Forecasting Multifractal Volatility*

Laurent Calvet
Department of Economics, Harvard University

Adlai Fisher
Stern School of Business, New York University

This Draft: November 10, 1999
First Draft: February 1999

*Littauer Center, Cambridge, MA 02138-3001, lcalvet@kuznets.harvard.edu; Kaufman Management Center, 44 West Fourth Street, Suite 9-190, New York, NY 10012-1126, afisher@stern.nyu.edu. We would like to thank David Backus, Rohit Deo, Francis Diebold, Robert Engle, Clifford Hurvich, Kai Lee, Benoit Mandelbrot, Joshua Rosenberg, Stanley Zin, and seminar participants at Carnegie Mellon University, New York University, and the NBER Summer Institute for helpful comments. Any remaining errors are our own.

Abstract

This paper develops analytical methods to forecast the distribution of future returns for a new continuous-time process, the Poisson multifractal. Our model captures the thick tails and volatility persistence exhibited by many financial time series. We assume that the forecaster knows the true generating process with certainty, but only observes past returns. The challenge in this environment is long memory and the corresponding infinite dimension of the state space. We show that a discretized version of the model has a finite state space, which allows an analytical solution to the conditioning problem. Further, the discrete model converges to the continuous-time model as time scale goes to zero, so that forecasts are consistent. The methodology is implemented on simulated data calibrated to the Deutschemark / US Dollar exchange rate. Applying these results to option pricing, we find that the model captures both volatility smiles and long-memory in the term structure of implied volatilities.

Keywords: Forecasting, Implied Volatility, Long Memory, Multifractal Model of Asset Returns, Option Pricing, Poisson Multifractal, Trading Time, Volatility Smile

1. Introduction

This paper develops analytical methods to forecast the distribution of future returns in a new class of continuous-time processes, Poisson multifractals. Our model parsimoneously captures the thick tails and volatility persistence exhibited by many financial time series. It provides a fully-stationary version of the multifractal model introduced in Mandelbrot, Fisher, and Calvet (1997), which contains residual effects of a grid-based construction. We model volatility as the multiplicative product of an infinite sequence of random functions. Each random function has innovations with Poisson arrivals of a different frequency, where the full set of time horizons progresses from a suitably long time scale to approach zero by a power law in time. This ensures that volatility clustering will be present at all time scales, and corresponds to the economic intuition that economic factors such as technological shocks, business cycles, earnings cycles, and liquidity shocks all have different time scales.

To provide parsimony, we assume that each volatility component has identically distributed innovations, so that they differ only by their respective timescales. We propose that such patterns may arise in fully rational equilibrium models, either because of exogenous shocks with scaling properties, or endogenously due to market incompleteness or informational cascades. The multiplicative interaction of the volatility components implies that the effect of an innovation is proportional to volatility at other time scales, as could be the case in models where trade depends on the availability of willing counterparties. Additionally, because the interactions are non-linear, past data is informative regarding the decomposition of volatility among its various components. Thus, with full knowledge of the current volatility level, forecasts may differ considerably depending upon the past history.

We first introduce the continuous-time version of the model, which has an infinite state space. The model compounds a standard Brownian Motion with a random time-deformation process that is obtained from the volatility model described above. This construction implies semi-martingale prices and uncorrelated returns, and thus precludes arbitrage in a standard two-asset economy. Squared returns have long-memory, and the highest finite moment of returns is permitted to have any value greater than two. Additionally, this flexible range of tail behavior is fully provided by intermittent bursts of volatility modifying the standard Brownian Motion, with no need to incorporate jumps or otherwise separately model the conditional distribution of returns. The unconditional distribution of returns changes with the time scale, and sample histograms may appear more thin-tailed at longer horizons. However, the distribution of returns does not generally converge to a Gaussian with increasing horizon, and never converges to a Gaussian with decreasing horizon.

To facilitate forecasting, we then introduce a discretized version of the process with finite state space and a simple Markovian structure. A recursive algorithm allows us to calculate the current probabilities of the volatility state conditional on past data and initial beliefs. A second recursive algorithm computes analytical multi-step forecasts of the distribution of the process at future dates, conditional on beliefs about the current state. We then show that refinements of the discretized process to progressively smaller time scales converge to the continuous time process, so that forecasts from the discrete model are consistent as the time scale goes to zero.

The forecasting methodology is numerically implemented for a multifractal process calibrated to Deustchemark/US Dollar exchange rate returns. We obtain a reduced form option pricing formula that is consistent with absence of arbitrage, and use this formula to calculate implied volatilities for simulated data. The model captures both volatility smile and long memory in the term structure of implied volatilities.

Our methodology has several features that distinguish it from forecasting in other long-memory models. For the FIGARCH process, the set of state variables relevant to forecasting future return densities consists of all past price changes, and all of these state variables are observed. The conditioning problem is thus straightforward, and reduces to a choice of how much past data to use. Multi-step density forecasts, however, are computationally problematic because of the large state space, and are usually formed by intensive simulation. For standard Long Memory Stochastic Volatility Models (LMSV), the problem is further complicated since the relevant state variables are past volatility levels, which are unobserved, i.e., latent. The forecaster thus must first solve a difficult conditioning problem, and then ideally compute forecasts conditional upon each possible set of past volatilities. The multifractal model simplifies forecasting because it greatly reduces the volatility state space. For FIGARCH and LMSV models, incorporating frequencies of size as low as $1/n$ typically requires n state variables. In the multifractal model, $\log_b n$ state variables capture frequencies of the same range, where b is a constant of the model.

Section 2 provides a brief review of the multifractal model. Section 3 introduces a new, grid-free multifractal measure in continuous time. Section 4 defines the new financial model and presents its economic intuition. Section 5 provides a discretized version of the process, presents the forecasting algorithm, and demonstrates the consistency of forecasts obtained from the discretized model. Section 6 discusses the numerical implementation of this method. Section 7 discusses option pricing and the calculation of implied volatilities.

2. A Review of the Multifractal Model

This section presents a brief review the Multifractal Model of Asset Returns (MMAR), which was introduced in Mandelbrot, Fisher, and Calvet (1997). The MMAR is constructed by compounding a Brownian Motion $B(t)$ with a stochastic trading time $\theta(t)$:

$$\ln P(t) - \ln P(0) = B[\theta(t)],$$

where $\theta(t)$ is a random increasing function or, equivalently, the cumulative distribution function (c.d.f.) of a random measure μ. The model captures the outliers and volatility persistence of many financial time series by specifying the measure μ to be multifractal, a concept which we now recall.

2.1. The Binomial Measure

Multifractal measures are built by iterating an elementary procedure, called a *multiplicative cascade*. The simplest multifractal is the binomial measure on the compact interval $[0,1]$. Consider the uniform probability measure μ_0 on $[0,1]$, and two positive numbers m_0 and m_1 adding up to 1. In the first step of the cascade, we define a new measure μ_1 by uniformly spreading the mass m_0 on the *left* subinterval $[0,1/2]$, and the mass m_1 on the *right* subinterval $[1/2,1]$. The density of μ_1 is now a step function.

In the second stage of the cascade, we split the interval $[0,1/2]$ into two subintervals of equal length, $[0,1/4]$ and $[1/4,1/2]$. The left subinterval $[0,1/4]$ receives a fraction m_0 of the mass $\mu_1[0,1/2]$, while the right subinterval receives a fraction m_1. Applying this again to the interval $[1/2,1]$, we obtain a new measure μ_2, which satisfies

$$\begin{aligned} \mu_2[0,1/4] &= m_0 m_0, & \mu_2[1/4,1/2] &= m_0 m_1, \\ \mu_2[1/2,3/4] &= m_1 m_0, & \mu_2[3/4,1] &= m_1 m_1. \end{aligned}$$

Iteration of this procedure generates an infinite sequence of measures (μ_k), which weakly converges to the binomial measure μ. Like many multifractals, the binomial is a continuous but singular probability measure; it thus has no density and no point mass. We also observe that since $m_0+m_1=1$, each stage of the construction preserves the mass of split dyadic intervals.

This construction can be extended in several ways. For instance at each stage of the cascade, intervals can be split into $b>2$ intervals of equal size. Subintervals, indexed from left to right by β $(0 \leq \beta \leq b-1)$, receive fractions of the total mass equal to $m_0, .., m_{b-1}$, and we preserve mass by imposing that these fractions, also called *multipliers*, add up to one: $\sum m_\beta = 1$. This defines the class of *multinomial* measures. Another extension randomizes the allocation of mass between subintervals at each step of the iteration. The multiplier of each subinterval is a discrete random variable M_β that takes values $m_0, m_1, ..., m_{b-1}$ with probabilities $p_0, .., p_{b-1}$. The preservation of mass imposes the additivity constraint: $\sum M_\beta = 1$.

2.2. Multiplicative Measures

Another extension of the multinomial allows non-negative multipliers M_β $(0 \leq \beta \leq b-1)$ with arbitrary probability distributions. For simplicity, we assume identically distributed multipliers drawn from a random variable M. The limit *multiplicative measure* is called *conservative* when mass is conserved at each stage of the construction: $\sum M_\beta = 1$, or *canonical* when it is only preserved "on average": $\mathbb{E}(\sum M_\beta) = 1$ or $\mathbb{E}M = 1/b$.

Consider the generating cascade of a *conservative* measure μ. In the first stage, we partition the unit interval $[0,1]$ into b-adic cells of length $1/b$, and allocate random masses $M_0, .., M_{b-1}$ to each of these cells. By a repetition of this scheme, the b-adic cell of length $\Delta t = b^{-k}$, starting at $t = 0.\eta_1...\eta_k = \sum \eta_i b^{-i}$, has measure

$$\mu(\Delta t) = M(\eta_1)M(\eta_1,\eta_2)...M(\eta_1,...,\eta_k). \tag{2.1}$$

Multipliers defined at different stages of the cascade are chosen to be statistically independent, and relation (2.1) implies that $\mathbb{E}[\mu(\Delta t)^q] = [\mathbb{E}(M^q)]^k$, or equivalently

$$\mathbb{E}[\mu(\Delta t)^q] = (\Delta t)^{\tau(q)+1}, \tag{2.2}$$

where $\tau(q) = -\log_b \mathbb{E}(M^q) - 1$. The moment of an interval's measure is thus a power functions of the length Δt. This important scaling rule characterizes multifractals.

Modifying the previous construction, we generate a *canonical* measure μ by imposing that the multipliers M_β be statistically independent *within* each stage of the cascade. The mass of the unit interval is now a random variable Ω,[1] and the mass of a b-adic cell takes the form

$$\mu(\Delta t) = \Omega(\eta_1, ..., \eta_k) M(\eta_1) M(\eta_1, \eta_2) ... M(\eta_1, ..., \eta_k).$$

We note that $\Omega(\eta_1, ..., \eta_k)$ has the same distribution as Ω. The measure μ thus satisfies the scaling relationship

$$\mathbb{E}\left[\mu(\Delta t)^q\right] = \mathbb{E}(\Omega^q)\,(\Delta t)^{\tau(q)+1}, \tag{2.3}$$

which generalizes (2.2). We note that multiplicative measures constructed so far are *grid-bound,* in the sense that the scaling rule (2.3) holds only when the length Δt is of the form b^{-k}. In Section 3, we will see how to construct grid-free measures, for which relation (2.3) asymptotically holds for small values of Δt.

2.3. Multifractal Processes

We now recall the formal defintion of the MMAR. Consider the price of a financial asset $P(t)$ on a bounded interval $[0, T]$, and define the *log-price* process

$$X(t) \equiv \ln P(t) - \ln P(0).$$

We model $X(t)$ by compounding a Brownian Motion with a multifractal trading time:

Assumption 1. *$X(t)$ is a compound process*

$$X(t) \equiv B[\theta(t)]$$

where $B(t)$ is a Brownian Motion, and $\theta(t)$ is a stochastic trading time.

Assumption 2. *Trading time $\theta(t)$ is the c.d.f. of a multifractal measure μ defined on $[0, T]$.*

Assumption 3. *The processes $\{B(t)\}$ and $\{\theta(t)\}$ are independent.*

[1] The random variable Ω has interesting distributional and tail properties that are discussed in Mandelbrot (1989a).

Calvet, Fisher, and Mandelbrot (1999) show that the process $X(t)$ is martingale with respect to its natural filtration. The price process $P(t)$ is therefore a semi-martingale, which allows use of stochastic integration to calculate gains from trade. Moreover, there is no arbitrage in a two asset economy containing a riskless asset with a fixed rate of return, and a risky asset with a multifractal price $P(t)$.

The MMAR also has long memory in volatility. For any stochastic process Z with stationary increments, the *autocovariance in levels*

$$\delta_Z(t,q) = Cov(|Z(a,\Delta t)|^q, |Z(a+t,\Delta t)|^q),$$

quantifies the dependence in the size of the process's increments. It is well-defined when $\mathbb{E}|Z(a,\Delta t)|^{2q}$ is finite. For a fixed q, we say that the process has *long memory in the size of increments* if the autocovariance in levels is hyperbolic in t when $t/\Delta t \to \infty$. When the process Z is multifractal, this concept does not depend on the particular choice of q.[2] Under these definitions, both trading time $\theta(t)$ and the process $X(t)$ have long memory in the size of increments.

3. Poisson Multifractals

This section introduces a grid-free multifractal measure in continuous time. The main ingredient of the construction consists of randomizing the times at which the multipliers are changing. More specifically, we assume that the arrival times of the multipliers are following Poisson processes.

3.1. Construction of a Poisson Multifractal Measure

We construct the Poisson multifractal measure on the interval $[0,T]$ as the limit of a multiplicative cascade. In the first stage of the procedure, consider a sequence $\{T_n\}_{n=1}^{\infty}$ of independent random variables with identical exponential density $f(x;\lambda) = \lambda \exp(-\lambda x)$. We wish to randomize arrivals of the first stage multipliers, and use the sequence T_i as the interval between arrivals. Denoting by $N = \max\{n : \sum_{i=1}^{n} T_i < T\}$ the number of arrivals on $[0,T]$, we then define the first stage random instants

$$t_n = \begin{cases} 0 & \text{if } n = 0 \\ \sum_{i=1}^{n} T_i & \text{if } 1 \leq n \leq N \\ T & \text{if } n = N+1 \end{cases}.$$

[2]Provided that $\mathbb{E}|Z(a,\Delta t)|^{2q} < \infty$, as is implicitly assumed in the rest of the paper.

The choice of an exponential distribution guarantees that the probability of arrival at any instant t is independent of past history.

The intervals $\{I_n = [t_n, t_{n+1}] : 0 \leq n \leq N\}$ form a random partition of $[0, T]$. Consider the function $\ell([t, t']) = t' - t$, which provides the length of a given interval. The measure μ_1 is defined by drawing independent random multipliers M_n and uniformly spreading within each interval the masses

$$\mu_1(I_n) = M_n \times \ell(I_n)/T,$$

In order to obtain a non-degenerate limit, we impose that mass be preserved *on average* at each stage of the cascade: $\mathbb{E}\,\mu_1[0, T] = 1$. Since

$$\mathbb{E}\,\mu_1[0, T] = \mathbb{E}\left[\sum_{n=0}^{N} M_n \ell(I_n)\right]/T = \mathbb{E}M,$$

the preservation of mass is equivalent to $\mathbb{E}M = 1$.

