Pricing Interest Rate Caps in a Generalized Affine Model with Stochastic Volatility and Correlation: Empirical Evidence

Andrey Baliakin[a], Alexei Egorov[b], and Haitao Li[c]

December, 2005

[a]Department of Economics, Cornell University, Ithaca, NY 14853, USA

[b]Division of Economics and Finance,West Virginia University, Morgantown, WV 26506, USA

[c]J Stephen M. Ross School of Business, University of Michigan, Ann Arbor, MI 48109, USA

Abstract

We provide a comprehensive empirical analysis of several generalized affine models with unspanned stochastic volatility and stochastic correlation. Unspanned stochastic volatility represents an important feature of the data generating process that volatility in bond prices cannot be spanned by bond prices alone. Stochastic correlation across different maturities is modeled as Ornstein-Uhlenbeck random field with Markov switching between two regimes. We compare these models along with the parsimonious string model with constant volatility and constant correlation in bond yields in terms their pricing interest rate caps. We find that the parsimonious string model with constant volatility and correlation of bond yields cannot price caps well. Stochastic volatility provides the most significant improvement in model performance and stochastic correlation is also important. The advantage of our implementation of the generalized affine models is that we do not sacrifice the important feature of the models - their infinite factor structure. It can be straightforwardly extended to incorporate multiple sources of volatility, different structure for stochastic correlation as well as jumps. It also shows that the generalized affine models are very flexible and as tractable as the traditional affine models with additional advantages of being infinite factor models.

JEL classification: C4, C5, G1

Keywords: interest rate derivatives, stochastic volatility, stochastic correlation, random field, affine models of term structure.

Corresponding Author: Alexei Egorov, phone: (304)293-7868, email: avegorov@mail.wvu.edu

Pricing Interest Rate Caps in a Generalized Affine Model with Stochastic Volatility and Correlation: Empirical Evidence

Abstract

We provide a comprehensive empirical analysis of several generalized affine models with unspanned stochastic volatility and stochastic correlation. Unspanned stochastic volatility represents an important feature of the data generating process that volatility in bond prices cannot be spanned by bond prices alone. Stochastic correlation across different maturities is modeled as Ornstein-Uhlenbeck random field with Markov switching between two regimes. We compare these models along with the parsimonious string model with constant volatility and constant correlation in bond yields in terms their pricing interest rate caps. We find that the parsimonious string model with constant volatility and correlation of bond yields cannot price caps well. Stochastic volatility provides the most significant improvement in model performance and stochastic correlation is also important. The advantage of our implementation of the generalized affine models is that we do not sacrifice the important feature of the models - their infinite factor structure. It can be straightforwardly extended to incorporate multiple sources of volatility, different structure for stochastic correlation as well as jumps. It also shows that the generalized affine models are very flexible and as tractable as the traditional affine models with additional advantages of being infinite factor models.

JEL classification: C4, C5, G1

Keywords: interest rate derivatives, stochastic volatility, stochastic correlation, random field, affine models of term structure.

1 Introduction

Interest rate caps and swaptions collectively represent the largest class of interest rate options. They are widely used by firms for managing interest rate risk. They are also the most liquid over-the-counter interest rate derivatives traded. Indeed, according to the Bank of International Settlement, by the end of 2001 the combined notional values of interest rate caps and swaptions was well over 10 trillion dollars. This notional value is many times larger than that of comparable exchange traded interest rate derivatives. Consequently, accurate and efficient pricing of caps is an important topic for academic research. As pointed out by Dai and Singleton (2003), there is also an "enormous potential for new insights from using derivatives data in (dynamic term structure) model estimations."

Despite the fact that these markets are so voluminous, the majority of the existing literature uses only at-the-money (ATM) caps and swaptions. The current caps and swaptions pricing literature has mainly focused on two issues.[1] The first issue is the so-called "unspanned stochastic volatility" puzzle documented by Collin-Dufresne and Goldstein (2002) and Heidari and Wu (2003) (see also Fan, Gupta, and Ritchken (2002), Li, and Zhao (2006), Casassus, Collin-Dufresne, and Goldstein (2005)). The "unspanned stochastic volatility" puzzle is that there appear to be risk factors that drive caps and swaptions prices not spanned by the factors explaining LIBOR or swap rates. The second issue is the relative pricing between caps and swaptions. A number of recent papers, including Longstaff, Santa-Clara, Schwartz (2001) and Jagannathan, Kaplin, Sun (2003), find significant and systematic mispricing between caps and swaptions using various multi-factor term structure models.

There are very few studies documenting the relative pricing of caps with different strike prices. Jarrow, Li and Zhao (2006) provide a comprehensive documentation of volatility smile in the caps market and develop a multifactor LIBOR model with stochastic volatility and

[1] For a review of the current term structure literature, see Dai and Singleton (2002, 2003).

jumps.[2] Their three factor model can capture volatility smile when significant negative jumps are allowed. Casassus, Collin-Dufresne and Goldstein (2005) develop a model where the drift and quadratic variation in the short rate affine in three state variables (the short rate, its long term mean and variance) while the bond prices are exponential affine functions of only two state variables, independent of the current interest rate volatility level. They fit their model to the cross section of cap prices for a one day and show that their model can capture volatility smile reasonably well.[3] In contrast, the attempt to capture the volatility smile in equity option markets is voluminous and it has been the driving force behind the development of the equity option pricing literature for the past quarter of a century (see Bakshi, Cao, and Chen (1997) and references therein).[4] Analogously, studying caps with different strike prices seems promising in delivering new insights about existing term structure models and developing superior models.

We use the same date set as in Li and Zhao (2006) and Jarrow, Li and Zhao (2006). It comprises more than two years of daily cap price data with different strikes (from August 1, 2000 to November 2, 2002) from SwapPX.[5] We study the importance of stochastic volatility and stochastic correlation of the pricing interest caps. Our data set contains rich cross-sectional information. For example, we have deep ITM and OTM caps with ten different strike prices and fifteen different maturities ranging from six months to ten years.

The literature on term structure of interest rates is currently dominated by two different

[2] Several studies have provided anecdotal evidence for the existence of a volatility smile in interest rate caps and have developed theoretical models to capture this phenomenon. See Hull and White (2000), Andersen and Andreasen (2000), Andersen and Brotherton-Ratcliffe (2001), and Glasserman and Kou (2003).

[3] Gupta and Subrahmanyam (2005) also consider caps with different strikes. As shown below, however, their data is more limited than that used herein. They also test different term structure models.

[4] For reviews of the equity option literature, see Duffie (2002) and Campbell, Lo and MacKinlay (1997).

[5] Jointly developed by GovPX and Garban-ICAP, SwapPX is the first widely distributed service delivering 24 hour real-time rates, data and analytics for the world-wide interest rate swaps market. GovPX was established in early 1990s by the major U.S. fixed-income dealers as a response to regulators' demands to increase the transparency of the fixed-income markets. It aggregates quotes from most of the largest fixed-income dealers in the world. Garban-ICAP is the world's leading swap broker specializing in trades between dealers and between dealers and large customers. According to Harris (2003), "Its securities, derivatives, and money brokerage businesses have daily transaction volumes in excess of 200 billion dollars".

frameworks. The first framework is originated by Vasicek (1977) and extended among others by Cox, Ingersoll and Ross (1985). It assumes that a finite number of latent factors drive the whole dynamics of term structure. Dai and Singleton (2003) provide an extensive review of the theory, estimation and performance of dynamic term structure models (DTSMs). The prominent class of DTSMs was introduced by Duffie and Kan (1996) and has recently become the dominant framework for modeling term structure of interest rates. Affine models assume that the spot rate, the risk neutral drift and instantaneous covariance matrix of the state vector are linear in the state vector. There are many advantages of using affine models. Bond prices have a simple exponential-affine structure. Analytic solutions exist for the prices of many fixed income derivatives, such as options on zero coupon bonds. However, the tractability of affine models comes at the potential cost of limiting its flexibility in explaining empirical observations. Jaganathan, Kaplin and Sun (2003) find that low dimensional affine models are unable to capture the joint dynamics of caps, swaptions and bonds. Affine models quickly become intractable as the number of factors increases. The other framework comprises curve models. These models are calibrated to the relevant forward curve. There are three groups of such models: forward rate models pioneered by Heath, Jarrow and Morton (1992), LIBOR market models developed by Brace, Garatek and Musiela (1997) and Miltersen, Sandmann and Sondermann (1997), and random field models introduced by Kennedy (1994,1997), and further developed by Goldstein (2000), Santa Clara and Sornette (2001) and Kimmel (2004).

In this paper, a generalized affine model with stochastic volatility and correlation (SCSV) is implemented and used to price interest rate caps given LIBOR rates. This framework was introduced by Collin-Dufresne and Goldstein (2003). The advantage of the generalized affine framework is that it circumvents the limitations of the finite-dimensional models while retaining the tractability of the traditional affine class of models. The log-bond prices themselves are considered as state variables, hence the state vector is of infinite dimension as it includes all bonds $P^T(t)$ for the continuum of maturities $T > t$, and all the state variables that drive innovations in volatility and correlation structure. Closed form solutions are readily obtained for many fixed income derivatives. The model can be calibrated to fit the initial term structures of both interest rates and volatility. Finally, generalized affine framework naturally includes both unspanned stochastic volatility and stochastic correlation. In particular, for a "switching" correlation structure estimated in this paper there are either closed form solutions or a system of ODEs can be solved numerically. All that makes generalized affine models a superior framework for pricing fixed income derivatives.

Despite all this attractive features of generalized affine models, no general implementation technique is developed and no comprehensive testing of these models has been done so far. This paper closes this gap in the literature. We provide the implementation procedure that actually delivers accurate and computationally efficient estimation of generalized affine models. In analyzing the generalized affine model, we reach the following conclusions. First, the simple version of the generalized affine model, the parsimonious string model, has large pricing errors and performs poorly. Second, its more advanced versions have much smaller pricing errors and thus can generate a volatility smile. We describe our estimation techniques and show how certain implementation difficulties can be overcome providing an accurate and computationally efficient procedure. We show that stochastic volatility is especially important for capturing volatility smile. Stochastic correlation is also useful. Although all considered models still cannot capture the entire smile. More research is warranted in particular multiple sources of stochastic volatility might significantly improve the fit of a model.

