Markov Chain Monte Carlo Algorithms for Bayesian Estimation of Microstructure Models

Computational Appendix to "Liquidity in the futures pits: Inferring market dynamics from incomplete data"

Joel Hasbrouck
May, 2002

1. Introduction

This appendix describes the Markov Chain Monte Carlo (MCMC) estimation of microstructure models with bid/ask spreads, discreteness, clustering and trade impacts. In all cases, the data are presumed to consist solely of trade prices and (optionally) trade volumes. The exposition discusses models of increasing complexity. The appendix is distributed in two forms: a text document, and a Mathematica notebook in which derivations are interspersed with the text.

The appendix presumes some exposure to Bayesian MCMC techniques. Textbook level discussions are given in Carlin and Louis (1996), Tanner (1996), Gamerman (1997), and Kim and Nelson (2000). Casella and George (1992), Chib and Greenberg (1995), and Chib and Greenberg (1996) also provide useful expositions.

(*Mathematica* initializations)

2. Notation and conventions

A time-series variable written without a time subscript denotes the full sample, e.g., $q = \{q_1, q_2, \ldots, q_T\}$. The subscript "$\backslash t$" indicates the full sample except for the observation at time t, e.g., $q_{\backslash t} = \{q_1, \ldots, q_{t-1}, q_{t+1}, \ldots, q_T\}$.

"Draw $x \mid y, z$" is shorthand for "draw x from $f(x \mid y, z)$."

The expression $\phi[\mu, \sigma, x]$ denotes the normal density function for random variable x, with mean μ and standard deviation σ. For example, we denote the normal density for a random variable z with mean m and standard deviation s as $\phi(m, s, z)$. Fully expanded,

$$\phi(m, s, z) = \frac{e^{-\frac{(z-m)^2}{2 s^2}}}{\sqrt{2\pi}\, s}$$

Similarly, the expression $\Phi[\mu, \sigma, a, b] \equiv \int_a^b \phi[\mu, \sigma, x]\, dx$, i.e., the cumulative normal distribution function evaluated between a and b.

3. Basic Roll model

a. Model description

The basic Roll model is:

$$\begin{aligned} m_t &= m_{t-1} + u_t \text{ where } u_t \text{ is distributed as } N(0, \sigma_u^2) \\ p_t &= m_t + c\, q_t \end{aligned} \tag{1}$$

where m_t is the log efficient price, q_t is the buy/sell (trade direction) indicator variable, $q_t \in \{-1, +1\}$, c is the is the half-spread, and p_t is the log transaction price. The increment to the log efficient price, $u_t \sim N(0, \sigma_u^2)$; c and σ_u^2 are constant parameters. The q_t are independent of $u_t = \Delta m_t$, so trade directions are not informative.

The data consist of a sample of (log) transaction prices $p = \{p_1, \ldots, p_T\}$. The Gibbs sampler iterates over parameter draws and latent data draws. There are two parameters in the model, $\{\sigma_u, c\}$. There are $2T$ latent data items, $\{q, m\}$.

b. Moment estimates

The price difference is

$$\Delta p_t = u_t + c \Delta q_t$$

The variance and first-order autocovariance are:

$$\begin{aligned} \gamma_0 &\equiv \text{Var}(\Delta p_t) = \sigma_u^2 + c^2\, \text{Var}(\Delta q_t) = \sigma_u^2 + 2c^2 \\ \gamma_1 &\equiv \text{Cov}(\Delta p_t, \Delta p_{t-1}) = c^2\, \text{E}\Delta q_t \Delta q_{t-1} = -c^2 \end{aligned}$$

Solving:

$$\sigma_u^2 = \gamma_0 + 2\gamma_1 \text{ and } c = \sqrt{-\gamma_1}$$

c. Overview of the Bayesian sampler

In the Gibbs sampler, we start with any feasible set of values (a sample path) for q. We then iterate over

1. Parameter draw. Generate a random draw of $c, \sigma_u \mid p, m, q$.

2. Latent data draw. Generate a random draw of $m, q \mid c, \sigma_u, p$.

In the parameter draw, we condition on the most-recently drawn values of the latent data; in the latent data draw, we condition on the most recent parameter draw. We now discuss each of these draws in greater detail.

d. Parameter draws

The prior and posterior densities for the parameters are standard Bayesian results covered in most basic treatments, such as Carlin and Louis (1997), Tanner (1997). Kim and Nelson (2000) discuss these results in a state-space context.

(i) The draw for *c*.

Given all of the other parameters and latent data $\Delta p_t = \Delta m_t + c\Delta q_t$. Therefore, c can be interpreted as the coefficient in the regression specification

$$\Delta p_t = c\Delta q_t + u_t. \qquad (2)$$

The classic Bayesian regression model is

$$y = X\beta + u, \text{ where } Euu' = \Omega_u.$$

If the prior distribution on the coefficients is normal: $\beta \sim N(\mu_\beta^{\text{Prior}}, \Omega_\beta^{\text{Prior}})$, then the posterior distribution for the coefficients is $\beta \mid y \sim N(\mu_\beta^{\text{Posterior}}, \Omega_\beta^{\text{Posterior}})$, where $\mu_\beta^{\text{Posterior}} = Dd$, $\Omega_\beta^{\text{Posterior}} = D^{-1}$, $D^{-1} = X'\,\Omega_u^{-1}\,X + (\Omega_\beta^{\text{Prior}})^{-1}$ and $d = X'\,\Omega_u^{-1}\,y + (\Omega_\beta^{\text{Prior}})^{-1}\,\mu_\beta^{\text{Prior}}$.

In the present case, $y = [\Delta p_t]$, the column vector of price changes, $X = [\Delta q_t]$, and $\Omega = \sigma_u^2$. We might take $\mu_\beta^{\text{Prior}} = 0$, and $\Omega_\beta^{\text{Prior}}$ =some large number like 10^6. We would then compute $\mu_\beta^{\text{Posterior}}$ and $\Omega_\beta^{\text{Posterior}}$, and make a random draw from the coefficient posterior.

On economic grounds, it is sensible to impose a non-negativity restriction on c. The easiest way to do this is to let the prior be $c \sim N^+(0, \Omega_c^{\text{Prior}})$, where the "+" superscript denotes restriction of the density to $[0, +\infty)$. The posterior is then $N^+(\mu_c^{\text{Posterior}}, \Omega_c^{\text{Posterior}})$, where the parameters are computed exactly as in the unrestricted case.

(The nonnegativity restriction on c is attractive from an economic perspective. It is also necessary for identification. The regression in eq. (1) is observationally equivalent if c and q are replaced with $-c$ and $-q$.)

(ii) The draw for σ_u

With normal u_t, a natural prior for σ_u^2 is the inverted gamma distribution, denoted IG[$\alpha^{\text{Prior}}, \beta^{\text{Prior}}$]. Then, given $u = \{u_1, \ldots, u_T\}$, the posterior is IG[$\alpha^{\text{Posterior}}, \beta^{\text{Posterior}}$], where $\alpha^{\text{Posterior}} = \alpha^{\text{Prior}} + T/2$ and $\beta^{\text{Posterior}} = \beta^{\text{Prior}} + \Sigma u_t^2/2$.

e. Latent data draws

There are two latent series, q and m, but one of these is redundant. (The p are known, and $p = m + cq$.) It suffices therefore to draw $q \mid p, c, \sigma_u$. At this stage, c and σ_u are known, so for notational economy, these parameters will be dropped from the explicit conditioning set. To set up the draw:

$$\Pr(q \mid p) = f(p \mid q) \times \Pr(q) \times \frac{1}{f(p)} \tag{3}$$

Since the q are discrete random variables and the p are continuous, the joint distribution is written as $f(p \mid q) \times (\Pr(q)$ to avoid writing an improper density function like "$f(p, q)$". In eq. (3) the last factor does not depend on q. Furthermore, since buys and sells are unconditionally equally probable, $\Pr(q) = (1/2)^T$. Therefore:

$$\Pr(q \mid p) \propto f(p \mid q) = f(m)|_{m=p-cq} \tag{4}$$

where $f(m)$ is the unconditional distribution of the m. From the structure of the model, $f(m) = f(m_1) \prod_{t=2}^{T} f(u_t)$ where $u_t = \Delta m_t$. The distribution of the initial realization of m is taken as uniform in the region of interest, so this can be impounded into the constant of proportionality.

The direct approach to making the draw would be to enumerate all possible sample paths for q, compute the unnormalized (i.e., up to the factor of proportionality) probabilities; normalize. Finally, we'd make the draw on the resulting discrete probability space. Given that the number of sample paths is 2^T, the direct approach is not feasible.

