International Journal of Theoretical and Applied Finance
Vol. 14, No. 3 (2011) 407–432

DOI: 10.1142/S0219024911006590

World Scientific
www.worldscientific.com

FORWARD AND FUTURE IMPLIED VOLATILITY

PAUL GLASSERMAN
Columbia Business School
pg20@columbia.edu

QI WU*
Department of APAM
Columbia University
200 S.W. Mudd Building
New York, NY 10027, USA
qw2107@columbia.edu

Received 14 June 2010
Accepted 12 January 2011

We address the problem of defining and calculating forward volatility implied by option prices when the underlying asset is driven by a stochastic volatility process. We examine alternative notions of forward implied volatility and the information required to extract these measures from the prices of European options at fixed maturities. We then specialize to the SABR model and show how the asymptotic expansion of the bivariate transition density in Wu (forthcoming) allows calibration of the SABR model with piecewise constant parameters and calculation of forward volatility. We then investigate empirically whether current option prices at multiple maturities contain useful information in predicting future option prices and future implied volatility. We undertake this investigation using data on options on the euro-dollar, sterling-dollar, and dollar-yen exchange rates. We find that prices across maturities do indeed have predictive value. Moreover, we find that model-based forward volatility extracts this predicative information better than a standard "model-free" measure of forward volatility and better than spot implied volatility. The enhancement to out-of-sample forecasting accuracy gained from model-based forward volatility is greatest at longer forecasting horizons.

Keywords: Forward volatility; implied volatility surface; time-dependent SABR model; currency options; volatility forecasting.

1. Introduction

This paper investigates the concept of forward implied volatility in option prices with a specific application to stochastic volatility and currency markets. The term "forward implied volatility" or simply "forward vol" is used, broadly, to refer to

*Corresponding author.

future levels of volatility consistent with current market prices of options. The values of many path-dependent options, including cliquets and barrier options, are commonly interpreted through levels of volatility implied at some future date and, often, at some future level of the underlying asset. Similarly, the price of a forward-starting option is sensitive to the anticipated level of volatility at the forward start date and strike price of the option.

Forward volatility builds on the notions of spot implied volatility and forward rates, but with important differences. Spot implied volatility is unambiguous, in the sense that, given all other contract terms and market parameters, there is precisely one value of volatility at which the Black-Scholes formula will match a finite, strictly positive price observed in the market. In a term structure setting, bond prices at any two maturities completely determine the forward rate between those maturities through a static arbitrage argument without any assumptions on the evolution of interest rates.

But forward vol is not uniquely determined by the absence of arbitrage and the market prices of standard calls and puts at a finite set of maturities. This is a special case of the fundamental ill-posedness of the derivatives valuation problem: calls and puts determine, at best, the marginal distribution of the underlying asset at fixed dates, but pinning down a specific model requires the joint distribution across multiple dates. Resolving the ambiguity requires additional information or assumptions. For example, in fitting a volatility surface based on Dupire [13], practitioners often add smoothness constraints to select a calibration. The main alternative is to posit a model for the dynamics of the underlying asset, and this is the approach we follow. Once calibrated to standard European options at two or more maturities, the model implies a value of forward volatility at all but the last maturity, for each level of the model's state variables. Thus, a forward implied volatility is implied by a combination of market prices and model choice.

We work within a stochastic volatility setting generally and the SABR model of Hagan, Kumar, Lesniewski, and Woodward [20] specifically. The SABR model is widely used to fit slices of the volatility surface, particularly for currency and interest rate options, so it is natural to extend this application to extract forward volatilities. Our approach is made feasible by the asymptotic expansion derived in Wu [26] for the bivariate transition density of the underlying and its stochastic volatility in the SABR model. The original expansion of Hagan *et al.* [20] provides highly accurate implied volatilities at a single maturity; but calibration to multiple maturities and the calculation of forward volatilities requires the joint transition density of the two state variables analyzed in Wu [26]. Henry-Labordere [22] and Hagan, Lesniewski, and Woodward [21] also derived approximations to the bivariate transition density, but these are less amenable to the computations we undertake here.

Before focusing on the SABR model, we examine alternative notions of forward implied volatility in the presence of stochastic volatility. Alternative notions differ, for example, in the information on which they condition. Passing from a forward

implied volatility at a given level of the underlying and its stochastic volatility to one conditioned only on the underlying requires integrating over the conditional distribution of the stochastic volatility. We also examine various ways of taking expectations of future levels of implied volatility as measures of forward volatility. This discussion clarifies alternative interpretations of "forward vol," which is often used loosely in practice. The calculations required for the alternative interpretations are, here again, made possible in the SABR setting by the results in Wu [26].

Having defined and calculated notions of forward volatility, we test our approach on market data and examine whether forward vol has predictive power in forecasting future levels of implied volatility. We use data from August 2001 to June 2009, provided by a major derivatives dealer, on over-the-counter currency options on the euro-dollar, sterling-dollar, and dollar-yen exchange rates. The data is in the form of constant-maturity quotes. From quotes on 6-month options and 9-month options, for example, we calculate 3-month volatilities, six months forward. We then compare these with spot implied volatilities observed six months later. As one benchmark, we use the current level of implied volatility to forecast future volatility. Another benchmark is provided by a "model-free" forward volatility associated with a deterministic but time-varying volatility function. We compare in-sample forecasts and out-of-sample forecasts using rolling regressions. For both, we find that model-based forward vol outperforms the benchmarks.

There is no simple theoretical link between forward volatility and future volatility. Forward volatility depends only on a model's dynamics under a pricing measure, whereas the evolution of implied volatility depends on the dynamics under both the pricing measure and the empirical measure. A link between the two requires, at a minimum, a model that describes dynamics under both measures. One would then expect that the relationship between forward vol and expected future vol involves a risk premium and a convexity correction, as is usually the case when a forward value is used to predict a future value. We therefore do not suggest that forward volatility should be an unbiased predictor of future implied volatility, even in theory. Nevertheless, the fact that the forward vols we calculate enhance the ability to predict future vols lends support and adds value to the approach we take. It also confirms the view that market prices of options at different maturities contain information relevant to predicting future option prices. More details on modeling implied volatility in general can be found in Gatheral's book [19]. Specific examples of dynamic volatility models can be found in Schönbucher [25], Cont and Da Fonseca [11], Buehler [7], Carr and Wu [10], Schweizer and Wissel [24], Carmona and Nadtochiy [9], Gatheral [18], and Bergomi's series of papers on "smile dynamics" [3]–[6].

The rest of this paper is organized as follows. In Sec. 2, we detail various notions of forward volatility. In Sec. 3, we present our calibration method, using time-varying parameters to fit the SABR model to multiple maturities. Section 4 applies Wu [26] and numerical integration to calculate forward vol in the calibrated SABR model. Section 5 presents our empirical results.

2. Notions of Forward Implied Volatility

Depending the set of information utilized for evaluation of future state variables, we introduce three notions of model-based forward implied volatilities, namely the fully-conditional, the partially-conditional, and the expected forward implied volatility. Throughout, we focus on stochastic volatility models and define all model-based notions through the Black model's volatility parameter. Expectations in this section are all taken under relevant forward measures.

2.1. *Spot and forward Black implied volatility*

Let the forward price process of an underlying asset be $F(t)$, and let its instantaneous volatility process be $\alpha(t)$. Further let the parameters of the concerned stochastic volatility model be θ and let the model's joint transition density from t to T be $p(t, f, \alpha; T, F, A)$ with f, α and F, A denoting values of random variables $F(t), \alpha(t)$ at t and T respectively.

A standard European call on $F(t)$ has payoff

$$(F(T) - K)^+, \quad K > 0, \quad t < T$$

where t is the evaluation time, usually understood as "now" and set to zero, and T is the option maturity. Under the Black model, F follows geometric Brownian motion, and the option price is given by:

$$\mathbb{E}[(F(T) - K)^+ | F(t)] = F(t)N(d_+) - KN(d_-)$$

$$\text{with } d_\pm = \frac{\log(F(t)/K) \pm \frac{1}{2}\sigma^2(T-t)}{\sigma\sqrt{T-t}} \tag{2.1}$$

Let V^{Mkt} be the market price of the call and let σ be its Black implied volatility, the unique volatility parameter under which V^{Mkt} matches the call price in the Black model (2.1).

Now consider a forward starting European call on $F(t)$ with payoff

$$(F(T_2) - kF(T_1))^+, \quad k > 0, \quad t < T_1 < T_2$$

where k is the strike ratio, t is still the evaluation time, T_1 is the option starting time and T_2 is the option maturity. Conditional on T_1 and denoted by $V^{\text{Black}}(T_1, T_2, k, \sigma)$, its price under the Black model becomes:

$$V^{\text{Black}}(T_1, T_2, k, \sigma) = F(T_1)N(d_+) - kF(T_1)N(d_-)$$

$$\text{with } d_\pm = \frac{\log(1/k) \pm \frac{1}{2}\sigma^2(T_2 - T_1)}{\sigma\sqrt{T_2 - T_1}}$$

Let $V^{\text{Model}}(T_1, T_2, k, \theta)$ be the T_1-conditional expectation of this payoff under the concerned stochastic volatility model. The model-based notion of T_1-into-T_2 forward Black implied volatility is defined as the unique volatility parameter in the Black model under which $V^{\text{Model}}(T_1, T_2, k, \theta)$ matches $V^{\text{Black}}(T_1, T_2, k, \sigma)$:

$$\sigma \text{ s.t. } V^{\text{Model}}(T_1, T_2, k, \theta) = V^{\text{Black}}(T_1, T_2, k, \sigma)$$

where

$$V^{\text{Model}}(T_1, T_2, k, \theta) := \mathbb{E}[(F(T_2) - kF(T_1))^+ | F(T_1), \alpha(T_1)]$$

is calculated under the dynamics of the stochastic volatility model with calibrated parameter θ. When the concerned model is the SABR stochastic volatility model, $V^{\text{Model}}(T_1, T_2, k, \theta)$ becomes $V^{\text{SABR}}(T_1, T_2, k, \alpha, \beta, \rho, \nu)$. The case of a put can be defined analogously.