The rest of this paper is organized as follows. In Section 2, we introduce a "generalized affine" model with unspanned stochastic volatility and stochastic correlation and provide its important special cases. In Section 3, we describe the data set and our estimation procedure. In Section 4, we provide the details of computational implementation of our procedure. In section 5 we discuss the estimation results. Section 6 concludes.

2 A ‘Generalized Affine’ Model with Unspanned Stochastic Volatility and Stochastic Correlation (SCSV)

We consider here the most general "generalized affine" model with unspanned stochastic volatility and stochastic correlation (SCSV). The specialized models with stochastic correlation (SC), unspanned stochastic volatility (SV) and the parsimonious string model (PSM) are derived as reductions of the SCSV model.

In the SCSV model the risk-neutral bond price P^T dynamics are assumed to follow

$$\frac{dP^T(s)}{P^T(s)} = r_s\, ds - \sigma^T(s)\, dZ_Q^T(s) - B^T(s)\sqrt{\Sigma(s)}\, d\omega_Q^1(s), \tag{1}$$

where T is a maturity date, s is a current date, r_s, $\sigma^T(s)$ and $B^T(s)$ are arbitrary deterministic functions. The volatility state variable $\Sigma(s)$ follows the diffusion

$$d\Sigma(s) = \kappa\left(\theta - \Sigma(s)\right) ds + \vartheta\sqrt{\Sigma(s)}\left(\nu d\omega_Q^1(s) + \sqrt{1-\nu^2} d\omega_Q^2(s)\right), \tag{2}$$

Brownian field $dZ_Q^T(s)$ and two Brownian motions $d\omega_Q^1(s)$ and $d\omega_Q^2(s)$ are assumed to be mutually independent and have the following correlation structure

$$dZ_Q^T(s) dZ_Q^U(s) \equiv c(s, U, T, \rho)\, ds. \tag{3}$$

We refer to Kennedy (1994, 1997), Goldstein (2000) and Santa-Clara and Sornette (2001) for Brownian fields application in fixed income literature. In particular different functional forms for the correlation function $c(s, U, T, \rho)$ were proposed. Note that in contrast to the standard affine framework, generalized affine framework does not have a finite state representation and, moreover, instantaneous correlation matrix has full rank for any number of bonds considered. Correlation structure $c(s, T, U, \rho)$ is driven by a single state variable ρ that follows a continuous time two-state Markov chain with transition probabilities given by

$$\begin{aligned} P\big(\rho_{s+ds} = \rho_H \,\big|\, \rho_s = \rho_L\big) &= p_{LH}\, ds, \\ P\big(\rho_{s+ds} = \rho_L \,\big|\, \rho_s = \rho_H\big) &= p_{HL}\, ds. \end{aligned}$$

Correlation may jump between deterministic forms $c_L(s,T,U)$ and $c_H(s,T,U)$ independently from $dZ_Q^T(s)$, $d\omega_Q^1(s)$ and $d\omega_Q^2(s)$.

The date-t price of a European bond-option with exercise date T and underlying zero-coupon bond maturing at date U can be written as

$$\begin{aligned} C(t,T,U) &= P^U(t)\,\Pi_t^U(\log K) - KP^T(t)\,\Pi_t^T(\log K), \\ \Pi_t^W(k) &= E_t^W\left[\mathbf{1}_{\log P^U(T)>K}\right]. \end{aligned}$$

If the characteristic function of the random variable $\log P^U(T)$ under the W-forward neutral measure is

$$G_t^W(i\lambda) = E_t^W\left[e^{i\lambda \log P^U(t)}\right], \tag{4}$$

then it follows from Fourier inversion theorem that

$$\begin{aligned} C(t,T,U) &= P^U(t)\left(\frac{1}{2}+\frac{1}{\pi}\int_0^\infty d\lambda\,\mathrm{Re}\left(\frac{e^{-i\lambda\log K}G_t^U(i\lambda)}{i\lambda}\right)\right) \\ &\quad -KP^T(t)\left(\frac{1}{2}+\frac{1}{\pi}\int_0^\infty d\lambda\,\mathrm{Re}\left(\frac{e^{-i\lambda\log K}G_t^T(i\lambda)}{i\lambda}\right)\right) \end{aligned} \tag{5}$$

If $G_t^W(i\lambda)$ can be written in closed-form, then so can the bond-option price, for any specifications of the deterministic functions $B^T(s)$, $\sigma^T(s)$, $c_H(s,T,U)$, $c_L(s,T,U)$.

We cite here Proposition 4 proven by Collin-Dufesne and Goldstein (2002) for their special case $m=1$.[6]

Proposition 1 *The characteristic function of the log-bond price takes the form*

$$G_s^W(\lambda) = e^{\lambda\log\left(P^{T_1}(s)/P^{T_0}(s)\right)+M_S(s)+N(s)\Sigma(s)}, \quad S\in\{L,H\},$$

where the deterministic function $N(s,W)$ satisfies the 'final condition' $N(T_0,W)=0$ and the

[6]In this paper we are studying pricing of interest rate caps. For this purpose it is enough for us to use Proposition 4 from Collin-Dufesne and Goldstein (2002) only for their case when $m=1$. Below we made an extension of this proposition also for this special case. We note though that similar extensions hold for $m>1$. This is important in pricing such derivatives as swaptions which however is not a focus of this paper.

ODE

$$
\begin{aligned}
-\frac{\partial}{\partial s}N(s,W) &= \frac{\vartheta^2}{2}N^2(s,W) - N(s,W)\left(\kappa + \nu\vartheta\left(B^W(s) + \lambda\left(B^{T_1}(s) - B^{T_0}(s)\right)\right)\right) \\
&\quad + \frac{\lambda}{2}\left(B^{T_1}(s) - B^{T_0}(s)\right)\left(2B^W(s) - B^{T_1}(s) - B^{T_0}(s)\right) + \frac{\lambda^2}{2}\left(B^{T_1}(s) - B^{T_0}(s)\right)^2,
\end{aligned}
$$

while $M_L(s,W)$ and $M_H(s,W)$ satisfy the 'final condition' $M_L(T_0,W) = M_H(T_0,W) = 0$ and the coupled ODEs

$$
\begin{aligned}
-\frac{\partial}{\partial s}M_{S_1}(s,W) &= \lambda\Big(\sigma^W(s)\sigma^{T_1}(s)c_{S_1}(s,T_1,W) - \frac{1}{2}\left(\sigma^{T_1}(s)\right)^2 \\
&\quad - \sigma^W(s)\sigma^{T_0}(s)c_{S_1}(s,T_0,W) + \frac{1}{2}\left(\sigma^{T_0}(s)\right)^2\Big) \\
&\quad + \frac{\lambda^2}{2}\left(\left(\sigma^{T_1}(s)\right)^2 + \left(\sigma^{T_0}(s)\right)^2 - 2\sigma^{T_0}(s)\sigma^{T_1}(s)c_{S_1}(s,T_0,T_1)\right) \\
&\quad + \kappa\theta N(s,W) + p_{S_1S_2}\left(e^{M_{S_2}(s,W) - M_{S_1}(s,W)} - 1\right),
\end{aligned}
$$

where S_1 and S_2 are correlation states, $(S_1,S_2) \in \{(H,L),(L,H)\}$.

For the $W = T_j$, $j \in \{0,1\}$ these ODEs can be written as

$$
\begin{aligned}
-\frac{\partial}{\partial s}N(s,T_j) &= \frac{\vartheta^2}{2}N^2(s,T_j) - N(s,T_j)\left(\kappa + \nu\vartheta\left(B^{T_j}(s) + \lambda\left(B^{T_1}(s) - B^{T_0}(s)\right)\right)\right) \\
&\quad + \frac{\lambda}{2}\left(\lambda + (-1)^{j+1}\right)\left(B^{T_1}(s) - B^{T_0}(s)\right)^2,
\end{aligned}
\tag{6}
$$

and

$$
\begin{aligned}
-\frac{\partial}{\partial s}M_{S_1}(s,T_j) &= \frac{\lambda}{2}\left(\lambda + (-1)^{j+1}\right)\left(\sigma_{f,S}^{T_0,T_1}(s)\right)^2 \\
&\quad + \kappa\theta N(s,T_j) + p_{S_1S_2}\left(e^{M_{S_2}(s,T_j) - M_{S_1}(s,T_j)} - 1\right),
\end{aligned}
\tag{7}
$$

where

$$
\left(\sigma_{f,S}^{T_0,T_1}(s)\right)^2 = \left(\sigma^{T_0}(s)\right)^2 - 2\sigma^{T_0}(s)\sigma^{T_1}(s)c_S(s,T_0,T_1) + \left(\sigma^{T_1}(s)\right)^2.
$$

As we show below, pricing of interest rate caps improves when we allow for correlation parameter ν. For this purpose we now extend the Proposition 2 proven to the correlation ν.

