Special Issue in honor of John W. Neuberger
Electronic Journal of Differential Equations, Special Issue 02 (2023), pp. 161–173.
ISSN: 1072-6691. URL: https://ejde.math.txstate.edu or https://ejde.math.unt.edu

INVERSE VOLATILITY PROBLEM FOR CURRENCY OPTIONS

IAN KNOWLES, SUNDAR TAMANG

Abstract. In transactions associated with future-oriented financial instruments, such as options, a huge amount of data is available buried inside of which is the market's best guess as to what the future holds. We consider here the possibility of extracting future foreign exchange volatility information from foreign exchange option data with the aid of a new computational inverse algorithm using minimization of a convex functional.

1. Introduction

Financial entities such as stocks, bonds and even foreign exchange (FX) rates, are often modeled by the stochastic differential equation

$$\frac{dS_t}{S_t} = m(S_t, t)dt + \sigma(S_t, t)dB_t, \tag{1.1}$$

where, for each time t, $S_t(\omega)$ is a random variable representing the price of the financial entity for the sample path ω, m is the drift, which relates to the "trend" of the entity, σ is the volatility ("wobble"), and $B_t(\omega)$ is the Brownian motion stochastic process used to model this log-normal randomness.

Financial derivatives are contracts that derive their value from such an underlying entity. In particular, a *European call option* on a stock, is a derivative financial instrument that, when purchased, gives the holder the right but not the obligation to *buy* a designated number of stock shares at a pre-agreed price (the strike price) on a specified date, and a *European put option* provides the right to *sell* under the same circumstances.

In their seminal work [2] Black and Scholes showed that, for dividend-free stocks, the arbitrage-free price $v(S, t; K, T)$ of a European option contract satisfies the deterministic partial differential equation

$$\frac{\partial v}{\partial t} + \frac{1}{2}\sigma^2(S, t)S^2\frac{\partial^2 v}{\partial S^2} + \mu s\frac{\partial v}{\partial S} - rv = 0 \tag{1.2}$$

in time t and the value S of the underlying asset, where $\sigma(S, t)$ is the volatility, K is the strike price, T is the expiration time, μ is the risk-neutral drift, and r is the short-term risk-free interest rate.

2020 *Mathematics Subject Classification.* 35K10, 65N21.
Key words and phrases. Inverse volatility; European FX options; convex functional.

Published March 27, 2023.

Analogously, in the case of FX options, a European call option on a foreign exchange rate (currency) pair, is a derivative financial instrument that, when purchased, gives the holder the right but not the obligation to *buy* a designated amount of money, denominated in one currency, in another currency at a pre-agreed exchange rate on a specified date, and a European put option provides the right to sell under the same circumstances.

Assume that S is the price of the deliverable currency (domestic units per foreign unit), K is the exercise price of the option (domestic units per foreign unit), t is the current time, $u(S,t;K,T)$ is the current price of an FX call option (domestic units per foreign unit), r_D is the domestic (riskless) interest rate, r_F is the foreign (riskless) interest rate, and σ is volatility of the spot currency price. The corresponding Black Scholes partial differential equation

$$\frac{\partial u}{\partial t} + \frac{1}{2}\sigma^2(S,t)S^2\frac{\partial^2 u}{\partial S^2} + (r_D - r_F)S\frac{\partial u}{\partial S} - r_D u = 0 \tag{1.3}$$

for this case was provided by Garman and Kohlhagen in their classic 1983 paper [8], where for consistency we have replaced the time to maturity T in [8, equ. 6] with the current time t, and we have allowed the volatility to vary with S and t. We note also the final condition

$$u(S,T) = \begin{cases} S - K, \text{if } S > K \\ 0, \text{if } S \le K, \end{cases}$$

and the boundary condition $u(0,t) = 0$, $0 \le t \le T$ hold for this call option.

Now, financial data specifying the market price of an option is readily available in quantity at various strike values K around the current price of the underlying (the “spot” price), and for values of the maturity T up to around twelve months or so into the future. Given that the computer projections currently used by financial analysts are at best only valid for a rather limited time into the future, one is led quite naturally to the so-called *inverse volatility problem*: determine a market-inspired estimate of the future volatility function $\sigma(S,t)$ from a knowledge of current market prices of options with different strikes and future maturities.

In the case of stocks the solution of this problem generally goes as follows. The value v of an option contract depends crucially on the exercise (strike) price K, and the maturity date T of the contract. In 1994 Bruno Dupire [7] noticed that the function $v(S,t;K,T)$ satisfies the “dual” Black-Scholes equation

$$\frac{\partial v}{\partial T} - \frac{1}{2}K^2\sigma(K,T)\frac{\partial^2 v}{\partial K^2} + \mu K\frac{\partial v}{\partial K} - (\mu - r)v = 0, \tag{1.4}$$

known also as the Dupire equation. A number of approaches have been proposed subsequent to the appearance of [7]. Minimization methods using regularized least-squares fitting have been proposed in [1, 3, 16]; the possible presence of spurious local minima is always an issue here. An integral equation approach is presented in [4, 5], where convergence problems are possible given the underlying ill-posedness, and in [6, 10] linearization of the inherently non-linear inverse problem is discussed. In [12] the volatility is recovered by the minimization of a convex functional, thereby avoiding the spurious local minima problem that plagues generic least square methods.