We now present the construction of the measure at stage $k \geq 2$. Consider a stage $k-1$ interval

$$I_{j_1,..,j_{k-1}} = [t_{j_1,..,j_{k-1}};\ t_{j_1,..,j_{k-1}+1}],$$

with evenly spread mass $\mu_{k-1}(I_{j_1,..,j_{k-1}})$. Let $\{T_n^{j_1,..,j_{k-1}}\}_{n=1}^{\infty}$ be a sequence of independent random variables with identical exponential density $f(x; 2^k\lambda)$. The sequence $\{T_n^{j_1,..,j_{k-1}}\}_{n=1}^{\infty}$ is moreover assumed to be independent of all other random variables defined up to stage k. We note that the frequency of arrival doubles at each stage[3] and $\mathbb{E}\left(T_n^{j_1,..,j_{k-1}}\right) = (2^k\lambda)^{-1}$. Because we again want to ensure that arrival times are contained in the interval $I_{j_1,..,j_{k-1}}$, it is convenient to define

$$N^j = \max\left\{n : \sum_{i=1}^{n} T_i^j < \ell(I_j)\right\}$$

where $j = (j_1, .., j_{k-1})$. We then define the corresponding increasing set of arrival times

$$t_{j,n} = \begin{cases} t_j & \text{if } \ n = 0 \\ t_j + \sum_{i=1}^{n} T_i^j & \text{if } \ 1 \leq n \leq N^j \\ t_{j_1,..,j_{k-1}+1} & \text{if } \ n = N^j + 1 \end{cases}$$

[3]More specifically, the number $2^k\lambda$ represents the frequency of arrival *within a stage $k-1$ interval*. An alternative specification, which we do not explore on this paper, assumes independent arrival times at each stage of the construction. The forecasting results developed of Section 5 directly extend to this new construction.

On each subinterval $I_{j,n} = [t_{j,n}; t_{j,n+1}]$, we randomly draw a multiplier $M_{j,n}$ and uniformly spread the mass

$$\mu_k\left(I_{j,n}\right) = \left(M_{j_1} ... M_{j_1,...,j_{k-1}}\, M_{j,n}\right) \times \ell(I_{j,n})/T. \tag{3.1}$$

We note that $\mathbb{E}\left[\mu_k(I_j) | \, \mu_{k-1}(I_j)\right] = \mu_{k-1}(I_j)$. By the martingale convergence theorem, $\mu_k(I_j)$ thus converges to a limit $\mu(I_j)$ when $k \to \infty$. The limit measure μ is therefore well defined, and called a *Poisson multifractal measure.* Its total mass is random and satisfies $\mathbb{E}\,\mu[0,T] = 1$.

3.2. Properties of the Measure

The random mass $\mu[0,T]$ defines a random variable $\Omega(T, \lambda)$ whose distribution depends on T, λ and M. We note that $\mathbb{E}\,\Omega(T,\lambda) = 1$, and observe

Proposition 1. *The mass of a subinterval* $[0,t]$ *satisfies*

$$\mu[0,t] \stackrel{d}{=} \frac{t}{T}\Omega(t,\lambda).$$

for all $t < T$.

The random mass $\Omega(t,\lambda)$ plays a crucial role for the scaling properties of the measure.

Proposition 2. *The random variable* $\Omega(T,\lambda)$ *satisfies the invariance relations*

$$\Omega(T,\lambda) \stackrel{d}{=} \Omega(1, \lambda T) \tag{3.2}$$

and

$$\Omega(T,\lambda) = \sum_{j=0}^{N} \frac{t_{j+1} - t_j}{T}\, M_j\, \Omega_j(t_{j+1} - t_j, 2\lambda) \tag{3.3}$$

for all T, λ *and* M.

Proof. See Appendix ■

Equation (3.2) shows that the distribution of the random mass $\Omega(T,\lambda)$ depends on a single parameter. It is thus convenient to define $\Omega(\lambda) \equiv \Omega(1,\lambda)$ and henceforth use a single index to represent it. We also note that equation (3.3) generalizes the invariance relation $\Omega = \sum M_\beta \Omega_\beta$ satisfied by the random mass Ω of a grid-bound measure. We observe

Proposition 3. *If there exists $\lambda > 0$ such that $\mathbb{E}[\Omega(\lambda)^q] < \infty$, then the q-th moments $\mathbb{E}[\Omega(\lambda')^q]$ is finite for all $\lambda' \in (0, \infty)$.*

Proof. See Appendix. ■

The critical moment $q_{crit}(\lambda) = \sup\{q : \mathbb{E}[\Omega(\lambda)^q] < \infty\}$ is thus independent of λ.

As seen in Section 2, grid-bound measures have very specific moment scaling properties. In particular, the moment $\mathbb{E}\mu[t, t+\Delta t]^q$ of an interval's mass varies as a power function of the length Δt. We now show that this property is preserved in the grid free case for small values of Δt. Proposition 1 suggests analysis of the moments of $\Omega(\lambda)$ as $\lambda \to 0$. For this reason, it is convenient to introduce the notation $f(\lambda) \sim g(\lambda)$ to indicate that $f(\lambda)/g(\lambda) \to 1$ as $\lambda \to 0$. When the parameter λ is small, we expect that the multiplier associated with the frequency λ is constant over the interval $[0, T]$. The random mass $\Omega(\lambda)$ is thus aproximately equal to $M\,\Omega(2\lambda)$, and therefore that $\mathbb{E}[\Omega(\lambda)^q] \sim \mathbb{E}(M^q)\,\mathbb{E}[\Omega(2\lambda)^q]$. This implies that $\mathbb{E}[\Omega(\lambda)^q]$ behaves like a power function for small values of λ.

Proposition 4. *The q^{th} moment of the random mass $\Omega(\lambda)$ satisfies*

$$\mathbb{E}\left[\Omega(\lambda)^q\right] \sim c_q \lambda^{\tau(q)+q+1} \textit{as } \lambda \to 0$$

where $\tau(q) = -\log_2(\mathbb{E}\,M^q) - q - 1$ and c_q is a positive constant.

Proof. See Appendix. ■

As in the grid-bound case, the relation $\tau(q) = -\log_2(\mathbb{E}\,M^q) - q - 1$ defines the scaling function $\tau(q)$. The slight difference in the definition comes from a different normalization of the multiplier. We observe that by Hölder's inequality, the function $\tau(q)$ is concave.

This allows derivation of the scaling behavior of a random measure μ associated with a *fixed* parameter λ.

Corollary 1. *For any $q \geq 0$, the q^{th} moment of the measure satisfies*

$$\mathbb{E}\left(\mu[0,t]^q\right) \sim c_{\lambda,q} t^{\tau(q)+1} \text{ as } t \to 0$$

where $c_{\lambda,q}$ is a positive constant.

We can thus apply the quantitative techniques developed in earlier work to the Poisson multifractal.

4. The Financial Model

We now define our model of asset returns. As in the previous section, we consider a random Poisson multifractal measure μ, and define $\theta(t) = \mu[0, t]$ as a trading time. The Poisson multifractal process $X(t) = \ln P(t) - \ln P(0)$ is again defined by compounding a Brownian Motion $B(t)$ with a stochastic trading time $\theta(t)$:

$$X(t) = \ln P(t) - \ln P(0) = B[\theta(t)].$$

This process has properties which makes it very similar to the MMAR. We see immediately that the process $X(t)$ is a martingale, and the price process $P(t)$ is therefore a semi-martingale. By Corollary 1, its moments satisfy the asymptotic scaling rule

$$\mathbb{E}\left[|X(t)|^q\right] = \mathbb{E}[\theta(t)^{q/2}]\mathbb{E}\left(|B(1)|^q\right) \sim C_q t^{\tau(q)+1} \text{ as } t \to 0,$$

where $\tau(q) \equiv \tau_\theta(q/2)$.

The model is now grid-free, in the sense that it makes no assumption on the time instants when the multipliers are changing. The process can also be discretized easily, which facilitates forecasting. Another advantage lies in the treatment of long memory. For a given forecasting problem over the time interval $[0, T]$, we can choose a parameter λ corresponding to the lowest frequency shock we want to consider. In practice, this choice is very similar to the truncation problem encountered when forecasting fractionally integrated processes such as ARFIMA (Baillie, 1996).

4.1. Economic Intuition

Economically, the volatility state vector corresponds to a set of volatility components, each with a different frequency for innovations. Thus, the highest frequencies might correspond to short lived liquidity shocks, while other frequencies could correspond to earnings cycles, business cycles, technological shocks, and so on. This closely corresponds to the economic intuition that different types of volatility shocks have different degrees of persistence. In contrast, most standard models impose that all volatility innovations are statistically identical, and thus use a single decay function to fit many different types of volatility shocks.

Enriching a volatility model with different types of volatility shocks would quickly become unwieldy without additional structure. The multifractal approach builds on the organizing principle that volatility shocks have no favored time scale. All shocks look important at their own natural time scale, appear like low-frequency variation at smaller time scales, and appear like high frequency noise at larger time scales. If we take the model literally, we might search for equilibrium models that produce this type of scaling behavior in volatility. Alternatively, we can view the scaling restrictions as a pragmatic approach to modelling heterogeneity in volatility shocks.

An added feature of the multifractal model that may be attractive from some points of view is that the process is non-ergodic, so that a single observed price path can be consistent with different beliefs about some parameters of the model depending on one's priors. In particular, note that taking the model to data requires a renormalization to fit the average volatility level of the data at the preferred modelling frequency. If we write

$$\ln P(t) - \ln P(0) = \sigma B[\theta(t)],$$

this renormalization is captured by the new parameter σ. A first approach might be to estimate σ with the average volatility over the history of the data on the interval $[0, T]$. However, recalling that $\theta(T) \sim \Omega$, we see that the average volatility over this time period is distributed like $\sigma\Omega^{1/2}$. This suggests that many prior beliefs about σ can be supported by an observed time path. To understand this intuitively, imagine that one knows that no stopping time has occured for the lowest frequency volatility component M_1, so that this component is constant over the life of the data. An individual with a relatively high point mass prior on σ can support this belief by inferring that the realization of M_1 was low, while an individual with a relatively low point mass prior on σ might infer that M_1 was high. Neither belief would be ruled out be the data, but the first investor would forecast volatility to increase over a very long run average, while the second investor would forecast a volatility decrease.

This is a feature of the model that deserves further attention, particularly in the context of equilibrium models that support the price process. It is compatible with the view that investors learn from data, but can never learn perfectly all parameters of the true model (Kurz, 1994; 1997). In the remainder of this paper, we avoid this issue by assuming that all investors know the true process with certainty. Future extensions should be able to incorporate parameter uncertainty via standard methods.

In addition to assuming that the forecaster knows the true price process with certainty, the remainder of the paper assumes that the forecaster uses only the series of past prices to form conditional probabilities over the current volatility state. The intuition behind the conditioning formula is straightforward. When the econometrician first observes a large positive spike to volatility, she weights her beliefs towards states with large realized volatility components at higher frequencies. If the volatility spike persists, however, it becomes more likely that the shock originated in a single large innovation to a low frequency volatility component, rather than a series of shocks to higher frequency components. There are many interesting extensions to this framework. For example, one potential source of value added in the financial services industry is explanation of volatility shocks. Analyst research can potentially refine the conditional probabilites of the volatility state beyond the information given by past prices. For example, analyst research might point towards a long-term macroeconomic explanation for a relatively recent volatility shock, which might lead one to adjust the conditioning formula we present. One obvious place to look for this additional information is in derivative security prices. We discuss some of these extensions in the Conclusion.

5. Computing the Conditional Density

We now examine the forecasting problem in the Poisson multifractal model. Consider a Poisson multifractal process

$$X(t) = \ln P(t) - \ln P(0) = B[\theta(t)]$$

defined on a time interval $[0, T]$. We observe the process X on a time interval $[0, t]$, and want to predict the conditional distribution of $X(s)$ at any future instant $s \geq t$. To solve this problem, we first assume that both X and θ are observed on the interval $[0, t]$, and denote the corresponding information set by $\mathcal{I}_t$. Let $f_{X_s}(x|\mathcal{I}_t)$ and $f_{\theta_s}(\theta|\mathcal{I}_t)$ denote the conditional densities of $X(s)$ and $\theta(s) - \theta(t)$ at date $s \geq t$. By Bayes' rule, the log-price process $X(t)$ satisfies

$$f_{X_s}(x|\mathcal{I}_t) = \int n[x; X(t), \theta] f_{\theta_s}(\theta|\mathcal{I}_t) d\theta, \tag{5.1}$$

where $n(x; \mu, \sigma^2)$ is the density of a Gaussian random variable with mean μ and variance σ^2. This suggests that we calculate the conditional density $f_{\theta_s}(\theta|\mathcal{I}_t)$ of trading time, and then derive from (5.1) the conditional density of the process.

The estimation of the conditional density $f_{\theta_s}(\theta|\mathcal{I}_t)$ is carried out as follows. We first introduce a discretized version of the process, in which stopping times take values on a finite grid and the construction of the measure stops after $\bar{k} + 1$

iterations. We then solve the conditioning problem of the discretized process, and show that the renormalized solution $f_{\theta^*_{\bar{k},s}}(\theta|\mathcal{I}_t)$ converges to the conditional density $f_{\theta_s}(\theta|\mathcal{I}_t)$ of the continuous process as $\bar{k} \to \infty$.

5.1. A Discretized Version of Trading Time

This section defines a continuous, piecewise linear trading time θ^* that permits non-differentiability on a finite grid of regularly spaced instants. The grid corresponds to the set of instants at which innovations to volatility may occur, and the values of θ^* on the grid define a discretized trading-time. We view the trading time θ of Section 3.1 as the true process, and the discretized version as a filter that facilitates forecasting.

We first define a random measure μ^* through a recursive construction with a *finite* number $\bar{k}+1$ of stages, where $\bar{k} \geq 1$ is an integer. The measure is defined on an interval $\left[0, b^{\bar{k}}\right]$ with the integer base $b > 1$, and the construction uses the regular grid $0, 1, ..., T^* = b^{\bar{k}}$. The corresponding trading time θ^* is then defined on the time interval $[0, T]$ by

$$\theta^*(s) = \mu^* \left[0, b^{\bar{k}} s/T\right] \qquad 0 \leq s \leq T.$$

The process θ^* is thus the c.d.f of a homothetic transform of the measure μ^*. This allows us to model trading time on an arbitrary time interval $[0, T]$, while the grid points of μ^* take integer values.

The construction of μ^* begins at stage zero with a unit mass spread evenly on the interval $[0, T^*]$. In the first stage, let $\{T_i^*\}_{i=1}^{\infty}$ be a sequence of iid random variables each having a *geometric* distribution with parameter $\gamma_1 = 1 - \exp(-\lambda T b^{-\bar{k}})$ and mean $1/\gamma_1$.[4] This sequence will provide the lengths between arrival times. To keep the arrival times bounded below T^*, we denote by P the largest integer such that $\sum_{i=1}^n T_i^* < T^*$. The random integers

$$t_n = \begin{cases} 0 & n = 0 \\ \sum_{i=1}^n T_i^* & 1 \leq n \leq P \\ T^* & n = P+1 \end{cases}$$

define a random partition $I_n = [t_n, t_{n+1}]$ of $[0, T^*]$. We define the first stage measure by drawing independent random multipliers M_n and uniformly spreading mass $M_n \times \ell(I_n)/T^*$ on each interval I_n.

[4] The distribution of T_1^* satisfies $\mathbb{P}\{T_1^* = n\} = \gamma_1(1-\gamma_1)^{n-1}$ for all $n \geq 1$.

We proceed similarly in stages $k = 2, .., \bar{k}$ of the cascade by considering intervals $I_{j_1,..,j_{k-1}}$, and their corresponding geometric random variables $\{T_n^{*\, j_1,..,j_{k-1}}\}_{n\geq 1}$ with parameter $\gamma_k = 1 - \exp\left[-\lambda T b^{-(\bar{k}+1-k)}\right]$. The random variable $P^{j_1,..,j_{k-1}}$ gives the number of arrival times that will fall strictly within the corresponding interval, and is used to truncate the series of random arrival times $\left\{t_{j_1,..,j_{k-1},n}\right\}_{0\leq n\leq P^{j_1,..,j_{k-1}}+1}$. Details of stages $k = 2, .., \bar{k}$ can be filled in by referring to the continuous time construction.

