CORC Report 2008-03

A behavioral finance based tick-by-tick model for price and volume*

Garud Iyengar †   Alfred Ka Chun Ma ‡

First version: July 7, 2008
This version: October 23, 2008

Abstract

We propose a model for jointly predicting stock price and volume at the tick-by-tick level. We model the investors' preferences by a random utility model that incorporates several important behavioral biases such as the status quo bias, the disposition effect, and loss-aversion. Our model is a logistic regression model with incomplete information; consequently, we are unable to use the maximum likelihood estimation method and have to resort to Markov Chain Monte Carlo (MCMC) to estimate the model parameters. Moreover, the constraint that the volume predicted by the MCMC model *exactly* match the observed volume v_t introduces serial correlation in the stock price. Thus, the standard MCMC methods for calibrating parameters do not work. We develop new modifications of the Metropolis-within-Gibbs method to estimate the parameters in our model. Our primary goal in developing this model is to predict the market impact function and VWAP (volume weighted average price) of individual stocks.

1 Introduction

In spite of the fact there is growing consensus that the traded volume and the price of a stock are correlated, these two quantities are traditionally modeled separately. Recently, there has been some attempts to link price and volume (see for example, Lee and Swaminathan, 2000; Hiemstra and Jones,

*Submitted to *Journal of Computational Finance*.
†Department of Industrial Engineering and Operations Research, Columbia University, New York, NY 10027. Email: garud@ieor.columbia.edu
‡Department of Industrial Engineering and Operations Research, Columbia University, New York, NY 10027. Email: km2207@columbia.edu

1994; Gallant et al., 1992; Hasbrouck, 1991). All of this work mainly focuses on empirically testing the correlation between the change in price and the volume traded. There is no explanation for the observed price-volume correlation (Karpoff, 1987). In this paper, we propose a behavioral finance based model that attempts to explain the correction between the price and volume observed in high frequency trading data.

The improvement in CPU speed and reduction in the price of memory has rapidly increased the ability to process large data set at very high speed. This development has led to an increased interest in using high frequency financial data to model and exploit local price dynamics. Hausman et al. (1992) build an ordered probit model which takes the discreteness of price change into account, Dufour and Engle (2000) use a vector autoregressive model to empirically test the price-volume relationship and McCulloch and Tsay (2001) propose using a nonlinear hierarchical model. In our model, we attempt to link the price and volume together using concepts from behavioral finance instead of relying purely on historical data.

Although behavioral finance theory based models have been studied in the context of asset pricing (Benartzi and Thaler, 1995; Barberis et al., 2001; Barberis and Huang, 2001; Coval and Shuway, 2005) and portfolio management (Berkelaar et al., 2004; Jin and Zhou, 2008), to the best of our knowledge, this is first time these concepts have been applied in the context of high frequency trading data. We focus on the disposition effect (Shefrin and Statman, 1985). Our model attempts to capture this effect at the tick by tick level. Disposition effect refers to the tendency of the investors to recognize profit more than loss. Investors exhibiting disposition effect evaluate their portfolios using mental accounts (Barberis and Huang, 2001). We hypothesize that these individual decisions have an impact on the aggregate price process at the tick-by-tick level. In order to model the aggregate effect, we keep track the reference price of every share of stock instead of using one reference price for all shares held by an investor (Grinblatt and Han, 2005). We use the random utility model (McFadden, 1973) to evaluate the probability that a particular share is traded as a function of the market price and the reference price of the share. This model can be used to empirically test the aggregate disposition effect in the market (Kaustia, 2004). Our model can also be used to jointly simulate sample paths of price and volume which can then be used to estimate execution costs (Berkowitz et al., 1988) and the price impact function.

The contribution of this paper is twofold. First, it provides a framework to incorporate concepts from behavioral finance and estimate the parameters characterizing the behavioral biases from tick-by-tick data. Testable hypotheses can then be formulated. Second, it provides practitioners who believe that behavioral biases exist in the market with a tool to exploit these effects in the context of portfolio management and algorithmic trading.

The remaining of the paper is organized as follows. Section 2 describes the model in details. Estimation techniques are given in Section 3 followed by numerical examples in Section 4.

2 Model

2.1 Notation

In this paper, we use upper case letter to denote random variables and lower case letters to denote realizations of random variables or deterministic quantities. We suppress subscripts to simplify the presentation. For example, suppose $y_{i,t}$ are defined for $i = 1, \ldots, M$ and $t = 0, \ldots, T$. Then we use $\mathbf{y}_t$ to represent the vector $(y_{1,t}, \ldots, y_{M,t})$, $\mathbf{y}_i = (y_{i,0}, \ldots, y_{i,T})$, and $\mathbf{y} = (\mathbf{y}_0, \ldots, \mathbf{y}_T)$.

2.2 Random utility model

We use the random utility model (McFadden, 1973) to model the behavior of each shareholder. We assume that for each share of stock, the shareholder decides whether or not to sell the share by maximizing the utility associated with the share. The holder of the stock has two actions available: $j = 0$ denotes the choice of holding the stock, and $j = 1$ denotes selling the stock. Let $U_{i,j,t}$ denote the utility when the holder of the i-th share chooses alternative j at time t, where $j = 0, 1$. Following McFadden (1973), we assume that

$$\begin{aligned} U_{i,0,t} &= \alpha + \beta_0 \left(\frac{r_i - p_t}{r_i} \right)^+ + \varepsilon_{i,0,t}, \\ U_{i,1,t} &= \beta_1 \left(\frac{p_t - r_i}{r_i} \right)^+ + \varepsilon_{i,1,t}, \end{aligned} \tag{1}$$

where $\varepsilon_{i,j,t}$ are independently and identically distributed according to the Gumbel distribution with CDF $F(x) = \exp(-\exp(-x))$, r_i is the *reference* price of the stock held by the i-th investor, and p_t is the market price at time t. In this model, the constant α represents the status quo bias (Samuelson and Zeckhauser, 1988) for holding the stock, and the constants β_0 and β_1 capture the disposition effect (Shefrin and Statman, 1985). The reference price r_i is defined below. It can be shown (McFadden, 1973) that a share with reference price r_i is sold at time t with probability

$$\mathbb{P}(U_{i,1,t} > U_{i,0,t}) = \frac{e^{\beta_1 \left(\frac{p_t - r_i}{r_i} \right)^+}}{e^{\beta_1 \left(\frac{p_t - r_i}{r_i} \right)^+} + e^{\alpha + \beta_0 \left(\frac{r_i - p_t}{r_i} \right)^+}} = \frac{1}{1 + e^{\alpha + \beta_0 \left(\frac{r_i - p_t}{r_i} \right)^+ - \beta_1 \left(\frac{p_t - r_i}{r_i} \right)^+}}. \tag{2}$$

From the description above it follows that two shares that have the same reference price have equal probability of being sold in the market. Consequently, we can collect all the shares of stock with the same reference price and call this collection a *bin*. It is clear that binning significantly reduces the dimension of the problem. In order to further reduce the problem dimension, we associate an interval of price with each bin and set the reference price of all the shares in the bin to the midpoint of the interval. Let M denote the number of bins, and for $i = 1, 2, \dots, M$, let B_i denote the price interval associate with bin i, R_i denote the reference price associated with bin i, and $S_{i,t}$ denote the number of shares in bin i at time t. We assume that the reference price of each share sold at time t is reset to R_i if $p_t \in B_i$. Thus, it follows that

$$S_{i,t+1} = \begin{cases} S_{i,t} - Y_{i,t} + v_t & \text{if } p_t \in B_i, \\ S_{i,t} - Y_{i,t} & \text{otherwise,} \end{cases} \tag{3}$$

where the random variable $Y_{i,t}$ denote the number of shares of stock in bin i traded at time t for $i = 1, 2, \dots M$. Let

$$\rho_i(p_t, \alpha, \beta_0, \beta_1) = \frac{1}{1 + e^{\alpha + \beta_0 \left(\frac{R_i - p_t}{R_i}\right)^+ - \beta_1 \left(\frac{p_t - R_i}{R_i}\right)^+}}, \tag{4}$$

denote the probability that a share in bin i is sold at price p_t. We suppress the parameters $(\alpha, \beta_0, \beta_1)$ when there is no ambiguity. Note that conditioned on $S_{i,t}$, each $Y_{i,t}$ is a binomial random variable with parameters $S_{i,t}$ and $\rho_i(p_t)$. Since the volume traded $v_t = \sum_i Y_{i,t}$, the model specifies that the distribution of v_t, conditional on the current price p_t, is the sum of binomial random variables with non-homogeneous parameters.