For FX options, observing that (1.3) may be obtained from (1.2) by replacing μ with $r_D - r_F$ and r with r_D. On using the same replacements in (1.4) we then obtain the Dupire-type equation

$$\frac{\partial u}{\partial T} - \frac{1}{2}K^2\sigma(K,T)\frac{\partial^2 u}{\partial K^2} + (r_D - r_F)K\frac{\partial u}{\partial K} + r_F u = 0, \tag{1.5}$$

for the price $u(S,t;K,T)$ of a European FX call option.

If, at a certain time t_0 (when the spot price of the underlying is $S_0 = S(t_0)$), the prices $u(S_0,t_0,K,T)$ are known for all strikes K and maturities T then, as was noted first in [7], the volatility $\sigma(K,T)$ is uniquely determined in principle from the equation (1.5). But such a formula for σ is of little use in practice, as the market data for u is not only noisy (which would make the estimation of these derivatives highly ill-posed), but even worse, the data is typically both discrete and sparse in the variables T,K.

We present here a new variational algorithm for recovering, via the Dupire equation (1.5), the volatility $\sigma(K,T)$ from a knowledge of European FX option prices at various strikes and maturities, that is, from a knowledge of $u(S_0,t_0;K,T)$ for some time $t_0 \le T$. Here, as the variable K represents prices around the current spot price and $T \ge t_0$ represents future times, we have in effect recovered $\sigma(S,t)$ for future times t and FX prices S in a neighborhood of the spot price S_0.

The method used is an adaption, and improvement in the FX case, of the variational approach used in [12] involving the minimization of convex functionals as we discuss below. We begin by outlining the approach used in [12] to set the stage for our new method.

2. Reconstruction of FX volatility

Let $T_0 < T_1 < \dots < T_n$ be maturity times, and for each maturity T_i, $0 \le i \le n$, let

$$K_{i,1} < K_{i,2} < \dots < K_{i,m_i}$$

be the known associated strike prices. From the known price data $u(S_0,t_0;K_{i,j},T_i)$, $0 \le i \le n$ and $1 \le j \le m_i$ obtained at some time t_0, we compute the function $u(S_0,t_0;K,T)$ by linear interpolation, and for convenience, we make use of the notation $u(S_0,t_0;K,T) = u(K,T)$. Next, we assume that the volatility is piecewise constant in time, so that, for $1 \le i \le n$,

$$\sigma = \sigma_i(K)$$

over the i-th time sub-interval $[T_{i-1},T_i]$. Fixing i, set $\sigma_i(K) = \sigma(K)$ and

$$w_\lambda(K) = \int_{T_{i-1}}^{T_i} e^{-\lambda T} u(K,T)\,dT, \tag{2.1}$$

where $\lambda > 0$ is a parameter. For each such fixed i, $1 \le i \le n$, we now Laplace transform the Dupire equation (1.5) over $[T_{i-1},T_i]$ to obtain

$$\begin{aligned} 0 = \int_{T_{i-1}}^{T_i} e^{-\lambda T}u_T\,dT - \frac{1}{2}K^2\sigma^2 \underbrace{\int_{T_{i-1}}^{T_i} e^{-\lambda T}u_{KK}\,dT}_{w''_\lambda} \\ + (r_D - r_F)K \underbrace{\int_{T_{i-1}}^{T_i} e^{-\lambda T}u_K\,dT}_{w'_\lambda} + r_F w_\lambda, \end{aligned}$$

where the primes indicate differentiation with respect to K. On integrating the first term by parts we obtain

$$[e^{-\lambda T}u]_{T_{i-1}}^{T_i} + \lambda \underbrace{\int_{T_{i-1}}^{T_i} e^{-\lambda T} u\, dT}_{w_\lambda} - \frac{1}{2}K^2\sigma^2 w_\lambda'' + (r_D - r_F)Kw_\lambda' + r_F w_\lambda = 0,$$

and rearranging terms gives

$$-\frac{1}{2}K^2\sigma^2 w_\lambda'' + (r_D - r_F)Kw_\lambda' + (\lambda + r_F)w_\lambda = -u(K,T_i)e^{-\lambda T_i} + u(K,T_{i-1})e^{-\lambda T_{i-1}}.$$

Next, dividing by $\frac{1}{2}K^2\sigma^2$ throughout, we see that

$$-(w_\lambda'' - \frac{2(r_D - r_F)}{K\sigma^2}w_\lambda') + \frac{\lambda + r_F}{\frac{1}{2}K^2\sigma^2}w_\lambda = \frac{-u(K,T_i)e^{-\lambda T_i} + u(K,T_{i-1})e^{-\lambda T_{i-1}}}{\frac{1}{2}K^2\sigma^2}.$$

Finally, on multiplying by the integrating factor

$$p(K) = e^{-2(r_D - r_F)\int_a^K \frac{dk}{k\sigma^2(k)}}, \tag{2.2}$$

we now have an equation in Sturm-Liouville form:

$$-(p(K)w_\lambda')' + (\lambda + r_F)q(K)w_\lambda = \beta(K,\lambda)q(K), \tag{2.3}$$

where

$$q(K) = \Big(\frac{2}{K^2\sigma^2(K)}\Big)p(K), \tag{2.4}$$

$$\beta(K,\lambda) = -u(K,T_i)e^{-\lambda T_i} + u(K,T_{i-1})e^{-\lambda T_{i-1}}. \tag{2.5}$$