Proposition 2 *If the volatility structure for zero-coupon bonds is assumed to be of the Gaussian-Vasicek type*

$$B^T(s) = \frac{1}{\beta_B}\left(1 - e^{-\beta_B(T-s)}\right),$$

then (6) possesses a unique solution given by

$$N(s,T_j) = \frac{\beta_B e^{-\beta_B(T_0-s)} w_j'(e^{-\beta_B(T_0-s)})}{c_2 w_j(e^{-\beta_B(T_0-s)})}, \tag{8}$$

where

$$w_j(u) = e^{hu}\left(\alpha\Phi(\tilde{a},\tilde{b};\eta u) + (\eta u)^{1-\tilde{b}}\,\Phi(\tilde{a}-\tilde{b}+1, 2-\tilde{b};\eta u)\right),$$

$\Phi(\tilde{a},\tilde{b};\xi)$ *denotes the Kummer (or confluent hypergeometric) function* ${}_1F_1(\tilde{a},\tilde{b};\xi)$*,*

$$\alpha = \eta^{1-\tilde{b}} \frac{h\Phi\left(\tilde{a}-\tilde{b}+1, 2-\tilde{b};\eta\right) + \eta\Phi_3'\left(\tilde{a}-\tilde{b}+1, 2-\tilde{b};\eta\right) + \left(1-\tilde{b}\right)\Phi\left(\tilde{a}-\tilde{b}+1, 2-\tilde{b};\eta\right)}{h\Phi\left(\tilde{a},\tilde{b};\eta\right) + \eta\Phi_3'\left(\tilde{a},\tilde{b};\eta\right)},$$

$$\begin{aligned}
h &= \frac{D-a}{2\beta_B}, \\
\tilde{a} &= \frac{(\beta_B + c_1)(D-a)}{2\beta_B D}, \\
\tilde{b} &= 1 + \frac{c_1}{\beta_B}, \\
\eta &= -\frac{D}{\beta_B}, \\
D &= \sqrt{a^2 - 4c_0c_2},
\end{aligned}$$

and

$$\begin{aligned}
c_2 &= \frac{\vartheta^2}{2}, \\
c_1 &= -\kappa - \frac{\nu\vartheta}{\beta_B}, \\
a &= \nu\vartheta\left(\frac{e^{-\beta_B(T_j-T_0)}}{\beta_B} - \lambda b^\delta_{\beta_B}\right), \\
c_0 &= \frac{\lambda\left(\lambda - (-1)^j\right)}{2}\left(b^\delta_{\beta_B}\right)^2, \\
b^\delta_\beta &= \frac{1 - e^{-\beta_B\delta}}{\beta_B}.
\end{aligned}$$

For the special case $\nu = 0$ ODE (6) is reduced to the ODE

$$-\frac{\partial}{\partial s}N(s,T_j) = \frac{\vartheta^2}{2}N^2(s,T_j) - \kappa N(s,W) + \frac{\lambda}{2}\left(\lambda + (-1)^{j+1}\right)\left(B^{T_1}(s) - B^{T_0}(s)\right)^2. \qquad (9)$$

If we replace β_B with β in (8), than it will coincide with the solution obtained by Collin-Dufesne and Goldstein (2003):

$$N(s,T_j) = \frac{1}{\vartheta^2}\left(\kappa + e^{-\beta_B \tau}\sqrt{2\phi}\frac{\alpha J'_{\frac{\kappa}{2\beta_B}}\left(\frac{e^{-\beta_B\tau}}{\beta_B}\sqrt{\phi/2}\right) + Y'_{\frac{\kappa}{2\beta_B}}\left(\frac{e^{-\beta_B\tau}}{\beta_B}\sqrt{\phi/2}\right)}{\alpha J_{\frac{\kappa}{2\beta_B}}\left(\frac{e^{-\beta_B\tau}}{\beta_B}\sqrt{\phi/2}\right) + Y_{\frac{\kappa}{2\beta_B}}\left(\frac{e^{-\beta_B\tau}}{\beta_B}\sqrt{\phi/2}\right)}\right), \qquad (10)$$

where J and Y are Bessel functions, $\tau = T_0 - s$,

$$\begin{aligned}
\phi &= \frac{\lambda}{2}\left(\lambda + (-1)^{j+1}\right)\frac{\vartheta^2}{\beta_B^2}\left(1 - e^{-\beta_B(T_1 - T_0)}\right)^2 e^{-2\beta_B\tau}, \\
\alpha &= -\frac{\kappa Y_{\frac{\kappa}{2\beta_B}}\left(\frac{1}{\beta_B}\sqrt{\phi/2}\right) + \sqrt{2\phi}Y'_{\frac{\kappa}{2\beta_B}}\left(\frac{1}{\beta_B}\sqrt{\phi/2}\right)}{\kappa J_{\frac{\kappa}{2\beta_B}}\left(\frac{1}{\beta_B}\sqrt{\phi/2}\right) + \sqrt{2\phi}J'_{\frac{\kappa}{2\beta_B}}\left(\frac{1}{\beta_B}\sqrt{\phi/2}\right)}.
\end{aligned}$$

2.1 Stochastic correlation model (SC)

To obtain the model with only stochastic correlation we eliminate the stochastic volatility component from the SCSV model by assuming $B^T(s) \equiv 0$. This can be obtained in the limit as $\beta_B \to \infty$. The risk-neutral bond price dynamics (1) are reduced to

$$\frac{dP^T(s)}{P^T(s)} = r_s\, ds - \sigma^T(s)\, dZ_Q^T(s), \qquad (11)$$

while the system of coupled ODEs (7) remains intact, and ODE (6) simplifies to

$$-\frac{\partial}{\partial s}N(s,T_j) = \frac{\vartheta^2}{2}N^2(s,T_j) - \kappa N(s,T_j). \qquad (12)$$

ODE (12) possesses the unique solution $N(s,T_j) \equiv 0$. This further simplifies this case.

2.2 Stochastic volatility model (SV)

We obtain the stochastic volatility model if we eliminate stochastic correlation from the SCSV model by assuming $\rho_H = \rho_L = \rho$. This restriction does not change ODE (6). From coupled

ODEs (7) it follows that

$$\frac{\partial}{\partial s}\left(M_H(s,T_j) - M_L(s,T_j)\right) = p_{LH}\left(e^{M_H(s,T_j)-M_L(s,T_j)} - 1\right) - p_{HL}\left(e^{M_L(s,T_j)-M_H(s,T_j)} - 1\right), \tag{13}$$

hence for any p_{LH} and p_{HL} we can conclude that $M_H(s,T_j) = M_L(s,T_j)$ for any $s \leq T_0$. Therefore

$$-\frac{\partial}{\partial s}M(s,T_j) = \kappa\theta N(s,T_j) + \frac{\lambda}{2}\left(\lambda + (-1)^{j+1}\right)\left(\sigma_f^{T_0,T_1}(s)\right)^2.$$

From (8) it follows that

$$N(s,T_j) = \frac{2}{\vartheta^2}\frac{\partial}{\partial s}\log w_j(e^{-\beta_B(T_0-s)}), \tag{14}$$

hence integration from t to T_0 and boundary conditions $M(T_0,T_j) = N(T_0,T_j) = 0$ yield

$$M(s,T_j) = -\frac{2\kappa\theta}{\vartheta^2}\log w_j(e^{-\beta_B(T_0-s)}) + \frac{\lambda}{2}\left(\lambda + (-1)^{j+1}\right)\Omega(s,T_0,T_1), \tag{15}$$

where

$$\Omega(s,T_0,T_1) = \int_s^{T_0}\left(\sigma_f^{T_0,T_1}(t)\right)^2 dt.$$

2.3 Parsimonious 'string' model (PSM)

The parsimonious string model can be obtained from the SV model in a limiting case $B^T(s) \equiv 0$. This is done by setting $\beta_B \to \infty$. This framework has been empirically investigated by Longstaff, Santa Clara and Schwartz (2001). They demonstrated the practical nature of this framework. Though both paper approximate their models with low dimentional factor approximations. This significantly simplifies calibration of the model, although significant advantage of infinite factor specification of PSM model (and the whole class of generalized affine models) has been sacrificed. We, however, do not make here this approximation and rather study the original infinite version of PSM as we do for all other models in this paper. The risk-neutral bond price dynamics is reduced to

$$\frac{dP^T(s)}{P^T(s)} = r_s\,ds - \sigma^T(s)\,dZ_Q^T(s). \tag{16}$$

In that case

$$
\begin{aligned}
N(s,T_j) &= 0, \\
M(s,T_j) &= \frac{\lambda}{2}\left(\lambda+(-1)^{j+1}\right)\Omega(t,T_0,T_1),
\end{aligned}
\tag{17}
$$

hence from (5) we have

$$
\begin{aligned}
C(t,T_0,T_1) &= P^{T_0}(t)\left(\frac{1}{2}+\frac{1}{\pi}\int_0^\infty \mathrm{Re}\left(\frac{e^{-i\lambda\mu+\frac{i\lambda}{2}(i\lambda+1)\Omega}}{i\lambda}\right)d\lambda\right) \\
&\quad -KP^{T_1}(t)\left(\frac{1}{2}+\frac{1}{\pi}\int_0^\infty \mathrm{Re}\left(\frac{e^{-i\lambda\mu+\frac{i\lambda}{2}(i\lambda-1)\Omega}}{i\lambda}\right)d\lambda\right) \\
&= P^{T_0}(t)\left(\frac{1}{2}+\frac{1}{\pi}\int_0^\infty e^{-\frac{\lambda^2}{2}\Omega}\frac{\sin\left(-\lambda\left(\mu-\frac{\Omega}{2}\right)\right)}{\lambda}d\lambda\right) \\
&\quad -KP^{T_1}(t)\left(\frac{1}{2}+\frac{1}{\pi}\int_0^\infty e^{-\frac{\lambda^2}{2}\Omega}\frac{\sin\left(-\lambda\left(\mu+\frac{\Omega}{2}\right)\right)}{\lambda}d\lambda\right) \\
&= P^{T_0}(t)N\left(\frac{-\mu+\frac{1}{2}\Omega}{\sqrt{\Omega}}\right)-KP^{T_1}(t)N\left(\frac{-\mu-\frac{1}{2}\Omega}{\sqrt{\Omega}}\right),
\end{aligned}
$$

where $\Omega=\Omega(t,T_0,T_1)$, $T_1=T_0+\delta$ and

$$
\mu=\log\left(\frac{\tilde{K}P^{T_1}(t)}{P^{T_0}(t)}\right)=\log\left((1+\delta K)\,P_f^{T_0,T_1}(t)\right).
$$

The indexed price of a caplet given by

$$
\frac{\mathrm{Cpl}(t,T_0,T_1)}{P^{T_0}(t)}=N\left(\frac{-\mu+\frac{1}{2}\Omega(t,T_0,T_1)}{\sqrt{\Omega(t,T_0,T_1)}}\right)-e^{\mu}N\left(\frac{-\mu-\frac{1}{2}\Omega(t,T_0,T_1)}{\sqrt{\Omega(t,T_0,T_1)}}\right).
\tag{18}
$$

is now a function of μ and $\Omega(t,T_0,T_1)$ only.