Note that the price interval $\bigcup_{i=1}^{M} B_i$ spanned by the M bins needs not be a partition of $\mathbb{R}^+$, all we require is that $p_t \in B_i$ for some bin i for all t. The number of bins M can, in principle, vary over time. However, in our numerical calculation, M is a sufficiently large number. Having a large number of bins is not an issue since we never need to consider bins with $S_{i,t} = 0$. We discuss the joint distribution of prices and volumes in Section 2.3.

Next, we give a simple example of our proposed dynamical model for price and volume. Suppose the stock is trading at 15.1, the collection $\{B_k = \{k/100\} | k = 1, \dots, 5000\}$ is sufficient to cover all possible future prices. Suppose we know that $N = 10000$ outstanding shares of the stock were offered at the IPO offer price of $p = 15.0$. Then $S_{1500,0} = 10000$. If $(p_0, v_0) = (15.1, 100)$, then we have $Y_{1500,0} = 100$ because $S_{i,0} = 0$ for $i \neq 1500$. At time $t = 1$, we will have

$$\begin{aligned} S_{1500,1} &= 9900, \\ S_{1510,1} &= 100, \end{aligned} \tag{5}$$

by (3). All $Y_{i,1}$ are all zero except for

$$\begin{aligned}
&Y_{1500,1} \sim \text{Binomial}\big(9900, \rho_{1500}(p_1)\big), \\
&Y_{1510,1} \sim \text{Binomial}\big(100, \rho_{1510}(p_1)\big),
\end{aligned} \tag{6}$$

and $v_1 = \sum_i Y_{i,1} = Y_{1500,1} + Y_{1510,1}$.

In our model we do not explicitly include the time effect. This is for two reasons. First, the time effect was not shown to be a major factor in the behavior finance literature. Investors are more concerned with losses than the time needed to recover from losses. Second, while there is a result (see for example, Dufour and Engle, 2000) showing that time may improve trade information, our model does not need the exogeneity assumption of time (Easley and O'Hara, 1992); the arrival time for the next trade can be introduced as an external component in our model.

2.3 Joint distribution of price and volume

Suppose we have all the information up to time t, i.e., we know $S_{i,t} = s_{i,t}$, for all $i = 1, \ldots, M$, as well as the parameters α, β_0, β_1. Since $v_t = \sum_{i=1}^M y_{i,t}$, we will focus on the joint distribution of $\mathbf{y}_t$ and p_t. The conditional density function $f(\mathbf{y}_t|p_t)$ is given by

$$f(\mathbf{y}_t|p_t) = \prod_{i=1}^M \binom{s_{i,t}}{y_{i,t}} \rho_i(p_t)^{y_{i,t}} (1 - \rho_i(p_t))^{s_{i,t} - y_{i,t}}. \tag{7}$$

From Bayes' theorem, it follows that the conditional density function $f(p_t|\mathbf{y}_t)$ satisfies

$$\begin{aligned}
f(p_t|\mathbf{y}_t) &\propto f(p_t) f(\mathbf{y}_t|p_t) \\
&= f(p_t) \prod_{i=1}^M \binom{s_{i,t}}{y_{i,t}} \rho_i(p_t)^{y_{i,t}} (1 - \rho_i(p_t))^{s_{i,t} - y_{i,t}},
\end{aligned} \tag{8}$$

where $f(p_t)$ is the prior density for the market price. Although, we allow any prior $f(p_t)$ for market prices, it it unlikely that we will have a closed form solution for $f(p_t|v_t)$ for any prior. We can, however, use the Metropolis-within-Gibbs algorithm (Tierney, 1994) to draw samples from the joint distribution.

Suppose the prior for the market price introduces serial dependence, i.e., $(p_1, \ldots, p_T)$ are *not* independent. Then

$$
\begin{aligned}
f(p_t,\ldots,p_T|\mathbf{y}_t,\ldots,\mathbf{y}_T) &\propto f(p_t,\ldots,p_T) f(\mathbf{y}_t,\ldots,\mathbf{y}_T|p_t,\ldots,p_T) \\
&= f(p_t,\ldots,p_T)\prod_{\tau=t}^{T}\prod_{i=1}^{M}\binom{s_{i,\tau}}{y_{i,\tau}}\rho_i(p_\tau)^{y_{i,\tau}}(1-\rho_i(p_\tau))^{s_{i,\tau}-y_{i,\tau}},
\end{aligned} \tag{9}
$$

where $\mathbf{s}_\tau$ are defined in (3) for $\tau = t+1,\ldots,T$.

2.4 Simulation strategies

We use the Metropolis-within-Gibbs approach (Tierney, 1994) to simulate the sample path for prices and volumes. First, we simulate sample paths of prices from the prior distribution. Although any prior distribution can be used; for concreteness, we take discreteness of prices into account (See for example Hausman et al., 1992; McCulloch and Tsay, 2001) and use prior distribution:

$$
f(p_{t+1}|p_t) = \begin{cases}
p_t - 3q & \text{with probability } p_{-3}, \\
p_t - 2q & \text{with probability } p_{-2}, \\
p_t - q & \text{with probability } p_{-1}, \\
p_t & \text{with probability } p_0, \\
p_t + q & \text{with probability } p_1, \\
p_t + 2q & \text{with probability } p_2, \\
p_t + 3q & \text{with probability } p_3,
\end{cases} \tag{10}
$$

where q is the pricetick. Another possibility is that $f(p_{t+1}|p_t)$ has lognormal prior distribution. The parameters in the prior distribution can either be estimated using historical data or be specified *a priori*.

Given a sample path $(\hat{p}_t,\ldots,\hat{p}_T)$ for the price, we simulate a new sample path $(p_t,\ldots,p_T)$ using the prior distribution, and accept it with probability

$$
\kappa(\hat{p}_t,\ldots,\hat{p}_T,p_t,\ldots,p_T) = \min\left(1, \frac{\prod_{\tau=t}^{T}\prod_{i=1}^{m}\binom{s_{i,\tau}}{y_{i,\tau}}\rho_i(p_\tau)^{y_{i,\tau}}(1-\rho_i(p_\tau))^{s_{i,\tau}-y_{i,\tau}}}{\prod_{\tau=t}^{T}\prod_{i=1}^{m}\binom{s_{i,\tau}}{y_{i,t}}\rho_i(\hat{p}_\tau)^{y_{i,\tau}}(1-\rho_i(\hat{p}_\tau))^{s_{i,\tau}-y_{i,\tau}}}\right). \tag{11}
$$

Given the prices $(p_t,\ldots,p_T)$, we simulate volumes $(v_t,\ldots,v_T)$ exactly, i.e., the acceptance probability for the volume is always 1. It is easy to see that the resulting Markov Chain is irreducible and aperiodic.