Finally in stage $\bar{k}+1$, we draw iid random variables Ω_t for each interval $[t-1, t]$. We assume that each Ω_t is distributed like the limit mass $\Omega\left(\lambda T\right)$ of the continuous time multifractal measure, as defined in Section 3.2. This assumption ties the high-frequency components of the discrete time and continuous time models. The mass of a cell therefore satisfies

$$\mu^*[t-1, t] = b^{-\bar{k}} M_{j_1} ... M_{j_1..j_{\bar{k}}} \Omega_t,$$

where $[t-1, t] \subseteq I_{j_1..,j_{\bar{k}}}$. It convenient to denote by $M_{1,t}, ..., M_{\bar{k},t}$ the multipliers $M_{j_1}, ..., M_{j_1..,j_{\bar{k}}}$, so that $M_{k,t}$ can be viewed as the value of the k-th multiplier at date t.

The random measure μ^* has a finite state space, which greatly simplifies the forecasting problem. We now show that the state space can be represented as a Markov chain. For a given t, define the sigma-field $\mathcal{F}_{t-1}$ generated by the set $\{(t_{j_1,..,j_k}, M_{j_1,..,j_k})\}_{t_{j_1,..,j_k}<t}$ of past stopping times and multipliers. Denote by κ_t the lowest frequency change between $t-1$ and t, or formally $\kappa_t = \bar{k}+1$ if $t \notin \{t_{j_1,..,j_k}\}$ and $\kappa_t = \inf\{k : t = t_{j_1,..,j_k}\}$ otherwise. Since time increments are geometrically distributed, κ_t is independent of past stopping times and satisfies

$$\mathbb{P}\{\kappa_t = k\} = \begin{cases} \gamma_1 & k = 1 \\ \gamma_k(1-\gamma_1)..(1-\gamma_{k-1}) & k = 2, .., \bar{k} \\ 1 - \sum_{k=1}^{\bar{k}} \mathbb{P}\left(\kappa_t = k\right) & k = \bar{k}+1 \end{cases}$$

The distribution of the future multipliers $(M_{1,t+1}, ..., M_{\bar{k},t+1})$ thus only depends on the current multipliers $(M_{1,t}, ..., M_{\bar{k},t})$, and the vector

$$X_t = (M_{1,t}, .., M_{\bar{k},t})$$

is thus a Markov chain. This property will greatly facilitate our treatment of the conditioning problem.

We assume in the rest of this section that a multiplier takes a finite number b_m of values. The vector X_t can therefore take $d = b_m^{\bar{k}}$ values $x^1, .., x^d \in \mathbb{R}^{\bar{k}}$.

The dynamics of the Markov chain X_t are characterized by the transition matrix $A = (a_{i,j})_{1\leq i,j\leq d}$ with components $a_{ij} = \mathbb{P}(X_{t+1} = x^j \,|\, X_t = x^i)$. Moreover, we know that

$$\begin{aligned} a_{ij} &= \sum_{\kappa=1}^{\bar{k}+1} \mathbb{P}(\kappa_t = \kappa)\, \mathbb{P}\left(X_{t+1} = x^j \,\middle|\, X_t = x^i,\ \kappa_t = \kappa\right) \\ &= \sum_{\kappa=1}^{\bar{k}+1} \mathbb{P}(\kappa_t = \kappa)\, 1_{\{x_1^i = x_1^j, \ldots, x_{\bar{k}-1}^i = x_{\bar{k}-1}^j\}} \left[\prod_{k=\kappa}^{\bar{k}} \mathbb{P}\left(M = x_k^j\right)\right] \end{aligned}$$

where x_k^i denotes the kth component of vector x^i.

This suggests an alternative method to construct and simulate the measure μ^*. Instead of going down frequency levels, we can iteratively build the measure through time. At date $t = 0$, we draw random variables $(M_{1,1}, ..., M_{\bar{k},1}, \Omega_1)$, and set $\mu^*[0,1] = b^{-\bar{k}} M_{1,1} ... M_{\bar{k},1} \Omega_1$. Given X_t, we generate the next increment of trading time $\mu^*[t, t+1]$ and the next state vector X_{t+1} by drawing the index κ_{t+1}, new multipliers $M_{\kappa_t, t+1}$, ..., $M_{\bar{k}, t+1}$, and a high frequency component Ω_{t+1}. Multipliers corresponding to frequencies lower than κ_{t+1} remain unchanged: $(M_{1,t+1}, ..., M_{\kappa_{t+1}-1,t+1}) = (M_{1,t}, ..., M_{\kappa_{t+1}-1,t})$. The mass $\mu^*[t, t+1]$ is then equal to the product $b^{-\bar{k}} M_{1,t+1} ... M_{\bar{k},t+1} \Omega_{t+1}$.

Constructing the measure through time rather than through stages more closely parallels our economic intuition of trading time. Each period, the economy can receive innovations to each volatility component. The volatility components differ only by the expected arrival frequency of their respective innovations. If no innovation is received at a given frequency, the volatility component at that frequency does not change from the previous period. If an innovation is received, the new volatility component at that frequency is obtained by drawing a new multiplier with distribution M. This draw is independent of all other random variables in the model. The assumption that all draws have distribution M independent of time and frequency gives the model its scaling properties.

One point to note is that the transition matrix defined above assumes a type of cascade effect in volatility innovations. If a volatility innovation arrives at a given frequency, all higher frequencies are assumed to also experience volatility innovations. Economically, this corresponds to the idea that relatively low frequency shocks to technology or demographics cause revisions in higher frequency economic variates such as the earnings cycle or aggregate liquidity. It also implies that lower frequency volatility shocks result in correspondingly larger average changes in volatility.[5]

Another interesting point is that we assume each volatility state variable is constant over time, except when an innovation arrives, at which time the new value is independent of the past value. We note that this implies an identical stationary distribution for each component of the volatility state vector, and mean reversion in each component at each instant. However, because the rule for total volatility $\mu^*[t, t+1] = b^{-\bar{k}} M_{1,t+1} ... M_{\bar{k},t+1} \Omega_{t+1}$ is multiplicative, volatility itself need not be mean reverting at each instant, although it will be over long time periods. This phenomenon will be observed in our simulations.

5.2. Inferring Current Multipliers

We now turn to the inference problem of a market participant who observes the trading time increments $\mu_t = \mu^*[t-1, t]$, but not the exact value of the multipliers $X_t = (M_{1,t}, .., M_{\bar{k},t})$. The investor has a conditional distribution over the multipliers that can be calculated recursively since X_t is a Markov chain. Throughout the remainder of the paper, we assume the investor knows the true process with certainty.

We previously imposed that Ω_t be distributed like the limit mass $\Omega(\lambda T)$ of the continuous time measure. In the remainder of the paper, we choose a specification of the multiplier M such that the random mass Ω_t has a density $f_\Omega(\omega)$ with respect to Lebesgue measure. Given $a = (a_1, .., a_k) \in \mathbb{R}^k$, we find it convenient to denote by $g(a)$ the product $a_1..a_k$. Since $\mu_t = g(X_t)\,\Omega_t$, we note that the conditional density $f_{\mu_t}(\mu | X_t = x^i)$ satisfies

$$f_{\mu_t}\left(\mu | X_t = x^i\right) = \frac{1}{g(x^i)} f_\Omega \left[\frac{\mu}{g(x^i)}\right]. \tag{5.2}$$

Let $\mathcal{I}_t$ denote the set of past trading time increments $(\mu_i)_{1 \leq i \leq t}$. The investor has conditional probabilities

$$\Pi_t^j = \mathbb{P}\left(X_t = x^j \left| \mathcal{I}_t \right.\right)$$

over $x^1, .., x^d$, which are concatenated in the row vector $\Pi_t = \left(\Pi_t^1, .., \Pi_t^d\right) \in \mathbb{R}_+^d$. We note that

$$\begin{aligned} \Pi_{t+1}^j &= \mathbb{P}\left(X_{t+1} = x^j \left| \mathcal{I}_t, \mu_{t+1} \right.\right) \\ &= f_{\mu_{t+1}}\left(\mu_{t+1} | X_{t+1} = x^j\right) \mathbb{P}\left(X_{t+1} = x^j \left| \mathcal{I}_t \right.\right) / f_{\mu_{t+1}}\left(\mu_{t+1} \left| \mathcal{I}_t \right.\right), \end{aligned}$$

[5] An alternative model that we do not explore in this paper assumes that the arrival of volatility innovations is independent across frequencies.

which, by (5.2), can be rewritten

$$\Pi_{t+1}^{j}=\frac{\left(\sum_{i=1}^{d}a_{ij}\Pi_{t}^{i}\right)\,f_{\Omega}[\mu_{t+1}/g\left(x^{j}\right)]/g\left(x^{j}\right)}{f_{\mu}\left(\mu_{t+1}\left|\mathcal{I}_{t}\right.\right)}$$

Since the conditional probabilities Π_{t+1}^{j} add up to one, we conclude that

$$\Pi_{t+1}^{j}=\frac{\left(\sum_{i=1}^{d}a_{ij}\Pi_{t}^{i}\right)\;f_{\Omega}[\mu_{t+1}/g\left(x^{j}\right)]/g\left(x^{j}\right)}{\sum_{j'=1}^{d}\left\{\left(\sum_{i=1}^{d}a_{ij'}\Pi_{t}^{i}\right)\;f_{\Omega}\left[\mu_{t+1}/g\left(x^{j'}\right)\right]/g\left(x^{j'}\right)\right\}}. \tag{5.3}$$

Let $\iota=(1,.,1)\in\mathbb{R}^{d}$, and let

$$\omega(\mu_{t+1})=\left[\frac{f_{\Omega}\left(\mu_{t+1}/g\left(x^{1}\right)\right)}{g\left(x^{1}\right)},...,\frac{f_{\Omega}\left(\mu_{t+1}/g\left(x^{d}\right)\right)}{g\left(x^{d}\right)}\right]$$

be a row vector of conditional densities. For any $x,y\in\mathbb{R}^{d}$, denote the Hadamard product $x*y=(x_{1}y_{1},..,x_{d}y_{d})$. We infer that

$$\Pi_{t+1}=\frac{\omega(\mu_{t+1})*\Pi_{t}A}{\left[\omega(\mu_{t+1})*\Pi_{t}A\right]\iota'}.$$

We can therefore recursively calculate Π_{t} given an initial prior Π_{0}. We select the prior Π_{0} that assigns $\Pi_{0}^{i}=\mathbb{P}\left(X_{t}=x^{i}\right)$, which is the unconditional (i.e., stationary) distribution of X_{t}.

5.3. Forecasting Future Volatility

We now consider the forecasting problem of an investor who observes at date t the set $\mathcal{I}_{t}$ of past trading time increments $(\mu_{i})_{1\leq i\leq t}$. As in the previous section, the agent assigns probabilities Π_{t} on the *current* vector X_{t}. The conditional probabilities of *future* multipliers $\widehat{\Pi}_{t,n}=\mathbb{P}\left(X_{n}\left|\mathcal{I}_{t}\right.\right)$ are then obtained by multiplication of the transition matrix:

$$\widehat{\Pi}_{t,n}=\Pi_{t}A^{n-t}$$

for all $n\in\{t,..,T\}$.

In order to forecast prices, we also want to estimate the conditional density at date t of the trading time increment $\sigma_{n}^{2}\equiv\theta_{n}-\theta_{t}$. By Bayes' rule, the conditional density $f_{\sigma_n^2}\left(\sigma^2|\mathcal{I}_t\right)$ satisfies

$$f_{\sigma_n^2}\left(\sigma^2|\mathcal{I}_t\right)=\sum_{i=1}^{d} f_{\theta_n}\left(\theta_t+\sigma^2\middle|\, X_n=x^i,\mathcal{I}_t\right)\widehat{\Pi}^i_{t,n}. \tag{5.4}$$

The conditional densities $f_{\theta_n}\left(\theta|\, X_n=x^i,\mathcal{I}_t\right)$ in the sum can be computed by a recursive algorithm. We know that for $n=t+1$,

$$f_{\theta_{t+1}}\left(\theta|\, X_{t+1}=x^i,\mathcal{I}_t\right)=f_\Omega\left[\frac{\theta-\theta_t}{g(x^i)}\right]/g(x^i),$$

and observe that for all $n\geq t+1$,

$$\begin{aligned} f_{\theta_{n+1}}\left(\theta|\, X_{n+1}=x^j,\mathcal{I}_t\right) &= \int f_{\theta_{n+1}}\left(\theta|\,\theta_n=\theta',X_{n+1}=x^j\right)\, f_{\theta_n}\left(\theta'|\, X_{n+1}=x^j,\mathcal{I}_t\right)d\theta' \\ &= \int_0^\theta \frac{f_\Omega\left[(\theta-\theta')/g(x^j)\right]}{g(x^j)}\, f_{\theta_n}\left(\theta'|\, X_{n+1}=x^j,\mathcal{I}_t\right)d\theta'. \end{aligned}$$

Since by Bayes' rule

$$\begin{aligned} f_{\theta_n}\left(\theta|\, X_{n+1}=x^j,\mathcal{I}_t\right) &= \sum_{i=1}^{d} f_{\theta_n}\left(\theta|\, X_{n+1}=x^j,X_n=x^i,\mathcal{I}_t\right)a_{ij} \\ &= \sum_{i=1}^{d} f_{\theta_n}\left(\theta|\, X_n=x^i,\mathcal{I}_t\right)a_{ij}\widehat{\Pi}^i_{t,n}/\widehat{\Pi}^j_{t,n+1}. \end{aligned}$$

we can recursively calculate the conditional densities $f_{\theta_n}\left(\theta|\, X_n=x^i,\mathcal{I}_t\right)$ in every period. Applying (5.4), we then infer the forecast density $f_{\sigma_n^2}\left(\sigma^2|\mathcal{I}_t\right)$ of trading time increments.

5.4. Consistency

Section 5.1 defined a piecewise linear trading time θ^* using a construction with a fixed number of stages $\bar{k}+1$. We now index the piecewise linear trading times $\theta^{\bar{k}}$ by the number of stages $\bar{k}+1$ in their construction, and provide conditions under which sequences converge to the non-differentiable trading time θ described in Section 3.1. This has implications for the consistency of forecasts obtained from the discretized process.

Without loss of generality, we assume that θ and $\theta^{\bar{k}}$, $\bar{k} \in \mathbb{N}$, are constructed on the interval $[0,1]$. In order to extend the set of convergence conditions, we introduce a generalized version of the discrete construction described in Section 5.1. In particular, we previously required the discrete construction to occur on a grid of size $\Delta t = b^{-\bar{k}}$. We now allow the grid size to be provided by a different constant $c > 1$, so that $\Delta t = c^{-\bar{k}}$, and the grid takes values $t = 0, 1/c^k, ..., 1$. The construction of $\theta^{\bar{k}}$ is now identical to section 5.1, except that the distribution of geometric random variables $\left\{T_i^{j_1,...,j_k}\right\}$ is now characterized by $\gamma_{k,\bar{k}} = 1 - \exp(-\lambda T b^{k-1}/c^{\bar{k}})$.