3 Estimation

For estimation purposes we need to make additional assumptions on the behavior of the volatility $\sigma^T(s)$ and correlation structure $c_S(s,T,U)$. We assume here that

$$
\sigma^T(s)=\frac{\sigma}{\beta}\left(1-e^{-\beta(T-s)}\right),
$$

and correlation structure in state $S \in \{H, L\}$ is generated by an integrated Ornstein-Uhlenbeck process

$$c_S(s,T,U) = e^{-\rho_S|T-U|}\left(1 + \rho_S\,|T-U|\right).$$

Note that, of course, more general specifications may improve our empirical results. Further research is warranted to see the pricing of interest rate caps with other specifications. In general the task of finding the best specifications for $\sigma^T(s)$ and $c_S(s,T,U)$ seems to be daunting given enormous possible behavior dynamics for the volatility and correlation in general. Though we believe that this fact should rather inspire future research. While for the purpose of our paper we restrict ourselves to these specifications since as we show below they render the main conlusion of our paper that stochastic volatility and stochastic correlation are important for pricing ITM and OTM interest rate caps suffices and models that have them price interest rate caps better that the basic PSM model .

3.1 Data set

Our data set contains caplet market prices $\widetilde{\mathrm{Cpl}}(t, t+m_{i-1}\delta, t+m_i\delta, K_j(t))$, where $\delta = 1/4$ corresponds to 3 month period, are maturity multipliers m_i are taken from the vector $m = (2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 24, 28, 32, 36, 40)$. These values of m correspond to 6 month, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5, 6, 7, 8, 9 and 10 year maturities respectively. Date t belongs to the set T of $n = 557$ trading days between 08/01/00 and 11/07/02. For each day t, 10 different strikes $K_j(t)$ are available in the form of annual rates. Caplet prices are censored to satisfy no-arbitrage conditions and outliers are removed. Let us denote by $T_{i,j}$ the subset of all dates t such that caplet price with tenor $[m_{i-1}, m_i]$ and strike $K_j(t)$ was not ruled out.

Our data set also contains LIBOR rates $L(t, t+l_{i-1}\delta, t+l_i\delta)$ for the periods $[l_{i-1}\delta, l_i\delta]$ observed at dates $t \in T$, where maturity multipliers l_i belong to the vector $l = (0, 1, 2, 4, 8, 12, 16, 20, 28, 40)$.

It is standard to define LIBOR rate as

$$
\begin{aligned}
L(t+l_{i-1}\delta, t+l_i\delta) &= \frac{1}{\delta\left(l_i - l_{i-1}\right)}\left(\frac{1}{P^{t+l_i\delta}(t+l_{i-1}\delta)} - 1\right), \\
P^{t+l_i\delta}(t+l_{i-1}\delta) &= \exp\left\{-\int_{t+l_{i-1}\delta}^{t+l_i\delta} f(t,s)ds\right\}.
\end{aligned}
$$

For pricing purpose we need forward rates

$$
P_f^{t+m_{i-1}\delta, t+m_i\delta}(t) = \frac{P^{t+m_i\delta}(t)}{P^{t+m_{i-1}\delta}(t)}, \tag{19}
$$

where $P_f^{T_0,T_1}(t)$ is defined as a forward price quoted at date t and paid at date T_0 for delivery at date T_0 of a zero-coupon bond that matures at T_1.

$$
\begin{aligned}
1+\left(l_i - l_{i-1}\right)\delta L(t+l_{i-1}\delta, t+l_{i-1}\delta) &= \exp\left\{\int_{t+l_{i-1}\delta}^{t+l_i\delta} f(t,s)ds\right\} \\
&= \exp\left\{\delta\left(l_i - l_{i-1}\right)\frac{f(t,t+l_{i-1}\delta)+f(t,t+l_i\delta)}{2}\right\}
\end{aligned}
$$

3.2 Objective function

For the given subset I of tenor indices our objective function is

$$
WMSE = \frac{1}{N_I}\sum_{i\in I}\sum_{j=1}^{10}\sum_{t\in T_{i,j}}\frac{e_{i,j}^2(t)}{\hat{\sigma}_{i,j}^2(t)}, \tag{20}
$$

where $N_I = \sum_{i\in I}\sum_{j=1}^{10}\sum_{t\in T_{i,j}} 1$ is a number of caplet prices and indexed pricing error $e_{i,j}(t)$ for caplet with tenor $[m_{i-1}, m_i]$ and strike $K_j(t)$ are defined by

$$
e_{i,j}(t) = \frac{\widetilde{\mathrm{Cpl}}(t, t+m_{i-1}\delta, t+m_i\delta, K_j(t)) - \mathrm{Cpl}(t, t+m_{i-1}\delta, t+m_i\delta, K_j(t))}{P_f^{t+m_{i-1}\delta, t+m_i\delta}(t)}. \tag{21}
$$

Indexed pricing error variances.$\hat{\sigma}_{i,j}^2(t)$ are estimated nonparametrically using Bartlett kernel. First, nonparametric estimates of indexed caplet prices are obtained as

$$
\frac{\widehat{\mathrm{Cpl}}(t, t+m_{i-1}\delta, t+m_i\delta, K_j(t))}{P_f^{t+m_{i-1}\delta, t+m_i\delta}(t)} = \frac{\sum_{j'=1}^{10}\sum_{t'\in T_{i,j'}} k(\mu_{i,j}(t)-\mu_{i,j'}(t'))\frac{\widetilde{\mathrm{Cpl}}(t', t'+m_{i-1}\delta, t'+m_i\delta, K_{j'}(t'))}{P_f^{t'+m_{i-1}\delta, t'+m_i\delta}(t')}}{\sum_{j'=1}^{10}\sum_{t'\in T_{i,j}} k\left(\mu_{i,j}(t)-\mu_{i,j'}(t')\right)},
$$

where moneyness is defined by

$$\mu_{i,j}(t) = \log\left(\left(1 + \delta\left(m_{i-1} - m_i\right) K_j(t)\right) P_f^{t+m_{i-1}\delta, t+m_i\delta}(t)\right).$$

Note that $\mu_{i,j}(t)$ differs from another definition of moneyness

$$k_{i,j}(t) = \log\left(\frac{K_j(t)}{\left(1 - P_f^{t+m_{i-1}\delta, t+m_i\delta}(t)\right) / \left(m_i - m_{i-1}\right)\delta}\right),$$

but both definitions are equivalent as δ tends to zero.

Alternatively, additional smoothing dimension of maturity might be used. Here maturity specific averaging is chosen. This is done since there is a well pronounced maturity-driven shift in observed caplet market price curves plotted as functions of caplet moneyness. Under assumptions made, there should be no shifts like that.

Estimated indexed caplet pricing errors are given by

$$\hat{e}_{i,j}(t) = \frac{\widetilde{\mathrm{Cpl}}(t, t + m_{i-1}\delta, t + m_i\delta, K_j(t)) - \widehat{\mathrm{Cpl}}(t, t + m_{i-1}\delta, t + m_i\delta, K_j(t))}{P_f^{t+m_{i-1}\delta, t+m_i\delta}(t)}, \qquad (22)$$

and their estimated variances are computed as

$$\hat{\sigma}^2_{i,j}(t) = \frac{\sum_{j'=1}^{10}\sum_{t'\in T_{i,j'}} k(\mu_{i,j}(t) - \mu_{i,j'}(t'))\hat{e}^2_{i,j}(t)}{\sum_{j'=1}^{10}\sum_{t'\in T_{i,j'}} k(\mu_{i,j}(t) - \mu_{i,j'}(t'))}. \qquad (23)$$

3.3 Estimation procedure

We limit our study to the caplets with maturities ranging from 2 years up to 10 years, i.e. $I = 4, ..., 15$. One reason for that is a number of caplet prices for 6 month and 1 year maturities that satisfy no-arbitrage conditions is as half as smaller than for other maturities. Also, as figure 1 suggests, caplets with maturities shorter that 2 years do not fit well into the surface with prices of caplets with longer maturities. This might be rigorously measured by increase in MSE as additional short maturity is added to the set of longer maturities.

After variances of indexed caplet pricing errors are estimated, 40 outliers from the set of indexed caplet prices are removed for each maturity, that accounts for less than 1%. Outliers are selected by the ratio of the corresponding estimated indexed pricing error to the estimated variance of the indexed pricing error. Remaining 59153 caplet prices are used as an input to the estimation procedure.

In the PSM only time-invariant parameters $\theta \in \Theta$ need to be estimated. For SC and SCSV models, we also need to estimate a time profile of binary 0,1 correlation state latent variables. We require that sample ratio of number of low-correlation days and high-correlation days coincides with its theoretical value. In other words, if we define

$$n_S = |\{t \in T : \rho_t = \rho_S\}|$$

for $S \in \{H, L\}$, then $n_L + n_H = n$ and we limit $(\rho_t)_{t \in T}$ to the constraint set

$$R(\theta) = \left\{ (\rho_t)_{t \in T} : n_L = \left[n \frac{p_{HL}}{p_{HL} + p_{LH}} \right] \right\}.$$

For the SV and SCSV model a time profile of real-valued nonnegative stochastic volatilities $(\Sigma_t)_{t \in T}$ has to be estimated. For the given value θ of the time-invariant parameters and for each day t we minimize daily $\mathit{WMSE}_t(\theta, \rho, \Sigma_t)$ (in SV case) or both $\mathit{WMSE}_t(\theta, \rho_L, \Sigma_t)$ and $\mathit{WMSE}_t(\theta, \rho_H, \Sigma_t)$ independently (in SCSV case) with respect to Σ_t. Then the sum of $\mathit{WMSE}_t(\theta, \rho_t, \hat{\Sigma}_t)$ over entire sample is minimized with respect to $(\rho_t)_{t \in T} \in R(\theta)$. Finally, for $\mathit{WMSE}(\hat{\theta}, (\hat{\rho}_t, \hat{\Sigma}_t)_{t \in T})$ is obtained by minimization with respect to time-invariant parameters $\theta \in \Theta$. The procedure might be summarized as:

$$\begin{aligned}
\mathit{WMSE}(\hat{\theta}, (\hat{\rho}_t, \hat{\Sigma}_t)_{t \in T}) &= \min_{\theta \in \Theta} \mathit{WMSE}(\theta, (\hat{\rho}_t, \hat{\Sigma}_t)_{t \in T}), \\
\mathit{WMSE}(\theta, (\hat{\rho}_t, \hat{\Sigma}_t)_{t \in T}) &= \min_{(\rho_t)_{t \in T} \in R(\theta)} \sum_{t \in T} \mathit{WMSE}_t(\theta, \rho_t, \hat{\Sigma}_t), \\
\mathit{WMSE}_t(\theta, \rho_t, \hat{\Sigma}_t) &= \min_{\Sigma_t \geq 0} \sum_{i \in I} \sum_{j=1}^{10} \mathbf{1}_{\{t \in T_{i,j}\}} \frac{e_{i,j}^2(t)}{\hat{\sigma}_{i,j}^2(t)}.
\end{aligned}$$

