

Local-Momentum Autoregression and the Modeling of Interest Rate Term Structure

Jin-Chuan Duan*

(First Draft: January 19, 2015; This Draft: February 26, 2015)

Abstract

A parsimonious autoregressive model that is globally mean-reverting but locally driven by momentum is proposed. The local-momentum autoregression (LM-AR) model carries one extra parameter, and depending on the sign of this extra parameter, it can be either local momentum-preserving or momentum-building. The LM-AR model is motivated by observing US interest rate movement over many decades, which over a long time span seems to mean revert but over a period of several months or years can actually exhibit a momentum-like behavior. We use the LM-AR model with a stochastic central tendency factor as the dominant global risk factor in interest rates and add a local variation component of the standard mean-reverting type to create a 3-factor risk environment. We then derive its corresponding term structure model and empirically implement the model on US interest rates of seven maturities from January 1954 to December 2013 on a weekly frequency to establish the presence of local momentum building.

*Duan is with the Department of Finance in NUS Business School and the Risk Management Institute of National University of Singapore. E-mail: bizdjc@nus.edu.sg. The author would like to thank Rong Chen, Yangru Wu and Oldrich Vasicek for their valuable inputs. The paper also benefits from the comments by the participants of many conferences.

1 Introduction

Treasury debts form a class of financial instruments that are essential to the functioning of financial markets. On the supply side, governments typically use them to finance fiscal deficits. For the demand side, foreign central banks may through which hold foreign exchange reserves. Fixed-income funds and financial institutions alike need them for investment and funding operations. The yields reflected in treasury debts also play a critical role in the pricing of other financial assets, because their yields serve as the benchmark for defaultable fixed-income investments and other assets such as equities. The treasury yield curve reflects the state of economy and is often used as a policy instrument by government to fine tune a nation's economy. It is therefore natural to think that yield curve offers a wealth of information.

Economists have long been studying treasury yields. Numerous theories have been put forward attempting to explain the formation of yield curve. Econometric analysis of interest rates also abounds. In the finance literature, the focus has been on the pricing and hedging side, dated back to Vasicek (1977). Models have been advanced with the aim of pricing the longer-dated notes and bonds using the short-term interest rates. Models have also been developed to price interest rate derivatives like bond options, interest caps/floors, interest rate futures, and so on. Statistical methodologies for estimating term structure models abound. The literature on interest rate term structure is simply too vast to list and categorize them all. Some examples are Cox, *et al* (1985), Chen and Scott (1993), Duffie and Kan (1996), Dai and Singleton (2000), Duffee (2002), Ang and Piazzesi (2003), Hong and Li (2005), Ait-Sahalia and Kimmel (2010), Joslin, *et al* (2011), Vasicek (2013), and Hamilton and Wu (2012&2014).

This paper contributes to the econometric modeling of interest rates and the pricing of debt instruments. We propose a new 3-factor interest rate model that comprises two components – a global risk factor and a local variation factor. Our contribution rests on devising a local-momentum autoregression process for the global risk factor that concurrently exhibits stochastic central tendency and local momentum. This local-momentum autoregressive process comprises two risk factors – central tendency and stochastic deviation. The model is mean-reverting on a larger timescale and displays local momentum on a smaller timescale. We also derive the corresponding interest rate term structure model which can then be implemented on the interest rates of multiple maturities.

This new model is motivated by observing the 3-month US Treasury yields shown in Figure 1, which plots the rates on a daily basis from January 4, 1954 to December 31, 2013 using the data from the US Federal Reserve Board's website. We note that interest rates over a long time span seem to mean revert albeit very slowly; for example, the US Treasury 3-month yield hit its all-time high of 17.14% on December 11, 1980 according to the Federal Reserve dataset, and then steadily declined over a long period of time. Likewise, it will be unreasonable to expect the low interest rate regime post 2008-09 financial crisis to last forever, and hence the yield will sooner or later move up to a more sustainable level. This mean-reversion phenomenon over a long time span is considered to be a global movement in this paper. When one zeros in on a finer timescale, however, quite a different picture emerges. Interest rates seem to run momentum as opposed to mean-reversion, i.e.,

steady climb (or decline) for a fairly long period of time such as the post dotcom bubble decline over 2001-2004.

The co-existence of global mean-reversion and local momentum is intriguing, but quite understandable because local movements of interest rate tend to be small and directional, but over a longer time span interest rates show large variations through different phases of business cycle and/or over various interest rate policy regimes. We contend that the joint behavior can simply be modeled by a local-momentum autoregression (LM-AR) process proposed in this paper.

The LM-AR process combines a typical mean-reversion term with a new momentum term in the conditional mean. The local momentum is created by allowing the process to either deflect away or converge to its moving average computed over a certain fixed window length. In the case of deflecting, the process is likely to continue its recent upward (or downward) movement, i.e., momentum building, whereas the converging case makes the process to stay around its recent level, i.e., momentum preserving. The LM-AR process is parsimonious because it only adds one extra parameter to the standard mean-reversion autoregressive process. We show that albeit with the local momentum property, the LM-AR model can still be stationary and provides simple conditions for checking stationarity and ergodicity. Moreover, we provide formulas for conditional mean and variance for an arbitrary number of periods forward that may be needed for various applications.

We implement the new 3-factor interest rate term structure model on seven US Treasury yield series that are sampled once a week on Wednesday over the period from January 4, 1954 to December 31, 2013, totaling 3130 data points inclusive of missing values. Since the model involves three latent variables and there are seven observable yield series, we employ a state-space approach to estimation. Specifically, we devise a Kalman filter suitable for the model and carry out maximum likelihood estimation and inference. Our empirical findings show highly significant local momentum in US interest rate data. While the global risk factor with local-momentum largely tracks short-term interest rates, its central tendency component is clearly stochastic and basically reflective of the 20-year yield. This is supportive of the interesting finding by Balduzzi, *et al* (1998) that central tendency is closely tied to long-term interest rates even if two papers employ different models.

2 Local momentum autoregression

In order to understand the local-momentum autoregression model, we need to begin with the standard AR(1) model in its mean-reversion form:

$$\begin{aligned} \Delta X_t &= \kappa_x(\mu - X_{t-1}) + \sigma_x \varepsilon_t & (1) \\ \varepsilon_t | \mathcal{F}_{t-1} &\sim D(0, 1) & (2) \end{aligned}$$

where $\Delta X_t = X_t - X_{t-1}$, $\sigma_x > 0$, and \mathcal{F}_t denotes the filtration generated by $\{X_s; s \leq t\}$. $D(0, 1)$ stands for some distribution with mean 0 and variance 1. It is well known that under $0 < \kappa_x < 2$, X_t is a strictly stationary and ergodic process and μ is its stationary mean. When $\kappa_x = 0$, it becomes a unit-root process. Empirically, many financial time series, such as interest rates, are found to be

very close to unit-root when applying the AR(1) model, i.e., κ_x takes on a small positive value. The AR(1) model possesses an exponential shock decaying property. The long-memory literature such as Bakus and Zin (1993) and Duan and Jacobs (2008) provides evidence that decaying rate should be hyperbolic as opposed to exponential. If the shock decaying rate is indeed hyperbolic, it will force κ_x to be extremely small so as to prolong the shock's influence.

It may be desirable to turn the mean-reverting level into a latent stochastic process, μ_t , which mean-reverts to some constant. Balduzzi, *et al* (1998) employed such an idea to model interest rates, and they referred to this feature as central tendency because empirical evidence suggests that the short-term interest rate tends to move towards the long term interest rate. We will refer to a stationary AR(1) model that moves towards another stationary AR(1) process as the central tendency AR (hereafter, CTAR). For interest rates, CTAR easily dominates the standard AR(1) model, which we will be shown later in the paper.

A less obvious way is to set this moving level as a function of a fixed-window moving average of the past values of X_t to create local momentum. Here, we propose such a simple and yet powerful model whose time-varying level is determined by a combination of two types of stochastic shocks – a latent central tendency factor and a self-generating component. A rather simple modification to equation (1) can arrive at a global mean-reverting autoregressive process that moves around a latent central tendency factor, μ_t , while exhibiting a local momentum behavior. This local momentum CTAR (LM-CTAR) model is intuitively appealing, parsimonious, easily implementable and yet delivers significantly better performance than CTAR. The LM-CTAR process is defined as follows:

$$\Delta X_t = \kappa_x (\mu_t - X_{t-1}) + \omega (\bar{X}_{(t-1)|n} - X_{t-1}) + \sigma_x \varepsilon_t \quad (3)$$

$$\Delta \mu_t = \kappa_\mu (\bar{\mu} - \mu_{t-1}) + \sigma_\mu \epsilon_t \quad (4)$$

$$\bar{X}_{(t-1)|n} = \sum_{i=t-n}^{t-1} b_{t-i} X_i \quad (5)$$

$$\varepsilon_t | \mathcal{G}_{t-1} \sim D(0, 1), \quad \epsilon_t | \mathcal{G}_{t-1} \sim D(0, 1) \quad (6)$$

where $0 < \kappa_\mu < 2$, $\sigma_\mu > 0$, $\kappa_x \geq 0$, $\sigma_x > 0$, and $\sum_{i=1}^n b_i = 1$ with $b_i \geq 0$ for $i = 1, 2, \dots, n$; \mathcal{G}_t denotes the filtration generated by $\{(X_s, \mu_s); s \leq t\}$; and ε_t and ϵ_t are independent. Note that the parameter restriction of $0 < \kappa_\mu < 2$ is to ensure stationarity of μ_t process.

Note that $\bar{X}_{(t-1)|n}$ is meant to be some sort of moving weighted sample mean. If we, for example, let $b_1 = b_2 = \dots = b_n = 1/n$, $\bar{X}_{(t-1)|n}$ becomes the simple sample mean at time $t-1$ for the sample size of n . Using exponentially or hyperbolically decaying weights is also possible, but one needs to normalize the weights to reflect the fact that n is finite.

When exponentially decaying weights are used, it would make more sense to let n go to infinity as well, because the seemingly infinite-dimensional system can actually be reduced to a three-dimensional Markov process. This result can be obtained by setting $b_i = (1-\alpha)\alpha^{i-1}$ with $0 < \alpha < 1$ so that $\sum_{i=1}^{\infty} b_i = 1$. With this weight choice, $\Delta \bar{X}_{t|\infty} = (1-\alpha)[X_t - \bar{X}_{(t-1)|\infty}]$, which can be used

to replace equation (5). As a result, $(X_t, \bar{X}_{t|\infty}, \mu_t)$ forms a three-dimensional Markov system with one extra parameter, α , and one extra latent variable, $\bar{X}_{t|\infty}$. However, one cannot expect a similar dimension-reduction result with hyperbolically decaying weights, because there exists no common multiplier enabling the simplification.