Let $\mathcal{C}$ denote the space of real-valued continuous functions defined on the interval $[0,1]$. Random functions θ and θ^k can be viewed as probability measures on $\mathcal{C}$. We provide two alternative conditions under which the sequence of trading times $\{\theta^k\}_{k=1}^{\infty}$ weakly converges to the process θ.[6]

Condition 1. $b < c$

Condition 2. $\mathbb{E}(M^2)b < c^2$

The first condition requires that the discretization becomes fine more rapidly than the growth rates of the volatility frequencies. The second condition allows that the rate of discretization may be equal to or even slower than the rate of increase of the highest frequency volatility component. In particular, for the leading case $b = c$, we require that $\mathbb{E}(M^2) = 1 + Var(M) < b$. This condition holds for the distribution of multipliers calculated to the DM/USD exchange rate. We can then show

Theorem 1. *Under Condition 1 or Condition 2, the sequence $\{\theta^k\}_{k=1}^{\infty}$ of discretized trading times weakly converges to the continuous time process θ.*

Proof. See Appendix B. ■

This implies the consistency of the density forecasts presented in Section 5.3.

6. Implementation

We demonstrate implementation of this methodology for a base four ($b = 4$) trinomial ($b_m = 3$) model calibrated to a daily series of DM / USD exchange rates.

[6] Billingsley (1999), Pollard (1984), and Davidson (1994) provide excellent presentations of the weak convergence of stochatic measures and processes.

All of the work that follows uses simulated data. We iterate to $k = 5$ stages with frequencies of $b^{-1}, ..., b^{-k}$ in units of days. Thus, the lowest frequency multiplier has an expected time to arrival of approximately four years. The number of volatility states is $d = b_m^k = 243$. The random variable M is assumed to take the values $(m_1, m_2, m_3) = (0.8, 1.0, 2.0)$ with probabilities $(p_1, p_2, p_3) = (0.5, 0.4, 0.1)$. The distribution of the high frequency component Ω is approximated by repeated Monte Carlo sampling from a $k = 10$ stage measure. Throughout this exercise, we assume that the specification of the process and its parameters is known with certainty. We also assume that volatility itself is observable. This can be justified by assuming that we have access to a very high frequency data set that allows calculation of realized volatilities as in Andersen, Bollerslev, Diebold, and Labys (1999). Alternatively, we note that the extension to conditioning on returns rather than volatility is straightforward, and given in equation (5.1).

Figure 1 shows the conditional probabilities of states across time for a simulated data set. Computing the forecast of future volatility is complicated by a dimensionality problem. At the first iteration, calculating $\mathbb{P}(\theta_{t+1}|X_{t+1})$ can be accomplished by approximating the distribution of Ω with a discrete probability over a fixed number of points n_m. At the next iteration, however, for each value of X_{t+2}, there would be $d \times n_m$ possible values of $\mathbb{P}(\theta_{t+1}|X_{t+2})$ and then $d \times n_m^2$ possible values of $\mathbb{P}(\theta_{t+2}|X_{t+2})$, with the problem growing like a power of $d \times n_m$ at each iteration. To address this problem, we approximate both of these conditional probabilities with an n_m point discretization at each iteration. Using this methodology, we can quickly calculate analytical multistep ahead forecasts for the conditional distribution total volatility. To demonstrate the forecasting method, we choose a time period from the simulated data with a very high level of recent volatility. This allows us to observe the memory of the model as the future volatility forecasts slowly decay.

Figure 2 shows our forecast density of average volatility over multiple time periods. Since the current level of volatility is relatively high, we expect our forecast of average volatility to gradually decay with the forecast horizon. This is evident over the long run, but note that initially both the conditional distribution and the point forecast shift upward before beginning to decline. This is a consequence of the multiplicative interaction of the volatility components. If the current state of the low frequency frequency components is relatively high, and the high frequency components have low values, then one would forecast volatility to initially rise before the low frequency components begin reverting to their mean. The model can thus support an interesting variety of beliefs about the behavior of future volatility.

7. Application to Option Pricing

We now apply our conditioning methodology to option pricing with the goal of comparing implied volatilities obtained from the multifractal model with stylized facts observed in the empirical literature. In previous empirical work, we focus on exchange rate returns because they are more likely to be conditionally symmetric with respect to the level of volatility.[7] For the same reason, we now emphasize stylized facts of currency option implied volatilities, for which conditional skewness is not as large a concern as with equity options. We focus on implied volatility smile, the term structure of the implied volatility smile, and, implicitly, the term structure of at the money implied volatilities. In effect, we will be assessing stylized facts on the entire implied volatility surface, across moneyness and maturity, of currency options.

Since the Brownian Motion B and the trading time θ are hit by independent shocks, derivative assets are not dynamically redundant. Thus, in contrast with the deterministic trading time $\theta(t) = \sigma^2 t$ considered by Merton and Scholes (1973), absence of arbitrage does not give a unique pricing rule. We instead choose a pricing rule that is consistent with absence of arbitrage and has reasonable implications for implied volatilities. The pricing rule has also been widely used in the literature for general stochastic volatility models.

We first consider a real multifractal asset P whose dynamics are described by the continuous time multifractal model, and seek to evaluate at date t a European derivative asset with contingent payoff $\widetilde{a}_\tau = a[P(t+\tau)]$ at date $t+\tau$. Let $\sigma_\tau^2 \equiv \theta(t+\tau) - \theta(t)$ denote the random total volatility over the period $[t, t+\tau]$. Conditional on σ_τ^2, the log-price has a Gaussian distribution $\ln P_{t+\tau} \sim \mathcal{N}(\ln P_t, \sigma_\tau^2)$, and the forecasting method allows us to calculate the conditional density $g_\tau\left(\sigma_\tau^2 | \mathcal{I}_t\right)$over volatility levels σ_τ^2. We assume a constant riskless rate of return r. In the absence of arbitrage, there exists an equivalent probability measure $\mathbb{P}^*$ such that $\pi(\widetilde{a}_\tau) = e^{-r\tau}\mathbb{E}^*(\widetilde{a}_\tau)$. We consider the pricing rule

$$\pi(\widetilde{a}_\tau) = e^{-r\tau} \int_0^{+\infty} \left[\int_0^{+\infty} a(P) f^*_{P,\sigma,\tau}(P) dP \right] g_\tau(\sigma|\mathcal{I}_t) d\sigma, \tag{7.1}$$

[7] Capturing leverage effects in the multifractal model would require removing the assumption that $B(t)$ is independent of $\theta(t)$ to allow negative correlation between the two.

where $f^*_{P,\sigma,\tau}(P)$ is the density of a random variable $P^*_{\tau,\sigma}$ with distribution

$$\ln P^*_{\tau,\sigma} \sim \mathcal{N}(\ln P_t + r\tau - \sigma^2/2, \sigma^2). \tag{7.2}$$

Under this pricing rule, the price of a European call with strike K can be written as a simple integral over Black-Scholes prices:

$$C_{BS}\left(P_t, K, \tau, \mathcal{I}_t\right) = \int_0^{+\infty} C_{BS}(P_t, K, \tau, \sigma\tau^{-1/2}) g_\tau(\sigma|\mathcal{I}_t) d\sigma,$$

which is consistent with the observed prices of riskless bonds and the risky asset at date t. We also note that this is the same pricing rule obtained by Hull and White (1987) for stochastic volatility following a geometric Brownian Motion. Hull and White show that this pricing rule is appropriate when volatility is uncorrelated with aggregate consumption. While this assumption is questionable when P is an equity price, it is more reasonable for pricing options on exchange rates, which we now consider.

When P is an exchange rate, we maintain equation (7.1), but adjust the distribution of $P^*_{\tau,\sigma}$ to

$$\ln P^*_{\tau,\sigma} \sim \mathcal{N}(\ln P_t - \sigma^2/2, \sigma^2). \tag{7.3}$$

This modification is necessary since investors do not hold the foreign currency itself but the foreign bond. We assume for simplicity that the foreign bond yields the same rate of return r as the domestic bond.[8] Among many adjustments to (7.2) that would prevent arbitrage when P is an exchange rate, we choose (7.3) because both are based on the principle that $\pi_\tau\left(P_{t+\tau}|\sigma_\tau\right)$ is independent of σ_τ. The price of a European call written on the exchange rate $P_{t+\tau}$ is now an integral over Garman-Kohlhagen (1983) prices:[9]

$$C_{GK}\left(P_t, K, \tau, \mathcal{I}_t\right) = \int_0^{+\infty} C_{GK}(P_t, K, \tau, \sigma) g_t(\sigma|\mathcal{I}_t) d\sigma, \tag{7.4}$$

where

$$C_{GK}\left(P_t, K, \tau, \sigma\right) = e^{-r\tau}\left[P_0\Phi(d_1) - X\Phi(d_2)\right], \tag{7.5}$$

and d_1 and d_2 take the same values as in the Black-Scholes equation.

[8]It is straightforward to generalize to different constant riskless returns for foreign and domestic bonds, but this generalization does not affect implied volatilities under our assumptions.

[9]The Garman-Kohlhagen formula with constant volatility is a simple adjustment to Black-Scholes that is appropriate when the underlying asset is an exchange rate.

To calculate an implied volatility we first obtain the call price (7.4) at a given strike K and maturity τ using the condtional density $g_\tau(\sigma_\tau|\mathcal{I}_t)$. We then use the basic Garman-Kohlhagen formula (7.5) to back out the single implied volatility level that is consistent with this price, i.e., $\sigma_{GKIV} = C_{GK}^{-1}(P_t, K, \tau, C_{GK}(P_t, K, \tau, \mathcal{I}_t))$.

Figure 3 shows implied volatilities from simulated data for many strikes and several maturities, conditioning on the set of past prices $\mathcal{I}_t$. The implied volatilities show the symmetric smile we might expect to find in currency option prices. Long-memory is evident in the slow decay in the term structure of at the money implied volatilities. Figure 4 shows implied volatilities for strikes that are rescaled by the square root of maturity; i.e., $\ln P_t - \ln\left[K_\tau \tau^{-1/2}\right]$ is held constant. The implied volatility smile decays slowly with maturity, indicating that the tails of the conditional distribution gradually become thinner as the horizon increases.

8. Conclusion

This paper has developed analytical forecasting methods for a new class of stochastic processes, the Poisson multifractals. We modify the multifractal model developed by Mandelbrot, Fisher, and Calvet (1997) to provide a version that is fully time stationary. The model captures the volatility persistence and thick tails that characterize many financial time-series. We specify volatility to be the multiplicative product of an infinite sequence of random functions, each having innovations with Poisson arrivals of different frequencies. The model thus incorporates volatility clustering at all horizons, a feature we relate to economic factors with different time scales such as technology shocks, business cycles, and liquidity shocks.

The continuous time version of the model compounds a standard Brownian Motion with a random time-deformation process that is obtained from the multiplicative construction of volatility. The model implies semi-martingale prices, and thus precludes arbitrage in a standard two-asset setting. Squared returns have long-memory, and the highest finite moment of returns may have any value greater than two. This wide range of tail behaviors is fully provided by intermittent bursts of volatility modifying a standard Brownian Motion.

Forecasting is facilitated by a discretized version of the process with a finite state space and simple Markovian structure. We show that the discretized version converges to the continuous process as the time scale goes to zero, and that forecasts from the discrete process are consistent. We apply the forecasting algorithm to simulated data from a process calibrated to DM/USD exchange rates. The results demonstrate that current volatility is not a sufficient statistic for forecasting. Past data permits conditioning over the full set of volatility state variables, each of which has different persistence. While volatility is always mean reverting in the long run, we show by example that short run forecasts can vary considerably. If volatility is high primarily because of contributions from long run volatility components, one may forecast volatility to increase at short horizons before mean reverting over long horizons. The model thus provides a flexible non-linear structure, but is analytically quite tractable.

We also develop option pricing methods consistent with the absence of arbitrage, and show through simulation that the model captures implied volatility smiles that slowly decay as the time horizon grows longer. Future empirical work can apply these methods to a variety of problems. Because the model parsimoneously incorporates both thick tails and long-memory, it is appropriate to test on options of any maturity or degree of moneyness.

For extensions to equity data, future work will need to consider the treatment of expected returns. In particular, permitting negative autocorrelation between the Brownian Motion and trading time provides one way of incorporating conditional skewness. Another intriguing property is that the model supports differing beliefs regarding the long-run behavior of volatility. In particular, average volatility over very long sample spans need not converge to its expected value. Thus individuals with different priors on the unconditional mean volatility can support these beliefs by inferring different values of long horizon volatility components. The model thus offers challenges in both empirical work and financial theory.

9. Appendix A

9.1. Proof of Proposition 2

We separately establish these two results.

A. We prove (3.2) by jointly contructing two measures μ and μ' defined on intervals $[0, T]$ and $[0, 1]$. In the first stage of the cascade, we define μ_1 by drawing exponential variables T_i^1 with mean $1/\lambda$, instants $t_n = \sum_{i=1}^n T_i^1$, and random multipliers M_n. We define μ_1' by considering the instants $t_n' = t_n/T$, and uniformly spreading mass $\mu_1'(I_n') = M_n(t_{n+1}' - t_n')$ on each interval $I_n' = [t_n', t_{n+1}']$. We note that $\mu_1'(I_n') = \mu_1(I_n)$ and thus $\mu_1' = \mu_1 \circ \varphi$, where φ denotes the homothetic transformation $\varphi(x) = x/T$. Note that each instant $t_n' = \sum_{i=1}^n (T_i^1/T)$ is the sum of random exponentials with mean $1/(\lambda T)$. The measure μ_1' is therefore the first stage in the construction of a multifractal measure on $[0, 1]$ with parameters λT and M. This argument easily generalizes to $\mu_k' = \mu_k \circ \varphi$ at each stage of the construction, implying $\mu'[0, 1] = \mu[0, T]$ or $\Omega(T, \lambda, M) \overset{d}{=} \Omega(1, \lambda T, M)$.

B. We now turn to the proof of (3.3). Recursive relation (3.1) implies

$$\begin{aligned} \mu\left(I_{j_1..,j_{k-1}}\right) &= \mu_{k-1}\left(I_{j_1..,j_{k-1}}\right) \Omega[\ell(I_{j_1..,j_{k-1}}), 2^k\lambda] \\ &= M_{j_1}...M_{j_1,..,j_{k-1}}\ \Omega[\ell(I_{j_1..,j_{k-1}}), 2^k\lambda]\ \ell(I_{j_1..,j_{k-1}})/T. \end{aligned}$$

Moreover, since

$$\mu_k\left(I_{j_1,..,j_{k-1}}\right) = M_{j_1}...M_{j_1,..,j_{k-1}} \sum_{n=0}^{N^{j_1,..,j_{k-1}}} M_{j_1,..,j_{k-1},n}\, \ell(I_{j_1..,j_{k-1},n})/T$$

we infer that

$$\begin{aligned} \mu\left(I_{j_1,..,j_{k-1}}\right) &= M_{j_1}...M_{j_1,..,j_{k-1}} \\ &\quad \sum_{n=0}^{N^{j_1,..,j_{k-1}}} \frac{\ell(I_{j_1..,j_{k-1},n})}{T} M_{j_1,..,j_{k-1},n}\, \Omega[\ell(I_{j_1..,j_{k-1},n}), 2^{k+1}\lambda] \end{aligned}$$

Apply this equation when $k = 0$.

9.2. Proof of Proposition 3

We first show that

$$c^q \mathbb{E}\left[\Omega(c\lambda)^q\right] \leq \mathbb{E}\left[\Omega(\lambda)^q\right] \leq n^{\max(0,1-q)} \mathbb{E}\left[\Omega(\lambda/n)^q\right] \tag{9.1}$$

for every integer $n \geq 1$ and all real numbers $\lambda > 0$, $0 < c < 1$, $q \geq 0$. Let μ denote a measure defined on $[0,1]$ such that $\mu[0,1] \stackrel{d}{=} \Omega(\lambda)$. By Proposition 1, the random variable $\mu[0,c]$ is distributed like $c\Omega(c\lambda)$ and satisfies $\mu[0,c] \leq \mu[0,1]$, which implies $c^q \mathbb{E}\left[\Omega(c\lambda)^q\right] \leq \mathbb{E}\left[\Omega(\lambda)^q\right]$. Moreover since $\mu[0,1] = \mu[0,1/n] + ... + \mu[(n-1)/n, 1]$, we infer that

$$(\mu[0,1])^q \leq \max(n^{q-1}, 1)\left(\mu[0,1/n]^q + ... + \mu[(n-1)/n, 1]^q\right)$$

and therefore $\mathbb{E}\left[\Omega(\lambda)^q\right] \leq n^{\max(0,1-q)} \mathbb{E}\left[\Omega(\lambda/n)^q\right]$. We conclude that $\mathbb{E}\left[\Omega(\lambda)^q\right] < \infty$ for a given λ implies that $\mathbb{E}\left[\Omega(\lambda')^q\right] < \infty$ for all $\lambda' \in (0, \infty)$.

