Developing bid-ask probabilities for high-frequency trading

Lester Ingber

Abstract:
Methods of path integrals are used to develop multi-factor probabilities of bid-ask variables for use in high-frequency trading (HFT). Adaptive Simulated Annealing (ASA) is used to fit the nonlinear forms so developed to a day of BitMEX tick data. Maxima algebraic code is used to develop the path integral codes into C codes, and sampling code is used for the fitting process. After these fits, the resultant C code is very fast and useful for forecasting upcoming ask, bid, midprice, etc., when narrow and wide windows of incoming data are used. A bonus is the availability of canonical momenta indicators (CMI) useful to forecast direction and strengths of these variables.

Key words: path integral; financial markets; high-frequency trading

1 Introduction

Although high-frequency trading (HFT) is a relatively new development in financial markets, it has become a primary force in market pricing. A large scientific literature has developed to understand the nature of these forces (Ait-Sahalia & Saglam, 2017; Avellaneda & Stoikov, 2008; Baradely et al., 2018; Cartea et al., 2014; Cont, 2011; Cont et al., 2009; Podra & Labadie, 2012; Gueant, 2017; Huang et al., 2014).

In this context, this paper applies a previously developed statistical mechanics of financial markets (SMFM) (Ingber, 1984, 1990, 1996a,b, 2000, 2010, 2017a; Ingber et al., 2001; Ingber & Mondescu, 2001, 2003; Ingber et al., 1991b; Ingber & Wilson, 1999, 2000), here applied to developing joint bid-ask probabilities to high-frequency data, using two methods of fitting price data or returns data to (a) the distribution and (b) fitting the returns. The returns are also developed into closed-form algebra using the path-integral formalism.

The path-integral algebra behind the algorithms used is introduced in Section 2. This Section also details how three equivalent methods of treating stochastic systems are developed: (a) path integrals, (b) sets of stochastic differential equations, and (c) Fokker-Planck or Chapman-Kolmogorov partial differential equations.

Section 3 describes the author’s numerical optimization algorithm, Adaptive Simulated Annealing (ASA).


As was true for these other disciplines, here too the path-integral methodology is used to develop canonical momenta indicators (CMI), useful to forecast direction and strengths of these variables.

Section 4 describes the development of the forecast code fit to data, using the algebraic language Maxima (maxima.sourceforge.net) to develop C code which is used both to fit data to parameters, as well as to run very fast in real time for HFT. The fitting process uses a sampling code developed for this project, to effectively sample and run a random subset of the data.

1Corresponding author: Lester Ingber
email: lester@ingber.com or ingber@caa.caltech.edu
Additional information may be obtained at https://www.ingber.com.
2 Path Integral

2.1 Path Integral in Stratonovich (Midpoint) Representation

The path integral in the Feynman (midpoint) representation can be developed in time-dependent nonlinear systems (Langouche et al., 1979, 1982; Schulman, 1981). The Einstein summation convention is often used wherein repeated indices signify summation; bars | ... | imply no summation.

\[ P[M_t|M_{t_0}]dM(t) = \int \ldots \int DM \exp \left(-\min_{t_0} \int_{t_0}^{t} dt'L \right) \delta(M(t_0) = M_0)\delta(M(t) = M_t) \]

\[ DM = \lim_{u \to \infty} \prod_{\rho=1}^{u+1} g^{1/2} \prod_{G} (2\pi\theta)^{-1/2}dM^G_{\rho} \]

\[ L(\dot{M}^G, M^G, t) = \frac{1}{2}(\dot{M}^G - h^G)g_{GG'}(\dot{M}^G' - h^G') + \frac{1}{2}h^G_{,G} + R/6 - V \]

\[ M^G(t) \to \frac{1}{2}(M^G_{\rho+1} + M^G_{\rho}) \]

\[ \dot{M}^G(t) \to M^G_{\rho+1} - M^G_{\rho}, [...], G = \frac{\partial[...]}{\partial M^G} \]