Fortunately, it is not necessary to fully characterize $\Pr(q \mid p)$ to make the draw. The next section describes a Gibbs sampler approach to the problem.

f. The Gibbs sampler

At this stage, we assume that c and σ_u are known, so for notational economy, these parameters will be dropped from the explicit conditioning set. Furthermore, since the p are known, given q, the m are redundant. In the Gibbs sampler, q_t draws are made sequentially. To distinguish the newly drawn q_t from those left over from the previous draw, I denote the newly drawn values q_t^*. The steps are then:

1. Draw q_1^* from $\Pr(q_1 \mid p, q_2, q_3, \ldots, q_T)$

2. Draw q_2^* from $\Pr(q_2 \mid p, q_1^*, q_3, \ldots, q_T)$

$\vdots$

3. Draw q_T^* from $\Pr(q_T \mid p, q_1^*, q_2^*, \ldots, q_{T-1}^*)$

Note that at each step, we use the fresh q_t^* as they become available. The individual draws all follow a similar pattern. We will first consider the general case, the draw for $q_t \mid p, q_{\setminus t}$ for $1 < t < T$. The endpoint draws (at $t = 1$ and $t = T$) follow as simple modifications.

Given the structure of the model, $\Pr(q_t \mid p, q_{\setminus t}) = \Pr(q_t \mid p_t, m_{t-1}, m_{t+1})$. That is, once we know p_t, m_{t-1}, and m_{t+1}, the additional variables in the full conditioning set are not informative. To analyze $\Pr(q_t \mid p_t, m_{t-1}, m_{t+1})$, note that by conventional conditional probability calculations

$$\Pr(q_t \mid p_t, m_{t-1}, m_{t+1}) = \frac{f(p_t \mid q_t, m_{t-1}, m_{t+1}) \times \Pr(q_t \mid m_{t-1}, m_{t+1})}{f(p_t \mid m_{t-1}, m_{t+1})} \tag{5}$$

When we actually need to evaluate this, p_t is known, so the denominator can be treated as a proportionality constant. Next, because the trade directions are independent of the efficient price evolution, $\Pr(q_t \mid m_{t-1}, m_{t+1}) = \Pr(q_t) = 1/2$. Therefore,

$$\Pr(q_t \mid p_t, m_{t-1}, m_{t+1}) \propto f(p_t \mid q_t, m_{t-1}, m_{t+1}).$$

To evaluate the r.h.s., note that

$$f(m_t \mid m_{t-1}, m_{t+1}) \propto f(m_{t+1} \mid m_t)\, f(m_t \mid m_{t-1}) = \phi(0, \sigma_u, m_t - m_{t-1})\, \phi(0, \sigma_u, m_{t+1} - m_t)$$

The last expression is:

$$\frac{e^{-\frac{(m_t - m_{t-1})^2}{2\sigma_u^2} - \frac{(m_{t+1} - m_t)^2}{2\sigma_u^2}}}{2\pi\sigma_u^2}$$

This implies

$$f(m_t \mid m_{t-1}, m_{t+1}) = \phi(\mu_t, \sigma_t, m_t) \tag{6}$$

where

$$\mu_t = \left(\frac{1}{2}(m_{t-1} + m_{t+1})\right)$$
$$\sigma_t = \frac{\sigma_u}{\sqrt{2}}$$

Since $p_t = m_t + c q_t$,

$$f(p_t \mid q_t, m_{t-1}, m_{t+1}) = \phi(\mu_t, \sigma_t, p_t - c q_t) \tag{7}$$

The unnormalized probabilities of a buy and a sell are given by evaluating the above quantity at $q_t = +1$ and $q_t = -1$ respectively. The normalized probability of a buy is therefore

$$\frac{\phi\left(\frac{1}{2}(m_{t-1}+m_{t+1}), \frac{\sigma_u}{\sqrt{2}}, p_t - c\right)}{\phi\left(\frac{1}{2}(m_{t-1}+m_{t+1}), \frac{\sigma_u}{\sqrt{2}}, p_t - c\right) + \phi\left(\frac{1}{2}(m_{t-1}+m_{t+1}), \frac{\sigma_u}{\sqrt{2}}, c + p_t\right)}$$

This simplifies to

$$\frac{e^{\frac{4 c p_t}{\sigma_u^2}}}{e^{\frac{2 c (m_{t-1}+m_{t+1})}{\sigma_u^2}} + e^{\frac{4 c p_t}{\sigma_u^2}}}$$

Given p_t, m_{t-1}, m_{t+1}, c and σ_u^2, we can compute this and make the draw. For example, with $\{m_{t-1} = 5,\ m_{t+1} = 5.1,\ p_t = 5.2,\ c = 0.2,\ \sigma_u = 0.4\}$, the probability of a buy is 0.679. Intuitively, we can rationalize this value in the following way. Given that $m_{t-1} = 5$ and $m_{t+1} = 5.1$, $E m_t = 5.05$. That the transaction price lies above this makes it more likely that the trade was a "buy".

(i) $t = 1$

At $t = 1$, the relevant conditional density is $f(m_1 \mid m_2)$, which is proportional to $f(m_2 \mid m_1) = \phi(0, \sigma_u, m_2 - m_1)$. Substituting in for m_1 gives $\phi(0, \sigma_u, m_2 - p_1 + c\, q_1)$. The unnormalized probabilities for a buy and sell are given by evaluating this at $q_1 = +1$ and $q_1 = -1$. The normalized probability of a buy is:

$$\frac{\phi(0, \sigma_u, c + m_2 - p_1)}{\phi(0, \sigma_u, -c + m_2 - p_1) + \phi(0, \sigma_u, c + m_2 - p_1)}$$

This simplifies to:

$$\frac{e^{\frac{2 c p_1}{\sigma_u^2}}}{e^{\frac{2 c m_2}{\sigma_u^2}} + e^{\frac{2 c p_1}{\sigma_u^2}}}$$

(ii) $t = T$

At time T we have $f(m_T \mid m_{T-1}) = \phi(0, \sigma_u, m_T - m_{T-1})$. Evaluated at $m_T = p_T - c q_T$, this becomes $\phi(0, \sigma_u, -m_{T-1} + p_T - c\, q_T)$. The (unnormalized) probabilities for a buy and sell are given by evaluating this at $q_T = +1$ and $q_T = -1$. The normalized probability of a buy is:

$$\frac{\phi(0,\sigma_u,-c-m_{T-1}+p_T)}{\phi(0,\sigma_u,-c-m_{T-1}+p_T)+\phi(0,\sigma_u,c-m_{T-1}+p_T)}$$

This evaluates and simplies to:

$$\frac{e^{\frac{2\,c\,p_T}{\sigma_u^2}}}{e^{\frac{2\,c\,m_{T-1}}{\sigma_u^2}}+e^{\frac{2\,c\,p_T}{\sigma_u^2}}}$$

4. Contemporaneous trade impacts on the efficient price

a. Description

This model incorporates trade impacts on the efficient price, which are presumed to reflect the informationcontent of the trades. The evolution of the efficient price is:

$$m_t = m_{t-1} + q_t\, V_t\, \lambda + u_t \tag{8}$$

where V_t is the volume of the trade, λ is the impact coefficient, and p_t is the observed log transaction price. The quantity $V_t\,\lambda$ can be interpeted as a product of scalars, or a vector product, as in: $V_t = \left[1\ \text{Vol}_t\ \ \sqrt{\text{Vol}_t}\,\right]$ and λ a 3×1 vector of coefficients. As in the basic model, q_t is the trade direction indicator, +1 for a buy and -1 for a sell. Thus $q_t\, V_t$ is the signed volume. The disturbance u_t is public information, and $u_t \sim N(0,\sigma_u^2)$. The mapping to the observed price is the same as in the Roll model:

$$p_t = m_t + c\, q_t$$

Relative to the basic Roll case discussed in the last section, this model is complicated by an additional parameter, λ, and dependence of the m_t on q_t.

The parameter draw is relatively straightforward, as the model implies

$$\Delta\, p_t = \Delta m_t + c\Delta q_t = \lambda q_t\, V_t + c\Delta q_t + u_t \tag{9}$$

At the parameter draw stage, $q_t V_t$ and Δq_t are known. This specification therefore fits in the regression framework, and λ and c may be drawn from the regression coefficient posterior.

