US and Canadian term structures of interest rates: A forecasting comparison

Van Dellen, S., Juneja, J.A., Nowman, K.B. and Zhao, G.

This is an electronic version of a paper presented atthe *23rd Annual Conference of the Multinational Finance Society,* Stockholm, Sweden, 26 to 29 June 2016.

US and Canadian Term Structures of Interest Rates: A Forecasting Comparison

Januj A. Juneja[a]
jjuneja@mail.sdsu.edu

Khalid B. Nowman[b*]
nowmank@westminster.ac.uk

Stefan van Dellen[b]
S.Vandellen@westminster.ac.uk

Gang Zhao[c]
gz249@bath.ac.uk

This Version: January 2016

Affiliations:

[a] College of Business Administration, San Diego State University, 5500 Campanile Drive, San Diego, CA 92182-8230, United States of America.

[b] Department of Accounting Finance and Governance, Westminster Business School, University of Westminster, 35 Marylebone Road, London NW1 5LS, United Kingdom.

[c] Bath School of Management, University of Bath, Claverton Down, Bath, Somerset BA2 7AY, United Kingdom.

Abstract

This paper provides empirical evidence for the US and Canadian yield curves using a one- and two-factor Generalised Vasicek model, using a data set comprised of daily panel data over the period between 2003 and 2011, which includes the recent global financial crisis. The two-factor model is found to have a good fit for both the US and Canadian yield curves. We also compare the forecasting performance of the term structure model with those from ARIMA, ARFIMA and Nelson-Siegel models. We find that for Canada the Nelson-Siegel model dominates, while for the US the ARFIMA model has a satisfactory performance.

Key Words: Yield Curve; Kalman Filter; Nelson-Siegel; ARFIMA

JEL Code: E43; C58; G17

[*] **Corresponding author:** Prof K. B. Nowman, Westminster Business School, 35 Marylebone Road, London NW1 5LS, United Kingdom. Email: nowmank@wmin.ac.uk. Tel: +44 (0) 207 911 5000.

1. Introduction

The modelling of the term structure of interest rates is of great importance given its use in the fixed income markets and other areas of finance. One important approach in terms of estimating term structure models uses panel data and Kalman filtering methods, where examples of this include Jegadeesh and Pennacchi (1996), Geyer and Pichler (1999), Duan and Simonato (1999), De Jong (2000), Chen and Scott (2003), Chatterjee (2004), Driessen, *et al*. (2005) and O'Sullivan (2007). A general multi-factor Gaussian term structure model was developed by Babbs and Nowman (1999) and is a subclass of the general model outlined by Langetieg (1980). Babbs and Nowman (1999) derived the general bond price and estimated their model using US data and the Kalman filter approach and found that the model provided a good description of the yield curve. In a more recent application, Nowman (2010) demonstrated that the model provided a good description of UK and Euro yield curves.

In this paper, we apply the Generalised Vasicek (GVT) model to the Canadian and US markets over the period 2000-2011, which includes the recent global financial crisis period. We first carry out a principal component analysis of the data, and then estimate one- and two-factor versions of the model. We find evidence that the two-factor model provides a good description of the Canadian and US yield curves. We also carry out a test of the forecasting performance of the one- and two-factor models and compare the forecast results with from the Nelson and Siegel (2007) three-factor model. The Nelson-Siegel (*hereafter,* NS) model decomposes the term structure of interest rates into three factors, namely the level, slope and curvature factors. The NS three-factor model is popular among practitioners as it uses a flexible and smooth parametric function to replicate the term structure at any given time (Svensson, 1995; Bank for International Settlements, 2005; Gurkaynak, *et al.*, 2007; Christensen, *et al.*, 2011; Sekkel, 2011). Empirically, Diebold and Li (2006), De Pooter (2007) and Yu and Zivot

(2011) all should that NS class models provide a good fit the real term structure in terms of both the in-sample and out-of-sample dynamics.

We also compare the forecasting performance of the above models with that for the ARIMA (Box and Jenkins, 1976) and ARFIMA (Granger and Joyeux, 1980; Granger, 1980, 1981; Hosking, 1981) models, which have been developed for and applied in forecasting time series. We find that for Canada, the NS models dominate, while in the case of the US the ARFIMA model has a satisfactory performance.

The remainder of this paper is organised as follows: Section 2 presents the methodologies for the GVT, as well as the state space model and the Kalman filter, and finally the NS, ARIMA and ARFIMA models. Section 3 presents the data, Section 4 reports and discusses the empirical results, and Section 5 presents the forecast results. Concluding remarks are provided in Section 6.

2. Methodology

2.1 The Generalised Vasicek Model

Babbs and Nowman (1999) assume that the spot interest rate (*r*) is given by

$$r(t) = \mu(t) - \sum_{j=1}^{J} X_j(t) \tag{1}$$

where μ is the long-run average rate and $X_1(t),\ldots,X_j(t)$ represent the current effect of *J* streams of economic 'news'. Babbs and Nowman (1999) interpreted the 'news' streams as rumours in the financial markets and short- and long-term economic 'news', both of which affect the yield curve. Examples of economic 'news' may include 'rumours' of interest rate decisions from the Federal Open Market Committee as well as monthly and quarterly economic statistics news. The arrival of each type of 'news' is modelled by the innovations of Brownian motions, which may be correlated, while the impact of a piece of 'news' dies away exponentially as the time since it was received increases. In equation form this is expressed as

$$dX_j = -\xi_j X_j dt + c_j dW_j \tag{2}$$

where each ξ_j and c_j are mean reversion and diffusion coefficients, and $W_1,\ldots,W_J$ are standard Brownian motions with correlations $\rho_{jk} : j,k = 1,\ldots,J$. Equation (2) can equivalently be expressed as

$$dX_j = -\xi_j X_j dt + \sum_{q=1}^{Q} \kappa_{jq} dZ_q \qquad \left(Q \le J\right) \tag{3}$$

where $Z_1,\ldots,Z_Q$ are independent standard Brownian motions and:

$$\sum_{q=1}^{Q} \kappa_{jq}\kappa_{kq} = \rho_{jk} c_j c_k \tag{4}$$

In the case where the long run level $\left(\mu\right)$, the mean-reversion speeds (ξ_j), the diffusion coefficients (κ_{jq}) and the market price of risk processes (θ_q) are all constant, the key pricing formula for a pure discount bond was derived by Babbs and Nowman (1999) and is given by

$$B\left(M,t\right) = \exp\left\{-\tau\left[R\left(\infty\right) - w\left(\tau\right) - \sum_{j=1}^{J} H\left(\xi_j \tau\right) X_j\left(t\right)\right]\right\} \tag{5}$$

with

$$R\left(\infty\right) = \mu + \sum_{q=1}^{Q}\theta_q \sum_{j=1}^{J}\frac{\kappa_{jq}}{\xi_j} - \frac{1}{2}\sum_{q=1}^{Q}\left(\sum_{j=1}^{J}\frac{\kappa_{jq}}{\xi_j}\right)^2 \tag{6}$$

and

$$w(\tau)=\sum_{j=1}^{J}H(\xi_j\tau)\left[\sum_{q=1}^{Q}\theta_j\frac{\kappa_{jq}}{\xi_j}-\sum_{q=1}^{Q}\sum_{i=1}^{J}\frac{\kappa_{iq}\kappa_{jq}}{\xi_i\xi_j}\right]+\dots$$
$$\dots\frac{1}{2}\sum_{i=1}^{J}\sum_{j=1}^{J}H((\xi_i+\xi_j)\tau)\sum_{q=1}^{Q}\frac{\kappa_{iq}\kappa_{jq}}{\xi_i\xi_j} \quad (7)$$

where

$$\tau\equiv M-t \quad (8)$$

and

$$H(x)=\frac{1-e^x}{x} \quad (9)$$

In the special case of a one-factor model, the bond pricing formula above reduces to the well-known model developed by Vasicek (1977).

2.2 *Estimation Method for the Generalised Vasicek Model*

We begin by presenting the state space model formulation of the term structure model and Kalman filter. The theoretical yield curve is given by

$$R(t+\tau,t)\equiv-\log B(t+\tau,t)/\tau=A_0(\tau)-A_1(\tau)'X_t \quad (10)$$

where $A_0(\tau)=R(\infty)-w(\tau)$ and $A_1(\tau)=H(\xi_j\tau)$ is a $J\times1$ vector (where the superscript denotes transpose. The scalar $A_0(\tau)$ and the vector $A_1(\tau)$ are functions of the time to maturity τ and the parameters of the model. We have N observed interest rates at time t_k, for $k=1,2,\dots,n$, which are denoted as $R_k=(R_{1k},\dots,R_{Nk})$, where $R_{ik}=-\log B(t_k+\tau_i,t_k)/\tau_i$.

We assume that measurement errors in the interest rates are additive and normally distributed. The measurement equation is the given by

$$R_k=d(\psi)+Z(\psi)X_k+\varepsilon_k\ ;\ R_k=d(\psi)+Z(\psi)X_k+\varepsilon_k \quad (11)$$

where ψ contains the unknown parameters of the mode including the parameters from the distribution of the measurement errors. The $i'th$ row of the matrices $d(N\times 1)$ and $Z(N\times J)$ are given by $A_0(\tau;\psi)$ and $-A_1(t;\psi)'$, respectively. The error terms (ε_k) are measurement errors to allow for noise in the sampling process of the data. Following Babbs and Nowman (1999), the measurement errors variances are assumed to be $H = h_1,\dots,h_N$ along the diagonal. The transition equation is the exact discrete-time distribution of the state variables

$$X_k = \Phi(\psi) X_{k-1} + \eta_k \tag{12}$$

where $\Phi(\psi) = e^{-\xi_J(t_k - t_{k-1})}$. The error term η_k is normally distributed with $E[\eta_k] = 0$ and $Cov[\eta_k] = V(\psi)$, where for a definition of V see Bergstrom (1984, Theorem 3). The measurement and transition equations represent the state space formulation of our model. The Kalman filter algorithm and the exactly likelihood function are now presented.