The thus-defined forward Black implied volatility is a T_1-dependent quantity. In particular, it is a function of future state variables $F(T_1)$ and $\alpha(T_1)$ in the concerned stochastic volatility model, which are unknown at time t. From now on, we shorten it to "Blk-Fwd-IV" and denote it by

$$\Sigma(T_1, T_2, k | F(T_1), \alpha(T_1)) \tag{2.2}$$

to explicitly reflect its T_1 dependency.

Depending on the set of information utilized to evaluate $F(T_1)$ and $\alpha(T_1)$ in (2.2), we arrive at three notions of forward implied volatilities — fully conditional, partially conditional, and expected.

2.2. *Fully-conditional*

The "Fully-Conditional Forward Implied Volatility" refers to the Blk-Fwd-IV when $F(T_1), \alpha(T_1)$ in Σ are evaluated at some chosen positive real values F_1, A_1. Denoted by Σ^{flcd} and shortened to "Flcd-Fwd-IV", the fully-conditional forward implied volatility is:

$$\Sigma^{\text{flcd}}(T_1, T_2, k, F_1, A_1) := \Sigma(T_1, T_2, k \,|\, F(T_1) = F_1, \alpha(T_1) = A_1) \tag{2.3}$$

Given a fixed strike ratio k and a fixed future period $[T_1, T_2]$, fully-conditional forward implied volatility is a function of future underlying level F_1 and future instantaneous volatility A_1. Its two dimensional property is analogous to that of a spot implied volatility surface which is understood as a function of today's underlying level and option tenor.

When computing Σ^{flcd} in (2.3), one first fixes both F_1 and A_1 at some chosen positive real values, then uses the model's full joint transition density from T_1 to T_2 to obtain an option price under the concerned model as

$$\iint_{\mathbb{R}^2_+} (F_2 - kF_1)^+ p(T_1, F_1, A_1; T_2, F_2, A_2) dF_2 dA_2$$

and inverts this price to a Black volatility using the same value of F_1.

Upon calibrating the model to the market's spot implied volatility curves at both T_1 and T_2, the Flcd-Fwd-IV incorporates market information regarding the T_2 states conditional on starting at F_1, A_1, and it does so through the model's full joint transition density from T_1 to T_2.

2.3. *Partially-conditional*

The "Partially-Conditional Forward Implied Volatility" is defined similarly through the Blk-Fwd-IV where only $F(T_1)$ in Σ is evaluated at some chosen positive real value F_1 and the volatility state variable $\alpha(T_1)$ is integrated out using the model's conditional joint transition density from t to T_1, conditional on $F(T_1)$ being F_1.

Denoted by Σ^{ptcd} and shortened to "Ptcd-Fwd-IV", the partially-conditional forward implied volatility is:

$$\Sigma^{\text{ptcd}}(t, T_1, T_2, k, F_1) := \mathbb{E}[\Sigma(T_1, T_2, k \,|\, F(T_1), \alpha(T_1)) \,|\, F(T_1) = F_1] \tag{2.4}$$

Alternatively, it could be defined by first taking the expectation of forward Black implied variance Σ^2 and then taking the squared root of the resulting expectation:

$$\Sigma^{\text{ptcd}}(t, T_1, T_2, k, F_1) := \sqrt{\mathbb{E}[\Sigma^2(T_1, T_2, k \,|\, F(T_1), \alpha(T_1)) \,|\, F(T_1) = F_1]} \tag{2.5}$$

(2.4) and (2.5) differ from each other only by a convexity adjustment. However, our focus will be on the alternative dependence on the variables F and A, rather than on this distinction.

Given a fixed strike ratio k and a fixed future period $[T_1, T_2]$, partially-conditional forward implied volatility is a one-dimensional function of future underlying level F_1 with future instantaneous volatility integrated out.

When computing Σ^{ptcd}, one first fixes F_1, then, for each value of the volatility state A_1 in its support $\mathbb{R}_+$, one obtains the corresponding Flcd-Fwd-IV using the joint transition density from T_1 to T_2 as in the fully-conditional case, and finally the Ptcd-Fwd-IV is obtained by integrating these Flcd-Fwd-IVs against the F_1-conditional joint transition density from t to T_1:

$$p(t, f, \alpha; T_1, A_1 \,|\, F_1) := \frac{p(t, f, \alpha; T_1, F_1, A_1)}{\int_{\mathbb{R}_+} p(t, f, \alpha; T_1, F_1, A_1) dA_1}$$

as:

$$\int_{\mathbb{R}_+} \Sigma^{\text{flcd}}(t, T_1, T_2, k, F_1, A_1) p(t, f, \alpha; T_1, A_1 \,|\, F_1) dA_1$$

Upon model calibration, the Ptcd-Fwd-IV not only reflects market information regarding the T_2 states conditional on starting at F_1, A_1 as in the Flcd-Fwd-IV case, but also partially incorporates through the model market information from the period $[t, T_1]$ to eliminate the uncertainty, from an evaluation point of view, of the future volatility state $\alpha(T_1)$ at T_1. It does so through the full joint transition density from T_1 to T_2 and the F_1-conditional joint transition density from t to T_1.

2.4. *Expected*

The "Expected Forward Implied Volatility" goes one step further than the partially conditional case, and is defined by integrating both of the T_1 state variables $F(T_1), \alpha(T_1)$ in Σ against the model's full joint transition density from t to T_1.

Denoted by Σ^{expt} and shortened to "Expt-Fwd-IV", the expected forward implied volatility is no longer a function of T_1 states and it is:

$$\Sigma^{\text{expt}}(t, T_1, T_2, k) := \mathbb{E}[\Sigma(T_1, T_2, k \,|\, F(T_1), \alpha(T_1))] \tag{2.6}$$

Similar to the partially-conditional case, it could also be defined alternatively as:

$$\Sigma^{\text{expt}}(t, T_1, T_2, k) := \sqrt{\mathbb{E}[\Sigma^2(T_1, T_2, k \,|\, F(T_1), \alpha(T_1))]} \tag{2.7}$$

up to a convexity difference.

When computing Σ^{expt}, one follows the same procedure as in the partially-conditional case except neither F_1 nor A_1 is fixed when computing Flcd-Fwd-IV. Instead, one obtains Flcd-Fwd-IV for each pair of F_1 and A_1 in its support $\mathbb{R}^2_+$, and then integrates these Flcd-Fwd-IVs against the full joint transition density from t to T_1 as:

$$\iint_{\mathbb{R}_+} \Sigma^{\text{flcd}}(t, T_1, T_2, k, F_1, A_1) p(t, f, \alpha; T_1, F_1, A_1) dF_1 dA_1$$

instead of using the F_1-conditional one.

The Expt-Fwd-IV utilizes full joint transition probabilities from both periods $[T_1, T_2]$ and $[t, T_1]$, thus fully incorporates market information regarding state variables at both the starting point T_1 and the ending point T_2.

2.5. *Model-free*

The following model-free quantity is sometimes used (see, e.g., Corte, Sarno, and Tsiakas [12], Egelkraut and Garcia [14] and Egelkraut, Garcia, and Sherrick [15]) as a measure of market implied forward volatility

$$\begin{aligned}&\Sigma^{\text{mdfr}}(t, T_1, T_2, K)\\ &\quad := \sqrt{\frac{(\sigma^{\text{Mkt}}(t, T_2, K))^2 \times (T_2 - t) - (\sigma^{\text{Mkt}}(t, T_1, K))^2 \times (T_1 - t)}{T_2 - T_1}}\end{aligned} \tag{2.8}$$

where σ^{Mkt} denote market quotes of spot implied volatilities and Σ^{mdfr} denotes this model-free notion which is shortened to "Mdfr-Fwd-IV". As a "model-free" concept, Σ^{mdfr} in (2.8) is not a function of state variables at any of the future times. When computing it, one simply follows its definition.

3. Model Specification and Calibration

Having introduced various notions of forward volatility, we turn to the SABR model, its tenor-dependent parameters and model calibration.