When $\omega = 0$, the LM-CTAR process reduces to CTAR. Since μ_t process in equation (4) can have an MA(∞) representation, the CTAR model can be interpreted as ARMA(1, ∞). If μ_t is a constant, i.e., $\sigma_\mu = 0$, this restricted version of LM-CTAR will be referred to as local momentum AR (LM-AR). If $\omega > 0$ and $\kappa_x > 0$, the process will move towards a time-varying target level that is jointly determined by an external latent central tendency factor and the process' own past n -period moving average. In such case, the model exhibits a *local momentum-preserving* feature. If $\omega < 0$ and κ_x is large enough to ensure stationarity, then the model exhibits a *local momentum-building* characteristic, meaning that the process tends to continue its current local trend (up or down).

Momentum is a much studied issue in finance. In Wu and Zhang (1996) and Balvers and Wu (2006), for example, momentum is modeled through a linear combination of many past one-period changes. If all coefficients are identical, this linear combination reduces to the most recent term minus the earliest term, cancelling out all terms in between. In empirical implementation, the coefficients in the combination almost necessarily need to differ for different past periods, and thus their momentum model entails more parameters. Although their specification has an intuitive merit of its own, our local momentum formulation significantly differs from theirs and is far more parsimonious.

The idea behind our local momentum formulation is quite intuitive by considering interest rates. The latent stochastic process, μ_t , defining the central tendency level, can be interpreted as reflecting different monetary regimes over time. The local-momentum feature is meant to capture the local behavior within a monetary regime. When the rate has been high (low) as reflected in the n -period moving average, the process will revolve around a higher (lower) level vis-a-vis the AR(1, ∞) model. In fact, our later analysis shows that the local behavior is of the local momentum-building type. In short, the LM-CTAR model is better in preserving (or building) momentum in interest rates beyond the extent possible under the standard model. The degree of momentum preservation (or building) is determined by the magnitude and sign of ω . Naturally, the volatility of the LM-CTAR model is larger vis-a-vis the otherwise identical standard AR(1, ∞) model.

We now characterize stationarity and ergodicity for the LM-CTAR process. Let

$$\mathbf{X}_t = \begin{bmatrix} X_t \\ X_{t-1} \\ \vdots \\ X_{t-n+1} \end{bmatrix} \quad \mathbf{Z}_t = \begin{bmatrix} \kappa_x(\mu_t - \bar{\mu}) + \sigma_x \varepsilon_t \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad \mathbf{A} = \begin{bmatrix} \kappa_x \bar{\mu} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$\mathbf{B} = \begin{bmatrix} 1 - \kappa_x - \omega(1 - b_1) & \omega b_2 & \dots & \omega b_{n-1} & \omega b_n \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix}$$

where \mathbf{X}_t , \mathbf{Z}_t and \mathbf{A} are n -dimensional column vectors, and \mathbf{B} is an $n \times n$ matrix. Express the LM-CTAR model in a matrix-vector form:

$$\mathbf{X}_t = \mathbf{A} + \mathbf{B}\mathbf{X}_{t-1} + \mathbf{Z}_t. \quad (7)$$

Again, μ_t in equation (4) has an MA(∞) representation with exponentially decaying coefficients, the LM-CTAR model thus looks like a typical ARMA(n, ∞) model except with a unique way of restricting its AR and MA parameters. Since the MA(∞) component is deduced from inverting a stationary AR(1) process, the spectral radius of \mathbf{B} , denoted by $\rho(\mathbf{B})$, alone determines the stationarity and ergodicity of the LM-CTAR process; that is, $\rho(\mathbf{B}) < 1$ ensures stationarity and ergodicity. The general condition and its two interesting special cases are given in Proposition 1 in Appendix A.

The LM-CTAR model can have a unit root in an interesting way. We note that $\kappa_x = 0$ gives rise to a unit root process regardless of the value of ω which governs the local moment behavior. We define

$$\mathbf{F} = \begin{bmatrix} \omega(b_1 - 1) & \omega(b_1 + b_2 - 1) & \dots & \omega(\sum_{i=1}^{n-2} b_i - 1) & \omega(\sum_{i=1}^{n-1} b_i - 1) \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix}$$

When $\kappa_x = 0$, differencing the series once, i.e., ΔX_t , gives rise to a stationary and ergodic series if $\rho(\mathbf{F}) < 1$ (see Proposition 2 in Appendix A).

When X_t is strictly stationary, we can compute its stationary mean and variance. It is straightforward to obtain $E(X_t) = \bar{\mu}$. For the stationary variance, we rely on the conditional mean and variance formulas given below.

The k -period ($k \geq 1$) ahead forecast and forecasting variance can be computed using the vector form in equation (7); that is,

$$E[\mathbf{X}_{t+k}|\mathcal{G}_t] = \mathbf{B}^k \mathbf{X}_t + (\mathbf{I} - \mathbf{B})^{-1}(\mathbf{I} - \mathbf{B}^k) \mathbf{A} + 1_{\{\kappa_\mu \neq 1\}} \kappa_x (1 - \kappa_u)^k (\mu_t - \bar{\mu}) \left[\mathbf{I} - \frac{\mathbf{B}}{1 - \kappa_\mu} \right]^{-1} \left[\mathbf{I} - \left(\frac{\mathbf{B}}{1 - \kappa_\mu} \right)^k \right] \mathbf{e} \quad (8)$$

$$\begin{aligned} Var[\mathbf{X}_{t+k}|\mathcal{G}_t] &= \sum_{i=0}^{k-1} \mathbf{B}^i E \{ [\mathbf{Z}_{t+k-i} - E(\mathbf{Z}_{t+k-i}|\mathcal{G}_t)] [\mathbf{Z}_{t+k-i} - E(\mathbf{Z}_{t+k-i}|\mathcal{G}_t)]' | \mathcal{G}_t \} (\mathbf{B}^i)' \\ &= \sum_{i=0}^{k-1} \mathbf{B}^i \mathbf{V}_i (\mathbf{B}^i)' \end{aligned} \quad (9)$$

where

$$\mathbf{e} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad \mathbf{V}_i = \begin{bmatrix} \sigma_x^2 + \kappa_x^2 \sigma_\mu^2 \left(\frac{1 - (1 - \kappa_\mu)^{2(i+1)}}{1 - (1 - \kappa_\mu)^2} \right) & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix}$$

When the LM-CTAR model is stationary, all element of $E[\mathbf{X}_{t+k}|\mathcal{G}_t]$ converges to $\bar{\mu}$, the stationary mean, as k goes to infinity, because its limit, $(\mathbf{I} - \mathbf{B})^{-1} \mathbf{A}$, is a constant vector of $\bar{\mu}$. The stationary covariance matrix for \mathbf{X}_t can be obtained by sending k to infinity, i.e., $Var(\mathbf{X}_t) = \sum_{i=0}^{\infty} \mathbf{B}^i \mathbf{V}_i (\mathbf{B}^i)'$. The term $\left[\mathbf{I} - \frac{\mathbf{B}}{1 - \kappa_\mu} \right]^{-1} \left[\mathbf{I} - \left(\frac{\mathbf{B}}{1 - \kappa_\mu} \right)^k \right]$ in equation (8) should be understood as equal to $\sum_{i=0}^{k-1} \mathbf{B}^i$ when $\mathbf{I} - \frac{\mathbf{B}}{1 - \kappa_\mu}$ is singular. The forecasting mean and variance formulas can be used for forecasting, and also become handy for treating missing data in estimation which is in effect used in our empirical analysis later.

To better understand local-momentum, we consider the LM-AR model, i.e., fixing the central tendency factor to a constant. We plot in Figure 2a a simulated sample path using a 5-period moving average and the parameter values: $\bar{\mu} = 0$, $\sigma_\mu = 0$, $\kappa_x = 0.001$, $\omega = -0.5$, and $\sigma_x = 0.002$. In addition, we use the conditional normality. The constraint, $\sigma_\mu = 0$, turns μ_t into a constant. The spectral radius corresponding to these parameters is 0.99981248 which implies stationarity and ergodicity. The stationary mean and standard deviation are 0 and 1.6318207, respectively. It is evident from the plot that the LM-AR model is capable of generating large and fairly regular cycles.

In Figure 2b, we construct the LM-AR model using a 10-period moving average and the parameter values: $\bar{\mu} = 0$, $\sigma_\mu = 0$, $\kappa_x = 0.001$, $\omega = -0.2$, and $\sigma_x = 0.02$. In this case, the spectral radius becomes 0.98572727. The stationary mean and standard deviation equal 0 and 1.4013125, respectively. The plot reveals that the LM-AR model in this case takes less frequent variations and is also not as regular in shape when compared to Figure 2a.

If a continuous-time model is preferred, the LM-CTAR model can be stated as

$$dX_t = [\kappa_x(\mu_t - X_t) + \omega(\bar{X}_t(\tau) - X_t)] dt + \sigma_x dW_{xt} \quad (10)$$

$$d\mu_t = \kappa_\mu(\bar{\mu} - \mu_{t-1}) dt + \sigma_\mu dW_{\mu t} \quad (11)$$

$$\bar{X}_t(\tau) = \int_{t-\tau}^t b(t-s)X_s ds \quad (12)$$

where $\kappa_\mu > 0$, $\sigma_\mu > 0$, $\kappa_x \geq 0$, $\sigma_x > 0$, and $\int_0^\tau b(s)ds = 1$ with $b(s) \geq 0$ for $0 \leq s \leq \tau$; and W_{xt} and $W_{\mu t}$ are two independent Wiener processes. Setting $b(s) = \frac{1}{\tau}$ for all s is, for example, an equal-weight choice. Note that $\kappa_\mu > 0$ is sufficient to ensure stationarity of μ_t process in continuous time.

The above system is obviously a non-Markovian stochastic process, and cannot be turned into a finite-dimensional Markov process due to the integral in defining $\bar{X}_t(\tau)$. This becomes quite clear by analyzing the time dynamic of $\bar{X}_t(\tau)$. Note that $\bar{X}_t(\tau)$ is of finite variation. If $b(s)$ is differentiable, one can take its derivative per usual to arrive at

$$\frac{d\bar{X}_t(\tau)}{dt} = b(0)X_t - b(\tau)X_{t-\tau} + \int_{t-\tau}^t b'(t-s)X_s ds. \quad (13)$$

With the presence of $X_{t-\tau}$, the relevant information set must be the entire history up to $t - \tau$. If equal weighting is the choice, i.e., $b(s) = \frac{1}{\tau}$, the third term in equation (13) vanishes. Even if we let τ go to infinity with $b(\infty) = 0$, the system in general cannot be reduced to a lower-dimensional Markov process due to the third term in equation (13). When the weight function is of an exponentially decaying type, i.e., $b(s) = \alpha e^{-\alpha s}$ with $\alpha > 0$, the third term can be simplified back to $-\alpha \bar{X}_t(\infty)$, which in turn gives rise to $d\bar{X}_t(\infty) = \alpha[X_t - \bar{X}_t(\infty)]dt$. Thus, the continuous-time LM-CTAR model becomes a three-dimensional Markov system, i.e., $(X_t, \bar{X}_t(\infty), \mu_t)$, which not surprisingly agrees with its discrete-time counterpart discussed earlier.