Let $\hat{X}_{k/k-1}$ and $\hat{X}_k$ denote the optimal estimator (in a mean square error sense) of the unknown state vector X_k based on the available information (i.e. the observed interest rates) up to time t_{k-1} and t_k, respectively. The optimal estimator is the condition mean of X_k in both cases, denoted $E_{k-1}[\cdot]$ and $E_k[\cdot]$, respectively. The prediction step is given by

$$\hat{X}_{k/k-1} = E_{k-1}(X_k) = \Phi \hat{X}_{k-1} \tag{13}$$

with the mean square error (MSE) matrix

$$\Sigma_{k/k-1} = E_{k-1}\left[\left(X_k - \hat{X}_{k/k-1}\right)\left(X_k - \hat{X}_{k/k-1}\right)'\right] = \Phi \Sigma_{k-1} \Phi' + V \tag{14}$$

In the update step, the addition information given by R_k is used to obtain a more precise estimator of X_k

$$\hat{X}_k = E_k\left(X_k\right) = \hat{X}_{k/k-1} + \Sigma_{k/k-1} Z' F_k^{-1} v_k \tag{15}$$

and

$$\begin{aligned}\Sigma_k &= E_k\left[\left(X_k - \hat{X}_k\right)\left(X_k - \hat{X}_k\right)'\right] = \Sigma_{k/k-1} - \Sigma_{k/k-1} Z' F_k^{-1} Z \Sigma_{k/k-1} \\ &= \left(\Sigma_{k/k-1}^{-1} + Z' H^{-1} Z\right)^{-1}\end{aligned} \tag{16}$$

where

$$v_k = R_k - \left(d + Z\hat{X}_{k/k-1}\right) \tag{17}$$

and

$$F_k = Z\Sigma_{k/k-1} Z' + H \tag{18}$$

This new estimate of X_k is called the *filtered* estimate. The log-likelihood function is given by (apart from a constant)

$$\log L\left(R_1, \ldots, R_n; \psi\right) = -\frac{1}{2}\sum_{i=1}^{n} \log\left|F_k\right| - \frac{1}{2}\sum_{k=1}^{n} v_k' F_k^{-1} v_k \tag{19}$$

where v_k and F_k are given by equations (17) and (18), respectively. We can also use the formulae for computing the inverse and determinant of F_k given by

$$F_k^{-1} = H^{-1} Z \left(\Sigma_{k/k-1}^{-1} + Z' H^{-1} Z\right)^{-1} Z' H^{-1}$$

and

$$\left|F_k\right| = \left|H\right| \cdot \left|\Sigma_{k/k-1}\right| \cdot \left|\Sigma_{k/k-1}^{-1} + Z' H^{-1} Z\right|$$

2.3 The Nelson-Siegel Term Structure Model

In the current study, we also investigate the goodness-of-fit of the NS three-factor model and its estimation method. The model, developed by Nelson and Siegel (1987), decomposes the term structure of interest rates into three factors, namely the level, slope and curvature factors. The NS three-factor model is popular among both practitioners and policy-makers as it uses a flexible and smooth parametric function to replicate the term structure at any given time (Svensson, 1995; Bank for International Settlements, 2005; Gurkaynak, *et al.*, 2007; Christensen, *et al.*, 2011; Sekkel, 2011). Although the NS model may lack the solid theoretical foundation of the affine-class models, it provides an excellent fit to the term structure of interest rates.[1]

Diebold and Li (2006), De Pooter (2007) and Yu and Zivot (2011) all show, empirically, that the NS-class models provide a good fit to the real term structure both in- and out-of-sample. Despite the desirable arbitrage-free property enjoyed by the affine-class models introduced by Vasicek (1977) and Cox, *et al.* (1985), Duffee (2002) argues that the affine models perform poorly when compared with real yield curve data. This being said, Coroneo, *et al.* (2011), using US Treasury yield curves, show that the NS model, in the case of the US market, is compatible with the arbitrage-free constraints. In other words, even without the arbitrage-free setting built-in, the NS-class models are capable of providing a yield curve estimation which is free from arbitrage.[2]

The NS model is based on *Laguerre* functions, which consist of the product between polynomial and exponential decay terms. The basic three-factor NS model can be treated as the solution to a second-order differential equation with equal roots for spot rates. The spot rate curve can be illustrated as

---

[1] The NS model has no restrictions to eliminate opportunities for riskless arbitrage. As the technical detail is beyond the scope of the current paper, we refer interested readers to studies by Filipovic (1999), Diebold, *et al.* (2005) and Christensen, *et al.* (2011), among others. Recently, Christensen, *et al.* (2011) proposed a new set of NS models with an additional 'yield-adjustment' term, which ensure the arbitrage-free property.

[2] Svensson (1995) proposes an extended four-factor model, based on the original NS three-factor model, by adding an additional curvature factor. In this study, we choose to use the NS three-factor model so as to avoid any potential difficulties in interpreting the two curvature factors. In addition, Diebold, *et al.* (2008) show that even adopting a NS model with only the level and slope factors would adequately explain the dynamics of the term structure of interest rates.

$$y_t(\tau) = \beta_{1t} + \beta_{2t}\left[\frac{1-\exp\left(-\frac{\tau}{\lambda_t}\right)}{\left(\frac{\tau}{\lambda_t}\right)}\right] + \beta_{3t}\left[\frac{1-\exp\left(-\frac{\tau}{\lambda_t}\right)}{\left(\frac{\tau}{\lambda_t}\right)} - \exp\left(-\frac{\tau}{\lambda_t}\right)\right] \quad (20)$$

where $y_t(\tau)$ us the spot rate with a maturity of τ at time t; β_{1t}, β_{2t} and β_{3t} are the three factor loadings estimated by the NS model at time t; and, λ_t is the decay factor that optimises the model fitting at time t.

There are three reasons for the NS-class models' popularity. The first of these is that it provides a parsimonious approximation of the yield curve in that it uses only four parameters, detailed below, to estimate the shape of the yield curve. The three components $\left\{1,\left[1-\exp\left(-\tau/\lambda_t\right)\right]/\left(\tau/\lambda_t\right),\left[1-\exp\left(-\tau/\lambda_t\right)\right]/\left(\tau/\lambda_t\right)-\exp\left(-\tau/\lambda_t\right)\right\}$ provide the model with enough flexibility to capture a range of monotonic S-type shapes commonly observed in the yield curve data. The second reason is that the model enjoys the desirable property of starting off at an easily computed instantaneous short-rate value of $\left[\beta_{1t}+\beta_{2t}\right]$ and levelling off at a finite infinite-maturity value of $\left[\beta_{1t}\right]$, which is constant, hence

$$\lim_{\tau \to 0} y_t(\tau) = \beta_{1t} + \beta_{2t} \;\; ; \;\; \lim_{\tau \to \infty} y_t(\tau) = \beta_{1t}$$

The final reason is that the three components provide a clear interpretation in terms of *long-*, *short-*, and *medium-term* components, which can also be identified as the level $\left(\beta_{1t}\right)$, slope $\left(\beta_{2t}\right)$ and curvature $\left(\beta_{3t}\right)$ factor loadings, respectively.[3]

The component attached to β_{1t} is assigned as the *long-term* component as it is constant and therefore the same for every maturity. The component attached to β_{2t} is assigned as the

[3] See Diebold and Li (2006) for further discussion.

short-term component since it starts at 1 but then decays to 0 at ab exponential rate as the maturity increases. The component attached to β_{3t} is the *medium-term* component, which starts at 0, increases for medium-term maturities, and then decays to 0 thereafter, thereby creating a hump-shape. λ_t is the decay factor that determines the maturity at which the *medium-term* component reaches is peak. The average curve resulting from the three-factor NS model is upward sloping and concave, with long-term rates being more persistent than short-term rates. Furthermore, the NS model indicates that the variance of the interest rate decreases as the maturity increases, which is also consistent with the main empirical findings. There are two different approaches that can be employed to estimate the NS model. The first is a simple OLS approach, while second is a non-linear least squares (NLS) approach.

The OLS approach estimates the term structure of interest rates for any given t while fixing the decay factor λ_t at a pre-specified figure, which at constant for every t. In this way, the non-linear exponential measurement equation reduces to a linear framework (Diebold and Li, 2006). Therefore, the NS model can be estimated using a standard cross-sectional OLS approach over the estimation period. The decay factor λ_t determines the maturity at which the *curvature* factor loading reaches its maximum point. When estimating the model, Diebold and Li (2006) use a pre-specified decay factor of $\lambda_t = 16.42$, which means that the *curvature* factor loading reaches its peak at a 30-month maturity. It is worth highlighting that a smaller (larger) value for λ_t produces faster (slower) decaying factor loadings, hence the *curvature* factor will reach its maximum value at a shorter (longer) maturity.

Moving on to the NLS approach, the fact that this approach estimates the decay parameters alongside the other factors makes the estimation procedure more challenging. This being said, it also increases the flexibility of the model, since the assumption of a constant decay parameter over time is eliminated. For this reason, we also use a NLS approach to estimate the parameters for the NS model in this study. One should be aware, however, that the non-linear estimation procedure can occasionally produce extreme results (Gimeno and Nave, 2006; Bolder and Streliski, 1999). The non-linear structure of the model seems to pose serious difficulties in terms of the optimisation procedure arriving at reasonable estimates. De Pooter (2007) showed that, when the decay parameters take on extreme values, the behaviour of the factor loadings will introduce multicollinearity problems; therefore, some of the factors are no longer uniquely identified. The demonstration of this extreme decay parameter problem is shown as

$$\begin{aligned}
&\lim_{\lambda\to 0}\left[\frac{1-\exp\left(-\frac{\tau}{\lambda_{1t}}\right)}{\left(\frac{\tau}{\lambda_{1t}}\right)}\right]=0 \ ; \ \lim_{\lambda\to 0}\left[\frac{1-\exp\left(-\frac{\tau}{\lambda_{1t}}\right)}{\left(\frac{\tau}{\lambda_{1t}}\right)}-\exp\left(-\frac{\tau}{\lambda_{1t}}\right)\right]=0 \\
&\lim_{\lambda\to 0}\left[\frac{1-\exp\left(-\frac{\tau}{\lambda_{1t}}\right)}{\left(\frac{\tau}{\lambda_{1t}}\right)}\right]=1 \ ; \ \lim_{\lambda\to 0}\left[\frac{1-\exp\left(-\frac{\tau}{\lambda_{1t}}\right)}{\left(\frac{\tau}{\lambda_{1t}}\right)}-\exp\left(-\frac{\tau}{\lambda_{1t}}\right)\right]=1
\end{aligned} \tag{21}$$

The above results imply that, for very small values of λ_t, the *slope* and *curvature* factors will be near non-identifiable, which can results in extreme estimation results; while, for large values of λ_t, the *curvature* factors are nearly non-identified. In addition, this means that the *level* and *slope* factors can only be estimated jointly and no longer individually.

In this study, we use both the OLS and NLS approaches to estimate our NS model. For the NLS approach, we follow De Pooter (2007) in that we impose restrictions on the estimation of the decay parameters so as to prevent the aforementioned unfavourable extreme estimates. Given that the mature of the rates in our sample set spans a horizon extending from the 1-month to 30-year rates; we assume that the *curvature* factor loading will reach its peak during the period between the 1-month and around the 20-year horizon. For this reason, the decay parameter for the NS three-factor model is restricted to lie within the $[2,120]$ domain. For the OLS approach, we define the fixed decay factor as the average decay factor estimated using the NLS approach over the sample period.[4]

2.4 The ARIMA and ARFIMA Models

An additional dimension is added to the study through the estimation of standard discrete time models, namely the ARIMA and ARFIMA models. These standard models differ in terms of their underlying assumptions regarding the degree of stationarity of the underlying data series, where the ARIMA model assumes that the underlying data series used in the estimation are non-stationary, while the ARFIMA model assumes that these are fractionally integrated.

