. Empirical comparisons in short-term interest rate models using nonparametric methods. J. Financ. Econ. 4, 310–345.
Balduzzi, P., Das, S., Foresi, S., 1998. The central tendency: A second factor in bond yields. Rev. Econ. Stat. 80, 62–72.
Bali, T., Wu, L., 2006. A comprehensive analysis of the short-term interest-rate dynamics. J. Bank. Financ. 30, 1269–1290.
Bandi, M., 2002. Short-term interest rate dynamics: a special approach. J. Financ. Econ. 65, 73–110.
Bao, Y., Ullah, A., Wang, Y., Yu, J., 2014. Bias in the mean reversion estimator in continuous-time Gaussian and Lévy processes. Working paper, Singapore Management University.
Bauer, M., Hamilton, D., 2017. Robust bond risk premia. Rev. Financ. Stud. 31, 399–448.
Bauer, M., Rudebusch, G., Wu, J., 2012. Correcting estimation bias in dynamic term structure models. J. Bus. Econom. Statist. 30, 454–467.
Bergstrom, A., 1986. The estimation of open higher-order continuous time dynamic models with mixed stock and flow data. Econ. Theory 2, 350–373.
Bergstrom, A., 1989. Optimal forecasting of discrete stock and flow data generated by a higher order continuous time system. Comput. Math. Appl. 17, 1203–1214.
Bierens, H., 1997. Testing the unit root with drift hypothesis against nonlinear trend stationary, with an application to the US price level and interest rate. J. Econometrics 81, 29–64.
Cai, L., Swanson, N., 2011. In-and out-of-sample specification analysis of spot rate models: Further evidence for the period 1982–2008. J. Empir. Financ. 18, 743–764.
Chan, K., Karolyi, A., Longstaff, F., Sanders, A., 1992. An empirical comparison of alternative models of the short-term interest rate. J. Finance 47, 1209–1227.
Chapman, D., Pearson, N., 2000. Is the short rate drift actually nonlinear? J. Finance 55, 355–388.
Cieslak, A., Povala, P., 2015. Expected returns in Treasury bonds. Rev. Financ. Stud. 28, 2859–2901.
Cochrane, J., Piazzesi, M., 2005. Bond risk premia. Amer. Econ. Rev. 95, 138–160.

Conley, T., Hansen, L., Luttmer, E., Scheinkman, J., 1997. Short-term interest rates as subordinated diffusions. Rev. Financ. Stud. 10, 525–577.
Cooper, I., Priestley, R., 2008. Time-varying risk premiums and the output gap. Rev. Financ. Stud. 22, 2801–2833.
Cox, J., Ingersoll, J., Ross, S., 1985. A theory of the term structure of interest rates. Econometrica 53, 385–407.
Czellar, V., Karolyi, G., Ronchetti, E., 2007. Indirect robust estimation of the short-term interest rate process. J. Empir. Financ. 14, 546–563.
Dai, Q., Singleton, K.J., 2000. Specification analysis of affine term structure models. J. Finance 55 (5), 1943–1978.
Danthine, J., Girardin, M., 1989. Business cycles in Switzerland: a comparative study. Eur. Econ. Rev. 33 (1), 31–50.
Diebold, F.X., Mariano, R.S., 1995. Com paring predictive accuracy. Journal of Business and Economic Statistics 13 (3), 253–263.
Duffee, G.R., 2002. Term premia and interest rate forecasts in affine models. J. Finance 57 (1), 405–443.
Duffee, G.R., 2011. Forecasting with the term structure: the role of no-arbitrage. Manuscript, Johns Hopkins University.
Duffee, G.R., 2018. Expected inflation and other determinants of Treasury yields. J. Finance 73 (5), 2139–2180.
Duffee, G., Stanton, R., 2012. Estimation of dynamic term structure models. Q. J. Financ. 2, 1–51.
Duffie, D., Kan, R., 1996. A yield-factor model of interest rates. Math. Financ. 6 (4), 379–406.
Durham, G., 2003. Likelihood-based specification analysis of continuous-time models of the short-term interest rate. J. Financ. Econ. 70, 463–487.
Elerian, O., 1998. A note on the existence of a closed-form conditional transition density for the Milstein scheme. Economics discussion paper 1998-W18, Nuffield College, Oxford.
Fama, E., 2006. The behavior of interest rates. Rev. Financ. Stud. 19, 359–379.
Gallant, A., 1987. Nonlinear Statistical Models. Wiley, New York.
Gallant, A., Tauchen, G., 1996. Which moments to match? Econ. Theory 12, 657–681.
Gourieroux, C., Monfort, A., Renault, E., 1993. Indirect inference. J. Appl. Econometrics 8, 85–118.