3.1. *Piecewise-constant parameters*

The original SABR model has constant parameters and specifies a forward price process $F(t)$ and an instantaneous volatility $\alpha(t)$ process under the T-forward measure

$\mathbb{Q}^T$. For $0 < t < T$,

$$\begin{aligned} dF(t) &= \alpha(t)F^{\beta}(t)dW_1(t); \qquad F(0) = f \\ d\alpha(t) &= \nu\alpha(t)dW_2(t); \qquad \alpha(0) = \alpha \\ \mathbb{E}^{\mathbb{Q}^T}&[dW_1(t)dW_2(t)] = \rho dt; \end{aligned}$$

where the model parameters are:

$$\theta := (\alpha, \beta, \rho, \nu)$$

We further denote the dependence of the model's joint transition density on the model parameters explicitly by:

$$p(t, f, \alpha; T, F, A; \theta)$$

The constant-parameter setting is adequate for calibrating θ to an implied volatility curve at one fixed tenor T. In practice, market participants use different parameters for different maturities and recalibrate frequently, so parameters depend on T and t.

Let us denote the parameters' dependence on T by:

$$\theta(T) := (\alpha, \beta, \rho, \nu)(T)$$

Plain vanilla options are traded only on a finite set of tenors

$$0 < T_1 < T_2 < \cdots < T_N$$

at any time t, so we use a piecewise-constant (in T) parameterization. For a fixed date t, where $0 < t < T_1$, the parameter vector becomes a parameter matrix,

$$\theta(T) := \begin{cases} \theta_0, & t < T \le T_1 \\ \theta_{i-1}, & T_{i-1} < T \le T_i \\ \theta_{N-1} & T_{N-1} < T \le T_N \end{cases} \tag{3.1}$$

$$\text{where } \theta_i = (\alpha_i, \beta_i, \rho_i, \nu_i), \;\; i = 0, 1, \ldots, N-1$$

and the tenor-dependent SABR model then reads:

$$\begin{aligned} dF(t, T_i) &= \alpha(t)[F(t, T_i)]^{\beta_i}dW_1(t); \qquad F(0, T_i) = f_i \\ d\alpha(t, T_i) &= \nu_i\alpha(t, T_i)dW_2(t); \qquad \alpha(0, T_i) = \alpha_i \\ \mathbb{E}^{\mathbb{Q}^{T_i}}&[dW_1(t)dW_2(t)] = \rho_i dt; \end{aligned} \tag{3.2}$$

Accordingly, the dependence of the model's transition density on parameters becomes piecewise-constant:

$$\begin{aligned} &p(t, f, \alpha; T_1, F_1, A_1; \theta_0) \\ &p(T_{i-1}, F_{i-1}, A_{i-1}; T_i, F_i, A_i; \theta_{i-1}), \quad i = 2, \ldots, N-1 \\ &p(T_{N-1}, F_{N-1}, A_{N-1}; T_N, F_N, A_N; \theta_{N-1}) \end{aligned} \tag{3.3}$$

In the case where only the parameter dependence need to be stressed, we use the shortened notion:

$$p(t, T_1; \theta_0) \quad \text{and} \quad p(T_{i-1}, T_i; \theta_{i-1}), \quad i = 2, \ldots, N$$

3.2. *Synchronizing underlying and measure*

A standard SABR model describes the dynamics of a forward price process $F(t, T_i)$ maturing at a particular T_i. Forward prices associated with different maturities are martingales with respect to different forward measures defined by different zero-coupon bond prices $B(t, T_i)$ as numéraires. This raises consistency issues on both the underlying and the pricing measure when we work with multiple option maturities simultaneously.

We address this issue by consolidating all dynamics into those of $F(t, T_N), \alpha(t, T_N)$ in (3.2) whose tenor is the longest among all, and express all option prices at different tenors in one terminal measure $\mathbb{Q}^{T_N}$ which is the one associated with the zero-coupon bond price $B(t, T_N)$. We may do so because we assume

- No-arbitrage between spot price of a security $S(t)$ and all of its forward prices $F(t, T_i), i = 1, \ldots, N$ at all trading time t;
- Zero-coupon bonds are risk-less assets whose prices $B(t, T_i), i = 1, 2, \ldots, N$ have positive values.

A consequence of the first assumption is the standard relation

$$F(t, T_i) = \frac{S(t)}{B(t, T_i)}, \quad i = 1, 2, \ldots, N,$$

at any trading time t. Forward prices $F(t, T_i)$ at shorter maturities $T_1, \ldots, T_{N-1}$ could then be expressed in terms of the common underlying $F(t, T_N)$ as:

$$F(t, T_i) = F(t, T_N) \frac{B(t, T_N)}{B(t, T_i)}, \quad 1 \leq i \leq N - 1. \tag{3.4}$$

The second assumption reflects the that fact that we do not model interest rates as stochastic processes. This is a simplifying assumption in the context of currency options. For models that do model interest rates stochastically in the currency option context, see Amin and Jarrow [1] and Piterbarg [23]. Although volatilities of domestic and foreign interest rates could be important forces affecting currency options, we deem it reasonable to make this assumption for option tenors that are not too long.

To price a call option on $F(\cdot, T_i)$ with strike price K_j and maturity T_i, we use (3.4) and some simple algebra to arrive at

$$\begin{aligned} V(t, T_i, K_j) = B(t, T_N) \mathbb{E}^{\mathbb{Q}^{T_N}} [(F(T_i, T_N) - \mathcal{K}_j^i)^+ \mid \mathcal{F}_t] \\ \text{where } \mathcal{K}_j^i := \frac{K_j}{B(T_i, T_N)}. \end{aligned} \tag{3.5}$$

In other words, we convert an option on $F(\cdot, T_i)$ to an option on $F(\cdot, T_N)$.

3.3. *Bootstrapping vs. global optimization*

Since model parameters are piecewise constant in maturity, at each date t they can be calibrated through either a bootstrapping algorithm (in which the piecewise constant parameters are calibrated sequentially) or one based on global optimization (in which they are calibrated simultaneously). In either case, we recalibrate at each date t and do not seek to calibrate simultaneously across different dates t, only across different maturities T.

At any time of calibration t, let σ^{Mkt} be the market-quoted implied volatilities of liquid European options and let σ^{Model} be implied volatilities obtained by inverting option prices under the SABR model. For a given set of model parameters $\theta(T)$ as in (3.1), let $f(t, \theta(T))$ be the l_2 difference of implied volatility surface between market quotes and those obtained from the SABR model,

$$f(t, \theta(T)) := \sum_{i=1}^{N} \sum_{j=1}^{M_i} |\sigma^{\text{Model}}(t, T_i, K_j) - \sigma^{\text{Mkt}}(t, T_i, K_j)|^2(t, \theta(T))$$

The model is globally calibrated (at date t) when a set of optimal parameters, denoted by $\theta^*(T)$,

$$\theta^*(T) := \begin{cases} \theta_0^*, & t < T \le T_1 \\ \theta_{i-1}^*, & T_{i-1} < T \le T_i \\ \theta_{N-1}^* & T_{N-1} < T \le T_N \end{cases}$$

$$\text{where } \theta_i^* = (\alpha_i^*, \beta_i^*, \rho_i^*, \nu_i^*), \quad i = 0, 1, \ldots, N-1$$

is found over its feasible region Ω^N with Ω being

$$\Omega := (0, +\infty) \times (0, 1) \times (-1, 1) \times (0, +\infty)$$

such that f is minimized:

$$\theta^*(T) := \underset{\theta(T) \in \Omega^N}{\text{argmin}} f(t, \theta(T))$$

• Bootstrapping

The bootstrapping calibration is carried out sequentially in N steps according to option tenors. First obtain θ_0^* such that $f(t, \theta_0)$ is minimized for implied volatility curve at T_1 quoted at time t. Next given θ_0^*, obtain θ_1^* such that $f(t, [\theta_0^*; \theta_1])$ is minimized for tenor T_1. The procedure goes on until the last tenor T_N where θ_{N-1}^* is obtained such that $f(t, [\theta_0^*; \ldots; \theta_{N-2}^*; \theta_{N-1}])$ is minimized given the previous $N-1$ optimal parameter curves $[\theta_0^*; \ldots; \theta_{N-2}^*]$.

Step 1: At tenor T_1, find θ_0^* such that:

$$\theta_0^* := \underset{\theta_0 \in \Omega}{\text{argmin}} f(t, \theta_0)$$

Step 2: At tenor T_2, find θ_1^* given θ_0^* from step 1 such that:

$$\theta_1^* := \underset{\theta_1 \in \Omega}{\operatorname{argmin}} f(t, [\theta_0^*; \theta_1])$$

$$\vdots$$

Step N: At tenor T_N, find θ_{N-1}^* given $[\theta_0^*; \ldots; \theta_{N-2}^*]$ from step 1 to $N-1$ such that:

$$\theta_{N-1}^* := \underset{\theta_{N-1} \in \Omega}{\operatorname{argmin}} f(t, [\theta_0^*; \ldots; \theta_{N-2}^*; \theta_{N-1}])$$

• Global Optimization

Calibration based on global optimization seeks $\theta_0^*, \theta_1^*, \ldots, \theta_{N-1}^*$ all at the same time such that

$$[\theta_0^*, \theta_1^*, \ldots, \theta_{N-1}^*] := \underset{[\theta_0, \theta_1, \ldots, \theta_{N-1}] \in \Omega^N}{\operatorname{argmin}} f(t, [\theta_0, \theta_1, \ldots, \theta_{N-1}])$$

In the empirical studies we undertake later in Sec. 5, we use global optimization for model calibrations. When evaluating f in either the bootstrapping calibration or that based on global optimization, σ^{Model} are obtained by inverting model prices of European options according to (3.5) which, in practice, are calculated by integrating payoffs against transition densities in (3.3). At tenor T_1,

$$V(t, T_1, K_j) = B(t, T_N) \iint_{R_+^2} (F_1 - \mathcal{K}_j^i)^+ p(t, T_1; \theta_0) dF_1 dA_1 \tag{3.6}$$

and generally at tenor $T_i, i = 2, \ldots, N$,

$$V(t, T_i, K_j) = B(t, T_N) \iint_{R_+^2} \left[\cdots \left[\iint_{R_+^2} (F_i - \mathcal{K}_j^i)^+ p(T_{i-1}, T_i; \theta_{i-1}) dF_i dA_i \right] \cdots \right] \times p(t, T_1; \theta_0) dF_1 dA_1 \tag{3.7}$$

4. Computing Conditional Expectations

Have specified the model and articulated its calibration, we now provide details on using the truncated series expression of the density obtained in Wu [26] of the model's joint transition density for fast computation of expectations and conditional expectations that arise in both model calibration and calculation of various forward implied volatility measures.