We begin the discussion here with the $\text{ARIMA}(p,d,q)$ model (Box and Jenkins, 1976), which, as discussed above, assumes that the underlying data series follow an $I(d)$ non-stationary process. This model has p autoregressive and q moving average terms, where the autoregressive terms measure the impact of the lagged variable and the moving average terms measure the impact of the lagged error. The d parameter measures the level of integration, i.e. the number of times that the underlying data series have to be differenced in order to make the process stationary, where $d \geq 1$ and an integer. We therefore specify the ARIMA model as

$$\phi(L)\left[(1-L)^d y_t\right] = \mu + \theta(L)\varepsilon_t \tag{22}$$

where $\phi(L)$ and $\theta(L)$ denote the polynomials in the lag operator. Therefore, $\phi(L) = 1 - \phi_1 L - \phi_2 L^2 \ldots - \phi_p L^p$, where p denotes the number of autoregressive terms in the

[4] For the OLS approach, the decay factor is fixed at 24.12 and 23.64 for the Canadian and US rates, respectively.

model; $\theta(L)=1-\theta_1 L-\theta_2 L^2 \ldots-\theta_q L^q$, where q denotes the number of moving average terms; $(1-L)^d=\Delta^d y_t$ is the dth difference of y_t; and ε_t denotes white noise.

The alternative model, i.e. the $\text{ARFIMA}(p,d,q)$ model, was first introduced by Granger and Joyeux (1980), Granger (1980, 1981) and Hosking (1981). The assumption underlying this model is that, while the underlying data series follow a mean reverting process, the Wold decomposition of the autocorrelation coefficients for this process will exhibit a very slow hyperbolic rate of decay, where, the higher the value of the d component, the slower the rate of decay. As was the case of the ARIMA model, this model has p autoregressive and q moving average terms as well as a d component, which again measures the order of integration, however, in this case $0<d<1$. The ARFIMA model parameterises the conditional mean of the series generating process as an $\text{ARFIMA}(p,d,q)$ process, which is specified as

$$\phi(L)(1-L)^d\left(y_t-\mu\right)=\theta(L)\varepsilon_t \tag{23}$$

where $\phi(L)$ and $\theta(L)$ denote the polynomials in the lag operator, as described for equation (22), above, where all the roots of $\phi(L)$ and $\theta(L)$ lie outside the unit root circle; p and q denote the number of autoregressive and moving average terms, respectively; d denotes the fractional differencing parameter; and ε_t denotes white noise. This model is estimated using the Maximum Likelihood Estimation (MLE) method outlined in Sowell (1992), hence the proposed likelihood function is

$$\log L=-\frac{T}{2}\log(2\pi)-\frac{1}{2}\log|\Omega|-\frac{1}{2}Y'\Omega^{-1}Y \tag{24}$$

where $\{\Omega\}_{ij}=\gamma_{|i-j|}$, where γ denotes the autocovariances of the ARFIMA process, and Y denotes a T-dimensional vector of the observations on the process y_t. It is worth highlighting again that the Wold decomposition and the autocorrelation coefficients for this process will exhibit a very slow rate of decay, where the higher (lower) the value of d, the slower (faster) the decay. Furthermore, in the case of the first-difference of the series, where $-0.5 < d < 0.5$, the process is covariance stationary, while, should $0.5 < d < 1$, the process would be fractionally integrated. This being said, as long as $d < 1$, the process will exhibit mean-reversion.

The ARFIMA model is included in the analysis as Shea (1991) appeared to provide evidence of long memory in interest rate spreads and some interest rates. Furthermore, although Backus and Zin (1993) noted that estimation of various ARFIMA models for bond series was relatively inconclusive, Crato and Rothman (1994) concluded that when the full MLE method was used to estimate an $\text{ARFIMA}(0,d,1)$ model of annual bond yields, the estimated d component, i.e. $d = 0.81$, was found to be significantly different from $d = 1$, which would be the assumption under the ARIMA model. This paper therefore adds another dimension to the debate as to the existence of long memory in interest rates.

3. Data

The dataset used in the empirical work consists of daily zero yields for Canada and the US, obtained from Datastrean. In particular, the 1-month and 1, 5, 7, 10, 15, 20 and 30-year maturities are used. The interest rates are sampled from January 2003 until December 2011, with data also being collected for January 2012, which is used for out-of-sample forecasts. There is a total of 2,348 observation dates, where at each date there are N interest rates $(N = 8)$. Table 1 reports the summary statistics for the Canadian and US rates, while Figures 1 to 4 display the term structure evolutions over the period. In particular, the means of the Canadian rates range from 2.5410%, for the 1-month rate, to 4.9474%, for the 20-year rate, while the standard deviations range from 1.4987%, for the 1-year rate, 0.6905%, for the 15-

year rate. In the case of the US rates, the means range from 2.2825%, for the 1-month rate, to 4.9459%, for the 30-year rate, with the standard deviations ranging from 2.0083%, in the case of the 1-month rate, to 0.8593%, for the 30-year rate. Finally, the results from the Augmented Dickey-Fuller tests (Dickey and Fuller, 1981) indicate that the Canadian and US rates are first-difference stationary across all maturities.

[Insert Table 1 about here]

[Insert Figures 1 to 4 about here]

Having outlined the initial characteristics of the data, we perform a principal components analysis (PCA) on the sample covariance matrix of the rates to identify the factors that explain the majority of the variation in each dataset. This should provide insight into the number of factors to use in a full-blow estimation of an interest rate model (e.g. Egorov, *et al.*, 2011). PCA transforms the original dataset into variables that maximise the explained variance of the group where each variable is orthogonal to one another. Since the variables are orthogonal, each factor is uniquely determined, up to a sign change.

PCA starts from the assumption that the covariance matrix for the data (Σ) can be decomposed to $\Gamma\Lambda\Gamma^{\mathrm{T}}$, where Γ is a $N \times N$ orthogonal matrix containing factor loadings and Λ is a $N \times N$ diagonal matrix containing N eigenvalues, with N being the number of interest rates. Denoting our original dataset as X, each subsequent variable is defined to be $\Gamma^{\mathrm{T}} X$. As the variance of each factor is given by its corresponding eigenvalue, each variable is ordered based on the relative size of its eigenvalue (see Flury (1988) for more details).[5] The variable with the largest eigenvalue is the first principal component, while the variable with the second

[5] To see this, we denote each variable as M. Since $M = \Gamma^{\mathrm{T}} X$, $\operatorname{var}(M) = \operatorname{var}(M) = \operatorname{var}(\Gamma^{\mathrm{T}} X) = \Gamma^{\mathrm{T}} \operatorname{var}(X)\Gamma$. Since $\operatorname{var}(X) = \Sigma$, $\operatorname{var}(\Gamma^{\mathrm{T}} X) = \Gamma^{\mathrm{T}}\Sigma\Gamma = \Lambda$ owing to the orthogonality of the Γ matrix. Here, Λ us a $N \times N$ matrix containing the eigenvalues of the sample covariance matrix of the group.

largest eigenvalues is the second principal component, and so on. As they are mathematical constructs, principal component factors are latent or unobservable in nature. The simplest way to interpret factors is to examine the effects of a shock to each factor on each yield. To accomplish this task, we plot the factor loading coefficients and provide a description of their shape.

The factor loadings of the first three factors for the Canadian and US rates, respectively, which were estimated using PCA, are presented in Table 2. We also plot the coefficients for the first three factors in Figures 5 and 6 for the Canadian and US rates, respectively. It is worth noting the factor loadings also correspond to the coefficients on an ordinary least squares (OLS) regression of the zero coupon yields on the factors. Each principal component coefficient measures the relative change in the rates given a shock to the corresponding factor.

[Insert Table 2 about here]

[Insert Figures 5 and 6 about here]

The patterns of the factor loadings for the first principal component indicate that a shock to the first factor moves all rates corresponding to each maturity in the same direction.[6] These patterns hold for both the Canadian and US rates. For the rates in both countries, a shock to the second factor moves rates corresponding to short-term maturities (i.e. 1-month and 1-year) in the opposite direction to the rates corresponding to the remaining maturities. Finally, although it only explains 0.1% of the total variation in each group, we examine the third factor since it has a clear interpretation in that, for both the Canadian and US rates, the third factor is a curvature factor. In the case of the Canadian rates, the third factor shifts the 1-month, 1-year, 15-year, 20-year and 30-year yields in the opposite direction to the 5-year, 7-year and 10-year

[6] As each is uniquely determined up to a sign change, we can only conclude that a shock to the first factor moves all yields up or down. Similar conclusions are made for the second and third factors.

yields; while for the US rates, it shifts the 1-month, 15-year, 20-year and 30-year yields in the opposite direction to the 1-year, 5-year, 7-year and 10-year yields.

For our sample of Canadian yields, the first two factors explain 98.54% of the cumulative variation in the sample of yields, with the first factor explaining about 82.59% and the second factor explaining about 15.95% of the variation in the sample. For the sample of US yields, the first two factors explain approximately 98.73% of the variation in the sample of yields, with the first factor explaining about 86.50% and the second factor explaining about 12.23% of the variation in the sample. For each country, the remaining six factors would be regarded as noise. This highlights that PCA is a powerful to that enables us to summarise the data while at the same time minimising the number of factors or variables.

4. Empirical Results

4.1 The Generalised Vasicek Model

We now discuss the empirical results of the one- and two-factor models, presented in Tables 3 and 4 for the Canadian and US rates, respectively. These tables therefore contain the parameter estimates of $\mu, \xi_j, c_j, \rho, \theta_q$, the estimated standard deviations of the measurement errors $\left(\sqrt{h_1}, \ldots, \sqrt{h_N}\right)$ as well as the respective log-likelihood and Bayesian Information Criterion (BIC) (Schwarz, 1978).

[Insert Table 3 about here]

Beginning with the results for the one-factor model of the Canadian rates, the mean reversion (ξ_1) and the diffusion (c_1) parameters are significant, while the long-run average rate (μ) and the market price of risk (θ_1) parameters are insignificant. The estimated standard deviations of the measurement errors are significant and, when compared, are larger for the one-factor model than in the case of the two-factor model. In particular, in the case of the one-factor model, these standard deviations are 74 basis points (bps) for the one-month rate, less than 1 bps for the 1-year rate, 59 bps for the 5-year rate, 63 bps for the 7-year rate, 61 bps for the 10-year rate, 63 bps for the 15-year rate, 69 bps for the 20-year rate, and 74 bps for the 30-year rate.

In the two-factor model, the mean reversion and the diffusion parameters as well as the measurement errors are significant. The long-run average rate is plausible within the data range, and the market price of risk is both positive and significant. The standard deviations of the measurement errors in the two-factor model are generally very small. In particular, these standard deviations are 46 bps for the 1-month rate, less than 1 bps for the 1-year rate, 13 bps for the 5-year rate, less than 1 bps for the 7- and 10-year rates, 31 bps for the 15-year rate, 30 bps for the 20-year rate, and 31 bps for the 30-year rate.