9.3. Proof of Proposition 4

We divide the proof in three steps.

Step 1. We want to show that $\mathbb{E}\left[\Omega(\lambda)^q\right] / \mathbb{E}[\Omega(2\lambda)^q] \to \mathbb{E}(M^q)$ as $\lambda \to 0$. The invariance property (3.3) implies

$$\mathbb{E}\left[\Omega(\lambda)^q\right] = X_1 \mathbb{P}(N = 0) + X_2 \mathbb{P}(N > 0),$$

where $X_1 = \mathbb{E}(M^q)\, \mathbb{E}[\Omega(2\lambda)^q]$, and

$$X_2 = \mathbb{E}\left[\left\{\sum_{j=0}^{N} M_j \left(t_{j+1} - t_j\right) \Omega_j[2\lambda(t_{j+1} - t_j)]\right\}^q \middle| N > 0\right].$$

Since $\mathbb{P}(N = 0) = \exp(-\lambda) \to 1$, the first component $X_1 \mathbb{P}\left[N_\lambda(1) = 0\right]$ is equivalent to $\mathbb{E}(M^q)\, \mathbb{E}[\Omega(2\lambda)^q]$ as $\lambda \to 0$. Second, recall that

$$\left(\sum_{i=1}^{n} x_i\right)^q \leq \max(n^{q-1}, 1)\left(\sum_{i=1}^{n} x_i^q\right)$$

for any $n \geq 1$, $(x_1, .., x_n) \in R_+^n$, $q \geq 0$. Relation (9.1) then implies

$$X_2 \leq K_q(\lambda)\, \mathbb{E}(M^q)\, \mathbb{E}[\Omega(2\lambda)^q]$$

where $K_q(\lambda) = \mathbb{E}\left[N^{\max(q,1)}|N>0\right]$. When $\lambda \to 0$, the coefficient $K_q(\lambda)\mathbb{P}(t_1 < 1) = \mathbb{E}\left[N^{\max(q,1)}\right]$ is locally equivalent to λ,[10] and the ratio $X_2\mathbb{P}(t_1<1)/\mathbb{E}[\Omega(2\lambda)^q]$ converges to 0.

Step 2. The function $g(\lambda) = \mathbb{E}[\Omega(\lambda)^q]$ is continuous and satisfies $g(\lambda)/g(2\lambda) \to \mathbb{E}(M^q)$ as $\lambda \to 0$. When $q \neq 1$, Jensen's inequality implies $\mathbb{E}(M^q) \neq 1$, and therefore there exists a slowly varying function $c_q(\lambda)$ such that $g(\lambda) \sim c_q(\lambda)\lambda^{-\log_2 \mathbb{E}(M^q)}$. Finally, we note that this relation also holds when $q = 1$.

Step 3. Since $\mathbb{E}[\Omega(\lambda)^q] \geq \mathbb{E}(M^q)\,\mathbb{E}[\Omega(2\lambda)^q]\exp(-\lambda)$, the function $\psi(\lambda) = e^{-\lambda}g(\lambda)\lambda^{\log_2 \mathbb{E}(M^q)}$ is decreasing, and thus converges to a (possibly infinite) limit as $\lambda \to 0$. We note moreover that

$$g(\lambda) \leq \mathbb{E}(M^q)\,\mathbb{E}[\Omega(2\lambda)^q]\,\left[e^{-\lambda} + K_q(\lambda)\mathbb{P}(t_1<1)\right],$$

implying that

$$e^{\lambda}g(\lambda)\lambda^{\log_2 \mathbb{E}(M^q)} \leq e^{2\lambda}g(2\lambda)(2\lambda)^{\log_2 \mathbb{E}(M^q)}[e^{-2\lambda} + K_q(\lambda)\mathbb{P}(t_1<1)e^{-\lambda}].$$

Since $e^{-2\lambda} + K_q(\lambda)\mathbb{P}(t_1<1) = 1-\lambda+o(\lambda) \leq 1$ for small values of λ, the function $e^{\lambda}g(\lambda)\lambda^{\log_2 \mathbb{E}(M^q)}$ is increasing and locally bounded. We conclude that $\psi(\lambda)$ is locally bounded and thus converges to a limit c_q when $\lambda \to 0$. ■

10. Appendix B

Consider a continuous Poisson multifractal θ, and the corresponding sequence $\{\theta^k\}$ of piecewise linear trading times constructed in Section 5.1. We prove in this Appendix that the sequence $\{\theta^k\}$ converges to θ in the weak topology of $\mathcal{C}$. For any continuous function $x \in \mathcal{C}$, it is convenient to consider the modulus of continuity

$$w(x,\delta) = \sup_{|t-s|<\delta} |x(t)-x(s)|.$$

Our proof is based on the following

[10] This stems form the fact that

$$\mathbb{E}(N^q) = \sum_{n=0}^{+\infty} \frac{e^{-\lambda}\lambda^n}{n!}n^q = \lambda \sum_{n=0}^{+\infty} \frac{e^{-\lambda}\lambda^n}{n!}(n+1)^{q-1} \sim \lambda$$

Theorem. *If*

$$[\theta^k(t_1), .., \theta^k(t_p)] \xrightarrow{d} [\theta(t_1), .., \theta(t_p)] \tag{10.1}$$

holds for all $t_1, .., t_p$, *and if f*

$$\lim_{\delta \to 0} \limsup_{k \to \infty} \mathbb{P}\{w(\theta^k, \delta) \geq \varepsilon\} = 0 \tag{10.2}$$

or each positive ε, *then* θ^k *weakly converges to* θ.

We refer the reader to Billingsley (1999) and Davidson (1994) for a proof of this theorem. We succesively prove the convergence of the marginals (10.1) and the tightness condition (10.2).

10.1. Convergence of the Marginals

10.1.1. Discretization of Arrival Times

First, consider a sequence $\{T_i\}_{i=1}^{\infty}$ of iid random variables with exponential density $f(x; \lambda) = \lambda \exp(-\lambda x)$, and arrival times $t_n = \sum_{i=1}^{n} T_i$, which can take any value on the line $[0, \infty)$. Given an integer $c > 1$, we discretize $\{t_n\}$ on the uniform grid $0,\ 1/c^k,\ 2/c^k,\ ..\ , \infty$. For all $x \in \mathbb{R}$, denote by $[x]$ the unique integer such that $[x] \leq x < [x] + 1$. We may discretize the sequence $\{t_n\}$ on the grid by letting $s_0^{(k)} = 0$, and

$$s_n^{(k)} = s_{n-1}^{(k)} + \frac{[c^k(t_n - t_{n-1})] + 1}{c^k}$$

for all $n \geq 1$. Note that

$$s_n^{(k)} = \sum_{j=1}^{n} \frac{[c^k(t_j - t_{j-1})] + 1}{c^k},$$

and thus $t_n \leq s_n^{(k)} \leq t_n + n/c^k$. Observe that $s_n^{(k)} \to t_n$ almost surely as $k \to \infty$, i.e. as the grid becomes infinitely fine.

Additionally, the arrival time $s_1^{(k)}$ has a geometric distribution defined on the grid. More specifically for all $n \geq 0$ and $i \geq 1$,

$$\begin{aligned} P\left\{s_{n+1}^{(k)} - s_n^{(k)} = i/c^k\right\} &= P\left\{\frac{i-1}{c^k} \leq t_{n+1} - t_n < \frac{i}{c^k}\right\} \\ &= \left[1 - \exp(-\frac{\lambda}{c^k})\right] \exp\left[-\frac{(i-1)\lambda}{c^k}\right] \\ &= (1 - \gamma_{1,k})^{i-1} \gamma_{1,k} \end{aligned}$$

where $\gamma_{1,k} = 1 - \exp(-\lambda/c^k)$. The random variable $c^k\left(s_{n+1}^{(k)} - s_n^{(k)}\right)$ has thus geometric distribution with parameter $\gamma_{1,k}$.

10.1.2. Construction of Coupled Trading Times

Consider the sequence $\{\mu_k\}_{k=1}^{\infty}$ used in the construction of the continuous time measure define on $[0,1]$. This construction relies on the exponential variables $\{T^{j_1,..,j_k}\}$, the stochastic arrival times $t_{j_1,..,j_k}$ and the multipliers $M_{j_1,..,j_k}$. We note that

$$\mu_k[0,1] = \sum_{j_1=0}^{N} M_{j_1} \sum_{j_2=0}^{N^{j_1}} M_{j_1,j_2} ... \sum_{j_k=0}^{N^{j_1,..,j_{k-1}}} M_{j_1..j_k} \Delta t_{j_1..j_k}$$

where $\Delta t_{j_1..j_k} = (t_{j_1..j_k+1} - t_{j_1..j_k})$.

We can also construct a measure based on discretized arrival times. For a given number $c > 1$, consider the grid $0,\ 1/c^k, .., 1$. In the first stage of the cascade, we consider the arrival times $s_0^{(k)} = 0,\ s_1^{(k)}, .., s_P^{(k)},\ s_{P+1}^{(k)} = 1$. We proceed as in Section 5.1, and thus define the numbers $P^{j_1,..,j_k}$ and the measure μ_k^{**}. Consistent with previous notation,

$$\mu_k^{**}[ic^{-k}, (i+1)c^{-k}] = b^{-k} M_{j_1} ... M_{j_1..j_k}$$

where $[ic^{-k}, (i+1)c^{-k}] \subseteq I_{j_1..j_k}$. We infer that

$$\mu_k^{**}[0,1] = \sum_{j_1=0}^{P} M_{j_1} \sum_{j_2=0}^{P^{j_1}} M_{j_1,j_2} ... \sum_{j_k=0}^{P^{j_1,..,j_{k-1}}} M_{j_1..j_k} \Delta s_{j_1..j_k},$$

where $\Delta s_{j_1..j_k} = s_{j_1..j_k+1} - s_{j_1..j_k}$.

One of the characteristics of this method is that the number of arrivals at each stage may differ in the continuous and discretized measures. It is notationally convenient to extend the definition of $\Delta t_{j_1..j_p}$ and $N^{j_1..j_p}$ by letting $\Delta t_{j_1..j_p} = 0$, $N^{j_1..j_p} = 0$ when these numbers are not already defined. We can similarly extend the definitions of $\Delta s_{j_1..j_p}$ and $P^{j_1..j_p}$. It is then convenient to introduce

$$\Delta_{j_1..j_k} = \Delta s_{j_1..j_k} - \Delta t_{j_1..j_k},$$

which quantifies the difference in length of corresponding intervals. We can also define the integers $L^{j_1,..,j_p} = \min(P^{j_1,..,j_p}, N^{j_1,..,j_p})$, and $H^{j_1,..,j_p} = \max(P^{j_1,..,j_p}, N^{j_1,..,j_p})$.

With this notation, the masses of the two measures differ by

$$\mu_k^{**}[0,1]-\mu_k[0,1]=\sum_{j_1=0}^{H}M_{j_1}\sum_{j_2=0}^{H^{j_1}}M_{j_1,j_2}\cdots\sum_{j_k=0}^{H^{j_1,\dots,j_{k-1}}}M_{j_1..j_k}\Delta_{j_1..j_k}. \tag{10.3}$$

In the rest of the proof, we show that the first or second moment of $|\mu_k^{**}[0,1]-\mu_k[0,1]|$ converges to zero as $k\to\infty$. This implies that $\mu_k^{**}[0,1]=\mu_k[0,1]+o_p(1)\xrightarrow{d}\mu[0,1]$.

10.1.3. Convergence under Condition 1

Relation (10.3) implies

$$\begin{aligned}
\mathbb{E}\left|\mu_k^{**}[0,1]-\mu_k[0,1]\right| &\le \mathbb{E}\left(\sum_{j_1=0}^{H}M_{j_1}\sum_{j_2=0}^{H^{j_1}}M_{j_1,j_2}\cdots\sum_{j_k=0}^{H^{j_1,\dots,j_{k-1}}}M_{j_1..j_k}\left|\Delta_{j_1..j_k}\right|\right)\\
&= \mathbb{E}\left(\sum_{j_1=0}^{H}\sum_{j_2=0}^{H^{j_1}}\cdots\sum_{j_k=0}^{H^{j_1,\dots,j_{k-1}}}\left|\Delta_{j_1..j_k}\right|\right)\\
&= \mathbb{E}\left(A_k\right)+\mathbb{E}\left(B_{k,1}\right)+..+\mathbb{E}\left(B_{k,k}\right)
\end{aligned}$$

where

$$\begin{aligned}
A_k &= \sum_{j_1=0}^{L}\sum_{j_2=0}^{L^{j_1}}\cdots\sum_{j_k=0}^{L^{j_1,\dots,j_{k-1}}}\left|\Delta_{j_1..j_k}\right|, \quad \text{and}\\
B_{k,p} &= \sum_{j_1=0}^{L}\cdots\sum_{j_{p-1}=0}^{L^{j_1,\dots,j_{p-2}}}\ \sum_{j_p=L^{j_1,\dots,j_{p-1}}+1}^{H^{j_1,\dots,j_{p-1}}}\left|\Delta_{j_1..j_p}\right|, \quad 1\le p\le k.
\end{aligned}$$

The sum A_k quantifies the difference in length between intervals that are used in the kth stage of the construction of both measures. In constrast, $B_{k,p}$ corresponds to the length of intervals that are only used in one construction.

Given $j=(j_1,..,j_p)$, consider the intervals (j,n) in the continuous and discretized constructions. We know that when $n<L^j$, we have $s_{j,n+1}-s_{j,n}=(1+[c^k(t_{j,n+1}-t_{j,n})])/c^k$ and therefore

$$0\le\Delta_{j,n}=(s_{j,n+1}-s_{j,n})-(t_{j,n+1}-t_{j,n})\le 1/c^k. \tag{10.4}$$

For $n\ge L^j$, we can show

Lemma 1. *The inequalities*

$$\left|\Delta_{j,L^j}\right| \le \left|\Delta_j\right| + \frac{1+L^j}{c^k} \tag{10.5}$$

and

$$\left|\Delta_{j,L^j+1}\right| + \ldots + \left|\Delta_{j,H^j}\right| \le \left|\Delta_j\right| + \frac{1+L^j}{c^k} \tag{10.6}$$

hold for all $p \ge 0$, $j = (j_1, .., j_p)$.

Proof. To simplify notation, let $j+1 = (j_1, .., j_{p-1}, j_p + 1)$. We note that

$$\begin{aligned} \Delta_{j,L^j} &= (s_{j,L^j+1} - s_{j,L^j}) - (t_{j,L^j+1} - t_{j,L^j}) \\ &= \Delta'_{j,L^j} + (t_{j,L^j} - t_j) - (s_{j,L^j} - s_j), \end{aligned}$$

where $\Delta'_{j,L^j} = (s_{j,L^j+1} - s_j) - (t_{j,L^j+1} - t_j)$. Since $(t_{j,L^j} - t_j) - (s_{j,L^j} - s_j) \in [-c^{-k}L^j, 0]$, we infer

$$\Delta'_{j,L^j} - c^{-k}L^j \le \Delta_{j,L^j} \le \Delta'_{j,L^j}. \tag{10.7a}$$

We now distinguish three cases depending on the values of the endpoints.