4.1. *Conditional expectations*

In model calibration, computing spot implied volatilities from the model requires computing option prices as in (3.6) and (3.7):

$$\begin{aligned} &\mathbb{E}^{\mathbb{Q}^{T_N}} [(F(T_i) - \mathcal{K}_j)^+ \mid F_{i-1}, A_{i-1}] \\ &\quad = \iint_{R_+^2} (F_i - \mathcal{K}_j)^+ p(T_{i-1}, T_i; \theta_{i-1}) dF_i dA_i \end{aligned} \tag{4.1}$$

at each tenor $T_i, i = 1, \ldots, N$ for each equivalent strike $\mathcal{K}_j, j = 1, \ldots, M_i$.

Once the model is calibrated, computing model-based forward implied volatilities in (2.3), (2.4)–(2.5) and (2.6)–(2.7) also boils down to computing:

$$\begin{aligned}&\mathbb{E}^{\mathbb{Q}^{T_N}}[(F(T_i)-kF(T_{i-1}))^+ \mid F_{i-1}, A_{i-1}]\\ &\quad = \iint_{R_+^2} (F_i - kF_{i-1})^+ p(T_{i-1}, T_i; \theta_{i-1}) dF_i dA_i\end{aligned} \tag{4.2}$$

over any speriod $[T_{i-1}, T_i], i = 2, \ldots, N-1$.

4.2. *Numerical integration*

In both (4.1) and (4.2), we need to evaluate two-dimensional integrations of payoff functions that depend on state variables at T_i or at both T_{i-1} and T_i. To include both cases, we consider payoff functions depending on two arbitrary temporal points t and T with $t < T$ and denote values of state variables by f, α at t and F, A at T.

Let

$$\Phi(F(T), F(t))$$

be an arbitrary payoff function of $F(T)$ and $F(t)$ and let $I_{(f,\alpha)}(t, T, \theta)$ be its time-t conditional expectation which is:

$$\begin{aligned}I_{(f,\alpha)}(t,T,\theta) &:= \mathbb{E}[\Phi(F(T), F(t)) | F(t) = f, \alpha(t) = \alpha]\\ &= \iint_{\Omega(F,A)} \Phi(F, f) p(t, f, \alpha; T, F, A; \theta) dF dA\end{aligned} \tag{4.3}$$

Both (4.1) and (4.2) could then be cast as instances of $I_{(f,\alpha)}(t, T, \theta)$.

Asymptotic expansions of the joint transition density $p(t, f, \alpha; T, F, A; \theta)$ have been obtained analytically in Wu [26] to the nth order as:

$$p_n(t,f,\alpha;T,F,A;\theta) = \frac{1}{\nu T F^\beta A^2} \sum_{k=0}^{n} (\nu\sqrt{T})^k \hat{p}_k(\tau, u, v) \quad \text{where}$$

$$\tau := \frac{T-t}{T}; \quad u := \eta(F) = \frac{f^{1-\beta} - F^{1-\beta}}{\alpha(1-\beta)\sqrt{T}}; \quad v := \zeta(v) = \frac{\ln(\alpha/A)}{\nu\sqrt{T}}$$

where $\hat{p}_k(\tau, u, v)$ are obtained recursively in [26]. As demonstrated in [26], the expansion to its second order p_2 is a quite accurate approximation of p for a wide range of model parameters and values of state variables as well.

When computing (4.3), we will first replace $p(t, f, \alpha; T, F, A; \theta)$ by the asymptotic expansion to 2nd order whose analytical form is $p_2(t, f, \alpha; T, F, A; \theta)$ obtained in [26] for an approximation of (4.3) and then evaluate the resulting approximation either by a direct numerical integration or through quadrature rules. We do not recommend using just the first-order expansion. The difference between p_1 and p_2 is that most of the tail details of the density function are captured by p_2 but not p_1,

without which skew and/or smile features of implied volatility are hard to capture correctly. Quoted from Eq. (40) in [26], $p_2(t, f, \alpha; T, F, A; \theta)$ reads:

$$\begin{aligned} p_2(t,f,\alpha;T,F,A;\theta) &= \frac{1}{\nu T F^{\beta} A^2}[\hat{p}_0 + \nu\sqrt{T}\hat{p}_1 + \nu^2 T\hat{p}_2](\tau,u,v,\theta), \quad \text{with} \\ \hat{p}_0(\tau,u,v,\theta) &= \frac{1}{2\pi\tau\sqrt{1-\rho^2}}\exp\left[-\frac{u^2-2\rho uv+v^2}{2\tau(1-\rho^2)}\right] \\ \hat{p}_1(\tau,u,v,\theta) &= g_1(\tau,u,v,\theta)\hat{p}_0(\tau,u,v,\theta) \\ \hat{p}_2(\tau,u,v,\theta) &= g_2(\tau,u,v,\theta)\hat{p}_0(\tau,u,v,\theta) \end{aligned} \tag{4.4}$$

where g_1 and g_2 are given by:

$$\begin{aligned} g_1(\tau,u,v) &= \frac{a_{11}+a_{10}/\tau}{2(-1+\rho^2)} \\ g_2(\tau,u,v) &= \frac{a_{23}\tau + a_{22} + a_{21}/\tau + a_{20}/\tau^2}{24(1-\rho^2)^2} \end{aligned}$$

To save space, we refer the reader to Eq. (42) in [26] for explicit expressions for the polynomial functions a_{11}, a_{10} and $a_{23}, a_{22}, a_{21}, a_{20}$.

A straightforward way of using this analytical result is to plug (4.4) into (4.3) and use one's favorite routine for a direct numerical integration, such as a trapezoidal rule with sub-intervals of equal size where the integral is taken over a truncated domain $\Omega(F, A)$ of the support of the original variable F, A, which is $\mathbb{R}^2_+$. The computing time required for each approximate double integral in (3.7) is quadratic in the number of grid points, and the total time required for (3.7) is linear in the number of maturities N.

If one is not tightly constrained by CPU time or computer memory, direct numerical integration with a sufficiently fine discretization of the domain of integration achieves high accuracy and is reasonably fast. This is how we carry out our model calibrations in the empirical studies we undertake in Sec. 5. It takes about 1–10 milliseconds for an evaluation of (4.3) on a 1000 by 1000 grid. Similar integrals have to be calculated repeatedly in the course of calibration.

To accelerate the calculation, the Gaussian structure in the density expansion can be exploited. In the transformed variables τ, u, v, the leading order density $\hat{p}_0$ is a bivariate Gaussian distribution and corrections at the next two orders $\hat{p}_1, \hat{p}_2$ are of a product form of polynomial functions g_1, g_2 and the leading order $\hat{p}_0$.

In the following, we derive expressions of $I_{(f,\alpha)}(t, T, \theta)$ in the transformed variables τ, u, v that are convenient for Gaussian-quadrature-based numerical integration to compute (4.3).

First, let us express the original integration variables F, A in terms of u, v as:

$$\begin{aligned} F &= \eta^{-1}(u) = [f^{1-\beta} - \alpha(1-\beta)\sqrt{T}\cdot u]^{\frac{1}{1-\beta}} \\ A &= \zeta^{-1}(v) = \alpha\cdot\exp(-\nu\sqrt{T}\cdot v) \end{aligned} \tag{4.5}$$

Plugging (4.5) into (4.4) and after some algebra, we consolidate p_2 in τ, u, v as:

$$\begin{aligned} p_2(t,f,\alpha,T,F,A,\theta) &= J^2_{(f,\alpha)}(\tau,u,v,\theta) \times \hat{p}_0(\tau,u,v) \quad \text{where,} \\ J^2_{(f,\alpha)}(\tau,u,v,\theta) &:= \frac{1+\nu\sqrt{T}\times g_1(\tau,u,v,\theta)+(\nu\sqrt{T})^2\times g_2(\tau,u,v,\theta)}{\nu T\times(\eta^{-1}(u))^{\beta}\times(\zeta^{-1}(v))^2} \end{aligned} \tag{4.6}$$

With (4.6) plugged into (4.3), we then express $I_{(f,\alpha)}(t,T,\theta)$ in τ, u, v as:

$$\begin{aligned} I_{(f,\alpha)}(t,T,\theta) &\approx \iint_{\Omega(u,v)} H_{(f,\alpha)}(\tau,u,v,\theta)\times\hat{p}_0(\tau,u,v)dudv \quad \text{where,} \\ H_{(f,\alpha)}(\tau,u,v,\theta) &= \Phi(\eta^{-1}(u),f)\times J^2_{(f,\alpha)}(\tau,u,v,\theta)\times(\eta^{-1}(u))'\times(\zeta^{-1}(v))' \end{aligned} \tag{4.7}$$