\[ h^G = g^G - \frac{1}{2}g^{-1/2}(g^{1/2}g^{GG'}), h^G_{,G} = h^G_{,G} + \Gamma^F_{GG'}h^G = g^{-1/2}(g^{1/2}h^G), g \]

\[ g_{GG'} = (g^{GG'})^{-1}, g = \det(g_{GG'}) \]

\[ \Gamma^F_{JK} \equiv g^{LF}[JK, L] = g^{LF}(g_{JL,K} + g_{KL,J} - g_{JK,L}) \]

\[ R = g^{JL}R_{JL} = g^{JL}g^{JK}R_{FJKL} \]

\[ R_{FJKL} = \frac{1}{2}(g_{FK,JL} - g_{JK,FL} - g_{FL,JK} + g_{JL,FK}) + g_{MN}(\Gamma^M_{FK}\Gamma^N_{JL} - \Gamma^M_{FLL}\Gamma^N_{JK}) \]

A Riemannian-curvature potential \( R/6 \) is induced for dimension > 1. \( N \) is the dimensionality of the space being considered. Boundary conditions may enter as a “potential” \( V \). \( g^\dagger \) in \( DM \) implies a prepoint evaluation.
2.2 Path Integral in Ito (Prepoint) Representation

An Ito prepoint discretization for the same probability distribution $P$ gives a simpler algebraic form than the above midpoint discretization,

$$M^G(t_s) = M^G_{\rho}$$

$$L = \frac{1}{2}(dM^G/dt - g^G)g_{GG'}(dM^{G'}/dt - g^{G'}) - V$$  \hspace{1cm} (2)

This Ito Lagrangian $L$ does not satisfy a variational principle, and often finer meshes are required.

2.3 Three Approaches Are Mathematically Equivalent

Three basic different approaches are mathematically equivalent:

(a) Fokker-Planck/Chapman-Kolmogorov partial-differential equations

(b) Langevin coupled stochastic-differential equations

(c) Lagrangian or Hamiltonian path-integrals

The path-integral approach is useful to define intuitive physical variables from the Lagrangian $L$ in terms of variables $M^G$:

Momента : $\Pi^G = \frac{\partial L}{\partial (\partial M^G/\partial t)}$

Mass : $g_{GG'} = \frac{\partial L}{\partial (\partial M^G/\partial t)\partial (\partial M^{G'}/\partial t)}$

Force : $\frac{\partial L}{\partial M^G}$

$$F = ma : \delta L = 0 = \frac{\partial L}{\partial M^G} - \frac{\partial}{\partial t} \frac{\partial L}{\partial (\partial M^G/\partial t)}$$  \hspace{1cm} (3)

The Momenta are used here as canonical momenta indicators (CMI).

Differentiation especially of noisy systems introduces more noise, and the path-integral often gives superior numerical performance because integration is a smoothing process.

2.3.1 Stochastic Differential Equation (SDE)

The Stratonovich midpoint discretized Langevin equations are given in terms of the Wiener process $dW^i$, or equivalently Gaussian noise $\eta^i = dW^i/dt$.

$$dM^G = f^G(t, M(t)) dt + \dot{g}_i^G(t, M(t)) dW^i$$

$$\dot{M}^G(t) = f^G(t, M(t)) + \dot{g}_i^G(t, M(t))\eta^i(t)$$

$$dW^i \rightarrow \eta^i dt$$

$$M = \{M^G; G = 1, \ldots, \Lambda\}$$
\[ \eta = \{ \eta^i; i = 1, \ldots, N \} \]

\[ \dot{M}^G = dM^G / dt \]

\[ < \eta^j(t) >_{\eta} = 0, < \eta^j(t), \eta^j(t') >_{\eta} = \delta_{jj} \delta(t - t') \]  \hspace{1cm} (4)

\( \eta^i \) represents Gaussian white noise.

As used here, this Langevin representation of a set of stochastic differential equations (SDE) is a convenient starting point to define the \( g \)-moments of the distribution.