1. If $N^j = P^j = L^j$, we know that $s_{j,L^j+1} = s_{j+1}$, $t_{j,L^j+1} = t_{j+1}$, and therefore $\Delta'_{j,L^j} = \Delta_j$. Relation (10.7a) then implies (10.5).

2. If $N^j > P^j = L^j$, we know that $s_{j,L^j+1} = s_{j+1}$ and therefore

$$\Delta'_{j,L^j} = (s_{j+1} - s_j) - (t_{j,L^j+1} - t_j) \ge \Delta_j.$$

We also know that

$$s_{j+1} \le s_{j,L^j} + \frac{[c^k(t_{j,L^j+1} - t_{j,L^j})] + 1}{c^k} \le s_{j,L^j} + (t_{j,L^j+1} - t_{j,L^j}) + \frac{1}{c^k}, \tag{10.8}$$

which implies $\Delta_{j,L^j} \le c^{-k}$. We now infer from (10.7a) that $\Delta_j - c^{-k}L^j \le \Delta_{j,L^j} \le c^{-k}$, and conclude that (10.5) holds. Moreover, inequality (10.8) implies $-t_{j,L^j+1} \le -t_{j,L^j} + 1/c^k - (s_{j+1} - s_{j,L^j})$, and therefore

$$\begin{aligned} t_{j+1} - t_{j,L^j+1} &\le t_{j+1} - t_{j,L^j} + 1/c^k - (s_{j+1} - s_{j,L^j}) \\ &= [(t_{j+1} - t_j) - (s_{j+1} - s_j)] + 1/c^k + [(s_{j,L^j} - s_j) - (t_{j,L^j} - t_j)] \\ &\le -\Delta_j + (1+L^j)/c^k, \end{aligned}$$

which implies (10.6).

3. If $P^j > N^j = L^j$, we know that $t_{j,L^j+1} = t_{j+1}$, and therefore

$$\Delta'_{j,L^j} = (s_{j,L^j+1} - s_j) - (t_{j+1} - t_j) \leq \Delta_j.$$

We also know that

$$t_{j+1} - t_{j,L^j} \leq T^j_{L^j} < \frac{[c^k T^j_{L^j}] + 1}{c^k} = s_{j,L^j+1} - s_{j,L^j} \tag{10.9}$$

and therefore $\Delta_{j,L^j} > 0$. Relation (10.7a) implies $\left|\Delta_{j,L^j}\right| = \Delta_{j,L^j} \leq \Delta'_{j,L^j} \leq \Delta_j$, and inequality (10.5) holds. Moreover, we infer from (10.9) that

$$\begin{aligned} s_{j+1} - s_{j,L^j+1} &\leq s_{j+1} - s_{j,L^j} - (t_{j+1} - t_{j,L^j}) \\ &= (s_{j+1} - s_j) - (t_{j+1} - t_j) + (t_{j,L^j} - t_j) - (s_{j,L^j} - s_j) \\ &\leq \Delta_j, \end{aligned}$$

which implies (10.6). ■

For every $p \in \{1, .., k\}$, let $u_p = \mathbb{E}\left(\sum_{j_1=0}^{N} \cdots \sum_{j_p=0}^{N^{j_1,\dots,j_{p-1}}} 1\right)$ denote the average number of subintervals in the pth stage of the construction. Simple conditioning implies

$$u_p = \mathbb{E}\left[\sum_{j_1=0}^{N} \cdots \sum_{j_{p-1}=0}^{N^{j_1,\dots,j_{p-2}}} \lambda b^p l(I_{j_1,..,j_{p-1}})\right] = \lambda b^p.$$

It is then useful to introduce $\alpha_p = \mathbb{E}\left(\sum_{j_1=0}^{L} \sum_{j_2=0}^{L^{j_1}} \cdots \sum_{j_p=0}^{L^{j_1,\dots,j_{p-1}}} \left|\Delta_{j_1..j_p}\right|\right)$, $p \geq 1$. This implies

$$\mathbb{E}\left[\sum_{j_1=0}^{N} \cdots \sum_{j_{p-1}=0}^{N^{j_1,\dots,j_{p-2}}} N^{j_1,..,j_{p-1}}\right] = \lambda b^p \tag{10.10}$$

We can show

Lemma 2. *The inequality $\alpha_p \leq \lambda(1+2b)b^p/c^k$ holds for all $p \geq 1$.*

Proof. For any $(j_1, .., j_p)$, inequality (10.4) and Lemma 1 imply

$$\sum_{j_{p+1}=0}^{L^{j_1,\dots,j_p}} \left|\Delta_{j_1..j_{p+1}}\right| \leq \left|\Delta_{j_1..j_p}\right| + \frac{1 + 2L^{j_1,..,j_p}}{c^k}$$

and thus $\alpha_{p+1} \leq \alpha_p + c^{-k}\mathbb{E}\left[\sum_{j_1=0}^{L}\sum_{j_2=0}^{L^{j_1}} \ldots \sum_{j_p=0}^{L^{j_1,\ldots,j_{p-1}}} \left(1+2L^{j_1,\ldots,j_p}\right)\right]$. We infer

$$\alpha_{p+1} \leq \alpha_p + \frac{u_p + 2u_{p+1}}{c^k} = \alpha_p + \lambda(1+2b)\frac{b^p}{c^k}, \tag{10.11}$$

and therefore $\alpha_p \leq \lambda(1+2b)\left(\sum_{i=0}^{p-1} b^p\right)/c^k \leq \lambda(1+2b)b^p/c^k$. ■

The lemma directly implies that

$$\mathbb{E}\, A_k \leq \lambda(1+2b)(b/c)^k \to 0$$

as $k \to \infty$. We also infer from Lemma 1 and Lemma 2 that

$$\begin{aligned} \mathbb{E}\, B_{k,p} &\leq \mathbb{E}\left[\sum_{j_1=0}^{L} \ldots \sum_{j_{p-1}=0}^{L^{j_1,\ldots,j_{p-2}}} \left(\left|\Delta_{j_1..j_{p-1}}\right| + \frac{1+L^{j_1..j_{p-1}}}{c^k}\right)\right] \\ &\leq \alpha_{p-1} + (u_{p-1}+u_p)/c^k \end{aligned}$$

and therefore

$$\mathbb{E}\, B_{k,p} \leq \frac{\lambda(1+2b)b^{p-1}}{c^k} + \frac{\lambda b^{p-1}(1+b)}{c^k} = \lambda(3b+2)\frac{b^{p-1}}{c^k}.$$

Hence

$$\sum_{p=1}^{k} \mathbb{E}\, B_{k,p} \leq \lambda(3b+2)\left(\frac{b}{c}\right)^k \to 0$$

as $k \to \infty$. This establishes that $\mathbb{E}\,|\mu_k^{**}[0,1] - \mu_k[0,1]|$ converges to zero as $k \to \infty$, and therefore $\mu_k^{**}[0,1] = \mu_k[0,1] + o_p(1) \xrightarrow{d} \mu[0,1]$. A straightforward adaptation of this proof implies that $\mathbb{E}\,|\mu_k^{**}[0,t] - \mu_k[0,t]| \to 0$ for all t, and therefore

$$\begin{aligned} [\theta^k(t_1), \theta^k(t_2), \ldots, \theta^k(t_p)] &= (\mu_k[0,t_1], \ldots, \mu_k[0,t_p]) + o_p(1) \\ &\xrightarrow{d} [\theta(t_1), \theta(t_2), \ldots, \theta(t_p)] \end{aligned}$$

for all $t_1, t_2, \ldots, t_p \in [0,1]$. The convergence condition (10.1) is therefore satisfied.

10.1.4. Convergence under Condition 2

By (10.3), the difference $X_k = \mu_k^{**}[0,1] - \mu_k[0,1]$ satisfies

$$\begin{aligned} X_k^2 &= \sum_{j_1=0}^{H} \sum_{j_2=0}^{H^{j_1}} \cdots \sum_{j_k=0}^{H^{j_1,\ldots,j_{k-1}}} \\ &\left(\sum_{i_1=0}^{H} M_{i_1} M_{j_1} \sum_{i_2=0}^{H^{i_1}} M_{i_1,i_2} M_{j_1,j_2} \cdots \sum_{i_k=0}^{H^{i_1,\ldots,i_{k-1}}} M_{i_1..i_k} M_{j_1..j_k} \Delta_{i_1..i_k} \Delta_{j_1..j_k} \right). \end{aligned}$$

Counting how many multipliers the intervals $(i_1,..,i_k)$ and $(j_1,..,j_k)$ have in common, we can write the second moment of X_k as

$$\mathbb{E}(X_k^2) = \left[\mathbb{E}(M^2)\right]^k \mathbb{E}\left[\sum_{j_1=0}^{H} \sum_{j_2=0}^{H^{j_1}} \cdots \sum_{j_k=0}^{H^{j_1,\ldots,j_{k-1}}} \left(\Delta_{j_1,..,j_{k-1},j_k}\right)^2 \right] + \sum_{p=1}^{k} \left[\mathbb{E}(M^2)\right]^{p-1} \mathbb{E}(C_p)$$

where

$$C_p = \sum_{j_1=0}^{H} \sum_{j_2=0}^{H^{j_1}} \cdots \sum_{j_k=0}^{H^{j_1,\ldots,j_{k-1}}} \left(\sum_{i_p \neq j_p}^{H^{j_1,\ldots,j_{p-1}}} \sum_{i_{p+1}=0}^{H^{j_1,\ldots,j_{p-1},i_p}} \cdots \sum_{i_k=0}^{H^{j_1,\ldots,j_{p-1},i_p,\ldots,i_{k-1}}} \Delta_{i_1..i_k} \Delta_{j_1..j_k} \right)$$

We know that for all $p \geq 0$,

$$\sum_{j_{p+1}=0}^{H^{j_1,\ldots,j_p}} \Delta_{j_1..j_p,j_{p+1}} = \Delta_{j_1..j_p}.$$

This implies

$$\begin{aligned} \sum_{i_2 \neq j_2} \sum_{i_3=0}^{H^{j_1,i_2}} \cdots \sum_{i_k=0}^{H^{i_1,\ldots,i_{k-1}}} \Delta_{j_1,i_2..i_k} &= \Delta_{j_1} - \sum_{i_3=0}^{H^{j_1,j_2}} \cdots \sum_{i_k=0}^{H^{i_1,\ldots,i_{k-1}}} \Delta_{j_1,j_2,i_3,..,i_k} \\ &= \Delta_{j_1} - \Delta_{j_1,j_2} \end{aligned}$$

and therefore

$$\mathbb{E}(C_2) = \mathbb{E}\left[\sum_{j_1=0}^{H} \sum_{j_2=0}^{H^{j_1}} \cdots \sum_{j_k=0}^{H^{j_1,\ldots,j_{k-1}}} (\Delta_{j_1} - \Delta_{j_1,j_2}) \Delta_{j_1,j_2..j_k} \right]$$

$$
\begin{aligned}
&= \mathbb{E}\left[\sum_{j_1=0}^{H}\sum_{j_2=0}^{H^{j_1}}(\Delta_{j_1}-\Delta_{j_1,j_2})\Delta_{j_1,j_2}\right] \\
&= \mathbb{E}\left[\sum_{j_1=0}^{H}(\Delta_{j_1})^2\right] - \mathbb{E}\left[\sum_{j_1=0}^{H}\sum_{j_2=0}^{H^{j_1}}(\Delta_{j_1,j_2})^2\right].
\end{aligned}
$$

Let

$$
D_p = \sum_{j_1=0}^{H}\sum_{j_2=0}^{H^{j_1}} \cdots \sum_{j_p=0}^{H^{j_1,\dots,j_{p-1}}} (\Delta_{j_1,\dots,j_p})^2, \ 1 \leq p \leq k.
$$

With this notation, we can rewrite $\mathbb{E}(C_2) = \mathbb{E}(D_1) - \mathbb{E}(D_2)$. More generally, it is easy to show that $\mathbb{E}(C_1) = -\mathbb{E}(D_1)$ and $\mathbb{E}(C_p) = \mathbb{E}(D_{p-1}) - \mathbb{E}(D_p)$ for all $p \geq 2$. This implies

$$
\frac{\mathbb{E}(X_k^2)}{\mathbb{E}(M^2)-1} = \sum_{p=1}^{k}\left[\mathbb{E}(M^2)\right]^{p-1}\mathbb{E}(D_p),
$$

which allows us to bound the second moment.

We heuristically expect that $[\mathbb{E}(M^2)]^p\,\mathbb{E}(D_p) \sim [\mathbb{E}(M^2)]^p\, b^p(c^{-k})^2$, and therefore that the second moment

$$
\mathbb{E}(X_k^2) \sim \sum_{k=1}^{p}\left[\mathbb{E}(M^2)\right]^p b^p(c^{-k})^2 \sim \left[\mathbb{E}(M^2)\right]^k b^k(c^{-2})^k.
$$

converges to zero if $\mathbb{E}(M^2)bc^{-2} < 1$. In order to rigorously prove this, it is useful to consider the decomposition

$$
D_p = F_p + G_1 + \dots + G_p,
$$

where

$$
\begin{aligned}
F_p &= \sum_{j_1=0}^{L}\sum_{j_2=0}^{L^{j_1}} \cdots \sum_{j_p=0}^{L^{j_1,\dots,j_{p-1}}} (\Delta_{j_1..j_p})^2, \quad \text{and} \\
G_q &= \sum_{j_1=0}^{L} \cdots \sum_{j_{q-1}=0}^{L^{j_1,\dots,j_{q-2}}} \sum_{j_q=L^{j_1,\dots,j_{q-1}}+1}^{H^{j_1,\dots,j_{q-1}}} (\Delta_{j_1..j_q})^2, \quad 1 \leq q \leq k.
\end{aligned}
$$

Given $j = (j_1, .., j_{p-1})$, Lemma 1 implies

$$\begin{aligned}\sum_{j_p=0}^{L^j} (\Delta_{j,j_p})^2 &= \sum_{j_p=0}^{L^j-1} (\Delta_{j,j_p})^2 + (\Delta_{j,L^j})^2 \\ &\leq L^j \left(\frac{1}{c^k}\right)^2 + \left(|\Delta_j| + \frac{1+L^j}{c^k}\right)^2 \\ &\leq \frac{L^j}{c^{2k}} + 2\left[|\Delta_j|^2 + \frac{(1+L^j)^2}{c^{2k}}\right].\end{aligned}$$

We infer that

$$F_p \leq \sum_{j_1=0}^{L} \sum_{j_2=0}^{L^{j_1}} \cdots \sum_{j_{p-1}=0}^{L^{j_1,\cdots,j_{p-2}}} \left\{ \frac{L^{j_1,..,j_{p-1}}}{c^{2k}} + 2\left[\left|\Delta_{j_1,..,j_{p-1}}\right|^2 + \frac{(1+L^{j_1,..,j_{p-1}})^2}{c^{2k}}\right]\right\}$$

and thus

$$\mathbb{E}(F_p) \leq 2\mathbb{E}(F_{p-1}) + c^{-2k}\gamma_p', \tag{10.12}$$

where $\gamma_p' = \mathbb{E}\left\{\sum_{j_1=0}^{N} \sum_{j_2=0}^{N^{j_1}} \cdots \sum_{j_{p-1}=0}^{N^{j_1,\cdots,j_{p-2}}} [N^{j_1,..,j_{p-1}} + 2(1+N^{j_1,..,j_{p-1}})^2]\right\}$. Relation (10.10) implies

$$\begin{aligned}\gamma_p' &= 5\lambda b^p + 2\lambda b^{p-1} + 2\mathbb{E}\left[\sum_{j_1=0}^{N} \sum_{j_2=0}^{N^{j_1}} \cdots \sum_{j_{p-1}=0}^{N^{j_1,\cdots,j_{p-2}}} (N^{j_1,..,j_{p-1}})^2\right] \\ &\leq 7\lambda b^p + 2\mathbb{E}\left\{\sum_{j_1=0}^{N} \sum_{j_2=0}^{N^{j_1}} \cdots \sum_{j_{p-1}=0}^{N^{j_1,\cdots,j_{p-2}}} \lambda b^p l(I_{j_1,..,j_{p-1}}) + \lambda^2 b^{2p} \left[l(I_{j_1,..,j_{p-1}})\right]^2\right\} \\ &= 9\lambda b^p + 2\lambda^2 b^{2p}\mathbb{E}\left\{\sum_{j_1=0}^{N} \sum_{j_2=0}^{N^{j_1}} \cdots \sum_{j_{p-1}=0}^{N^{j_1,\cdots,j_{p-2}}} \left[l(I_{j_1,..,j_{p-1}})\right]^2\right\}.\end{aligned}$$

We must now use

Lemma 3. *Consider a Poisson process with frequency* λ, *which partitions the interval* $[0,t]$ *into* $N+1$ *intervals:* $I_0 = [0,t_1]$, $I_1 = [t_1,t_2], ..., I_N = [t_N,t]$. *Then* $\mathbb{E}\left[\sum_{n=0}^{N} l(I_n)^2\right] \leq 2t/\lambda$.