Finally, we slightly enlarge the domain of integration $\Omega(u,v)$ to a rectangular one such that it covers $\Omega(u,v)$ which is usually a curved one. Then we are ready to apply a two-dimensional Gaussian quadrature rule to (4.7) as $I_{(f,\alpha)}(t,T,\theta)$ now becomes an integration of a bivariate Gaussian kernel $\hat{p}_0$ against an analytical function H whose expression is explicit and differentiable with respect to u, v. For instance, let the slightly-enlarged rectangular domain be

$$\Omega'(u,v) = [\underline{u},\overline{u}]\times[\underline{v},\overline{v}]$$

then (4.7) can be evaluated as

$$\begin{aligned} I_{(f,\alpha)}(t,T,\theta) &\approx \int_{\underline{v}}^{\overline{v}}\int_{\underline{u}}^{\overline{u}} H_{(f,\alpha)}(\tau,u,v,\theta)\times\hat{p}_0(\tau,u,v)dudv \\ &\approx \sum_i\sum_j W_{ij}\times H_{f,\alpha}(\tau,u_i,v_j,\theta)\times\hat{p}_0(\tau,u_i,v_j) \end{aligned}$$

where one could choose the weights W_{ij} according to a particular quadrature rule.

5. Application to Currency Options

Having laid out the framework, we now report empirical findings based on currency option market data. We first examine whether the term structure of today's option prices across multiple tenors contains predictive information about future option prices. We then examine the relative predictive merits of spot volatility, model-based forward volatility, and model-free forward volatility. We test both in-sample and out-of-sample performance.

5.1. *Data description*

Option quotes in the foreign exchange market are expressed as Garman-Kohlhagen [17] implied volatilities for fixed tenors and for fixed Garman-Kohlhagen deltas. The deltas at which implied volatilities are quoted can be converted to strikes.

Our data covers options on the euro-dollar, sterling-dollar and dollar-yen rates from September 24, 2001 to June 16, 2009 and includes daily quotes at four tenors — 3 months, 6 months, 9 months, 1 year — and five strikes — the 10 and 25 put delta,

the at-the-money call, and the 10 and 25 call deltas. For each of the three currency pairs, the data set consists of 20 time series, each corresponding to a fixed tenor and delta, and we have 60 such quoted time series in total.

Our data also includes domestic and foreign LIBOR rates, spot exchange rates, forward exchange rates, and the relevant option strikes calculated according to the Garman-Kohlhagen deltas, at the same calender dates as those of the option quotes. Table 1 summarizes the statistics of the data set.

5.2. *Model calibration*

We use the global optimization method of Sec. 3.3 for calibration. On each calendar date, we use quoted implied volatility curves at two relevant option tenors to calibrate the model. Within the calibration algorithm, we use the trust-region-reflective method in [8] for global optimization with the objective function being the l_2 difference between the ten implied volatility quotes, five at each of the two tenors, and those from the model-generated ones. We further supplement the objective function with a Tikhonov regularization term, see Chapter 5 of [16], in terms of the l_2 norm of the model parameters, to stabilize convergence.

The optimization algorithm is stopped when either one of the following three criteria is met: the residual norm is less than 10^{-6}, relative changes in the Jacobian of the objective function are less than 10^{-6}, or the maximum number of iterations (set at 10000) is reached.

When calculating the model-based implied volatilities, we use a 1000 by 1000 grid for numerical integration of joint transition densities against payoff functions, whenever conditional expectations are involved. On a 2.93 GHz Xeon workstation, each calibration takes 30 seconds to 3 minutes to converge, and each 2-dimensional numerical integration takes 1–10 milliseconds.

The most CPU-intensive components within the optimization process are the total number of searching iterations for the optimization to converge, the inversion of option prices to implied volatilities whenever needed, and the calculation of model prices at the second tenor given a set of tenor-dependent parameters, due to the nested nature of this step. Table 2 reports performances of all calibrations.

5.3. *Regressions*

After calibrating the model, we calculate both the model-based expected forward implied volatilities, Expt-Fwd-IV, and the model-free ones, Mdfr-Fwd-IV, at all five deltas (P10d, P15d, ATM, C25d, C10d), all three forward horizons (3-into-6 months, 6-into-9 months, and 9-into-12 months), and for all three currency pairs on a given calender day.

When calculating forward implied volatilities using today's calibrated model, we use absolute strikes from those of the future spot implied volatilities at the relevant Garman-Kohlhagen deltas. Next, we assemble time series from future spot quotes

Table 1. Summary statistics of currency option spot implied volatility quotes.

	P10d				P25d				ATM				C25d				C10d			
	Mean	Std	Min	Max	Mean	Std	Min	Max	Mean	Std	Min	Max	Mean	Std	Min	Max	Mean	Std	Min	Max
EURUSD																				
3M	10.81	3.75	5.57	28.75	10.22	3.31	5.23	25.66	10.06	3.11	5.04	24.50	10.42	3.22	5.10	25.64	11.17	3.56	5.41	28.22
6M	10.99	3.52	5.96	26.01	10.32	3.02	5.60	23.11	10.14	2.82	5.42	22.25	10.55	2.94	5.57	23.31	11.37	3.35	5.96	25.98
9M	11.09	3.40	6.06	24.73	10.38	2.87	5.66	21.89	10.18	2.66	5.55	20.90	10.60	2.80	5.79	22.00	11.48	3.23	6.25	24.96
1Y	11.18	3.37	6.23	24.36	10.41	2.78	5.76	21.10	10.19	2.57	5.61	20.00	10.64	2.70	5.91	21.12	11.57	3.17	6.42	24.21
GBPUSD																				
3M	10.20	4.48	5.66	32.61	9.47	3.80	5.23	28.86	9.13	3.34	5.00	26.00	9.29	3.21	5.11	24.57	9.83	3.34	5.48	24.32
6M	10.31	4.15	6.04	28.09	9.53	3.42	5.60	24.52	9.17	2.97	5.39	22.00	9.35	2.85	5.57	20.96	9.93	3.02	5.99	22.54
9M	10.39	4.05	6.23	27.47	9.58	3.28	5.77	23.24	9.19	2.81	5.57	20.60	9.40	2.70	5.78	20.15	10.01	2.92	6.23	22.13
1Y	10.46	4.04	6.33	27.22	9.61	3.20	5.85	22.65	9.20	2.72	5.65	20.00	9.43	2.62	5.93	19.97	10.08	2.88	6.42	22.16
USDJPY																				
3M	12.96	5.04	7.56	42.63	11.21	3.74	6.85	33.73	9.97	2.78	6.13	27.65	9.53	2.33	5.81	24.73	9.76	2.30	5.72	24.63
6M	13.08	4.68	8.08	37.22	11.08	3.22	7.14	28.08	9.71	2.26	6.36	22.05	9.24	1.89	6.00	19.34	9.48	1.95	6.01	19.36
9M	13.18	4.51	8.48	34.24	11.02	2.96	7.35	24.88	9.57	2.01	6.46	18.92	9.09	1.70	6.07	16.30	9.38	1.83	6.10	16.29
1Y	13.31	4.45	8.76	32.29	11.00	2.80	7.34	22.94	9.48	1.86	6.53	17.17	9.00	1.60	6.15	14.67	9.35	1.77	6.09	14.77

Note: The four columns under each currency pair report the mean (Mean), standard deviation (Std), minium (Min) and maximum (Max) of the quoted implied volatilities at 10-delta put (P10d), 25-delta put (P25d), at-the-money (ATM), 25-delta call (C25d), and 10-delta call (C10d). The units are absolute values of implied volatility, expressed in percentage points. The range of the data is from August 9, 2001 to June 16, 2009, consisting 2049 data points. The first column denotes option tenors under relevant currency pair and the first row denotes deltas.

Table 2. Summary statistics of model calibrations.