Consider the constraint set $R(\theta)$. Expression $n_L = \left[\frac{n p_{HL}}{p_{HL} + p_{LH}} \right]$ reflects the fact that number of low correlation days is always integer. This restriction might result in discontinuous *WMSE* in SC and SCSV cases since for arbitrary close θ' and θ'' we might get different optimal n_L' and

n_L''. To avoid that complication, we assume for *WMSE* calculation purposes only that at the boundary date $t^* = t^*(\theta, (\hat{\rho}_t, \hat{\Sigma}_t)_{t\in T})$ defined as

$$t^* = \arg\min_{t\in T} \left\{ WMSE_t(\theta, \rho_t, \hat{\Sigma}_t) : \rho_t = \rho_H \right\}$$

we have an $WMSE_{t^*}(\theta, \rho_{t^*}, \hat{\Sigma}_{t^*})$ to be a weighted sum of $WMSE_{t^*}(\theta, \rho_L, \hat{\Sigma}_{t^*})$ and $WMSE_{t^*}(\theta, \rho_H, \hat{\Sigma}_{t^*})$ given by

$$\begin{aligned} WMSE^*_{t^*}(\theta, \rho_{t^*}, \hat{\Sigma}_{t^*}) \quad = \quad & \left(\frac{np_{HL}}{p_{HL} + p_{LH}} - n_L \right) WMSE_{t^*}(\theta, \rho_L, \hat{\Sigma}_{t^*}) \\ & + \left(n_L + 1 - \frac{np_{HL}}{p_{HL} + p_{LH}} \right) WMSE_{t^*}(\theta, \rho_H, \hat{\Sigma}_{t^*}). \end{aligned} \tag{24}$$

Since $WMSE_{t^*}$ is weighted in the objective function with a relative weight which is approximately $1/n = 1/557$, hence the impact of that change to the value of the objective function might be considered as negligible.

To estimate PSM, BFGS is used. Combination of BFGS and directional search is used for SV,SC and SCSV models. Both SVSC and SV models with daily calibrated $\Sigma(s)$ can be approximated by SVSC and SV with $\Sigma(s) \equiv \theta$ and $\vartheta = 0$.

Both SC and SCSV error surfaces might exhibit nonsmoothness since for arbitrary close θ' and θ'' minimization procedure might switch between different profiles (ρ_t') and (ρ_t'') that would result in discontinuity in gradient. A problem of the same sort might be brought by switching from boundary $\Sigma_t = 0$ to internal $\Sigma_t > 0$ in SV and SCSV models. We are not taking any precautions in that aspect because we expect only one switching of that type for any sufficiently close θ' and θ'', and nonsmoothness that it introduces is summed in the objective function with relative weight of magnitude $1/n = 1/557$. Efficiencies of both BFGS and directional search are not expected to be seriously affected, and numerical evidence confirms that.

4 Computational Implementation

For each set of parameters 59153 caplets should be priced, hence more than 100000 computationally costly numerical Fourier transforms should be performed. We need to reduce computational costs substantially. For these purpose, to solve ODEs for $N(\lambda), M_H(\lambda)$ and $M_L(\lambda)$ efficiently, three ranges of the Fourier inversion parameter λ are considered separately.

- For small λ, $N(\lambda), M_H(\lambda)$ and $M_L(\lambda)$ are analytic in λ, and coefficients of their power series expansions can be recursively derived from ODEs.

- For big λ, ODE for $N(\lambda)$ can be solved in closed form while coupled ODEs for $M_H(\lambda)$ and $M_L(\lambda)$ do not seem to have a closed form solution. It turns out that $(1/\lambda)\, N(\lambda)$ is analytic in $1/\lambda$, and coefficients of its power series expansion can be recursively derived from the ODE. The closed form solution helps to determine the radius of convergence of that expansion. It is trick to derive the expansions of $M_H(\lambda)$ and $M_L(\lambda)$ from the system os coupled ODEs. Numerical methods are used to detect convergence.

- For intermediate λ numerical ODE solvers deliver acceptable performance. For non-correlated case the closed-form solution might be used as well.

- Given that $N(\lambda), M_H(\lambda)$ and $M_L(\lambda)$ can be accurately evaluated for any $\lambda \geq 0$ and their asymptotics as $\lambda \longrightarrow 0$ and $\lambda \longrightarrow \infty$ are known, we approximate them after proper normalization as functions of λ on a certain finite segment using cubic splines. It reduces the number of costly $N(\lambda), M_H(\lambda)$ and $M_L(\lambda)$ evaluations in the integration step. Splines as a smoothing device helps integration routines.

- Now $N(\lambda), M_H(\lambda)$ and $M_L(\lambda)$ are accurate, smooth and can be obtained efficiently for any $\lambda \geq 0$, but 100000 integrations over λ are still too costly. Cubic splines are used to approximate a caplet price as a function of moneyness. When the SV component is present, cubic splines are also used to approximate a caplet price as a function of initial stochastic volatility Σ_t. This approximation is especially useful for optimization with respect to Σ_t.

5 Estimation Results

Parameter estimates and their standard errors are listed in Table 1. Apart from the above PSM, SC, SV and SCSV models we also estimate SV and SCSV models with estimated correlation parameter ν. We denote these models SVc and SCSVc correspondingly. Table 2 shows the goodness of fit using two measures: square root MSE and WMSE. MSE in this table correspond to the non-indexed pricing errors normalized to single quarterly caplet. When nonparametric estimates $\widehat{e}_{i,j}(t)$ are plugged into objective function instead of pricing errors $e_{i,j}$, we get the benchmark case $\widehat{WMSE} \approx 1$.

Figure 1 shows $\sqrt{MSE}$ and $\sqrt{MSRE}$ of non-indexed caplet prices for separate maturities. Nonparametrically smoothed pricing errors for non-indexed caplet prices as functions of moneyness on Figure 2 clearly emphasizes that the previous literature concentration on comparison of pricing errors only ATM is misleading. Figure 2 shows that those models that are best ATM may be really poor ITM or OTM. Caplets with adjacent maturities are pooled together if their pricing error curves are similar. To further explore whether some important dynamics has not been priced, we analyze principle component of pricing errors for all models under consideration. The results are summarized in Tables 3 and 4 and Figure 3. Table 3 shows eigenvalues of principal components of pricing errors while Table 4 shows the explanation power of this components. Figure 3 displays eigenvectors that correspond to the 3 largest eigenvalues. Eigenvectors are normalized to the unit length. Their direction is inverted when it is needed to make the coordinate with largest modulus positive. It is clear that PSM is dominated by its more sophisticated counterparts. However, neither graphical analysis, nor principal component analysis leads to clear cut conclusion about the best model. We perform statistical analysis to further differentiate between models.

For any model M and day t we define $WMSE_t^M$ as a weighted mean squared pricing error for indexed caplets of all maturities at the day t. The null hypothesis

$$H_0 : \frac{1}{T}\sum_{t=1}^{T} WMSE_t^{M_1} = \frac{1}{T}\sum_{t=1}^{T} WMSE_t^{M_2} \tag{25}$$

against the alternative

$$H_1 : \frac{1}{T}\sum_{t=1}^{T} WMSE_t^{M_1} > \frac{1}{T}\sum_{t=1}^{T} WMSE_t^{M_2} \tag{26}$$

is tested for each pair of models M_1 and M_2. The Newey-West statistic is $N(0,1)$ under the null. The upper triangle is obtained using the quadratic spectral kernel, the lower triangle corresponds to the Bartlet kernel. Bandwidth is chosen to be $h = 0.1, T = 557$. The results of this test are summarized in Table 5.To show the robustness of DM results with respect to bandwidth, the null is tested for 100 equidistant bandwidths $0.01 \leq h < 1$ and the significance levels $\alpha_{\min}$ are presented in Table 5. For any bandwidth $0.01 \leq h < 1$ the null hypothesis H_0 is rejected in favor of the alternative hypothesis H_1 at any significance level up to $1 - \alpha_{\min}^K$. It follows from Table 5 that the models with stochastic volatility term improve over PSM and SC at any reasonable significance level. We cannot conclude that SVc improves upon SCSV. The null can be always rejected for the pair SCSVc, SCSV only at 69% level. All other pairs of models always show improvement at 82% significance level.

6 Conclusion

Using more than two years of daily ITM and OTM interest rate cap price data, we provide probably the first comprehensive empirical analysis of performance of the generalized affine models of Collin-Dufresne and Goldstein (2003) in pricing interest rate derivatives. We develop efficient implementation procedure for estimation of these models. In analyzing the generalized affine model, we reach the following conclusions. First, we find that a popular version of generalized affine models, parsimonious string model with constant volatility and correlation of bond yields, cannot price caps well. Second, stochastic volatility provides the most significant improvement in model performance and stochastic correlation is also important. Finally we show that the generalized affine model is very flexible and as tractable as the traditional affine models. Our analysis demonstrates the empirical importance of incorporating stochastic volatility and correlation in bond yields for pricing interest rate derivatives. Although all considered models still cannot capture the entire smile. More research is warranted. In particular multiple sources of stochastic volatility might significantly improve the fit of a model. This paper show that stochastic correlation indeed matters for pricing interest rate caps. Thus study of more flexible correlation structures should be performed in future research. Another important aspect is out-of-sample performance of generalized affine models in pricing interest rate caps. Finally similar analysis of swaption would be desirable. The advantage of generalized affine framework in allowing consistently price different interest rate derivatives, especially caps and swaptions should be fully utilized.