Proof. See Section 10.1.5. ■

The Lemma implies that

$$\gamma'_p \leq 9\lambda b^p + 2\lambda^2 b^{2p}\mathbb{E}\left\{\sum_{j_1=0}^{N}\sum_{j_2=0}^{N^{j_1}}\cdots\sum_{j_{p-2}=0}^{N^{j_1,\cdots,j_{p-3}}} 2\frac{l(I_{j_1,..,j_{p-2}})}{\lambda b^{p-1}}\right\} \leq 13\lambda b^{p+1}.$$

We then infer from (10.12) that

$$\mathbb{E}(F_p) \leq c^{-2k}\left(\gamma'_p + 2\gamma'_{p-1} + \ldots + 2^{p-1}\gamma'_1\right) \leq \frac{13\lambda b}{c^{2k}}\sum_{i=1}^{p} 2^{p-i}b^i$$

or equivalently

$$\mathbb{E}(F_p) \leq \frac{13\lambda b 2^p}{c^{2k}}\sum_{i=1}^{p}(b/2)^i.$$

Thus $\mathbb{E}(F_p) \leq 13\lambda p 2^{p+1}/c^{2k}$ when $b = 2$, and $\mathbb{E}(F_p) \leq 13\lambda b^{p+2}/c^{2k}$ when $b > 2$. In both cases, the inequality $\mathbb{E}(F_p) \leq 13\lambda p b^{p+2}/c^{2k}$ thus holds. Similarly, we observe that

$$\begin{aligned} G_q &\leq \sum_{j_1=0}^{L}\cdots\sum_{j_{q-1}=0}^{L^{j_1,\cdots,j_{q-2}}}\left(\sum_{j_q=L^{j_1,\cdots,j_{q-1}}+1}^{H^{j_1,\cdots,j_{q-1}}}\left|\Delta_{j_1..j_q}\right|\right)^2 \\ &\leq \sum_{j_1=0}^{L}\cdots\sum_{j_{q-1}=0}^{L^{j_1,\cdots,j_{q-2}}}\left(|\Delta_j| + \frac{1+L^j}{c^k}\right)^2 \end{aligned}$$

and therefore $\mathbb{E}(G_q) \leq 13\lambda q b^{q+2}/c^{2k}$ for all $q \geq 1$.

We now infer that

$$\mathbb{E}(D_p) \leq 13\lambda c^{-2k}\left(pb^{p+2} + \sum_{q=1}^{p} q b^{q+2}\right) = 26\lambda b^2 c^{-2k}\sum_{q=1}^{p} q b^q$$

and therefore[11] $\mathbb{E}(D_p) \leq 26\lambda b^2 c^{-2k}(p+1)b^{p+2}$. This implies

$$\frac{\mathbb{E}(X_k^2)}{\mathbb{E}(M^2)-1} \leq \frac{26\lambda b^2}{c^{2k}}\sum_{p=1}^{k}(p+1)b^{p+2}\left[\mathbb{E}(M^2)\right]^{p-1}$$

[11] We recall that

$$\sum_{q=1}^{p} q z^q = \frac{pz^{p+2} - (p+1)z^{p+1} + z}{(z-1)^2}$$

which can be obtained by taking the derivative of $f(z) = \sum_{q=1}^{p} z^q = (z^{p+1} - z)/(z-1)$. This implies $\sum_{q=1}^{p} q z^q \leq (p+1)z^{p+2}$ for any $z \geq 2$.

$$\leq \quad \frac{26\lambda b^2}{c^{2k}} \sum_{p=1}^{k+2} p \left[b\mathbb{E}(M^2)\right]^p$$

Since $\sum_{p=1}^{k+2} p \left[b\mathbb{E}(M^2)\right]^p \sim (k+2)[b\mathbb{E}(M^2)]^{k+4}/[b\mathbb{E}(M^2)-1]^2$, we infer that

$$c^{-2k} \sum_{p=1}^{k+2} p \left[b\mathbb{E}(M^2)\right]^p \sim \frac{[b\mathbb{E}(M^2)]^4}{[b\mathbb{E}(M^2)-1]^2}(k+2)\left[\frac{b\mathbb{E}(M^2)}{c^2}\right]^k$$

and $\mathbb{E}(X_k^2)$ converge to zero whenever $b\mathbb{E}(M^2)/c^2 < 1$. As in Section 10.1.3, we then conclude that condition (10.1) therefore holds.

10.1.5. Proof of Lemma 3

A heuristic argument is that $\mathbb{E}\left[\sum_{n=0}^{N} l(I_n)^2\right] \sim (\lambda t)\frac{1}{\lambda^2} = \frac{t}{\lambda}$. More concretely, we can write $l(I_n)^2 = T_n^2 1_{\{n<N_t\}} + \left(t - \sum_{i=1}^n T_i\right)^2 1_{\{n=N_t\}}$, and therefore

$$\mathbb{E}\left[\sum_{n=0}^{N} l(I_n)^2\right] = \mathbb{E}\left[\sum_{n=0}^{+\infty} T_n^2 1_{\{n<N_t\}}\right] + \mathbb{E}\left[\sum_{n=0}^{+\infty}\left(t-\sum_{i=1}^{n} T_i\right)^2 1_{\{n=N_t\}}\right].$$

We separately analyze these two series.

We first note that

$$\begin{aligned}\mathbb{E}\left(T_n^2 1_{\{n<N_t\}}\right) &= \mathbb{E}\left(T_n^2 \,\middle|\, \sum_{i=1}^{n} T_i < t\right) P(N_t > n) \\ &\leq \mathbb{E}\left(T_1^2\right)\mathbb{P}(N_t > n),\end{aligned}$$

and thus $\mathbb{E}\left[\sum_{n=0}^{+\infty} T_n^2 1_{\{n<N_t\}}\right] \leq \mathbb{E}(T_1^2)\sum_{n=0}^{+\infty}\mathbb{P}(N_t > n) = \mathbb{E}(T_1^2)\mathbb{E}(N_t)$, implying

$$\mathbb{E}\left[\sum_{n=0}^{+\infty} T_n^2 1_{\{n<N_t\}}\right] \leq \frac{1}{\lambda^2}(\lambda t) = \frac{t}{\lambda}.$$

Second, we define $U_n = \sum_{i=1}^{n} T_i$ and consider

$$\mathbb{E}\left[\sum_{n=0}^{+\infty}\left(t-\sum_{i=1}^{n} T_i\right)^2 1_{\{n=N_t\}}\right] = \sum_{n=0}^{+\infty} P(N_t = n)\mathbb{E}\left[\left.(t-U_n)^2\right| N_t = n\right].$$

The random variables T_{n+1} and $U_n$are independent and have respective densities $f(u) = \lambda \exp(-\lambda u)$ and $f_n(u) = \lambda \exp(-\lambda u)(\lambda u)^{n-1}/(n-1)!$. When $N_t = n$, the vector (U_n, T_{n+1}) has conditional density

$$f_n(u)f(v)1_{\{0\leq u<t;u+t>v\}}/\mathbb{P}(N_t = n).$$

The marginal density of U_n is therefore

$$\begin{aligned}\frac{f_n(u)}{\mathbb{P}(N_t = n)}\int_{t-u}^{+\infty} f(v)dv &= \frac{\lambda\exp(-\lambda u)(\lambda u)^{n-1}\exp[-\lambda(t-u)]}{(n-1)!\mathbb{P}(N_t = n)}1_{\{u<t\}} \\ &= \frac{\lambda^n\exp(-\lambda t)u^{n-1}}{(n-1)!\mathbb{P}(N_t = n)}1_{\{u<t\}}.\end{aligned}$$

We thus infer that

$$\mathbb{E}\left[\,(t-U_n)^2\middle|\,N_t = n\right] = \frac{\lambda^n\exp(-\lambda t)}{(n-1)!\mathbb{P}(N_t = n)}\int_0^t (t-u)^2u^{n-1}du$$

or equivalently

$$\begin{aligned}\mathbb{P}(N_t &= n)\mathbb{E}\left[\,(t-U_n)^2\middle|\,N_t = n\right] = \frac{\lambda^n\exp(-\lambda t)}{(n-1)!}\frac{2t^{n+2}}{n(n+1)(n+2)} \\ &= \frac{2\exp(-\lambda t)}{\lambda^2}\frac{(\lambda t)^{n+2}}{(n+2)!}.\end{aligned}$$

Therefore

$$\begin{aligned}\mathbb{E}\left[\sum_{n=0}^{+\infty}\left(t-\sum_{i=1}^{n}T_i\right)^2 1_{\{n=N_t\}}\right] &= \frac{2\exp(-\lambda t)}{\lambda^2}\sum_{n=0}^{+\infty}\frac{(\lambda t)^{n+2}}{(n+2)!} \\ &= \frac{2\exp(-\lambda t)}{\lambda^2}\left[\exp(\lambda t) - 1 - \lambda t\right] \\ &\leq \frac{\lambda t}{\lambda^2} = \frac{t}{\lambda}.\end{aligned}$$

We thus conclude that $\mathbb{E}\left[\sum_{n=0}^{N} l(I_n)^2\right] \leq 2t/\lambda$.

10.2. Tightness

Since the function $\limsup_{k\to\infty} P\{w(\theta^k,\delta)\geq\varepsilon\}$ is increasing in δ, we can restrict our attention to step sizes of the form $\delta = c^{-l}$, $l = 1, 2, .., \infty$. For a given $\delta = c^{-l}$, we consider the regularly spaced grid $t_0 = 0 < t_1 = \delta < ... < t_n = 1$, where $n = 1/\delta$. For any $k \geq 1$, we know[12] that

$$\begin{aligned}
\mathbb{P}\{w(\theta^k,\delta) &\geq \varepsilon\} = \mathbb{P}\left\{\sup_{|t-s|<\delta}\left|\theta^k(t)-\theta^k(s)\right|\geq\varepsilon\right\} \\
&\leq \sum_{i=0}^{n-1}\mathbb{P}\left\{\sup_{t_i\leq s\leq t_{i+1}}\left|\theta^k(s)-\theta^k(t_i)\right|\geq\frac{\varepsilon}{3}\right\} \\
&= \sum_{i=0}^{n-1}\mathbb{P}\left\{\theta^k(t_{i+1})-\theta^k(t_i)\geq\frac{\varepsilon}{3}\right\},
\end{aligned}$$

since the process θ_k is increasing. For every $k \geq l$, the trading time increment $\theta^k(t_{i+1}) - \theta^k(t_i)$ is distributed like $\theta^k(\delta)$, and therefore

$$\mathbb{P}\{w(\theta^k,\delta)\geq\varepsilon\}\leq\delta^{-1}\mathbb{P}\left\{\theta^k(\delta)\geq\frac{\varepsilon}{3}\right\}.$$

The convergence of the marginals implies that $\theta^k(\delta)\xrightarrow{d}\theta(\delta)$ and therefore $\mathbb{P}\left\{\theta^k(\delta)\geq\varepsilon/3\right\}\to\mathbb{P}\{\theta(\delta)\geq\varepsilon/3\}$. We now infer that

$$\limsup_{k\to+\infty}\mathbb{P}\{w(\theta^k,\delta)\geq\varepsilon\}\leq\delta^{-1}\mathbb{P}\left\{\theta(\delta)\geq\frac{\varepsilon}{3}\right\}.$$

Given a number $q > 0$ satisfying $\tau(q) > 0$, Chebyshev's inequality implies

$$\limsup_{k\to+\infty}\mathbb{P}\{w(\theta^k,\delta)\geq\varepsilon\}\leq\left(\frac{3}{\varepsilon}\right)^q\delta^{-1}\mathbb{E}[\theta(\delta)^q].$$

Letting $\delta\to 0$, we know that $\delta^{-1}\mathbb{E}[\theta(\delta)^q]\sim c_{\lambda,q}\delta^{\tau(q)}\to 0$ and conclude that condition (10.2) is satisfied.

[12] See Theorem 7.4 in Billingsley (1999).