2.3.2 Partial Differential Equation (PDE)

The Fokker-Planck or Chapman-Kolmogorov partial differential equation is:

\[ P_{,t} = \frac{1}{2}(g^{GG} P)_{,G} - (g^{G} P)_{,G} + VP \]

\[ P = < P_{,\eta} >_{\eta} \]

\[ g^{G} = f^{G} + \frac{1}{2} \hat{g}^{iG} \hat{g}_{i,G} \]

\[ g^{GG'} = \hat{g}^{i} \hat{g}_{i,G} \]

\[ (...)_{,G} = \partial(...) / \partial M^G \]  \hspace{1cm} (5)

\( g^{G} \) replaces \( f^{G} \) in the SDE if the Ito (prepoint discretized) calculus is used.

2.4 PATHINT Applications

Path integrals and PATHINT have been applied across several disciplines, including combat simulations (Ingber et al., 1991a), neuroscience (Ingber, 1994, 2017b; Ingber & Nunez, 1995, 2010), finance (Ingber, 2000, 2016, 2017a,b,c; Ingber et al., 2001; Ingber & Wilson, 2000), and other nonlinear systems (Ingber, 1995, 1998b; Ingber et al., 1996).

2.5 PATHINT/qPATHINT Code

qPATHINT is an N-dimensional code developed to calculate the propagation of quantum variables in the presence of shocks. Many systems propagate in the presence of sudden changes of state dependent on time. qPATHINT is based on the classical-physics code, PATHINT. Applications have been made to SMNI and Statistical Mechanics of Financial Markets (SMFM) (Ingber, 2017a,b,c).

The PATHINT C code of about 7500 lines of code using the GCC C-compiler was rewritten to use double complex variables instead of double variables, and further developed for arbitrary N dimensions, creating qPATHINT. The outline of the code is described here for classical or quantum systems, using generic coordinates \( q \) (Ingber, 2016, 2017a,c).

The distribution (probabilities for classical systems, wave-functions for quantum systems) is numerically approximated to a high degree of accuracy using a histogram procedure, developing sums of rectangles of height \( P_i \) and width \( \Delta q^i \) at points \( q^i \).
2.5.1 PATHINT/qPATHINT Histograms

A one-dimensional path-integral in variable \( q \) in the prepoint Ito discretization is developed in terms of the kernel/propagator \( G \), for each of its intermediate integrals, as

\[
P(q; t + \Delta t) = \int dq'[g^{1/2}(2\pi \Delta t)^{-1/2}] \exp(-L \Delta t) P(q'; t) = \int dq' G(q, q'; \Delta t) P(q'; t)
\]

\[
P(q; t) = \sum_{i=1}^{N} \pi(q - q^i) P_i(t)
\]

\[
\pi(q - q^i) = 1, \ (q^i - \frac{1}{2} \Delta q^{i-1}) \leq q \leq (q^i + \frac{1}{2} \Delta q^i); 0, \text{ otherwise} \tag{6}
\]

This yields

\[
P_i(t + \Delta t) = T_{ij}(\Delta t) P_j(t)
\]

\[
T_{ij}(\Delta t) = \frac{2}{\Delta q^{i-1} + \Delta q^i} \int_{q^i - \Delta q^{i-1}/2}^{q^i + \Delta q^i/2} dq \int_{q^j - \Delta q^{j-1}/2}^{q^j + \Delta q^j/2} dq' G(q, q'; \Delta t) \tag{7}
\]

\( T_{ij} \) is a banded matrix representing the Gaussian nature of the short-time probability centered about the drift.

2.5.2 Meshes For [q]PATHINT

Explicit dependence of \( L \) on time \( t \) can be included. The mesh \( \Delta q^i \) is strongly dependent on diagonal elements of the diffusion matrix, e.g.,

\[
\Delta q^i \approx (\Delta t |g_{ii}|)^{1/2} \tag{8}
\]

By considering the contributions to the first and second moments, conditions on the time and variable meshes can be derived. \( \Delta t \) can be measured by the diffusion divided by the square of the drift.