While there is no closed form formula for the joint distribution of prices and volumes, the prices $(p_t,\ldots,p_T)$ and volumes $(v_t,\ldots,v_T)$ generated by the Metropolis-within-Gibbs algorithm will have the following properties.

Property 1. *At any given iteration with $(\hat{p}_t,\hat{\mathbf{y}}_t)$, $\mathbb{E}(v_t|\hat{p}_t)$ increases with the parameter β_1 and decreases* *with the parameters α and β_0.*

Property 2. *At any given iteration with $(\hat{p}_t, \hat{\mathbf{y}}_t)$, if $\hat{y}_{i,t} > \mathbb{E}(Y_{i,t})$ ($\hat{y}_{i,t} < \mathbb{E}(Y_{i,t})$) for all i, there exists R_{min} and R_{max} with $R_{min} \leq R_{max}$ such that $\mathbb{P}(p_t \geq R_{max}|\hat{\mathbf{y}}_t)$ increases (decreases) with the parameter β_0 and $\mathbb{P}(p_t \leq R_{min}|\hat{\mathbf{y}}_t)$ increases (decreases) with the parameter β_1 within a small neighborhood of (β_0, β_1).*

The proofs are given in Appendix A. From Property 1 and 2, it follows that different values of the parameters $(\alpha, \beta_0, \beta_1)$ can result in vastly different market behavior. The estimation procedure will be discussed in Section 3.

2.5 Application

Since we are able to simulate the joint sample paths of prices and volumes, an immediate application of our model is to predict measures which depends on both prices and volumes. For example, we are able to estimate the volume weighted average price (VWAP) which is a popular benchmark used in estimation of cost of execution of orders (Berkowitz et al., 1988). The ability to estimate VWAP is extremely important for algorithmic trading.

Another application of our model is to estimate the market impact of a trade. Suppose a trader wants to purchase V shares of stock at the next trade and is interested in estimating the price change induced by this trade. We can answer this question by computing $\mathbb{E}_{t-1}(p_t|v_t = V)$, i.e., the expected price at time t conditional on $v_t = V$ and all information up to time $t-1$. In Section 4.3 of the paper, we show how to estimate this using the Metropolis-Hasting algorithm. Property 2 in the previous section implies that the parameters (β_0, β_1) will have an impact on $\mathbb{E}_{t-1}(p_t|v_t = V)$: one is more likely to see high prices for larger values of β_0 and low prices for larger values of β_1.

Note that our model computes the price impact of a *buy* order under the assumption that the stock holders have behavioral biases. We need additional assumptions on the behavior of the buyers in order to compute the price impact of a *sell* order.

3 Estimation procedure

3.1 Metropolis-within-Gibbs

We observe the time series of price and volume $(p_0, v_0), \dots (p_T, v_T)$. Since the samples $y_{i,t}$ of the number of shares sold from bin i are not observable, the standard maximum likelihood estimation cannot be used. We employ a Bayesian approach (Choi et al., 2007). The main idea here is to treat all missing information $y_{i,t}$ as parameters and assume they all have a flat prior distribution. Then we have

$$\begin{aligned} &f(\alpha, \beta_0, \beta_1, \mathbf{s}_0, \mathbf{y} | p_0, \ldots, p_T, v_0, \ldots, v_T) \\ &\propto f(\alpha) f_0(\beta_0) f_1(\beta_1) f(\mathbf{s}_0) \prod_{t=0}^{T} \mathbf{1}_{\{\sum_{i=1}^M y_{i,t} = v_t\}} \prod_{i=1}^{M} \mathbf{1}_{\{s_{i,t} \geq 0\}} \binom{s_{i,t}}{y_{i,t}} \rho_i(p_t)^{y_{i,t}} (1 - \rho_i(p_t))^{s_{i,t} - y_{i,t}}. \end{aligned} \tag{12}$$

The unobservable variables $\mathbf{y}_t$, $t = 0, \ldots, T$, and $\mathbf{s}_0$ are highly dependent. Consequently, it is more efficient to simulate them together as a block. Given the parameters $(\alpha, \beta_0, \beta_1)$, we first simulate $\mathbf{s}_0$ according to a prior distribution. We use the independent Metropolis proposal kernel

$$q(\hat{\mathbf{y}}_t, \mathbf{y}_t) = \frac{v_t!}{y_{1,t}! \ldots y_{M,t}!} \prod_{i=1}^{M} Q_{i,t}^{y_{i,t}}, \tag{13}$$

for the time t trade vector $\mathbf{y}_t$, where

$$Q_{i,t} = \frac{\rho_i(p_t) s_{i,t}}{\sum_{i=1}^M \rho_i(p_t) s_{i,t}}, \tag{14}$$

and $\hat{\mathbf{y}}_t$ is the current trade vector. The rationale for choosing this kernel is as follows. We expect $\rho_i(p_t) \ll 1$ and $s_{i,t} \gg 1$, thus we can approximate $\mathbf{y}_t$ by a Poisson distribution with mean $s_{i,0}\rho_i(p_t)$. Suppose each $Y_{i,t} \sim \text{Poisson}(s_{i,0}\rho_i(p_t))$. Then the distribution of $Y_{i,t}$ given $\sum_i Y_{i,0} = v_t$ is given by (13). After simulating $\mathbf{y}_t$, we compute $\mathbf{s}_{t+1}$ using (3). Next, we simulate $\mathbf{y}_t$ recursively for all $t = 1, 2, \ldots, T$, and accept $(\mathbf{s}_0, \mathbf{y})$ with probability

$$\min\left(1, \frac{f(\mathbf{y}|\cdot)}{f(\hat{\mathbf{y}}|\cdot)} \prod_{t=0}^{t} \frac{q(\mathbf{y}_t, \hat{\mathbf{y}}_t)}{q(\hat{\mathbf{y}}_t, \mathbf{y}_t)}\right), \tag{15}$$

where

$$f(\mathbf{y}|\cdot) = \prod_{t=0}^{T} \mathbf{1}_{\{\sum_{i=1}^M y_{i,t} = v_t\}} \prod_{i=1}^{M} \mathbf{1}_{\{s_{i,t} \geq 0\}} \binom{s_{i,t}}{y_{i,t}} \rho_i(p_t)^{y_{i,t}} (1 - \rho_i(p_t))^{s_{i,t} - y_{i,t}}. \tag{16}$$

For parameters $(\alpha, \beta_0, \beta_1)$, we use random walk approach, i.e., $\alpha \sim N(\hat{\alpha}, \sigma_\alpha^2)$ and $\beta_i \sim n(\hat{\beta}_i, \sigma_{\beta_i}^2)$, and accept with probability

$$\kappa(\hat{\alpha}, \alpha) = \min\left(1, \frac{f(\alpha) f(\alpha|\cdot)}{f(\hat{\alpha}) f(\hat{\alpha}|\cdot)}\right), \quad \kappa(\hat{\beta}_i, \beta_i) = \min\left(1, \frac{f_i(\beta) f(\beta_i|\cdot)}{f_i(\hat{\beta}_i) f(\hat{\beta}_i|\cdot)}\right), \tag{17}$$

where

$$
\begin{aligned}
f(\alpha|\cdot) &= f(\alpha)\prod_{t=0}^{T}\prod_{i=1}^{M}\binom{s_{i,t}}{y_{i,t}}\rho_i(p_t,\alpha,\beta_0,\beta_1)^{y_{i,t}}\Big(1-\rho_i(p_t,\alpha,\beta_0,\beta_1)\Big)^{s_{i,t}-y_{i,t}},\\
f(\beta_i|\cdot) &= f_i(\beta_i)\prod_{t=0}^{T}\prod_{i=1}^{M}\binom{s_{i,t}}{y_{i,t}}\rho_i(p_t,\alpha,\beta_0,\beta_1)^{y_{i,t}}\Big(1-\rho_i(p_t,\alpha,\beta_0,\beta_1)\Big)^{s_{i,t}-y_{i,t}},
\end{aligned}
\tag{18}
$$

The variance σ^2_α and $\sigma^2_{\beta_i}$ are set using the EM algorithm described in Section 3.2. The proof that the Markov Chain is irreducible and aperiodic is given in Appendix B.