Greenwood, R., Vayanos, D., 2014. Bond supply and excess bond returns. Rev. Financ. Stud. 27, 663–713.
Hansen, L., 1982. Large sample properties of Generalized Method of Moments estimators. Econometrica 50, 1029–1054.
Hansen, L., Scheinkman, J., 1995. Back to the future: generating moment implications for continuous-time Markov processes. Econometrica 63, 767–804.
Harvey, D., Leybourne, S., Newbold, P., 1997. Testing the equality of prediction mean squared errors. Int. J. Forecast. 13, 281–291.
Harvey, A., Trimbur, T., 2008. Trend estimation and the Hodrick-Prescott filter. J. Japan Statist. Soc. 38, 41–49.
Hodrick, R., Prescott, E., 1997. Postwar US business cycles: an empirical investigation. J. Money Credit Bank. 1–16.
Hong, Y., Li, H., 2005. Nonparametric specification testing for continuous-time models with applications to spot interest rates. Rev. Financ. Stud. 18, 37–84.
Inoue, A., Shintani, M., 2006. Bootstrapping GMM estimators for time series. J. Econometrics 133, 531–555.
Jamshidian, F., 1989. An exact bond option formula. J. Finance 44, 205–209.
Jiang, G., 1998. Nonparametric modeling of US term structure of interest rates and implications of the prices of derivative securities. J. Financ. Quant. Anal. 33, 465–497.
de Jong, R., Sakarya, N., 2016. The econometrics of the Hodrick-Prescott filter. Rev. Econ. Stat. 98, 310–317.
Joslin, S., Priebsch, M., Singleton, K., 2014. Risk premiums in dynamic term structure models with unspanned macro risks. J. Finance 69, 1197–1233.
Kozicki, S., Tinsley, P., 2001. Shifting endpoints in the term structure of interest rates. J. Monetary Econ. 47, 613–652.
Liu, R., 2018. Forecasting bond risk premia with unspanned macroeconomic information. Q. J. Financ. 1940001.
Lo, A., 1988. Maximum likelihood estimation of generalized Itô processes with discretely sampled data. Econ. Theory 4, 231–247.
MacKinnon, J., 1994. Approximate asymptotic distribution functions for unit-root and cointegration tests. J. Bus. Econom. Statist. 2, 167–176.
McElroy, T., Trimbur, T., 2007. Continuous time extraction of a nonstationary signal with illustrations in continuous low-pass and band-pass filtering. Statistics 13.
McElroy, T., Trimbur, T., 2011. On the discretization of continuous-time filters for nonstationary stock and flow time series. Econometric Rev. 30, 475–513.
Mehra, Y., 2004. The output gap, expected future inflation and inflation dynamics: another look. B.E. J. Macroecon. 4, http://dx.doi.org/10.2202/1534-5998.1194.
Merton, R., 1980. On estimating the expected return on the market: An exploratory investigation. J. Financ. Econ. 8, 323–361.
Nowman, K., 1997. Gaussian estimation of single-factor continuous time models of the term structure of interest rates. J. Finance 52, 1695–1703.
Orphanides, A., Wilcox, D., 2002. An opportunistic approach to disinflation. Int. Financ. 5, 47–71.
Pantula, S., Gonzalez-Farias, G., Fuller, G., 1994. A comparison of unit-root test criteria. J. Bus. Econom. Statist. 12, 449–459.
Perron, P., 1989. The great crash, the oil-price shock, and the unit-root hypothesis. Econometrica 57, 1361–1402.
Phillips, P., Yu, J., 2005. Jackknifing bond option prices. Rev. Financ. Stud. 18, 707–742.
Phillips, P., Yu, J., 2009. Maximum likelihood and Gaussian estimation of continuous time models in finance. In: Handbook of Financial Time Series. Springer Berlin Heidelberg, pp. 497–530.
Ravn, M., Uhlig, H., 2002. On adjusting the Hodrick-Prescott filter for the frequency of observations. Rev. Econ. Stat. 84, 371–376.
Stanton, R., 1997. A nonparametric model of term structure dynamics and the market price of interest rate risk. J. Finance 52, 1973–2002.
Tang, C., Chen, S., 2009. Parameter estimation and bias correction for diffusion processes. J. Econometrics 149, 65–81.
Vasicek, O., 1977. An equilibrium characterization of the term structure. J. Financ. Econ. 5, 177–188.
Wecker, W., Ansley, C., 1983. The signal extraction approach to nonlinear regression and spline smoothing. J. Amer. Statist. Assoc. 78, 81–89.
Yu, J., 2012. Bias in the estimation of the mean reversion parameter in continuous time models. J. Econometrics 169, 114–122.