The idea in [12] is that if one can recover the functions $p(K)$ and $q(K)$ from (2.3), one can then find the volatility $\sigma(K)$ from the formula

$$\sigma(K) = \sqrt{\frac{2p(K)}{K^2 q(K)}}. \tag{2.6}$$

So, we first focus attention on a variational approach to the recovery of the pair of positive coefficient functions p, q defined on the generic interval $a \le K \le b$. It is assumed that we are given the functions $w_\lambda(K)$ for K in $[a,b]$ and all $\lambda > 0$. For positive functions r and s also defined on $[a,b]$, let $c = (r,s)$. Define $w_{c,\lambda}(K)$ to be the solution to the boundary value problem

$$\begin{gathered} L_{c,\lambda}w_{c,\lambda} = -(r(K)w_{c,\lambda}')' + (\lambda + r_F)s(K)w_{c,\lambda} = \beta(K,\lambda)q(K), \\ w_{c,\lambda}(a) = w_\lambda(a), \quad w_{c,\lambda}(b) = w_\lambda(b). \end{gathered} \tag{2.7}$$

Let $\mathcal{D}$ be the set of all positive function pairs $c = (r,s)$ such that the boundary value problem (2.7) is *disconjugate* on $[a,b]$, i.e. every non-trivial solution has at most one zero on $[a,b]$. It is known [9, Theorem 6.1, p. 351] that (2.7) is disconjugate if and only if the boundary value problem (2.7) can always be solved uniquely. It is also known (c.f. [15, Proposition 2.1]) that this set is open and convex in $\mathcal{L}[a,b] \times \mathcal{L}[a,b]$ and $\mathcal{L}^2[a,b] \times \mathcal{L}^2[a,b]$.

For each $\lambda > 0$ define the functional G_λ on the convex set $\mathcal{D}$ by

$$\begin{aligned} G_\lambda(c) = \int_a^b & r(K)(w_\lambda'^2 - w_{c,\lambda}'^2) + (\lambda + r_F)s(K)(w_\lambda^2 - w_{c,\lambda}^2) \\ & - 2\beta s(K)(w_\lambda - w_{c,\lambda})\, dK. \end{aligned} \tag{2.8}$$

2.1. **Properties of the functional G_λ.** The main properties of the functional G_λ are summarized in the following theorem, which is [12, Theorem 3.1].

Theorem 2.1. (a) *For any $c = (r, s)$ in $\mathcal{D}$,*

$$G_\lambda(c) = \int_a^b r(w'_\lambda - w'_{\lambda,c})^2 + (\lambda + r_F)s(w_\lambda - w_{c,\lambda})^2. \tag{2.9}$$

(b) *$G_\lambda(c) \geq 0$ for all $c = (r, s)$ in $\mathcal{D}$, and $G_\lambda(c)$=0 if and only if $w_\lambda = w_{c,\lambda}$.*

(c) *The first Gâteaux derivative of G_λ is given by*

$$G'_\lambda(r, s)[h_1, h_2] = \int_a^b \underbrace{(w'^2_\lambda - w'^2_{c,\lambda})}_{\text{L}^2\text{ gradient in r}} h_1 + \underbrace{[(\lambda + r_F)(w^2_\lambda - w^2_{c,\lambda}) - 2\beta(w_\lambda - w_{c,\lambda})]}_{\text{L}^2\text{ gradient in s}} h_2. \tag{2.10}$$

(d) *The second Gâteaux derivative of G_λ is given by*

$$G''_\lambda(c)[h, k] = 2(L^{-1}_{c,\lambda}(e(h)), e(k)), \tag{2.11}$$

where $h = (h_1, h_2)$, $k = (k_1, k_2)$,

$$e(h) = -(h_1 w'_{\lambda,c})' + (\lambda + r_F)h_2 w_{c,\lambda} - \beta h_2,$$

and $(\cdot\,,\cdot)$ denotes the usual inner product in $L^2[a, b]$.

As $L_{c,\lambda}$ is a positive operator on $W^1_0[a, b]$, we have from Theorem 2.1(d) that $G''_\lambda(c) \geq 0$ for all c in the convex set $\mathcal{D}$. By [18, Corollary 42.8] the functional G_λ is therefore convex on $\mathcal{D}$. We know from Theorem 2.1(b) that G_λ has a global minimum (zero) at $c = (r, s)$ if and only if $w_\lambda = w_{c,\lambda}$. Choose $N \geq 3$ positive distinct real numbers λ_j, $1 \leq j \leq N$, so that

$$0 < (\lambda_j + r_F)T < 2, \quad T \in [T_{i-1}, T_i].$$

In [12] the convex functional G was defined on the domain $\mathcal{D}$ (defined above) by

$$G(c) = \sum_{j=1}^{N} G_{\lambda_j}(c). \tag{2.12}$$

From the uniqueness theorem [11, Theorem 3.5] we know that, under certain (computer-verifiable) conditions on the nature of the flows of certain associated vector fields (which amount here to an admissibility restriction on the data $u(K, T)$), the condition that $w_\lambda = w_{c,\lambda}$ for at least three distinct values of λ implies that $c = (r, s) = (p, q)$. By [18, Proposition 42.6(1)] we know that if the convex functional G has a stationary point at (p, q) then it must have a global minimum there, and from the foregoing (assuming admissible data) that stationary point must uniquely occur at (p, q). So, the desired function pair (p, q) now appears as the unique global minimum of a convex functional with a unique stationary point. In practical numerics this is an important consideration, as many (if not most) least-square-type minimization methods suffer greatly from the minimization process getting stuck in spurious local minima. That this cannot happen here is one of the significant advantages of this approach.