The latent data draws are more complicated. Given the structure of the model,

$$\Pr[q_t \mid m_{\backslash t},\ q_{\backslash t},\ p_t] = \Pr[q_t \mid m_{t-1},\ m_{t+1},\ q_{t+1},\ p_t].$$

Analogous to eq. (5),

$$\Pr(q_t \mid p_t, m_{t-1}, m_{t+1}, q_{t+1}) = \frac{f(p_t \mid q_t, m_{t-1}, m_{t+1}, q_{t+1}) \times \Pr(q_t \mid m_{t-1}, m_{t+1}, q_{t+1})}{f(p_t \mid m_{t-1}, m_{t+1}, q_{t+1})} \quad (10)$$

As in the simple case, when we need to evaluate this, p_t is known, and so the denominator can be treated as a proportionality factor. Here, however, and in contrast with the Roll case, $f(q_t \mid m_{t-1}, m_{t+1}) \neq f(q_t) = 1/2$. Intuitively, if m_{t+1} is relatively high, this implies that q_t is more likely to be +1 (a buy). Therefore, we need to compute both of the factors in the numerator of eq. (10).

b. $\Pr(q_t \mid m_{t-1}, m_{t+1}, q_{t+1})$

The joint distribution of u_t and u_{t+1} is $\phi(0, \sigma_u, u_t)\,\phi(0, \sigma_u, u_{t+1})$. Substituting from (8) gives:

$$\phi(0, \sigma_u, -m_{t-1} + m_t - \lambda\, q_t\, V_t)\,\phi(0, \sigma_u, -m_t + m_{t+1} - \lambda\, q_{t+1}\, V_{t+1})$$

To compute $\Pr(q_t \mid m_{t-1}, m_{t+1}, q_{t+1})$, the dependence on m_t must be eliminated, which we do by integrating over m_t. That is,

$$\Pr(q_t \mid m_{t-1}, m_{t+1}, q_{t+1}) \propto \int_{-\infty}^{+\infty} \phi(0, \sigma_u, -m_{t-1} + m_t - \lambda\, q_t\, V_t)\,\phi(0, \sigma_u, -m_t + m_{t+1} - \lambda\, q_{t+1}\, V_{t+1})\, d m_t \equiv \Pr(q_t)$$

Evaluating the integral gives $\Pr(q_t) \equiv$

$$\frac{e^{-\frac{(m_{t-1} - m_{t+1} + \lambda\, q_t\, V_t + \lambda\, q_{t+1}\, V_{t+1})^2}{4\sigma_u^2}}}{2\sqrt{\pi}\,\sigma_u}$$

Performing the final normalization:

$$\Pr(q_t \mid m_{t-1}, m_{t+1}, q_{t+1}) = \frac{\Pr(q_t)}{\Pr(+1) + \Pr(-1)}$$

Therefore, $\Pr(q_t = 1 \mid m_{t-1}, m_{t+1}, q_{t+1}) =$

$$\frac{\Pr(1)}{\Pr(-1) + \Pr(1)}$$

For example, at the values

$$\{m_{t-1} = 5,\ m_{t+1} = 5.2,\ V_t = 1,\ q_{t+1} = 1,\ V_{t+1} = 2,\ \lambda = 0.01,\ \sigma_u = 0.05\}$$

the probability of a buy is 0.673. Intuitively, given the increase in the efficient price between m_{t-1} and m_{t+1}, it is more likely that the trade at time t was a buy.

c. $f(p_t \mid m_{t-1}, q_t, m_{t+1}, m_{t+1}, q_{t+1})$

As indicated above, the transition density is

$$\phi(0, \sigma_u, -m_{t-1} + m_t - \lambda\, q_t\, V_t)\, \phi(0, \sigma_u, -m_t + m_{t+1} - \lambda\, q_{t+1}\, V_{t+1}).$$

Expanding the normal densities and simplifying gives:

$$\frac{e^{-\frac{(m_{t-1} - m_t + \lambda\, q_t\, V_t)^2 + (m_t - m_{t+1} + \lambda\, q_{t+1}\, V_{t+1})^2}{2\,\sigma_u^2}}}{2\,\pi\,\sigma_u^2}$$

Therefore, m_t is normal: $f(m_t \mid m_{t-1}, m_{t+1}, q_t, q_{t+1}) = \phi(\mu_t, \sigma_t, m_t)$ with parameters

$$\mu_t = \left(\frac{1}{2}\,(m_{t-1} + m_{t+1} + \lambda\, q_t\, V_t - \lambda\, q_{t+1}\, V_{t+1})\right)$$
$$\sigma_t = \frac{\sigma_u}{\sqrt{2}}$$

Substituting in for m_t gives

$$f(p_t \mid m_{t-1}, m_{t+1}, q_t, q_{t+1}) = \phi(\mu_t, \sigma_t,\ p_t - c q_t)$$

d. Summary

The draw is made in the following steps.

1. Compute the normalized probabilities $\Pr(q_t \mid m_{t-1}, m_{t+1}, q_{t+1})$ for $q_t = -1$ and $q_t = +1$.

2. Compute $\phi(\mu_t, \sigma_t,\ p_t - c q_t)\Pr(q_t \mid m_{t-1}, m_{t+1}, q_{t+1})$ for $q_t = -1$ and $q_t = +1$.

3. Normalize the probabilities from step 2.

4. Make the draw.

Note that the overall calculation requires two normalizations, at step 1 (done, above, for the trial values) and step 3.

e. Endpoint modifications

(i) $t = 1$

At $t = 1$,

$$
\begin{aligned}
&\Pr(q_1 \mid m_{\backslash 1}, q_{\backslash 1}, p_1) \qquad\qquad\qquad .\\
&\quad = \Pr(q_1 \mid m_2, q_2, p_1)\\
&\qquad \propto f(p_1 \mid m_2, q_2, q_1) \times \Pr(q_1 \mid m_2, q_2)
\end{aligned}
$$

Here, however, $\Pr(q_1 \mid m_2, q_2) = \Pr(q_1) = 1/2$. To compute the first term on the r.h.s., the relevant density is $f(u_2) = \phi(0, \sigma_u, u_2)$. Substituting in for u_2 and $m_1$and solving indicates

$$f(m_1 \mid m_2, q_2, q_1) = f(m_1 \mid m_2, q_2) = \phi(\mu_1, \sigma_1, m_1)$$

where

$$
\begin{aligned}
&\mu_1 = (m_2 - \lambda\, q_2\, V_2)\\
&\sigma_1 = \sigma_u
\end{aligned}
$$

Thus,

$$f(p_1 \mid m_2, q_2, q_1) = \phi(\mu_1, \sigma_1, p_1 - c q_1)$$

We evaluate this. for $q_1 = \pm 1$, normalize, and make the draw of q_1.

(ii) *t = T*

At $t = T$,

$$
\begin{aligned}
&\Pr(q_T \mid m_{\backslash T}, q_{\backslash T}, p_T) \qquad\qquad\qquad .\\
&\quad = \Pr(q_T \mid m_{T-1}, q_{T-1}, p_T)\\
&\qquad \propto f(p_T \mid m_{T-1}, q_{T-1}, q_T) \times \Pr(q_T \mid m_{T-1}, q_{T-1})
\end{aligned}
$$

Again, $\Pr(q_T \mid m_{T-1}, q_{T-1}) = \Pr(q_{T-1}) = 1/2$. To compute the first term on the r.h.s., the relevant density is $f(u_T) = \phi(0, \sigma_u, u_T)$. Substituting in for u_T and m_T gives:

$$f(m_T \mid m_{T-1}, q_{T-1}, q_T) = \phi(\mu_T, \sigma_T, m_T)$$

where

$$
\begin{aligned}
&\mu_T = (m_{T-1} + \lambda\, q_T\, V_T)\\
&\sigma_T = \sigma_u
\end{aligned}
$$

We normalize and make the draw.

5. Discrete prices

a. Model Description

This model is a Roll model with price discreteness. The evolution of the (log) efficient price is:

$$m_t = m_{t-1} + u_t \text{ where } u_t \text{ is distributed as } N(0, \sigma_u^2)$$

The (level) efficient price is therefore $M_t = e^{m_t}$. (The units here are "dollars.") The half-spread in levels is C. The bid and ask quotes are:

$$B_t = \text{Floor}(M_t - C)$$
$$A_t = \text{Ceiling}(M_t + C)$$

Note that the rounding is asymmetric. The bid is rounded down to the next available increment; the ask is rounded up. It is assumed that the variables are scaled so that the tick size is unity. The buy/sell indicator is

$$q_t = \begin{cases} +1, \text{ a buy, with probability } 1/2 \\ -1, \text{ a sell, with probability } 1/2 \end{cases}$$

The obseved transaction price is

$$P_t = \begin{cases} A_t \text{ if } q_t = +1 \\ B_t \text{ if } q_t = -1 \end{cases}$$

In the model without discreteness, the cost parameter (here, C) can be estimated from the regression specification eq. (2). With the rounding transformations in the present case, there is no corresponding regression. We discuss the draw of C below.