The correlation coefficient $\left(\rho_{12}\right)$ in the two-factor model is -79% and significant. The log-likelihood values for the one- and two-factor models are 45,549 and 50,160, respectively. Based on the BIC, we find that moving from the one-factor to the two-factor model improves the BIC by 10%. The likelihood ratio test of the one- vs. two-factor models gives a value of 9,222, hence on can reject the null hypothesis of one-factor model at the 5% level of significance. The mean reversion parameters imply the mean half-lives of the interest rate process, i.e. the expected time for the process to return half-way to its long-term mean, defined as $-\ln\left(0.5\right)/\xi_j$. For the Canadian rates, for the one-factor model the mean half-life is 1.4 years, while the mean half-lives for the two-factor model are 1.4 years for the first factor and 7.1 years for the second factor.

[Insert Table 4 about here]

Turning to the results for the one-factor model of the US rates, the mean reversion and market price risk parameters are significant. The long-run average rate is also significant and within the average level. The estimated standard deviations of the measurement errors are significant and are larger for most of the rates for the one-factor than when compared to the two-factor model. In particular, in the one-factor model, these standard deviations are 47 bps for the 1-month rate, less than 1 bps for the 1-year rate, 48 bps for the 5-year rate, 53 bps for the 7-year rate, 57 bps for the 10-year rate, 65 bps for the 15-year rate, 72 bps for the 20-year rate, and 75 bps for the 30-year rate.

In the two-factor model, the mean reversion parameter, first diffusion parameter and the measurement errors are significant; and the long-run average is plausible within the data range. The market price of risk parameters are both positive and the first one is significant. The standard deviations of the measurement errors for the two-factor model are generally very small. In particular, these standard deviations are 46 bps for the 1-month rate, less than 1 bps for the 1-year rate, 23 bps for the 5-year rate, less than 1 bps for the 7- and 10-year rates, 31 bps for the 15-year rate, 40 bps for the 20-year rate, and 70 bps for the 30-year rate.

The correlation coefficient in the two-factor model is -80% and significant. The log-likelihood values for the one- and two-factor models are 46,615 and 54,584, respectively. Based on the BIC, we find that moving from the one-factor to the two-factor model improves the BIC by 17%. The likelihood ratio test of the one- vs. two-factor models gives a value of 15,938, hence on can reject the null hypothesis of one-factor model at the 5% level of significance. The mean reversion parameters imply that, for the US rates, for the one-factor model the mean half-life is 2.7 years, while the mean half-lives for the two-factor model are 1.4 years for the first factor and 7.1 years for the second factor.

We also look at the factor loadings for the two-factor model as a function of maturity, which should help determine the nature of the factors calculated by the Kalman filter. This is supported by Litterman and Scheinkman (1991) who, using PCA, investigated a number of US yields and identified three factors, which they interpreted as changes in level, steepness and curvature.

[Insert Figures 7 and 8 about here]

With this in mind, Figures 7 and 8 plot the Canadian and US factor loadings for the two-factor model, as a function of maturity, respectively. These indicate that, in the case of the two-factor model, the first factor's impact on changes in the yield, where Litterman and Scheinkman (1991) identified this as a level factor, for both the Canadian and US rates, has an increasing and positive effect on maturity of up to 8 years, beyond which it has an equal impact on the remaining maturities. Moving on, the figures indicate that the second factor, which Litterman and Scheinkman (1991) identified as a steepness factor, has a strong influence, for both the Canadian and US rates, on short-term rates of up to 10 years, where it lowers them, following which it then raises yields on longer-term maturities.

It is worth noting that, whereas the Litterman and Scheinkman (1991) approach is completely data-driven, the state-space approach imposes restrictions on the extracted factors that come from a formal term structure model used in pricing bonds. Further examples of using this approach of comparing the factors from the seminar Cox, Ingersoll and Ross model (Cox, *et al.*, 1985) using the state-space approach are given in Geyer and Pichler (1999) and Chen and Scott (2003).

4.2 The ARIMA and ARFIMA Models

Having completed the analysis of the results of the GVT model, we now examine those for the discrete time models. Given the fact that the unit root tests presented in Table 1 indicated that both the Canadian and US rates were non-stationary, $\text{ARIMA}(0,1,1)$ through $\text{ARIMA}(3,1,3)$

models were estimated, where the best specification was selected on the basis of the log-likelihood, Akaike Information Criterion (AIC) (Akaike, 1974) and BIC.

The results of the best model specifications for the Canadian rates are presented in Panel A of Table 5. The results here are somewhat mixed in that no specification for the 7-year rate was found to be significant, while the Canadian 1-month, 1-year, 5-year and 10-year rates were best specified as $\text{ARIMA}(2,1,0)$, $\text{ARIMA}(2,1,2)$, $\text{ARIMA}(1,1,1)$ and $\text{ARIMA}(0,1,1)$ processes, respectively. The results at the long-end of the curve are more consistent in that the Canadian 15-year, 20-year and 30-year rates are best specified as $\text{ARIMA}(1,1,0)$ processes. We therefore conclude that, while past rates are found to have a significant impact on current rates across the yield curve, with the exception of the 10-year rate, the impact of past shocks is found to vary at the short- and medium-ends of the curve, while at the long-end of the curve they do not appear to have any real impact.

[Insert Table 5 about here]

Changing focus to the results for the US rates, presented in Panel B of Table 5, these are somewhat more uniform in that only the ARIMA specifications at the very short-end of the curve, i.e. for the 1-month and 1-year rates, are found to be significant. As was the case for the Canadian rates, past rates are found to have a significant impact on current rates for the 1-month and 1-year rates, although past shocks are only found to have an impact on the prevailing 1-month rate.

As stated previously, the underlying assumption of the ARIMA model is that the underlying data series follows a non-stationary process. Given the discussion in the extant literature, and as stated previously, an interesting approach would be to extend this debate by arguing that US and Canadian rates are fractionally integrated, hence shocks to these would not persist indefinitely, as would be the case under the assumption of non-stationarity, but would instead decay hyperbolically, thereby indicating that these rates are mean-reverting, however, there would be a delay in the mean-reversion process. In order to investigate this alternate hypothesis, $\text{ARFIMA}(0,d,1)$ through $\text{ARFIMA}(3,d,3)$ models are estimated across both sets of rates, where, as was the case for the ARIMA models, the best model is then selected on the basis of the log-likelihood, AIC and BIC measures.

Examining the results of these models of the Canadian rates, present in Panel A of Table 6, no ARFIMA specification is found to be significant for any of the 1-year, 7-year and 10-year rates. As above, results for the long-end of the curve indicate that rates in the previous period have a significant impact on the current prevailing rates, while past shocks are found to have no significant effect. At the shorter-end of the curve, both past rates and past shocks are found to have a significant impact on the current rates.

[Insert Table 6 about here]

If one looks at the results for the US rates, presented in Panel B of Table 6, the results are almost uniform in that, with the exception of the 1-month rate, the best model specification is found to be an $\text{ARFIMA}(1,d,0)$. Interestingly, past shocks are found to have no significant impact on current rates, regardless of the time horizon examined. We therefore conclude that this may be an indication that including the fractional component in the process may enable us to capture more of the dynamics of the data.

5. Forecast Results

Having estimated these models, ex-post dynamic forecasts were performed for each of these models using the rates during January 2012, which corresponds to a period of 22 days. These forecasts were then compared using the Root Mean Squared-Error (RMSE) forecast metric, which is calculated as

$$\text{RMSE} = \sqrt{\frac{1}{FH}\left[\sum_{i=1}^{FH}\left(r_i^{\,a} - r_i^{\,f}\right)^2\right]} \tag{25}$$

where $r_i^{\,a}$ denotes the actual observed value at time i, $r_i^{\,f}$ denotes the forecasted value at time i, and FH denotes the forecast horizon.

Beginning with the forecast metrics for the forecasts of the Canadian rates, presented in Panel A of Table 7. At the short-end of the curve, the NS model is found to perform best in terms of forecasting the 1-month rate, followed by the two-factor GVT model; while, in the case of the 1-year rate, the two-factor GVT model is found to perform best, followed by the NS model. For the medium-term of the curve, the NS model is found to perform best for both the 5-year and 7-year rates, again followed by the two-factor GVT model. Finally, for the longer-end of the curve, although the two-factor model outperforms the NS model in terms of forecasting the 15-year rate, the NS model outperforms all other models for the 20-year and 30-year rates. Overall, however, the NS model is found to outperform the other models in terms of forecasting the yield curve.

[Insert Table 7 about here]

Moving onto the forecast metrics for the forecasts of the US rates, presented in Panel B of Table 7, at the short-end of the curve, the ARIMA model is found to perform best, followed by the ARFIMA model, in terms of forecasting the 1-month rate, while in the case of the 1-year rate, the ARIMA and ARFIMA models are joint best. Across all other horizons, however, the ARFIMA model uniformly outperforms the other models. In terms of the remaining models, the NS model is found to outperform the one-and two-factor GVT model in terms of forecasting the 1-month, 7-year, 10-year, 15-year, 20-year and 30-year rates, while the two-factor GVT model outperforms the NS model for the 1-year and 5-year rates. We can therefore conclude that introducing a fractional component, at least in terms of the US rates, definitely allows us to better capture the overall dynamics of the yield curve.

6. Conclusions

In this paper, we have compared empirical evidence for the Canadian and US yield curves using a one- and two-factor GVT yield curve model, using daily panel data, which were then compared to the NS and standard discrete time (ARIMA and ARFIMA) models. We then compared the forecasting performance of these models so as to determine which best fits the dynamics of the respective yield curves.

The choice of model comparison was justified by the argument that, although the NS model may lack the theoretical foundation of the affine-class models, introduced by Vasicek (1977) and Cox, *et al.* (1985), Duffee (2002) argues that the affine models perform poorly when compared with real yield curve data, whereas the NS provides an excellent fit to the term structure of interest rates. The inclusion of the ARIMA and ARFIMA models was justified by the fact that, although Backus and Zin (1993) did not found any conclusive evidence of interest rates following a fractionally integrated process, this argument was counteracted by Shea (1991) and Crato and Rothman (1994). We therefore felt that it would be interesting to examine whether the Canadian and US rates are fractionally integrated, whereby shocks to these would not persist indefinitely, as would be the case under the assumption of non-stationarity inherent in the ARIMA model and as suggested by the unit root tests performed, but would instead decay hyperbolically, thereby indicating that the rates are mean-reverting, however there would be a delay in the mean-reversion process.