As it happens, in the FX option case the relevant functional G takes on quite tiny starting values ($\approx 10^{-4}$ for the data in our example) and thus even though we have a well-posed minimization (so errors arising from ill-posedness are not an issue here [13]), this still constitutes a difficult inverse problem, because we are minimizing to zero and there is not much wiggle room.

Our solution in this paper is to pursue the volatility $\sigma(K)$ directly, and rather than first minimizing the two-variable functional $G(r,s)$ to obtain (p,q) followed by computing σ from (2.6), we now construct from G a new convex functional $H(\rho)$ with a unique global minimum at $\rho=\sigma$. The philosophy here is that a one variable minimization is expected to be better behaved numerically than a two variable minimization. In the sequel we show that this is indeed the case for the function $H(\rho)$ defined below.

3. The functional $H(\rho)$

For a generic volatility $\rho(K)$ define

$$p_\rho(K) = e^{-2(r_D-r_F)\int_a^K \frac{d\tau}{\tau\rho^2(\tau)}}, \tag{3.1}$$

$$q_\rho(K) = \frac{2p_\rho(K)}{K^2\rho^2(K)}\,. \tag{3.2}$$

Note that from (2.2) and (2.4) we have $p=p_\sigma$ and $q=q_\sigma$.

For $\lambda>0$ we define the non-negative functional $H_\lambda(\rho)$ by $c=(p_\rho,q_\rho)$ and

$$H_\lambda(\rho) = G_\lambda(c) = \int_a^b p_\rho.(w'_\lambda - w'_{\lambda,c})^2 + (\lambda+r_F)q_\rho.(w_\lambda - w_{c,\lambda})^2. \tag{3.3}$$

Theorem 3.1. (a) *$H_\lambda(\rho)\ge 0$ for all ρ, and $H_\lambda(\rho)=0$ if and only if*

$$w_\lambda = w_{c,\lambda},$$

where $c=(p_\rho,q_\rho)$.

(b) *The first Gâteaux derivative of H_λ is given by*

$$H'_\lambda(\rho)[h] = \int_a^b (w'^2_\lambda - w'^2_{c,\lambda})h_1 + [(\lambda+r_F)(w^2_\lambda - w^2_{c,\lambda}) - 2\beta(w_\lambda - w_{c,\lambda})]h_2, \tag{3.4}$$

where

$$h_1(K) = 4(r_D-r_F)e^{-2(r_D-r_F)\int_a^K \frac{dk}{k\rho^2(k)}}\int_a^K \frac{h(\tau)d\tau}{\tau\rho^3(\tau)},$$

and

$$h_2 = -\frac{4h}{K^2\rho^3}e^{-2(r_D-r_F)\int_a^K \frac{d\tau}{\tau\rho^2(\tau)}} + \frac{8}{K^2\rho^2}e^{-2(r_D-r_F)\int_a^K \frac{d\tau}{\tau\rho^2(\tau)}}(r_D-r_F)\int_a^K \frac{h}{\tau\rho^3}\,d\tau.$$

(c) *The second Gâteaux derivative of H_λ is given by*

$$H''_\lambda(\rho)[h,k] = 2(L^{-1}_{c,\lambda}(e(h_1,h_2)), e(k_1,k_2)), \tag{3.5}$$

where

$$e(h_1,h_2) = -(h_1 w'_{\lambda,c})' + h_2[(\lambda+r_F)w_{c,\lambda} - \beta],$$

and $(\cdot,\cdot)$ denotes the usual inner product in $L^2[a,b]$ and h_1 and h_2 are calculated above from h and ρ.

Proof. (a) This follows from Theorem 2.1(a).

(b) We have

$$\begin{aligned} H'_\lambda(\rho)[h] &= \lim_{\epsilon\to 0} \frac{H_\lambda(\rho+\epsilon h) - H_\lambda(\rho)}{\epsilon} \\ &= \lim_{\epsilon\to 0} \frac{G_\lambda(p_{\rho+\epsilon h}, q_{\rho+\epsilon h}) - G_\lambda(p_\rho, q_\rho)}{\epsilon} \\ &= \lim_{\epsilon\to 0} \frac{G_\lambda(p_\rho+\epsilon h_1^*, q_\rho+\epsilon h_2^*) - G_\lambda(p_\rho, q_\rho)}{\epsilon} \\ &= G'_\lambda(p_\rho, q_\rho)[h_1, h_2]. \end{aligned}$$

where we define h_1^* and h_2^* by $p_{\rho+\epsilon h} = p_\rho + \epsilon h_1^*$ and $q_{\rho+\epsilon h} = q_\rho + \epsilon h_2^*$ and note that