The latent data draw is also complicated. Without discreteness, $p_t = m_t + c\, q_t$, so (given c) any variable in $\{p_t, m_t, q_t\}$ is determined by the other two. This is not the case here: q_t and m_t must be drawn separately. The steps are:

1. Draw $q_t \mid m_{\backslash t}, q_{\backslash t}, P_t$

2. Draw $m_t \mid m_{\backslash t}, q, P_t$

Note that the conditioning set at step 2 includes q_t.

b. Drawing $q_t \mid m_{\backslash t}, q_{\backslash t}, P_t$

From the properties of conditional probabilities:

$$\begin{aligned}
&\Pr(q_t \mid m_{t-1}, m_{t+1}, P_t)\\
&= \frac{\Pr(P_t \mid m_{t-1}, m_{t+1}, q_t) \times \Pr(q_t, \mid m_{t-1}, m_{t+1}) \times f(m_{t-1}, m_{t+1})}{\Pr(P_t \mid m_{t-1}, m_{t+1})\, f(m_{t-1}, m_{t+1})}\\
&\quad \propto \Pr(P_t \mid m_{t-1}, m_{t+1}, q_t) \times \Pr(q_t, \mid m_{t-1}, m_{t+1})\\
&\quad = \Pr(P_t \mid m_{t-1}, m_{t+1}, q_t) \times \Pr(q_t)
\end{aligned} \tag{11}$$

where we have used the fact that the denominator will be a constant of proportionality. The last equality follows from the fact that q_t is independent of the evolution of m_t. Next,

$$\Pr(P_t \mid m_{t-1}, m_{t+1}, q_t) = \int_{\underline{m}}^{\overline{m}} f(m_t \mid m_{t-1}, m_{t+1})\, dm_t \tag{12}$$

where $\underline{m}$ and $\overline{m}$ are the limits established by q_t and the Floor and Ceiling rounding functions (see below). $f(m_t \mid m_{t-1}, m_{t+1})$ is given in eq. (6). Using (12), we compute unnormalized probabilities for $q_t = +1$ and $q_t = -1$. We normalize these probabilities and draw q_t. We then proceed to the m_t draw.

The limits of integration in eq. (12) are computed as follows. From the definition of the Floor and Ceiling functions,

$$C + B_t < M_t < C + B_t + 1$$

and

$$-C + A_t - 1 < M_t < A_t - C$$

When $q_t = -1$, $P_t = B_t$ and:

$$C + P_t < M_t < C + P_t + 1$$

When $q_t = +1$, $P_t = A_t$ and:

$$-C + P_t - 1 < M_t < P_t - C$$

As long as $q_t \in \{-1, +1\}$, the lower and upper limits may be expressed as functions:

$$\begin{aligned}
&\underline{M}(q_t) = \left(\tfrac{1}{2}\,(2\,P_t - 2\,C\,q_t - q_t - 1)\right)\\
&\text{and}\\
&\overline{M}(q_t) = \left(P_t - \tfrac{1}{2}\,(2\,C + 1)\,q_t + \tfrac{1}{2}\right)
\end{aligned} \tag{13}$$

The limits on the log efficient price are then $\underline{m}(q_t) = \log[\underline{M}(q_t)]$ and $\overline{m}(q_t) = \log[\overline{M}(q_t)]$.

Recall that

$$f(m_t \mid m_{t-1}, m_{t+1}) = \phi[\mu_m, \sigma_m, m_t] \text{ with } \left\{\mu_m = (m_{t-1} + m_{t+1})/2,\ \sigma_m = \sigma_u \Big/ \sqrt{2}\right\}.$$

From eq. (12), $\Pr(q_t \mid m_{t-1}, m_{t+1}, P_t) \propto \Phi[\mu_m, \sigma_m, \underline{m}(q_t), \overline{m}(q_t)]$ for $q_t = \pm 1$. Therefore, the normalized probability of a buy is:

$$\frac{\Phi(\mu_m, \sigma_m, \underline{m}(+1), \overline{m}(+1))}{\Phi(\mu_m, \sigma_m, \underline{m}(-1), \overline{m}(-1)) + \Phi(\mu_m, \sigma_m, \underline{m}(+1), \overline{m}(+1))}$$

At the values $\{m_{t-1} = \log(100),\ m_{t+1} = \log(104),\ P_t = 101,\ C = 0.2,\ \sigma_u = 0.01\}$ this becomes 0.092. Intuitively, given the locations of m_{t-1} and m_{t+1}, $Em_t \approx 102$. Since P_t lies below this, it is more likely that the transaction was a sale, and the buy probability is below one-half.

c. Drawing $m_t \mid m_{\backslash t}, q, P_t$

As noted, $f(m_t \mid m_{t-1}, m_{t+1}) = \phi[\mu_m, \sigma_m, m_t]$ where μ_m and σ_m are given above. Conditioning on q_t and P_t merely restricts the range of the distribution. Therefore, the draw is from the normal distribution $\phi[\mu_m, \sigma_m, m_t]$ truncated to the range $(\underline{m}[q_t], \overline{m}[q_t])$.

d. Drawing $C \mid P, m, q$

The upper and lower bounds on C, given everything else, can be computed in a fashion similar to that used above for M_t. If $q_t = -1$,

$$M_t - P_t - 1 < C < M_t - P_t$$

If $q_t = +1$,

$$-M_t + P_t - 1 < C < P_t - M_t$$

Written as a linear function of q_t, the lower and upper bounds on C are:

$$\underline{C_t} = (-M_t + P_t - \tfrac{1}{2})\, q_t - \tfrac{1}{2}$$

and

$$\overline{C_t} = (-M_t + P_t + \tfrac{1}{2})\, q_t - \tfrac{1}{2}$$

These bounds cause difficulty because, if we are taking P and M as fixed when we make the C draw, then the new value C^* is bounded by:

$$\operatorname*{Max}_t \underline{C_t} \le C^* \le \operatorname*{Min}_t \overline{C_t}$$

As the maxima and minima are over the full sample, these bounds are likely to be extremely confining. This is likely to prevent the C draws from mixing well.

e. Alternative approach: A joint draw of $C, m \mid P, q$

The preceding discussion demonstrated the difficulty of drawing C from its full conditional distribution. One way of viewing the problem is to note that if a hypothetical new draw of C were to move more than a very small step away from the present value, it would be likely that for some observation in the sample, the existing value of m_t lie outside of the $(\underline{m_t}, \overline{m}_t)$ bounds implied by the newly drawn C, i.e., that the new draw would not be feasible. This suggests that a new draw of C would have greater latitude if new m_t were also drawn as well.

This is feasible. An overview of the procedure is as follows. We start with the current values of C and m. Suppose for the moment that we have some new candidate draw C^*. We assume only that $C^* > 0$. It might be the case that C^* implies infeasibility for some or all elements of the current m. A simple way of generating new feasible m is via a deterministic shift. Consider the upper and lower bounds for the level efficient price given in eq. (13), holding P_t and q_t fixed. In moving from C to C^*, these bounds shift from $(\underline{M_t}, \overline{M_t})$ to $(\underline{M_t^*}, \overline{M_t^*})$. The shifts in bounds are identical: $\underline{M_t^*} - \underline{M_t} = \overline{M_t^*} - \overline{M_t}$. This suggests setting the new value

$$M_t^* = \underline{M_t^*} + (M_t - \underline{M_t})$$

(and setting $m_t^* = \log(M_t^*)$). In doing this, we are keeping the relative position of the efficient price the same within the new set of bounds.

There are two remaining problems. First, how should the C^* be generated? Second, how can this procedure be integrated into the Gibbs sweep in a valid manner?

The first issue is straightforward. In the basic Roll model, the new draws for the cost parameter c arose from the linear regression $\Delta p_t = c\Delta q_t + u_t$ (where the variables in the model were in logs). This equation doesn't hold in the present case due to the discreteness transformations. It does suggest, however, a way for generating new values of C. We could randomly generated $U_t \sim N(0, \sigma_U^2)$, and then set up the regression:

$$(C\Delta q_t + U_t) = C^* \Delta q_t + U_t \qquad (14)$$

That is, we use the current value C to simulate a dependent variable, regress it against the Δq_t, and draw C^* from the coefficient posterior. Obviously, this regression loosely corresponds to the one used in the basic model.