Our in-sample results suggest that, out of the two forms of GVT models; the two-factor model has a good fit for both the Canadian and US yield curves. The in-sample results for the discrete-time models, suggest that the US rates could not be estimated using the standard

ARIMA approach, but could using the ARFIMA approach, thereby lending support to the argument of fractional integration, at least for these rates. In terms of the ex-post forecasts, we find that, overall, the NS model is found to outperform the other models in terms of forecasting the Canadian yield curve. This being said, when examining the forecasts of the US rates, we found that the ARFIMA model generally outperformed the other models, where the NS model was generally found again to outperform the GVT models.

Given that the yield curve provides crucial information for both policymakers and other players in the fixed-income instrument markets, in terms of providing a preliminary indication of the future direction of interest rates and yields, we can draw two important conclusions. The first of these is that NS-class models do indeed outperform the affine-class models when compared with real data, which, together with the fact that these have the advantage of being parsimonious in terms of its parameterisation, the fact that it is comparatively easy to compute, and that is has a clear interpretation, may lend support to the further use of these models. The second conclusion is that further investigation is needed to definitively conclude whether interest rates follow a fractionally integrated process.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transaction on Automatic Control*, 19(6), 716-723.

Babbs, S. H., & Nowman, K. B. (1999). Kalman filtering of Generalised Vasicek term structure models. *Journal of Financial & Quantitative Analysis*, 34(1), 115-130.

Backus, D. K., & Zin, S. E. (1993). Long-memory inflation uncertainty: Evidence from the term structure of interest rates. *Journal of Money, Credit & Banking*, 25(3), 681-700.

Bank for International Settlements (2005). Zero-coupon yield curves: Technical documentation. *BIS Papers*, No. 25, Monetary & Economic Department, Bank for International Settlements, October 2005.
[Available at: http://www.bis.org/publ/bppdf/bispap25.pdf]

Bergstrom, A. R. (1984). Continuous time stochastic models and issues of aggregation over time. In: Gilriches, Z. & Intrilgator, M. D. (Eds.), Handbook of Econometrics, Amsterdam: Elsevier Science, pp. 1146-1212.

Bolder, D., & Strelitski, D. J. (1999). Yield curve modeling at the Bank of Canada. *Bank of Canada Technical Report*, No. 84, Bank of Canada, February 1999.

Box, G. E. P., & Jenkins, G. M. (1976). Time Series Analysis: Forecasting and Control. San Francisco: Holden-Day.

Chatterjee, S. (2005). Application of the Kalman filter for estimating continuous time term structure models: The case of UK and Germany. *Department of Economics Discussion Paper Series*, 2005-2, University of Glasgow.
[Available at: http://hdl.handle.net/1905/355]

Chen, R.-R., & Scott, L. (2003). Multi-factor Cox-Ingersoll-Ross models of the term structure: Estimates and tests from a Kalman filter model. *Journal of Real Estate Finance & Economics*, 27(2), 143-172.

Christensen, J. H. E., Diebold, F. X., & Rudebusch, G. D. (2011). The affine arbitrage-free class of Nelson-Siegel term structure models. *Journal of Econometrics*, 164(1), 4-20.

Coroneo, L. Nyholm, K., & Vidova-Koleva, R. (2011). How arbitrage-free is the Nelson-Siegel model?, *Journal of Empirical Finance*, 18(3), 393-407.

Cox, J. C., Ingersoll, J. E., & Ross, S. A. (1985). A theory of the term structure of interest rates. *Econometrica*, 53(2), 385-407.

Crato, N., & Rothman, P. (1994). Fractional integration analysis of long-run behavior for US macroeconomic time series. *Economics Letters*, 45(3), 287-291.

De Jong, F. (2000). Time series and cross-section information in affine term-structure models.

Journal of Business & Economic Statistics, 18(3), 300-314.

De Pooter, M. (2007). Examining the Nelson-Siegel class of term structure models. *Tinbergen Institute Discussion Paper*, TI 07-043/4, Tinbergen Institute, Erasmus University Rotterdam, June 2007.

Dickey, D. A., & Fuller, W. A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica*, 49(4), 1057-1072.

Diebold, F. X., & Li, C. (2006). Forecasting the term structure of government bond yields. *Journal of Econometrics*, 130(2), 337-364.

Diebold, F. X., Li, C., & Yue, V. Z. (2008). Global yield curve dynamics and interaction: A dynamic Nelson-Siegel approach. Journal of Econometrics, 146(2), 351-363.

Diebold, F. X., Piazzesi, M., & Rudebusch, G. D. (2005). Modeling bond yields in finance and macroeconomics. *American Economic Review*, 95(2), 415-420.

Driessen, J., Melenberg, B., & Nijman, T. (2005). Testing affine term structure models in case of transaction costs. *Journal of Econometrics*, 126(1), 201-232.

Duan, J.-C., & Simonato, J.-G. (1999). Estimating and testing exponential affine term structure models by Kalman filters. *Review of Quantitative Finance & Accounting*, 13(2), 111-135.

Duffee, G. R. (2002). Term premia and interest rate forecasts in affine models. *Journal of Finance*, 57(1), 405.443.

Egorov, A. V., Li, H., & Ng, D. (2011). A tale of two yield curves: Modeling the joint term structure of Dollar and Euro interest rates. *Journal of Econometrics*, 162(1), 55-70.

Filipovic, D. (1999). A note on the Nelson-Siegel family. *Mathematical Finance*, 9(4), 349-359.

Flury, B. (1988). Common Principal Components and Related Mutlivariate Models. New York: Wiley.

Geyer, A. J. L., & Pichler, S. (1999). A state-space to estimate and test multi-factor Cox-Ingersoll-Ross models of the term structure. *Journal of Financial Research*, 22(1), 107-130.

Gimeno, R., & Nave, J. M. (2006). Genetic algorithm estimation of interest rate term structure. Bank of Spain Working Paper, No. 0636, Banco de Espana, December 2006. [Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=950982]

Granger, C. W. J. (1980). Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics*, 14(2), 227-238.

Granger, C. W. J. (1981). Some properties of time series data and their use in econometric model specification. *Journal of Econometrics*, 16(1), 121-130.

Granger, C. W. J., & Joyeux, R. (1980). An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis*, 1(1), 15-29.

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

Hosking, J. R. M. (1981). Fractional differencing, *Biometrika*, 68(1), 165-176.

Jegadeesh, N., & Pennacchi, G. G. (1996). The behaviour of interest rates implied by the term structure of Eurodollar futures. *Journal of Money, Credit & Banking*, 28(3), 426-446.

Langetieg, T. C. (1980). A multivariate model of the term structure. *Journal of Finance*, 35(1), 71-91.

Litterman, R., & Scheinkman, J. (1991). Common factors affecting bond returns. *Journal of Fixed Income*, 1(1), 64-61.

Nelson, C. R., & Siegel, A. F. (1987). Parsimonious modelling of yield curves. *Journal of Business*, 60(4), 473-589.

Nowman, K. B. (2010). Modelling the UK and Euro yield curves using the Generalized Vasicek model: Empirical results from panel data for one and two factor models. *International Review of Financial Analysis*, 19(5), 334-341.

O’Sullivan, C. (2007). Parameter uncertainty in Kalman filter estimation of the CIR term structure model. *Centre for Financial Markets Working Paper Series*, WP-07-18, University College Dublin.
[Available at: http://hdl.handle.net/10197/1193]

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2), 461-464.

Sekkel, R. (2011). International Evidence on bond risk premia. *Journal of Banking & Finance*, 35(1), 174-181.

Shea, G. S. (1991). Uncertainty and implied variance bounds in long-memory models of the interest rate term structure. *Empirical Economics*, 16(3), 287-312.

Sowell, F. (1992). Maximum likelihood estimation of stationary univariate fractionally integrated time series models. *Journal of Econometrics*, 53(1-3), 165-188.

Svensson, L. E. O. (1995). Estimating forward interest rates with the extended Nelson-Siegel method. *Sveriges Riksbank Quarterly Review*, 1995(3), 13-26.

Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics*, 5(2), 177-188.

Table 1: Summary Statistics for the Canadian and US Rates (2003 to 2011)

Panel A - Levels of the Canadian Rates								
	1-Month	**1-Year**	**5-Year**	**7-Year**	**10-Year**	**15-Year**	**20-Year**	**30-Year**
Mean	2.5410	2.6261	3.6071	3.9319	4.3193	4.7760	4.9474	4.8465
Standard Deviation	1.4605	1.4987	0.9772	0.9106	0.8098	0.6905	0.6916	0.7991
Augmented Dickey-Fuller Test	-0.9525	-1.1378	-2.0408	-2.1984	-2.5114	-2.8606	-2.7197	-2.8780
	(0.9484)	(0.9211)	(0.5780)	(0.4897)	(0.3225)	(0.1758)	(0.2286)	(0.1699)
Panel B - First-Differences of the Canadian Rates								
	1-Month	**1-Year**	**5-Year**	**7-Year**	**10-Year**	**15-Year**	**20-Year**	**30-Year**
Mean	-0.0007	-0.0010	-0.0012	-0.0012	-0.0012	-0.0012	-0.0013	-0.0015
Standard Deviation	0.0257	0.0351	0.0562	0.0526	0.0507	0.0500	0.0482	0.0478
Augmented Dickey-Fuller Test	-23.2209	-24.4489	-48.2241	-47.1650	-46.9510	-46.3894	-46.1243	-50.2181
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Panel A - Levels of the US Rates								
	1-Month	**1-Year**	**5-Year**	**7-Year**	**10-Year**	**15-Year**	**20-Year**	**30-Year**
Mean	2.2825	2.5168	3.6307	4.0074	4.3754	4.7431	4.8939	4.9459
Standard Deviation	2.0083	1.8549	1.2321	1.0594	0.9334	0.8678	0.8789	0.8593
Augmented Dickey-Fuller Test	-1.3708	-1.2568	-1.9531	-2.1679	-2.4421	-2.6475	-2.6490	-2.7054
	(0.8693)	(0.8975)	(0.6260)	(0.5068)	(0.3575)	(0.2591)	(0.2584)	(0.2344)
Panel B - First-Differences of the US Rates								
	1-Month	**1-Year**	**5-Year**	**7-Year**	**10-Year**	**15-Year**	**20-Year**	**30-Year**
Mean	-0.0005	-0.0003	-0.0009	-0.0009	-0.0010	-0.0011	-0.0012	-0.0012
Standard Deviation	0.0709	0.0424	0.0705	0.0700	0.0699	0.0675	0.0661	0.0673
Augmented Dickey-Fuller Test	-11.0939	-44.7287	-48.5677	-48.1719	-48.1902	-48.3365	-48.3350	-49.6208
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)

Note: Figures in parentheses denote the *p*-values for the Augmented Dickey-Fuller test (Dickey and Fuller, 1981), which tests $H_0 : \phi = 1$ vs. $H_A : \phi < 1$.