3.2 EM algorithm

The Expectation Maximization (EM) algorithm (Dempster et al., 1977) or Expectation Conditional Maximization (ECM) algorithm (Meng and Rubin, 1993) are popular techniques for the estimation of the parameters when there is missing information. However, the EM algorithm is not directly applicable in our model because the E-step is intractable. We, therefore, use a modified version of EM algorithm. Although the general MCMC method described in Section 3.1 does converge, the EM approach has several advantages. The EM algorithm is helpful in generating good initial values for the MCMC method, for detecting if the target distribution is multimodal (Gelman and Rubin, 1992) and also for setting the variances σ^2_α, $\sigma^2_{\beta_0}$ and $\sigma^2_{\beta_1}$.

The E-step in the EM algorithm involves computing the conditional expectation

$$
\mathbb{E}(Y_{i,t}|\sum_{i=1}^{M}Y_{i,\tau}=v_\tau,\tau=0,\ldots,T,\alpha,\beta_0,\beta_1,\mathbf{s}_0)
\tag{19}
$$

of the missing data.

The time dependence in $\mathbf{Y}$ makes this computation intractable. We propose an approximate approach to circumvent these computation issues. We assume that the posterior distribution of Y is so sharply peaked at its posterior mode that the posterior mean coincides with the posterior mode. Therefore, we can treat $\mathbf{y}$ as a parameter instead of missing information. Thus, our estimation problem reduces to

$$
\max_{\mathbf{s}_0,\mathbf{y},\alpha,\beta_0,\beta_1}\log f(\mathbf{p}|\mathbf{y},\mathbf{s}_0,\alpha,\beta_0,\beta_1),
\tag{20}
$$

or equivalently

$$
\max_{\mathbf{s}_0,\mathbf{y},\alpha,\beta_0,\beta_1}\log\Big(f(\mathbf{y}|\mathbf{s}_0,\alpha,\beta_0,\beta_1,\mathbf{p})f(\mathbf{p})\Big).
\tag{21}
$$

To proceed further, we use Stirling's formula to approximate the likelihood function as follows.

$$\log f(\mathbf{z}|\mathbf{s}_0, \alpha, \beta_0, \beta_1, \mathbf{p}) \approx \sum_{t=0}^{T}\sum_{i=1}^{M} s_{i,t}\left(z_{i,t}\log\left(\frac{\rho_i(p_t)}{z_{i,t}}\right) + (1-z_{i,t})\log\left(\frac{1-\rho_i(p_t)}{1-z_{i,t}}\right)\right), \tag{22}$$

where $z_{i,t} = \frac{y_{i,t}}{s_{i,t}}$, $i = 1, \ldots, M$, $t = 0, \ldots, T$, denote the proportion of shares in bin i traded at time t.

We further relax $\mathbf{z}$ to be continuous over $[0, 1]$. Now, our problem reduces to

$$\begin{aligned}
\max \quad & \textstyle\sum_{t=0}^{T}\sum_{i=1}^{M} s_{i,t}\left(z_{i,t}\log\left(\frac{\rho_i(p_t)}{z_{i,t}}\right) + (1-z_{i,t})\log\left(\frac{1-\rho_i(p_t)}{1-z_{i,t}}\right)\right) \\
\text{s.t.} \quad & s_{i,t+1} = s_{i,t}(1-z_{i,t}) + J_{i,t}, \\
& \textstyle\sum_{i=1}^{M} s_{i,t}z_{i,t} = v_t, \\
& \textstyle\sum_{i=1}^{M} s_{i,0} = N, \\
& s_{i,t} \geq 0, \\
& 0 \leq z_{i,t} \leq 1,
\end{aligned} \tag{23}$$

where

$$J_{i,t} = \begin{cases} v_t & \text{if } p_t \in B_i, \\ 0 & \text{otherwise.} \end{cases} \tag{24}$$

We will solve (23) by coordinate descent, i.e., the CM (Meng and Rubin, 1993) method, with $\mathbf{z}_t$, $(\alpha, \beta_0, \beta 1)$, and $\mathbf{s}_0$ as separate blocks. Since in all practical scenarios, the components of $\mathbf{z}_t$ are likely to be small, (23) can be approximated by the following series of optimization problems,

$$\begin{aligned}
\mathcal{P}_t \equiv \max \quad & \textstyle\sum_{i=1}^{M} s_{i,t}\left(z_{i,t}\log(\frac{\rho_i(p_t)}{z_{i,t}}) + (1-z_{i,t})\log(\frac{1-\rho_i(p_t)}{1-z_{i,t}})\right) \\
\text{s.t.} \quad & \textstyle\sum_{i=1}^{M} s_{i,t}z_{i,t} = v_t, \\
& 0 \leq z_{i,t} \leq 1,
\end{aligned} \tag{25}$$

for $t = 0, 1, \ldots, T$. In Appendix C we show that (25) problem can be solved very efficiently. After solving for the optimal solution $\mathbf{z}_t^*$, one can compute $\mathbf{s}_{t+1}$ using (3) and solve (25) for z_{t+1}^*.

The CM-step for parameters $(\alpha, \beta_0, \beta_1)$ involves solving the maximization problem.

$$\max \quad \textstyle\sum_{t=0}^{T}\sum_{i=1}^{M} s_{i,t}\left(z_{i,t}^*\log(\frac{\rho_i(p_t,\alpha,\beta_0,\beta_1)}{z_{i,t}^*}) + (1-z_{i,t}^*)\log(\frac{1-\rho_i(p_t,\alpha,\beta_0,\beta_1)}{1-z_{i,t}^*})\right). \tag{26}$$

Note that the (26) is convex in α, β_0, and β_1 (see Appendix C for details), and can, therefore, be solved very efficiently by numerical methods such as the Damped Newton Method.

Finally, we solve the maximization problem in $\mathbf{s}_0$:

$$\begin{aligned} \max \quad & \sum_{t=0}^{T} \sum_{i=1}^{M} s_{i,t} \left(z_{i,t}^* \log\left(\frac{\rho_i(p_t)}{z_{i,t}^*}\right) + (1 - z_{i,t}^*) \log\left(\frac{1-\rho_i(p_t)}{1-z_{i,t}^*}\right)\right) \\ \text{s.t.} \quad & s_{i,t+1} = s_{i,t}(1 - z_{i,t}^*) + J_{i,t}, \\ & \sum_{i=1}^{M} s_{i,t} z_{i,t}^* = v_t, \\ & \sum_{i=1}^{M} s_{i,0} = N, \\ & s_{i,t} \geq 0. \end{aligned} \tag{27}$$

Note that the maximization problem (27) is linear, and, therefore, can be solved quickly. We iteratively solve (25), (26), and (27) to compute an estimate for the parameter $(\alpha, \beta_0, \beta_1)$ and $\mathbf{s}_0$.

The ECM algorithm is a good complementary pilot for the MCMC method. The estimates of the parameters from the ECM method can be used as the initial values of the random walk Metropolis algorithm and the estimated standard error can be used as the standard deviations of the random walk.