This procedure defines what is generally termed a candidate or proposal distribution for generating a new parameter value from an existing one. It defines a conditional distribution $g(C^* \mid C)$. The convention, however, is to write this as $g(C, C^*)$, where g is a transition density in a Markov chain. Since a new value C^* implies new values m^*, this candidate/proposal/transition density is more properly written $g(C, m, C^*, m^*)$. (In this case, though, once we've chosen the new value C^*, however, the new m^* are automatically determined.)

We now turn to the second question, viz., how to use this candidate density. The Gibbs sampler turns out to be a special case of a more general sampling scheme called the Metropolis-Hastings sampler. The program for this sampler can be described as follows. Suppose that we enter sweep $j+1$ with values $C^{(j)}$ and $m^{(j)}$. Then:

1. Draw C^* and m^* from the candidate density $g(C, m, C^*, m^*)$

2. Compute

$$\alpha = \text{Min}\left(1, \frac{f(C^*, m^*)\, g(C^*, m^*, C^{(j)}, m^{(j)})}{f(C^{(j)}, m^{(j)})\, g\left(C^{(j)}, m^{(j)}, C^*, m^*\right)}\right) \tag{15}$$

where $f(C^*, m^*)$ is the true ("target") density evaluated at the proposal draws, $f(C^{(j)}, m^{(j)})$ is the true density evaluated at the sweep-j values, $g\left(C^{(j)}, m^{(j)}, C^*, m^*\right)$ is the transition probability associated with generating the new values (conditional on the old ones), and $g(C^*, m^*, C^{(j)}, m^{(j)})$ is the reverse transition probability (i.e., the probability of generating the old values if we'd started with the new ones).

3. Generate z a uniform random number between zero and one.

4. If $z < \alpha$, "accept the proposal". That is, set $C^{(j+1)} = C^*$ and $m^{(j+1)} = m^*$.
If $z \geq \alpha$, "reject the proposal". That is, set $C^{(j+1)} = C^{(j)}$ and $m^{(j+1)} = m^{(j)}$.

As with the Gibbs sampler, in the limit as $j \to \infty$, $(C^{(j)}, m^{(j)})$ are random numbers distributed in accordance with the target density f.

In the present case, the target density is the conditional density $f(C, m \mid \sigma_u, q, P)$, i.e., the full conditional density given the observed data, the latent data and remaining parameter.

$$f(C, m \mid \sigma_u, q, P) = \frac{\Pr(P \mid m, q, C, \sigma_u)\, f(m \mid q, C, \sigma_u)\, \Pr(q \mid C, \sigma_u)\, f(C, \sigma_u)}{\Pr(P, q \mid \sigma_u)\, f(\sigma_u)} \tag{16}$$

Since the mapping from m, q, and C to P is nonstochastic, $\Pr(P \mid m, q, C, \sigma_u) = 1$. From the assumed independence of the prior distributions,

$$\Pr(q \mid C, \sigma_u)\, f(C, \sigma_u) = \Pr(q)\, f(C)\, f(\sigma_u) \tag{17}$$

Therefore

$$f(C, m \mid \sigma_u, q, P) \propto f(m \mid q, C, \sigma_u)\, f(C) = f(m \mid \sigma_u)\, f(C) \tag{18}$$

(since C and q are independent of m). These two quantities are simple to evaluate for the existing and proposed draws.

In drawing from the proposal density, it is not really necessary to simulate the U_t and actually run the regression given in eq. (14). Suppose that the prior for C is $N^{+}(0, \Omega_C^{\text{Prior}})$. Using the classic Bayesian regression results from section 1.1, if we simulated and ran the regression, the posterior would be $N^{+}(\mu_C^{\text{Posterior}}, \Omega_C^{\text{Posterior}})$ with

$$\mu_C^{\text{Posterior}} = \left(\frac{\sum_t \Delta q_t^2}{\sigma_U^2} + \frac{1}{\Omega_C^{\text{Prior}}} \right)^{-1} \left(\frac{\sum_t \Delta q_t (C \Delta q_t + U_t)}{\sigma_U^2} \right)$$

and

$$\Omega_C^{\text{Posterior}} = \left(\frac{\sum_t \Delta q_t^2}{\sigma_U^2} + \frac{1}{\Omega_C^{\text{Prior}}} \right)^{-1}$$

A sensible expedient is to take

$$\mu_C^{\text{Posterior}} = \left(\frac{\sum_t \Delta q_t^2}{\sigma_U^2} + \frac{1}{\Omega_C^{\text{Prior}}} \right)^{-1} \left(\frac{C \sum_t \Delta q_t^2}{\sigma_U^2} \right)$$

Since C is a cost parameter in levels (rather than logs), U_t is also a level disturbance. A logical choice for σ_U^2 is $\overline{P}\,\sigma_u^2$ where σ_u is the current draw of this parameter and $\overline{P}$ is the average price level over the sample. In taking these shortcuts, we are departing from the full conditional distribution. This is permissible because $g(\ldots)$ is only a proposal density.

Although this draw is joint over both C and m, it does not replace the m draw described above. The reason is that this joint draw cannot generate all values of m; it only shifts the values in a rather limited fashion.

6. Clustered prices

a. Model description

This is an extension of the simple discreteness model. The evolution of the (log) efficient price is:

$$m_t = m_{t-1} + u_t \text{ where } u_t \sim N(0, \sigma_u^2)$$

The (level) efficient price is therefore $M_t = e^{m_t}$. (The units here are "dollars.") The half-spread in levels is C. The bid and ask quotes are:

$$B_t = \text{Floor}(M_t - C,\ K_t)$$
$$A_t = \text{Ceiling}(M_t + C,\ K_t)$$

The rounding functions here round down or up to the nearest K-multiple of the "official" tick size. It is assumed that the variables are scaled so that the minimum tick size is unity. The K_t are then i.i.d. Bernoulli variates:

$$K_t = \begin{cases} 1, & \text{with probability}\,(1-k) \\ \kappa, & \text{with probability}\,k \end{cases}$$

κ is a "natural multiple" of the basic tick size, like 2, 5 or 10. k is the clustering probability. The buy/sell indicator is

$$q_t = \begin{cases} +1, & \text{a buy, with probability } \frac{1}{2} \\ -1, & \text{a sell, with probability } \frac{1}{2} \end{cases}$$

The obseved transaction price is

$$P_t = \begin{cases} A_t \text{ if } q_t = +1 \\ B_t \text{ if } q_t = -1 \end{cases}$$

In the latent data draw, at each time t, we need to draw $m_t,\ q_t,\ K_t \mid m_{\backslash t},\ q_{\backslash t},\ P_t$. The draw is made in two steps:

1. Draw $q_t,\ K_t \mid m_{t-1},\ m_{t+1},\ P_t$.

2. Draw $m_t \mid m_{t-1},\ m_{t+1},\ q_t,\ K_t,\ P_t$.

The parameter draw for C proceeds exactly as in discreteness model. (That is, we make a joint draw of C and m. See section 5.) There is also an additional parameter draw, $k \mid K$.

b. The draw for $q_t,\ K_t \mid m_{t-1},\ m_{t+1},\ P_t$

We have:

$$\begin{aligned} &\Pr(q_t,\ K_t \mid m_{t-1},\ m_{t+1},\ P_t) \\ &\quad = \frac{\Pr(P_t \mid m_{t-1},\ m_{t+1},\ q_t,\ K_t) \times \Pr(q_t,\ K_t \mid m_{t-1},\ m_{t+1}) \times f(m_{t-1},\ m_{t+1})}{\Pr(P_t \mid m_{t-1},\ m_{t+1})\ f(m_{t-1},\ m_{t+1})} \\ &\quad \propto \Pr(P_t \mid m_{t-1},\ m_{t+1},\ q_t,\ K_t) \times \Pr(q_t,\ K_t \mid m_{t-1},\ m_{t+1}) \\ &\quad = \Pr(P_t \mid m_{t-1},\ m_{t+1},\ q_t,\ K_t) \times \Pr(q_t) \Pr(K_t) \end{aligned} \tag{19}$$

The last equality follows from the fact that q_t and K_t are mutually independent and also independent of the evolution of m_t. Next,

$$\Pr(P_t \mid m_{t-1},\ m_{t+1},\ q_t,\ K_t) = \int_{\underline{m}}^{\overline{m}} f(m_t \mid m_{t-1},\ m_{t+1})\, dm_t$$

where $\underline{m}$ and $\overline{m}$ are the limits established by q_t and K_t (see below). The general plan is as follows. Using (19), we compute unnormalized probabilities for the four possible combinations of q_t and K_t. We normalize these probabilities and make the draw of the $\{q_t, K_t\}$ pair. We then proceed to the m_t draw. We turn now to the details.