The LM-CTAR model can be used to model endogenous cycles. If further variations around the endogenous cycles are needed, one can simply add to the LM-CTAR model with an AR(1) process where the LM-CTAR component is meant to capture global movements and the extra AR(1) component addresses local variations. In the next section, we will adopt this strategy to model interest rates. For now, we will just estimate the LM-CTAR model to the observed interest rate series without an extra AR(1) component. We will compare its performance with that of the standard mean-reverting model.

The term structure data used in our study are the US treasury constant maturity yields of seven maturities: 3 months, 6 months, 1 year, 2 years, 5 years, 10 years and 20 years. The data for the period from January 4, 1954 to December 31, 2013 are taken from the website of the US Federal Reserve Board. We use weekly data on Wednesday. The 3-month yield has the longest history which dates back to January 4, 1954, and it is also most complete with only 50 missing data points over 3130 weeks. There are many missing data for other maturities, particularly in the earlier time period; for example, the one-year yield becomes available only after January 2, 1962.

We follow the standard practice of converting interest rates to the continuously compounded form. The summary statistics for the continuously compounded yields for the seven maturities are given in Table 1.

The 3-month yield series is for now assumed to follow the LM-CTAR model. In its general form, the LM-CTAR process is a state-space model because μ_t is latent. When local momentum is switched off, CTAR is still a state-space model, again due to latency of μ_t . Both the LM-CTAR and CTAR models can be estimated with the help of the Kalman filter by imposing a further assumption of normality on ε_t and ϵ_t .¹ The measurement equation is (3) whereas the transition equation is (4). If μ_t is set to a constant, there is no latency and the model can be straightforwardly estimated. The results for four versions of the LM-CTAR model are reported in Table 2. The 3-month US Treasury constant maturity yields from January 4, 1954 to December 31, 2013 (sampled once a week on Wednesday, totaling 3130 data points including 50 missing values) are used in the analysis. When a missing value is encountered, the appropriate likelihood will need to reflect the fact that the missing data point is skipped over. If, say, only one week is skipped over, the conditional expected value will be the first entry of $\mathbf{B}^2\mathbf{X}_{t-2} + (\mathbf{I} + \mathbf{B})\mathbf{A} + 1_{\{\kappa_\mu \neq 1\}}\kappa_x(1 - \kappa_u)^2(\mu_{t-2} - \bar{\mu})\left[\mathbf{I} + \frac{\mathbf{B}}{1 - \kappa_\mu}\right]\mathbf{e}$ and the conditional variance will be the first entry of $\mathbf{V}_0 + \mathbf{B}\mathbf{V}_1\mathbf{B}'$. When μ_t is latent, we apply the Kalman filter that starts with the first filtered value of μ_1 equal to X_1 and the filtered variance equal to zero.² Missing data are handled by applying the Kalman filter in a natural way.

The results reported in Table 2 use seven weeks to compute the moving average if the local momentum feature is allowed. Our analysis reveals that the LM-AR model with the seven-week moving average performs better than any other number of weeks. The estimate for ω shows a highly significant value of -0.071, suggesting that the 3-month interest rate exhibits a local momentum-building feature. The rate tends to continue its upward (or downward) movement until going too far away from its long-run mean so that the mean-reversion term associated with parameter κ_x effectively kicks in. Along with the results for the LM-AR model, we report the standard mean-reverting model results in the second column. It is clear by judging from the log-likelihood values that the standard AR(1) model as compared to the LM-AR model performs rather poorly.

The central tendency feature identified by Balduzzi, *et al* (1998) and reflected in the CTAR model is clearly evident in Table 2 either by comparing the log-likelihood of the CTAR and AR(1) models or examining individual significance of the two parameters (i.e., κ_μ and σ_μ) that govern the central tendency feature. The level that the CTAR process reverts to is clearly stochastic (significant σ_μ) and also mean-reverting (significant κ_μ). Worth noting is the fact that CTAR has a lower log-likelihood value than LM-AR even though the latter has one less parameter. By the AIC criterion, LM-AR dominates CTAR. Similarly, central tendency can be incorporated into

¹If innovations are not normally distributed, the mixture Kalman filter of Chen and Liu (2000), which approximates innovations by a Gaussian mixture, can be deployed in estimation. One can also use a particle filter such as localized particle filter of Duan and Fulop (2009) to conduct estimation.

²Using stationary mean and variance as the starting filtered mean and variance is not a desirable option, because the observed time series can start far from the stationary mean, and that is the case for the three-month US interest rate series used in this study.

the local momentum model. Doing so, the increase in log-likelihood from LM-AR to LM-CTAR with two degrees of freedom is significant at the 5% level. The local momentum building feature (i.e., negative ω) is clearly not affected by introducing central tendency, because the estimate of ω remains hardly changed. In short, local momentum in the 3-month Treasury yield series appears to be genuine and robust. It is worth noting that the spectral radius of \mathbf{B} has dropped substantially from 0.9957 to 0.8259, i.e., a faster mean reversion, when the central tendency factor is introduced into the local momentum model. This is understandable because with the central tendency factor setting the global trend, the local momentum mechanism is free to capture the local behavior without the burden of matching the global trend.

3 Modeling interest rates

In the section, we use the discrete-time version of the LM-CTAR process to construct the interest rate term structure model. It will become fairly clear that one can also construct the interest rate term structure model with the continuous-time equivalent version of the LM-CTAR process in equations (10)-(11).

3.1 A state-space model for interest rates

We now specify an interest rate process driven by three risk factors. A LM-CTAR process is taken as the factor to describe for the global movement of the interest rate dynamic, which contains two risk factors. Adding onto the global movement is a local variation component which comprises one shock governed by a independent autoregressive process of order 1. This local variation exhibits the typical transient effect, decaying at the exponential rate determined by the autoregression coefficient.

Since the interest rate is a function of maturity, we add τ to the notation and use $r_t(\tau)$ to denote the risk-free zero-coupon yield (continuously compounded) at time t with maturity τ periods of length h . To simplify the notation, we will just use r_t to denote the interest rate with the one-period maturity.

Specifically,

$$r_t = X_t + v_t \tag{14}$$

The local-momentum system driving X_t has been described earlier in equations (3)-(6). In addition, we assume that the innovation terms are governed by normal random variables. The local variation process is first-order autoregressive so that local shocks can still have lingering effects. Specifically,

$$\Delta v_t = -\kappa_v v_{t-1} + \sigma_v \xi_t \tag{15}$$

$$\xi_t | (\mathcal{G}_t \cup v_{t-1}) \sim N(0, 1) \tag{16}$$

where $0 < \kappa_v < 2$, and $(\mathcal{G}_t \cup v_{t-1})$ denotes the minimum σ -algebra generated by \mathcal{G}_t and v_{t-1} . Naturally, we make the local variation process to evolve around zero mean.

It should be noted that under the above formulation, \mathbf{X}_t is no longer directly observable. The state-space interest rate model needs to be recast using the following matrix-vector system:

$$r_t = \mathbf{H}'\mathbf{X}_t^* \quad (17)$$

$$\mathbf{S}\mathbf{X}_t^* = \mathbf{C} + \mathbf{D}\mathbf{X}_{t-1}^* + \mathbf{W}_t \quad (18)$$

where

$$\mathbf{H} = \begin{bmatrix} 1 \\ \mathbf{0}_{n-1} \\ 0 \\ 1 \end{bmatrix} \quad \mathbf{X}_t^* = \begin{bmatrix} \mathbf{X}_t \\ \mu_t - \bar{\mu} \\ v_t \end{bmatrix} \quad \mathbf{W}_t = \begin{bmatrix} \sigma_x \varepsilon_t \\ \mathbf{0}_{n-1} \\ \sigma_\mu \epsilon_t \\ \sigma_v \xi_t \end{bmatrix} \quad \mathbf{C} = \begin{bmatrix} \mathbf{A} \\ 0 \\ 0 \end{bmatrix}$$

$$\mathbf{D} = \begin{bmatrix} \mathbf{B} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & 1 - \kappa_\mu & 0 \\ \mathbf{0} & 0 & 1 - \kappa_v \end{bmatrix} \quad \mathbf{S} = \begin{bmatrix} 1 & \mathbf{0} & -\kappa_x & 0 \\ \mathbf{0} & \mathbf{I}_{(n-1) \times (n-1)} & \mathbf{0} & \mathbf{0} \\ 0 & \mathbf{0} & 1 & 0 \\ 0 & \mathbf{0} & 0 & 1 \end{bmatrix}$$

$$\mathbf{U} = \begin{bmatrix} \sigma_x^2 & \mathbf{0} & 0 & 0 \\ \mathbf{0} & \mathbf{0}_{(n-1) \times (n-1)} & \mathbf{0} & \mathbf{0} \\ 0 & \mathbf{0} & \sigma_\mu^2 & 0 \\ 0 & \mathbf{0} & 0 & \sigma_v^2 \end{bmatrix}$$

and \mathbf{U} is the covariance matrix for \mathbf{W}_t . Note that \mathbf{X}_t , \mathbf{A} and \mathbf{B} have been previously defined. Similar to Proposition 1, the interest rate process, i.e., r_t , is strictly stationary and ergodic if the spectral radius of $\mathbf{S}^{-1}\mathbf{D}$ is less than 1.

Since X_t , μ_t and v_t are latent stochastic processes with some dynamic structure, all of them can be regarded signals. If r_t is directly observable, equation (18) serves as a measurement equation, and then there are only two latent processes. In general, there are many interest rates with different maturities concurrently available, one will need a term structure model to define measurement equations, and there will be three latent processes.

3.2 Term structure of interest rates

For our interest rate model, it is possible to derive an analytical term structure formula much like the standard exponential affine term structure model pioneered by Vasicek (1977) and further developed later by others. For this, we need to deal with risk premiums associated with the three innovation terms in the interest rate model. Let h be the length of one period measured as the fraction of a year. We assume the following stochastic discount factor from time $t + \tau$ back to time t is $\exp[-r_t(\tau)\tau h]M_{t,t+\tau}$ where for $s \geq t$,

$$M_{t,s} = \alpha(t, s) \exp \left[\sum_{j=t+1}^s [(\lambda_0 + \lambda_1 X_{j-1})\varepsilon_j + (\psi_{\mu 0} + \psi_{\mu 1} \mu_{j-1})\epsilon_j + (\psi_{v 0} + \psi_{v 1} v_{j-1})\xi_j] \right] \quad (19)$$

Note that $\alpha(t, s)$ is a factor (can be stochastic but must be measurable with respect to time- $(s - 1)$ information) for making $M_{t,s}$ a martingale for $s \geq t$, and $\alpha(t, t) = 1$ so that $M_{t,t} = 1$. This way of defining the stochastic discount factor basically uses the default-free zero-coupon bond of a relevant maturity as the numeraire asset.