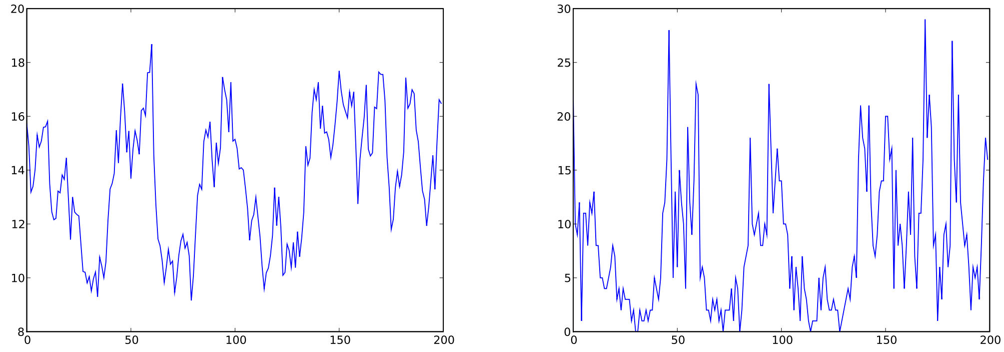
4 Numerical results

In this section, we report the result of the two sets of numerical experiments.

4.1 Simulation experiment

In this section we discuss an example that illustrates the details of our proposed method. For this example, we set the prior $f(p_{t+1}|p_t) = \log\mathcal{N}(0.0005, 0.075)$, the parameters $\alpha = 9.0379$, $\beta_0 = 6$, $\beta_1 = 3$, and $N = 166180$. Also, we mimic a IPO situation where all the shares have the reference price at time $t = 0$, and this reference price was set equal 15. Since the prior for the market price is a continuous, the collection of all possible bins can be an uncountable set. However, we only need to keep track of the bins with non-zero value of $s_{i,t}$ for the estimation described in Section 3 to work. Without loss of generality, we can reorder the bins and set $s_{1,0} = N$, and $B_1 = \{15\}$. Moreover, we set $|B_i| = 0$, i.e., we keep track of actual trade prices. Next, we simulated a time series of price and volume and this time series is displayed in Figure 1. We use the results from the ECM method to run the MCMC (4,000 samples, with 2,000 burning-in and thinning equal to 4) with flat prior distributions for $(\alpha, \beta_0, \beta_1)$. The results of the estimation are displayed in Table 1.

Although, we make a number of simplifying approximation in developing the ECM algorithm, the algorithm is still able to compute reasonably good estimates for the parameters $(\alpha, \beta_0, \beta_1)$. Also, since ECM is a deterministic algorithm it is significantly more efficient when compared to the MCMC algorithm. However, the MCMC result is likely to be more reliable.



(a) simulated prices. (b) simulated volumes.

Figure 1: simulated prices and volumes.

	α	β_0	β_1
True value	9.0379	6	3
ECM estimate	9.0734	5.5368	3.2696
95% confidence interval	$(9.0349, 9.1317)$	$(5.3709, 6.0073)$	$(1.1251, 4.1770)$
99% confidence interval	$(9.0197, 9.1469)$	$(5.2709, 6.1073)$	$(0.6455, 4.6566)$
MCMC estimate	9.0779	5.7581	2.4878
95% confidence interval	$(8.9909, 9.1645)$	$(5.3721, 6.1790)$	$(-1.7695, 6.4869)$
99% confidence interval	$(8.9724, 9.1852)$	$(5.3008, 6.3302)$	$(-2.1761, 6.9163)$

Table 1: Estimation results, assuming $T = 200$, $N = 166, 180$, and there is one bin at time 0 with $B_1 = \{15\}$.

Next we considered the situation where the initial bin population $\mathbf{S}_0$ is not deterministic. We used the same simulated data as in the previous experiment. We assumed that we begin our estimation at time $t = 50$. We still observe $(p_0, v_0), \ldots, (p_{49}, v_{49})$, but $\mathbf{S}_{50}$ is no longer known. We use a multinomial distribution with p_i proportional to the total volume turnover traded at R_i for $t < 50$ as a prior distribution for $\mathbf{S}_{50}$. Table 2 displays the estimation results. Note that the ECM confidence interval for β_0 does not cover the true value while the confidence interval generated by the MCMC method does cover the true parameters. Thus, the statistical properties of the MCMC method are superior to that of the ECM algorithm.

	α	β_0	β_1
True value	9.0379	6	3
ECM estimate	9.1217	5.5170	3.2490
95% confidence interval	$(9.0675, 9.1759)$	$(5.1389, 5.8952)$	$(1.8612, 4.6367)$
99% confidence interval	$(9.0504, 9.1929)$	$(5.0200, 6.0140)$	$(1.4251, 5.0729)$
MCMC estimate	9.1462	5.4079	3.5477
95% confidence interval	$(9.0468, 9.2209)$	$(4.7717, 6.0164)$	$(0.6593, 6.6003)$
99% confidence interval	$(9.0224, 9.2375)$	$(4.6297, 6.0985)$	$(0.2015, 7.2922)$

Table 2: Estimation results, assuming $T = 150$, $N = 166, 180$, with $\mathbf{s}_0$ unknown.

4.2 Empirical results

The Trade and Quote (TAQ) database contains all the trade data starting from 1993. For better estimation, we consider cases where the uncertainty of $\mathbf{S}_0$ can be removed, i.e., the IPO scenario. Since the TAQ database does not contain the IPO offer price, we use the SDC database for this information. The TAQ database contains more than 500 stocks with IPO date after 1993. Table 3 in Section 2 shows the information of the stock and the description of the data we chose for our numerical study. Since stocks are typically traded in lots, we define N to be the number of lots outstanding. For CKI, $N = 304000$. Table 4 shows the estimation results for the parameters using flat prior distributions for $(\alpha, \beta_0, \beta_1)$. The results show that the parameters corresponding to the disposition effect, i.e., β_0 and β_1, are both positive. And the holding-the-loser effect is observed to be stronger that the selling-the-winner effect, which is consistent with the prospect theory (Kahneman and Tversky, 1979).

Stock symbol	CKI
IPO offer date	2003-03-04
IPO offer price	15.0
Share outstanding	30,400,000
Lot size	100
Number of trades (T)	1678
Total turnover	$4,569,500$

Table 3: Information of the stock

	α	β_0	β_1
ECM estimate	9.2547	5.2475	2.6113
95% confidence interval	$(9.2456, 9.2639)$	$(4.9858, 5.5092)$	$(2.2842, 2.9383)$
99% confidence interval	$(9.2427, 9.2668)$	$(4.9035, 5.5914)$	$(2.1814, 3.0411)$
MCMC estimate	9.2498	5.3426	2.4766
95% confidence interval	$(9.2376, 9.2603)$	$(5.0898, 5.5974)$	$(2.0627, 2.8020)$
99% confidence interval	$(9.2348, 9.2634)$	$(5.0113, 5.7976)$	$(1.9426, 2.9009)$

Table 4: Estimation results for CKI.

Note that it is important to see if the ECM algorithm is sensitive to the initial values. The result can be sensitive to initial values if the posterior distribution is multimodal. Table 5 shows that the ECM algorithm is very robust and not sensitive to the initial values for the parameters $(\alpha, \beta_0, \beta_1)$. This also implies that there is no evidence of multi-modality of the posterior distribution.

4.3 Price impact

In this section we show how to use our method to estimate the price impact. When the future v_t is not fixed, we can simulate $\mathbf{y}_t$ and set v_t can be defined as $\sum_i y_{i,t}$. However, exactly simulating $\mathbf{y}_t$ conditional on $\sum_i y_{i,t} = V$ is not feasible. Consequently, another Metropolis step is required.