$$\begin{aligned} h_1 &= \lim_{\epsilon\to 0} h_1^* \\ &= \lim_{\epsilon\to 0} \frac{1}{\epsilon}[p_{\rho+\epsilon h} - p_\rho] \\ &= \lim_{\epsilon\to 0} \frac{1}{\epsilon}\Big[e^{-2(r_D-r_F)\int_a^K \frac{d\tau}{\tau(\rho(\tau)+\epsilon h(\tau)^2}} - e^{-2(r_D-r_F)\int_a^K \frac{d\tau}{\tau\rho^2(\tau)}}\Big] \\ &= \lim_{\epsilon\to 0} \frac{1}{\epsilon} e^{-2(r_D-r_F)\int_a^K \frac{d\tau}{\tau\rho^2(\tau)}}\Big\{e^{-2(r_D-r_F)\int_a^K \frac{1}{\tau}[\frac{1}{(\rho+\epsilon h)^2}-\frac{1}{\rho^2}]\,d\tau} - 1\Big\} \\ &= \lim_{\epsilon\to 0} \frac{1}{\epsilon} e^{-2(r_D-r_F)\int_a^K \frac{d\tau}{\tau\rho^2(\tau)}}\Big\{e^{2(r_D-r_F)\epsilon\int_a^K \frac{1}{\tau}\frac{2\rho h+\epsilon h^2}{\rho^2(\rho+\epsilon h)^2}\,d\tau} - 1\Big\} \\ &= \lim_{\epsilon\to 0}\ e^{-2(r_D-r_F)\int_a^K \frac{d\tau}{\tau\rho^2(\tau)}} 2(r_D-r_F)\int_a^K \frac{1}{\tau}\frac{2\rho h+\epsilon h^2}{\rho^2(\rho+\epsilon h)^2}\,d\tau \\ &\quad \times\Big\{\frac{e^{2(r_D-r_F)\epsilon\int_a^K \frac{1}{\tau}\frac{2\rho h+\epsilon h^2}{\rho^2(\rho+\epsilon)^2}\,d\tau} - 1}{2(r_D-r_F)\epsilon\int_a^K \frac{1}{\tau}\frac{2\rho h+\epsilon h^2}{\rho^2(\rho+\epsilon h)^2}\,d\tau}\Big\} \\ &= 4e^{-2(r_D-r_F)\int_a^K \frac{d\tau}{\tau\rho^2(\tau)}}(r_D-r_F)\int_a^K \frac{h}{\tau\rho^3}\,d\tau, \end{aligned}$$

using that

$$\lim_{z\to 0}\frac{e^z-1}{z} = 1;$$

and in similar fashion

$$\begin{aligned} h_2 &= \lim_{\epsilon\to 0} h_2^* \\ &= \lim_{\epsilon\to 0} \frac{1}{\epsilon}[q_{\rho+\epsilon h} - q_\rho] \\ &= \lim_{\epsilon\to 0} \frac{2}{\epsilon K^2}\Big[p_{\rho+\epsilon h}\Big\{\frac{1}{(\rho+\epsilon h)^2} - \frac{1}{\rho^2}\Big\} + \frac{1}{\rho^2}\{p_{\rho+\epsilon h} - p_\rho\}\Big] \\ &= -\frac{4h}{K^2\rho^3}p_\rho + \frac{2}{K^2\rho^2}\lim_{\epsilon\to 0}\frac{1}{\epsilon}[p_{\rho+\epsilon} - p_\rho] \\ &= -\frac{4h}{K^2\rho^3}e^{-2(r_D-r_F)\int_a^K \frac{d\tau}{\tau\rho^2(\tau)}} \\ &\quad + \frac{8}{K^2\rho^2}e^{-2(r_D-r_F)\int_a^K \frac{d\tau}{\tau\rho^2(\tau)}}(r_D-r_F)\int_a^K \frac{h}{\tau\rho^3}\,d\tau. \end{aligned}$$

(c) This follows from Theorem 2.1(d). □

We next calculate the $L^2[a,b]$-gradient of the functional H_λ.

Theorem 3.2. *Further to* (3.4) *we have*

$$H'_\lambda(\rho)[h] = \int_a^b \Big[\frac{1}{K\rho^3(K)}\int_K^b m(\tau)\,d\tau - n(K)\Big]h(K)\,dK, \tag{3.6}$$

where

$$\begin{aligned} m(K) &= 4(r_D - r_F)e^{-2(r_D-r_F)\int_a^K \frac{d\tau}{\tau\rho^2(\tau)}} \\ &\quad\times \Big[[(\lambda + r_F)(w_\lambda^2 - w_{c,\lambda}^2) - 2\beta(w_\lambda - w_{c,\lambda})]\frac{2}{K^2\rho^2} + (w_\lambda'^2 - w_{c,\lambda}'^2)\Big], \end{aligned} \tag{3.7}$$

and

$$n(K) = [(\lambda + r_F)(w_\lambda^2 - w_{c,\lambda}^2) - 2\beta(w_\lambda - w_{c,\lambda})]\frac{4}{K^2\rho^3}e^{-2(r_D-r_F)\int_a^K \frac{d\tau}{\tau\rho^2(\tau)}}, \tag{3.8}$$