Define a martingale measure $Q_{t,T}$ by setting $dQ_{t,T}/dP = M_{t,T}$. We show in Appendix B that the distributional characteristics of ε_s , ϵ_s and ξ_s under measure $Q_{t,T}$ actually remain the same for different t or T . We will thus drop the subscript and use Q instead. Furthermore, $\varepsilon_t^Q = \varepsilon_t - \lambda_0 - \lambda_1 X_{t-1}$, $\epsilon_t^Q = \epsilon_t - \psi_{\mu 0} - \psi_{\mu 1} \mu_{t-1}$ and $\xi_t^Q = \xi_t - \psi_{v 0} - \psi_{v 1} v_{t-1}$ are three independent standard normal random variables under measure Q . Measure Q is in essence characterized by the six risk premium parameters – λ_0 , λ_1 , $\psi_{\mu 0}$, $\psi_{\mu 1}$, $\psi_{v 0}$ and $\psi_{v 1}$, and is unique once these risk premium parameters are fixed. Also shown in Appendix B is an equilibrium restriction between r_t and $\alpha(t, t + 1)$. We have adopted $r_t = X_t + v_t$ in equation (14). Alternatively, one can choose to set $\alpha(t, t)$ as a function of state variables and deduce r_t accordingly.

Let $f_t(\tau)$ denote the one-period forward rate at time t starting at time $t + \tau$, where each of the τ periods is of length h . Define \mathcal{H}_t as the filtration generated by $\{(X_s, \mu_s, v_s); s \leq t\}$. By the pricing result above, it must be the case that, for $\tau \geq 1$,

$$f_t(\tau) = -\frac{\ln E^Q(e^{-r_{t+\tau}h}|\mathcal{H}_t)}{h} \quad (20)$$

and for $\tau = 0$, $f_t(0) = r_t$. Forward rates for different forward starting times can in turn be used to compute spot interest rate rates such as, for $\tau \geq 1$,

$$r_t(\tau) = \frac{1}{\tau} \sum_{j=0}^{\tau-1} f_t(j). \quad (21)$$

In order to compute the forward rate, we need to restate the interest rate model in equations (17)-(18) in an alternative form:

$$r_t = \mathbf{H}'\mathbf{X}_t^* \quad (22)$$

$$\mathbf{S}\mathbf{X}_t^* = \mathbf{C}^* + \mathbf{D}^*\mathbf{X}_{t-1}^* + \mathbf{W}_t^* \quad (23)$$

$$\mathbf{C}^* = \begin{bmatrix} \kappa_x \bar{\mu} + \sigma_x \lambda_0 \\ \mathbf{0}_{n-1} \\ \sigma_\mu \psi_{\mu 0} \\ \sigma_v \psi_{v 0} \end{bmatrix} \quad \mathbf{W}_t^* = \begin{bmatrix} \sigma_x \varepsilon_t^Q \\ \mathbf{0}_{n-1} \\ \sigma_\mu \varepsilon_t^Q \\ \sigma_v \xi_t^Q \end{bmatrix}$$

$$\mathbf{D}^* = \begin{bmatrix} d & \omega b_2 & \dots & \omega b_{n-1} & \omega b_n & 0 & 0 \\ 1 & 0 & \dots & 0 & 0 & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 & 1 - \kappa_\mu + \sigma_\mu \psi_{\mu 1} & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 & 1 - \kappa_v + \sigma_v \psi_{v 1} \end{bmatrix}$$

where $d = 1 - \kappa_x - \omega(1 - b_1) + \sigma_x \lambda_1$.

Thus, the continuously compounded forward rate can be derived using the above system under measure Q . The future latent state $\mathbf{X}_{t+\tau}^*$ conditional on \mathcal{H}_t is Gaussian, which gives rise to the following:

$$\begin{aligned} f_t(\tau) &= -\frac{\ln E^Q(e^{-r_{t+\tau}h} | \mathcal{H}_t)}{h} \\ &= \mathbf{H}' E^Q(\mathbf{X}_{t+\tau}^* | \mathcal{H}_t) - \frac{h}{2} \mathbf{H}' \text{Var}^Q(\mathbf{X}_{t+\tau}^* | \mathcal{H}_t) \mathbf{H} \end{aligned} \quad (24)$$

where

$$E^Q(\mathbf{X}_{t+\tau}^* | \mathcal{H}_t) = (\mathbf{S}^{-1} \mathbf{D}^*)^\tau \mathbf{X}_t^* + (\mathbf{I} - \mathbf{S}^{-1} \mathbf{D}^*)^{-1} [\mathbf{I} - (\mathbf{S}^{-1} \mathbf{D}^*)^\tau] \mathbf{S}^{-1} \mathbf{C}^* \quad (25)$$

$$\text{Var}^Q(\mathbf{X}_{t+\tau}^* | \mathcal{H}_t) = \sum_{i=0}^{\tau-1} (\mathbf{S}^{-1} \mathbf{D}^*)^i \mathbf{S}^{-1} \mathbf{U} (\mathbf{S}^{-1})' [(\mathbf{S}^{-1} \mathbf{D}^*)^i]^\prime \quad (26)$$

Note that the covariance matrix of \mathbf{W}_t in equation (23) under measure Q is still \mathbf{U} , because it is per usual unaffected by the measure change..

The above results can be combined to yield the spot rate of any maturity as follows: for $\tau \geq 1$,

$$r_t(\tau) = \Phi_1(\tau) + \Phi_2(\tau) \mathbf{X}_t^* \quad (27)$$

where

$$\begin{aligned} \Phi_1(\tau) &= \mathbf{H}'(\mathbf{I} - \mathbf{S}^{-1}\mathbf{D}^*)^{-1} \left(\mathbf{I} - \frac{1}{\tau} \sum_{j=0}^{\tau-1} (\mathbf{S}^{-1}\mathbf{D}^*)^j \right) \mathbf{S}^{-1}\mathbf{C}^* \\ &\quad - \mathbf{H}' \left(\frac{h}{2\tau} \sum_{j=0}^{\tau-1} \sum_{i=0}^{j-1} (\mathbf{S}^{-1}\mathbf{D}^*)^i \mathbf{S}^{-1}\mathbf{U}(\mathbf{S}^{-1})' [(\mathbf{S}^{-1}\mathbf{D}^*)^i]' \right) \mathbf{H} \end{aligned} \quad (28)$$

$$\Phi_2(\tau) = \mathbf{H}' \left(\frac{1}{\tau} \sum_{j=0}^{\tau-1} (\mathbf{S}^{-1}\mathbf{D}^*)^j \right) \quad (29)$$

Note that the sum should be understood as $\mathbf{0}$ when its bottom index is larger than the upper index.

The above term structure model gives rise to a 3-factor Vasicek (1977) model if we turn off the local momentum feature, i.e., $\omega = 0$. The local-momentum feature fundamentally changes the nature of term structure; for example, the LM-AR factor alone is able to produce the hump-shaped term structure. The reason is quite intuitive. When the base rate is at the upward momentum phase, the interest rate will increase with maturity up to some point, but will eventually decline with maturity as the global mean-reversion force increases strength due to the fact that the base rate has deviated too far away from its global mean level. Using the LM-CTAR factor (i.e., adding central tendency), our 3-factor interest rate term structure is even more flexible as will be seen later.

With the term structure model in place, we are able to explicitly address the term effect and to estimate with many rates of different maturities concurrently. When multiple rates are used in estimation, we need to introduce pricing errors. Suppose we have k observed rates $\{\tilde{r}_t(\tau_1), \dots, \tilde{r}_t(\tau_k)\}$. Then, the following measurement equation follows:

$$\begin{aligned} \tilde{r}_t(\tau_1) &= \Phi_1(\tau_1) + \Phi_2(\tau_1)\mathbf{X}_t^* + \epsilon_{1t} \\ \tilde{r}_t(\tau_2) &= \Phi_1(\tau_2) + \Phi_2(\tau_2)\mathbf{X}_t^* + \epsilon_{1t} \\ &\quad \vdots \\ \tilde{r}_t(\tau_k) &= \Phi_1(\tau_k) + \Phi_2(\tau_k)\mathbf{X}_t^* + \epsilon_{kt} \end{aligned} \quad (30)$$

where the measurement errors $\{\epsilon_{1t}, \dots, \epsilon_{kt}\}$ are assumed to be normally distributed with mean zero so that we can run the Kalman filter.³ In addition to the above measurement equation, we have the transition equation in (18) to complete the specification of the state-space model. The specifics of the Kalman filter for this model is given in Appendix C.

We need to deal with pricing errors when different rates from the observed term structure are used in estimation. If we set rate-specific error magnitudes, there will be seven parameters to

³Again, normality is not a necessity, because one can use the mixture Kalman filter of Chen and Liu (2000) or the localized particle filter of Duan and Fulop (2009) to estimate a model with non-Gaussian measurement errors.

correspond to seven rate series. The three latent factor processes are governed by eight parameters under the physical probability, i.e., $\{\bar{\mu}, \kappa_x, \omega, \sigma_x, \kappa_\mu, \sigma_\mu, \kappa_v, \sigma_v\}$. In addition, there are six parameters specifically arising from risk-neutralization, which are $\{\lambda_0, \lambda_1, \psi_{\mu 0}, \psi_{\mu 1}, \psi_{v 0}, \psi_{v 1}\}$. However, there is only one identifiable parameter among $\lambda_0, \psi_{\mu 0}$ and $\psi_{v 0}$, because these three enter into the same constant in the risk neutral system. Our estimation using seven US Treasury yields reveals that these risk premium parameters are not significantly different from zero. With interest derivatives such as caplets and bond options added to the data set, one may be in a better position to estimate the four identifiable risk premium parameters.

Table 3 reports our findings for two versions of our 3-factor term structure model using the seven weekly US Treasury yield series described in the previous section. The first model, denoted by CTAR+AR(1), sets ω to zero whereas the second, denoted by LM-CTAT+AR(1), has no restriction. The results corroborate with our earlier findings by using the 3-month yield series alone, that ω is highly significantly and negative. Local momentum is clearly present, and it is of local momentum building type. Adding an extra AR(1) factor to the base rate and estimating with multiple yields have, however, changed the magnitude of ω substantially from -0.075 to -0.0071. Also different from the earlier result is the best fixed-window length for computing the moving average equal to 28 periods as opposed to seven. The stochastic and mean-reverting central tendency factor is clearly present, and the result is not affected by whether local momentum is permitted. Likewise, the measurement errors of the seven yield series are not affected by the presence of local momentum. Importantly, the spectral radius of matrix \mathbf{B} which is intimately associated with local momentum behavior reveals a fast mean reversion to the central tendency factor, because 0.8913 is lower than 0.9277 (i.e., $1-\kappa_x$ under CTAR+AR(1)). This happens because the local momentum mechanism no longer needs to play the role of setting the global trend as in the earlier discussion related to Table 2.

Figure 3a displays the filtered LM-CTAR and central tendency series along with the 3-month US Treasury yields over the sample period. It is evident that the LM-CTAR process revolves around the central tendency factor, which in turn tracks the overall interest rate level over a long time span quite well. Recall that the base rate (one week in our implementation) is a direct sum of LM-CTAR factor and another AR(1) process. Although the base interest rate is not the same as the 3-month rate due to the term structure effect, its filtered values (not reported here to conserve space) are numerically close to their corresponding 3-month rates over the sample period. The graph shows that the filtered LM-CTAR factor (part of the filtered base rate) can be away from the 3-month yield, and the gap is filled in by the extra AR(1) factor, i.e., v_t .