These mesh considerations constrain the forecast time \( dt \) if consistency with the Action \( A \) is required, where

\[
A = L dt \tag{9}
\]

3 Adaptive Simulated Annealing (ASA) Algorithm

3.1 Importance Sampling

Nonlinear and/or stochastic systems often require importance-sampling algorithms to scan or to fit parameters. Methods of simulated annealing (SA) are often used.

The ASA code is open-source software, and can be downloaded and used without any cost or registration at https://www.ingber.com/#ASA [Ingber, 1993b, 2012c].

This algorithm fits empirical data to a cost function over a \( D \)-dimensional parameter space, adapting for varying sensitivities of parameters during the fit [Ingber, 1989].
3.2 Outline of ASA Algorithm

For parameters

\[ \alpha^i_k \in [A_i, B_i] \]

sampling with the random variable \( x^i \)

\[ x^i \in [-1, 1] \]

\[ \alpha^i_{k+1} = \alpha^i_k + x^i(B_i - A_i) \]

the default generating function is

\[ g_T(x) = \prod_{i=1}^{D} \frac{1}{2 \ln(1 + 1/T_i)(|x^i| + T_i)} \equiv \prod_{i=1}^{D} g^i_T(x^i) \]

in terms of parameter “temperatures”

\[ T_i = T_{i0} \exp(-c_i k^{1/D}) \] (10)

This ASA algorithm is faster than fast Cauchy annealing, which has schedule \( T_i = T_0/k \), and much faster than Boltzmann annealing, which has schedule \( T_i = T_0/\ln k \).

3.3 ASA Applications

The ASA code (Ingber, 1993b) and the original Very Fast Simulated Reannealing (VFSR) code (Ingber, 1989) have been used by many researchers, including the author in a range of disciplines, as referenced in the ASA-README.pdf that comes with the code, or in other papers in the ingber.com archive.

4 Forecast Code

Tick data was cleaned on-the-fly in C code, since this is a practical real-time consideration. The data was volume-weighted, and level-weighted by 0.9, using 20 levels each of bid and ask; the last level was weighted by 0.9^{20} = 0.122.

It is clear that in many HFT markets such as the one used for this study, competitive traders’ tactics, that “game” the book of bids and asks, cause the volume at different levels to be much more volatile than the prices per se.

At first, both prices \( S \) and returns \( R \) of bids and asks were fit. The returns are simply

\[ R(t) = S(t)/S(t-1) \] (11)

Since the trends are the first entities to be swallowed by such markets, making them useless for trading, the ASA fits had to work longer for fits with prices than with returns, so returns were used for most of the runs.
4.1 Analytic Returns

The path integral formalism permits an analytic calculation of the most probable state (Dekker, 1980),

\[ \frac{dM^G}{dt} = g^G - g^{1/2}(g^{1/2}g^{GG'})G' \]  \hspace{0.5cm} (12)

4.2 MAXIMA to C Code

The 2-factor functional forms used, with ask as \(x^1\) and bid as \(x^2\), are indexed here by \(j\), without the Einstein convention,

\[ g^j = A_j + B_j x^j \]
\[ \hat{g}^j = C_j + D_j x^j \]
\[ \hat{g}_1^1 = E_1, \hat{g}_2^1 = E_2 \]  \hspace{0.5cm} (13)

where the parameters \(\{A_i, B_i, C_i, D_i, E_i\}\) are fit to the data. Note that these simple functional forms, especially after inverses of the covariance matrix are taken, form Padé approximates, ratios of polynomials, known to be quite robust functional forms to fit data.

The cost function used is the effective Action,

\[ A_{eff} = Ldt + N \log(2\pi dt)/2 + \log(g)/2 \]  \hspace{0.5cm} (14)

\(A_{eff}\) the analytic expression above for returns can be calculated using the algebraic language Maxima. Since such expressions can be about 1000 characters (without spaces), this is advisable.

