

# An Extended Model of Effective Bid-ask Spread

Hao Zhang<sup>1</sup>, Stewart Hodges

Cass Business School, City University, London

March 20, 2012

<sup>1</sup>**Correspondence Information:** Hao Zhang, Cass Business School, 106 Bunhill Row  
London, EC1Y 8TZ, tel: +44(0)7853360208, <mailto:Hao.Zhang.1@cass.city.ac.uk>

## **Abstract**

In this paper we present an extended model for the estimation of effective bid-ask spread that improves the existing models and offers a new direction of generalisation. The quoted bid-ask spread represents the prices available at a given time for transactions only up to some relatively small amount. Moreover, it is observed that large trades are usually made at “worse” prices. Thus, we extend Roll’s model to include multiple spreads of different sizes and their associated probabilities. Bayesian model estimation and comparison methods are used to assess the fit of the model to a time series of a year of corporate bond transaction data.

**Keywords:** bid-ask spread, market microstructure, liquidity, Gibbs sampler, Bayesian model comparison

**JEL:** C58, G10, D80

# 1 Introduction

In a simple security market, at any point in time trades take place at prices posted by the specialist. Specifically, since potential traders purchase securities at the ask and sell securities at the bid, those traders pay the spread between the bid and the ask. The spread arises due to the costs of maintaining inventories, order processing and because traders act on private information, as market makers require compensation for bearing those risks.

However, trades are sometimes executed either inside or outside the quoted bid-ask spread. Schultz (2001) finds that trading costs are lower for larger trades in the over-the-counter corporate bond market. In contrast, Easley et al (1987) argue that, given that they wish to trade, informed traders prefer to trade larger amounts at any given price. As a result, market makers' pricing strategies must depend on trade size, with large trades being made at less favorable prices for the traders. In a more realistic model setting there should exist at least two different magnitudes of spreads, although, there is a large number of potential spreads in the real market. Therefore, in this paper we extend Roll's model of estimating bid-ask spread to include multiple spreads of different sizes and their associated probabilities.

In the market microstructure literature of estimating and decomposing the bid-ask spread, there are two classes of models: the serial covariance spread estimation model, and the order flow spread estimation model.

In the serial covariance spread estimation model the spread measures are derived from the serial covariance properties of transaction price changes. The first such model was developed by Roll (1984). Transaction costs are inferred from serial covariance of daily equity returns. Choi, Salandro and Shastri (1988) modify Roll's estimator and introduce a serial correlation assumption regarding transaction type. Stoll (1989)

models the relation between the bid-ask spread and the serial covariance of transaction price changes as a function of the probability of a price reversal and the magnitude of a price reversal in order to decompose the bid-ask spread into three components: adverse information costs, order processing costs, and inventory holding costs. George, Kaul and Nimalendran (1991) construct an alternative spread measure based on the serial covariance of the difference between transaction returns and returns calculated using bid prices. They show that this spread measure provides unbiased and efficient estimators of the components of the bid-ask spread.

In another class of models, the bid-ask spread is estimated via order flow regression models. Glosten and Harris (1988) applied this idea to extract the adverse selection spread component by developing an order flow transaction costs estimation model in which order sign and efficient price are treated as latent variables. Their model is estimated by Maximum Likelihood. Recently, Hasbrouck (2004) suggests a Bayesian Gibbs approach to estimate those latent variables and applies it to commodity futures transaction data. Extensions along the similar lines includes Madhavan, Richardson and Roomans (1997), Huang and Stoll (1997) and Ball and Chordia (2001).