The difference between the local-momentum and more traditional 3-factor AR term structure model is evident in the difference between the filtered LM-CTAR factor in the former and the filtered CTAR in the latter as displayed in Figure 3b. In 1960s, the difference was as high as over 30 basis points. Throughout the sample period, the filtered values of the LM-CTAR factor were typically lower than those of the CTAR factor, but they were usually within 10 basis points of each other.

3.3 Central tendency and local momentum and their impacts on yields

As argued in Balduzzi, et al (1998), the central tendency factor should be more reflective of longer-term interest rates. Here, we take a closer look by plotting the filtered central tendency factor under two 3-factor term structure models along with the 20-year constant maturity US Treasury yield in Figure 4. The 20-year rate was not always available over the sample period. For the available periods, the three series are fairly close to one another, confirming that the central tendency factor is indeed reflective of the longest-term interest rate in our dataset. The fact that the two central tendency estimates are fairly close to each other regardless of having local momentum or not suggests that central tendency is a long-run phenomenon that differs from the local behavior of momentum. In short, the momentum captured by our model and contributes the term structure model’s performance is indeed local in nature.

As Figure 4 suggests, the three factors in our local momentum term structure model should reflect different segments of term structure of interest rates. We now run the following diagnostic regression for changes in seven yield series to gain further insights about the roles of these three factors.

$$\Delta r_t(\tau) = a_\tau + b_\tau \Delta \hat{\mu}_{t|t} + c_\tau \Delta \hat{X}_{t|t} + d_\tau \Delta \hat{v}_{t|t} + \epsilon_t(\tau) \quad (31)$$

The above regression should be understood as only suggestive, because the explanatory variables are filtered estimates that are subjected to obvious endogeneity and measurement errors.

Table 4 summarizes the results of the seven regression runs. The one striking feature is the universally high R^2 for all maturities. This is hardly surprising because the explanatory variables are filtered estimates from these seven yield series. The interesting fact is about the increasing coefficients of the change in $\hat{\mu}_{t|t}$. In addition to what has been revealed by Figure 4, it suggests that short-term yields are locally far less responsive to the change in the central tendency factor.

The regression results for the filtered value of the LM-CTAR factor suggest a completely opposite relationship relating to maturity. The change in $\hat{X}_{t|t}$ causes short-term rates to react in a pronounced way, but its impact diminishes quickly when maturity goes up. Locally, the effect of the LM-CTAR factor is clearly confined to shorter-term yields. As to the extra AR(1) factor, it acts pretty much like the LM-CTAR factor, but the effect of maturity lingers on more and its impact on the change in interest rate drops at a much slower pace.

4 Conclusion

We propose a new autoregressive process with a local-momentum feature and establish its basic statistical properties. Using this local momentum autoregression model, we then construct an interest rate term structure model, devise a state-space estimation technique, and show that local momentum is indeed present in US Treasury interest rates. This paper adds to the vast literature of term structure of interest rates in a unique way. In contrast to the literature that mainly relies on deploying several first-order Markov processes to model interest rates in different ways, the

local-momentum autoregression model or its continuous-time counterpart offers an intuitive and parsimonious non-Markovian formulation that opens up many new possibilities.

Along the line of Hamilton and Wu (2012&2014), for example, one can work out the implications on identification and testing for a canonical formulation of a general lower-dimensional, say 3-factor, term structure model with one or two factors being replaced by the local-momentum autoregressive process where the actual Markovian dimension is actually much higher. One can also introduce observable macroeconomic factors into the model like in Ang and Piazzesi (2003) along with some latent factors in an additive fashion. Alternatively and perhaps more interestingly, macro factors can be incorporated into the local-momentum autoregression process directly such as adding them to the central tendency factor.

The local-momentum autoregression process exhibits interesting endogenous cycles and opens up a new way for modeling other economic series. Exogenous variables can also be built into the central tendency factor to allow for policy analysis of all kinds. Regimes such as in Hamilton (1988) can also be incorporated into the local-momentum model to give extra flexibility to allow for changes in volatility and others. All these are naturally subjects for future research.

References

- [1] At-Sahalia, Y. and R.L. Kimmel, 2010, Estimating Affine Multifactor Term Structure Models Using Closed-form Likelihood Expansions, *Journal of Financial Economics* 98, 113-144.
- [2] Ang, A. and M. Piazzesi, 2003, A No-arbitrage Vector Autoregression of Term Structure Dynamics with Macroeconomic and Latent Variables, *Journal of Monetary Economics* 50, 745-787.
- [3] Backus, D. and S. Zin, 1993, Long-memory Inflation Uncertainty: Evidence from the Term Structure of Interest Rates, *Journal of Money, Credit and Banking* 25, 681-700.
- [4] Balduzzi, P., S.R. Das and S. Foresi, 1998, The Central Tendency: A Second Factor in Bond Yields, *Review of Economics and Statistics* 80, 62-72.
- [5] Balvers, J.R. and Y. Wu, 2006, Momentum and Mean Reversion Across National Equity Markets, *Journal of Empirical Finance* 13, 24-48.
- [6] Chen, R. and J.S. Liu, 2000, Mixture Kalman Filters, *Journal of the Royal Statistical Society Series B* 62, 493-508.
- [7] Chen, R.-R., L. Scott, 1993, Maximum Likelihood Estimation for a Multifactor Equilibrium Model of the Term Structure of Interest Rates, *Journal of Fixed Income* 3, 1431.
- [8] Cox, J.C., J.E. Ingersoll, and S.A. Ross, 1985, A Theory of the Term Structure of Interest Rates, *Econometrica* 53, 385-408.

- [9] Dai, Q. and K.J. Singleton, 2000, Specification Analysis of Affine Term Structure models, *Journal of Finance* 55, 1943-1978.
- [10] Duffee, G.R., 2002, Term Premia and Interest Rate Forecasts in Affine Models, *Journal of Finance* 57, 405-443.
- [11] Duffie, D. and R. Kan, 1996, A Yield-Factor Model of Interest Rates, *Mathematical Finance* 6, 379-406.
- [12] Duan, J.-C. and A. Fulop, 2009, Estimating the Structural Credit Risk Model When Equity Prices Are Contaminated by Trading Noises, *Journal of Econometrics* 150, 288-296.
- [13] Duan, J.-C. and K. Jacobs, 2008, Is Long Memory Necessary? An Empirical Investigation of Nonnegative Interest Rate Processes, *Journal of Empirical Finance* 15, 567-581.
- [14] Hamilton, J.D., 1994, Time Series Analysis, Princeton University Press.
- [15] Hamilton, J.D., 1988, Rational-Expectations Econometric Analysis of Changes in Regime – An Investigation of the Term Structure of Interest Rates, *Journal of Economic Dynamics and Control* 12, 385-423.
- [16] Hamilton, J.D. and J.C. Wu, 2014, Testable Implications of Affine Term Structure Models, *Journal of Econometrics* 178, 231-242.
- [17] Hamilton, J.D. and J.C. Wu, 2012, Identification and Estimation of Gaussian Affine Term Structure Models, *Journal of Econometrics* 168, 315-331.
- [18] Hong, Y. and H. Li, 2005, Nonparametric Specification Testing for Continuous-time Models with Applications to Term Structure of Interest Rates, *Review of Financial Studies* 18, 3784.
- [19] Joslin, S., K.J. Singleton and H. Zhu, 2011, A New Perspective on Gaussian Dynamic Term Structure Models, *Review of Financial Studies* 24, 926-970.
- [20] Vasicek, O., 2013, General Equilibrium with Heterogeneous Participants and Discrete Consumption Times, *Journal of Financial Economics* 108, 608-614.
- [21] Vasicek, O., 1977, An Equilibrium Characterization of the Term Structure, *Journal of Financial Economics* 5, 177-188.
- [22] Wu, Y. and H. Zhang, 1996, Mean Reversion in Interest Rates: New Evidence from a Panel of OECD Countries, *Journal of Money, Credit and Banking* 28, 604-621.

Appendices

A. Propositions

Proposition 1 X_t as defined in equations (3)-(6) is strictly stationary and ergodic if and only if $\rho(\mathbf{B}) < 1$. Furthermore, X_t is strictly stationary and ergodic under either of the two sets of sufficiency conditions: (1) $\kappa_x > 0$, $\omega \geq 0$ and $\kappa_x + \omega(1 - b_1) \leq 1$; (2) $\kappa_x > -2\omega(1 - b_1)$, $\omega < 0$ and $\kappa_x + \omega(1 - b_1) \leq 1$.

Proof: First note that X_t can be expressed as an ARMA(n, ∞) process with its MA component being deduced from inverting, μ_t , which is a stationary AR(1) process. The MA coefficients are thus absolutely summable. The first statement about strict stationarity and ergodicity is a standard result for the ARMA(n, ∞) process, and we will skip the proof. We now derive the two sufficiency conditions. Spectral radius is the largest absolute value of a matrix's eigenvalues, and the eigenvalues of \mathbf{B} must satisfy the following equation (Proposition 1.1; Hamilton (1994), page 10):

$$\lambda^n = [1 - \kappa_x - \omega(1 - b_1)] \lambda^{n-1} + \omega b_2 \lambda^{n-2} + \dots + \omega b_{n-1} \lambda + \omega b_n.$$

Case 1: The first set of sufficiency conditions: $\kappa_x > 0$, $\omega \geq 0$ and $\kappa_x + \omega(1 - b_1) \leq 1$.

Under these conditions, we can deduce

$$\begin{aligned} |\lambda|^n = |\lambda^n| &= |[1 - \kappa_x - \omega(1 - b_1)] \lambda^{n-1} + \omega b_2 \lambda^{n-2} + \dots + \omega b_{n-1} \lambda + \omega b_n| \\ &\leq [1 - \kappa_x - \omega(1 - b_1)] |\lambda^{n-1}| + \omega b_2 |\lambda^{n-2}| + \dots + \omega b_{n-1} |\lambda| + \omega b_n \\ &= [1 - \kappa_x - \omega(1 - b_1)] |\lambda|^{n-1} + \omega b_2 |\lambda|^{n-2} + \dots + \omega b_{n-1} |\lambda| + \omega b_n \end{aligned}$$

Substitute $\rho(\mathbf{B})$ into the above inequality to obtain

$$\rho(\mathbf{B})^n \leq [1 - \kappa_x - \omega(1 - b_1)] \rho(\mathbf{B})^{n-1} + \omega b_2 \rho(\mathbf{B})^{n-2} + \dots + \omega b_{n-1} \rho(\mathbf{B}) + \omega b_n.$$

Suppose $\rho(\mathbf{B}) \geq 1$. Divide both sides by $\rho(\mathbf{B})^n$ to yield

$$\begin{aligned} 1 &\leq [1 - \kappa_x - \omega(1 - b_1)] \rho(\mathbf{B})^{-1} + \omega b_2 \rho(\mathbf{B})^{-2} + \dots + \omega b_{n-1} \rho(\mathbf{B})^{-n+1} + \omega b_n \rho(\mathbf{B})^{-n} \\ &\leq [1 - \kappa_x - \omega(1 - b_1)] + \omega b_2 + \dots + \omega b_{n-1} + \omega b_n \\ &= 1 - \kappa_x \\ &< 1. \end{aligned}$$

That is obviously a contradiction. Thus, $\rho(\mathbf{B}) < 1$, and the conditions are sufficient to ensure strict stationarity of X_t .