	Tenor T_1										Tenor T_2									
	P10d		P25d		ATM		C25d		C10d		P10d		P25d		ATM		C25d		C10d	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
EURUSD																				
3M6M	0.84	0.88	0.59	0.77	0.52	0.49	0.47	0.42	0.61	0.60	2.37	1.70	5.91	5.68	4.52	5.64	0.95	1.17	1.34	1.61
6M9M	0.93	1.00	0.65	0.89	0.62	0.63	0.66	0.60	0.79	0.86	1.96	1.47	4.70	4.70	3.83	4.14	1.00	0.99	1.23	1.03
9M1Y	1.01	1.04	0.66	0.92	0.60	0.69	0.69	0.64	0.89	0.93	1.53	1.25	3.76	3.92	3.14	3.39	1.06	1.35	1.09	0.89
GBPUSD																				
3M6M	1.93	1.81	1.79	1.48	0.97	0.83	0.46	0.42	1.38	1.05	0.76	1.11	1.64	3.06	1.74	2.61	1.47	1.85	0.93	1.10
6M9M	1.85	1.88	1.70	1.53	0.87	0.79	0.50	0.59	1.40	1.32	0.80	1.14	1.61	2.96	1.62	2.53	1.40	1.76	0.88	1.10
9M1Y	1.75	1.21	1.96	1.14	1.67	0.90	0.98	0.72	0.79	0.95	0.74	0.77	1.43	1.76	1.22	1.30	1.30	1.34	0.61	0.50
USDJPY																				
3M6M	0.51	0.69	0.12	0.20	0.25	0.29	0.21	0.32	0.23	0.19	0.23	0.14	0.94	0.46	0.85	0.52	0.19	0.30	0.17	0.27
6M9M	0.95	0.76	0.50	0.93	0.58	0.88	0.35	0.46	0.47	0.68	0.28	0.46	0.94	0.56	1.05	0.49	0.78	1.04	0.37	0.63
9M1Y	1.10	0.87	0.62	1.10	0.70	1.04	0.43	0.62	0.62	0.78	0.32	0.64	0.83	0.68	1.06	0.52	0.95	1.11	0.44	0.73

Note: Each column reports the mean (Mean) and standard deviation (Std) of residual errors of model calibrations in units of absolute values of implied volatilities, expressed percentage points. Residual errors are from daily calibrations from August 9, 2001 to April 8, 2009, consisting of 2000 days. At each day, implied volatility quotes from two relevant tenors (Tenor T_1, Tenor T_2), each with 5 data points at 5 deltas (P10d, P25d, ATM, C25d, C10d), are inputs to the calibration. The first column denotes the relevant option tenors with "M" denoting month and "Y" denoting "year".

(Fut-Spot-IV), the two forward ones (Expt-Fwd-IV, Mdfr-Fwd-IV), and today's spot quotes (Tdy-Spot-IV) for regressions.

In both in-sample and out-of-sample tests, we take Fut-Spot-IV as dependent variable and use Expt-Fwd-IV, Mdfr-Fwd-IV and Tdy-Spot-IV as explanatory variables. For in-sample tests, regression performance is measured by the root-mean-square-absolute-error (RMSAE) and R^2 values. For out-of-sample tests, error metrics are root-mean-square-absolute-error (RMSAE) and root-mean-square-percentage-error (RMSPE).

In-sample tests are reported in Table 3. We carry out 135 regressions: for a given forecasting horizon, delta, and currency pair, we have three individual regressions. The dependent variable is the future spot implied volatility time series and the explanatory variables are the vector series of Expt-Fwd-IV curve, Mdfr-Fwd-IV curve and Tdy-Spot-IV curve, all involving five deltas for the same horizon and

Table 3. Summary statistics of in-sample regressions.

	Tdy-Spot-IV Curve		Mdfr-Fwd-IV Curve		Expt-Fwd-IV Curve	
	RMSAE	R^2	RMSAE	R^2	RMSAE	R^2
3M6M	0.0024	0.6495	0.0022	0.6738	0.0016	0.8521
6M9M	0.0046	0.4465	0.0036	0.5394	0.0016	0.7552
9M1Y	0.0065	0.3950	0.0058	0.5167	0.0021	0.6852
P10d	0.0105	0.4826	0.0094	0.5553	0.0042	0.7098
P25d	0.0044	0.4923	0.0038	0.5692	0.0017	0.7386
ATM	0.0025	0.4995	0.0020	0.5819	0.0010	0.7622
C25d	0.0022	0.5026	0.0018	0.5880	0.0009	0.7810
C10d	0.0029	0.5077	0.0023	0.5888	0.0012	0.7842
EURUSD	0.0042	0.5383	0.0030	0.6672	0.0013	0.8474
GBPUSD	0.0042	0.5015	0.0037	0.5745	0.0017	0.7662
USDJPY	0.0052	0.4511	0.0049	0.4883	0.0024	0.6519

Note: Entries report in-sample regression performance in terms of the root-mean-squared-absolute-error (RMSAE) in unites of absolute values of implied volatilities, expressed in percentage points, and its associated R^2 values. The dependent variable is the future spot implied volatility (Fut-Spot-IV) series. The explanatory variables are vector series of implied volatility curve from today's spot one (Tdy-Spot-IV curve), the model-free forward one (Mdfr-Fwd-IV curve), and the model-based expected forward one (Expt-Fwd-IV curve) at all 5 deltas for a relevant horizon and currency pair. The first 3 rows summarize statistics according to the forecasting horizons. Each entry of RMSAE and R^2 is the arithmetic mean of 15 individual ones from regression at all 5 deltas and all 3 currency pairs for a shared horizon. The second 5 rows summarize results according to the deltas. Each entry is the arithmetic mean of 9 individual ones from regressions at all 3 horizons and all 3 currency pairs for a shared delta. The last 3 rows summarize results according to the currency pairs and the average is taken over 15 individual ones from all 3 horizons and all 5 deltas for a shared currency pair. The first column denotes the relevant groups and the first row denotes relevant explanatory variables.

currency pair. The length of data are 1934 from March 16, 2009 to September 24, 2001 for all 3M6M cases, 1868 from December 12, 2008 to September 24, 2001 for all 6M9M cases and 1802 from September 11, 2008 to September 24, 2001 for all 9M1Y cases.

All in-sample regressions are summarized into three groups according to the forecasting horizon, delta and currency pair. In the horizon group, we report the arithmetic mean of RMSAE and R^2 over fifteen individual regressions from all five deltas and all three currency pairs for a shared horizon. In the delta group, we report the arithmetic mean over nine individual regressions from all three horizons and all three currency pairs at a shared delta. And in the currency pair group, the reported RMSAE and R^2 is taken over fifteen individual ones from all three horizons and all five deltas of the shared currency pair.

Out-of-sample tests are reported in Tables 4 and 5. The dependent variables and the explanatory variables are the same as in the in-sample case.

Table 4 reports out-of-sample forecasting results from 135 one-day-ahead rolling regressions with a one year rolling window where for a given forecasting horizon, delta, and currency pair, three regressions are carried out. In the rolling regressions, both absolute volatility differences and relative percentage changes between the one-day-ahead out-of-sample forecasts and the actual ones are recorded to compute the RMSAE in absolute differences of volatility points and RMSPE in relative percentage changes.

Table 5 then summarizes the forecasting enhancements of Expt-Fwd-IV over Mdfr-Fwd-IV and Tdy-Spot-IV into the same three groups as in the in-sample case and using the same arithmetic averages.

5.4. *Observations and conclusions*

5.4.1. *Predicative information embedded in the option quotes*

Findings regarding the predictive information embedded in the liquid options are summarized below. We begin with overall observations on the predictive information in current option prices before distinguishing the relative performance of alternative predictors:

- Today's option prices do contain predicative information about future prices, and this holds true for all forecasting horizons, at all deltas, and on all currency pairs;
- Forecasts of future implied volatility are more predictive, other things equal, for shorter forecasting horizons than longer horizons;
- Forecasts of future implied volatility are most predicative for the euro-dollar pair, with sterling-dollar second, and dollar-yen last.

The first point is supported by the overall observation in the tables that a significant amount of predicative information regarding Fut-Spot-IV has been extracted from today's option quotes using all three explanatory variables. This is evidenced by both in-sample and out-of-sample tests. In the in-sample tests, 76%, 58% and

Table 4. Summary statistics of out-of-sample regressions.

	Expt-Fwd-IV Curve					Mdfr-Fwd-IV Curve					Tdy-Spot-IV Curve				
	P10d	P25d	ATM	C25d	C10d	P10d	P25d	ATM	C25d	C10d	P10d	P25d	ATM	C25d	C10d
EURUSD															
3M6M	0.75	0.57	0.48	0.48	0.54	1.50	1.10	1.02	1.20	1.70	1.75	1.25	1.13	1.32	1.92
6M9M	0.53	0.39	0.32	0.31	0.34	4.16	2.44	1.96	2.35	3.87	3.02	1.82	1.50	1.82	3.00
9M1Y	0.66	0.54	0.50	0.51	0.60	3.23	1.86	1.45	1.58	2.43	3.52	2.10	1.58	1.62	2.19
GBPUSD															
3M6M	0.50	0.37	0.31	0.29	0.30	7.29	4.03	2.68	2.48	3.09	2.18	1.37	1.06	1.05	1.27
6M9M	0.60	0.44	0.36	0.33	0.34	1.01	0.70	0.59	0.64	0.81	1.21	0.79	0.64	0.66	0.83
9M1Y	1.90	1.10	0.72	0.60	0.63	11.15	3.82	1.72	1.21	1.11	15.50	4.67	1.94	1.45	1.42
USDJPY															
3M6M	1.34	0.65	0.41	0.35	0.40	6.52	3.08	1.59	1.16	1.11	5.77	2.82	1.55	1.15	1.14
6M9M	1.28	0.75	0.53	0.47	0.51	11.50	3.21	1.19	0.83	0.86	6.41	1.98	0.94	0.78	0.84
9M1Y	1.90	1.10	0.72	0.60	0.63	11.15	3.82	1.72	1.21	1.11	15.50	4.67	1.94	1.45	1.42

Note: Entries report out-of-sample regression performance in terms of the root-mean-squared-absolute-error (RMSAE) in unites of absolute values of implied volatilities, expressed in percentage points. The dependent variable is the future spot implied volatility at each of the 3 horizons, 5 deltas, and 3 currency pairs. The explanatory variables are vector series of implied volatility curve from today's spot one (Tdy-Spot-IV curve), the model-free forward one (Mdfr-Fwd-IV curve), and the model-based expected one (Expt-Fwd-IV curve) at all 5 deltas for a relevant horizon and currency pair. Each regression uses one year rolling window and the differences between the one-day-ahead forecasts and the actual ones are recorded to compute the RMSAE. The first column denotes explanatory variables at different horizon and currency pair and the first row denotes explanatory variables at different delta.