It is straightforward to convert Maxima output to Fortran code, but not so easy to covert to C. Previous projects have used f2c from netlib.org to compile mixed C and Fortran code, but here the main issue is that C requires powers to use the prefix operator-function pow(), whereas Maxima uses postfix operators. This was circumvented simply by writing all powers as multiple factors with different names, converting to Fortran, then renaming these the same after the calculations.

4.3 Dynamic Memory

Arrays for books of data over a day or more require 10’s of megabytes of memory. Static arrays that size crashes C code, so dynamic memory using calloc() was used.

4.4 Sampling Code

It is not necessary to use all tick data to get very good fits. Instead, a modification of ASA subroutines that calculate random states was used to randomly sample the data that is already volume-weighted and level-weighted. It was found that a sample size of 100K gave as good results as a sample size of several million for a day of BitMEX data. This random sample then was used for the ASA cost function calculations that ran over the sample for each generated state.
4.5 Windows of Data

The calculated distribution at any point is dependent on at least 2 points of weighted bid-ask data, and the calculation takes into account correlations and dynamics inherent in the functional form of the drifts and diffusions.

Therefore, the best use of this approach is to use narrow and wide windows for forecasting (Ingber & Mondescu, 2001). Although similar to a standard practice of using windows of raw data, the result is a different than would be calculated just using of raw incoming data (also volume-weighted and level-weighted), since results include most probable correlated future behavior of the market.

If $A_{eff}$ is fit directly, then the derived return equations were only accurate to a give $dt$, e.g., about 0.1 sec., in agreement with the expected mesh used in PATHINT as described above, so instead returns were directly fit to future return data, typically requiring a few million ASA generated states, by “looking ahead” to the next $dt$ point. Of course, PATHINT could be used to fold over the short-time distribution for many secs (as has been done in other systems), but most likely this would too slow in the context of HFT.

If the returns are fit directly to the data, then just about any $dt$, ranging from 0.1 sec to 5 sec give the same forecast value for given current data (returns are transformed back into prices — 6 significant figures), since a realistic set of fitted drifts and diffusion matrices can have parameters slip some from their ideal-$dt$ range to still get good fits at another $dt$, and then the probability calculation at any given point just reflects essentially the same distribution.

4.6 $dt$ Library

A $dt$ library is easily created, within a desired range of $dt$’s that are “reasonably” close to the ideal-$dt$, by doing multiple ASA fits to return data. This defines a library of probabilities that can be used as described here, yielding a range of choices to be made during HFT, e.g., as required to take into account latencies of trades actually being posted.

4.7 Updating Parameters

Fitted parameters can be updated overnight with new ASA runs. Alternatively, e.g., if there is a sudden change in context of the market and if sufficient recent data of the new context is available, the fast modified Nelder-Mead simplex code that comes with the ASA code typically used to efficiently gain some precision in fits, can be used to quickly update the parameters.

4.8 CMI

The forecast also includes the CMI. However, similar to limitations in fitting $A_{eff}$, which cannot be fit directly to the returns data for any $dt$, here too the forecast CMI at a different $dt$ than ideal-$dt$ need to have more ASA generated states for additional precision to get stable CMI (which often are useful guides on upcoming directions and strengths of bids, ask, midprices, etc.).

4.9 Volatility Risk

The risk associated with the probability distributions were straightforwardly calculated from the covariance matrix $g^{GG'}$. A quick measure is its determinant $g$. 
4.10 Additional Functional Complexity

Additional functional complexity, i.e., more terms added to the drifts and covariance matrix were added, but not much seemed to have been gained.

4.11 3-Factor Model

A 3-factor code of bid-ask-trades could have been processed the same way as the 2-factor bid-ask code, but it seemed dynamics were captured quite well by the 2-factor code, so only 2-factor runs were further developed.

4.12 Data

The actual format of the BitMEX data used may be considered proprietary, so no description is given in this paper.

5 Conclusion

A 2-factor probability distribution of bid-ask tick data was developed using Maxima to further develop C code. After parameters of the 2-factor drifts and diffusion matrix were fit, very fast calculations of narrow and wide windows of data are processed that are useful for high-frequency trading.
References