Table 2: Factor Loadings for the Canadian and US Rates (2003 to 2011)

Panel A - Factor Loadings for the Canadian Rates								
	1-Month	**1-Year**	**5-Year**	**7-Year**	**10-Year**	**15-Year**	**20-Year**	**30-Year**
Factor 1	0.4989	0.5307	0.3621	0.3303	0.2812	0.2171	0.2053	0.2466
Factor 2	-0.4953	-0.4062	0.1389	0.2290	0.2951	0.3503	0.3845	0.4004
Factor 3	0.0235	0.3145	-0.5871	-0.4056	-0.1611	0.1166	0.3020	0.5105
Panel B - Factor Loadings for the US Rates								
	1-Month	**1-Year**	**5-Year**	**7-Year**	**10-Year**	**15-Year**	**20-Year**	**30-Year**
Factor 1	0.5586	0.5237	0.3585	0.3034	0.2561	0.2170	0.2058	0.1952
Factor 2	-0.4586	-0.3578	0.1282	0.2128	0.2929	0.3807	0.4300	0.4336
Factor 3	0.6068	-0.3770	-0.4759	-0.2908	-0.0821	0.1250	0.2500	0.3062

Table 3: Estimates for the GVT Models of the Canadian Rates (2003 to 2011)

Panel A - Model Parameters		
	One-Factor	**Two-Factor**
ξ_1	0.4947	0.4947
	(0.1514)	(0.0124)
ξ_2		0.0973
		(0.0484)
c_1	0.0158	0.0157
	(0.0053)	(0.0037)
c_2		0.0170
		(0.0090)
ρ_{12}		-0.7976
		(0.0119)
μ	0.0435	0.0435
	(0.0307)	(0.0130)
θ_1	0.2163	0.2163
	(1.0227)	(0.0031)
θ_2		0.3504
		(0.0085)
Panel B - Standard Deviations of Measurement Errors		
	One-Factor	**Two-Factor**
$\sqrt{h_1}$	0.0074	0.0046
	(0.0015)	(0.0002)
$\sqrt{h_2}$	0.0000	0.0000
	(<0.0001)	(<0.0001)
$\sqrt{h_3}$	0.0059	0.0013
	(0.0006)	(0.0003)
$\sqrt{h_4}$	0.0063	0.0006
	(0.0005)	(0.0002)
$\sqrt{h_5}$	0.0061	0.0004
	(0.0002)	(<0.0001)
$\sqrt{h_6}$	0.0063	0.0031
	(0.0001)	(0.0017)
$\sqrt{h_7}$	0.0069	0.0030
	(0.0002)	(0.0014)
$\sqrt{h_8}$	0.0074	0.0030
	(0.0002)	(0.0014)
Panel C - Information Criteria		
	One-Factor	**Two-Factor**
log-likelihood	45,549	50,160
BIC	-91,004	

Note: Figures in parentheses are the standard errors of the coefficient estimate above.

Table 4: Estimates for the GVT Models of the US Rates (2003 to 2011)

Panel A - Model Parameters		
	One-Factor	**Two-Factor**
ξ_1	0.2536	0.4901
	(0.0222)	(0.0429)
ξ_2		0.0974
		(0.0129)
c_1	0.0078	0.0220
	(1.6449)	(0.0085)
c_2		0.0196
		(0.0135)
ρ_{12}		-0.7976
		(0.0788)
μ	0.0471	0.0488
	(0.0109)	(0.0458)
θ_1	0.2956	0.1734
	(0.0768)	(0.0018)
θ_2		0.3504
		(0.2927)
Panel B - Standard Deviations of Measurement Errors		
	One-Factor	**Two-Factor**
$\sqrt{h_1}$	0.0047	0.0046
	(0.0024)	(0.0002)
$\sqrt{h_2}$	0.0000	0.0000
	(<0.0001)	(<0.0001)
$\sqrt{h_3}$	0.0048	0.0023
	(0.0022)	(0.0003)
$\sqrt{h_4}$	0.0053	0.0006
	(0.0016)	(0.0001)
$\sqrt{h_5}$	0.0056	0.0004
	(0.0015)	(0.0001)
$\sqrt{h_6}$	0.0065	0.0031
	(0.0010)	(0.0006)
$\sqrt{h_7}$	0.0072	0.0040
	(0.0009)	(0.0013)
$\sqrt{h_8}$	0.0075	0.0070
	(0.0012)	(0.0002)
Panel C - Information Criteria		
	One-Factor	**Two-Factor**
log-likelihood	46,615	54,584
BIC	-93,137	-109,169

Note: Figures in parentheses are the standard errors of the coefficient estimate above.

Table 5: Estimates for the ARIMA Models of the Canadian and US Rates (2003 to 2011)

Panel A - ARIMA Model Results for the Canadian Rates								
	1-Month	**1-Year**	**5-Year**	**7-Year**	**10-Year**	**15-Year**	**20-Year**	**30-Year**
α	-0.0007	-0.0012	-0.0012	-0.0013	-0.0012	-0.0013	-0.0013	-0.0015
	(0.0007)	(0.0015)	(0.0012)	(0.0011)	(0.0011)	(0.0011)	(0.0010)	(0.0010)
ϕ_1	0.0888	0.5549	-0.4212	-0.1431		0.0444	0.0498	-0.0361
	(0.0205)	(0.1353)	(0.2176)	(0.4528)		(0.0206)	(0.0206)	(0.0206)
ϕ_2	0.1202	0.4149		0.1514				
	(0.0205)	(0.1285)		(0.2466)				
θ_1		-0.4932	0.4496	0.1739	0.0359			
		(0.1343)	(0.2147)	(0.4502)	(0.0206)			
θ_2		-0.4427		-0.1865				
		(0.1230)		(0.2505)				
log-likelihood	5,291	4,552	3,428	3,586	3,671	3,708	3,792	3,808
AIC	-7.3488	-6.7184	-5.7589	-5.8946	-5.9655	-5.9978	-6.0699	-6.0833
BIC	-7.3388	-6.7018	-5.7489	-5.8780	-5.9589	-5.9912	-6.0633	-6.0767
Panel B - ARIMA Model Results for the US Rates								
	1-Month	**1-Year**	**5-Year**	**7-Year**	**10-Year**	**15-Year**	**20-Year**	**30-Year**
α	-0.0005	-0.0004	-0.0010	-0.0011	-0.0012	-0.0013	-0.0014	-0.0013
	(0.0016)	(0.0009)	(0.0014)	(0.0014)	(0.0014)	(0.0014)	(0.0013)	(0.0013)
ϕ_1	-0.5235	0.0821	-0.7216	-0.7487	-0.7591	-0.7638	0.0412	0.2601
	(0.0872)	(0.0206)	(0.1578)	(0.1600)	(0.1562)	(0.1610)	(0.3756)	(0.5229)
ϕ_2	0.0749		0.1106	0.0825	0.1009	0.0689	0.1168	0.1201
	(0.0287)		(0.1270)	(0.1273)	(0.1280)	(0.1301)	(0.2099)	(0.2821)
ϕ_3	-0.1505		0.7225	0.7587	0.7686	0.7678	-0.0367	-0.2852
	(0.0205)		(0.1578)	(0.1605)	(0.1566)	(0.1617)	(0.3728)	(0.5207)
θ_1	0.7503		-0.1566	-0.1167	-0.1340	-0.0990	-0.1695	-0.1547
	(0.0868)		(0.1323)	(0.1328)	(0.1341)	(0.1333)	(0.2089)	(0.2923)
log-likelihood	2,952	4,098	2,908	2,922	2,925	3,004	3,051	3,007
AIC	-5.3545	-6.3305	-5.3157	-5.3275	-5.3301	-5.3978	-5.4383	-5.4004
BIC	-5.3379	-6.3239	-5.2992	-5.3109	-5.3135	-5.3813	-5.4217	-5.3839

Note: Figures in parentheses are the standard errors of the coefficient estimate above.

Table 6: Estimates for the ARFIMA Models of the Canadian and US Rates (2003 to 2011)

Panel A - ARIMA Model Results for the Canadian Rates								
	1-Month	**1-Year**	**5-Year**	**7-Year**	**10-Year**	**15-Year**	**20-Year**	**30-Year**
α	-0.0004	-0.0005	-0.0008	-0.0009	-0.0009	-0.0009	-0.0009	-0.0013
	(0.0013)	(0.0008)	(0.0012)	(0.0011)	(0.0010)	(0.0011)	(0.0010)	(0.0010)
d	1.0783	1.0417	1.0086	1.0152	1.0081	1.0093	1.0114	0.9852
	(0.0310)	(0.0000)	(0.0160)	(0.0000)	(0.0000)	(0.0480)	(0.0490)	(0.0500)
ϕ_1	0.9514	0.7524	-0.4010	-0.2924	0.3184	0.0357	0.0391	0.0196
	(0.0235)	(0.1484)	(0.2363)	(0.3599)	(0.3567)	(0.0206)	(0.0206)	(0.0074)
ϕ_2	0.0372	0.0205		-0.0433	-0.0539			
	(0.0212)	(0.0230)		(0.0232)	(0.0209)			
θ_1	-0.9714	-0.7349	0.4224	0.3068	-0.2945			
	(0.0113)	(0.1481)	(0.2342)	(0.3602)	(0.3571)			
log-likelihood	5,317	4,548	3,427	3,586	3,675	3,707	3,791	3,807
AIC	-7.3711	-6.7152	-5.7584	-5.8949	-5.9704	-5.9973	-6.0692	-6.0816
BIC	-7.3578	-6.7020	-5.7485	-5.8816	-5.9572	-5.9907	-6.0626	-6.0749
Panel B - ARIMA Model Results for the US Rates								
	1-Month	**1-Year**	**5-Year**	**7-Year**	**10-Year**	**15-Year**	**20-Year**	**30-Year**
α	-0.0002	-0.0003	-0.0008	-0.0009	-0.0010	-0.0011	-0.0012	-0.0011
	(0.0012)	(0.0009)	(0.0015)	(0.0015)	(0.0015)	(0.0015)	(0.0014)	(0.0014)
d	1.1130	1.0179	0.9785	0.9840	0.9854	0.9870	0.9886	0.9825
	(0.0540)	(0.0330)	(0.0420)	(0.0150)	(0.0440)	(0.0450)	(0.0460)	(0.0160)
ϕ_1	0.0979	0.0627	0.0618	0.0593	0.0528	0.0497	0.0469	0.0398
	(0.0205)	(0.0206)	(0.0149)	(0.0136)	(0.0124)	(0.0110)	(0.0104)	(0.0104)
ϕ_2	-0.1374							
	(0.0204)							
ϕ_3	-0.1162							
	(0.0205)							
log-likelihood	2,959	4,099	2,900	2,916	2,920	3,002	3,051	3,005
AIC	-5.3610	-6.3317	-5.3084	-5.3222	-5.3250	-5.3952	-5.4366	-5.3977
BIC	-5.3478	-6.3251	-5.3018	-5.3156	-5.3184	-5.3885	-5.4300	-5.3911

Note: Figures in parentheses are the standard errors of the coefficient estimate above.