Our model falls into the second category. Unlike other models, such as Glosten and Milgrom (1985) and Glosten and Harris (1988), we do not intend to decompose the bid-ask spread. Compared to Rydberg and Shephard (2003) where they decompose price movements by modelling sequentially price activity, direction of moves and size of moves, the purpose of our model is to estimate the bid-ask spread as well as the underlying return variance. In the original Roll model, it is not possible to distinguish spreads of different magnitudes. The Roll measure is essentially a weighted average of those spreads. Therefore, we extend the Roll model by adding an extra parameter  $\lambda$ , the so-called ‘spread multiplier’, which is constructed to separate spreads of different magnitudes. In other words, we generalize Roll’s spread estimator (a scalar) to include

a vector of spreads with associated probabilities. Our estimation procedure is based on a Bayesian Gibbs estimation method proposed by Hasbrouck (2004). Since we treat models with different values of  $\lambda$  as competitive models, the value of  $\lambda$  is determined via Bayesian model selection <sup>1</sup>.

Our extended model offers an alternative way of estimating the effective bid-ask spread more accurately than Roll's model, as well as the underlying return variance. An accurate measure of transaction costs is important for any test of market efficiency, as trading strategies must produce profits large enough to cover the transaction costs actually paid. It is also essential for tests of the role of transaction costs in asset pricing models. Transaction costs should appear as a characteristic in explaining expected returns, as investors may demand higher returns from securities with higher transaction costs [Brennan and Subrahmanyam (1996) and Hasbrouck (2009)]. Therefore, the bid-ask spread is a key measure of asset liquidity and plays a very important role in empirical asset pricing theory.

We organize our paper as follows. Section 2 presents the model specification of the original Roll model, followed by the extended model. In Section 3, we introduce the Bayesian model estimation and selection methodology, and the details of the computational procedures. Simulation analysis is carried out in Section 4 to assess the performance of the Bayesian estimators. An example of an application to the actual transaction data is given in Section 5. Section 6 contains our conclusions.

## 2 The model

This section first gives a brief introduction to the basic Roll model. Based on which, we present our extended model and discuss the key parameters.

---

<sup>1</sup>Bayes factors and various model selection criteria have been applied to compare financial models e.g. Osiewalski and Pipien (2004) and Deschamps (2011).

## 2.1 The Roll model

It has long been recognized that if trades fluctuate between bid and ask prices, then observed price changes will be negatively autocorrelated. Roll (1984) uses this property of transaction prices to derive an estimator of the bid-ask spread. One advantage of his model is that it is based only on published transaction prices.

In an efficient market the price dynamics may be stated as

$$m_t = m_{t-1} + \epsilon_t, \quad (1)$$

$$p_t = m_t + sq_t, \quad (2)$$

where  $m_t$  is the unobservable efficient price,  $p_t$  is the transaction price observed at time  $t$ ,  $s$  is one-half the bid-ask spread, and the  $q_t$  are trade direction indicators, which take the values  $+1$  for buy orders or  $-1$  for sell orders with equal probability.

The changes in transaction prices between two successive trades is

$$p_t - p_{t-1} = m_t + sq_t - (m_{t-1} + sq_{t-1}), \quad (3)$$

$$\Delta p_t = s\Delta q_t + \epsilon_t. \quad (4)$$

Thus,

$$\begin{aligned} Cov(\Delta p_t, \Delta p_{t-1}) &= E[\Delta q_t \Delta q_{t-1} s^2] + E[\Delta q_t \epsilon_{t-1} s] \\ &\quad + E[\Delta q_{t-1} \epsilon_t s] + E[\epsilon_t \epsilon_{t-1}], \end{aligned} \quad (5)$$

where  $Cov(\Delta p_t, \Delta p_{t-1})$  is the first-order autocovariance of the price changes.