The floor and ceiling functions in the model imply

$$C + B_t < M_t < C + B_t + K_t$$

and

$$-C + A_t - K_t < M_t < A_t - C$$

When $q_t = -1$, $B_t = P_t$, and

$$C + P_t < M_t < C + K_t + P_t$$

When $q_t = +1$, $A_t = P_t$, and

$$-C - K_t + P_t < M_t < P_t - C$$

The lower and upper limit functions for M are:

$$\underline{M}(q_t, K_t) = \frac{1}{2}(-q_t K_t - K_t + 2P_t - 2C q_t)$$

and

$$\overline{M}(q_t, K_t) = P_t + \frac{1}{2}(K_t - (2C + K_t) q_t)$$

The limits on the log efficient price are then $\underline{m}[q_t, K_t] = \log[\underline{M}(q_t, K_t)]$ and $\overline{m}[q_t, K_t] = \log[\overline{M}(q_t, K_t)]$.

From eq. (19), and using the fact that $\Pr(q_t) = 1/2$,

$$\Pr(q_t, K_t \mid m_{t-1}, m_{t+1}, P_t) \propto \Pr(P_t \mid m_{t-1}, m_{t+1}, q_t, K_t) \times \Pr(K_t) \qquad (20)$$

Recall that $f(m_t \mid m_{t-1}, m_{t+1}) = \phi(\mu_m, \sigma_m, m_t)$ with $\mu_m = (m_{t-1} + m_{t+1})/2$ and $\sigma_m = \sigma_u \big/ \sqrt{2}$. Thus,

$$\Pr(q_t, K_t \mid m_{t-1}, m_{t+1}, P_t) \propto \Phi[\mu_m, \sigma_m, \underline{m}(q_t, K_t), \overline{m}(q_t, K_t)] \Pr(K_t)$$

for $(q_t, K_t) \in \{+1, -1\} \times \{1, \kappa\}$. Let i index the four possible combinations of (q_t, K_t), and let

$$Z_i = \Phi[\mu_m, \sigma_m, \underline{m}(q_t, K_t), \overline{m}(q_t, K_t)] \Pr(K_t)$$

evaluated at the the *i*th values. Then the normalized probability of outcome i is $Z_i / \Sigma Z_i$.

c. The draw for $m_t \mid m_{t-1}, m_{t+1}, q_t, K_t, P_t$

This is simply a draw from $f(m_t \mid m_{t-1}, m_{t+1})$ truncated to $(\underline{m}, \overline{m})$.

d. The parameter draw for $k \mid K$

The clustering variables are defined by $K = \{K_1, \ ..., K_T\}$, where each $K_t \in \{1, \kappa\}$. Let n be the number of t for which $K_t = \kappa$. If the probability of $K_t = \kappa$ is k, then n is a binomial random variable. The beta distribution is conjugate to the binomal (Tanner, 1996), so Beta$[a, b]$ is a convenient prior for k. We let $a^{\text{Prior}} = b^{\text{Prior}} = 1/2$, for which the Beta density is uniform on the unit interval. The posterior for k is then Beta$[a^{\text{Posterior}}, b^{\text{Posterior}}]$ where $a^{\text{Posterior}} = a^{\text{Prior}} + n$ and $b^{\text{Posterior}} = b^{\text{Prior}} + T - n$.

7. Lagged trade dependencies

a. Model description

In this extension of the informative trade model, the evolution of the efficient price is allowed to depend on lagged signed trades (as well as the contemporaneous one).

$$m_t = m_{t-1} + \Lambda_t + u_t \tag{21}$$

where Λ_t is the impact term:

$$\Lambda_t = \sum_{j=0}^{\text{Min}(J,t-1)} q_{t-j}\, \lambda_j\, V_{t-j}$$

J is the order of the lagged dependence. The upper limit of the sum indicates that the lags do not extend before the beginning of the sample. The observed price is:

$$p_t = m_t + c q_t$$

The parameter draw for $\{\lambda_0, \ ..., \lambda_J, c, \sigma_u\}$ is (mostly) straightforward. Given the observed and latent data,

$$\Delta p_t = \Delta m_t + c\Delta q_t = \Lambda_t + c\Delta q_t + u_t = \sum_{j=0}^{\text{Min}(J,t-1)} \lambda_j\, q_{t-j}\, V_{t-j} + c\Delta q_t + u_t \tag{22}$$

This fits into the Bayesian multiple regression framework (with c and the λ_j the coefficients).

There is the potential for multicollinearity here, however. Suppose that $J = 1$ and $V_t = 1$. Then

$$\Delta p_t = \lambda_0\, q_t + \lambda_1\, q_{t-1} + c\Delta q_t + u_t$$

In economic terms, the impact of a trade on the efficient price depends only on the direction of the trade, and not its size. Since $\Delta q_t = q_t - q_{t-1}$, however, the cross-product matrix is singular. Some sort of additional structure is needed. (One possibility is to simply let $V_t = \text{Vol}_t$.)

The latent data draw is of $q_t \mid m_{\backslash t},\ p_t,\ q_{\backslash t}$. This follows the design of the draw for the contemporaneous impact model, but the lagged structure extends the q_t dependencies. From the structure of the sum, it is clear that q_t influences $m_t,\ \ldots,\ m_{t+P}$. Therefore $\Pr(q_t \mid m_{\backslash t},\ q_{\backslash t})$ depends on $\prod_{s=0}^{J} f(u_{t+s})$. From inspection of these terms,

$$\Pr(q_t \mid m_{\backslash t},\ p_t,\ q_{\backslash t}) = \Pr(q_t \mid m_{t-1},\ m_{t+1},\ \ldots m_{t+P},\ p_t,\ q_{t-P},\ \ldots,\ q_{t-1},\ q_{t+1},\ \ldots,\ q_{t+P}).$$

For notational simplicity, however, the conditioning set indicated on the l.h.s. will be used below.

Analogous to eq. (10),

$$\Pr(q_t \mid m_{\backslash t},\ p_t,\ q_{\backslash t}) = \frac{f(p_t \mid m_{\backslash t},\ q_{\backslash t},\ q_t) \times \Pr(q_t \mid m_{\backslash t},\ q_{\backslash t})}{f(p_t \mid m_{\backslash t},\ q_{\backslash t})} \tag{23}$$

Since $q = \{q_{\backslash t},\ q_t\}$, the first term in the numerator could have been written more concisely (but less clearly) as $f(p_t \mid m_{\backslash t},\ q)$. The denominator can be treated as a proportionality factor. The two terms in the numerator must be evaluated separately.

b. Evaluation of $\Pr(q_t \mid m_{\backslash t},\ q_{\backslash t})$

This probability is derived from the joint density of $u_t,\ \ldots,\ u_{\text{Min}(t+J,T)}$, which is $\prod_{s=0}^{\text{Min}(J,T-t)} f(u_{t+s})$. When $t < T$, it is convenient to write this as:

$$\prod_{s=0}^{\text{Min}(J,T-t)} f(u_{t+s}) = \left(\prod_{s=t+2}^{J+t} \phi(0,\ \sigma_u,\ u_s)\right) \phi(0,\ \sigma_u,\ u_t)\, \phi(0,\ \sigma_u,\ u_{t+1}) \tag{24}$$

Here, we've pulled out the two terms that depend on m_t. Substituting in for u_t and u_{t+1}, these terms are:

$$\phi(0,\ \sigma_u,\ -m_{t-1} + m_t - \Lambda_t)\, \phi(0,\ \sigma_u,\ -m_t + m_{t+1} - \Lambda_{t+1})$$

Integrating out m_t gives:

$$\frac{e^{-\frac{(m_{t-1} - m_{t+1} + \Lambda_t + \Lambda_{t+1})^2}{4\sigma_u^2}}}{2\sqrt{\pi}\,\sigma_u}$$

Replacing this in the joint density gives $\Pr(q_t \mid m_{\backslash t},\ q_{\backslash t}) \propto$

$$\frac{e^{-\frac{(m_{t-1}-m_{t+1}+\Lambda_t+\Lambda_{t+1})^2}{4\sigma_u^2}}\prod_{s=t+2}^{\text{Min}(J+t,T)}\phi(0,\sigma_u,u_s)}{2\sqrt{\pi}\,\sigma_u}$$

We evaluate this for $q_t = \pm 1$, normalize and make the draw.