Table 7: Forecast Comparison for the Canadian and US Rates Using the Root Mean Squared-Error Metric

Panel A - Forecast Metric for the Canadian Rates	1-Month	1-Year	5-Year	7-Year	10-Year	15-Year	20-Year	30-Year
ARIMA	0.3536	0.4752	0.5771		0.6154	0.6073	0.6329	0.7256
ARFIMA	0.3535		0.5772			0.6075	0.633	0.7256
1-Factor GVT	1.3719	0.0907	1.5484	1.6524	1.6026	1.4494	1.5266	1.9377
2-Factor GVT	0.6623	0.0046	0.0972	0.0655	0.0936	0.118	0.1529	0.7946
Nelson-Siegel [2,120]	0.0295	0.0096	0.0324	0.0496	0.0606	0.2195	0.1524	0.2578
Nelson-Siegel [2,120]	0.0377	0.0249	0.0353	0.0564	0.0552	0.2125	0.1458	0.2643
Nelson-Siegel [2,120]	0.1172	0.1761	0.0363	0.0365	0.0954	0.2194	0.1223	0.3219

Panel A - Forecast Metric for the US Rates	US1M	US1Y	US5Y	US7Y	US10Y	US15Y	US20Y	US30Y
ARIMA	0.0018	0.0186						
ARFIMA	0.002	0.0186	0.0372	0.0422	0.0471	0.0506	0.0527	0.0536
1-Factor GVT	0.1587	0.1693	1.3114	1.3776	1.4749	1.6144	1.7932	2.0277
2-Factor GVT	0.327	0.1001	0.1276	0.076	0.0718	0.2397	0.5784	1.1225
Nelson-Siegel [2,120]	0.0624	0.1116	0.1497	0.0515	0.0663	0.1022	0.0575	0.0994
Nelson-Siegel [2,120]	0.0611	0.1128	0.1481	0.0508	0.0682	0.1042	0.0586	0.098
Nelson-Siegel [2,120]	0.0945	0.1655	0.1374	0.0612	0.0546	0.1002	0.0657	0.0691

Figure 1: Canadian Yield Curve (2003 to 2011)

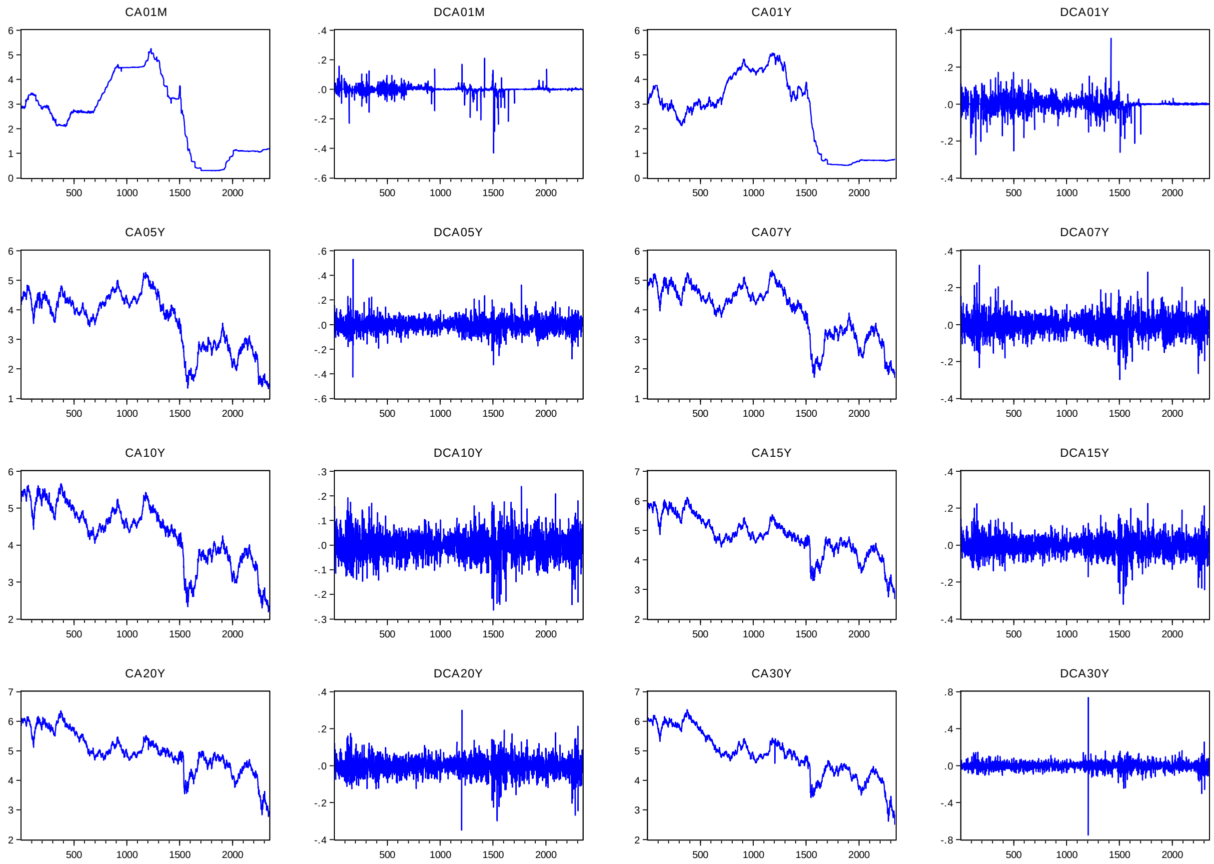

Figure 2: US Yield Curve (2003 to 2011)

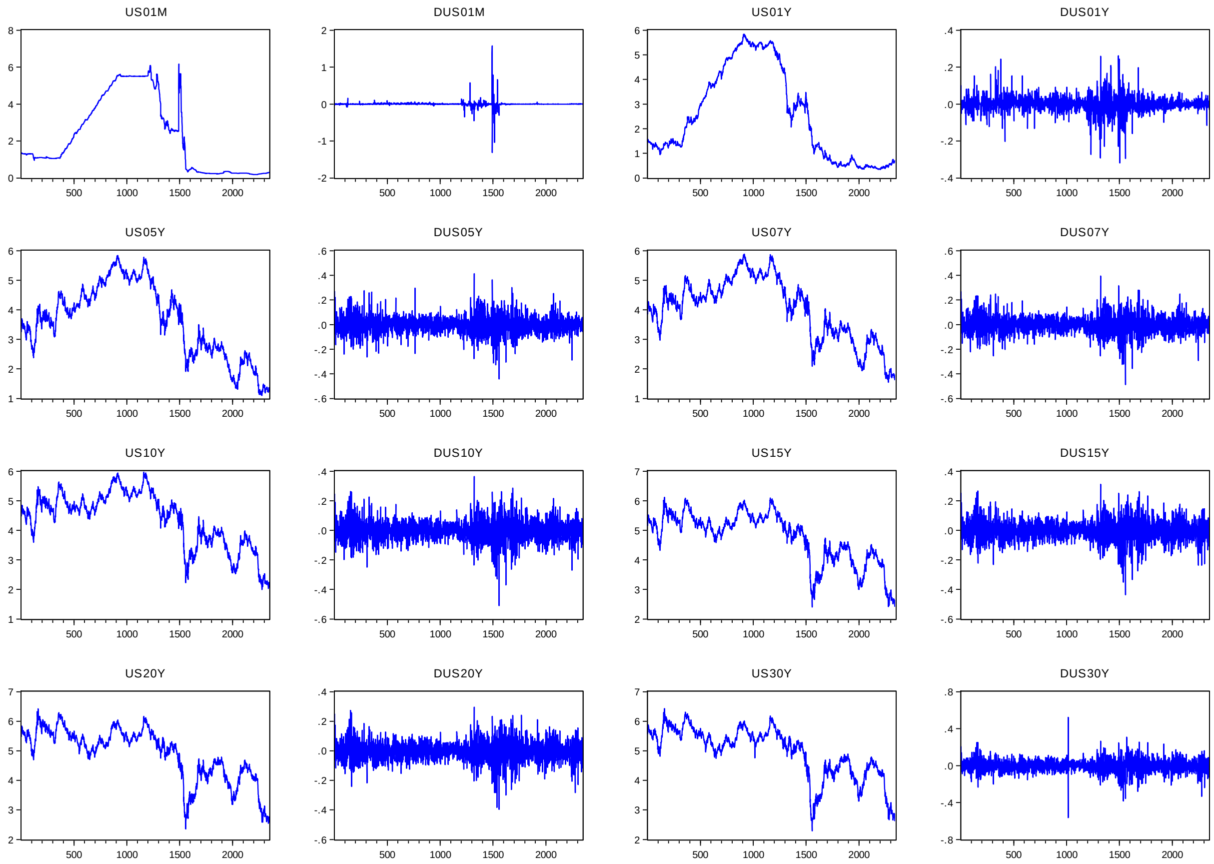

Figure 3: Canadian Term Structure (2003 to 2011)