Case 2: The second set of sufficiency conditions: $\kappa_x > -2\omega(1 - b_1)$, $\omega < 0$ and $\kappa_x + \omega(1 - b_1) \leq 1$.

Under these conditions, we can deduce

$$\begin{aligned} |\lambda|^n = |\lambda^n| &= |[1 - \kappa_x - \omega(1 - b_1)] \lambda^{n-1} + \omega b_2 \lambda^{n-2} + \dots + \omega b_{n-1} \lambda + \omega b_n| \\ &\leq [1 - \kappa_x - \omega(1 - b_1)] |\lambda^{n-1}| - \omega b_2 |\lambda^{n-2}| - \dots - \omega b_{n-1} |\lambda| - \omega b_n \\ &= [1 - \kappa_x - \omega(1 - b_1)] |\lambda|^{n-1} - \omega b_2 |\lambda|^{n-2} - \dots - \omega b_{n-1} |\lambda| - \omega b_n \end{aligned}$$

Substitute $\rho(\mathbf{B})$ into the above inequality to obtain

$$\rho(\mathbf{B})^n \leq [1 - \kappa_x - \omega(1 - b_1)]\rho(\mathbf{B})^{n-1} - \omega b_2 \rho(\mathbf{B})^{n-2} - \dots - \omega b_{n-1} \rho(\mathbf{B}) - \omega b_n.$$

Suppose $\rho(\mathbf{B}) \geq 1$. Divide both sides by $\rho(\mathbf{B})^n$ to yield

$$\begin{aligned} 1 &\leq [1 - \kappa_x - \omega(1 - b_1)]\rho(\mathbf{B})^{-1} - \omega b_2 \rho(\mathbf{B})^{-2} - \dots - \omega b_{n-1} \rho(\mathbf{B})^{-n+1} - \omega b_n \rho(\mathbf{B})^{-n} \\ &\leq [1 - \kappa_x - \omega(1 - b_1)] - \omega b_2 - \dots - \omega b_{n-1} - \omega b_n \\ &= 1 - \kappa_x - 2\omega(1 - b_1) \\ &< 1. \end{aligned}$$

It is obviously a contradiction. Thus, $\rho(\mathbf{B}) < 1$, and the conditions are sufficient to ensure strict stationarity and ergodicity of X_t .

Proposition 2 *If $\kappa_x = 0$, then X_t as defined in equations (3)-(6) has a unit root regardless of the value of ω , and ΔX_t is strictly stationary and ergodic if $\rho(\mathbf{F}) < 1$.*

Proof: When $\kappa_x = 0$, equation (3) can be written as

$$\begin{aligned} X_t &= X_{t-1} + \omega (\bar{X}_{(t-1)|n} - X_{t-1}) + \sigma_x \varepsilon_t \\ &= [1 + \omega(b_1 - 1)]X_{t-1} + \omega b_2 X_{t-2} + \dots + \omega b_n X_{t-n} + \sigma_x \varepsilon_t \end{aligned}$$

Clearly, the above time series has a unit root because $\sum_{i=1}^n b_i = 1$.

When $\kappa_x = 0$, we can also express equation (3) as

$$\begin{aligned} \Delta X_t &= \omega (\bar{X}_{(t-1)|n} - X_{t-1}) + \sigma_x \varepsilon_t \\ &= \omega(b_1 - 1)\Delta X_{t-1} + \omega(b_1 + b_2 - 1)\Delta X_{t-2} + \dots \\ &\quad + \omega(b_1 + b_2 + \dots + b_{n-1} - 1)\Delta X_{t-n+1} + \sigma_x \varepsilon_t \end{aligned}$$

There is no term beyond ΔX_{t-n+1} again because $\sum_{i=1}^n b_i = 1$.

Since the differenced series ΔX_t , stated in a matrix-vector form, is characterized by matrix \mathbf{F} . By the standard result, the differenced series is strictly stationary and ergodic when the spectral radius of matrix \mathbf{F} is strictly less than 1.

B. Innovation terms under the risk-neutral measure

Consider the following moment generating function: for $t \leq s \leq T$,

$$\begin{aligned}
& E_{s-1}^{Q_{t,T}} (e^{a\varepsilon_s}) \\
&= E_{s-1}^P \left(e^{a\varepsilon_s} \frac{M_{t,s}}{M_{t,s-1}} \right) \\
&= \frac{\alpha(t,s)}{\alpha(t,s-1)} \times \\
& \quad E_{s-1}^P \{ \exp [(a + \lambda_0 + \lambda_1 X_{s-1}) \varepsilon_s + (\psi_{\mu 0} + \psi_{\mu 1} \mu_{s-1}) \varepsilon_s + (\psi_{v 0} + \psi_{v 1} v_{s-1}) \xi_s] \} \\
&= \frac{\alpha(t,s)}{\alpha(t,s-1)} \exp \left[\frac{(\lambda_0 + \lambda_1 X_{s-1})^2}{2} + \frac{(\psi_{\mu 0} + \psi_{\mu 1} \mu_{s-1})^2}{2} + \frac{(\psi_{v 0} + \psi_{v 1} v_{s-1})^2}{2} \right] \\
& \quad \times \exp \left[a(\lambda_0 + \lambda_1 X_{s-1}) + \frac{a^2}{2} \right]
\end{aligned}$$

Setting $a = 0$ and note that $E_{s-1}^{Q_{t,T}} (1) = 1$, we have

$$E_{s-1}^{Q_{t,T}} (e^{a\varepsilon_s}) = \exp \left[a(\lambda_0 + \lambda_1 X_{s-1}) + \frac{a^2}{2} \right].$$

This implies that ε_s is a normal random variable under measure $Q_{t,T}$ with mean $\lambda_0 + \lambda_1 X_{s-1}$ and variance 1. Notice that the result does not depend on either t or T . Thus, we can drop the subscript and use Q . Define $\varepsilon_t^Q = \varepsilon_t - \lambda_0 - \lambda_1 X_{t-1}$, which is a standard normal random variable under measure Q . Applying the above relationship to the period of $[t, t+1]$ and recalling that $\alpha(t,t) = 1$, one period interest rate in equilibrium must obey

$$r_t = \ln[\alpha(t, t+1)] + \frac{(\lambda_0 + \lambda_1 X_{s-1})^2}{2} + \frac{(\psi_{\mu 0} + \psi_{\mu 1} \mu_{s-1})^2}{2} + \frac{(\psi_{v 0} + \psi_{v 1} v_{s-1})^2}{2}.$$

Recall that $\alpha(t, t+1)$ is measurable with respect to the information set at time t . One is free to specify either r_t or $\alpha(t, t+1)$ as a function of state variables at time t , but not both.

The argument also applies to μ_s and ξ_s . Define $\varepsilon_t^Q = \varepsilon_t - \psi_{\mu 0} - \psi_{\mu 1} \mu_{t-1}$ and $\xi_t^Q = \xi_t - \psi_{v 0} - \psi_{v 1} v_{t-1}$ which are also standard normal random variables under measure Q . Finally, ε_t^Q , ε_t^Q and ξ_t^Q are independent under measure Q can be established by showing that $E_{s-1}^{Q_{t,T}} (e^{a(\varepsilon_s + \varepsilon_s + \xi_s)}) = E_{s-1}^{Q_{t,T}} (e^{a\varepsilon_s}) E_{s-1}^{Q_{t,T}} (e^{a\varepsilon_s}) E_{s-1}^{Q_{t,T}} (e^{a\xi_s})$.

C. The Kalman filter implementation

Denote the covariance matrix of the measurement errors by Ω , the column vector of observed rates by \mathbf{R}_t , and the coefficient vector and matrix by Φ_1 and Φ_2 . We can apply the standard

Kalman filtering results to equation (18) to obtain the predicted mean and variance of the state variables:

$$\hat{\mathbf{X}}_{t|t-1}^* = \mathbf{S}^{-1}\mathbf{C} + \mathbf{S}^{-1}\mathbf{D}\hat{\mathbf{X}}_{t-1|t-1}^* \quad (32)$$

$$\hat{\mathbf{P}}_{t|t-1} = \mathbf{S}^{-1}\mathbf{D}\hat{\mathbf{P}}_{t-1|t-1}(\mathbf{S}^{-1}\mathbf{D})' + \mathbf{S}^{-1}\mathbf{U}(\mathbf{S}^{-1})' \quad (33)$$

Applying the measurement equations in (30), the filtered mean and variance become

$$\hat{\mathbf{X}}_{t|t}^* = \hat{\mathbf{X}}_{t|t-1}^* + \hat{\mathbf{K}}_t[\mathbf{R}_t - \Phi_1 - \Phi_2\hat{\mathbf{X}}_{t|t-1}^*] \quad (34)$$

$$\hat{\mathbf{P}}_{t|t} = (\mathbf{I}_{(n+1)\times(n+1)} - \hat{\mathbf{K}}_t\Phi_2)\hat{\mathbf{P}}_{t|t-1} \quad (35)$$

where $\hat{\mathbf{K}}_t = \hat{\mathbf{P}}_{t|t-1}\Phi_2'(\Phi_2\hat{\mathbf{P}}_{t|t-1}\Phi_2' + \Omega)^{-1}$.

To set the initial values, we let $\hat{\mathbf{X}}_{0|0}^* = [x_0^{(s)}, x_0^{(s)}, \dots, x_0^{(s)}, x_0^{(l)} - \bar{\mu}, 0]'$ where $x_0^{(s)}$ is the first observed interest rate of the shortest maturity and $x_0^{(l)}$ is the rate of the longest maturity at the same time. In our case, both turn out to be the three-month rate. The rationale is that X_t is meant as the key component of the short-term base rate whereas the central tendency should be more closely aligned with the longest-term rate. In fact, we have set the initial state of the entire lagged vector of the LM-CTAR process to $x_0^{(s)}$. The initial value for the second factor, $\mu_t - \bar{\mu}$, follows by assuming $\mu_0 = x_0^{(l)}$, and the third factor is initialized at zero. Consistent with this assumption, we set $\hat{\mathbf{P}}_{0|0} = \mathbf{0}$.

Figure 1. The 3-month US Treasury yields (continuously compounded) daily from January 4, 1954 to December 31, 2013.