References

Andersen, L., and J. Andreasen (2000). Volatility Skews and Extensions of the Libor Market Model. *Applied Mathematical Finance* 7, 1-32.

Andersen, L., and R. Brotherton-Ratcliffe (2001). Extended Libor Market Models with Stochastic Volatility. Working paper. Gen Re Securities.

Bakshi, G., C. Cao, and Z. Chen (1997). Empirical Performance of Laternative Option Pricing Models. *Journal of Finance* 52, 2003-2049.

Brace, A., M. Gatarek and M. Musiela (1997). The Market Model of Interest Rate Dynamics. *Mathematical Finance 7,* 127-155.

Campbell, J., A. Lo and C. MacKinlay (1997). *The Econometrics of Financial Markets.* Princeton University, New Jersey.

Casassus, J., P. Collin-Dufresne, and R. S. Goldstein (2005) Unspanned stochastic volatility and fixed income derivatives pricing. *Journal of Banking and Finance* 29, 2723-2749.

Collin-Dufresne, P., and R. S. Goldstein (2002). Do Bond Span the Fixed Income Markets? Theory and Evidence for Unspanned Stochastic Volatility. *Journal of Finance* 56, 1685-1730.

Collin-Dufresne, P., and R. S. Goldstein (2003). Generalizing the Affine Framework to HJM and Random Fields Models. Working paper. Carnegie Mellon University.

Cox, J. C., J. E. Ingersoll and S. A. Ross (1985). A Theory of the Term Structure of Interest Rates. *Econometrica 53,* 385-408.

Dai, Q., and K. Singleton (2002). Fixed Income Pricing. Working Paper, New York University.

Dai, Q., and K. Singleton (2003). Term Structure Dynamics in Theory and Reality. *Review of Financial Studies* 16, 631-678.

Duffie, D., and R. Kan (1996). A Yield-Factor Model of Interest Rates. *Mathematical Finance* 6, 379-406.

Duffie, D. (2002). *Dynamic Asset Pricing Theory.* Princeton University, New Jersey.

Egorov, A., H. Li, and Y. Xu, (2003). Maximum Likelihood Estimation of Time-Inhomogeneous Diffusions. *Journal of Econometrics* 114, 107-139.

Glasserman, P., and S. Kou (2003). The Term Structure of Simple Forward Rates with Jump Risk. *Mathematical Finance* 13, 383-410.

Goldstein, R. S. (2000). The Term Structure of Interest Rates as a Random Field. *The Review of Financial Studies* 2, 365-384.

Gupta, A., M. Subramanian (2005). Pricing and Hedging Interest Rate Options: Evidence from Cap-Floor Markets. *Journal of Banking and Finance* 29,701-733.

Harris, L. (2003). *Trading and Exchanges: Market Microstructure for Practitioners.* Oxford University Press.

Heath, D., R. Jarrow, and A. Morton (1992). Bond Pricing and the Term Structure of Interest Rates: A New Methodology for Contingent Claims Evaluation. *Econometrica* 60, 77-105.

Heidari, M., and L. Wu (2003). Are Interest Rate Derivatives Spanned by the Term Structure of Interest Rates? *Journal of Fixed Income*, 13, 75-86.

Hull, J., and A. White (2000). Forward Rate Volatilities, Swap Rate Volatilities and Implementation of the LIBOR market models. *Journal of Fixed Income* 10, 46-62.

Jagannathan R., A. Kaplin, and S. Sun (2003). An Evaluation of Multi-Factor CIR Models Using LIBOR, Swap Rates and Swaptions. *Journal of Econometrics*, 116, 113-146.

Jarrow, R., H. Li, and F. Zhao (2006), Interest Rate Caps "Smile" Too! But Can the LIBOR Market Models Capture It? *Journal of Finance*, forthcoming.

Kennedy, D. (1994). The Term Structure of Interest Rates as a Gaussian Random Field. *Mathematical Finance* 4, 247-258.

Kennedy, D. (1997). Characterizing Gaussian Models of Term Structure of Interest Rates. *Mathematical Finance* 7, 107-118.

Kimmel, R. (2004). Modeling the Term Structure of Interest Rates: A New Approach. *Journal of Financial Economics*, 72, 143-183.

Li, H., and F. Zhao (2006), Unspanned Stochastic Volatility: Evidence from Hedging Interest Rate Derivatives, *Journal of Finance*, forthcoming.

Longstaff, F., P. Santa-Clara and E. S. Schwartz (2001). The Relative Valuation of Caps and Swaptions: Theory and Empirical Evidence. *Journal of Finance* 56. 2067-2109.

Longstaff, F., P. Santa-Clara and E. S. Schwartz (2001). Throwing Away a Billion of Dollars: The Cost of Suboptimal Excercise Strategies in the Swaption Market. *Journal of Financial Economics* 62. 39–66.

Miltersen, K. R., K. Sandmann and D. Sondermann (1997). Closed Form Solutions for Term Structure Derivatives with Log-Normal Interest Rates. *Journal of Finance* 52, 409-430.

Santa-Clara, P. and D. Sornette (2001). The Dynamics of the Forward Interest Rate Curve with Stochastic String Shocks. *The Review of Financial Studies* 1, 149-185.

Vasicek, O. (1977). An Equilibrium Characterization of the Term Structure. *Journal of Financial Economics* 5, 177-188.

7 Appendix.

Proof of Proposition 2.

If we define $\tilde{N}\left(T_0 - s, T_j\right) = N\left(s, T_j\right)$ and $\tau = T_0 - s$, then ODE (6) equation can be written as

$$
\begin{aligned}
\tilde{N}'_{\tau}(\tau, T_j) &= \frac{\vartheta^2}{2}\tilde{N}^2(\tau, T_j) - \tilde{N}(\tau, T_j)\left(\kappa + \nu\vartheta\left(B^{T_j}(T_0 - \tau) + \lambda\left(B^{T_1}(T_0 - \tau) - B^{T_0}(T_0 - \tau)\right)\right)\right) \\
&\quad + \frac{\lambda\left(\lambda - (-1)^j\right)}{2}\left(b^{\delta}_{\beta_B}\right)^2 e^{-2\beta\tau} \\
&= \frac{\vartheta^2}{2}\tilde{N}^2\left(\tau, T_j\right) - \tilde{N}\left(\tau, T_j\right)\left(\kappa + \nu\vartheta\left(\frac{1 - e^{-\beta(T_j - T_0 + \tau)}}{\beta} + \lambda e^{-\beta\tau} b^{\delta}_{\beta_B}\right)\right) \\
&\quad + \frac{\lambda\left(\lambda - (-1)^j\right)}{2}\left(b^{\delta}_{\beta_B}\right)^2 e^{-2\beta\tau} \\
&= c_2\tilde{N}^2\left(\tau, T_j\right) + \tilde{N}\left(\tau, T_j\right)\left(c_1 + ae^{-\beta\tau}\right) + c_0 e^{-2\beta\tau}, \\
\tilde{N}(0, T_j) &= 0,
\end{aligned}
$$

where $\delta = T_1 - T_0$ and the coefficients are determined by

$$
\begin{aligned}
b^{\delta}_{\beta} &= \frac{1 - e^{-\beta_B\delta}}{\beta_B}, c_2 = \frac{\vartheta^2}{2}, c_1 = -\kappa - \frac{\nu\vartheta}{\beta_B}, \\
a &= \nu\vartheta\left(\frac{e^{-\beta_B(T_j - T_0)}}{\beta_B} - \lambda b^{\delta}_{\beta_B}\right), c_0 = \frac{\lambda\left(\lambda - (-1)^j\right)}{2}\left(b^{\delta}_{\beta_B}\right)^2.
\end{aligned}
$$

In correlated case $\nu > 0$ we have $\tilde{N}(\tau, T_0) \neq \overline{\tilde{N}(\tau, T_1)}$. If $x'_j(\tau)$ is defined by

$$
\begin{aligned}
c_2\tilde{N}(\tau, T_j) &= -\frac{x'_j(\tau)}{x_j(\tau)}, \\
x'_j(0) &= 0,
\end{aligned}
$$

then

$$
\begin{aligned}
x''_j(\tau) &= x'_j(\tau)\left(c_1 + ae^{-\beta_B\tau}\right) - c_0 c_2 x_j(\tau) e^{-2\beta_B\tau} = 0, \\
x'_j(0) &= 0.
\end{aligned}
$$

If we define $u = e^{-\beta_B \tau}$ and $w_j(u) = x_j(\tau)$ (p.149, 141, $k = 0$), then

$$\begin{aligned} \beta_B^2 u w_j''(u) + (\beta_B a u + \beta_B (\beta_B + c_1)) w_j'(u) + c_0 c_2 u w_j(u) &= 0, \\ w_j(1) &= 0. \end{aligned}$$

It has form

$$(a_2 u + b_2) w_j''(u) + (a_1 u + b_1) w_j'(u) + (a_0 u + b_0) w_j(u) = 0,$$

where

$$\begin{aligned} a_2 &= \beta^2 \neq 0, b_2 = 0, a_1 = \beta a \neq 0, b_1 = \beta (\beta + c_1), \\ a_0 &= c_0 c_2 \neq 0, b_0 = 0, D^2 = a_1^2 - 4 a_0 a_2. \end{aligned}$$

If $D \neq 0$, then

$$\begin{aligned} w_j(u) &= e^{hu} z(\xi(u)), \\ \xi(u) &= (u - \mu) \eta, \\ h &= \frac{D - a_1}{2a_2} = \frac{\sqrt{a^2 - 4c_0c_2} - a}{2\beta_B}, \\ \eta^{-1} &= -\frac{a_2}{A(h)} = -\frac{\beta_B}{\sqrt{a^2 - 4c_0c_2}}, \\ \mu &= -\frac{b_2}{a_2} = 0, \\ D^2 &= a_1^2 - 4a_0a_2 = \beta_B^2 \left(a^2 - 4c_0c_2\right), \\ A(h) &= 2a_2h + a_1 = \beta_B \sqrt{a^2 - 4c_0c_2}, \\ B(h) &= b_2h^2 + b_1h + b_0 = \frac{(\beta_B + c_1)\left(\sqrt{a^2 - 4c_0c_2} - a\right)}{2}, \\ z(\xi) &= \mathcal{J}\left(\frac{B(h)}{A(h)}, \frac{a_2b_1 - a_1b_2}{a_2^2}; \xi\right), \end{aligned}$$

where

$$\mathcal{J}(\tilde{a}, \tilde{b}; \xi) = C_1 \Phi(\tilde{a}, \tilde{b}; \xi) + C_2 \xi^{1-\tilde{b}} \Phi(\tilde{a} - \tilde{b} + 1, 2 - \tilde{b}; \xi)$$

and $\Phi(\tilde{a}, \tilde{b}; \xi) = {}_1F_1(\tilde{a}, \tilde{b}; \xi)$ is a Kummer (or confluent hypergeometric) function.