References

[1] Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (1999), The Distribution of Exchange Rate Volatility, Working Paper, University of Pennsylvania

[2] Baillie, R. T. (1996), Long Memory Processes and Fractional Integration in Econometrics, *Journal of Econometrics* **73**, 5-59

[3] Billingsley, P. (1999), *Convergence of Probability Measures*, 2nd ed., New York: John Wiley and Sons

[4] Black, F., and Scholes, M., The Pricing of Options and Corporate Liabilities, *Journal of Political Economy* **81**, 637-654

[5] Bollerslev, T., and Mikkelsen, H. O. (1996), Modeling and Pricing Long Memory in Stock Market Volatility, *Journal of Econometrics* **73**, 151-184

[6] Calvet, L., Fisher, A., and Mandelbrot, B. (1997), Large Deviation Theory and the Distribution of Price Changes, *Cowles Foundation Discussion Paper #1165*

[7] Calvet, L., Fisher, A., and Mandelbrot, B. (1999), A Multifractal Model of Asset Returns, Working Paper

[8] Davidson, J. (1994), *Stochastic Limit Theory*, Oxford University Press

[9] Diebold, F. X., Gunther, T. A., and Tay, A. S. (1998), Evaluating Density Forecasts, with Applications to Financial Risk Management, *International Economic Review* **39**, 863-883

[10] Fisher, A., Calvet, L., and Mandelbrot, B. (1997), Multifractality of Deutschemark/US Dollar Exchange Rate, *Cowles Foundation Discussion Paper #1166*

[11] Garman, M., and Kohlhagen, S. (1983), Foreign Currency Option Values, *Journal of International Money and Finance* **2**, 231-237

[12] Hull, J., and White, A. (1987), The Pricing of Options on Assets with Stochastic Volatilities, *Journal of Finance* **42**, 281-300

[13] Kurz, M. (1994), On Rational Belief Equilibria, *Economic Theory* **4**, 859-876

[14] Kurz, M. (1997) Ed., *Endogenous Economic Fluctuations: Studies in the Theory of Rational Beliefs*, Springer Verlag

[15] Mandelbrot, B., Fisher, A., and Calvet, L. (1997), A Multifractal Model of Asset Returns, *Cowles Foundation Discussion Paper #1164*

[16] Pollard, D. (1984), *Convergence of Stochastic Processes*, Springer-Verlag

Fig 1. Conditional Probabilities of States over Time

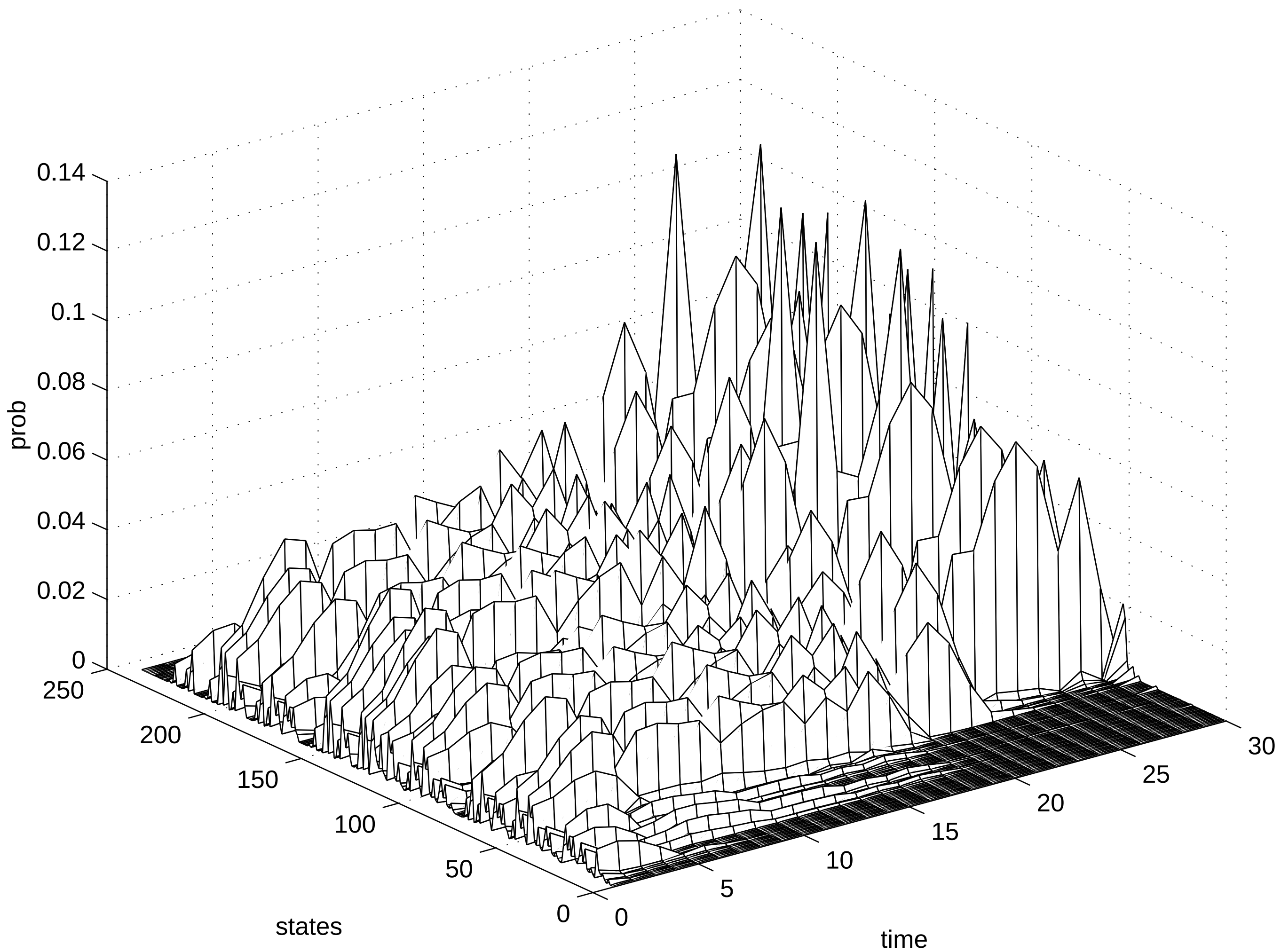


Fig 2. Smoothed Forecasts of Average Trading Time

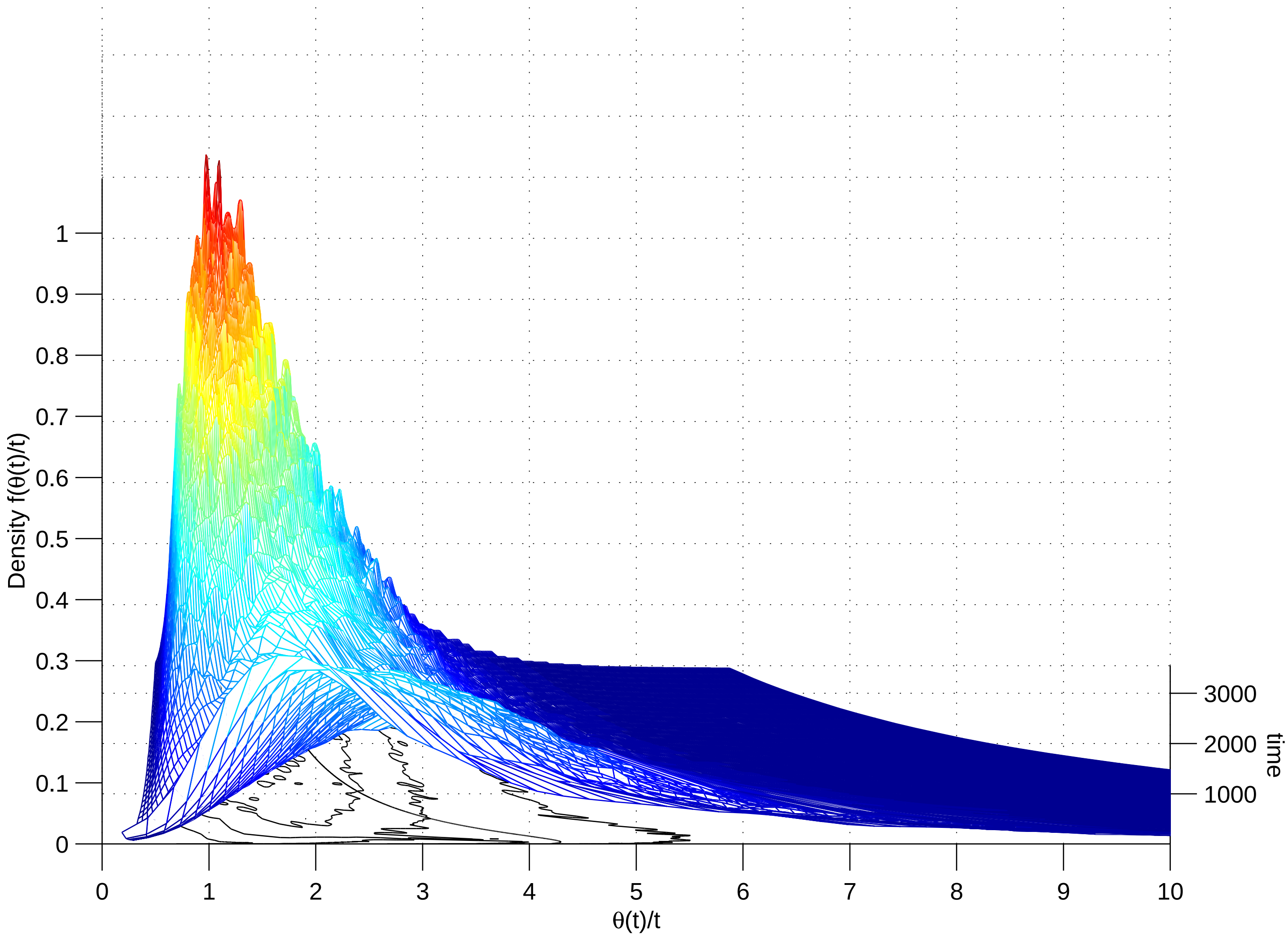
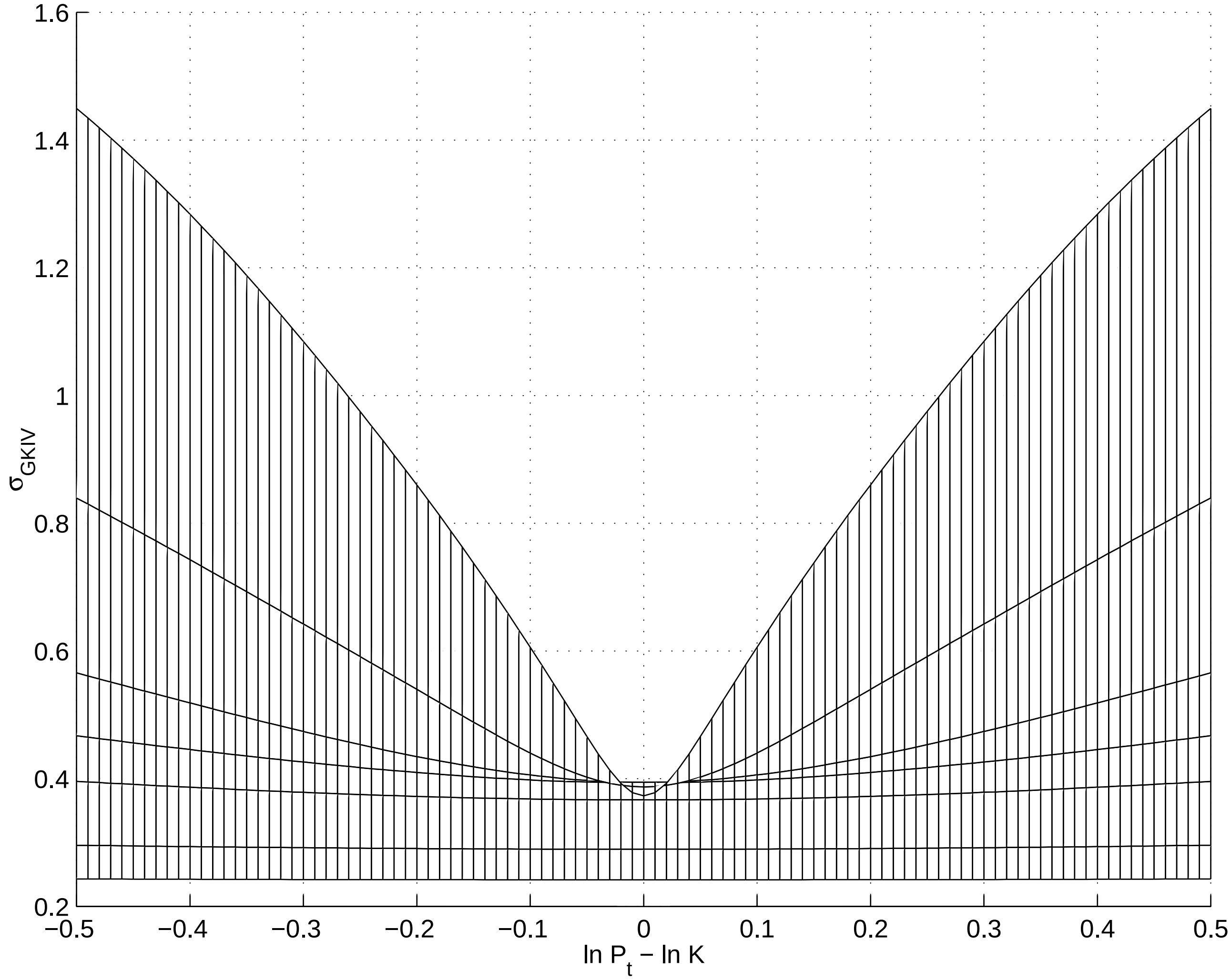


Fig 3. Implied Volatilities from Garman−Kohlhagen Currency Option Pricing Formula



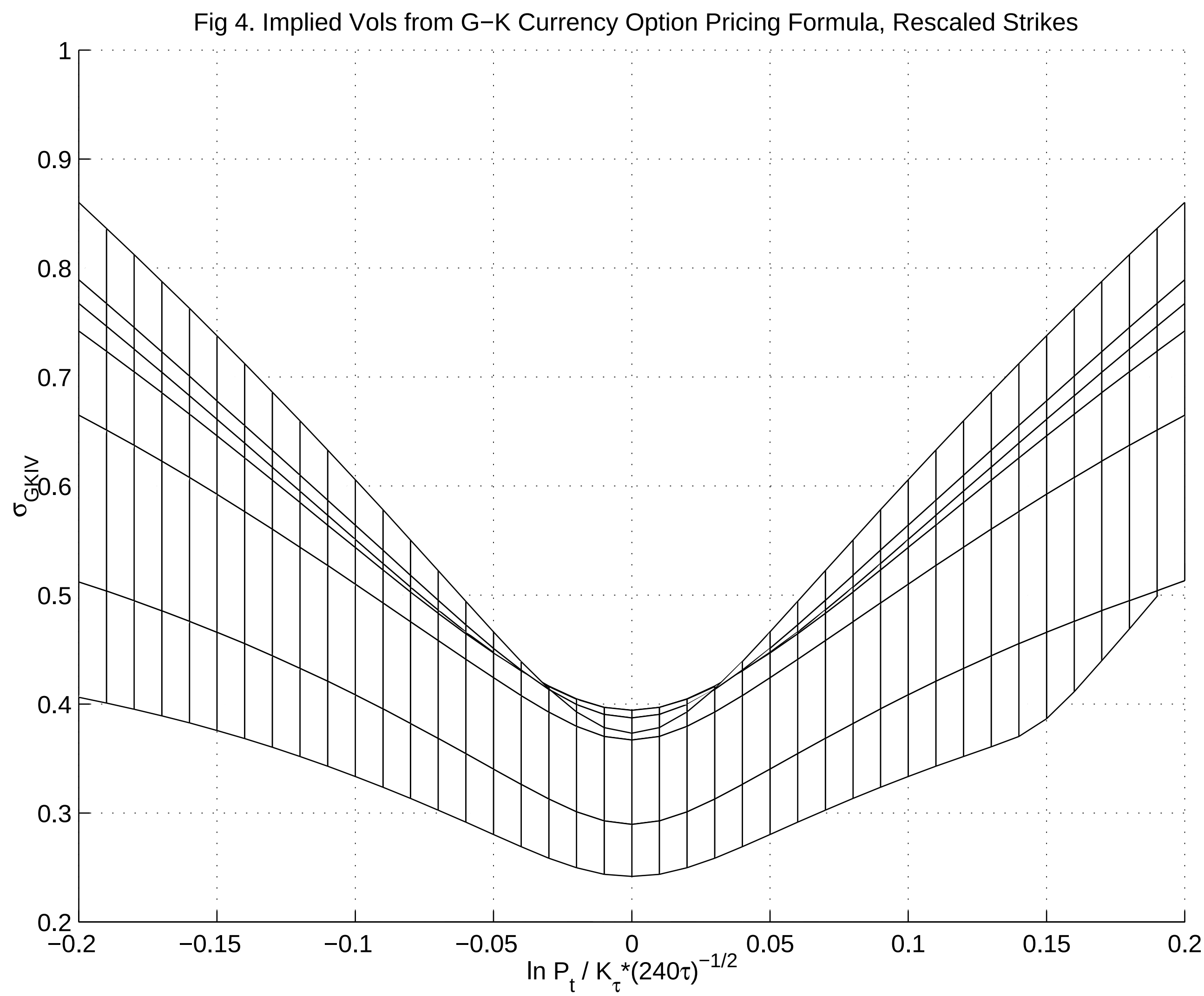


Fig 4. Implied Vols from G−K Currency Option Pricing Formula, Rescaled Strikes
σGKIV
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
−0.2
−0.15
−0.1
−0.05
0
0.05
0.1
0.15
0.2
ln Pt / Kτ*(240τ)−1/2