Initial values			ECM estimates			Standard error		
α	β_0	β_1	α	β_0	β_1	α	β_0	β_1
−1.8659	7.9288	−2.5704	9.2548	5.2475	2.6114	0.0047	0.1335	0.1669
−17.7336	4.4388	7.5808	9.2550	5.2449	2.6158	0.0047	0.1335	0.1669
−2.429	2.4725	−23.5016	9.2548	5.2473	2.6117	0.0047	0.1335	0.1669
−6.4644	−0.6125	11.8318	9.2550	5.2446	2.6162	0.0047	0.1335	0.1669
−2.1779	7.9513	−3.8555	9.2548	5.2473	2.6117	0.0047	0.1335	0.1669
−5.7451	8.4821	2.3337	9.2548	5.2475	2.6114	0.0047	0.1335	0.1669
6.9288	2.7617	−7.6676	9.2548	5.2477	2.6111	0.0047	0.1335	0.1669
22.0485	−13.1423	4.0943	9.2550	5.2448	2.6159	0.0047	0.1335	0.1669
−19.9901	−5.6886	−4.019	9.2550	5.2447	2.6161	0.0047	0.1335	0.1669
−8.1609	1.8267	4.1252	9.2550	5.2448	2.6160	0.0047	0.1335	0.1669
−9.2098	10.6505	−1.4303	9.2548	5.2473	2.6117	0.0047	0.1335	0.1669
−6.5353	11.4522	20.4315	9.2550	5.2447	2.6160	0.0047	0.1335	0.1669
0.6879	−17.6739	−7.489	9.2550	5.2449	2.6158	0.0047	0.1335	0.1669
0.9023	13.5104	9.7253	9.2550	5.2448	2.6159	0.0047	0.1335	0.1669
−9.4558	−2.1418	17.2409	9.2550	5.2448	2.6159	0.0047	0.1335	0.1669
−11.1299	3.2033	−19.3766	9.2548	5.2476	2.6112	0.0047	0.1335	0.1669
−2.0949	1.813	−8.698	9.2548	5.2475	2.6113	0.0047	0.1335	0.1669
13.0248	5.4837	9.3879	9.2550	5.2446	2.6162	0.0047	0.1335	0.1669
−4.6583	10.854	6.0414	9.2548	5.2476	2.6113	0.0047	0.1335	0.1669
5.2877	−3.9282	21.4714	9.2550	5.2447	2.6160	0.0047	0.1335	0.1669

Table 5: ECM Estimation results for CKI with various initial values.

In order to study the price impact as a function of the parameters $(\alpha, \beta_0, \beta_1)$, we use an IPO scenario, i.e., $\mathbf{S}_0$ has a single non zero component as in Section 4.1. To further reduce the uncertainty and make the result clearer to understand, we use the prior distribution in (10) with pricetick size $q = 0.125$ and

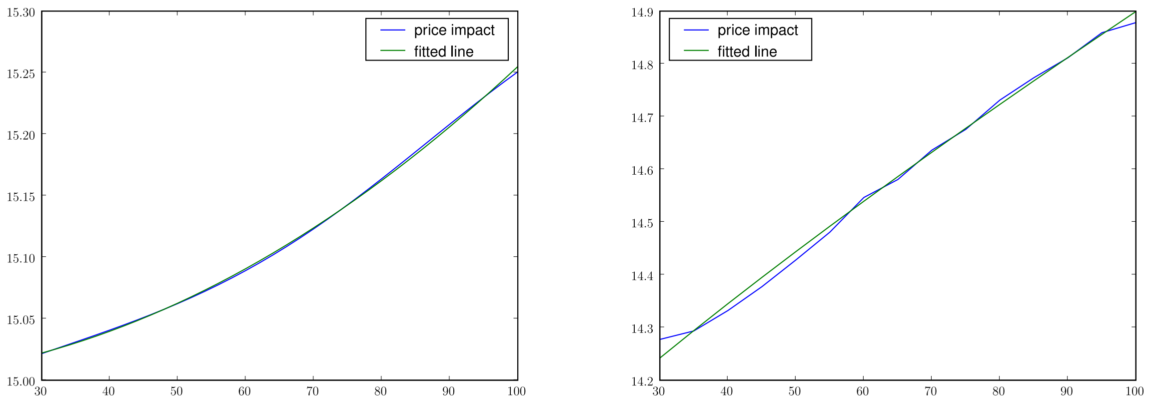
$$(p_{-3}, p_{-2}, p_{-1}, p_0, p_1, p_2, p_3,) = (0.0066, 0.0133, 0.1453, 0.6706, 0.1453, 0.0127, 0.0063). \quad (28)$$

The numbers are taken from McCulloch and Tsay (2001). With this simple prior density function, we can compute the density function at each price. Table 6 shows the result for various changes of parameters with the base values $(\alpha, \beta_0, \beta_1) = (9.0379, 6, 3)$ and $V = 100$. It is to note that the higher the numerical value of the disposition effect parameters (β_0, β_1) and the status quo bias parameter (α), the greater is the price impact. The result is also consistent with Property 2 in Section 2.4.

Figure 2(a) shows the price impact function $\mathbb{E}_{t-1}(p_t|v_t = V)$ for the base values. We also fit a line of the form $p = a + bV^c$ where $a = 15$, $b = 2.3467e^{-5}$ and $c = 2.0183$, with residual sum of squares equal to $7.30201e^{-05}$. Figure 2(b) shows the same for the price impact function at time $T = 1678$ with prior $f(p_{t+1}|p_t) \sim \log\mathcal{N}(0.0005, 0.075)$. The constants for the fitted line are $a = 13.862$, $b = 0.0222$ and $c = 0.8339$ with residual sum of squares equal to 0.0288975.

		$\mathbb{E}_{t-1}(p_t \mid v_t = V)$	f_{-3}	f_{-2}	f_{-1}	f_0	f_1	f_2	f_3
	9.0	15.25080	$1.2042e^{-11}$	$1.3049e^{-8}$	$6.8616e^{-5}$	0.1374	0.2201	0.1404	0.5018
β_0	6.0	15.25063	$6.3747e^{-9}$	$8.0999e^{-7}$	0.0005255	0.1374	0.2200	0.1403	0.5016
	3.0	15.24931	$3.0574e^{-6}$	$4.7935e^{-5}$	0.003965	0.1369	0.2192	0.1398	0.4998
	4.5	15.34580	$5.6902e^{-10}$	$7.2303e^{-8}$	$4.6909e^{-5}$	0.0122	0.0531	0.0902	0.8442
β_1	3.0	15.25063	$6.3747e^{-9}$	$8.0999e^{-7}$	0.0005255	0.1374	0.2200	0.1403	0.5016
	1.5	15.09299	$2.4138e^{-8}$	$3.0671e^{-6}$	0.001989	0.5203	0.3070	0.0728	0.0977
	9.05	15.25162	$6.0917e^{-9}$	$7.8220e^{-7}$	$5.1310e^{-4}$	0.1357	0.2186	0.1403	0.5047
α	9.04	15.25063	$6.3747e^{-9}$	$8.0999e^{-7}$	0.0005255	0.1374	0.2200	0.1403	0.5016
	9.02	15.24914	$6.8238e^{-9}$	$8.5353e^{-7}$	$5.4467e^{-3}$	0.1399	0.2221	0.1403	0.4969
	200	15.37142	$2.0955e^{-18}$	$3.9517e^{-14}$	$3.8050e^{-9}$	0.0015	0.0028	0.0223	0.9745
V	100	15.25063	$6.3747e^{-9}$	$8.0999e^{-7}$	0.0005255	0.1374	0.2200	0.1403	0.5016
	50	15.06230	$5.0039e^{-5}$	0.0005219	0.02779	0.5966	0.2737	0.0500	0.0512

Table 6: Price impact for various initial values, where the base values are $(\alpha, \beta_0, \beta_1) = (9.0379, 6, 3)$ and $V = 100$. Here we use the notation $f_k = f(15 + k \times q)$.