Therefore

$$\begin{aligned} w_j(u) &= e^{hu} z(\xi(u)) \\ &= e^{hu} \mathcal{J}\left(\frac{B(h)}{A(h)}, \frac{a_2 b_1 - a_1 b_2}{a_2^2}; \xi(u)\right) \\ &= e^{hu} \mathcal{J}\left(\frac{b_1 h}{2a_2 h + a_1}, \frac{b_1}{a_2}; u\eta\right) \\ &= e^{hu} \mathcal{J}\left(\frac{\beta\left(\beta + c_1\right) h}{2\beta^2 h + \beta a}, 1 + \frac{c_1}{\beta}; u\eta\right) \\ &= e^{\frac{\sqrt{a^2-4c_0c_2}-a}{2\beta_B} u} \mathcal{J}\left(\frac{\left(\beta_B + c_1\right)\left(\sqrt{a^2 - 4c_0c_2} - a\right)}{2\beta_B\sqrt{a^2-4c_0c_2}}, 1 + \frac{c_1}{\beta_B}; -\frac{\sqrt{a^2-4c_0c_2}}{\beta_B} u\right) \\ &= e^{hu} \mathcal{J}\left(\tilde{a}, \tilde{b}; u\eta\right), \end{aligned}$$

$$\begin{aligned} \tilde{N}(\tau, T_j) &= -\frac{x_j'(\tau)}{c_2 x_j(\tau)} = \frac{\beta_B e^{-\beta_B \tau} w_j'(e^{-\beta_B \tau})}{c_2 w_j(e^{-\beta_B \tau})}, \\ w_j'(1) &= \frac{\partial}{\partial u}\left(e^{hu}\mathcal{J}(\tilde{a}, \tilde{b}; u\eta)\right)\Big|_{u=1} = e^h\left(h\mathcal{J}(\tilde{a}, \tilde{b}; \eta) + \eta\mathcal{J}_3'\left(\tilde{a}, \tilde{b}; \eta\right)\right) = 0, \end{aligned}$$

and that can be written as

$$\begin{aligned} \mathcal{J}_3'\left(\tilde{a}, \tilde{b}; \eta\right) &= C_1 \Phi_3'\left(\tilde{a}, \tilde{b}; \eta\right) + C_2 \eta^{1-\tilde{b}} \Phi_3'\left(\tilde{a} - \tilde{b} + 1, 2 - \tilde{b}; \eta\right) \\ &\quad + C_2\left(1 - \tilde{b}\right)\eta^{-\tilde{b}}\Phi\left(\tilde{a} - \tilde{b} + 1, 2 - \tilde{b}; \eta\right)\Big). \end{aligned}$$

Hence

$$\alpha = \frac{C_1}{C_2} = \eta^{1-\tilde{b}} \frac{h\Phi\left(\tilde{a} - \tilde{b} + 1, 2 - \tilde{b}; \eta\right) + \eta\Phi_3'\left(\tilde{a} - \tilde{b} + 1, 2 - \tilde{b}; \eta\right) + \left(1 - \tilde{b}\right)\Phi\left(\tilde{a} - \tilde{b} + 1, 2 - \tilde{b}; \eta\right)}{h\Phi\left(\tilde{a}, \tilde{b}; \eta\right) + \eta\Phi_3'\left(\tilde{a}, \tilde{b}; \eta\right)}$$

Table 1. Parameter Estimates and Standard Errors

Parameter	PSM	SC	SV	SCSV
σ	0.016572 (0.0019)	0.036447 (0.0041)	0.013852 (0.0029)	0.014099 (0.0027)
β	0.095899 (0.0437)	1.258782 (0.0681)	0.108378 (0.0773)	0.095747 (0.0320)
θ_{VS}	-	-	0.000073 (0.0002)	0.000028 (0.0002)
κ_{VS}	-	-	3.584018 (6.5890)	7.576052 (0.3619)
ϑ_{VS}	-	-	2.436245 (2.7384)	2.753821 (0.3826)
β_B	-	-	0.026702 (0.32579)	0.034288 (0.03803)
ρ_H	0.014969 (0.3872)	0.000000 (0.0732)	0.097722 (0.1140)	0.041101 (0.0375)
ρ_L	-	0.569314 (0.0627)	-	0.141720 (0.0365)
p_{HL}	-	0.452110 (0.0715)	-	0.173817 (0.3186)
p_{LH}	-	0.545540 (0.07221)	-	0.280520 (0.3438)

Table 2. Goodness of Fit

Measure	NPM	PSM	SC	SV	SCSV
WMSE	0.998672	1.281636	0.959661	0.656805	0.595781
$\sqrt{MSE}$	0.028986	0.031648	0.026957	0.021137	0.020359

Table3. Eigenvalues of Principal Components of Pricing Errors

	PSM	SC	SV	SVc	$SCSV$	$SCSVc$
1	1.4583	0.6391	0.4235	0.4698	0.2536	0.3652
2	0.4514	0.5134	0.2252	0.2172	0.2323	0.2177
3	0.1395	0.1456	0.1566	0.1325	0.1776	0.1307
4	0.1395	0.0881	0.1088	0.0864	0.1014	0.0728
5	0.0601	0.0697	0.0662	0.0646	0.0640	0.0637
6	0.0513	0.0442	0.0505	0.0501	0.0606	0.0495
7	0.0292	0.0285	0.0290	0.0290	0.0335	0.0293
8	0.0206	0.0192	0.0215	0.0197	0.0229	0.0199
9	0.0167	0.0155	0.0163	0.0163	0.0177	0.0157
10	0.0153	0.0135	0.0137	0.0126	0.0147	0.0126
11	0.0113	0.0119	0.0117	0.0114	0.0137	0.0114
12	0.0079	0.0078	0.0083	0.0085	0.0081	0.0080

Table 4. Explanatory Power of Principal Components of Pricing Errors

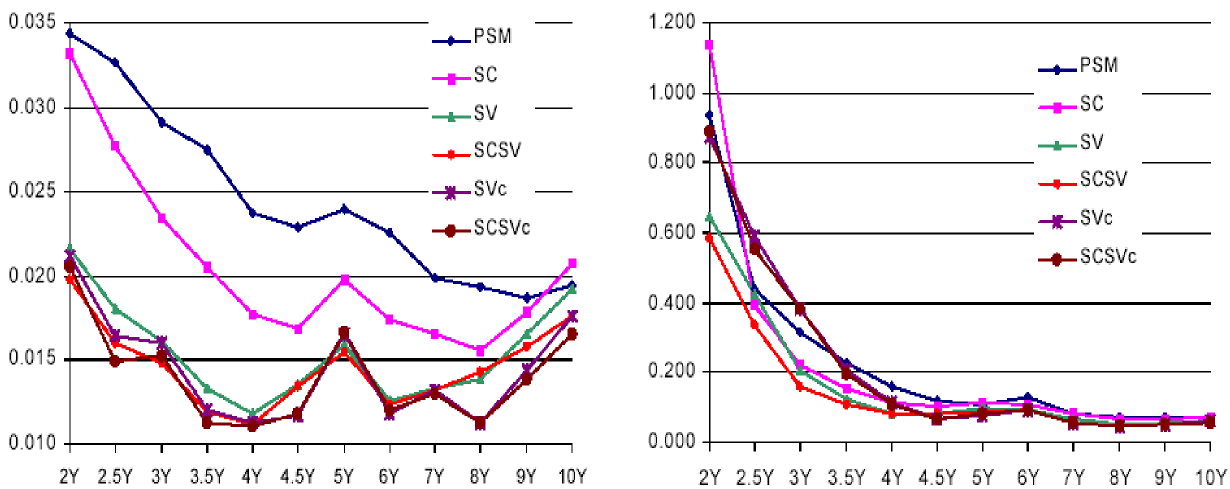
	PSM	SC	SV	SVc	$SCSV$	$SCSVc$
1	0.6172	0.4003	0.3744	0.4201	0.4135	0.3666
2	0.1911	0.3216	0.1991	0.1942	0.1942	0.2185
3	0.0591	0.0912	0.1384	0.1185	0.1228	0.1311
4	0.0428	0.0552	0.0962	0.0773	0.0651	0.0731
5	0.0254	0.0437	0.0585	0.0578	0.0542	0.0639
6	0.0217	0.0277	0.0447	0.0448	0.0385	0.0496
7	0.0124	0.0179	0.0256	0.0260	0.0325	0.0294
8	0.0087	0.0120	0.0190	0.0176	0.0279	0.0199
9	0.0071	0.0097	0.0144	0.0146	0.0175	0.0158
10	0.0065	0.0084	0.0121	0.0113	0.0134	0.0126
11	0.0048	0.0075	0.0103	0.0102	0.0113	0.0114
12	0.0033	0.0049	0.0073	0.0076	0.0060	0.0080

Table 5. Newey West Statistics

$Model$	PSM	SC	SV	$SCSV$	SVc	$SCSVc$
PSM	–	2.65	5.14	5.32	4.66	4.95
SC	2.94	–	7.80	6.88	5.20	5.98
SV	5.67	8.57	–	3.33	1.81	2.44
$SCSV$	5.86	7.51	3.53	–	0.63	1.17
SVc	5.16	5.77	2.03	0.70	–	2.49
$SCSVc$	5.47	6.61	2.72	1.30	2.70	–