Table 5. Summary statistics of forecasting enhancements of model-based forward implied volatilities.

	Enhancement of Expt-Fwd-IV over Tdy-Spot-IV						Enhancement of Expt-Fwd-IV over Mdfr-Fwd-IV					
	RMSAE			RMSPE			RMSAE			RMSPE		
	Expt	Spot	Ehmt	Expt	Spot	Ehmt	Expt	Mdfr	Ehmt	Expt	Mdfr	Ehmt
3M6M	0.51	1.78	1.27	35.60%	128.67%	93.06%	0.51	2.64	2.12	35.60%	149.10%	113.50%
6M9M	0.50	1.75	1.25	43.46%	115.18%	71.72%	0.50	2.41	1.91	43.46%	142.45%	99.00%
9M1Y	0.85	4.06	3.22	73.25%	267.08%	193.83%	0.85	3.24	2.39	73.25%	217.06%	143.81%
P10d	1.05	6.10	5.05	56.73%	258.08%	201.35%	1.05	6.39	5.34	56.73%	258.81%	202.08%
P25d	0.65	2.39	1.73	51.24%	179.11%	127.87%	0.65	2.67	2.02	51.24%	182.12%	130.88%
ATM	0.48	1.36	0.88	48.25%	142.73%	94.48%	0.48	1.55	1.07	48.25%	142.99%	94.75%
C25d	0.44	1.25	0.82	46.93%	134.55%	87.62%	0.44	1.41	0.97	46.93%	130.85%	83.92%
C10d	0.48	1.56	1.08	47.73%	137.06%	89.33%	0.48	1.79	1.31	47.73%	132.92%	85.19%
EURUSD	0.50	1.97	1.47	41.86%	129.86%	88.00%	0.50	2.12	1.62	41.86%	130.51%	88.65%
GBPUSD	0.58	2.40	1.82	52.07%	172.03%	119.96%	0.58	2.82	2.24	52.07%	168.91%	116.85%
USDJPY	0.77	3.22	2.45	56.61%	209.04%	152.43%	0.77	3.34	2.56	56.61%	209.19%	152.58%

Note: Entries report out-of-sample forecasting enhancements of model-based expected forward implied volatility measure (Expt-Fwd-IV), shortened to "Expt", over today's spot implied volatility (Tdy-Spot-IV), shortened to "Spot", and the model-free forward implied volatility measure (Mdfr-Fwd-IV), shortened to "Mdfr", in terms of both the root-mean-squared-absolute-error (RMSAE) in unites of absolute values of implied volatilities, expressed in percentage points and the root-mean-squared-percentage-error (RMSPE). Enhancements are summarized according to forecasting horizons (3M6M, 6M9M, 9M1Y), implied volatility deltas (P10d, P25d, ATM, C25d, C10d) and currency pairs (EURUSD, GBPUSD, USDJPY). For the horizon group, enhancements are calculated as the arithmetically average of absolute differences of RMSAE and RMSPE from table 3 over 15 individual ones across all 5 deltas and 3 currency pairs for the same horizon. For the delta group, the average is over 9 individual ones across all 3 horizons and 3 currency pairs. And for the currency pair group, it is over 15 individual ones across all 3 horizons and 5 deltas. The first column denotes the relevant groups. The left half of the table summarizes forecasting enhancements of Expt-Fwd-IV over Tdy-Spot-IV and the right half Expt-Fwd-IV over Mdfr-Fwd-IV.

50% of Fut-Spot-IV variance is explained respectively by Expt-Fwd-IV, Mdfr-Fwd-IV, and Tdy-Spot-IV, where, in each case, the results are averaged over all forward horizons, all deltas and all currency pairs. In the out-of-sample tests, 0.62, 2.53 and 2.76 volatility points, on average, in RMSAE occurred respectively using Expt-Fwd-IV, Mdfr-Fwd-IV, and Tdy-Spot-IV as explanatory variables.

The second point above is evidenced by the fact that regression performance is better at the short end of the forecasting horizons than at the long end, as one would expect. In the in-sample tests, 73%, 56% and 54% of Fut-Spot-IV variance is explained for 3M6M, 6M9M and 9M1Y respectively, each of which are averaged over all three explanatory variables, all five deltas, and all three currency pairs for a given forecasting horizon. In the out-of-sample tests, 1.64, 1.55 and 2.71 volatility points in RMSAE occurred respectively.

The third point is also evident from the regression results. In particular, we observe 68%, 61% and 53% of Fut-Spot-IV variance is explained in-sample for EURUSD, GBPUSD and USDJPY respectively and 1.53, 1.93 and 2.44 volatility points occurred respectively in out-of-sample RMSAE, each of which is averaged over all three explanatory variables and over all five deltas and all three forecasting horizons.

5.4.2. *Relative performance of forecasts*

Our findings regarding the relative merits of each of the explanatory variables on extracting this embedded inforation are summarized below.

- Model-based forward implied volatility extracts substantially more information from today's prices for prediction of future spot prices, and this holds true for all forecasting horizons, at all deltas, and in all currency pairs;
- The improvement from using model-based forecasts is greater at longer forecasting horizons;
- The improvement is greatest for the dollar-yen pair, with sterling-dollar next, and euro-dollar last.

The first point is evidenced by the observation that regressions using Expt-Fwd-IV as explanatory variable produce the largest R^2 values and the smallest RMSAE in in-sample tests and the smallest RMSAE and RMSPE in out-of-sample test, compared to those using Tdy-Spot-IV and Mdfr-Fwd-IV as explanatory variables.

In terms of in-sample performance, this enhancement of Expt-Fwd-IV over Mdfr-Fwd-IV is as large as 72% in RMSAE and 65% in R^2 across the three forecasting horizons, 55% in RMSAE and 37% in R^2 across five deltas, and finally 56% in RMSAE and 33% in R^2 across three currency pairs. Comparing Expt-Fwd-IV to Tdy-Spot-IV, this enhancement is as large as 75% in RMSAE and 91% in R^2 across horizons, 61% in RMSAE and 62% in R^2 across deltas, and finally 68% in RMSAE and 57% in R^2 across currency pairs.

Out-of-sample performance is consistent with this observation. Forecasts using Expt-Fwd-IV have the smallest RMSAE which is 0.62 in volatility points and the smallest RMSPE which is 50%, averaged over all out-of-sample forecasting horizons, all deltas and all currency pairs. The averaged RMSAE and RMSPE is 2.76 and 169% for Mdfr-Fwd-IV. And for Tdy-Spot-IV it is 2.53 and 170%. Thus on average, the out-of-sample forecasting enhancement of Expt-Fwd-IV over Mdfr-Fwd-IV is 2.14 volatility points in RMSAE and 119% in RMSPE. The enhancement of Expt-Fwd-IV over Tdy-Spot-IV and is 1.91 volatility points in RMSAE and 119% in RMSPE.

The second and third point are evidenced by the fact that the out-of-sample forecasting enhancements of Expt-Fwd-IV over Tdy-Spot-IV and Mdfr-Fwd-IV are the largest for the longest forecasting horizon (9M1Y) and are the largest for the dollar-yen pair among all three currency pairs under investigation.

Across horizons, these enhancements are 1.27, 1.25, 3.22 in RMSAE and 93%, 71%, 194% in RMSPE for Expt-Fwd-IV over Tdy-Spot-IV at 3M6M, 6M9M, 9M1Y respectively. And for Expt-Fwd-IV over Mdfr-Fwd-IV, they are $2.12, 1.91, 2.39$ in RMSAE and 114%, 99%, 144% in RMSPE respectively.

Across currency pairs, these enhancements are $1.47, 1.82, 2.45$ in RMSAE and 88.00%, 119.96%, 152.43% in RMSPE for Expt-Fwd-IV over Tdy-Spot-IV in EURUSD, GBPUSD, USDJPY respectively. And for Expt-Fwd-IV over Mdfr-Fwd-IV, they are 1.62, 2.24, 2.56 in RMSAE and 88.65%, 116.85%, 152.58% in RMSPE respectively.