and finally,

$$g(K) = \frac{1}{K\rho^3(K)}\int_K^b m(\tau)\,d\tau - n(K)$$

is the $L^2[a,b]$*-gradient of* $H_\lambda(\rho)$*.*

Proof. From Theorem 3.1(b) we have

$$\begin{aligned} &H'_\lambda(\rho)[h] \\ &= \int_a^b \Big\{(w_\lambda'^2 - w_{c,\lambda}'^2)4(r_D - r_F)e^{-2(r_D-r_F)\int_a^K \frac{d\tau}{\tau\rho^2(\tau)}}\int_a^K \frac{h(\tau)d\tau}{\tau\rho^3(\tau)} \\ &\quad - h(K)[(\lambda + r_F)(w_\lambda^2 - w_{c,\lambda}^2) - 2\beta(w_\lambda - w_{c,\lambda})]\frac{4}{K^2\rho^3}e^{-2(r_D-r_F)\int_a^K \frac{d\tau}{\tau\rho^2(\tau)}} \\ &\quad + [(\lambda + r_F)(w_\lambda^2 - w_{c,\lambda}^2) - 2\beta(w_\lambda - w_{c,\lambda})]\frac{8(r_D - r_F)}{K^2\rho^2}e^{-2(r_D-r_F)\int_a^K \frac{d\tau}{\tau\rho^2(\tau)}} \\ &\quad \times \int_a^K \frac{h}{\tau\rho^3}\,d\tau, \Big\}\,dK \\ &= \int_a^b \Big\{4(r_D - r_F)e^{-2(r_D-r_F)\int_a^K \frac{dk}{k\rho^2(k)}}\int_a^K \frac{h(\tau)d\tau}{\tau\rho^3(\tau)} \\ &\quad \times \Big[[(\lambda + r_F)(w_\lambda^2 - w_{c,\lambda}^2) - 2\beta(w_\lambda - w_{c,\lambda})]\frac{2}{K^2\rho^2} + (w_\lambda'^2 - w_{c,\lambda}'^2)\Big] \\ &\quad - h[(\lambda + r_F)(w_\lambda^2 - w_{c,\lambda}^2) - 2\beta(w_\lambda - w_{c,\lambda})]\frac{4}{K^2\rho^3}e^{-2(r_D-r_F)\int_a^K \frac{d\tau}{\tau\rho^2(\tau)}}\Big\}\,dK. \end{aligned}$$

Collecting the terms containing $\int_a^K \frac{h(\tau)d\tau}{\tau\rho^3(\tau)}$, the above expression becomes

$$\int_a^b \Big\{m(K)\int_a^K \frac{h(\tau)d\tau}{\tau\rho^3(\tau)} - h(K)n(K)\Big\}dK, \tag{3.9}$$

where

$$\begin{aligned} m(K) &= 4(r_D - r_F)e^{-2(r_D-r_F)\int_a^K \frac{dk}{k\rho^2(k)}} \\ &\quad\times \Big[[(\lambda + r_F)(w_\lambda^2 - w_{c,\lambda}^2) - 2\beta(w_\lambda - w_{c,\lambda})]\frac{2}{K^2\rho^2} + (w_\lambda'^2 - w_{c,\lambda}'^2)\Big], \end{aligned} \tag{3.10}$$

$$n(K) = [(\lambda + r_F)(w_\lambda^2 - w_{c,\lambda}^2) - 2\beta(w_\lambda - w_{c,\lambda})]\frac{4}{K^2\rho^3}e^{-2(r_D-r_F)\int_a^K \frac{d\tau}{\tau\rho^2(\tau)}}. \tag{3.11}$$

We now integrate the first term in the integrand in (3.9) by parts to isolate the h factor. Specifically,

$$\int_a^b m(K)\int_a^K \frac{h(\tau)d\tau}{\tau\rho^3(\tau)}\,dK = \int_a^b h(K)\frac{1}{K\rho^3(K)}\Big[\int_K^b m(\tau)\,d\tau\Big]dK,$$

and hence from (3.9)

$$H'_\lambda(\rho)[h] = \int_a^b h(K)\Big\{\frac{1}{K\rho^3(K)}\int_K^b m(\tau)\,d\tau - n(K)\Big\}dK, \tag{3.12}$$

which completes the proof. □

Finally, analogously to (2.12), we set

$$\mathcal{D}_H = \{\rho : (p_\rho, q_\rho) \in \mathcal{D}\},$$

and define the convex functional $H(\rho)$ on the domain $\mathcal{D}_H$ by

$$H(\rho) = \sum_{j=1}^N H_{\lambda_j}(\rho). \tag{3.13}$$

4. A recovery algorithm

$H(\rho)$ is a nonnegative convex functional since it is the sum of nonnegative convex functionals, and it also has a unique stationary point at σ. The idea here is that by using H rather than just one of the H_λ, in addition to gaining favourable uniqueness properties, we are blending additional time-based data into the inverse problem, and this is intended to improve the well-posedness of the problem, given that ill-posedness is basically due to a paucity of information. We note in passing from [13] that this inverse recovery is conditionally well-posed in the weak-L^2 sense, so the recoveries are expected to be, and indeed were, quite stable.

We minimize this functional for $N = 20$ using the steepest descent method to recover the coefficient $\sigma(T,K)$. The L^2-direction of steepest ascent for H at ρ is

$$\nabla_{L^2}H(\rho) = \sum_{i=1}^N \Big\{\frac{1}{K\rho^3(K)}\int_K^b m_i(\tau)\,d\tau - n_i(K)\Big\},$$

where, for $1 \le i \le N$,

$$\begin{aligned} m_i(K) &= 4(r_D - r_F)e^{-2(r_D-r_F)\int_a^K \frac{d\tau}{\tau\rho^2(\tau)}}\Big[\big[(\lambda_i + r_F)(w_{\lambda_i}^2 - w_{c,\lambda_i}^2) \\ &\quad - 2\beta(w_{\lambda_i} - w_{c,\lambda_i})\big]\frac{2}{K^2\rho(K)^2} + (w'^2_{\lambda_i} - w'^2_{c,\lambda_i})\Big], \end{aligned} \tag{4.1}$$

$$\begin{aligned} n_i(K) &= [(\lambda_i + r_F)(w_{\lambda_i}^2 - w_{c,\lambda_i}^2) - 2\beta(w_{\lambda_i} - w_{c,\lambda_i})] \\ &\quad \times \frac{4}{K^2\rho(K)^3}e^{-2(r_D-r_F)\int_a^K \frac{d\tau}{\tau\rho(K)^2(\tau)}}, \end{aligned} \tag{4.2}$$