The following intermediate calculations are used in the C++ code:

$$-\frac{(m_{t-1}-m_{t+1}+\Lambda_t+\Lambda_{t+1})^2}{4\sigma_u^2} - \log(\sigma_u) - \frac{\log(\pi)}{2} - \log(2)$$

$$-\frac{u_s^2}{2\sigma_u^2} - \log(\sigma_u) + \frac{1}{2}\,(-\log(2) - \log(\pi))$$

$$m_{t-1} - m_{t+1} + \Lambda_t + \Lambda_{t+1}$$

We evaluate this expression for $q_t = \pm 1$ (recognizing that the u_s for $t + 2 \le s \le t + P$ also depend on q_t), and normalize. We now turn to the first component in the numerator of (1).

c. Evaluation of $f(p_t \mid m_{\backslash t},\ q)$

Consider again from the joint density for u_t, ..., u_{t+P} in eq. (24). The first two terms are:

$$\phi(0,\ \sigma_u,\ -m_{t-1} + m_t - \Lambda_t)\,\phi(0,\ \sigma_u,\ -m_t + m_{t+1} - \Lambda_{t+1})$$

Using the definition for the normal density and simplifying gives:

$$\frac{e^{-\frac{(m_{t-1}-m_t+\Lambda_t)^2+(m_t-m_{t+1}+\Lambda_{t+1})^2}{2\sigma_u^2}}}{2\pi\sigma_u^2}$$

This is proportional to $\phi(\mu_t,\ \sigma_t,\ m_t)$ where

$$\mu_t = \left(\frac{1}{2}\,(m_{t-1} + m_{t+1} + \Lambda_t - \Lambda_{t+1})\right)$$
$$\sigma_t = \frac{\sigma_u}{\sqrt{2}}$$

Thus,

$$f(m_t \mid m_{\backslash t}, q_{\backslash t}, q_t) \propto$$ (25)

$$\left(\prod_{s=t+2}^{P+t} \phi(0, \sigma_u, u_s)\right) \phi(\mu_t, \sigma_t, m_t)$$

Substituting in for m_t gives $f(p_t \mid m_{\backslash t}, q_{\backslash t}, q_t) =$

$$\left(\prod_{s=t+2}^{P+t} \phi(0, \sigma_u, u_s)\right) \phi(\mu_t, \sigma_t, p_t - c\, q_t)$$

d. Summary

To draw $q_t \mid m_{\backslash t}, q_{\backslash t}, p_t$, the steps are:

1. Compute the normalized $\Pr(q_t = -1 \mid m_{\backslash t}, q_{\backslash t})$ and $\Pr(q_t = +1 \mid m_{\backslash t}, q_{\backslash t})$

2. Compute $f(p_t \mid m_{\backslash t}, q_{\backslash t}, q_t = -1)$ and $f(p_t \mid m_{\backslash t}, q_{\backslash t}, q_t = +1)$

3. Compute the unnormalized probabilities $\pi(-1) = \Pr(q_t = -1 \mid m_{\backslash t}, q_{\backslash t})\, f(p_t \mid m_{\backslash t}, q_{\backslash t}, q_t = -1)$ and $\pi(+1) = \Pr(q_t = +1 \mid m_{\backslash t}, q_{\backslash t})\, f(p_t \mid m_{\backslash t}, q_{\backslash t}, q_t = +1)$.

4. The normalized probability of a buy is $\Pr(q_t = +1 \mid m_{\backslash t}, q_{\backslash t}, p_t) = \frac{\pi(+1)}{\pi(+1) + \pi(-1)}$

5. Use this probability to make the Bernoulli draw.

e. Endpoint modifications

(i) *t* = 1

At $t = 1$, the relevant joint density is $\left(\prod_{s=3}^{J} \phi(0, \sigma_u, u_s)\right) \phi(0, \sigma_u, u_2)$. (The product term is dropped when $J < 3$.) Consider the $\phi(0, \sigma_u, u_2)$ term. Substituting in for u_2 gives:

$$\phi(0, \sigma_u, -m_1 + m_2 - \Lambda_2)$$

Integrating over $m_1$gives 1. Therefore $\Pr(q_1 \mid m_{\backslash 1}, q_{\backslash 1}) \propto \prod_{s=3}^{J} \phi(0, \sigma_u, u_s)$. Compute the r.h.s for $q_1 = \pm 1$and normalize. If $J < 3$, then $\Pr(q_1) = 1/2$.

The function above is equivalent to $\phi(\mu_1, \sigma_1, m_1)$ with

$$\mu_1 = (m_2 - \Lambda_2)$$
$$\sigma_1 = \sigma_u$$

Therefore $f(m_1 \mid m_{\backslash 1}, q_{\backslash 1}, q_1) =$

$$\left(\prod_{s=3}^{P}\phi(0,\sigma_u,u_s)\right)\phi(\mu_1,\sigma_1,m_1)$$

Or, $f(p_1 \mid m_{\backslash 1}, q_{\backslash 1}, q_1) =$

$$\left(\prod_{s=3}^{P}\phi(0,\sigma_u,u_s)\right)\phi(\mu_1,\sigma_1,\ p_1 - c\,q_1)$$

Since $\Pr(q_1 \mid m_{\backslash 1}, q_{\backslash 1},\ p_1) \propto f(p_1 \mid m_{\backslash 1}, q)\Pr(q_1 \mid m_{\backslash 1}, q_{\backslash 1})$, we may compute the r.h.s., normalize and make the draw.

(ii) ***t = T***

For $t = T$, the relevant joint density is $\phi(0, \sigma_u, u_T)$. Substituting for u_T:

$$\phi(0,\sigma_u,-m_{T-1}+m_T-\Lambda_T)$$

Integrating out m_T gives 1, so $\Pr(q_T \mid m_{\backslash T}, q_{\backslash T}) = 1/2$. From the above expression, $f(m_T \mid m_{\backslash T}, q_{\backslash T}, q_T) = \phi(\mu_T, \sigma_T, m_T)$ with

$$\mu_T = (m_{T-1} + \Lambda_T)$$
$$\sigma_T = \sigma_u$$

Therefore $f(p_T \mid m_{\backslash T}, q_{\backslash T}, q_T) = \phi(\mu_T, \sigma_T,\ p_T - cq_T)$.

Using these results, compute $f(p_T \mid m_{\backslash T}, q_{\backslash T}, q_T)$ at $q_T = \pm 1$ and normalize. (This is the first normalization.) Make the draw for q_T.

8. Combined model I: Trade dependencies, discreteness and clustering

a. Model description

This model combines the features considered separately in earlier sections. The efficient price evolution reflects a contemporaneous trade impact:

$$m_t = m_{t-1} + q_t\,\lambda V_t + u_t \tag{26}$$

The mapping to observed prices follows the discreteness/clustering model. The bid and ask quotes are:

$$B_t = \text{Floor}(M_t - C,\ K_t)$$
$$A_t = \text{Ceiling}(M_t + C,\ K_t)$$

where $M_t = e^{m_t}$; C is the half-spread; the rounding functions here round down or up to the nearest K-multiple of the "official" tick size. The clustering multiple is K_t , an i.i.d. Bernoulli variate:

$$K_t = \begin{cases} 1, & \text{with probability } (1-k) \\ \kappa, & \text{with probability } k \end{cases}$$

κ is a "natural multiple" of the basic tick size, like 2, 5 or 10. k is the clustering probability. The buy/sell indicator is

$$q_t = \begin{cases} +1, & a \text{ buy, with probability } \frac{1}{2} \\ -1, & a \text{ sell, with probability } \frac{1}{2} \end{cases}$$

The obseved transaction price is

$$P_t = \begin{cases} A_t \text{ if } q_t = +1 \\ B_t \text{ if } q_t = -1 \end{cases}$$

The parameter draws are minor modifications of those given above. That is, $\{\lambda, \sigma_u\}$ are modeled in the regression:

$$\Delta m_t = \lambda q_t + u_t \qquad (27)$$

The clustering parameter k is modeled in the beta/binomial framework. C is drawn using the Metropolis-Hastings algorithm described in section 5.