(a) Example with fitted line: $p = 15 + 2.3467e^{-5}V^{2.0183}$.

(b) CKI with fitted line: $p = 13.862 + 0.0222V^{0.8339}$.

Figure 2: Price impact function and the fitted lines

5 Conclusion

We propose a tick-by-tick model that jointly predicts price and volume. The model attempts to incorporate behavioral biases and applies robust estimation methods to estimate parameters. Future work includes a thorough analysis of the empirical data and extending the model to solve problems in portfolio management and algorithm trading.

Appendix A Proof of the properties of prices and volumes

Property 1 follows from the fact that

$$\mathbb{E}(v_t|\hat{p}_t) = \mathbb{E}(\sum_{i=1}^{M} y_{i,t}|\hat{p}_t) = \sum_{i=1}^{M} s_{i,t}\rho(\hat{p}_t, \alpha, \beta_0, \beta_1), \tag{29}$$

is an increasing function of β_1 and a decreasing function of α and β_0.

To prove Property 2, we take the first derivative of $f(v_t|p_t)$ with respect to ρ_i,

$$\frac{\partial f(p_t|\hat{\mathbf{y}}_t)}{\partial \rho_i} = f(p_t)\binom{\hat{s}_{i,0}}{\hat{y}_{i,t}}\rho_i^{\hat{y}_{i,t}-1}(1-\rho_i)^{\hat{s}_{i,t}-\hat{y}_{i,t}-1}(\hat{y}_{i,t}-\hat{s}_{i,t}\rho_i). \tag{30}$$

Thus $\frac{\partial f(p_t|\hat{\mathbf{y}}_t)}{\partial \rho_i} > 0$ if and only if $\hat{y}_{i,t} > E(Y_{i,t})$. We only prove Property 2 for β_0. The case for β_1 can be established in similar manner.

Let $\hat{\beta}_0 > \beta_0$, and take $R_{\max}$ be the maximum value over all the reference prices of all bins. By the definition of ρ_i in (4), it follows that $\rho_i(p_t, \alpha, \beta_0, \beta_1) = \rho_i(p_t, \alpha, \hat{\beta}_0, \beta_1)$ for $p_t > R_{\max}$. Therefore, we have

$$\begin{aligned} A &\triangleq \int_{R_{\max}}^{\infty} f(p_t)\prod_{i=1}^{M}\binom{\hat{s}_{i,t}}{\hat{y}_{i,t}}\rho_i(p_t,\alpha,\beta_0,\beta_1)^{\hat{y}_{i,t}}(1-\rho_i(p_t,\alpha,\beta_0,\beta_1))^{\hat{s}_{i,t}-\hat{y}_{i,t}}dp_t \\ &= \int_{R_{\max}}^{\infty} f(p_t)\prod_{i=1}^{M}\binom{\hat{s}_{i,t}}{\hat{y}_{i,t}}\rho_i(p_t,\alpha,\hat{\beta}_0,\beta_1)^{\hat{y}_{i,t}}(1-\rho_i(p_t,\alpha,\hat{\beta}_0,\beta_1))^{\hat{s}_{i,t}-\hat{y}_{i,t}}dp_t. \end{aligned} \tag{31}$$

On the other hand, $\rho_i(p_t, \alpha, \beta_0, \beta_1) > \rho_i(p_t, \alpha, \hat{\beta}_0, \beta_1)$ for $p_t < R_{\max}$ so we have

$$\begin{aligned} B &\triangleq \int_{-\infty}^{R_{\max}} f(p_t)\prod_{i=1}^{M}\binom{\hat{s}_{i,t}}{\hat{y}_{i,t}}\rho_i(p_t,\alpha,\beta_0,\beta_1)^{\hat{y}_{i,t}}(1-\rho_i(p_t,\alpha,\beta_0,\beta_1))^{\hat{s}_{i,t}-\hat{y}_{i,t}}dp_t \\ &> \int_{-\infty}^{R_{\max}} f(p_t)\prod_{i=1}^{M}\binom{\hat{s}_{i,t}}{\hat{y}_{i,t}}\rho_i(p_t,\alpha,\hat{\beta}_0,\beta_1)^{\hat{y}_{i,t}}(1-\rho_i(p_t,\alpha,\hat{\beta}_0,\beta_1))^{\hat{s}_{i,t}-\hat{y}_{i,t}}dp_t \triangleq C, \end{aligned} \tag{32}$$

for $\hat{\beta}_0$ close enough to β_0. Then the change of $P(p_t \geq R_{\max}|\hat{\mathbf{y}}_t)$ when β_0 increases to $\hat{\beta}_0$ is

$$\frac{A}{A+C} - \frac{A}{A+B} = \frac{A(B-C)}{(A+B)(A+C)} > 0. \tag{33}$$

Appendix B Convergence of the MCMC algorithm

Our MCMC method splits the space of parameters into four components, namely α, β_0, β_1, and $(s_{1,0}, \ldots, s_{M,0}, y_{1,0}, \ldots, y_{M,0}, \ldots, y_{1,T}, \ldots, y_{M,T})$. It suffices to show that each Markov kernel is irreducible and aperiodic. For α, β_0, and β_1, as we use random walk chains with normal distribution for the increment, the Markov kernel for these three components are irreducible and aperiodic (Tierney, 1994).

For the last component, if $f(\alpha, \beta_0, \beta_1, s_{1,0}, \ldots, s_{M,0}, y_{1,0}, \ldots, y_{M,0}, \ldots, y_{1,T}, \ldots, y_{M,T}) > 0$, we have

$$f(s_{1,0}, \ldots, s_{M,0})\prod_{t=0}^{T}\frac{v_t!}{y_{1,t}!\ldots y_{M,t}!}\prod_{i=1}^{M}Q_{i,t}^{y_{i,t}} = 0 \tag{34}$$

only if $Q_{j,\tau} = 0$ for some (j, τ) with $y_{j,\tau} > 0$. However, $Q_{j,\tau} = 0$ if and only if $s_{j,\tau} = 0$ and this, in turn implies that $y_{j,\tau} = 0$. Thus, we have established that this independence Metropolis kernel is irreducible and aperiodic.

Appendix C CM-step of the ECM algorithm

First we show that the continuously approximate loglikelihood function (22) is concave. It suffices to show that the function $f(x, y) = -y\log(\frac{x}{y}) - (1-y)\log(\frac{1-x}{1-y})$ is convex in $\mathbb{R}^2$. We state the following inequality without proof (see Cover and Thomas, 2006, for details). For any positive a_i, b_i,

$$\sum_i a_i \log\left(\frac{a_i}{b_i}\right) \geq \sum_i a_i \log\left(\frac{\sum_i a_i}{\sum_i b_i}\right). \tag{35}$$

For any $\lambda \in (0, 1)$,

$$\begin{aligned}
&f(\lambda(x_1, y_1) + (1-\lambda)(x_2, y_2)) \\
&= -(\lambda y_1 + (1-\lambda)y_2)\log\left(\frac{\lambda x_1 + (1-\lambda)x_2}{\lambda y_1 + (1-\lambda)y_2}\right) - (1 - (\lambda y_1 + (1-\lambda)y_2))\log\left(\frac{1-(\lambda x_1 + (1-\lambda)x_2)}{1-(\lambda y_1 + (1-\lambda)y_2)}\right) \\
&\leq \lambda y_1 \log\left(\frac{y_1}{x_1}\right) + (1-\lambda)y_2\log\left(\frac{y_2}{x_2}\right) + \lambda(1-y_1)\log\left(\frac{1-y_1}{1-x_1}\right) + (1-\lambda)(1-y_2)\log\left(\frac{1-y_2}{1-x_2}\right) \\
&= \lambda f(x_1, y_1) + (1-\lambda)f(x_2, y_2),
\end{aligned} \tag{36}$$

where the last step follows by applying (35) twice.