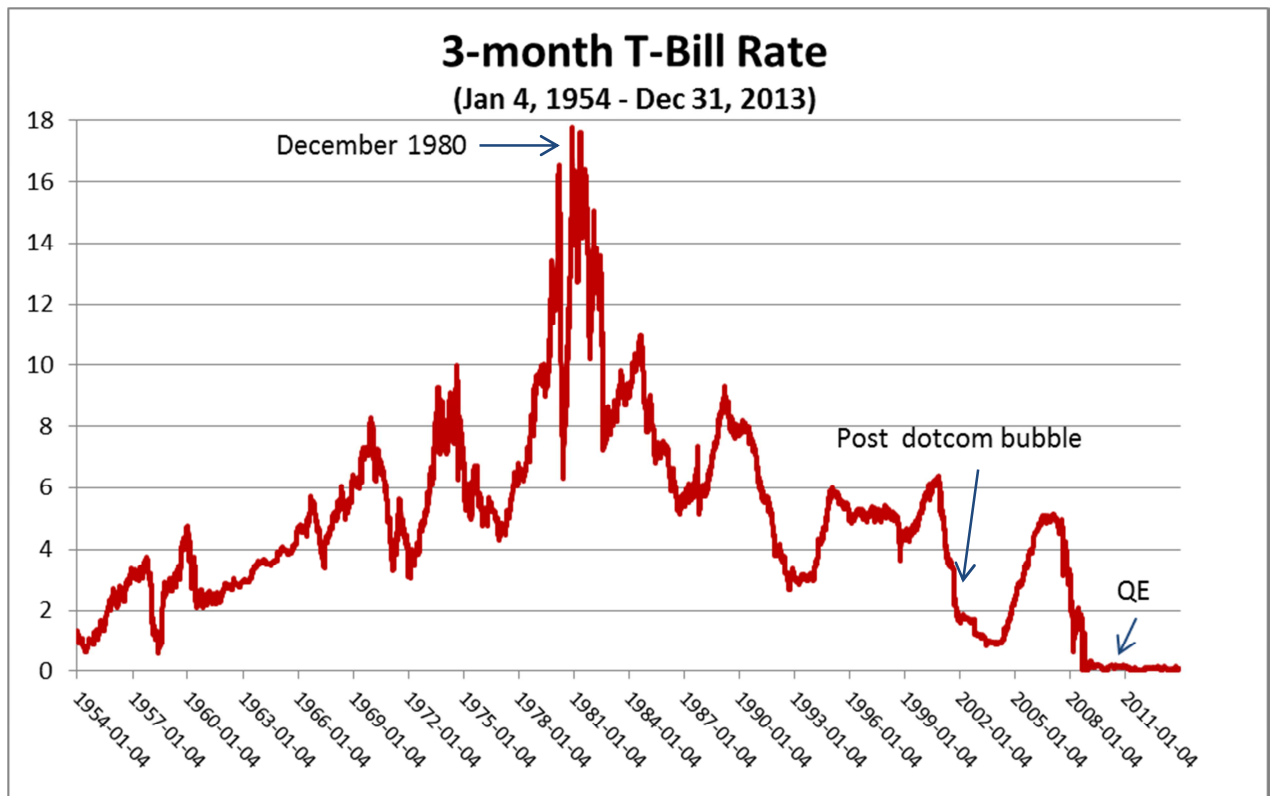


Figure 2a. A simulated sample path for the LM-AR with 5 lags. The parameters are: $\bar{\mu} = 0, \kappa_x = 0.001, \omega = -0.5, \sigma_x = 0.002$ and $\sigma_\mu = 0$.

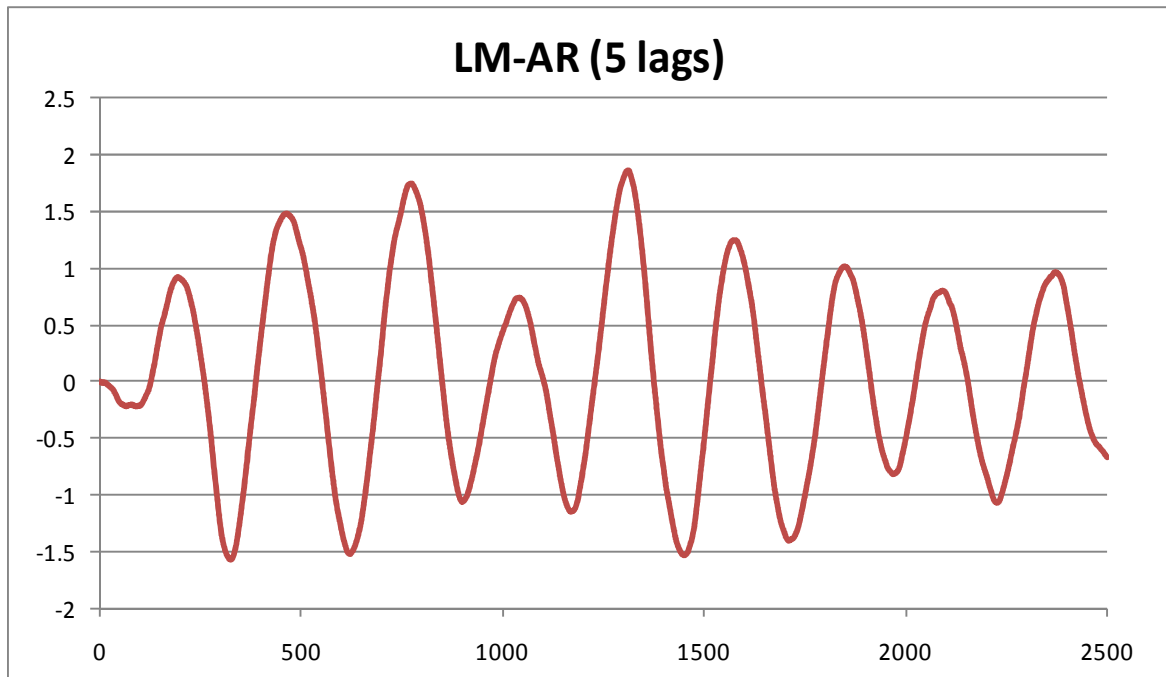


Figure 2b. A simulated sample path for the LM-AR with 10 lags. The parameters are: $\bar{\mu} = 0, \kappa_x = 0.001, \omega = -0.2, \sigma_x = 0.02$ and $\sigma_\mu = 0$.

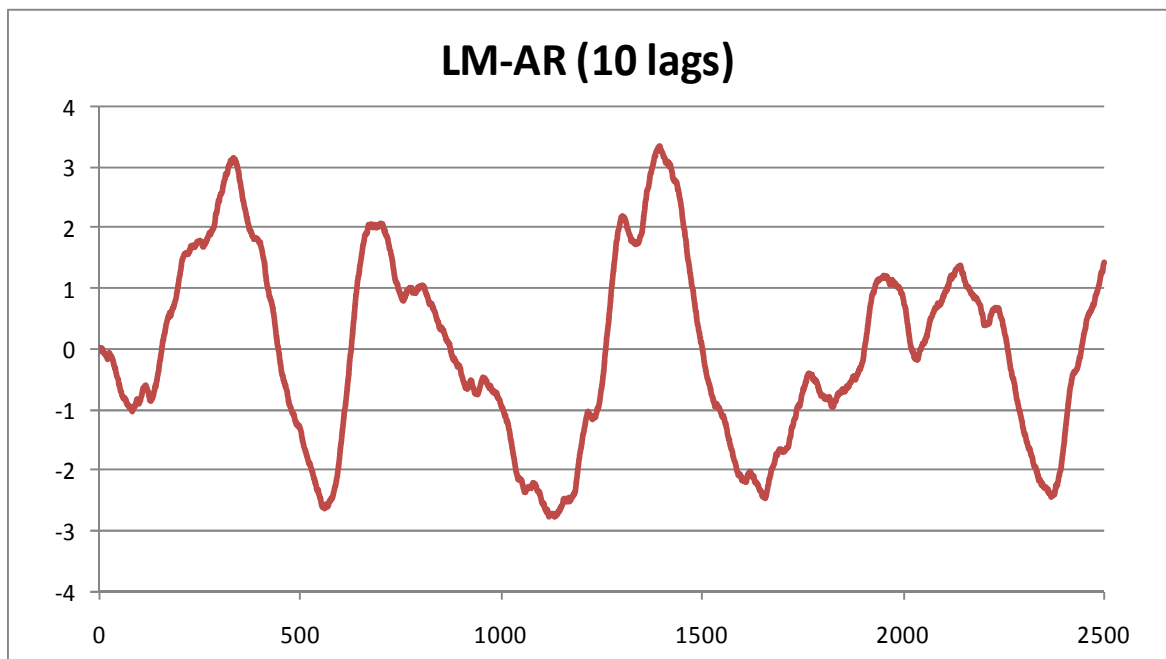


Figure 3a. A 3-factor Gaussian term structure model built on the LM-CTAR process is fitted to the US treasury constant maturity yields (continuously compounded) of seven maturities (1 month, 3 months, 6 months, 1 year, 5 years, 10 years and 20 years). The filtered estimate of the LM-CTAR and central tendency components are plotted along with the 3-month rate, weekly from January 4, 1954 to December 31, 2013. The vertical axis is in percentage points.

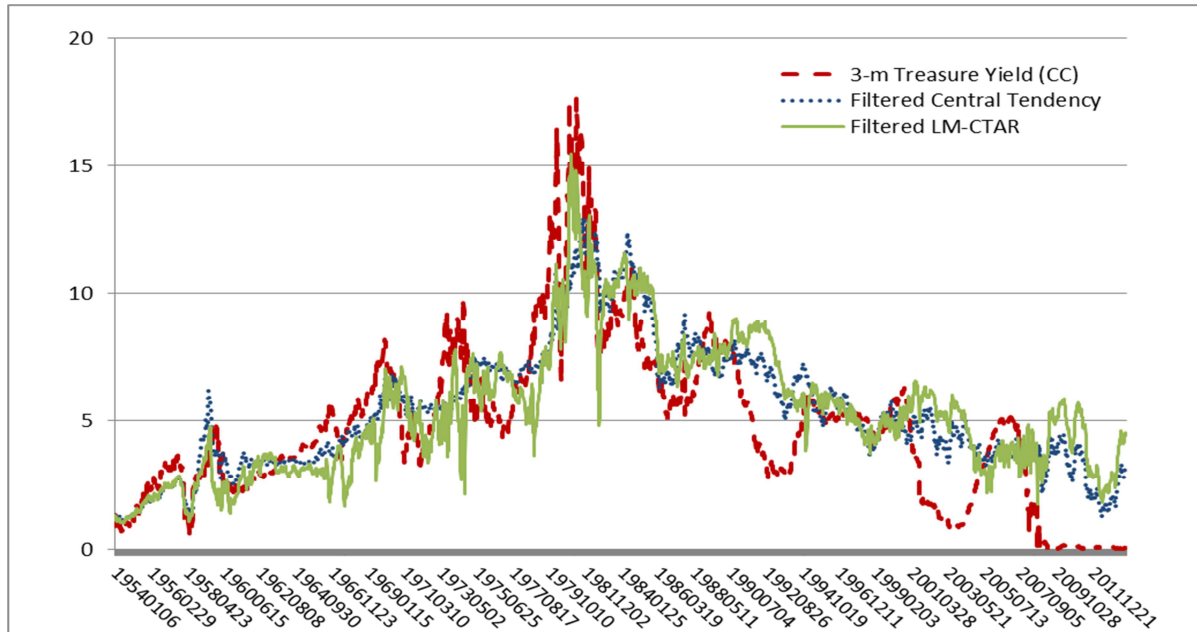


Figure 3b. The difference in the filtered LM-CTAR and CTAR factors from two 3-factor term structure models over the sample period from January 4, 1954 to December 31, 2013.