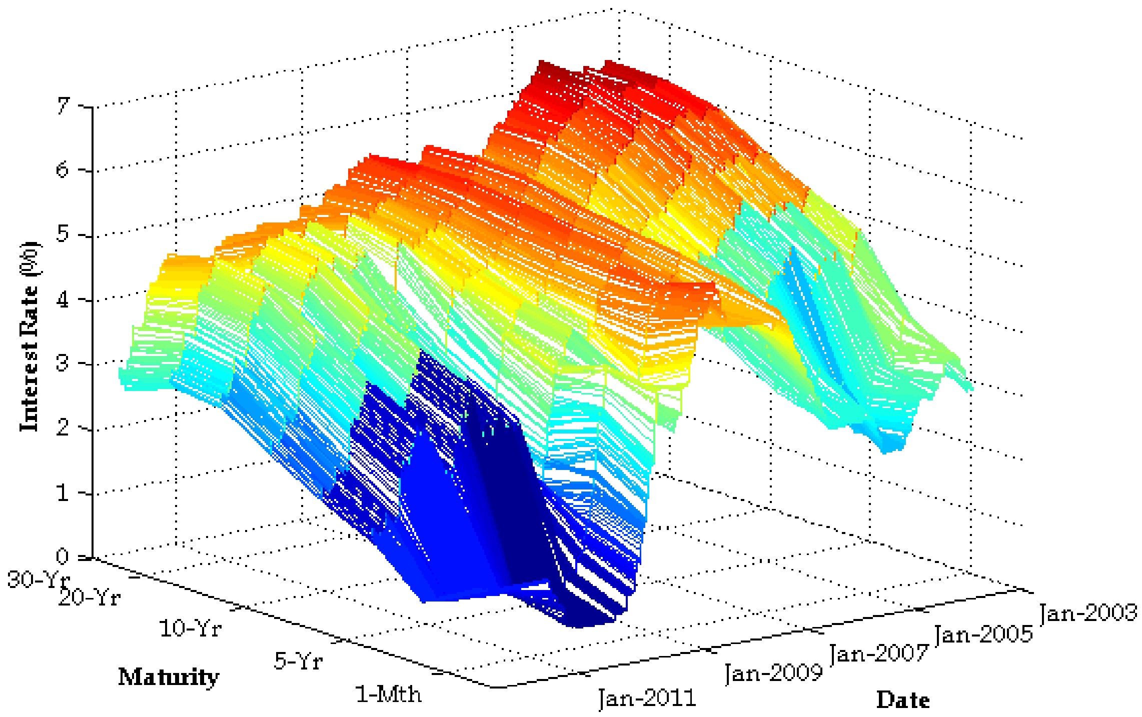

Figure 4: US Term Structure

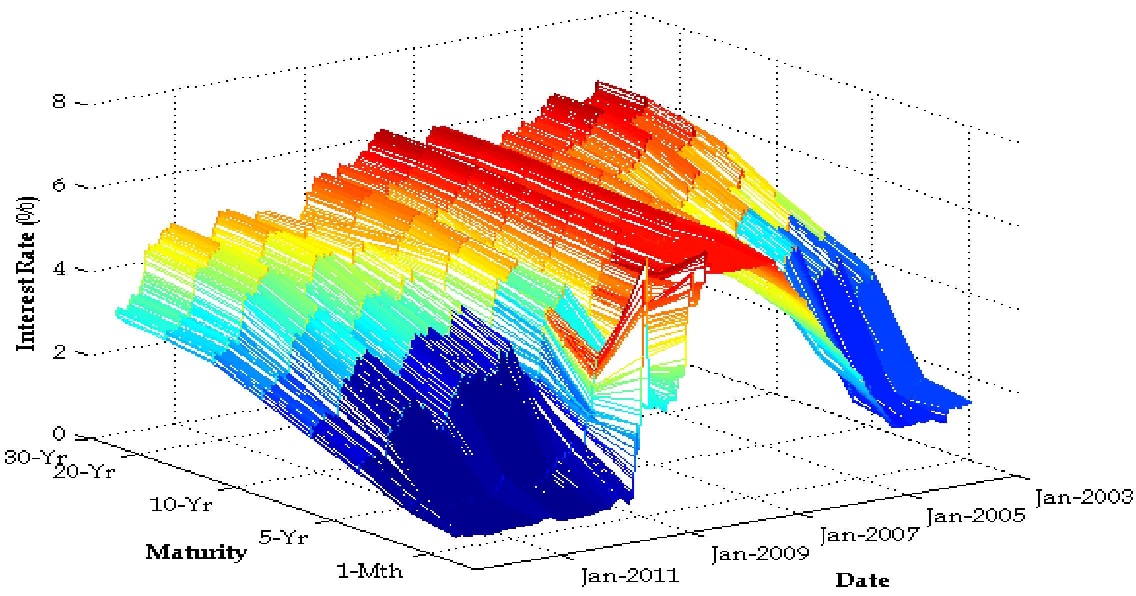

Figure 5: Factor Loadings for the Canadian Rates

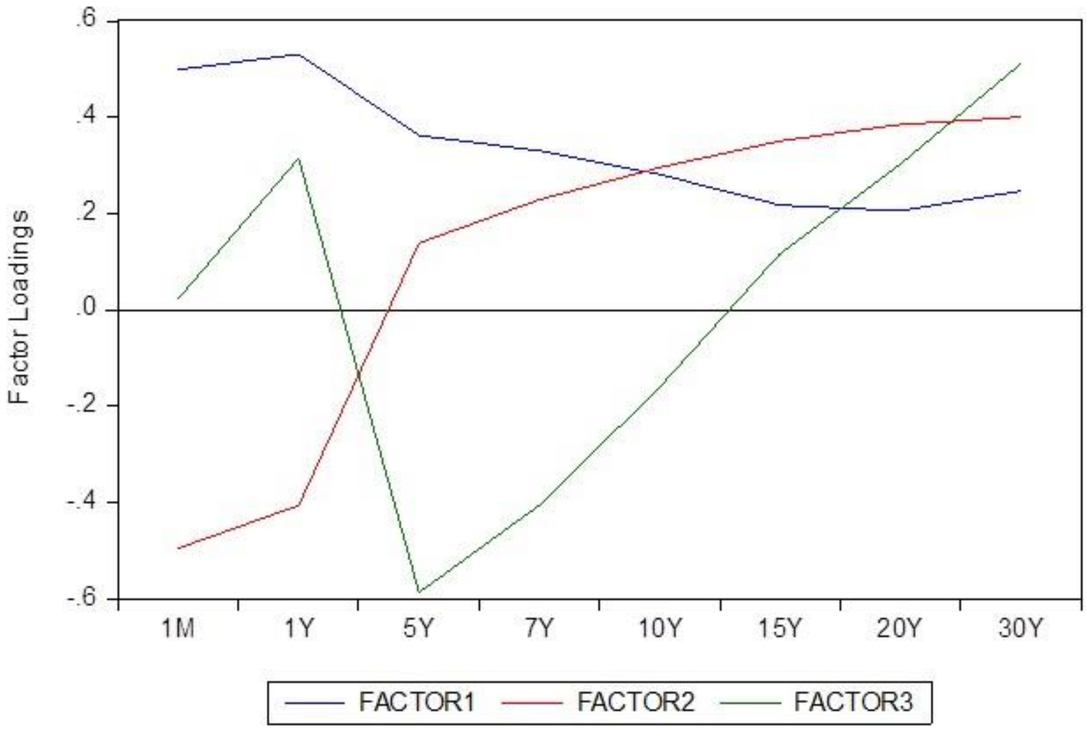

Figure 6: Factor Loadings for the US Rates

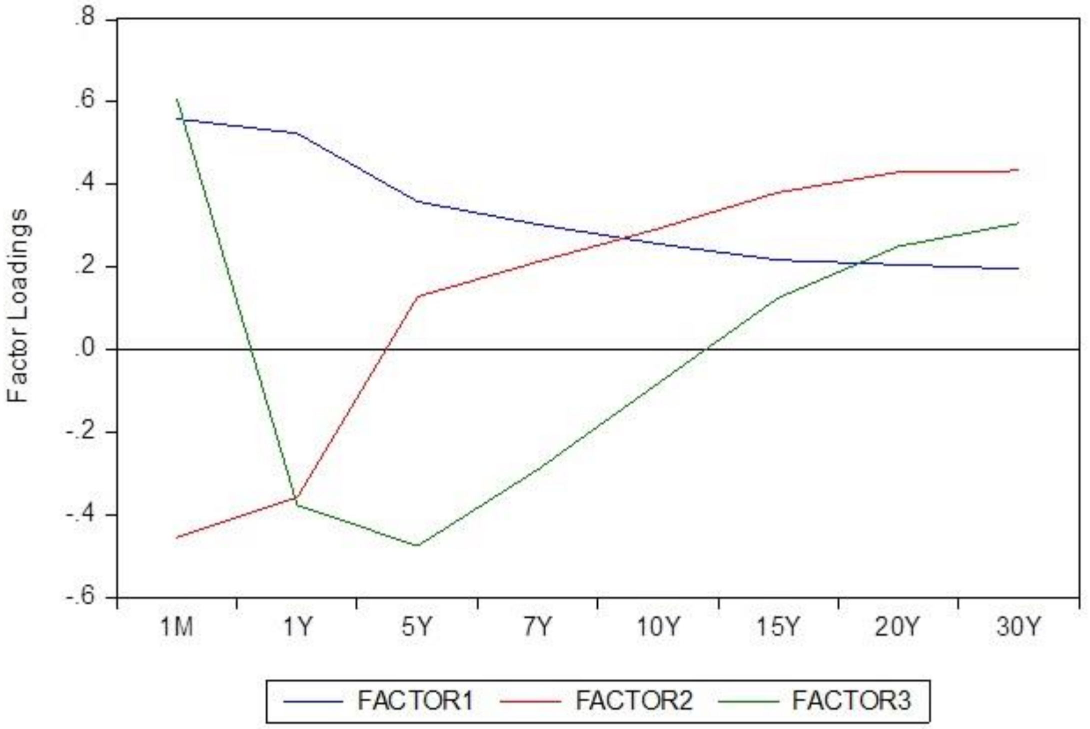

Figure 7: Two-Factor Model Factor Loadings (Maturity Years) for the Canadian Rates

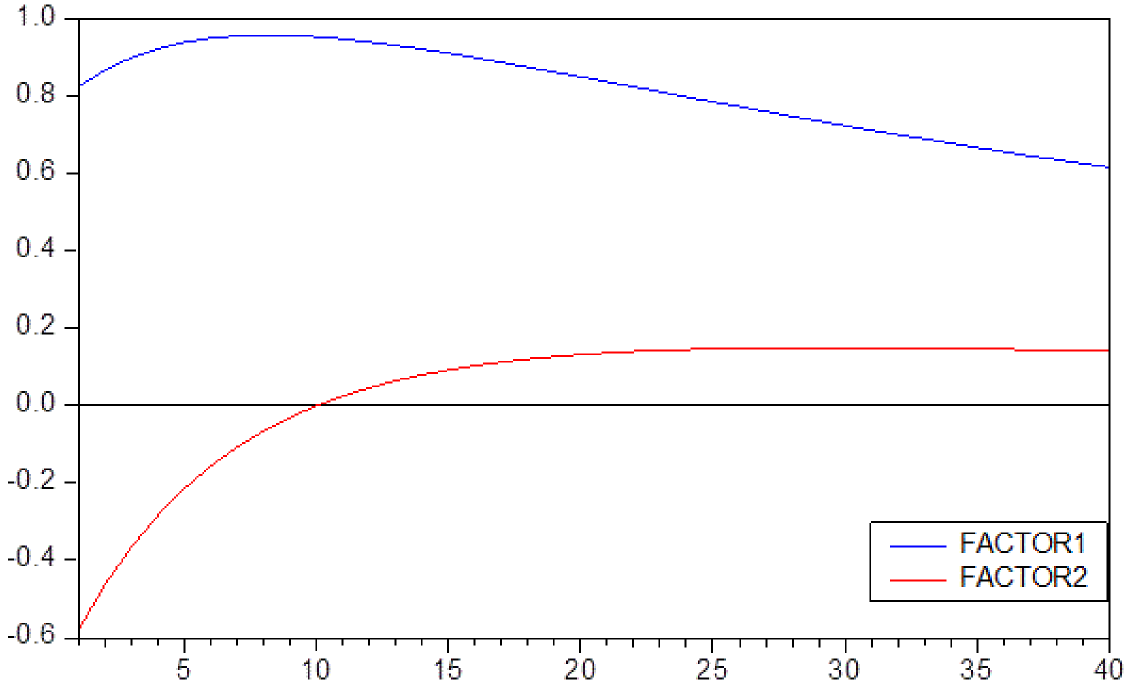

Figure 8: Two-Factor Model Factor Loadings (Maturity Years) for the US Rates