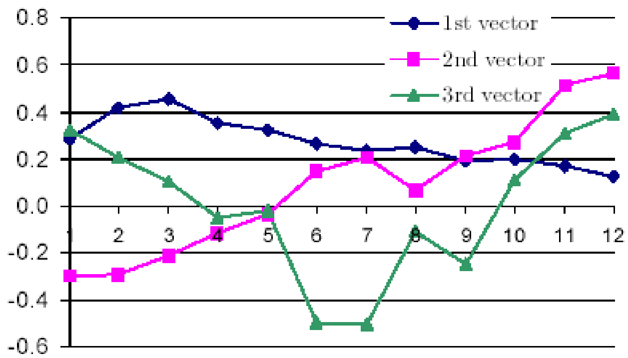
Table 6. Significance Levels $\alpha_{\min}$

$Model$	PSM	SC	SV	$SCSV$	SVc	$SCSVc$
PSM	$-$	0.0110	2×10^{-6}	2×10^{-6}	4×10^{-5}	1×10^{-5}
SC	0.0060	$-$	4×10^{-13}	1×10^{-8}	4×10^{-6}	4×10^{-7}
SV	4×10^{-7}	8×10^{-15}	$-$	0.0178	0.1160	0.0378
$SCSV$	4×10^{-7}	2×10^{-9}	0.0106	$-$	0.5738	0.3044
SVc	1×10^{-5}	1×10^{-6}	0.0880	0.5378	$-$	0.0164
$SCSVc$	2×10^{-6}	1×10^{-7}	0.0238	0.2582	0.0058	$-$

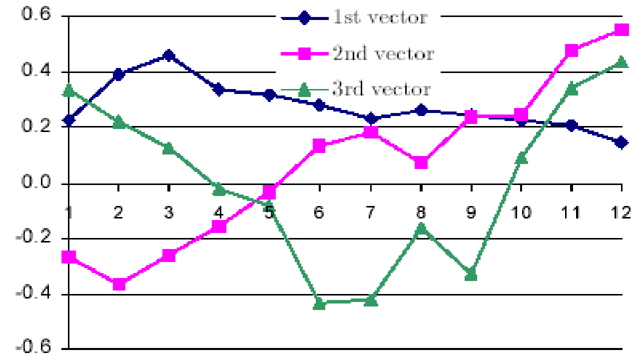


0.035
0.030
0.025
0.020
0.015
0.010
2Y 2.5Y 3Y 3.5Y 4Y 4.5Y 5Y 6Y 7Y 8Y 9Y 10Y
PSM
SC
SV
SCSV
SVc
SCSVc
1.200
1.000
0.800
0.600
0.400
0.200
0.000
2Y 2.5Y 3Y 3.5Y 4Y 4.5Y 5Y 6Y 7Y 8Y 9Y 10Y
PSM
SC
SV
SCSV
SVc
SCSVc

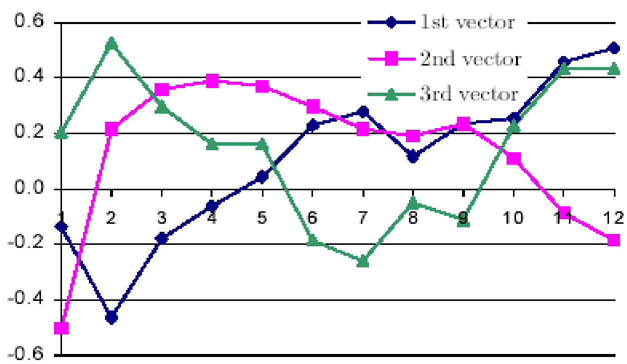
Figure 1:



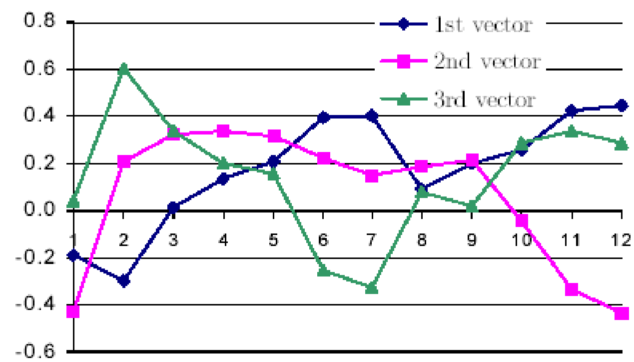
PSM



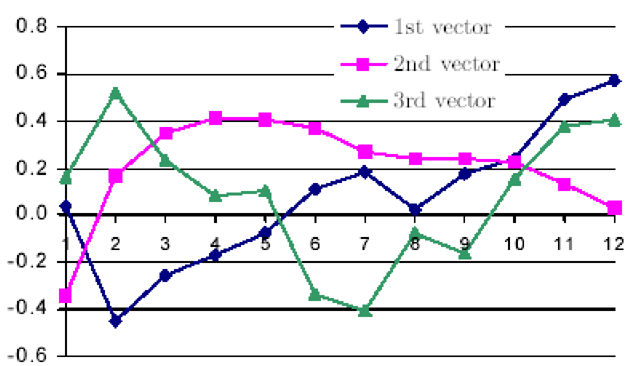
SC



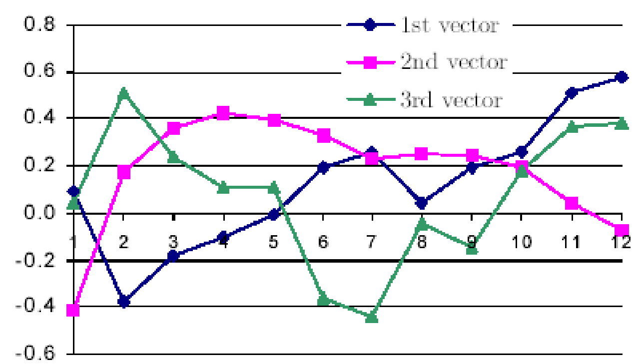
SV



SCSV



SVc



SCSVc

Figure 2:

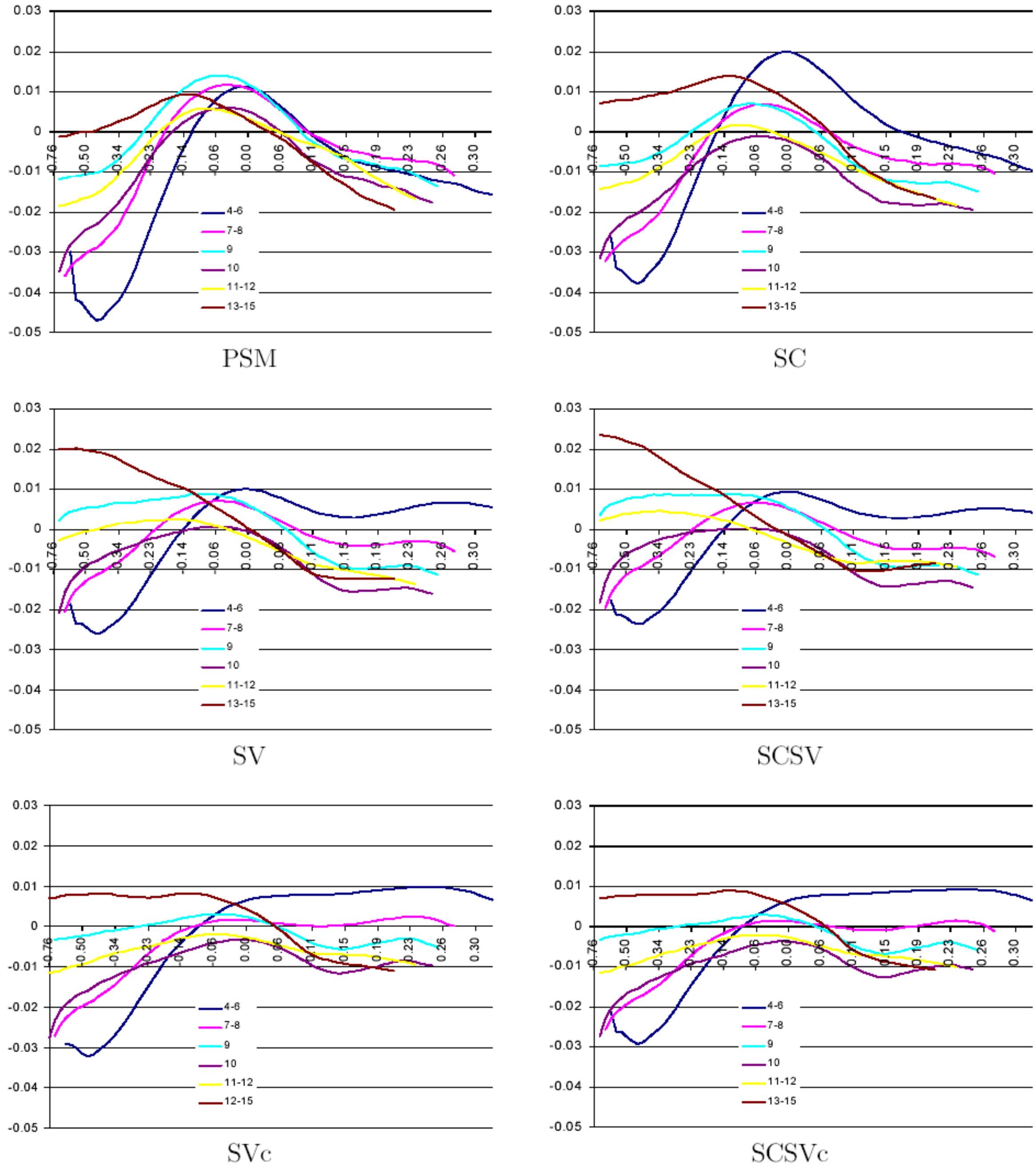

PSM
SC
SV
SCSV
SVc
SCSVc
4-6
7-8
9
10
11-12
13-15
12-15
0.03
0.02
0.01
0
-0.01
-0.02
-0.03
-0.04
-0.05
-0.76
-0.50
-0.34
-0.23
-0.14
-0.06
0.00
0.06
0.11
0.15
0.19
0.23
0.26
0.30

Figure 3: