Microstructure noise components of the S&P 500 index: variation, persistence and distributions[1]

Stephen J. Taylor

Department of Accounting and Finance,
Lancaster University Management School, United Kingdom

s.taylor@lancaster.ac.uk

December 2015

JEL classification: C13, C51, C52, C58, G14

Keywords: High-frequency, Microstructure noise, S&P 500 index, Spot/futures basis

Acknowlegements

I thank participants at the 2015 SoFiE annual conference for their comments. I also thank Ingmar Nolte for numerous helpful conversations and Tristan Linke for sharing data.

[1] This paper supercedes a 2014 working paper on "The distribution of microstructure noise for the S&P 500 index".

Microstructure noise components of the S&P 500 index: variation, persistence and distributions

Abstract

By studying the differences between exchange-traded fund prices and futures prices, new results are obtained about the distribution and persistence of the microstructure noise component created by bid/ask spreads and discrete price scales. The univariate distributions are shown to be time-varying and to depend on the minute of the day, on the year studied and on index volatility. The bivariate density is estimated from high-frequency prices, to provide estimates of the probabilities of one-tick bid/ask spreads, marginal noise densities and measures of noise dependence across the markets studied. Properties of the residual microstructure noise, created by factors other than discrete prices, are also estimated. The residual component has more variation and less persistence than the discrete-price component during the period examined, from January 2010 to December 2012.

1 Introduction

Microstructure noise (MN) is defined as the difference between an actual market price and the efficient price which would be observed if markets had perfect characteristics. We split MN into discrete-price and residual components and show that empirical inferences can be made about these components for the S&P 500 index, despite the impossibility of observing efficient prices. Our results exploit the no-arbitrage constraint on spot and futures prices, with spot prices obtained from an exchange traded fund. We provide the first estimates of the variance and persistence of each MN component. We also provide the first empirical

estimates of the bivariate distribution for the discrete-price components of spot and futures MN.

Continuous-time market prices (or their logarithms) are frequently described as the sum of a semimartingale process plus a MN term representing market frictions, as in the seminal high-frequency paper on MN by Hansen and Lunde (2006). The semimartingale process can include diffusion and jump terms and usually incorporates stochastic volatility (Barndorff-Nielsen, Hansen, Lunde and Shephard, 2008). MN then has many potential sources, which we separate into the two components noted by Hansen and Lunde (2006) and amplified by Ait-Sahalia, Mykland and Zhang (2011). The discrete-price component captures the two most obvious sources of MN, namely the positive tick size which defines the minimum gap between different feasible prices and the positive spread between bid and ask prices. The residual component covers the trading environment and includes effects from order flow, price pressure, inventory control, block trades and asymmetric information. Related insights from economic theory are discussed by Diebold and Strasser (2013), while the specific effect of price pressure is analyzed by Hendershott and Menkveld (2014).

We focus on the discrete-price MN component in this paper. We show the variability of discrete-price MN depends on both the clock time and the volatility of the index, and that appropriate distributions (conditional on the bid/ask spread width) are uniform for the spot asset but are less dispersed than uniform for futures. We also show there is positive although weak correlation between the spot and futures discrete-price MN. There is positive serial dependence in at least the futures component, which can be attributed to trades tending to cluster on either the bid or ask side of the market. We estimate that the appropriate persistence measure for the probability of a trade at the ask corresponds to a half-life of almost 30 minutes.

The no-arbitrage constraint and empirical evidence are used to argue that the spot and futures residual MN are essentially identical. We show that the auto- and cross-correlations of spot and futures price changes (beyond the first lag) are dominated by the residual component and then estimate that the residual component has more variation and less persistence than the discrete-price components. The residual half-life estimate is approximately 3.3 minutes. Typical standard deviations for the discrete-price components are 6 cents for the spot asset and 13 cents for the futures (when their bid/ask spreads are 10 and 25 cents respectively), compared with 30 cents for the common residual component, during the middle of the trading day.

There is now a considerable literature about understanding and mitigating the high-frequency econometric consequences of MN. Methods for measuring the realized variance of efficient prices are developed and compared by Zhang, Mykland and Ait-Sahalia (2005), Bandi and Russell (2008, 2011), Barndorff-Nielsen et al (2008) and Dahlhaus and Neddermeyer (2014), with realized covariance studied in Voev and Lunde (2007) and Corsi and Audrino (2012). Related work covers the impact of MN on volatility forecasting, as in Ait-Sahalia and Mancini (2008), Andersen, Bollerslev and Meddahi (2011) and Ghysels and Sinko (2011). Tests for jumps in efficient prices, robust against MN, are available in Ait-Sahalia, Jacod and Li (2011) and Lee and Mykland (2012).

There are, however, relatively few empirical studies of the statistical properties of MN. For the variance of MN, methods and results are provided by Bandi and Russell (2006), Ait-Sahalia and Yu (2009) and Nolte and Voev (2012), while the autocorrelations of MN are estimated by Ubukata and Oya (2009), Ait-Sahalia, Mykland and Zhang (2011) and Jacod, Li and Zheng (2014).

Researchers often need to make assumptions about the distribution and autocorrelations of MN, either to obtain theoretical results or to design Monte Carlo studies

which clarify the properties of econometric methods. The simplest assumptions are independent and identical Normal distributions, which are common in simulations. Our paper provides a framework for moving beyond these simple assumptions, which we presume are common because distributional and time series facts for MN have hitherto been neglected.

Our data are trade prices for S&P 500 e-mini futures and a spot, exchange-traded fund (ETF) which replicates the S&P 500 index. Section 2 describes the ETF and futures markets. Section 2 also presents no-arbitrage conditions which constrain the differences between spot and futures prices to fall within a narrow interval around the no-arbitrage expected difference. From a MN perspective, there is a fundamental difference between the spot and futures assets due to their different tick sizes. One tick for the ETF equals one cent and as the price of ten ETF shares tracks the index we can say the effective tick size is 10 cents. In contrast, the futures tick size is much larger at 25 cents. We deduce that the expectation-adjusted, spot/futures price difference will fall between -35 cents and 35 cents when (a) each asset has a one-tick, bid/ask spread and (b) the two spreads overlap after adjusting for the expected spot/futures basis. For prices recorded once a minute from January 2010 to December 2012, we find that almost 99% of the adjusted price differences do fall within ± 35 cents. However, we estimate that 12% of these differences occur when the ETF and/or the futures spread is two ticks wide.

Section 3 defines the MN components, states assumptions and presents theoretical results. Section 4 is a detailed exploratory analysis of prices recorded once a minute, including several results about the differences between ETF and futures prices. Section 5 presents and motivates assumptions about the distribution of discrete-price MN, which place strong restrictions on acceptable bivariate distributions. Section 6 then provides estimates of parametric distributions obtained by maximizing a log-likelihood criterion. Section 7 uses the

time-series dependence among one-minute price changes to estimate the variance and persistence of the residual MN component. Finally, Section 8 offers conclusions.

2 Relationships between ETF and futures prices

Several traded contracts provide payoffs proportional to changes in the S&P 500 index. We investigate the prices of an exchange-traded fund (ETF) and the prices of a series of futures contracts to learn about their microstructure properties.

2.1 Two markets for trading the S&P 500 index

The shares of the SPDR S&P 500 ETF represent ownership in a unit investment trust managed by State Street Global Advisors. These shares are traded electronically at several U.S. exchanges, with ticker symbol SPY. We use the ticker symbol as the name of the ETF throughout this paper. The price of 10 SPY shares tracks the level of the S&P 500 index. Dividends are paid quarterly, after the deduction of a management fee which is approximately 0.10% per annum.

On a typical day in 2013, more than 250,000 SPY trades were executed and more than 100 million shares were traded at an average price near to $160. The market capitalization of SPY was then $\$1.4 \times 10^{11}$, which was almost 1% of the total capitalization of the 500 component stocks. One price tick for SPY equals one cent. As may be expected from SPY's high liquidity, the usual width of the visible bid-ask spread is one cent. This corresponds to a 10-cent spread when we multiply all SPY prices by 10 throughout the remainder of this paper.

The e-mini S&P is a futures contract written on the S&P 500 and traded on the CME's Globex electronic trading platform, with ticker symbol ES. The mini contracts are for 50 index units and their trading volume far exceeds that of the original index futures contract,

which is for 250 index units. ES contracts expire in March, June, September and December and their prices are for one unit of the underlying index.

One ES price tick equals 25 cents, which is two-and-a-half times the SPY tick when the underlying quantity is equalized at one index unit. Consequently, it is anticipated that there is substantially more bid-ask spread noise in ES prices than in SPY prices.

The value traded for ES is approximately an order of magnitude higher than the SPY value traded. On a typical day in 2013, 2 million ES contracts were traded, representing 100 million index units at a price of approximately $1600 each. Andersen, Bondarenko, Kyle and Obizhaeva (2015) report an average trade size of 13 contracts and a trade frequency of 6 per second during the period from 09:30 to 16:15 EST. Thus ES trades are slightly less frequent than SPY trades (about 10 per second) but on average they are for a much larger quantity of index units.

SPY pays dividends once a quarter and the ex-dividend dates are identical to the expiration dates of the ES futures contracts. When bid-ask spreads are ignored, the SPY price equals the S&P 500 index level when SPY goes ex-dividend; the SPY price then equals the final futures settlement price at 09:30 EST on the third Friday of an expiry month, which is calculated from the opening prices of the constituent stocks.

2.2 Constraints on prices

Let S_t and F_t respectively denote prices at time t for SPY and the nearest-to-expiry ES contract, with SPY going ex-dividend at times 0 and T. Then there is a standard no-arbitrage pricing equation, which applies for perfect markets having zero spreads, zero transaction

costs, continuous price scales, continuous trading and correctly anticipated dividends, namely[2]:

$$F_t = [1 + r_t(T - t)]S_t - D \quad \text{for} \quad 0 \le t < T\,. \tag{1}$$

Here time is measured in years, r_t is the appropriate risk-free rate and D is the value at time T of the SPY dividend. The term D represents dividend income from the constituent stocks minus management fees, with adjustments for the timing of the constituent cash flows and the delayed payout a few weeks after the ex-dividend date.

As S_t exceeds F_t throughout our sample period, we define the theoretical *basis* as

$$B_t = S_t - F_t = D - r_t(T - t)S_t, \quad 0 \le t < T\,. \tag{2}$$

The basis then increases from $D - r_0 T S_0$ at time 0 to D just before time T. Some representative values for our recent sample period, after multiplying the SPY price and dividend by 10, are $S = 1200$, $D = 5$ and $r = 0.3\%$, for which the basis would increase over 3 months from 4.1 to 5.0. We note that intraday variation in the theoretical basis is expected to be very small.

The above no-arbitrage conditions are not exact when prices to buy exceed prices to sell. Let $S_{t,bid} < S_{t,ask}$ and $F_{t,bid} < F_{t,ask}$ denote the bid and ask prices. Arbitrageurs may be expected to buy one asset and sell the other whenever the price of this portfolio is sufficiently cheap relative to a fair value derived from dividend expectations, interest rates and the SPY price level. In particular, arbitrage trading may be expected when the bid-ask spreads, adjusted for the basis, do not overlap. A pair of no-arbitrage equations is provided by requiring *overlapping spreads*, which implies:

$$S_{t,bid} \le F_{t,ask} + B_t \quad \text{and} \quad S_{t,ask} \ge F_{t,bid} + B_t\,, \tag{3}$$

[2] Similar no-arbitrage equations are well-known and can be found, for example, in MacKinlay and Ramaswamy (1988), Duffie (1989) and Stoll and Whaley (1990).

with B_t the market's consensus value for the basis. If one of these constraints is broken then buying the cheap portfolio does not guarantee a profit, even when trading costs are ignored. Rather, a positive payoff occurs if the initial price advantage (e.g. $S_{t,bid} - (F_{t,ask} + B_t)$) exceeds the present value of the equivalent cost when the portfolio is sold (e.g. $S_{s,ask} - (F_{s,bid} + B_s)$ at some time $s > t$).

For contemporaneous trades agreed at prices S_t and F_t, each of which is either a bid or ask price, overlapping spreads imply that:

$$\left|S_t - (F_t + B_t)\right| \le S_{t,ask} - S_{t,bid} + F_{t,ask} - F_{t,bid} . \quad (4)$$

Thus the recorded magnitudes of differences in traded prices, adjusted for the basis, must not exceed the *spreads cost* which is the sum of the widths of the two bid-ask spreads. We anticipate magnitudes of at most 35 cents when all spreads are one tick wide. A magnitude of more than 35 cents may then attract arbitrageurs seeking positive expected payoffs; beyond 70 cents a profit is almost guaranteed as the initial price advantage exceeds 35 cents which is more than the spreads cost at any future time whose spreads overlap.

3 Components of microstructure noise

It is conventional in microstructure literature to assume there are unobservable, efficient prices which follow continuous-time semimartingale processes. We denote the efficient spot prices by $\{S_t^e, t \ge 0\}$ and assume that the efficient futures prices $\{F_t^e, T > t \ge 0\}$ follow from the no-arbitrage constraint, thus:

$$F_t^e = [1 + r_t(T - t)]S_t^e - D, \quad 0 \le t < T . \quad (5)$$

Microstructure noise (MN) is defined in this paper as the difference between a representative observed price and the efficient price. Using prices facilitates our discussion of

noise components, which would be more complicated if noise was defined as the difference between observed and efficient log prices. We rely on exceptionally high trading volume to assume that at a calendar time *t* there are contemporaneous trade prices S_t and F_t. With microstructure noise denoted by $N_{S,t}$ and $N_{F,t}$,

$$S_t = S_t^e + N_{S,t} \quad \text{and}$$

$$F_t = F_t^e + N_{F,t}. \tag{6}$$

We separate total MN into *discrete-price MN* caused by positive tick sizes and spreads and *residual MN* which accumulates all other price frictions.

3.1 Residual microstructure noise

We assume an asset's residual MN depends on a component shared by the spot and futures assets and possibly also on an idiosyncratic component. We denote by M_t the contribution of common, residual MN to spot prices. It is plausible to suppose that the sum $S_t^e + M_t$ and a corresponding futures sum satisfy the no-arbitrage constraint, which implies the futures sum is $F_t^e + (1 + r_t(T-t))M_t$. Denoting the idiosyncratic, residual MN terms by $m_{S,t}$ and $m_{F,t}$, the latent prices when there is no discrete-price MN are:

$$S_t^* = S_t^e + M_t + m_{S,t} \quad \text{and}$$

$$F_t^* = F_t^e + (1 + r_t(T-t))M_t + m_{F,t}. \tag{7}$$

We will assume the three residual components in (7) have zero expectations and zero cross-covariances. Also, the vectors of residual components are assumed to be stochastically independent of the vectors of efficient prices. The idiosyncratic components, if they do exist, must be small relative to spreads otherwise there will be arbitrage opportunities. We present

empirical evidence in Section 4 supporting the claim that (almost) all the residual MN can be explained by a common component.

3.2 Discrete-price microstructure noise

At all times the most competitive bid/ask spreads on offer are assumed to include the prices S_t^* and F_t^*, so $S_{t,bid} \le S_t^* \le S_{t,ask}$ and $F_{t,bid} \le F_t^* \le F_{t,ask}$. The discrete-price MN components U_t and V_t are then defined by:

$$S_t = S_t^* + U_t = S_t^e + M_t + m_{S,t} + U_t \qquad \text{and}$$

$$F_t = F_t^* + V_t = F_t^e + (1 + r_t(T-t))M_t + m_{F,t} + V_t . \tag{8}$$

Note that the continuous random variables S_t^* and U_t are not independent because their sum must be a multiple of the tick size and hence $U_t \left| S_t^* \right.$ has a time-varying, discrete distribution.

We define the discrete-price MN for trades at the ask prices by:

$$U_t^+ = S_{t,ask} - S_t^* \quad \text{and}$$

$$V_t^+ = F_{t,ask} - F_t^* . \tag{9}$$

As the widths of the bid-ask spreads are very small relative to the asset prices, it is safe to assume that the distributions of U_t^+ and V_t^+ conditional upon the spread widths $S_{t,ask} - S_{t,bid}$ and $F_{t,ask} - F_{t,bid}$ are uniform, respectively with positive densities on the intervals $[0, S_{t,ask} - S_{t,bid}]$ and $[0, F_{t,ask} - F_{t,bid}]$. We will also assume U_t^+ is independent of V_t^+ for the following reasons: (1) the basis is stochastic and it can be assumed to have a continuous distribution, (2) the variation in basis values is large relative to the spread widths,

and so (3) at a random time moment the location of F_t^* within its spread (measured by V_t^+) does not depend on the corresponding location of S_t^* within its spread (measured by U_t^+).

The distributions of U_t and V_t, conditional upon their spread widths, will only be uniform and independent *if* we make the additional assumptions that trades occur randomly at the bid and ask ends of the spreads, with chance one-quarter for all end combinations whatever the outcomes for S_t^* and F_t^*. For these special assumptions, and for a one-tick spot spread width w, $U_t^+ \sim \text{uniform}[0, w]$, $U_t \sim \text{uniform}[-w, w]$, $\text{var}(U_t^+) = w^2/12$ and $\text{var}(U_t) = w^2/3$; as the one-tick futures spread width is $2.5w$, $\text{var}(V_t) = 6.25w^2/3$ and $\text{var}(U_t - V_t) = 7.25w^2/3$, so the standard deviations of U_t, V_t and $U_t - V_t$ are respectively 5.77, 14.43 and 15.55 cents as w equals 10 cents.

There are, however, three assumptions which we should *not* automatically make about the discrete-price MN variables. First, we do not assume U_t and V_t have conditional uniform distributions. One good reason is that trades may be more likely to occur at the ends of the bid/ask spreads nearest to S_t^* and F_t^*. Let $X_t = \text{sign}(S_t - S_t^*)$, which is 1 for a spot trade at the ask price and -1 for a trade at the bid price. Then

$$U_t = U_t^+ + \tfrac{1}{2}(X_t - 1)(S_{t,ask} - S_{t,bid}). \tag{10}$$

As the outcome u_t^+ for U_t^+ decreases towards zero, S_t^* approaches the ask price and $P(X_t = 1 \mid u_t^+)$ may increase. Hence there might be a negative correlation between U_t^+ and X_t which decreases the variance of U_t below the level for a uniform distribution. To illustrate the possible effect, suppose the spot spread width is fixed at w and that

$$P(X_t = 1 \mid U_t^+) = \tfrac{1}{2}(1 + \lambda) - \lambda(U_t^+/w), \quad \text{with} \quad 0 \le \lambda \le 1. \tag{11}$$

Then it is easy to show that

$$E[U_t^+ X_t] = -\lambda w / 6$$

and

$$\text{var}(U_t) = (2 - \lambda) w^2 / 6 . \tag{12}$$

Second, we do not assume U_t is independent of V_t because both trades on the same side of the spread (two bids or two asks) may be more likely than trades at opposite ends (one bid and one ask); in these circumstances $Y_t = \text{sign}(F_t - F_t^*)$ will be positively correlated with X_t and thus V_t will be positively correlated with U_t.

Third, we do not assume that a time series of noise terms $\{U_t\}$ have independent distributions because a trade at the ask (bid) may be more likely if recent trades have tended to be at ask (bid) prices; then the trade direction process $\{X_t\}$ may be persistent leading to positive autocorrelation among the $\{U_t\}$.

Finally in this subsection we note that the difference between traded spot and futures prices, adjusted for the basis, is essentially the sum of two MN component differences. When the market's value of the basis is the difference between efficient prices, $B_t = S_t^e - F_t^e$, then

$$\begin{aligned} S_t - (F_t + B_t) &= (S_t - S_t^e) - (F_t - F_t^e) = N_{S,t} - N_{F,t} \\ &\cong m_{S,t} - m_{F,t} + U_t - V_t, \end{aligned} \tag{13}$$

when interest terms are ignored. The absolute value of $S_t - (F_t + B_t)$ must not exceed the sum of the widths of the two bid-ask spreads when the no-arbitrage equations apply, for all four combinations of feasible spot and futures prices, which constrains the variation of the idiosyncratic MN terms.

3.3 Components of variances and covariances

Some inferences about appropriate price models can be made by studying the variances, autocovariances and cross-covariances of price changes. For a time interval δ, let Δ be the price change operator, so, for example, $\Delta S_t = S_t - S_{t-\delta}$. Then the total variation in spot price changes is

$$\text{var}(\Delta S_t) = \text{var}(\Delta S_t^*) + \text{var}(\Delta U_t) + 2\,\text{cov}(\Delta S_t^*, \Delta U_t). \tag{14}$$

The covariance term is either zero or relatively small. This and similar covariance terms are assumed to be zero in the following material. To understand why, note from (10) that $\text{cov}(\Delta S_t^*, \Delta U_t)$ is the sum of $\text{cov}(\Delta S_t^*, \Delta U_t^+)$ and $\frac{1}{2} w\,\text{cov}(\Delta S_t^*, \Delta X_t)$ when the spread width is a constant *w*. The assumption that U_t^+ has a uniform distribution is sufficient to prove that $\text{cov}(\Delta S_t^*, \Delta U_t^+) = 0$, while there is no compelling reason for a substantial correlation between ΔS_t^* and ΔX_t. Ignoring the covariance term in (14), we have:

$$\text{var}(\Delta S_t) = \text{var}(\Delta S_t^e) + \text{var}(\Delta M_t) + \text{var}(\Delta m_{S,t}) + \text{var}(\Delta U_t). \tag{15}$$

The variance of efficient price changes is proportional to δ. For very short intervals almost all of the variation in price changes comes from noise changes and in particular from prices bouncing across the bid/ask spread. In contrast, for intervals which are long relative to persistence measures for MN, almost all variation is due to efficient price changes.

With,

$$R_{t,T} = r_t(T - t)$$

denoting the (generally very small) interest rate from now (time *t*) until futures expire (at *T*),

$$\text{var}(\Delta F_t) = (1 + R_{t,T})^2[\text{var}(\Delta S_t^e) + \text{var}(\Delta M_t)] + \text{var}(\Delta m_{F,t}) + \text{var}(\Delta V_t) \quad \text{and}$$

$$\text{cov}(\Delta S_t, \Delta F_t) = (1 + R_{t,T})[\text{var}(\Delta S_t^e) + \text{var}(\Delta M_t)] + \text{cov}(\Delta U_t, \Delta V_t). \tag{16}$$

When the idiosyncratic terms are negligible, we anticipate $\text{var}(\Delta F_t) > \text{var}(\Delta S_t)$ as we may expect $\text{var}(\Delta V_t) > \text{var}(\Delta U_t)$; and if discrete MN is (almost) uncorrelated across the two markets, then we may also expect $\text{var}(\Delta S_t) > \text{cov}(\Delta S_t, \Delta F_t)$.

Only the noise components contribute to covariances for non-overlapping price changes. If we simplify by assuming stationary processes for all noise components, with $\rho_{\tau,Z} = \text{cor}(Z_t, Z_{t+\delta\tau})$ denoting the autocorrelation at the integer lag τ of a process Z observed once every δ time units, then for $\tau > 0$:

$$\begin{aligned}\text{cov}(\Delta S_t, \Delta S_{t+\delta\tau}) &= \text{cov}(\Delta N_{S,t}, \Delta N_{S,t+\delta\tau}) \\ &= (2\rho_{\tau,M} - \rho_{\tau-1,M} - \rho_{\tau+1,M})\,\text{var}(M_t) \\ &\quad + (2\rho_{\tau,m_S} - \rho_{\tau-1,m_S} - \rho_{\tau+1,m_S})\,\text{var}(m_{t,S}) \\ &\quad + (2\rho_{\tau,U} - \rho_{\tau-1,U} - \rho_{\tau+1,U})\,\text{var}(U_t).\end{aligned}$$

For short time differences $\delta\tau$, during which changes in interest terms are trivial, and for $\tau > 0$,

$$\begin{aligned}\text{cov}(\Delta F_t, \Delta F_{t+\delta\tau}) &\cong (2\rho_{\tau,M} - \rho_{\tau-1,M} - \rho_{\tau+1,M})(1 + R_{t,T})^2\,\text{var}(M_t) \\ &\quad + (2\rho_{\tau,m_F} - \rho_{\tau-1,m_F} - \rho_{\tau+1,m_F})\,\text{var}(m_{t,F}) \\ &\quad + (2\rho_{\tau,V} - \rho_{\tau-1,V} - \rho_{\tau+1,V})\,\text{var}(V_t).\end{aligned} \tag{17}$$

Also, for cross-correlations $\gamma_\tau = \text{cor}(U_t, V_{t+\delta\tau})$, when $\tau \neq 0$,

$$\begin{aligned}\text{cov}(\Delta S_t, \Delta F_{t+\delta\tau}) &\cong (2\rho_{\tau,M} - \rho_{\tau-1,M} - \rho_{\tau+1,M})(1 + R_{t,T})\,\text{var}(M_t) \\ &\quad + (2\gamma_\tau - \gamma_{\tau-1} - \gamma_{\tau+1})\sqrt{\text{var}(U_t)\,\text{var}(V_t)}.\end{aligned}$$

The interest-rate terms make trivial contributions to the equations above for our sample, as the average values of r_t, $T - t$ and $R_{t,T}$ are respectively 0.3% per annum, 1/8 years and 0.0004 approximately. Consequently, in the remainder of this paper we assume $R_{t,T} \equiv 0$.

4 Exploratory data analysis

4.1 Data

SPY and ES prices are studied for the three-year period from 4 January 2010 to 31 December 2012. Only prices recorded during the primary trading sessions are considered. For both assets these sessions commence at 09:30 EST and conclude at 16:15 EST. Although trade does occur during all hours of the day, trading volumes are much lower before 09:30 and after 16:15; 95% of the total SPY volume occurred within the primary trading periods. We will show that the variance of discrete-price MN is lower mid-day and for this reason selected results are presented for the mid-day period from 11:00 until 15:00.

Transaction prices are analyzed at the one-minute frequency, because this is the highest frequency for which futures prices are available to us. Our primary data for each trading day are 405 SPY and ES prices defined by the last recorded trade prices before 09:31, 09:32, …, 16:15; we also briefly consider the first recorded trade prices after the minute marks. The time differences between pairs of SPY and ES prices are fractions of a second and therefore it will be assumed that the matched prices are contemporaneous. On most days the nearest-to-expiry ES contract is selected, but if some of its prices are missing then the second-nearest contract is used.

Only days when both assets have complete data are analyzed. There are 248, 249 and 247 complete days, respectively in 2010, 2011 and 2012. The Flash Crash day (6 May 2010) is discarded, when volatility was exceptionally high, and the following high-volatility day (7 May) is also discarded. The data investigated is then for $742 \times 405 = 300{,}510$ one-minute intervals.

The tick size for SPY is one cent at all exchanges. All quotes we have checked in the TAQ database are for an exact multiple of one cent. However, 17% of the recorded trade

prices for SPY are “off tick”; the dollar prices then have a non-zero digit in the third and/or fourth decimal place. This phenomenon is identically present in our primary data source and in the TAQ database. As explained by Buti, Consonni, Rindi, Wen and Werner (2014), off-tick trade prices occur in TAQ because trades by dark pools are included in the database. Off-tick prices are rounded to the nearest cent for our reported calculations; we are not aware of any important differences between using rounded and recorded prices for the exploratory data analysis.

Further discussion of the data and our processing methods is provided in Appendix A. For the remainder of this paper, the variable t now counts trading days, while j counts minutes within the day.

4.2 Trade-to-trade price changes

Some quick insights into discrete-price MN are provided by data about the price change from the last transaction price before a minute mark to the first transaction price after the same mark. For day t and minute j, the last trade prices for SPY and ES are respectively denoted $S_{t,j}$ and $F_{t,j}$, while the next prices are denoted $S_{t,j+}$ and $F_{t,j+}$. As the time gap between consecutive prices is of the order of 0.1 seconds, we can usually ignore any changes in efficient prices and residual MN during these short time periods and then state:

$$S_{t,j+} - S_{t,j} = U_{t,j+} - U_{t,j}\,.$$

If we also assume there are no changes in the bid and ask prices, then $U^{+}_{t,j+} = U^{+}_{t,j}$ and

$$S_{t,j+} - S_{t,j} = \tfrac{1}{2}(X_{t,j+} - X_{t,j})(S_{t,ask} - S_{t,bid})\,, \tag{18}$$

so the price change is then either zero or $\pm$ the bid-ask spread.

For the mid-day period from 11:00 to 15:00 inclusive, we find that 99.0% of the minute marks have both (1) trade-to-trade spot price changes equal to either zero or $\pm$ one

spot tick and (2) likewise for future price changes; these calculations exclude all minute marks having an off-tick spot price. Table 1 summarizes the bivariate distribution of the price changes when all changes beyond one tick are excluded. We obtain three conclusions from our brief review of trade-to-trade price changes. First, the width of the bid-ask spread is usually one tick during the mid-day period. Second, the chance of consecutive trades being on the same side of the market is estimated as 62.6% for SPY and 59.7% for ES. These numbers are useful upper bounds for prices recorded further apart, i.e. for $P(X_{t,j} = X_{t,k})$ and $P(Y_{t,j} = Y_{t,k})$ when $j \neq k$. Third, the spot and futures price changes are almost independent (their correlation equals 0.05) and from their bivariate distribution we can estimate that the chance of trades at the same ends of the spreads, i.e. $P(X_{t,j} = Y_{t,j})$, is between 51% and 63% based on the methods presented in Appendix B.

4.3 Differences between spot and futures prices

We assume the noise-free basis, $S^e_{t,j} - F^e_{t,j}$, is constant within each trading day. We estimate the common value, B_t, by the average of *N* observed price differences,

$$\hat{B}_t = \frac{1}{N} \sum_{i=1}^{N} S_{t,i} - F_{t,i}\,. \tag{19}$$

This is an unbiased estimate, i.e. $E[\hat{B}_t | B_t] = B_t$, as all the microstructure noise terms have zero expectations. *Adjusted price differences* are defined as price differences minus basis estimates:

$$q_{t,j} = S_{t,j} - F_{t,j} - \hat{B}_t\,. \tag{20}$$

In the following text we often drop the adjective “adjusted” and simply refer to the quantity $q_{t,j}$ as a *price difference*.

Ideally we could observe B_t and study the properties of

$$q^e_{t,j} = S_{t,j} - F_{t,j} - B_t \ = U_{t,j} - V_{t,j} + m_{S,t,j} - m_{F,t,j}\,. \tag{21}$$

The estimation error $\varepsilon_t = \hat{B}_t - B_t$ is, however, small relative to $q^e_{t,j}$ when N is large, since

$$\begin{aligned} q_{t,j} &= q^e_{t,j} - \varepsilon_t \\ &= (S_{t,j} - F_{t,j} - B_t) - \frac{1}{N}\sum_{i=1}^{N}(S_{t,i} - F_{t,i} - B_t). \end{aligned} \tag{22}$$

The typical magnitude of the estimation error is quantified in Section 4.3.5, after estimating the autocorrelations of the price differences.

It is convenient to select cents as the price units. The rounded SPY price is a multiple of 10 cents, remembering that this is the price of 10 shares, while the ES price is a multiple of 25 cents. Then $S_{t,j} - F_{t,j}$ is a discrete variable; it is a multiple of 5 cents and all multiples are possible. We may suppose the estimate $\hat{B}_t$ is a continuous variable for large N as the minimum difference between distinct feasible values equals $5/N$ cents. Consequently we may treat the price difference $q_{t,j}$ as continuous, although on each day t any pair of values will be separated by some multiple of 5 cents.

4.3.1 Most price differences are small

We expect the magnitude of the price differences to be less than 35 cents when the SPY bid-ask spread is one tick (and thus 10 cents wide), the ES spread is one tick (25 cents wide), the two spreads overlap after controlling for the basis, and the basis estimation error is negligible. A striking property of the price differences is that almost all are indeed between -35 and 35 cents. The overall frequency of this event is 98.8% when prices are measured at the $N = 405$ minutes from 09:31 to 16:15 inclusive.

Price differences within ± 35 can occur if a spread is wider than one tick and/or the spreads do not overlap. Thus the frequency of overlapping, one-tick spreads is less than 98.8%. In Section 6.3 we estimate the average probability of overlapping, one-tick spreads to be approximately 87% across the trading day, after fitting parametric density functions to the differences.

The frequency of differences inside ± 35 varies across years and across the time within the trading day. The annual averages are 99.3% in 2010, 98.3% in 2011 and 98.8% in 2012, with lower frequencies before 11:00 and after 15:00. Table 2 shows the frequency for each year outside ± 35, before 11:00, after 15:00, and between these times. The central time period from 11:00 to 15:00 contains 60% of the 405 minute-marks but only 33% of the differences outside ± 35.

The daily ranges of the price differences confirm the small sizes of the price differences. The modal category for the daily range from 09:31 to 16:00 inclusive is 70, 70 and 65 cents in 2010, 2011 and 2012. The medians are higher at 75, 85 and 80, because the daily ranges are skewed.

The dispersion of the price differences is summarized in Table 2 by mean absolute deviations (m.a.d.) and standard deviations. As there are some extreme outliers, we focus on the m.a.d. measure of dispersion. The m.a.d. for a single trading day is usually between 11 and 13 cents. Figure 1 shows there are subperiods having a higher level, notably from August to November 2011 and during December 2012.

When the discrete-price MN terms are independent and have uniform distributions with maximum values equal to one tick, the m.a.d. of the noise difference, $U_{t,j} - V_{t,j}$, equals 13.17 cents and the standard deviation is 15.55 cents.[3] These special values are higher than the estimates for our complete dataset, namely 12.02 and 15.30. As price differences are

[3] The m.a.d. equation is included in Section 5.4 after (40).

noise differences plus a term representing basis estimation error plus another term for differences in idiosyncratic MN, we conclude that the empirical dispersion is less than predicted by the above collection of assumptions. On Figure 1 we see that most days have a m.a.d. estimate below the independent uniform (IU) value of 13.17.

4.3.2 Large price differences

There is a tendency for price differences outside ± 35 to cluster. Only 1.2% of the differences are outside ± 35. The conditional chance of an outside event, when one occurs for the previous minute, is higher and equals 5%, 8% and 13% in 2010, 2011 and 2012. The longest consecutive periods of time outside, by year, are 5, 6 and 8 minutes. A few days have an exceptional number of price differences outside ± 35, such as 77 and 85 occurrences on 18 and 27 December 2012 respectively and a total of 216 occurrences on the 5 consecutive, high-volatility, market days from 5 to 11 August 2011 inclusive.

Approximately 9% of the differences outside ± 35 are also outside ± 70. Table 2 includes the frequencies of differences beyond ± 70. The overall frequency of this event is only 0.11%. There are 37 differences outside ± 140, with frequency 0.012%; of these 22 occur at 16:00 or later, 7 occur between 09:59 and 10:01 and 5 between 14:15 and 14:30.

4.3.3 Intraday variation in price differences

There is some variation across minutes in the average level of the price differences, but this variation is small compared with the standard deviations of the price differences. Most of the averages for each minute are within ± 1 cent. The averages tend to be negative for the first 15 or so minutes of the day, are then more likely to be positive until around 13:15 and then more likely negative. Standard tests on individual averages reveal weak evidence that expected

differences are not zero. Regressions of averages against three functions of time[4] provide strong evidence, however, that expectations vary intraday. These expectations are all within ± 0.5 cents. They generally fall as the day proceeds, so they cannot be explained by intraday variation in the theoretical basis. Instead, they might be explained by day traders holding positive quantities of SPY, which are purchased earlier in the day than they are sold.

Figure 2 displays the m.a.d. of the price differences for each minute of the trading day. The three highest values are at 10:00, 16:00 and 16:15, respectively coinciding with macroeconomic news releases, the end of trading at Wall Street and the end of the primary trading sessions for SPY and ES. The m.a.d. starts the day around 13 cents, declines gradually towards 11 cents and then rises to 13 cents at 15:59; the values are notably higher after 16:00. Once more we see that most m.a.d. values are below the IU level of 13.17. We may infer that the price differences are generally small across the Wall Street trading period, except around news announcements. The curve on Figure 2 shows fitted values from a quadratic regression.

4.3.4 Dependence on volatility

There is some association between the volatility of the index and the magnitude of price differences. This is hinted at by the U-shape of intraday, average magnitudes (see Figure 2) which resembles the well-known U-shape of intraday, average volatility. Association is observed directly between the daily realized variance and both the mean absolute deviation and the standard deviation of the day's price differences, with more correlation found with the standard deviation. The rank correlations between time series of realized volatility for

[4] The functions are j, j^2 and $\sin(2\pi(j-1)/389)$ and the regressions are estimated by OLS for the period 09:31 to 16:00. The p-value of each estimated slope coefficient is less than 10^{-6}.

SPY and the standard deviation of adjusted differences between SPY and ES prices[5] are 0.36 in 2010, 0.64 in 2011 and 0.24 in 2012, all of which are highly significant.

The average level of dispersion in price differences is almost constant when realized volatility is within the first three quartiles, as can be seen from the following averages:

Realized volatility	Days	Average m.a.d.	Average s.d.	Beyond[6] ±35 median	Beyond[6] ±35 average
<0.6	273	11.6	14.1	2	3.0
0.6-0.8	202	11.7	14.2	2.5	3.5
0.8-1.0	121	11.9	14.6	3	5.4
1.0-1.5	106	12.1	15.2	5	6.8
1.5-2.0	29	13.0	16.8	11	12.2
>2.0	11	15.5	21.9	27	28.8

It is seen that price differences outside ± 35 cents become more likely as the price volatility increases.

4.3.5 Autocorrelation among price differences

The autocorrelation between price differences, τ minutes apart on the same trading day, is estimated for a sample of *M* days by

$$\hat{\rho}_\tau = \frac{N \sum_{t=1}^{M} \sum_{j=1}^{N-\tau} q_{t,j} q_{t,j+\tau}}{(N-\tau) \sum_{t=1}^{M} \sum_{j=1}^{N} q_{t,j}^2}. \qquad (23)$$

Figure 3 shows the estimates for each year, when *N* equals 405 for the period from 09:31 until 16:15. It is seen that the estimates are small and always positive when the time gap τ is less than one hour. These small estimates are much larger than the negligible dependence

[5] Daily realized volatility is the square root of realized variance (RV), here calculated very simply from the sum of squared, one-minute, percentage returns for the period from 09:30 to 16:00. The standard deviations of the adjusted price differences are calculated for periods from 09:31 to 15:59.
[6] Counts of minutes from 09:31 to 16:15 with price differences outside -35 to 35.

created by firstly mis-measurement of the basis and secondly the intraday variation in the average level of the price differences[7].

The estimated autocorrelations decline slowly, resembling estimates from a general ARMA(1,1) process whose theoretical autocorrelations are $\rho_\tau = A\phi^\tau, \tau \geq 1$. The curves on Figure 3 show values for ρ_τ provided by minimizing $\sum(\rho_\tau - \hat{\rho}_\tau)^2$. The three estimates of the autoregressive parameter ϕ are similar, at 0.971 for 2010, 0.981 for 2011 and 0.982 for 2012. In contrast, the estimates of $\rho_1 = A\phi$ in consecutive years increase, from 0.062 to 0.102 to 0.129. Consequently, microstructure noise has some time-varying properties. Estimates of ϕ for the mid-day period from 11:00 to 15:00 remain similar, being 0.965, 0.975 and 0.980, while the estimates of ρ_1 are then slightly lower at 0.056, 0.079 and 0.098.

The price difference $q_{t,j}$ incorporates an estimate of the basis. The variance of the estimation error $\hat{B}_t - B_t$ equals the variance of $\hat{B}_t$ which depends on the autocovariances of the price differences, as there is the excellent approximation

$$\text{var}(\hat{B}_t) \cong \text{var}(\frac{1}{N}\sum_{j=1}^{N} q_{t,j}). \tag{24}$$

Assuming the price difference process is stationary, with positive autocorrelations $\rho_\tau = A\phi^\tau$,

$$\text{var}(\hat{B}_t) < \frac{1}{N}\text{var}(q_{t,j})(1 + 2\sum_{\tau>0}\rho_\tau) = \frac{1-\phi+2A\phi}{N(1-\phi)}\text{var}(q_{t,j}). \tag{25}$$

[7] When the terms $q^e_{t,j}$ are zero-mean and uncorrelated, the expected value of $\hat{\rho}_\tau$ equals $-1/N = -0.0025$. Replacing each term $q_{t,j}$ in the numerator of (23) by a plausible estimate of its intraday expectation μ_j gives an expected first-lag autocorrelation of 0.0004 when the terms $q_{t,j} - \mu_j$ are uncorrelated.

Substituting estimates from Table 2 and from the fitted autocorrelations $\hat{A}\hat{\phi}^{\tau}$ provides upper-bound estimates of the standard error of $\hat{B}_t$ equal to 1.72, 2.71 and 2.90 cents, respectively in 2010, 2011 and 2012.

4.4 Variances, covariances, autocorrelations and cross-correlations for price changes

As ES futures have wider bid/ask spreads than the SPY ETF, we expect the changes in our futures prices to be more variable than the changes in our spot prices; we also expect more negative covariation to be found between consecutive futures changes than between consecutive spot changes. The summary statistics in Table 3 confirm both of these expectations[8]. The tabulated numbers are for one-minute price changes, $\Delta S_{t,j} = S_{t,j} - S_{t,j-1}$ and $\Delta F_{t,j} = F_{t,j} - F_{t,j-1}$, and also for one-minute returns, $\ln(S_{t,j} / S_{t,j-1})$ and $\ln(F_{t,j} / F_{t,j-1})$. We find that the one-minute variances are 14% higher for the futures asset than the spot asset, while the one-minute covariances and correlations for futures are a factor 3 to 4 times the spot level. As the futures/spot covariance ratio is well below the ratio for squared one-tick sizes, namely $2.5^2 = 6.25$, it is probable that variation in residual MN makes a significant contribution to the negative, first-lag covariances.

Table 3 also lists the estimated autocorrelations and cross-correlations for one-minute price changes and returns for leads and lags from one to ten minutes. These correlations are very similar for price changes and returns. All the tabulated numbers are negative. Figure 4 plots these negative correlations and also displays some positive estimates for lags beyond ten minutes. Only correlations at lags 1, 2 and 4 are clearly outside the robust 95% intervals of Taylor (1984) and Lo and MacKinlay (1988) for an uncorrelated process, shown by dotted

[8] Results are tabulated for the mid-day period, from 11:00 to 15:00, to avoid distorting effects from the intraday pattern in volatility. All covariances are estimated from products of terms a fixed time apart on the same trading day, as illustrated by (23).

curves on Figure 4. Both Table 3 and Figure 4 show that the sample correlations are very similar for the four combinations of either $\Delta S_{t,j}$ or $\Delta F_{t,j}$ with either $\Delta S_{t,j+\tau}$ or $\Delta F_{t,j+\tau}$ when $\tau \geq 2$. This strongly suggests that these correlations reflect persistence in a common noise component, which can only be the residual MN component.

4.5 Component models consistent with the data

Five MN components were introduced in Section 3, namely spot discrete-price MN ($U_{t,j}$), futures discrete-price MN ($V_{t,j}$), residual MN identically present in spot and futures prices ($M_{t,j}$) and idiosyncratic residual MN for spot ($m_{S,t,j}$) and futures ($m_{F,t,j}$) prices. We now identify the simplest plausible models for these components which are compatible with (1) the variance of price differences $q_{t,j} = S_{t,j} - F_{t,j} - \hat{B}_t$, (2) the autocorrelations of the price differences and (3) the auto- and cross-covariances of the price changes $\Delta S_{t,j}$ and $\Delta F_{t,j}$.

First, we infer that it is credible to discard the idiosyncratic components because the level of variation in $q_{t,j}$ can be entirely explained by variation in the two discrete-price components. The sample standard deviation of $q_{t,j} = U_{t,j} - V_{t,j} + m_{S,t,j} - m_{F,t,j} - \varepsilon_t$ equals 15.30 cents, which is less than the standard deviation of 15.55 cents for $U_{t,j} - V_{t,j}$ when the discrete-price MN terms are independent and have uniform distributions with maximum values equal to one tick. As the basis estimation error ε_t can be assumed uncorrelated with all other terms, we can explain the variation in $q_{t,j}$ by nonuniform distributions and/or dependent noise terms without requiring idiosyncratic components. Of course we cannot prove idiosyncratic components do not exist, but if they do then no-

arbitrage considerations and the above remarks show they must have minor variation. From now on, we assume $m_{S,t,j} \equiv m_{F,t,j} \equiv 0$.

Second, the persistent autocorrelations of the price differences can only be explained by persistence in the futures noise term $V_{t,j}$ because this term accounts for most of the variation in $q_{t,j} = U_{t,j} - V_{t,j} - \varepsilon_t$. Positive dependence in $V_{t,j}$ will occur if trades tend to cluster on one side of the market for several minutes. To illustrate its possible magnitude, let $Y_{t,j}$ be 1 for a futures trade at the ask price and -1 for a trade at the bid price. Also let $Z_{t,j}$ be the latent probability of a trade at the ask, so $P(Y_{t,j} = 1 \big| Z_{t,j}) = Z_{t,j}$. With $v_{\max}$ the width of the futures spread, assumed constant, from (10):

$$V_{t,j} = V^+_{t,j} + \tfrac{1}{2} v_{\max} (Y_{t,j} - 1). \tag{26}$$

When we can assume $\{V^+_{t,j}\}$ is an i.i.d. process[9], stochastically independent of $\{Y_{t,j}\}$, it can be shown that the autocovariances of discrete-price noise are proportional to those of the probability process, with

$$\operatorname{cov}(V_{t,j}, V_{t,j+\tau}) = v^2_{\max} \operatorname{cov}(Z_{t,j}, Z_{t,j+\tau}). \tag{27}$$

A uniform distribution for $V_{t,j}$ follows from the assumptions above, when $V^+_{t,j}$ is uniform and $Y_{t,j}$ has zero expectation, and hence the proportional relationship for autocorrelations is

$$\rho_{\tau,V} = 3 \operatorname{var}(Z_{t,j}) \rho_{\tau,Z}. \tag{28}$$

The discrete-price noise process then has the ARMA(1,1)-style autocorrelations $A_V \phi_V^\tau$, seen on Figure 3, when $\{Z_{t,j}\}$ is AR(1) with autocorrelations ϕ_V^τ. A ballpark estimate of

[9] Independence is a reliable assumption for one-minute observations and our empirical level of volatility. It would not be reliable for much shorter times between prices.

$\text{var}(Z_{t,j})$ comes from supposing consecutive trades have the same probabilities of being at the ask, i.e. $Z_{t,j+} = Z_{t,j}$. Then

$$P(Y_{t,j+} = Y_{t,j}) = \tfrac{1}{2} + 2\,\text{var}(Z_{t,j}). \tag{29}$$

From Section 4.2 and Table 1, we estimate $\text{var}(Z_{t,j})$ to be approximately 0.05 giving an upper bound of 0.15 for the noise autocorrelations. This is consistent with the autocorrelations of the price differences shown on Figure 3.

Third, the covariance between $\Delta F_{t,j}$ and $\Delta F_{t,j+\tau}$ when $\tau \geq 2$ is dominated by the covariances of residual MN because the discrete-price MN is highly persistent. From (17),

$$\begin{aligned}\text{cov}(\Delta F_{t,j}, \Delta F_{t,j+\tau}) = {} & (2\rho_{\tau,M} - \rho_{\tau-1,M} - \rho_{\tau+1,M})\,\text{var}(M_{t,j}) \\ & + (2\rho_{\tau,V} - \rho_{\tau-1,V} - \rho_{\tau+1,V})\,\text{var}(V_{t,j}).\end{aligned} \tag{30}$$

Supposing the autocorrelations of the terms $V_{t,j}$ can be modelled as $A_V \phi_V^{\tau}$, $\tau \geq 1$, the final term can be approximated as $-A_V(1-\phi_V)^2 \phi_V^{\tau-1}\,\text{var}(V_{t,j})$ when $\tau \geq 2$. When $A_V = 0.08$ and $\phi_V = 0.975$ from Section 4.3.5, and $V_{t,j}$ is uniformly distributed between -25 and 25 cents, this term is approximately -0.010 at lag 2, compared with the empirical covariance of -38 derived from Table 3. We therefore ignore the discrete-price term for the futures asset when we estimate a process for the residual MN in Section 7 and it is plausible to also ignore comparable contributions to spot covariances and the cross-covariances. From Table 3, which shows negative autocorrelations for $\Delta F_{t,j}$, a natural model for $\{M_{t,j}\}$ is an AR(1) process with positive autocorrelations ϕ_M^{τ}. Then

$$\text{cov}(\Delta F_{t,j}, \Delta F_{t,j+\tau}) \cong -(1-\phi_M)^2 \phi_M^{\tau-1}\,\text{var}(M_{t,j}),\ \tau \geq 2. \tag{31}$$

5 Theoretical distributions for discrete-price microstructure noise

A general framework for the bivariate distribution of contemporaneous, discrete-price, microstructure noise at two markets is now presented, followed by three plausible assumptions which constrain specifications of the bivariate density.

The empirical results in Section 4 imply that satisfactory specifications must have nonuniform marginal distributions and/or positive dependence between the two noise terms. The first possibility can be motivated by expecting trades to be more likely to occur at the nearest feasible market prices to noise-free prices, while the second can be motivated by a common liquidity pressure which makes it more likely that both markets trade at the bid or at the ask than one at the bid and the other at the ask. Examples of densities compatible with the assumptions and these motivating remarks are provided in Sections 5.4 and 5.5, with the most general specification having six free parameters.

5.1 A general framework

Whenever possible we suppress all notation referring to time in Section 5, to simplify the presentation of relationships between random variables. Our goal is to use adjusted price differences q to estimate distributions for the spot noise $U = S - S^*$ and the futures noise $V = F - F^*$, with each discrete-price noise term equal to a trade price minus an unobservable latent price which incorporates common price components.

It is innocuous to suppose the latent prices have continuous distributions, so we assume the distribution of $(U,V)'$ is continuous. The bivariate density of U and V is denoted by $h(u,v)$.

As in Section 4.3, each price difference Q equals a noise difference D minus a basis estimation error ε:

$$\begin{aligned} D &= U - V, \\ Q &= D - \varepsilon. \end{aligned} \tag{32}$$

The density of D follows from the density of $(U,V)'$ and equals:

$$\psi(d) = \int_{-\infty}^{\infty} h(u, u-d)\, du. \tag{33}$$

As the estimation error equals the average of a large number of noise differences, we assume it is independent of D and Gaussian, with mean zero and variance σ^2. The price difference Q then has density:

$$\xi(q) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp(-\tfrac{1}{2}(x/\sigma)^2)\psi(q+x)\, dx\,. \tag{34}$$

The empirical challenge is to estimate the bivariate density h by using observations from the density ξ.

The noise term for an asset must be between $-w$ and w inclusive, when the asset's bid-ask spread is of width w. A natural strategy is to first specify distributions conditional on spread widths and then define h as a mixture of densities. Let $k = 1,2,...,K$ label pairs of spread widths denoted by $(u_{\max}^{(k)}, v_{\max}^{(k)})$, occurring with probabilities $p^{(k)}$. Then, for component densities $h^{(k)}$,

$$h(u,v) = \sum_{k=1}^{K} p^{(k)} h^{(k)}(u, v \,\big|\, u_{\max}^{(k)}, v_{\max}^{(k)}). \tag{35}$$

We identify $k = 1$ with the narrowest possible spreads, namely one tick wide for each asset when dark pool activity is excluded, and we anticipate that $p^{(1)}$ is near 1 based upon the exploratory data analysis (EDA).

The EDA shows that the distribution of Q depends on the minute j, may depend on the day t and is unlikely to be independent of variables such as realized variance. The most complicated specification estimated in this paper takes the form:

$$h_{t,j}(u,v|I_t) = \sum_{k=1}^{K} p^{(k)}(I_t, j)\, h^{(k)}(u,v \,|\, u_{\max}^{(k)}, v_{\max}^{(k)}) \tag{36}$$

with I_t representing auxiliary information such as a measure of contemporaneous or past volatility.

5.2 Three assumptions

We make assumptions about symmetry, uniformity and independence which apply to each component density $h^{(k)}$. To simplify the text, we omit *k* from the remaining notation and equations in Section 5.

Let $f(u)$ and $g(v)$ be the marginal densities of $h(u,v)$. It is pragmatic, and possibly also intuitive, to assume that all densities are symmetric, so $f(u) = f(-u),\ g(v) = g(-v)$ and $h(u,v) = h(-u,-v)$ for all u, v.

Suppose the latent price S^* falls within the unique[10] SPY bid-ask spread from S_{bid} to $S_{ask} = S_{bid} + u_{\max}$. As in Section 3.2, it is appropriate to assume that S^* is uniformly distributed between the feasible trade prices, i.e. the distribution of $U^+ = S_{ask} - S^*$ is uniform between 0 and $u_{\max}$. When U^+ has outcome u^+ the only possible outcomes for *U* are u^+ and $u^+ - u_{\max}$. Therefore, the symmetry and uniformity assumptions imply:

$$f(u) + f(u_{\max} - u) = \frac{1}{u_{\max}}, \quad \text{for } 0 \le u \le u_{\max}. \tag{37}$$

Then $f(u)$ is entirely determined by the function's values over the half-spread interval from 0 to $\frac{1}{2}u_{\max}$; similar remarks apply to $g(v)$.

[10] We ignore the possibility that the noise-free price exactly equals a feasible trade price because the assumed probability of this event is zero.

The bivariate distribution of $U^+ = S_{ask} - S^*$ and $V^+ = F_{ask} - F^*$, depends on properties of the basis $B = S^* - F^*$. For the reasons given in Section 3.2, we assume that U^+ and V^+ are independent. Any feasible outcome of $(U^+, V^+)'$ defines four possible outcomes for $(U,V)'$, located at the corners of a rectangle with the length of each side equal to either $u_{\max}$ or $v_{\max}$. Assuming independent, uniform distributions for U^+ and V^+, the *corner sum constraint* states that each set of rectangle corners appearing in the following equation has the same total density:

$$h(u,v) + h(u, v - v_{\max}) + h(u - u_{\max}, v) + h(u - u_{\max}, v - v_{\max}) = u_{\max}^{-1} v_{\max}^{-1},$$

$$\text{for } 0 \le u \le u_{\max} \text{ and } 0 \le v \le v_{\max}. \qquad (38)$$

This constraint severely restricts the specification of *h*, by excluding otherwise plausible specifications of dependence between *U* and *V*. In particular, the constraint rules out well-known copula functions.

5.3 Examples

As is appropriate for SPY and ES, we assume $u_{\max} \le v_{\max}$; for the numerical examples we select dollar price units and minimum spread widths, so $u_{\max} = 0.1$ and $v_{\max} = 0.25$.

The simplest example occurs when $(U,V)'$ has the *independent uniform* (IU) density:

$$h_{IU}(u,v) = \frac{1}{4u_{\max}v_{\max}}, \qquad \text{for } 0 \le |u| \le u_{\max} \text{ and } 0 \le |v| \le v_{\max}. \qquad (39)$$

The difference $D = U - V$ has density:

$$\begin{aligned} \psi_{IU}(d) &= \frac{1}{2v_{\max}} && \text{if } |d| \le v_{\max} - u_{\max}, \\ &= \frac{1}{4u_{\max}v_{\max}}(u_{\max} + v_{\max} - |d|) && \text{if } v_{\max} - u_{\max} < |d| \le v_{\max} + u_{\max}. \end{aligned} \qquad (40)$$

This distribution has expectation equal to zero, variance equal to $(u_{\max}^2 + v_{\max}^2)/3$ and mean absolute deviation (m.a.d.) equal to $(u_{\max}^2 + 3v_{\max}^2)/(6v_{\max})$.

Figure 5 includes the density $\psi_{IU}(d)$ for the minimum spread widths. This density is composed of three lines, passing through the coordinates $(d, \psi_{IU}(d))$ = (0, 2), ($\pm$0.15, 2) and ($\pm$0.35, 0). The standard deviation and m.a.d. are respectively 0.1555 and 0.1317. The m.a.d. exceeds most of the empirical estimates presented in Table 2 and displayed on Figures 1 and 2. From Figure 2 in particular, it is immediately seen that the IU density $\psi_{IU}(d)$ has too much dispersion to explain empirical price differences.

To reduce the dispersion of the difference *D* we must select non-uniform noise distributions and/or permit positive dependence between the two noise terms *U* and *V*. The simplest, parametric, candidate density for *U* is a general linear function of $|u|$:

$$f_P(u|\lambda) = \frac{1}{2u_{\max}}(1 + \lambda - 2\lambda\frac{|u|}{u_{\max}}) \qquad \text{for } 0 \le |u| \le u_{\max}. \tag{41}$$

This density is linear between $f_P(0|\lambda) = (1+\lambda)/(2u_{\max})$ and $f_P(\pm u_{\max}|\lambda) = (1-\lambda)/(2u_{\max})$, with $-1 \le \lambda < 1$. Its shape is a pentagon[11], hence the subscript *P* in its definition. The variance is equal to $\frac{1}{6}(2-\lambda)u_{\max}^2$. Trades are more likely at the end of the spread nearest the noise-free price when λ is positive. Density (41) occurs when (11) applies, i.e. when the chance of a trade at the ask price is a linear function of U^+.

The *independent pentagon* (IP) density for $(U,V)'$ is

$$h_{IP}(u,v|\lambda,\eta) = f_P(u|\lambda)\, f_P(v|\eta) \quad \text{for } 0 \le |u| \le u_{\max}, 0 \le |v| \le v_{\max}. \tag{42}$$

[11] On a density plot, the shape of the uniform density is often called a rectangle. The shape of the general pentagon density has two vertical sides, two sloping sides and a fifth side along the horizontal axis.

As U^+ is independent of V^+, the corner sum constraint applies. From (33) and (41), the density of $D = U - V$ is then the integral of a quadratic function of *u*, whose coefficients depend on the signs of *u* and $u - d$. Consequently, the IP density $\psi_{IP}(d)$ is a cubic function of *d*, with coefficients varying across the support of the distribution; computational methods for these coefficients are provided in Appendix C. Figure 5 includes a plot of the IP density when $\lambda = \eta = 0.3$; The variance of *D* for the IP distribution is $\frac{1}{6}(2-\lambda)u_{\max}^2 + \frac{1}{6}(2-\eta)v_{\max}^2$, so the standard deviation of the plotted density is 0.1433, which is close to the sample value of 0.142 for the mid-day period from 11:00 to 15:00.

Positive dependence between *U* and *V* will occur if contemporaneous trades for the two markets are more likely to be at either two bid prices or two ask prices than at one bid and one ask. Let $S(uv)$ be the usual sign function, so

$$\begin{aligned} S(uv) &= 1 \quad \text{if } uv > 0, \\ &= 0 \quad \text{if } uv = 0 \text{ and} \\ &= -1 \quad \text{if } uv < 0. \end{aligned} \tag{43}$$

Then the simplest parametric example of a dependent density for $(U,V)'$ is *dependent uniform* (DU):

$$h_{DU}(u,v|\omega) = \frac{1+\omega S(uv)}{4u_{\max}v_{\max}}, \qquad 0 \le |u| \le u_{\max}, 0 \le |v| \le v_{\max}, \tag{44}$$

with $|\omega| \le 1$. This density trivially satisfies the corner sum constraint because (almost surely) each of the four corners in (38) is in a different quadrant. The variance of $D = U - V$ now equals $\frac{1}{3}(u_{\max}^2 + v_{\max}^2) - \frac{1}{2}\omega u_{\max}v_{\max}$. Let $\psi_{DU}(d)$ denote the corresponding density for *D*. Then $\psi_{DU}(d) - \psi_{IU}(d)$ is made up of line segments, which can be computed from equations given in Appendix C. For our SPY and ES example, $\psi_{DU}(d)$ is composed of eight lines, passing through the coordinates $(d, \psi_{DU}(d)) =$ $(0,\ 2+2\omega)$, $(\pm 0.1,\ 2)$, $(\pm 0.15,\ 2)$, $(\pm$

0.25, $1-\omega$) and (± 0.35, 0). The density when $\omega = 0.3$ is shown on Figure 5 and then the standard deviation of *D* is 0.1429.

The density of price differences *Q* is similar to that of noise differences *D* when the standard deviation σ of the basis estimation error is small. Figure 5 compares the distributions of *D* and *Q* for IP densities when σ is set to a high level compared with the estimates given in Section 4.3.5. The curve labeled 'Q: IP 0.3' is for parameter values $\lambda = \eta = 0.3$ and $\sigma = 0.04$. It stays close to the curve 'D: IP 0.3' except when *d* is beyond ± 0.3.

5.4 A general specification

General parametric specifications of the bivariate density *h* can be constructed from linear combinations of products of univariate densities plus a term which creates dependence between *U* and *V*. For example, combining the general IP density with the special case when $\lambda = \eta = -1$ makes *h* a general linear function of $|u|$, $|v|$ and $|uv|$. Dependence can be included by adding a residual function which sums to zero across all sets of rectangle corners appearing in the corner sum constraint; three simple choices are functions proportional to $S(uv)$, $S(uv)|u|$ and $S(uv)|v|$.

We focus on the general parametric density defined by:

$$h_G(u,v|\Theta) = \frac{1}{4u_{\max}v_{\max}}[\theta_0 - \theta_1\frac{|u|}{u_{\max}} - \theta_2\frac{|v|}{v_{\max}} + \theta_3\frac{|uv|}{u_{\max}v_{\max}} + S(uv)[\theta_4 + \theta_5\frac{|u|}{u_{\max}} + \theta_6\frac{|v|}{v_{\max}}]]. \quad (45)$$

The parameters must be constrained to ensure *h* integrates to 1, so:

$$\theta_0 - \tfrac{1}{2}\theta_1 - \tfrac{1}{2}\theta_2 + \tfrac{1}{4}\theta_3 = 1. \quad (46)$$

We apply two sets of constraints when we estimate the parameters in Section 6. First, we do not permit the density to increase as the distance from the origin increases, i.e. we require $\partial h/\partial u \le 0$ for $u > 0$, $\partial h/\partial u \ge 0$ for $u < 0$ and likewise for $\partial h/\partial v$; then:

$$\theta_1 \ge \max(|\theta_5|, \theta_3 + |\theta_5|) \text{ and } \theta_2 \ge \max(|\theta_6|, \theta_2 + |\theta_6|). \tag{47}$$

Second, we require outcomes with identical signs for *u* and *v* to be at least as likely as those with different signs, i.e. $h(u,v) = h(-u,-v) \ge h(-u,v) = h(u,-v)$ for all $u, v \ge 0$, which implies:

$$\min(\theta_4, \theta_4 + \theta_5, \theta_4 + \theta_6, \theta_4 + \theta_5 + \theta_6) \ge 0. \tag{48}$$

When the above constraints apply, the function *h* is always non-negative when $h(u_{\max}, -v_{\max}) \ge 0$, i.e. when

$$\theta_0 - \theta_1 - \theta_2 + \theta_3 - \theta_4 - \theta_5 - \theta_6 \ge 0. \tag{49}$$

6 Estimated distributions for discrete-price microstructure noise

The general distributions of Section 5 are parametric and we now estimate their parameters within a likelihood framework, outlined in Section 6.1. Initially we assume stationary distributions and we do not condition on the level of index volatility. The results obtained in Section 6.2 for the mid-day trading period from 11:00 to 15:00 show that it is appropriate to simplify the bivariate density of microstructure noise by removing some parameters. We do this and then include the time-of-day and index volatility in the density specification, to provide estimates in Section 6.3 for the longer trading period from 09:31 to 15:59.

6.1 Methods

Distributions are estimated from sets of adjusted price differences, $q_{t,j} = S_{t,j} - F_{t,j} - \hat{B}_t$, for days t and minutes j. We provide results for subsets of the minutes during a set of trading days T; a subset of minutes is denoted by J and then the basis estimate $\hat{B}_t$ equals the average of the price differences $S_{t,j} - F_{t,j}$ across $j \in J$.

Each price difference equals a noise difference minus an independent basis estimation error: $q_{t,j} = d_{t,j} - \varepsilon_t$. The density of the noise differences is $\psi(d \mid \Theta, I_{t,j})$ which can depend on auxiliary information $I_{t,j}$, while $\varepsilon_t \sim N(0, \sigma^2)$. We adopt a two-stage estimation strategy: an estimate $\hat{\sigma}$ of σ is obtained from the sample variance and autocorrelations of the price differences (as in Section 4.3.5) and then the vector Θ of noise parameters is estimated by maximizing a log-likelihood criterion.

The log-likelihood criterion is defined for days $t \in T$ and sets of minutes $A_t \subseteq J$ by:

$$\log L(\Theta, \hat{\sigma}) = \sum_{t \varepsilon T} \log L_t(\Theta, \hat{\sigma}) \qquad \text{with}$$

$$L_t(\Theta, \hat{\sigma}) = \frac{1}{\hat{\sigma}\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp(-x^2 / 2\hat{\sigma}^2) \prod_{j \varepsilon A_t} \psi(q_{t,j} + x \mid \Theta, I_{t,j})\, dx . \qquad (50)$$

The function L is not an exact likelihood function, because the results in Section 4.3.5 show that the adjusted price differences are weakly autocorrelated. However, the maximum correlation between pairs of different price differences is less than 0.15 so it is plausible to estimate and compare models using likelihood methods, particularly if inferences are based upon very low significance levels.

We progress to preferred specifications for the bivariate density of microstructure noise $h_{t,j}(u, v \mid I_{t,j})$ by using results for simpler specifications to guide choices for more general specifications. The densities $h_{t,j}$ are linear combinations of K component densities

$h_{t,j}^{(k)}$, with each component having different maximum levels for *u* and *v* based upon the assumed widths of bid-ask spreads. We commence with one state ($K = 1$) and no auxiliary information, then add more states and finally consider the relevance of selected information $I_{t,j}$.

As $h_{t,j}^{(k)}(u,v)$ is zero when $|u| > u_{\max}^{(k)}$ or $|v| > v_{\max}^{(k)}$, it follows that the density $\psi(d)$ is zero whenever $|d| > q_{\max} = \max(u_{\max}^{(k)} + v_{\max}^{(k)}, 1 \le k \le K)$. Consequently, we exclude price differences from the likelihood criterion when they are outside $\pm q_{\max}$. Thus A_t contains all minutes *j* for which $j \in J$ and $|q_{t,j}| \le q_{\max}$. Some data censoring is inevitable for practical values of *K* and $q_{\max}$, otherwise there can be days in the sample for which the product term in (50) is zero so that $L_t = 0$. Some bias may occur in the parameter estimates because some data are excluded, although we expect any bias to be small as all estimates are based upon including more than 99% of the sample minutes.

All component densities are defined by (45) in Section 5.4 and they are all estimated with the constraints listed in (46)-(49). The noise difference density $\psi(d)$ is calculated exactly from (33) and the equations in Appendix C. The integral in (50) is calculated numerically.

We compare density specifications by comparing maximum values of log(*L*) and goodness-of-fit measures X^2 evaluated at the MLEs. To calculate X^2 we count the observed number of minutes O_i having $j \in J$, $|q_{t,j}| \le q_{\max}$ and $0.05(i-1) \le |q_{t,j}| < 0.05i$, for $1 \le i \le 7$, and obtain the expected number E_i when Θ equals the MLE. The fit for $|q| \le 0.35$ is then summarized by:

$$X^2 = \sum (O_i - E_i)^2 / E_i \,. \tag{51}$$

6.2 Estimates for the mid-day period

Distributions are initially estimated from prices recorded between 11:00 and 15:00 inclusive. We do so because Figure 2 shows there is less variation in the mean absolute deviations of the price differences during this mid-day period, so ignoring any intraday density variation is then more reasonable. The mid-day period also has a higher percentage of price differences inside ± 0.35, which enhances the usefulness of one-state specifications.

One state

Our first results are for one state, so $K = 1$, $u_{\max}^{(1)} = 0.1$, $v_{\max}^{(1)} = 0.25$ and $q_{\max} = 0.35$; then 0.43% of the minutes are excluded for 2010, 0.95% for 2011 and 0.51% for 2012. The parameter vector is $\Theta = (\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6)'$, with the terms θ_j defined by (45).

Panel A of Table 4 provides estimates of Θ and σ for each year, for three special cases and then for all seven parameters. The first special case is the independent uniform density, which has by far the lowest log-likelihoods and the highest goodness-of-fit statistics. This confirms that it is necessary to investigate nonuniform marginal densities and/or dependence between the two noise terms.

The most general estimates show that $\hat{\theta}_2$ exceeds 0.5 for all three years, that two values of $\hat{\theta}_6$ exceed 0.3 and that the maximum values of $|\hat{\theta}_1|$, $|\hat{\theta}_3|$, $|\hat{\theta}_4|$ and $|\hat{\theta}_5|$ are respectively 0.05, 0.05, 0.002 and 0.002. These facts motivate estimating the two-parameter special case when $\theta_1 = \theta_3 = \theta_4 = \theta_5 = 0$, which produces values of the log-likelihood near to the best for a general Θ. Table 4 lists values for $\log(L)$ minus the corresponding values for the two-parameter special case. The three log-likelihood differences are 1.18, 0.12 and 0 which provide no evidence to support including the additional four parameters in the

bivariate density function. Table 4 also shows that the additional restriction $\theta_6 = 0$ is inappropriate, as then double the reduction in log(L) always exceeds 39, which is far above the 0.1% critical point of χ_1^2, which equals 10.8. The standard errors of $\hat{\theta}_2$ and $\hat{\theta}_6$ are small for the two-parameter special case; these are approximately 0.031 for $\hat{\theta}_2$ and 0.017 for $\hat{\theta}_6$ when the i.i.d. assumption is made, and they increase to approximately 0.04 and 0.025 for block bootstrap estimates with the blocks defined by the trading days. The values of X^2 are 20.39, 30.18 and 70.44 for the preferred specification, which are rather large when comparisons are made with the χ_3^2 distribution; we defer graphical comparisons of empirical and fitted densities until after the results for four-state densities.

From Panel A of Table 4 we conclude that a parsimonious, single-state specification of the bivariate density is possible with only two parameters. Replacing θ_2 by γ and θ_6 by δ, the density is:

$$h_{S2}(u,v|\gamma,\delta,u_{\max},v_{\max}) = \frac{1}{4u_{\max}v_{\max}}[1+\tfrac{1}{2}\gamma-(\gamma-\delta\, S(uv))\frac{|v|}{v_{\max}}], \qquad (52)$$

with the label S2 indicating that this is a two-parameter special case. The constraints now simplify to $0 \le \delta \le \gamma$ and $\frac{1}{2}\gamma + \delta \le 1$. The marginal distribution of U, whose support is relatively narrow, is then uniform, while the marginal for V is the pentagon density $f_P(v|\eta)$ (see (41)) with $\eta = \gamma/2$ and with relatively wider support.

The two parameters can be interpreted separately. First, γ determines the unconditional probability that an ES trade occurs at the end of the bid-ask spread nearest to the noise-free ES price; this probability equals $\frac{1}{2}(1+\frac{1}{4}\gamma)$, which varies from 57% to 60% for the estimates in Table 4, with standard errors equal to 0.5% Second, δ controls the unconditional probability that both markets trade at the same end of the spread (either at two

bids or at two asks); this equals $\frac{1}{2}(1+\frac{1}{2}\delta)$ and its estimates vary from 53% to 58% (approximate standard errors are 0.6%). The covariance between U and V equals $\delta u_{\max} v_{\max}/6$ and the correlation is $\delta/\sqrt{4-\gamma}$; our correlation estimates vary from 0.06 to 0.17.

Two states

The range of feasible price differences can be widened from ± 0.35 to ± 0.45 by adding a second state with $u_{\max}^{(2)} = 0.2$ and $v_{\max}^{(2)} = 0.25$, which permits the possibility of a two-tick spread for SPY. Now $q_{\max} = 0.45$ and approximately three-quarters of the minutes previously excluded are now included; only 0.10% of the minutes remain excluded for 2010, with 0.28% for 2011 and 0.12% for 2012.

We estimate the following combination of S2-densities:

$$h(u,v|\Theta) = p^{(1)} h_{S2}(u,v|\gamma^{(1)},\delta^{(1)},0.1,0.25) + p^{(2)} h_{S2}(u,v|\gamma^{(2)},\delta^{(2)},0.2,0.25), \qquad (53)$$

with $\Theta = (p^{(1)},\gamma^{(1)},\delta^{(1)},\gamma^{(2)},\delta^{(2)})'$ and $p^{(2)} = 1 - p^{(1)}$. Maximizing the log-likelihood over all five parameters, we find that all estimates $\hat{\gamma}^{(2)}$ and $\hat{\delta}^{(2)}$ are either zero or small as can be seen in Panel B of Table 4. The maximum log-likelihood is reduced by less than 0.1, for each year, when the restriction $\gamma^{(2)} = \delta^{(2)} = 0$ is imposed so an appropriate two-state model is provided by the three-parameter special case $\Theta = (p^{(1)},\gamma^{(1)},\delta^{(1)},0,0)'$.

The estimates of $\gamma^{(1)}$ are similar for the one and two-state specifications, and equal 0.68, 0.79 and 0.55 when two states are estimated. The estimates of $\delta^{(1)}$ are higher when there are two states and these equal 0.39, 0.24 and 0.40. The estimated probabilities of a two-tick SPY spread are similar for 2010 and 2012, at 0.060 and 0.072, while the estimate is

notably higher in 2011 at 0.129. The higher estimate for 2011 reflects the higher dispersion of price differences in that year, seen in Table 2 and on Figure 1.

Four states

The possibility of a two-tick spread for ES is included by setting $v_{\max}^{(3)} = v_{\max}^{(4)} = 0.5$, with $u_{\max}^{(3)} = 0.1$ and $u_{\max}^{(4)} = 0.2$. Now $q_{\max} = 0.7$ which only excludes 0.027%, 0.075% and 0.012% of the minutes in 2010, 2011 and 2012 respectively.

The two-state results support selecting independent uniform densities whenever one of the spreads is more than one tick wide. Consequently, we now estimate parameters for the following combination of S2 and IU densities:

$$h(u,v|\Theta) = p^{(1)} h_{S2}(u,v|\gamma^{(1)},\delta^{(1)},u_{\max}^{(1)},v_{\max}^{(1)}) + \sum_{k=2}^{4} p^{(k)} h_{IU}(u,v|u_{\max}^{(k)},v_{\max}^{(k)}). \qquad (54)$$

with $\Theta = (p^{(1)}, p^{(2)}, p^{(3)}, \gamma^{(1)}, \delta^{(1)})'$ and $p^{(4)} = 1 - p^{(1)} - p^{(2)} - p^{(3)}$.

Panel C of Table 4 shows that the estimates of $\gamma^{(1)}$ and $\delta^{(1)}$ do not change much when the two additional states are included in the bivariate density specification. Two of the estimates of $p^{(3)}$ are zero, which is not surprising as the third state has an ES spread five times as wide as the SPY spread.[12] The probability of a two-tick ES spread is $p^{(3)} + p^{(4)}$, estimated to be 0.44% in 2010, 1.21% in 2011 and 0.79% in 2012. Compared with the two-state estimates, $\hat{p}^{(1)}$ is slightly higher and $\hat{p}^{(2)}$ is notably lower.

It is inevitably difficult to obtain accurate estimates of four state probabilities when three of them are small. Indicative standard errors (s.e.) have been calculated when $p^{(3)}$ is constrained to be zero and the i.i.d. assumption is made. We find that $\hat{p}^{(1)}$ and $\hat{p}^{(2)}$ are

[12] In Section 6.3 we assume the third state probability is zero. Applying this constraint reduces the maximum log-likelihood for the 2012 data by 2.40.

almost perfectly negatively correlated and their s.e. range from 0.004 to 0.006, while the s.e. for $\hat{p}^{(4)}$ are from 0.0007 to 0.0011. There is little correlation between either $\hat{\gamma}^{(1)}$ or $\hat{\delta}^{(1)}$ and either $\hat{p}^{(2)}$ or $\hat{p}^{(4)}$, and the s.e. of $\hat{\gamma}^{(1)}$ and $\hat{\delta}^{(1)}$ are respectively all approximately equal to 0.033 and 0.019.

Density comparisons for adjusted price differences

Plots of the fitted one, two and four-state densities for the adjusted price differences q all show a close agreement with kernel densities estimated from the data provided by all mid-day minutes in a selected year. Our kernel densities first apply a Gaussian kernel, with bandwidth equal to 0.03, to give $\mu(q)$ and are then converted to the symmetric function $(\mu(q)+\mu(-q))/2$. Figures 6a, 6b and 6c compare the four-state and kernel densities, for $0 \le q \le 0.4$; over this range the visual comparison is essentially identical for two-state densities and it is very similar for one-state densities as far as $q = 0.35$. The fitted densities tend to be slightly higher than the kernel densities for $0 < q < 0.2$ and slightly lower for $0.2 < q < 0.35$. Figures 7a, 7b and 7c show a reasonable match between the four-state and kernel densities in the tail region $0.4 \le q \le 0.7$, where the densities are very small. The densities are similar across years, although the tails are fatter for 2011 which can be seen from the higher values on the vertical axis of Figure 7b than on Figures 7a and 7c.

6.3 Estimates for the primary trading period

The primary trading period is from 09:30 to 16:15 inclusive. We exclude the period after 16:00 because Figure 2 shows there is substantially more noise after Wall Street closes. For the same reason we also exclude the minute marks at 09:30, 10:00 and 16:00. Thus the set J

contains the 388 integers j in the list $\{1,...,29,31,...,389\}$, corresponding to 09:31 until 09:59 and 10:01 until 15:59 inclusive.

Based upon the exploratory data analysis in Section 4.3.3, we expect the state probabilities to vary across the day with two-tick spreads more likely early and late in the trading day. We use a quadratic function of j to describe this intraday time-variation. From Section 4.3.4, we also expect the probabilities of two-tick spreads to be increasing functions of realized volatility. We provide presults when these probabilities depend on linear functions of SPY realized volatility, defined simply as the sum of squared, one-minute, percentage returns aggregated over the periods from 09:31 to 09:59 and 10:01 to 15:59.

Guided by the results for the mid-day period, three states are defined by (1) both spreads are one tick wide, (2) the SPY spread width is two ticks and the ES width is one tick, (3) both spreads are two ticks wide; thus $K = 3$, $u_{\max}^{(1)} = 0.1$, $u_{\max}^{(2)} = u_{\max}^{(3)} = 0.2$, $v_{\max}^{(1)} = v_{\max}^{(2)} = 0.25$, $v_{\max}^{(3)} = 0.5$ and $b_{\max} = 0.7$. The percentages of minutes excluded are 0.041% in 2010, 0.127% in 2011 and 0.019% in 2012.

Following (54) and the results in Table 4, the bivariate density for the two noise terms is a linear combination of one S2-density and two IU-densities with the state probabilities depending on the minute j and the day's realized variance, RV_t:

$$\begin{aligned} h_{t,j}(u,v\,|\,\Theta, RV_t) = {} & p^{(1)}(RV_t, j)\, h_{S2}(u,v\,|\,\gamma,\delta,u_{\max}^{(1)},v_{\max}^{(1)}) \\ & + \sum_{k=2}^{3} p^{(k)}(RV_t, j)\, h_{IU}(u,v\,|\,u_{\max}^{(k)},v_{\max}^{(k)}). \end{aligned} \tag{55}$$

Our choices for the state probabilities are:

$$p^{(k)}(RV_t, j) = p_{\min}^{(k)}(1+\alpha RV_t)(1+\beta(\frac{j-j_{\min}}{194})^2), \qquad k = 2,3 \tag{56}$$

and $$p^{(1)}(RV_t, j) = 1 - p^{(2)}(RV_t, j) - p^{(3)}(RV_t, j).$$

The parameter vector is $\Theta = (\gamma, \delta, p_{\min}^{(2)}, p_{\min}^{(3)}, \alpha, \beta, j_{\min})'$ and all parameters are non-negative. The terms $p_{\min}^{(k)}$ are the minimum probabilities for states 2 and 3, which occur when firstly $\alpha = 0$ or $RV_t = 0$ and secondly $\beta = 0$ or $j = j_{\min}$. The same multiplicative factors are applied to the two terms $p_{\min}^{(k)}$ to avoid estimating more than seven parameters. To prevent very high values of RV being overinfluential, all seven values above 5 have been truncated to 5; all the truncated values occur in the second half of 2011.

Table 5 contains the parameter estimates when all seven parameters are estimated and also for three special cases, which switch off one or both of the time-varying multipliers. The final column shows the adjusted log-likelihood *AL*, defined as the maximum log-likelihood in excess of the maximum when there is no time-variation, i.e. when $\alpha = \beta = 0$. We see large values of *AL* confirming that the state probabilities vary though time. When $\alpha = 0$, the log-likelihood is always more than 35 below the global log-likelihood for the same year, while for $\beta = 0$ the reduction is always more than 89. We conclude that both the time of the day and the level of price volatility influence the probabilities of spreads wider than one tick.

The average probabilities across the primary trading period are estimated as:

	State 1	State 2	State 3
2010	0.930	0.063	0.007
2011	0.807	0.165	0.028
2012	0.886	0.100	0.014

The estimated times when the first-state probabilities are maximized are 13:14, 13:24 and 13:32, respectively for 2010, 2011 and 2012. They are all at their minimum level at 09:31, when their averages are estimated to be:

	State 1	State 2	State 3
2010	0.841	0.144	0.015

2011	0.511	0.418	0.072
2012	0.683	0.277	0.040

We note that the estimates of α, which multiplies RV_t in (56), are highest for the 2011 data when realized variance is also generally higher.

Discrete-price microstructure noise densities

Finally, we present some representative estimated densities for discrete-price microstructure noise. For our preferred bivariate density, the marginal density $f(u)$ of the SPY noise variable for the exchange traded fund, $U = S - S^*$, is a weighted combination of two uniform densities with the weights depending on the clock and on index volatility. The first uniform density equals $1/0.2$ when $|u| \le 0.1$ and the spread is one tick wide, while the second for two-tick spreads is $1/0.4$ when $|u| \le 0.2$. The density of U is thus, for $|u| \le 0.2$:

$$f(u) = 5p^{(1)}1_{|u|\le 0.1} + 2.5(1 - p^{(1)})1_{|u|\le 0.2}. \tag{57}$$

For the e-mini futures contracts ES, the marginal density $g(v)$ of the ES noise variable, $V = F - F^*$, is a weighted combination of a polygon density and two uniform densities. The polygon density is used when both spreads are one tick wide. The density of V becomes, for $|v| \le 0.5$:

$$\begin{aligned} g(v|\gamma) &= 2p^{(1)}[1 + \tfrac{1}{2}\gamma - \gamma\left|\frac{v}{0.25}\right|]1_{|v|\le 0.25} + 2(1 - p^{(1)} - p^{(3)})1_{|v|\le 0.25} + p^{(3)}1_{|v|\le 0.5} \\ &= \{p^{(1)}\gamma[1 - 8|v|] + 2(1 - p^{(3)})\}1_{|v|\le 0.25} + p^{(3)}1_{|v|\le 0.5}. \end{aligned} \tag{58}$$

Figures 8a, 8b and 8c show the estimated marginal densities for each year when the state probabilities equal the average values presented above and when γ equals the values in the final row for each year in Table 5. These diagrams show two empirical conclusions: firstly that ES has more discrete-price noise variation than SPY, reflecting their different tick

sizes, and secondly that ES trades are more likely to occur at the end of the spread nearest to the noise-free price.

7 The magnitude and persistence of residual microstructure noise

There is no obvious method for estimating the distribution of the residual MN component, denoted $M_{t,j}$, so we conclude the empirical analysis by only estimating the variance and autocorrelations of the residual MN. We obtain these estimates by matching theoretical and empirical moments for the auto- and cross-covariances of the price changes $\Delta S_{t,j}$ and $\Delta F_{t,j}$ during the mid-day period from 11:00 to 15:00 inclusive.

Initially we match moments from lag 2 onwards, because the exploratory data analysis shows that the auto- and cross-covariances are then dominated by contributions from the residual MN. We have:

$$\begin{aligned}\operatorname{cov}(\Delta S_{t,j}, \Delta S_{t,j+\tau}) \cong \operatorname{cov}(\Delta F_{t,j}, \Delta F_{t,j+\tau}) \cong \operatorname{cov}(\Delta S_{t,j}, \Delta F_{t,j+\tau}) \cong \operatorname{cov}(\Delta F_{t,j}, \Delta S_{t,j+\tau}) \\ \cong (2\rho_{\tau,M} - \rho_{\tau-1,M} - \rho_{\tau+1,M}) \operatorname{var}(M_{t,j}), \quad \tau \geq 2.\end{aligned} \tag{59}$$

These covariances are negative for the simplest credible specification for $\{M_{t,j}\}$, which is an AR(1) process with positive autocorrelations ϕ_M^τ. Then

$$\operatorname{cov}(\Delta S_{t,j}, \Delta S_{t,j+\tau}) \cong .... \cong -(1-\phi_M)^2 \phi_M^{\tau-1} \operatorname{var}(M_{t,j}), \quad \tau \geq 2. \tag{60}$$

From Table 3 and Figure 4, the empirical auto- and cross-covariances are negative from lags 2 to 10 inclusive and their values are very similar for all combinations of either $\Delta S_{t,j}$ or $\Delta F_{t,j}$ with either $\Delta S_{t,j+\tau}$ or $\Delta F_{t,j+\tau}$ when $\tau \geq 2$. Let $\widehat{C}_\tau$ denote the average across the four sample values at lag τ, which equals the autocovariance of $\Delta(S_{t,j} + F_{t,j})/2$ at lag τ; also let $C_\tau = -(1-\phi_M)^2 \phi_M^{\tau-1} \operatorname{var}(M_{t,j})$. Method of moment (MM) estimates of

$\text{var}(M_{t,j})$ and ϕ_M are provided by minimizing $W_2 = \sum(\hat{C}_\tau - C_\tau)^2$. Using lags 2 to 20 inclusive, and all three years of mid-day prices (measured in cents), we estimate the variance and autoregressive parameter to be respectively equal to 963 and 0.807.

Potentially more precise estimates, as well as standard errors, can be obtained by the generalized method of moments (GMM). Following the notation and equations of Hamilton (1994, page 416), we minimize the quantity $g'\hat{S}^{-1}g$ where g is a 19×1 vector, with $g_{\tau-1} = \hat{C}_\tau - C_\tau, \tau \geq 2$, and $\hat{S}$ is an appropriate 19×19 matrix; we use the MM estimates to evaluate $\hat{S}$. For all three years the GMM estimate of $(\text{var}(M), \phi_M)'$ equals $(927, 0.813)'$ which is similar to the MM estimate. The estimated GMM standard errors are 471 and 0.071 and the estimated correlation between the two parameter estimates is a substantial 0.92. Year by year, the GMM estimates are $(508, 0.832)'$, $(2025, 0.827)'$ and $(254, 0.673)'$, respectively for 2010, 2011 and 2012. The standard diagnostic test compares the minimum value of $Tg'\hat{S}^{-1}g$ with the χ^2_{17} distribution, when *T* time periods are used. Our test values range from 14.85 to 19.06, none of which provides evidence of model mis-specification; we observe that the minimum value of 16.75 for all three years is less than half of the value of $Tg'\hat{S}^{-1}g$ when $C_\tau \equiv 0$, namely 41.51.

We note that similar estimates of $(\text{var}(M), \phi_M)'$ are obtained when first-lag covariances are also used, so that moments are matched from lags 1 to 20 inclusive, but it is then necessary to insert discrete-price MN parameter estimates into the calculations of the theoretical moments at lag 1. These are:

$$\begin{aligned} C_{SS} &= \text{cov}(\Delta S_{t,j}, \Delta S_{t,j+1}) = -(1-\phi_M)^2\,\text{var}(M_{t,j}) - (1 - 2A_U\phi_U + A_U\phi_U^2)\,\text{var}(U_{t,j}), \\ C_{FF} &= \text{cov}(\Delta F_{t,j}, \Delta F_{t,j+1}) = -(1-\phi_M)^2\,\text{var}(M_{t,j}) - (1 - 2A_V\phi_V + A_V\phi_V^2)\,\text{var}(V_{t,j}), \\ C_{SF} &= \text{cov}(\Delta S_{t,j}, \Delta F_{t,j\pm 1}) = -(1-\phi_M)^2\,\text{var}(M_{t,j}) - \text{cov}(U_{t,j}, V_{t,j}), \end{aligned} \qquad (61)$$

assuming the autocorrelations and cross-correlations of discrete-price MN are $\rho_{\tau,U} = A_U \phi_U^\tau$ and $\rho_{\tau,V} = A_V \phi_V^\tau$ for $\tau \geq 1$ and $\gamma_\tau = \text{cor}(U_{t,j}, V_{t,j+\tau}) = 0$ for $\tau \neq 0$. A second set of MM estimates for $\text{var}(M_{t,j})$ and ϕ_M are given by minimizing $W_1 = W_2 + (\hat{C}_{SS} - C_{SS})^2 + (\hat{C}_{FF} - C_{FF})^2 + (\hat{C}_{SF} - C_{SF})^2$, with $A_U = A_V = 0.08$ and $\phi_U = \phi_V = 0.975$ (from Section 4.3.5) and with $\text{var}(U_{t,j}) = 40.2$, $\text{var}(V_{t,j}) = 178.2$ and $\text{cor}(U_{t,j}, V_{t,j}) = 0.15$ (from Table 4, Panel C). These estimates are 1052 for $\text{var}(M_{t,j})$ and 0.827 for ϕ_M, which are very near the estimates given by minimizing W_2.

Using the lag 1 information also permits identification of an additional autocorrelation parameter for residual MN. Supposing $\rho_{\tau,M} = A_M \phi_M^\tau$, minimizing W_1 provides $\hat{A}_M = 1.006$ which supports the simple AR(1) specification for residual MN.

The GMM estimate of persistence for residual MN shocks during the mid-day period, $\hat{\phi}_M = 0.813$, corresponds to a short half-life equal to 3.35 minutes; from a symmetric 95% confidence interval for $\hat{\phi}_M$, the 95% interval for the half-life is found to be from 1.76 to 14.1 minutes. Note that it is conceivable that there also exist highly persistent components of residual MN, which cannot be identified from our high-frequency data.

The standard deviation of residual MN has a GMM estimate equal to 30.4 cents, which is notably higher than the estimated standard deviations of discrete-price MN. A direct comparison of standard deviations does not, however, totally clarify the relative importance of the residual and discrete-price components. High-frequency financial econometrics research makes frequent use of realized variance measures calculated from returns, which are often biased because of MN. As returns are essentially price changes divided by prices, the key comparison is between variances of changes such as between $\text{var}(M_{t,j} - M_{t,j-1})$ and

$\text{var}(V_{t,j} - V_{t,j-1})$. For *h*-minute returns, we may thus compare $2(1-\phi_M^h)\,\text{var}(M_{t,j})$ with $2(1-A_V\phi_V^h)\,\text{var}(V_{t,j})$. The former quantity disappears as $h \to 0$, while the latter quantity does not because $A_V < 1$. With $\phi_M = 0.8$, $A_V = 0.08$ and $\phi_V = 0.975$, we estimate the two terms are the same for futures when *h* is approximately 0.94, so that discrete-price MN is more influential for return intervals shorter than 56 seconds; it is more influential for the spot asset for intervals shorter than 12 seconds if we assume $A_U = A_V$ and $\phi_U = \phi_V$. We do not claim that these estimates (12 and 56 seconds) are accurate.

8 Conclusions

Microstructure noise (MN) can be modelled as the sum of discrete-price and residual components. By studying the differences between exchange traded fund prices and futures prices, we have obtained new results about the distribution of the discrete-price MN created by bid/ask spreads and minimum price changes. The distribution is time-varying and depends on the minute of the day, on the year studied and on index volatility. For our data, we can also infer some properties of the residual MN, finding that it contributes more to measures of price variation than discrete-price MN for the popular sampling frequency of five minutes.

The S&P 500 ETF trading under the ticker symbol SPY has a spread equal to 10 cents when the spread is one tick wide and the ETF prices are multiplied by 10 to match the level of the replicated index. The probability of a two-tick spread is relatively high before 10:00 and after 15:00 and it increases as volatility increases. Our average estimate of the one-tick probability is 87% for SPY, with an average 13% chance of a two-tick spread. The simplest credible distribution for the discrete-price component of SPY microstructure noise, conditional on the spread width, is uniform which our data supports.

S&P 500 e-mini futures have a minimum spread of 25 cents and consequently there is more discrete-price noise in futures prices than in the ETF prices. One-tick and two-tick spreads are estimated to have average frequencies respectively equal to 98.4% and 1.6%. Uniform distributions have more variation than our fitted pentagon distributions, whose representative densities are shown on Figure 8. The higher density values for smaller (in magnitude) noise levels are consistent with trade being more likely at the end of the spread which is nearest to the latent price which would occur if there were no spreads and no minimum price changes. The estimated probability of trade at the "better value" price averages 58%, both mid-day from 11:00 to 15:00 and across the longer trading period from 09:31 to 15:59.

There is some dependence between ETF discrete-price noise and futures discrete-price noise, attributable to dependence between the orders flowing to the two markets. As buying pressure relative to selling pressure increases at both markets there will be a higher chance that both markets trade at ask prices, while if selling pressure dominates more trades occur at bid prices. Our average estimate of the chance of trades at the same ends of the spreads (two bids or two asks) is 58% mid-day and 56% for the longer trading period. The dependence between discrete-price noise across the markets is thus weak, with average correlation estimates of 15% mid-day and 11% all-day.

There is persistence in the discrete-price component for futures, which reflects persistence in the latent probability that a trade occurs at the ask price. Our data supports the probability process having mean, standard deviation and persistence half-life respectively equal to 0.5, approximately 0.22 and approximately 30 minutes. We may conjecture similar results for the spot component, but cannot obtain estimates because the spot component is far less variable than the futures component.

The residual MN component must be very similar for spot and futures because these asset prices are constrained by the impossibility of arbitrage profits. We find no evidence to challenge the assumption of identical residual components for the two assets. The estimated standard deviation of residual MN, at 30 cents during the mid-day period, is high relative to the standard deviations of spot and futures discrete-price MN, around 6 and 13 cents during the same period.

The relative contributions of residual and discrete-price MN to the variance of returns depend on both the standard deviations and the autocorrelations of the MN components. The residual MN is estimated to have high autocorrelations for nearby terms, because it can be modelled by an AR(1) process with the AR parameter having a half-life of approximately 3.3 minutes. In contrast, all the autocorrelations of discrete-price MN are estimated to be between 0 and 0.15 for all positive lags. Consequently, the discrete-price MN contributes more than the residual MN to the variance of returns for return measurement intervals shorter than some threshold. Our very approximate estimates for the threshold are 12 seconds for the ETF and 56 seconds for the futures.

Appendix A: Data processing

The one-minute price records were bought from RC Research via www.price-data.com.

A majority of the ES futures contracts have a full set of prices available until 16:15 on the Thursday preceding final settlement at 09:30 on the next morning. For these contracts we use all the prices before final settlement; no evidence of unusual basis values during the final Thursdays has been observed. For the remaining five contracts we switch to the second-nearest contract on the first day that a full set is unavailable for the nearest contract; on average the switch is made one week before final settlement.

Partial trading days are excluded. There are six holidays in each year having two hours of ES trading but no SPY trading. Three hours of ES trades are unavailable on the Fridays following Thanksgiving Days and many SPY records are also then missing. Three hours of ES trades are also unavailable pre-holiday on 3 July and 24 December 2012. The ES data inexplicably ends early on 11 February and 25 May 2011. Six minutes of SPY data are missing after 16:00 on 26 March 2012 and prices have been created by adding a basis estimate to the ES prices. Finally, four minutes of SPY data are duplicated on 12 August 2011 and the price records in each pair having the lower volume have been deleted.

Large differences between SPY and ES prices, relative to basis estimates, have been checked to detect occasional large price errors. No evidence for any price errors has been found.

The percentage frequencies of off-tick SPY prices are 15.4%, 15.4% and 19.4% in 2010, 2011 and 2012. The most common off-tick final digits are 50 (as in a price of 109.125), with frequencies 4.9%, 5.8% and 7.1%. The final pairs 01, 10, 90 and 99 (as in 109.1201, 109.121, 109.129, 109.1299) all have average frequencies above 1.2%.[13] Rounded prices are defined by rounding to the nearest cent. The half-cent prices are arbitrarily rounded up for odd minutes (e.g. to 109.13 for the last trade before 09:31) and down for even minutes (to 109.12 at 09:32).

Appendix B: Chances of trades at the same ends of the spreads

The probability that spot and futures trades occur at the same ends of the spreads (two bids or two asks) is $P(X_{t,j} = Y_{t,j})$. Table 1 provides probabilities $q_{x,y}$ for the nine events

[13] Buti et al (2014) provide different dark-pool explanations for midpoint (final digits 50) and non-midpoint off-tick prices.

$\frac{1}{2}(X_{t,j+} - X_{t,j}) = x$ and $\frac{1}{2}(Y_{t,j+} - Y_{t,j}) = y$, for $x, y \in [-1,0,1]$. It is trivial to deduce $p_{x,y} = P(X_{t,j} = Y_{t,j} | x, y)$ for eight of the events, namely $p_{-1,-1} = p_{1,1} = 1$, $p_{-1,1} = p_{-1,1} = 0$ and $p_{-1,0} = p_{1,0} = p_{0,1} = p_{1,0} = 0.5$. For any estimate $p_{0,0}$, the same ends probability can then be estimated as $\sum\sum p_{x,y} q_{x,y}$.

To get an upper bound for $p_{0,0}$, we note that $q_{0,0}$ is the sum of 4 terms, from $X_{t,j+} = X_{t,j} = \pm 1$ and $Y_{t,j+} = Y_{t,j} = \pm 1$. We claim that the event $\{X_{t,j+} = X_{t,j} = 1, Y_{t,j+} = Y_{t,j} = -1\}$ is more likely than the event $\{X_{t,j+} = 1, X_{t,j} = -1, Y_{t,j+} = -1, Y_{t,j} = 1\}$ which has chance $q_{-1,1}$. Our upper bound for $p_{0,0}$ is then $(q_{0,0} - q_{-1,1} - q_{1,-1}) / q_{0,0}$ which is 82%. The logical lower bound is 50%.

Appendix C: Evaluating the density of noise differences

The density of $D = U - V$ is required for $0 \le d \le u_{\max} + v_{\max}$. It is provided by $\int h(u, u-d)\, du$, with the integral across all u for which $h(u, u-d)$ is positive. Without loss of generality we assume $u_{\max} \le v_{\max}$.

For the most general density specification in Section 5.4, we need to evaluate

$$\int_{L(d)}^{u_{\max}} a_0 + a_1|u| + a_2|u-d| + a_3|u(u-d)| + S(u(u-d))[a_4 + a_5|u| + a_6|u-d|]\, du \qquad \text{(C1)}$$

for seven coefficients a_i, with $L(d) = \max(d - v_{\max}, -u_{\max})$. The sign function is $S(x) = 1$ for $x > 0$ and $S(x) = -1$ for $x < 0$. The general integral can be rewritten as

$$\int_{L(d)}^{u_{\max}} a_0 + a_4 S(u)S(u-d) + (a_1 S(u) + a_5 S(u-d))u + (a_2 S(u-d) + a_6 S(u))(u-d) + a_3 S(u)S(u-d)u(u-d)\,du. \tag{C2}$$

The integral is evaluated by splitting the range of integration into intervals across which the sign functions are constant; within each interval the integrand is a quadratic function of *u*. The first interval has $S(u) = 1$ and $S(u-d) = -1$, for

$$\begin{aligned} &0 \le u \le d && \text{when } 0 \le d \le u_{\max}, \\ &0 \le u \le u_{\max} && \text{when } u_{\max} \le d \le v_{\max}, \\ &d - v_{\max} \le u \le u_{\max} && \text{when } v_{\max} \le d \le u_{\max} + v_{\max}. \end{aligned} \tag{C3}$$

The second interval contributes when $0 \le d \le v_{\max}$, and it has $S(u) = -1$ and $S(u-d) = -1$ for $L(d) \le u \le 0$. The third interval also contributes when $0 \le d \le u_{\max}$; it has $S(u) = 1$ and $S(u-d) = 1$, for $d \le u \le u_{\max}$.

References

Ait-Sahalia, Y., J. Jacod and J. Li, 2011, Testing for jumps in noisy high frequency data, *Journal of Econometrics* 160, 207-222.

Ait-Sahalia, Y. and L. Mancini, 2008, Out of sample forecasts of quadratic variation, *Journal of Econometrics* 147, 17-33.

Ait-Sahalia, Y., P.A. Mykland and L. Zhang, 2011, Ultra high frequency volatility estimation with dependent microstructure noise, *Journal of Econometrics* 160, 160-175.

Ait-Sahalia, Y. and J. Yu, 2009, High frequency market microstructure noise estimates and liquidity measures, *Annals of Applied Statistics* 3, 422-457.

Andersen, T.G., T. Bollerslev and N. Meddahi, 2011, Realized volatility forecasting and market microstructure noise, *Journal of Econometrics* 160, 220-234.

Andersen, T.G., O. Bondarenko, A.S. Kyle and A.A. Obizhaeva, 2015, high-frequency trading invariants for equity-index futures, working paper, Northwestern University.

Bandi, F.M. and J.R. Russell, 2006, Separating microstructure noise from volatility, *Journal of Financial Economics* 79, 655-692.

Bandi, F.M. and J.R. Russell, 2008, Microstructure noise, realized variance, and optimal sampling, *Review of Economic Studies* 75, 339-369.

Bandi, F.M. and J.R. Russell, 2011, Market microstructure noise, integrated variance estimators, and the accuracy of asymptotic approximations, *Journal of Econometrics* 160, 145-159.

Barndorff-Nielsen, O.E., P.R. Hansen, A. Lunde and N. Shephard, 2008, Designing realized kernels to measure the ex post variation of equity prices in the presence of noise, *Econometrica* 76, 1481-1536.

Buti, S., F. Consonni, B. Rindi, Y. Wen and I.M. Werner, 2014, Sub-penny and queue-jumping, working paper, University of Toronto.

Corsi, F. and F. Audrino, 2012, Realized covariance tick-by-tick in presence of rounded time stamps and general microstructure effects, *Journal of Financial Econometrics* 10, 591-616.

Dahlhaus, R. and J.C. Neddermeyer, 2014, Online spot volatility-estimation and decomposition with nonlinear market microstructure noise models, *Journal of Financial Econometrics* 12, 174-212.

Diebold, F.X. and G. Strasser, 2013, On the correlation structure of microstructure noise: a financial economic approach, *Review of Economic Studies* 80, 1304-1337.

Duffie, D., 1989, *Futures Markets* (Prentice Hall, Englewood Cliffs).

Ghysels, E. and A. Sinko, 2011, Volatility forecasting and microstructure noise, *Journal of Econometrics* 160, 257-271.

Hansen, P.R. and A. Lunde, 2006, Realized variance and market microstructure noise, *Journal of Business and Economic Statistics* 24, 127-161.

Hendershott, T. and A.J. Menkveld, 2014, Price pressures, *Journal of Financial Economics* 114, 405-423.

Jacod, J., Y.Li and X. Zheng, 2014, Statistical properties of microstructure noise, working paper, Université Pierre et Marie Curie and Hong Kong University of Science and Technology.

Lee, S.S. and P.A. Mykland, 2012, Jumps in equilibrium prices and market microstructure noise, *Journal of Econometrics* 168, 396-406.

Lo, A.W. and A.C. MacKinlay, 1988, Stock market prices do not follow random walks: evidence from a simple specification test, *Review of Financial Studies* 1, 41-66.

MacKinlay, A.C. and K. Ramaswamy, 1988, Index-futures arbitrage and the behavior of stock index futures prices, *Review of Financial Studies* 1, 137-158.

Nolte, I. and V. Voev, 2012, Least squares inference on integrated volatility and the relationship between efficient prices and noise, *Journal of Business and Economic Statistics* 30, 94-108.

Stoll, H.R. and R.E. Whaley, 1990, The dynamics of stock index and stock index futures returns, *Journal of Financial and Quantitative Analysis* 25, 441-468.

Taylor, S.J., 1984, Estimating the variances of autocorrelations calculated from financial time series, *Applied Statistics* 33, 300-308.

Ubukata, M. and K. Oya, 2009, Estimation and testing for market microstructure noise, *Journal of Financial Econometrics* 7, 106-151.

Voev, V. and A. Lunde, 2007, Integrated covariance estimation using high-frequency data in the presence of noise, *Journal of Financial Econometrics* 5, 68-104.

Zhang, L., P.A. Mykland and Y. Ait-Sahalia, 2005, A tale of two time scales: determining integrated volatility with noisy high-frequency data, *Journal of the American Statistical Association* 100, 1394-1411.

Table 1 Bivariate distribution of trade-to-trade price changes

The percentages are calculated from 122,605 minute marks, for which firstly the SPY price change from last-before-mark to first-after-mark is either zero or ± one spot tick and secondly likewise for the ES price change. All minute marks are between 11:00 and 15:00 inclusive. There are 1,251 minute marks during these hours for which either the SPY change or the ES change (or both) is (are) beyond ± one tick. Minute marks having off-tick spot prices are excluded from these counts.

	ES change, in ticks			
	−1	0	1	Total
SPY change, in ticks				
−1	4.13%	11.07%	3.31%	18.51%
0	12.28%	37.36%	12.93%	62.57%
1	3.31%	11.24%	4.37%	18.92%
Total	19.72%	59.67%	20.61%	100%

Table 2 Summary statistics for adjusted price differences

Statistics are provided for the adjusted price differences, $q_{t,j} = S_{t,j} - F_{t,j} - \hat{B}_t$, calculated once a minute, with *S*, *F*, *t* and *j* respectively denoting the SPY price, the ES price, the day and the minute, and with $\hat{B}_t$ the average for day *t* across minutes of the differences $S_{t,j} - F_{t,j}$. The units of $q_{t,j}$ are cents.

	2010	2011	2012	All years
Days	246	249	247	742
Mean absolute deviations				
09:31 to 10:59	11.89	13.06	12.49	12.48
11:00 to 15:00	11.33	11.86	11.62	11.61
15:01 to 16:15	12.39	13.20	12.83	12.81
All day	11.65	12.37	12.04	12.02
Standard deviations				
09:31 to 10:59	14.91	16.96	15.60	15.85
11:00 to 15:00	13.76	14.76	14.09	14.22
15:01 to 16:15	19.20	17.58	16.49	17.79
All day	15.16	15.82	14.90	15.30
Percentages beyond ± 35				
09:31 to 10:59	1.01	2.74	1.95	1.90
11:00 to 15:00	0.43	0.93	0.61	0.66
15:01 to 16:15	1.20	2.78	2.21	2.07
All day	0.70	1.67	1.20	1.19
Percentages beyond ± 70				
09:31 to 10:59	0.12	0.29	0.05	0.15
11:00 to 15:00	0.03	0.07	0.01	0.04
15:01 to 16:15	0.21	0.30	0.29	0.26
All day	0.08	0.16	0.07	0.11

Table 3 Variances, covariances, autocorrelations and cross-correlations for price changes

Variances and covariances for one-minute price changes $\Delta S_{t,j} = S_{t,j} - S_{t,j-1}$ and $\Delta F_{t,j} = F_{t,j} - F_{t,j-1}$, and one-minute returns, $\ln(S_{t,j} / S_{t,j-1})$ and $\ln(F_{t,j} / F_{t,j-1})$, with t counting days and j counting minutes. The first group of autocorrelations and cross-correlations are estimates of dependence between either $\Delta S_{t,j}$ or $\Delta F_{t,j}$ and either $\Delta S_{t,j+\tau}$ or $\Delta F_{t,j+\tau}$. The data are mid-day prices, from 11:00 to 15:00 on days from January 2010 to December 2012. Covariances are estimated from products of terms a fixed time apart on the same trading day. The units of the price changes are cents.

Covariances for price changes

	Spot	Futures
Spot	2154.2	2117.6
Futures		2447.8

$10^8 \times$ Covariances for returns

	Spot	Futures
Spot	14.65	14.47
Futures		16.67

Correlations for price changes

Correlations for returns

	j	Spot	Futures	Spot	Futures	Spot	Futures	Spot	Futures
	$j+\tau$	Spot	Futures	Futures	Spot	Spot	Futures	Futures	Spot
τ									
0		1	1	0.9222	0.9222	1	1	0.9258	0.9258
1		-0.0265	-0.0842	-0.0300	-0.0054	-0.0259	-0.0810	-0.0296	-0.0054
2		-0.0147	-0.0154	-0.0166	-0.0138	-0.0152	-0.0154	-0.0165	-0.0144
3		-0.0030	-0.0009	-0.0019	-0.0013	-0.0023	-0.0007	-0.0015	-0.0008
4		-0.0176	-0.0151	-0.0153	-0.0176	-0.0186	-0.0155	-0.0157	-0.0187
5		-0.0072	-0.0060	-0.0061	-0.0065	-0.0075	-0.0068	-0.0069	-0.0069
6		-0.0029	-0.0019	-0.0024	-0.0027	-0.0037	-0.0020	-0.0028	-0.0032
7		-0.0070	-0.0059	-0.0054	-0.0069	-0.0076	-0.0067	-0.0061	-0.0074
8		-0.0012	-0.0010	-0.0013	-0.0016	-0.0011	-0.0009	-0.0011	-0.0018
9		-0.0041	-0.0007	-0.0015	-0.0022	-0.0043	-0.0012	-0.0022	-0.0022
10		-0.0020	-0.0039	-0.0033	-0.0035	-0.0033	-0.0048	-0.0043	-0.0048

Table 4 Parameter estimates for the bivariate density of S&P 500 microstructure noise during the mid-day period

Bivariate densities $h(u,v)$ are defined by combining component densities across states, with each state having a different combination of bid/ask spread widths. Densities are defined by (45) for one state, by (52) and (53) for two states and by (39), (52) and (54) for four states; parameter constraints are stated in Sections 5.4 and 6.2. The standard deviation of the basis estimation error is denoted by σ. The column AL shows adjusted log-likelihoods, which are the maximum log-likelihood for a density specification minus the maximum log-likelihood for the selected specification having $\mathrm{AL} = 0$. The goodness-of-fit measure X^2, defined by (51), summarizes the fit for adjusted price differences inside ± 0.35. Results are shown by year for the mid-day period from 11:00 to 15:00 inclusive.

Panel A: One state

SPY spread 0.1
ES spread 0.25

	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	σ	AL	X^2
2010	-	-	-	-	-	-	0.0173	– 2008.06	3639.20
	-	1.1528	-	-	-	-		– 171.99	363.49
	-	0.6795	-	-	-	0.3188		0	20.39
	0.0484	0.6860	– 0.0181	0.0003	– 0.0003	0.3001		1.18	21.39
2011	-	-	-	-	-	-	0.0239	– 1321.11	2620.95
	-	0.9447	-	-	-	-		– 19.88	82.11
	-	0.7845	-	-	-	0.1116		0	30.18
	0.1034	0.8293	0.0950	0.0084	– 0.0084	0.1029		0.12	31.63

Table 4 continued…

	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	σ	AL	X^2
2012	-	-	-	-	-	-	0.0269	– 1589.12	3094.72
	-	1.0161	-	-	-	-		– 158.35	419.67
	-	0.5614	-	-	-	0.3135		0	70.44
	0.0000	0.5614	0.0000	0.0000	0.0000	0.3135		0	70.44

A dash (-) indicates that the parameter is constrained to be zero.

Panel B: Two states

SPY spread	0.1	0.2	
ES spread	0.25	0.25	

	$p^{(1)}$	$\gamma^{(1)}$	$\delta^{(1)}$	$p^{(2)}$	$\gamma^{(2)}$	$\delta^{(2)}$	X^2
2010	0.9402	0.6830	0.3864	0.0598	0.0321	0.0000	44.87
2011	0.8712	0.7861	0.2444	0.1288	0.1398	0.0000	64.40
2012	0.9283	0.5493	0.4013	0.0717	0.0000	0.0000	92.29

Panel C: Four states

SPY spread	0.1	0.2	0.1	0.2	
ES spread	0.25	0.25	0.5	0.5	

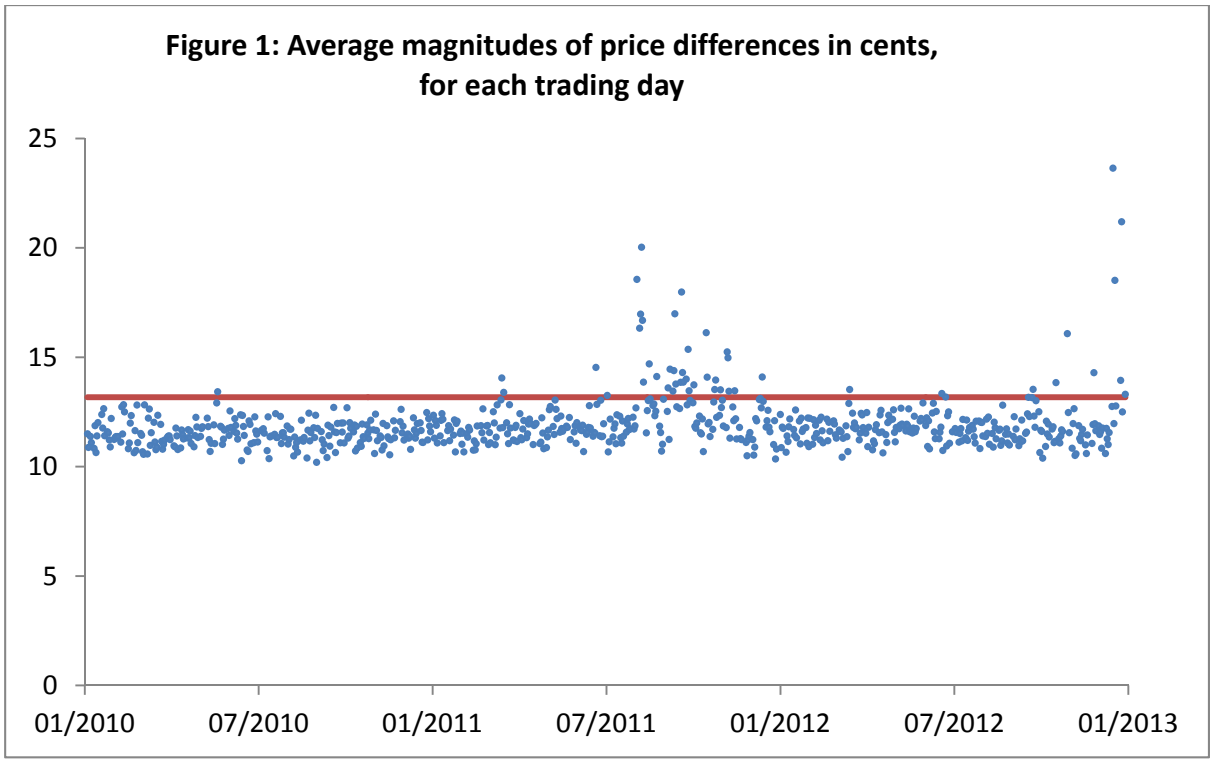
	$p^{(1)}$	$\gamma^{(1)}$	$\delta^{(1)}$	$p^{(2)}$	$p^{(3)}$	$p^{(4)}$	X^2
2010	0.9483	0.6842	0.3793	0.0473	0.0000	0.0044	42.33
2011	0.8967	0.7905	0.2264	0.0912	0.0000	0.0121	59.37
2012	0.9444	0.5587	0.3830	0.0478	0.0041	0.0038	88.91

Table 5 Parameter estimates for the bivariate density of S&P 500 microstructure noise during the primary trading period

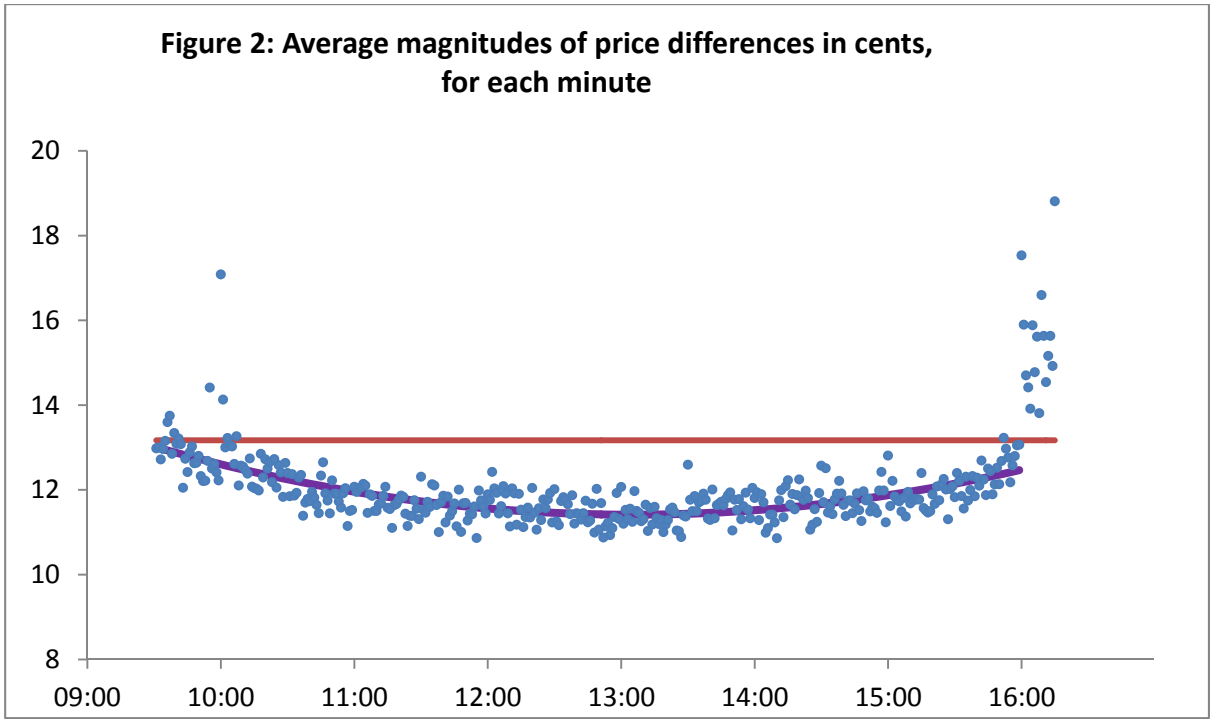
Bivariate densities are defined by (55) and combine component densities across three states, with each state having a different combination of bid/ask spread widths. The component densities are defined by (39) and (52). Their time-varying state probabilities depend on the minute *j* and the day's realized variance RV_t, as shown in (56). The standard deviation of the basis estimation error is first estimated and denoted by σ. The remaining parameters are all non-negative and estimated by maximizing the log-likelihood criterion stated in (50). The column AL shows adjusted log-likelihoods, which are the maximum log-likelihood for a density specification minus the maximum log-likelihood for the special case when there are no time-varying effects. Results are shown by year for the trading period from 09:31 to 15:59 inclusive, excluding 10:00.

	γ	δ	$p_{\min}^{(2)}$	$p_{\min}^{(3)}$	α	β	$j_{\min}$	σ	AL
2010	0.7200	0.3009	0.0585	0.0066	-	-	-	0.0163	0
	0.7175	0.2997	0.0277	0.0032	1.453	-	-		48.17
	0.7174	0.3146	0.0283	0.0028	-	3.872	203.1		88.98
	0.7151	0.3126	0.0139	0.0014	1.335	3.865	204.4		138.03
2011	0.8050	0.1846	0.1368	0.0220	-	-	-	0.0255	0
	0.8000	0.1729	0.0198	0.0034	5.483	-	-		557.91
	0.8038	0.1930	0.0527	0.0083	-	4.865	212.4		316.93
	0.7942	0.1842	0.0088	0.0014	4.958	4.609	213.5		856.91
2012	0.6610	0.2910	0.0937	0.0141	-	-	-	0.0273	0
	0.6593	0.2916	0.0498	0.0075	2.302	-	-		40.07
	0.6501	0.3105	0.0329	0.0047	-	5.831	221.6		272.62
	0.6489	0.3097	0.0193	0.0028	1.901	5.647	222.1		308.74

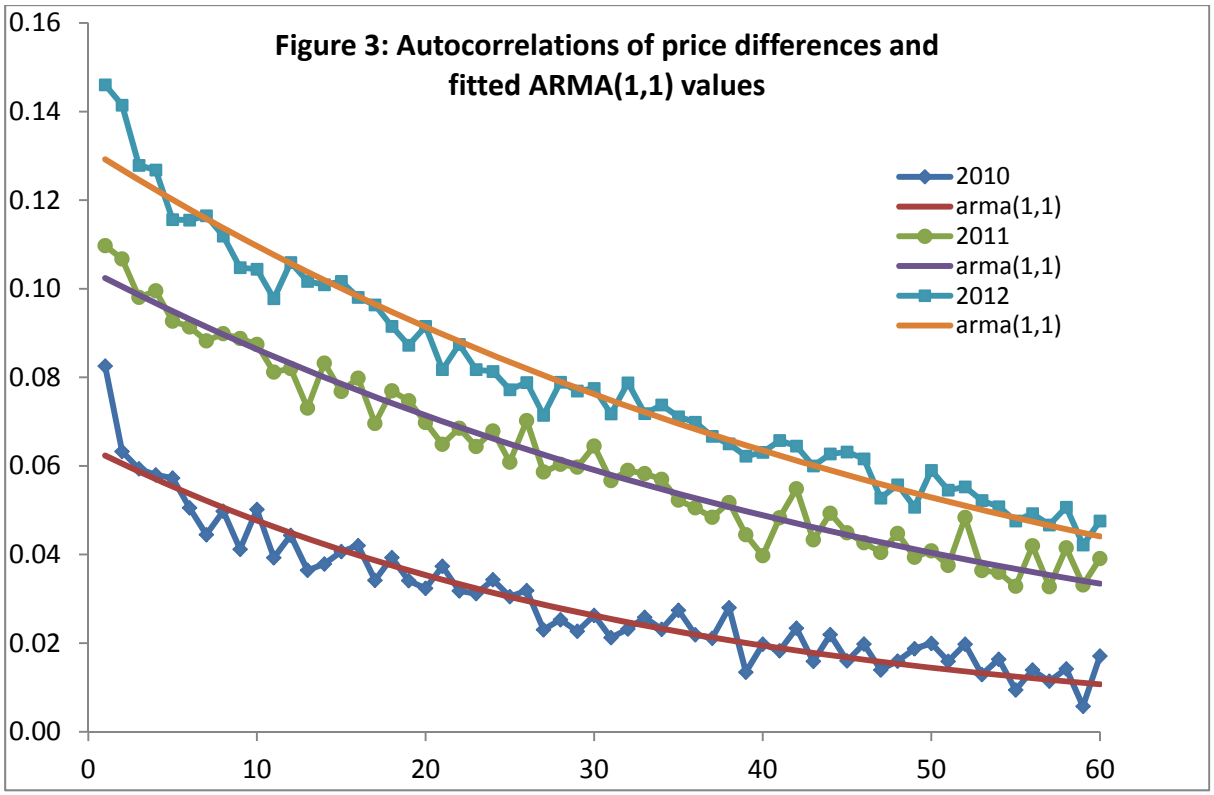
A dash (-) indicates that the parameter is constrained to be zero.



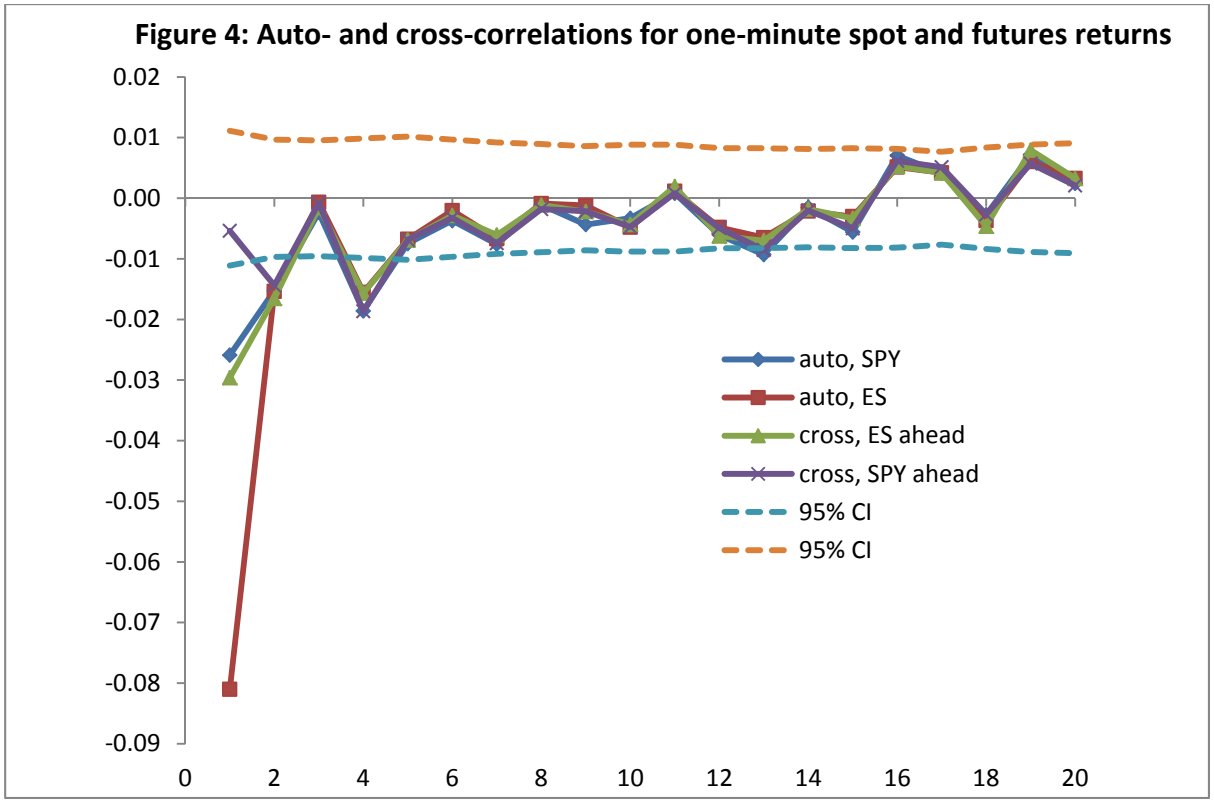
The blue dots mark mean absolute deviations of adjusted price differences. The red line is the theoretical m.a.d. when the discrete-price noise terms are independent and have uniform distributions, which have zero expectations and maximum values equal to one tick.



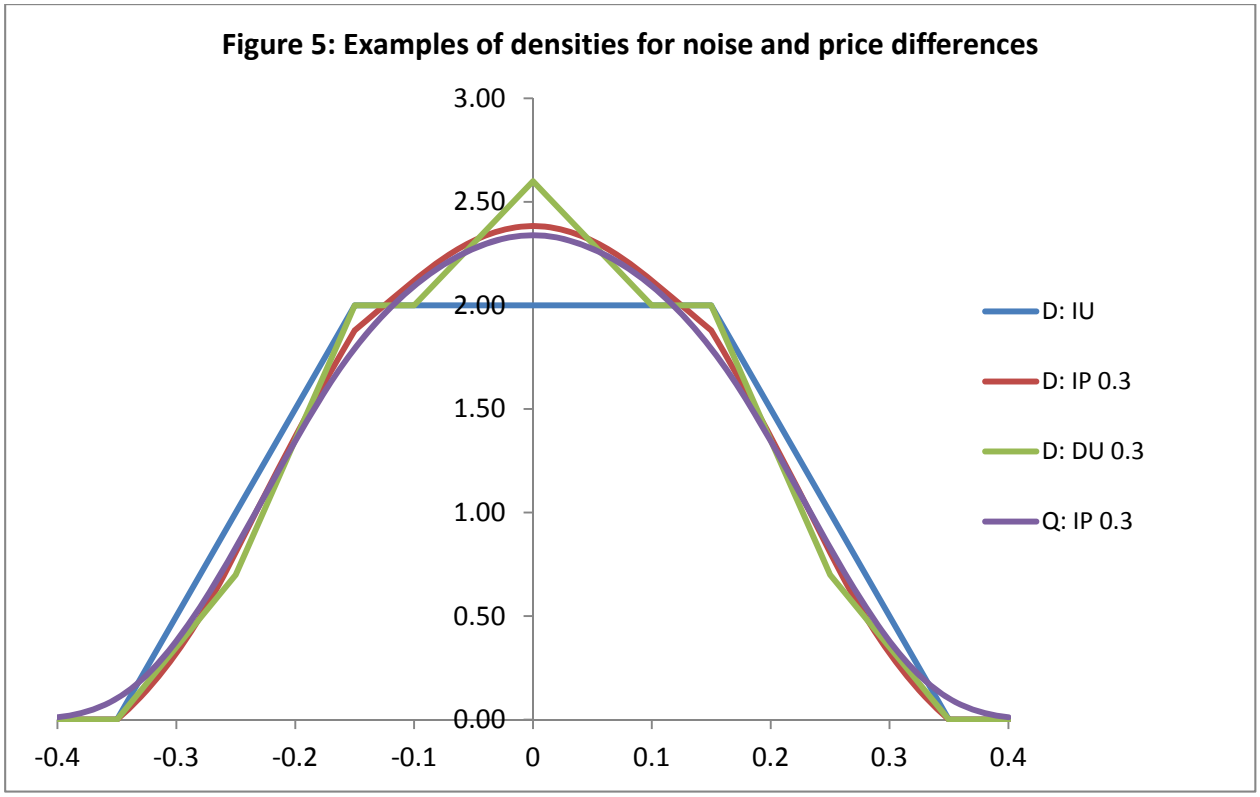
The purple curve shows fitted values when the mean absolute deviation of price differences is regressed on time and its square. The red line is the level for independent, uniform noise when all spreads are one tick wide.



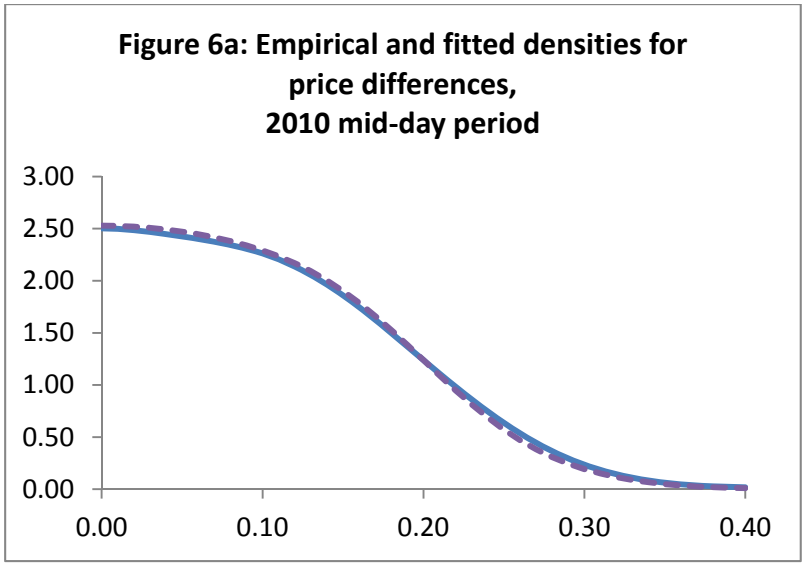
Autocorrelations are shown for each year. The time lag is measured in minutes. The fitted curves minimize the sum of squared differences between empirical and theoretical autocorrelations.

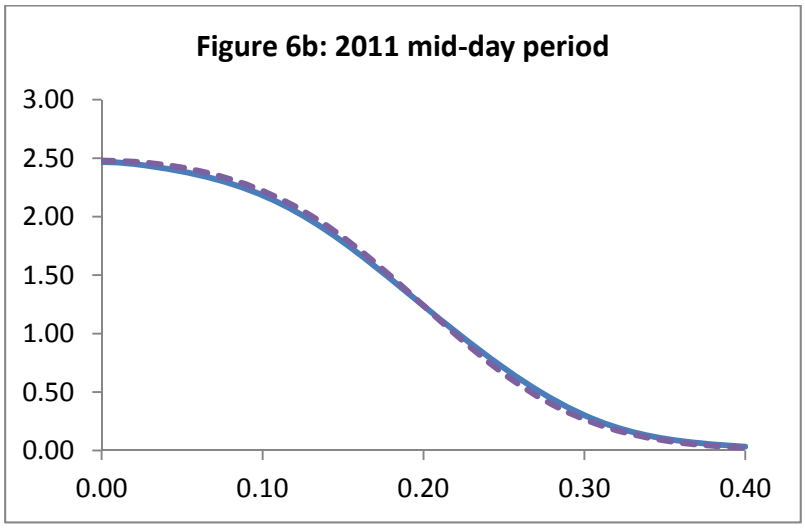


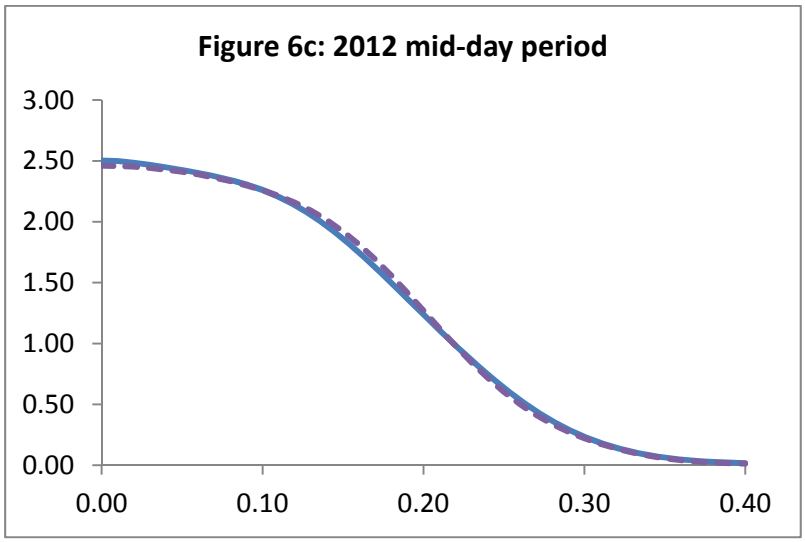
Auto- and cross-correlations are shown for one-minute returns, for the mid-day period from 11:00 to 15:00. The dotted lines connect robust 95% intervals for the individual correlation estimates under the hypothesis that the returns are generated by an uncorrelated process. All numbers are calculated by using the returns from 2010 to 2012.



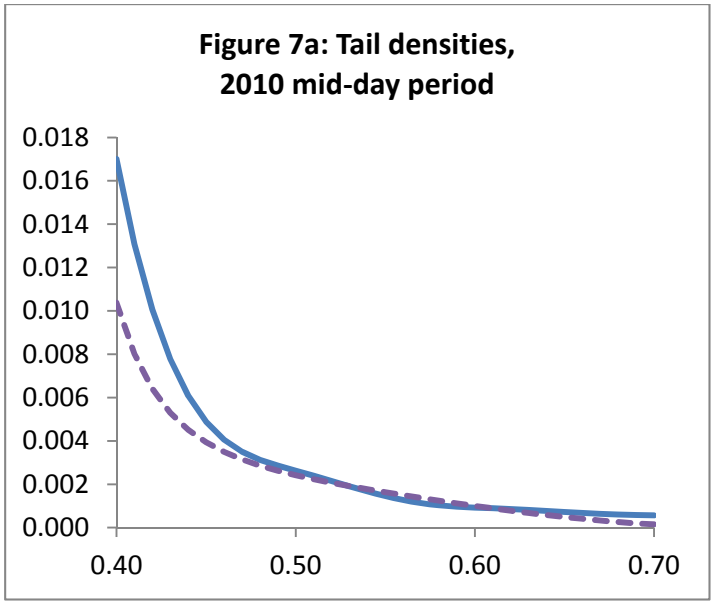
The density of noise differences D when the noise terms are independent and uniform (IU) has three line segments and passes through (0, 2). The independent polygon (IP) density is illustrated by the red and curve and is defined by (41) and (42). The dependent uniform (DU) density, defined by (44), is shown by the green curve and has the highest peak. The density of price differences Q is illustrated by the curve whose density extends beyond 0.35, and is calculated from (34). The price units are dollars.

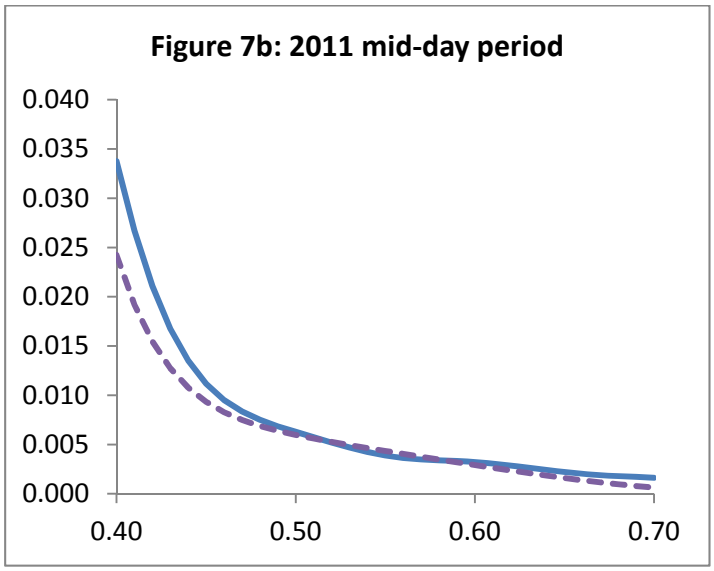


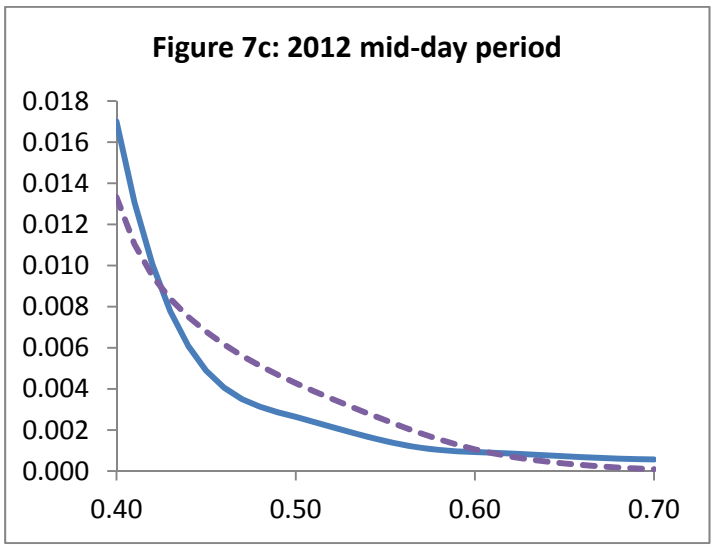




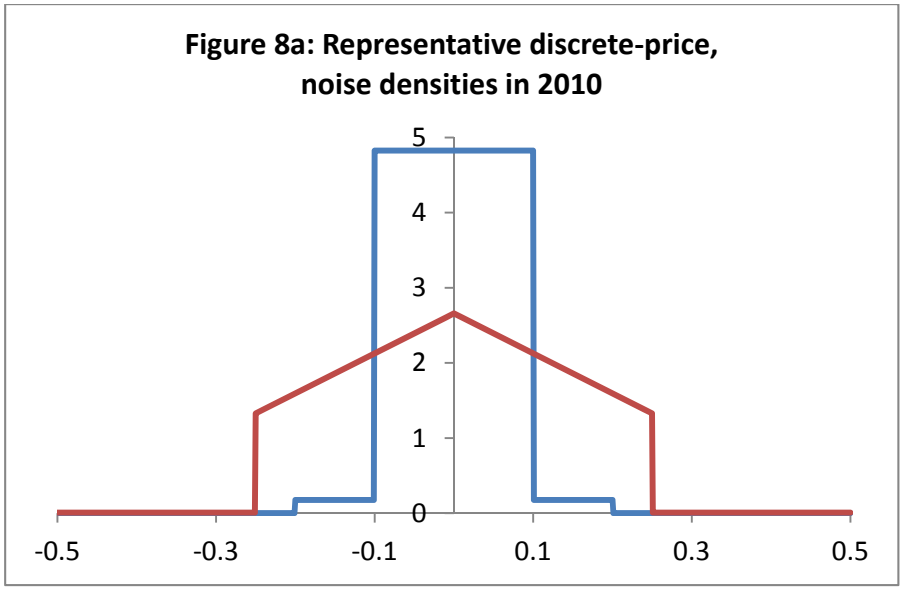
The mid-day period is from 11:00 to 15:00. The solid curves show kernel estimates of the densities of the adjusted price differences q . The dashed curves are the fitted densities for the four-state specification defined by (39), (52) and (54).

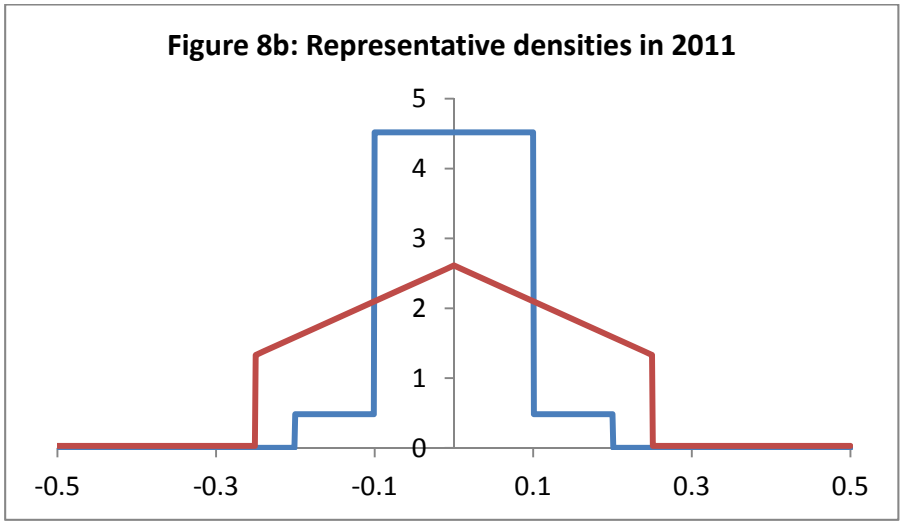


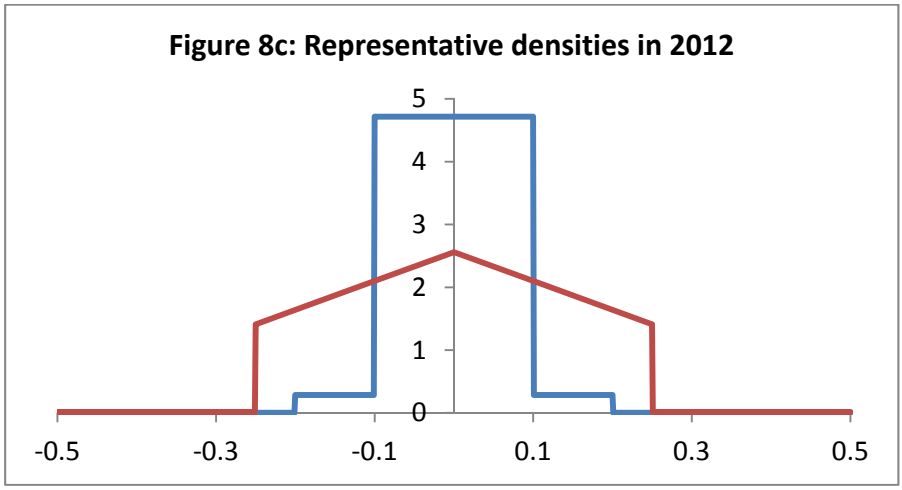




Solid curves are kernel estimates. Dashed curves are fitted densities for four states.







Blue lines are for the density of SPY noise, red lines are for ES noise. The three state probabilities used are averages of values which vary across the clock and also depend on realized volatility. The bivariate density is (55) and (56) and the marginal densities are (57) and (58). Densities are estimated for prices from 09:31 to 15:59, excluding 10:00.