To solve for $z^*_{i,t}$, we write down the Lagrangian function

$$\sum_{i=1}^M s_{i,t}\left(z_{i,t}\log(\frac{P_{i,t}}{z_{i,t}}) + (1-z_{i,t})\log(\frac{1-P_{i,t}}{1-z_{i,t}})\right) - \gamma\left(\sum_{i=1}^M s_{i,t}z_{i,t} - v_t\right) + \sum_{i=1}^M \mu_i s_{i,t}(z_{i,t} - 1). \tag{37}$$

Setting the first derivative with respect to $z_{i,t}$ to zero gives

$$s_{i,t}\left(\log(\frac{1-z_{i,t}}{z_{i,t}}) + \log(\frac{P_{i,t}}{1-P_{i,t}})\right) - \gamma s_{i,t} + \mu_i s_{i,t} = 0 \iff \frac{z_{i,t}}{1-z_{i,t}} = \frac{P_{i,t}}{1-P_{i,t}}e^{\gamma-\mu_i}. \tag{38}$$

For interior solution, $z_{i,t} < 1$ and so $\mu_i = 0$. Thus,

$$z_{i,t} = \frac{l_{i,t}e^{\beta}}{1 + l_{i,t}e^{\beta}}, \tag{39}$$

where $l_{i,t} = \frac{P_{i,t}}{1-P_{i,t}}$. Finally, we solve the following equation

$$\begin{aligned} \sum_{i=1}^{M} s_{i,t} z_{i,t} &= v_t \\ \sum_{i=1}^{M} \frac{s_{i,t} l_{i,t} e^{\beta}}{1 + l_{i,t} e^{\beta}} &= v_t, \end{aligned} \tag{40}$$

which is an equation with one unknown and it can be shown that (40) has unique solution.

To solve (26) for $(\alpha, \beta_0, \beta_1)$, we can use the Damped Newton Method. The gradient and Hessian can be shown to be

$$\begin{aligned} \nabla f &= \sum_{t=0}^{T} \sum_{i=1}^{M} \Big(\rho_i(p_t, \alpha, \beta_0, \beta_1) - s_{i,t} \Big) \mathbf{h}_{i,t}, \\ H(f) &= - \sum_{t=0}^{T} \sum_{i=1}^{M} \rho_i(p_t, \alpha, \beta_0, \beta_1) \Big(1 - \rho_i(p_t, \alpha, \beta_0, \beta_1) \Big) \mathbf{h}_{i,t} \mathbf{h}_{i,t}^T, \end{aligned} \tag{41}$$

where

$$\mathbf{h}_{i,t}^T = \left[1, \left(\frac{R_i - p_t}{R_t} \right)^+, \left(\frac{p_i - R_t}{R_t} \right)^+ \right]. \tag{42}$$

The Damped Newton Method we use is as follows.

---

repeat
  $\lambda \leftarrow \|H(f)^{-1} \nabla f\|$
  if $\lambda > 1$ **then**
    $x_{k+1} \leftarrow x_k - \frac{1}{1+\lambda} H(f)^{-1} \nabla f$
  else
    $x_{k+1} \leftarrow x_k - H(f)^{-1} \nabla f$
  end if
until $\|x_{k+1} - x_k\|_2 < \varepsilon$

References

Barberis, N. and Huang, M. (2001). Mental accounting, loss aversion, and individual stock returns. *Journal of Finance*, 56(4):1247–1291.

Barberis, N., Huang, M., and Santos, T. (2001). Prospect theory and asset prices. *The Quarterly Journal of Economics*, 116(1):1–53.

Benartzi, S. and Thaler, R. (1995). Myopic loss aversion and the equity premium puzzle. *The Quarterly Journal of Economics*, 110(1):73–92.

Berkelaar, A., Kouwenberg, R., and Post, T. (2004). Optimal portfolio choice under loss aversion. *Rev. Econ. Stat.*, 86:973–987.

Berkowitz, S., Logue, D., and Noser, E. (1988). The total cost of transactions on the NYSE. *Journal of Finance*, 43(1):97–112.

Choi, T., Schervish, M., Schmitt, K., and Small, M. (2007). A bayesian approach to a logistic regression model with incomplete information. *Biometrics*.

Coval, J. and Shuway, T. (2005). Do behavioral biases affect prices? *Journal of Finance*, 60(1):1–34.

Cover, T. and Thomas, J. (2006). *Elements of information theory*. Wiley-Interscience.

Dempster, A., Laird, N., and Rubin, D. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society*, 39(1):1–38.

Dufour, A. and Engle, R. (2000). Time and the price impact of a trade. *Journal of Finance*, 55(6):2467–2498.

Easley, D. and O'Hara, M. (1992). Time and the process of security price adjustment. *Journal of Finance*, 47(2):577–605.

Gallant, A., Rossi, P., and Tauchen, G. (1992). Stock prices and volume. *Review of Financial Studies*, 5(2):199–242.

Gelman, A. and Rubin, D. (1992). Interference from iterative simulation using multiple sequences. *Statistical Science*, 7(4):457–472.

Grinblatt, M. and Han, B. (2005). Prospect theory, mental accounting and momentum. *Journal of Financial Economics*, 78(2):311–339.

Hasbrouck, J. (1991). Measuring the information content of stock trades. *Journal of Finance*, 46(1):179–207.

Hausman, J., Lo, A., and MacKinlay, C. (1992). An ordered probit analysis of transaction stock prices. *Journal of Financial Economics*, 31(3):319–379.

Hiemstra, C. and Jones, J. (1994). Testing for linear and nonlinear Granger causality in the stock price-volume relation. *Journal of Finance*, 49(5):1639–1664.

Jin, H. and Zhou, X. (2008). Behavioral portfolio selection in continuous time. *Mathematical Finance*, 18(3):385–426.

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision making under risk. *Econometrica*, 47:263–291.

Karpoff, J. (1987). The relation between price changes and trading volume: A survey. *Journal of Financial and Quantitative Analysis*, 22(1):109–126.

Kaustia, M. (2004). Market-wide impact of the disposition effect: evidence from IPO trading volume. *Journal of Futures Markets*, 7(2):207–235.

Lee, C. and Swaminathan, B. (2000). Price momentum and trading volume. *Journal of Finance*, 55(5):2017–2069.

McCulloch, R. and Tsay, R. (2001). Nonlinearity in high-frequency financial data and hierarchical models. *Studies in Nonlinear Dynamics and Econometrics*, 5(1). Article 1.

McFadden, D. (1973). Conditional logit analysis of quantitative choice behavior. In Zarembka, P., editor, *Frontiers in econometrics*, chapter 4. New York: Academic Press.

Meng, X. and Rubin, D. (1993). Maximum likelihood estimation via the ECM algorithm: a general framework. *Biometrika*, 80(2):267–278.

Samuelson, W. and Zeckhauser, R. (1988). Status quo bias in decision making. *Journal of Risk and Uncertainty*, 1(1):7–59.

Shefrin, H. and Statman, M. (1985). The disposition to sell winners too early and ride losers too long: theory and evidence. *Journal of Finance*, 40(3):777–790.

Tierney, L. (1994). Markov Chain for exploring posterior distributions. *Annals of Statistics*, 22(4):1701–1728.