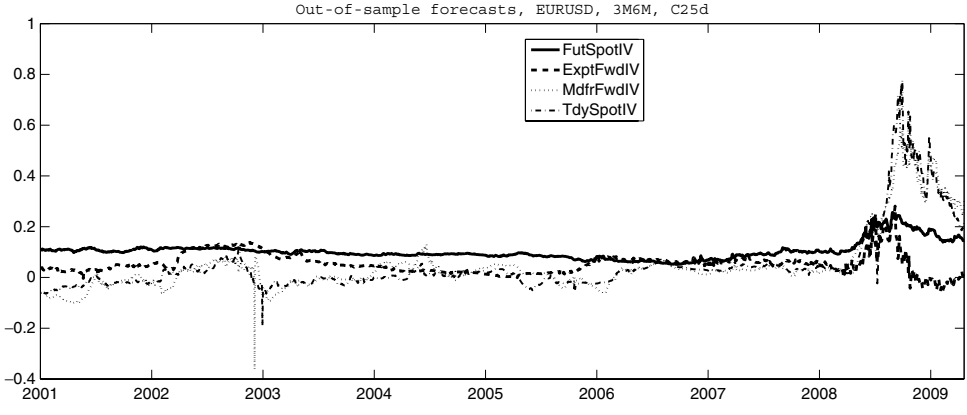

Fig. 1. The illustrative case is with respect to the euro-dollar pair (EURUSD) for 3 months-into-6 months (3M6M) forecasting horizon at 25 call delta (C25d). The picture shows time series plots of one-day-ahead one-year rolling out-of-sample forecasts using future spot implied volatility (FutSpotIV) as dependent variable. The explanatory variables are vector series of implied volatility curve at all five deltas from the model-based expected forward one, the model-free forward one and today's spot one.

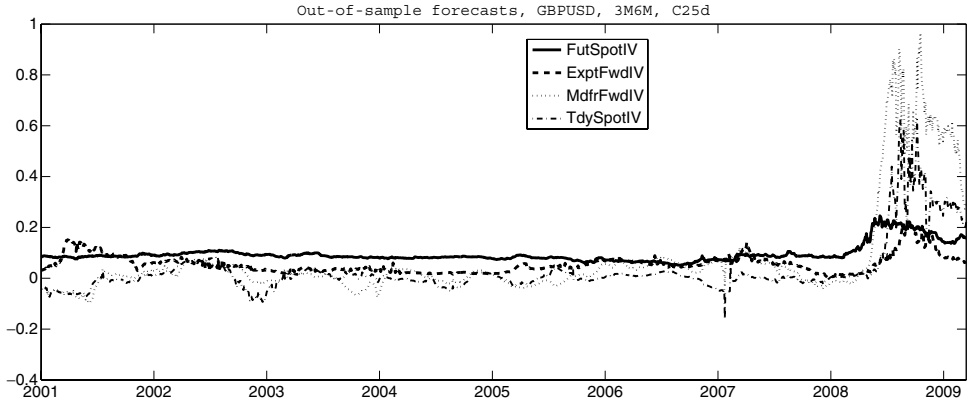

Fig. 2. The illustrative case is with respect to the sterling-dollar pair (GBPUSD) for 3 months-into-6 months (3M6M) forecasting horizon at 25 call delta (C25d). The picture shows time series plots of one-day-ahead one-year rolling out-of-sample forecasts using future spot implied volatility (FutSpotIV) as dependent variable. The explanatory variables are vector series of implied volatility curve at all five deltas from the model-based expected forward one, the model-free forward one and today's spot one.

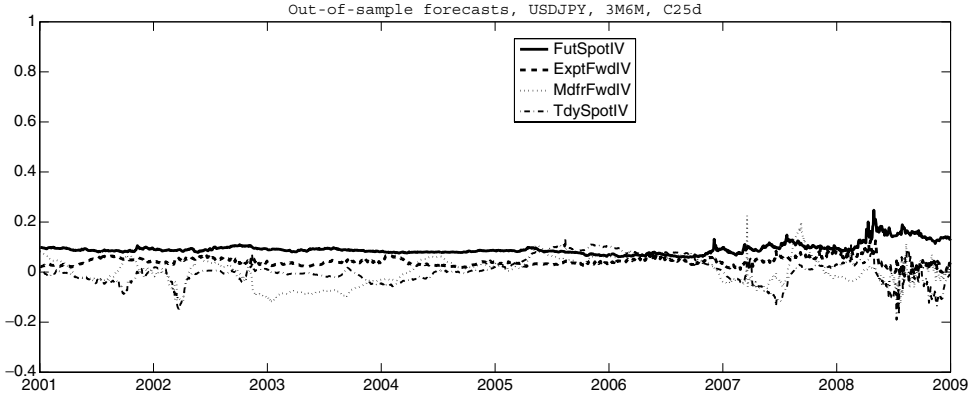

Fig. 3. The illustrative case is with respect to the dollar-yen pair (USDJPY) for 3 months-into-6 months (3M6M) forecasting horizon at 25 call delta (C25d). The picture shows time series plots of one-day-ahead one-year rolling out-of-sample forecasts using future spot implied volatility (FutSpotIV) as dependent variable. The explanatory variables are vector series of implied volatility curve at all five deltas from the model-based expected forward one, the model-free forward one and today's spot one.

As illustrative examples, Figs. 1–3 plot forecasts of 6-month implied volatility as forecast three months earlier against the actual spot 6-month implied volatility three months later. The figures show the EURUSD, GBPUSD, and USDJPY pairs, respectively, for 25 call delta options.

6. Conclusions

In this paper, we have addressed three aspects of forward implied volatility. First, we have formalized alternative definitions of forward implied volatility in the context of a stochastic volatility model, and distinguished three model-based notions that differ in the information they use from the term structure of current market prices of options. We then specialized to the SABR model and show how the asymptotic expansion of the bivariate transition density in Wu [26] allows calibration of the SABR model with piecewise constant parameters and calculation of forward implied volatility. A similar approach could be followed using the Heston model or any other tractable stochastic volatility model. We have focused on the SABR model because of its widespread use in fitting implied volatility smiles.

Finally, we have investigated empirically whether today's option prices contain predictive information regarding the future option prices and to what degree various notions of forward implied volatility extracts this information. Using currency option data, our first finding is that option prices across maturities do contain predictive information in forecasting future spot volatility. Our second finding is that model-based forward implied volatility measures extract this predicative information more effectively than a model-free forward measure and more effectively than today's spot implied volatility. The enhancement from using model-based forecasts is greater at longer horizons and greater for the dollar-yen pair than the euro-dollar and sterling-dollar pairs.

Acknowledgments

This work is supported by NSF grant DMS0914539. The authors would like to thank Credit Suisse for providing the data.

References

[1] K. I. Amin and R. A. Jarrow, Pricing foreign currency options under stochastic interest rates, *Journal of International Money and Finance* **10** (1991) 310–329.
[2] G. Bakshi, C. Cao and Z. Chen, Empirical performance of alternative option pricing models, *Journal of Finance* **52** (1997) 2003C2049.
[3] L. Bergomi, Smile Dynamics I, *Risk* April (2004).
[4] L. Bergomi, Smile Dynamics II, *Risk* March (2005).
[5] L. Bergomi, Smile Dynamics III, *Risk* March (2008).
[6] L. Bergomi, Smile Dynamics IV, *Risk* December (2009).
[7] H. Buehler, Consistent variance curve models, *Finance and Stochastics* **10** (2006) 178–203.
[8] R. H. Byrd, R. B. Schnabel and G. A. Shultz, Approximate solution of the trust region problem by minimization over two-dimensional subspaces, *Mathematical Programming* **40** (1988) 247–263.
[9] R. Carmona and S. Nadtochiy, Local volatility dynamic models, *Finance and Stochastics* **13** (2008) 1–48.
[10] P. Carr and L. Wu, Stochastic skew in currency options, *Journal of Financial Economics* **86** (2007) 213–247.

[11] R. Cont and J. da Fonseca, Dynamics of implied volatility surfaces, *Quantitative Finance* **12** (2002) 45–60.
[12] P. D. Corte, L. Sarno and I. Tsiakas, Spot and forward volatility in foreign exchange, FEA 2009 Bergen Meetings Paper (2009).
[13] B. Dupire, Pricing with A Smile, *Risk* January (1994).
[14] T. M. Egelkraut and P. Garcia, Intermediate volatility forecasts using implied forward volatility: The performance of selected agricultural commodity options, *Journal of Agricultural and Resource Economics* **31** (2006) 508–528.
[15] T. M. Egelkraut, P. Garcia and B. J. Sherrick, The term stucture of implied forward volatility: Recovery and informational content in the corn options market, *American Journal of Agricultural Economics*, February (2007) 1–11.
[16] H. W. Engl, M. H. and A. Neubauer, *Regularization of Inverse Problems* (Springer-Verlag, New York, 1996).
[17] M. Garman and S. Kohlhagen, Foreign currency option values, *Journal of International Money and Finance* **2** (1983) 231–237.
[18] J. Gatheral, Consistent modeling of SPX and VIX options, *5th World Congress of the Bachelier Finance Society*, London July (2008).
[19] J. Gatheral, *The Volatility Surface: A Practitioner's Guide* (John Wiley and Sons, New Jersey).
[20] P. S. Hagan, D. Kumar, A. S. Lesniewski and D. E. Woodward, Managing smile risk, *Wilmott Magazine*, September (2002) 84–108.
[21] P. S. Hagan, A. S. Lesniewski and D. E. Woodward, Probability distribution in the SABR model of stochastic volatility, preprint (2005).
[22] P. Henry-Labordere, A general asymptotic implied volatility for stochastic volatility models, working paper (2005).
[23] V. Piterbarg, A multi-currency model with FX volatility skew, working paper (2005).
[24] M. Schweizer and J. Wissel, Term structures of implied volatilities: Absence of arbitrage and existence results, *Mathematical Finance* **18** (2008) 77–114.
[25] P. Schönbucher, A market model for stochastic implied volatility, working paper, (1999).
[26] Q. Wu, Series expansion of SABR joint density, *Mathematical Finance*, forthcoming.