In the latent data draw, at each time t, we need to draw $m_t, q_t, K_t \mid m_{\backslash t}, q_{\backslash t}, P_t$. The draw is made in two steps:

1. Draw $q_t, K_t \mid m_{\backslash t}, q_{\backslash t}, P_t$.

2. Draw $m_t \mid m_{\backslash t}, q, K_t, P_t$.

Given the structure of the model, $\Pr(q_t, K_t \mid m_{\backslash t}, q_{\backslash t}, K_{\backslash t}, P) = \Pr(q_t, K_t \mid m_{t-1}, m_{t+1}, q_{t+1}, P_t)$. Analogous to eq. (19),

$$\Pr(q_t, K_t \mid m_{t-1}, m_{t+1}, q_{t+1}, P_t)$$

$$= \frac{\Pr(P_t \mid m_{t-1}, m_{t+1}, q_{t+1}, q_t, K_t) \times \Pr(q_t, K_t \mid m_{t-1}, m_{t+1}, q_{t+1})}{\Pr(P_t \mid m_{t-1}, m_{t+1}, q_{t+1})} \qquad (28)$$

$$\propto \Pr(P_t \mid m_{t-1}, m_{t+1}, q_{t+1}, q_t, K_t) \times \Pr(q_t \mid m_{t-1}, m_{t+1}, q_{t+1}) \times \Pr(K_t)$$

The denominator is subsumed into the proportionality constant; K_t is independent of the other variables. The first two terms must be evaluated separately.

b. $\Pr(q_t \mid m_{t-1}, m_{t+1}, q_{t+1})$

This development is identical to the one given for the asymmetric information model above. The joint distribution of u_t and $u_{t+1}$is $\phi(0, \sigma_u, u_t)\,\phi(0, \sigma_u, u_{t+1})$. Substituting from (26) gives:

$$\phi(0, \sigma_u, -m_{t-1} + m_t - \lambda\, q_t\, V_t)\,\phi(0, \sigma_u, -m_t + m_{t+1} - \lambda\, q_{t+1}\, V_{t+1})$$

Integrating out m_t gives:

$$\frac{e^{-\frac{(m_{t-1} - m_{t+1} + \lambda\, q_t\, V_t + \lambda\, q_{t+1}\, V_{t+1})^2}{4\,\sigma_u^2}}}{2\sqrt{\pi}\;\sigma_u}$$

Evaluating this at $q_t = +1$ and $q_t = -1$ and normalizing gives $\Pr(q_t = +1 \mid m_{t-1}, m_{t+1}, q_{t+1}) =$

$$\frac{e^{\frac{\lambda\, m_{t+1}\, V_t}{\sigma_u^2}}}{e^{\frac{\lambda\, m_{t+1}\, V_t}{\sigma_u^2}} + e^{\frac{\lambda\, V_t\, (m_{t-1} + \lambda\, q_{t+1}\, V_{t+1})}{\sigma_u^2}}}$$

We combine this with the fact that $\Pr(K_t = \kappa) = k$ to obtain the full set of $\Pr(q_t, K_t \mid m_{t-1}, m_{t+1}, q_{t+1})$ for $(q_t, K_t) \in \{-1, +1\} \times \{1, \kappa\}$.

c. $\Pr(P_t \mid m_{t-1}, m_{t+1}, q_{t+1}, q_t, K_t)$

This is obtained by integrating $f(m_t \mid q_t, K_t, m_{t-1}, m_{t+1}, q_{t+1})$ over the range implied by q_t and K_t:

$$\Pr(P_t \mid m_{t-1}, m_{t+1}, q_{t+1}, q_t, K_t) = \int_{\underline{m}}^{\overline{m}} f(m_t \mid q_t, K_t, m_{t-1}, m_{t+1}, q_{t+1})\, dm_t \qquad (29)$$

where $\underline{m} = \underline{m}(q_t, K_t)$ and $\overline{m} = \overline{m}(q_t, K_t)$ are the bounding functions derived for the clustering model. The integrand is:

$$\phi(0, \sigma_u, -m_{t-1} + m_t - \lambda\, q_t\, V_t)\,\phi(0, \sigma_u, -m_t + m_{t+1} - \lambda\, q_{t+1}\, V_{t+1})$$

This can be reworked as the normal density $\phi(\mu_m, \sigma_m, m_t)$ where the parameters are:

$$\mu_m = \left(\frac{1}{2}\,(m_{t-1} + m_{t+1} + \lambda\, q_t\, V_t - \lambda\, q_{t+1}\, V_{t+1})\right)$$
$$\sigma_m = \frac{\sigma_u}{\sqrt{2}}$$

We now use the result from eq. (28) that

$$\Pr(q_t, K_t \mid m_{t-1}, m_{t+1}, q_{t+1}, P_t) \propto \Pr(P_t \mid m_{t-1}, m_{t+1}, q_{t+1}, q_t, K_t) \times \Pr(q_t \mid m_{t-1}, m_{t+1}, q_{t+1}) \times \Pr(K_t) \quad (30)$$

We compute the r.h.s. for all values of $\{q_t, K_t\}$, normalize and make the draw.

9. Combined model II: Lagged trade dependencies, discreteness and clustering

This is the richest model estimated in the paper, involving all of the features considered to this point. The model is closest to the one considered in the last section. In this version, the efficient price evolution reflects lagged trade impacts:

$$m_t = m_{t-1} + \Lambda_t + u_t \quad (31)$$

where Λ_t is the impact term:

$$\Lambda_t = \sum_{j=0}^{\text{Min}(J,t-1)} q_{t-j} \lambda_j V_{t-j}$$

J is the order of the lagged dependence. The mapping to observed prices is the same as in the previous section.

The draw strategies are minor modifications of those encountered earlier.

For the parameter draws, $\{\lambda_j, \sigma_u\}$ are modeled in the regression:

$$\Delta m_t = \sum_{j=0}^{\text{Min}(J,t-1)} \lambda_j q_{t-j} V_{t-j} + u_t \quad (32)$$

In discussing the corresponding regression for the lagged trade impact model without discreteness or clustering (cf. Section 7), the potential for multicollinearity was noted. This arose from the presence (in eq. (22)) of a cΔq_t term. This term is absent in eq. (32). This alleviates the problem of multicollinearity in the regression. The problem persists in the broader model, however, if the tick size is small relative to the price changes.

The clustering parameter k is modeled in the beta/binomial framework. C is drawn using the Metropolis-Hastings algorithm.

In the latent data draw, at each time t, we need to draw $m_t, q_t, K_t \mid m_{\backslash t}, q_{\backslash t}, P_t$. The draw is made in two steps:

1. Draw $q_t, K_t \mid m_{\backslash t}, q_{\backslash t}, P_t$.

2. Draw $m_t \mid m_{\backslash t}, q, K_t, P_t$.

To derive the probabilities for the first draw, note:

$$\Pr(q_t, K_t \mid m_{\setminus t}, q_{\setminus t}, P_t)$$

$$= \frac{\Pr(P_t \mid m_{\setminus t}, q_{\setminus t}, q_t, K_t) \times \Pr(q_t, K_t \mid m_{\setminus t}, q_{\setminus t})}{\Pr(P_t \mid m_{\setminus t}, q_{\setminus t})} \qquad (33)$$

$$\propto \Pr(P_t \mid m_{\setminus t}, q_{\setminus t}, q_t, K_t) \times \Pr(q_t \mid m_{\setminus t}, q_{\setminus t}) \times \Pr(K_t)$$

$\Pr(q_t \mid m_{\setminus t}, q_{\setminus t})$ is the same as in the lagged trade impact model (see section 7), since the results of this analysis do not involve the clustering or discreteness effects.

$\Pr(P_t \mid m_{\setminus t}, q_{\setminus t}, q_t, K_t)$ is computed as:

$$\Pr(P_t \mid m_{\setminus t}, q_{\setminus t}, q_t, K_t) = \int_{\underline{m}}^{\overline{m}} f(m_t \mid m_{\setminus t}, q_{\setminus t}, q_t)\, dm_t$$

where the limits of the integral are the clustering limits discussed in section 6. The integrand is the m_t density for the lagged trade impact model.

To summarize, compute the r.h.s. of eq. (33) for the four possible values of $\{q_t, K_t\}$. Normalize and make the draw. Finally, draw m_t from $f(m_t \mid m_{\setminus t}, q_{\setminus t}, q_t)$ restricted to the range $(\underline{m}, \overline{m})$ implied by the (just-drawn) $\{q_t, K_t\}$.

10. References

Carlin, B.P., and T.A. Louis, 1996. Bayes and Empirical Bayes Methods for Data Analysis. (London: Chapman and Hall).
Casella, G., George, E. I., 1992. Explaining the Gibbs sampler. The American Statistician 46, 167-190.
Chib, S., Greenberg, E., 1995. Understanding the Metropolis-Hastings algorithm. The American Statistician 49, 327-335.
Chib, S., Greenberg, E., 1996. Markov chain Monte Carlo simulation methods in econometrics. Econometric Theory 12, 409-431.
Gamerman, D., 1997. Markov Chain Monte Carlo. Chapman and Hall, New York.
Kim, C.-J., Nelson, C. R., 2000. State-space models with regime switching. MIT Press, Cambridge, Massachusetts.
Tanner, M.A., 1996. Tools for Statistical Inference. Third ed., (New York: Springer-Verlag).