Instead of using this L^2-gradient directly, it is numerically preferable to use the corresponding Neuberger gradient (see [17]) given that the L^2-gradient has numerical problems that are extensively discussed in [15]. In particular, the L^2-gradient is zero on the boundary of [a,b] given that w_λ and $w_{c,\lambda}$ are equal there, and thus the algorithm is unable to properly recover σ. The Neuberger gradient smooths the L^2-gradient and preserves boundary data during the descent, an important property not shared by other descent techniques. Our Neuberger gradient $g =$

$\nabla_{H^1} H$ can be found from an L^2-gradient $\nabla_{L^2} H$ by solving the Dirichlet-Neumann boundary value problem

$$\begin{aligned} -g'' + g &= \nabla_{L^2} H, \\ g(a) &= g'(b) = 0, \end{aligned} \tag{4.3}$$

where $[a, b]$ represents the generic K-interval. The choice of the Neumann condition at b allows for some flexibility in the shape of the iterates. Forcing $g(b) = 0$ appears to be too confining and inhibits the descent.

We have option prices for discrete sets of strikes and maturities. We generated the function $v(K, T)$ by linearly interpolating the option price in both strike and maturity. The function $v(K, T)$ was mollified (c.f. [14, §6]) so that it could be differentiated, and the derivative $v_K(K, T)$ was found using central differences. For 20 fixed values of λ the functions $v(K, T)$ and $v'_K(K, T)$ were Laplace transformed using (2.1) to $w_\lambda(K)$ and $w'_\lambda(K)$ respectively. The functions $p_{\sigma_0}(K)$ and $q_{\sigma_0}(K)$ were initialized using (2.2) and (2.4) with the initial σ chosen to be the implied volatility. We performed a series of descents in σ using the aforementioned Neuberger steepest descent algorithm such that the functional could not be minimized any further. The line minimization of $f(\alpha) = H(\sigma_0 - \alpha \nabla_{H^1} H(\sigma_0))$ in α was done using the well known Brent minimization technique, by adapting the one-variable code in the Numerical Recipes in C function `brent()`. To avoid possible catastrophic cancellation in the Simpson rule formula used in the calculation of the integrals in the formula (2.8) for the functional H_λ, we used the alternate formula (3.3) instead.

Below is the steepest descent algorithm used to get one descent step in σ:

(i) Initialize $\sigma = \sigma_0$, for example using a known implied volatility.
(ii) Set $c_0 = (p_{\sigma_0}, q_{\sigma_0})$.
(iii) For all i, find w_{λ_i, c_0} and w'_{λ_i, c_0} by solving (2.7) with $\lambda = \lambda_i$.
(iv) Find the L^2 gradient of H at σ_0, $\nabla_{L^2} H(\sigma_0)$.
(v) Find the Neuberger gradient, $\nabla_{H^1} H(\sigma_0)$.
(vi) Find $\alpha = \alpha_{\min}$ that gives the lowest value of $f(\alpha) = H(\sigma_0 - \alpha \nabla_{H^1} H(\sigma_0))$.
(vii) Compute $\sigma_{\text{new}} = \sigma_0 - \alpha_{\min} \nabla_{H^1} H(\sigma_0)$.

One may then iterate by overwriting σ_0 with σ_{new} and repeating steps (ii) through (vii) until H fails to descend using the descent direction $-\nabla_{H^1} H(\sigma_{\text{new}})$.

5. Results

The most widely traded of the FX options is the EURUSD option. To recover the coefficient function $\sigma(T, K)$ we wrote a C program, which used, as the initial guess in the descent process, the "implied volatility" obtained directly from the standard option price formula of Garman and Kohlhagen [8] by substituting the known option price and solving for the implied volatility σ as an unknown. The implied and recovered volatility surfaces and the convergence graph are shown in Figures 1–3

6. Conclusions

We have presented herein a new inverse algorithm for the computation of foreign exchange volatility from foreign exchange option data. The major difficulty that we encountered in this particular inverse problem was that the typical values for our previously effective ([12]) minimizing functional $G(r, s)$ (see (2.12)) were quite tiny (of the order 10^{-4}) and as we were minimizing to zero the two-variable functional