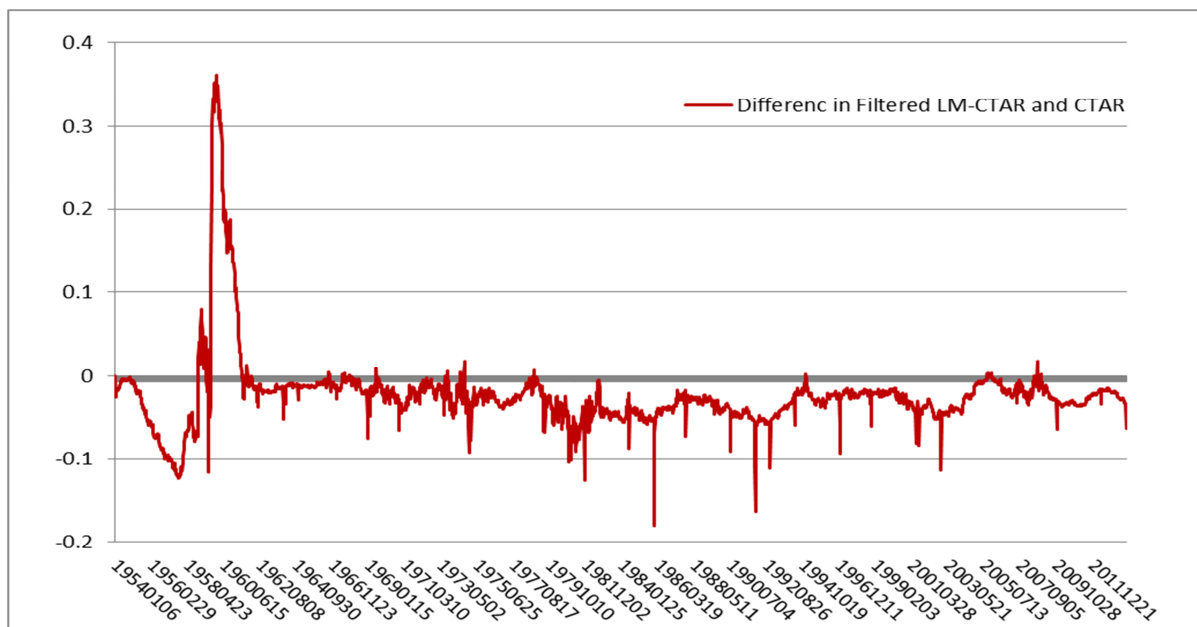


Figure 4. The filtered central tendency estimates corresponding to two versions of the 3-factor Gaussian term structure model built on, respectively, the CTAR and LM-CTAR processes from January 4, 1954 to December 31, 2013 on a weekly frequency. Also plotted is the 20-year US Treasury yields (continuously compounded) when they were available. The vertical axis is in percentage points.

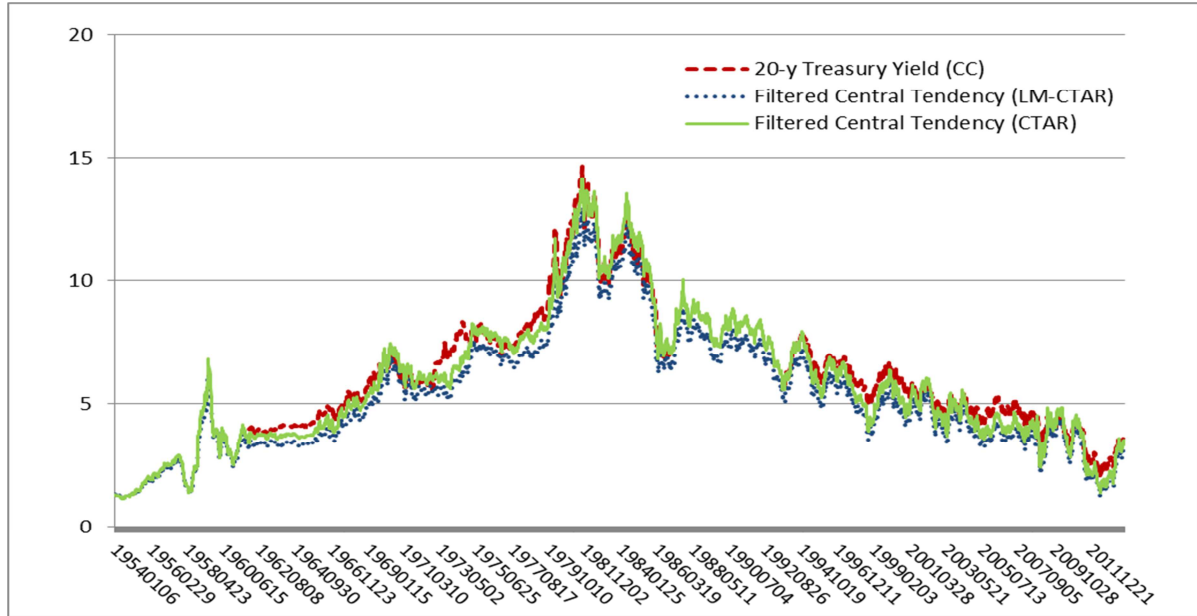


Table 1. Summary statistics for the US treasury constant-maturity yields (continuously compounded) from January 4, 1954 to December 31, 2013, sampled once a week on Wednesday.

Maturity	# of Points	Mean	Median	Standard Deviation
1 month	641	0.015144	0.009636	0.016606
3 months	3080	0.047902	0.047220	0.031239
6 months	2831	0.052203	0.051244	0.031604
1 year	2451	0.057484	0.055555	0.031759
5 years	2674	0.060156	0.058458	0.027835
10 years	2674	0.063239	0.060107	0.025515
20 years	2327	0.063808	0.058458	0.024891
	Skewness	Excess Kurtosis	Maximum	Minimum
1 month	0.942746	-0.397540	0.052626	0
3 months	0.885921	1.416064	0.176280	0
6 months	0.718757	1.086112	0.167857	0.000304
1 year	0.484282	0.759180	0.166050	0.000811
5 years	0.521315	0.396290	0.150745	0.005584
10 years	0.721610	0.391892	0.147040	0.014199
20 years	1.016944	0.596012	0.146522	0.020880

Table 2. Estimation results using the 3-month US treasury constant-maturity yields (continuously compounded) from January 4, 1954 to December 31, 2013, sampled once a week on Wednesday. Estimation does not factor in the term structure effect. The values in brackets are standard errors. The moving level in the LM-AR and LM-CTAR processes uses a 7-week average.

Parameter	AR(1)	CTAR	LM-AR	LM-CTAR
$\bar{\mu}$	0.046387 (0.024706)	0.046832 (0.019745)	0.047097 (0.019325)	0.045395 (0.022350)
k_x	0.002632 (0.000941)	0.004181 (0.001324)	0.003381 (0.000977)	0.105702 (0.025165)
ω			-0.071005 (0.004154)	-0.075354 (0.007143)
σ_x	0.002328 (0.000008)	0.002189 (0.000019)	0.002314 (0.000008)	0.002177 (0.000041)
k_μ		0.295870 (0.039138)		0.003284 (0.001234)
σ_μ		0.134946 (0.048605)		0.002511 (0.000132)
Log-likelihood	14231.12	14247.24	14249.08	14253.61
$\rho(\mathbf{B})$			0.995690	0.825882
Data points	3130			
Missing data	50			

Table 3. Estimation results for two 3-factor term structure models using the US Treasury constant-maturity yields (continuously compounded) of seven maturities. The data are from January 4, 1954 to December 31, 2013, sampled once a week on Wednesday. The values in brackets are standard errors. The moving level in the LM-CTAR process uses a 28-week average.

	CTAR+AR(1)	LM-CTAR+AR(1)
Physical process parameters		
$\bar{\mu}$	0.103950 (0.002614)	0.095287 (0.002504)
k_x	0.072262 (0.000993)	0.080571 (0.001112)
ω		-0.007134 (0.000690)
σ_x	0.002649 (0.000028)	0.002591 (0.000027)
k_μ	0.000189 (0.000011)	0.000188 (0.000011)
σ_μ	0.001440 (0.000017)	0.001317 (0.000018)
k_ν	0.008227 (0.000087)	0.008176 (0.000086)
σ_ν	0.002058 (0.000019)	0.002050 (0.000019)
Measurement errors		
1 month	0.000892 (0.000039)	0.000886 (0.000039)
3 months	0.000576 (0.000013)	0.000579 (0.000013)
6 months	0.000746 (0.000009)	0.000742 (0.000009)
1 year	0.000628 (0.000010)	0.000636 (0.000010)
5 years	0.001767 (0.000022)	0.001770 (0.000022)
10 years	0.000174 (0.000032)	0.000174 (0.000031)
20 years	0.002362 (0.000030)	0.002365 (0.000030)
Log-likelihood	96976.40	97012.06
$\rho(\mathbf{B})$		0.891281
Data points	3130	
Missing data	50	

Table 4. Changes in yields (continuously compounded) of seven maturities on changes in the three filtered factors of the local momentum term structure model. The data are from January 4, 1954 to December 31, 2013, sampled once a week on Wednesday. The values in brackets are standard errors.

Yield Change	Intercept	$\Delta\hat{\mu}_{t t}$	$\Delta\hat{X}_{t t}$	$\Delta\hat{V}_{t t}$	R^2
$\Delta r_t(1m)$	0.000109 (0.000083)	0.220220 (0.007196)	0.991148 (0.003920)	0.976043 (0.008784)	0.9148
$\Delta r_t(3m)$	0.000017 (0.000009)	0.351476 (0.000573)	0.732278 (0.000194)	0.952301 (0.000265)	0.9850
$\Delta r_t(6m)$	0.000067 0.000017	0.583060 0.000982	0.413108 0.000325	0.944362 0.000449	0.9511
$\Delta r_t(1y)$	-0.000042 (0.000018)	0.768468 (0.001023)	0.190847 (0.000319)	0.881598 (0.000424)	0.9590
$\Delta r_t(5y)$	-0.000352 (0.000017)	1.030643 (0.001013)	0.039280 (0.000327)	0.384023 (0.000454)	0.9181
$\Delta r_t(10y)$	-0.000011 (0.000001)	1.035115 (0.000074)	0.026693 (0.000024)	0.227487 (0.000033)	0.9995
$\Delta r_t(20y)$	0.000028 (0.000018)	0.902467 (0.001052)	0.038297 (0.000317)	0.191439 (0.000440)	0.9004