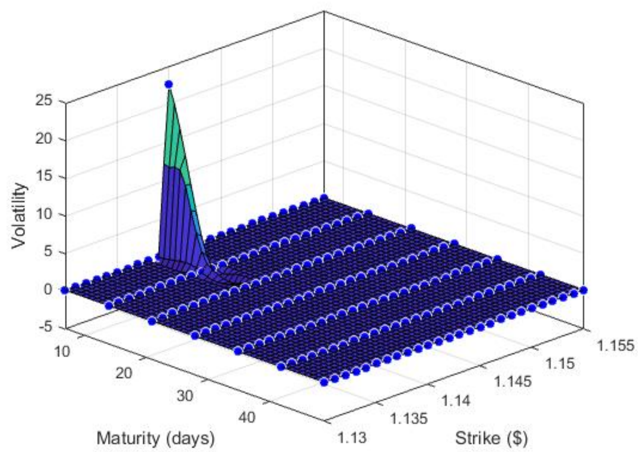

Figure 1. Implied volatility surface for EURUSD options

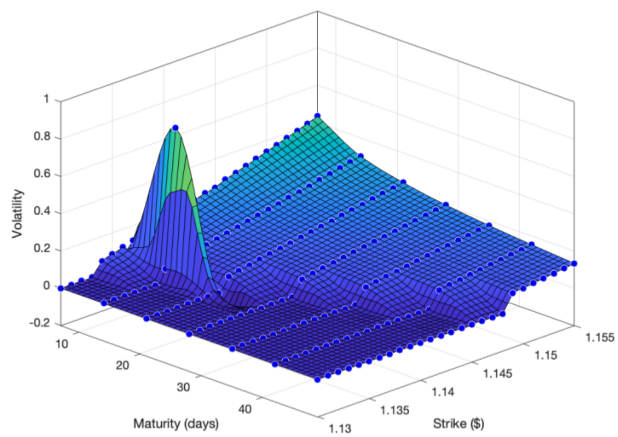

Figure 2. Recovered EURUSD option volatility surface $\sigma(K,T)$

G was too insensitive to handle this task. So we reformulated to a new one-variable functional $H(\rho)$ defined by (3.13) and this proved quite effective. It seems likely that this new approach would also improve the volatility recovery not only in [12] but elsewhere, such as in groundwater flow-parameter recovery, where the smallness of the functional values is not an issue.

References

[1] M. Avellaneda, C. Friedman, L. Holmes, L. Sampieri; Calibrating volatility surfaces via relative entropy minimization. *Appl. Math. Finance*, 4:37–64, 1997.

[2] Fischer Black, Myron Scholes; The pricing of options and corporate liabilities. *Journal of Political Economy*, 81:637–654, 1973.

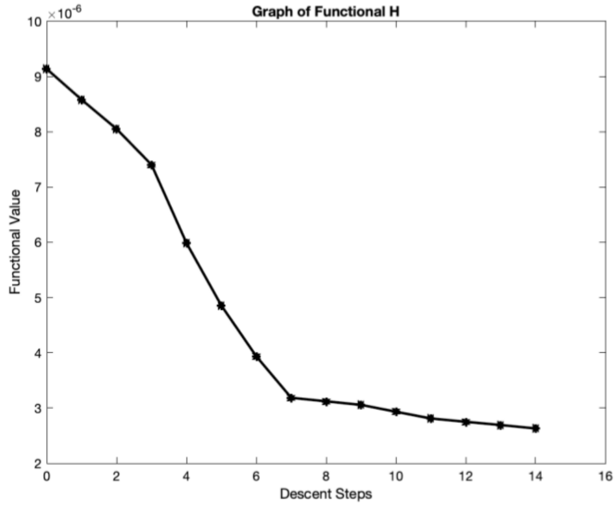

FIGURE 3. Minimization of the functional H

[3] J. N. Bodurtha, M. Jermakyan; Non-parameric estimation of an implied volatility surface. *J. Computational Finance*, 2:29–61, 1999.
[4] Ilia Bouchouev, Victor Isakov; The inverse problem of option pricing. *Inverse Problems*, 13:L11–L17, 1999.
[5] Ilia Bouchouev, Victor Isakov; Uniqueness, stability and numerical methods for the inverse problem that arises in financial markets. *Inverse Problems*, 15:R95–R116, 1999.
[6] Ilia Bouchouev, Victor Isakov; Recovery of volatility coefficient by linearization. *Quant. Finance*, 2:257–263, 2002.
[7] B. Dupire; Pricing with a smile. *RISK*, 7:18–20, 1994.
[8] Mark B. Garman, Steven W. Kohlhagen; Foreign currency option values. *Journal of International Money and Finance*, 2(3):231 – 237, 1983.
[9] Philip Hartman; *Ordinary differential equations.* S. M. Hartman, Baltimore, Md., 1973. Corrected reprint.
[10] Victor Isakov; The inverse problem of option pricing. Preprint, 2004.
[11] Ian Knowles; Uniqueness for an elliptic inverse problem. *SIAM J. Appl. Math.*, 59(4):1356–1370, 1999.
[12] Ian Knowles, Li Feng, Ajay Mahato; The inverse volatility problem for European options. In *Proceedings of the Variational and Topological Methods: Theory, Applications, Numerical Simulations, and Open Problems*, volume 21 of *Electron. J. Differ. Equ. Conf.*, pages 183–195. Texas State Univ., San Marcos, TX, 2014.
[13] Ian Knowles, Mary A. LaRussa; Conditional well-posedness for an elliptic inverse problem. *SIAM J. Appl. Math.*, 71:952–971, 2011.
[14] Ian Knowles, Tuan A. Le, Aimin Yan; On the recovery of multiple flow parameters from transient head data. *J. Comp. Appl. Math.*, 169:1–15, 2004.
[15] Ian Knowles, Robert Wallace; A variational method for numerical differentiation. *Numerische Mathematik*, 70:91–110, 1995.
[16] R. Lagnado, S. Osher; A technique for calibrating derivation of the security pricing models: numerical solution of the inverse problem. *J. Computational Finance*, 1:13–25, 1997.
[17] J. W. Neuberger; *Sobolev gradients and differential equations*, volume 1670 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, second edition, 2010.
[18] Eberhard Zeidler; *Nonlinear functional analysis and its applications. III.* Springer-Verlag, New York, 1985. Variational methods and optimization, Translated from the German by Leo F. Boron.

Ian Knowles
Department of Mathematics, University of Alabama at Birmingham, Birmingham AL 35294, USA
Email address: iknowles@uab.edu

Sundar Tamang
Department of Math and Computer Science, Western New Mexico University, Silver City, NM 88062, USA
Email address: sundar.tamang@wnmu.edu