In deriving the Method of Moments estimator Roll (1984) makes several assumptions:

1. Successive transaction types are independent. Thus,

$$E[q_t q_{t-1}] = 0.$$

2. The half-spread  $s$  is constant.

3. Order flows do not contain information about future fundamental price changes.

Thus,

$$E[\Delta q_{t-1} \epsilon_t] = 0.$$

4. Changes in fundamental value cannot predict order flows:

$$E[\Delta q_t \epsilon_{t-1}] = 0.$$

5. The innovations in the fundamental price process reflect public information and are assumed to be independent. Therefore,

$$E[\epsilon_t \epsilon_{t-1}] = 0.$$

Then, from the joint probabilities in Table 1 and Equation (5), we obtain

$$Cov(\Delta p_t, \Delta p_{t-1}) = -s^2. \tag{6}$$

This gives Roll's Method of Moments spread estimator

$$\hat{s} = \sqrt{-Cov(\Delta p_t, \Delta p_{t-1})}. \tag{7}$$

If the innovations in the efficient price process are assumed to follow a normal distribution

with mean 0 and constant variance  $\sigma_\epsilon^2$ , the Method of Moments variance estimator is

$$\hat{\sigma}_\epsilon^2 = \text{Var}(\Delta p_t) - s^2 \text{Var}(\Delta q_t) = \text{Var}(\Delta p_t) - 2s^2. \quad (8)$$

However, Roll's method of moments spread estimator has performed poorly, generating undefined spread estimates almost half of the time when applied to daily transaction data on equities [Harris (1990)]. It is also biased downward as a result of Jensen's inequality. Having said that, several refinements to the Roll's model are possible. Our extended model, which will be introduced in the next section, provides a richer structure, and does not have undefined spread estimates.

## 2.2 The extended Roll model

The fact that trades are sometimes executed either inside or outside the posted bid and ask indicates that the posted spread sometimes may not represent the transaction cost that investors can expect. Therefore, distinguishing spreads of different magnitudes is of importance for several theoretical and practical purposes mentioned earlier in this paper.

This motivates our extended Roll model aiming at separating spreads with two different magnitudes. In our model the trade direction indicators  $q_t$  are generalised as follow:

$$q_t \in \{-\lambda, -1, +1, +\lambda\}, \quad (9)$$

where  $\lambda$  is the so-called 'spread multiplier' which is used to distinguish different spreads. Thus, for expositional ease, we define  $\lambda \in [1, +\infty)^2$ .  $\lambda \neq 1$  means that there exist two different sizes of spreads.  $-\lambda$  and  $+\lambda$  are indicators of the abnormal spreads that

---

<sup>2</sup>Equivalently, we can define  $\lambda \in (0, 1]$ . For computational ease, we choose to use  $\lambda \in (0, 1]$  later in the numerical exercises and applications.

are outside normal bid-ask spreads. Roll's model is a special case of our extended model corresponding to  $\lambda = 1$ . For estimation purposes we also assume here that  $\epsilon \sim i.i.d.N(0, \sigma_\epsilon^2)$ . We shall not hesitate to drop the subscript of  $\sigma_\epsilon^2$ , writing  $\sigma^2$  wherever convenient.

In summary, the half-spread,  $s$ , and the variance of the efficient price changes,  $\sigma_\epsilon^2$ , are unknown parameters from the regression specification

$$\Delta p_t = s\Delta q_t + \epsilon_t \quad (10)$$

$$q_t \in [-\lambda, -1, +1, +\lambda] \quad (11)$$

$$\epsilon \sim i.i.d.N(0, \sigma_\epsilon^2), \quad (12)$$

where  $q_t$  and  $\lambda$  are latent variables.

A slightly simplified graphical illustration of the extended model is shown in Figure 2. At any point in time, there exist four positions for the security price to be located. They are: Outer bid, outer ask, inner bid and inner ask, for which the trade direction indicator  $q_t$  assign corresponding values, e.g.  $-1, +1, -\lambda$  and  $+\lambda$ , respectively. For instance, in Figure 2, the security price at time  $t - 1$  is at the quote setter's outer bid. Assume that no new information arrives in the market about the security, there are 16 possible paths of observed price between successive time periods. For convenient demonstration and comparison, we use  $\alpha$  and  $\theta$  to control the probability of each scenario.  $\alpha$  is the probability of existence of outer bid and ask. Then  $\theta$  determines whether the model has a symmetric structure.

It is also interesting to look at the only two non-zero second moments, namely autocovariance and variance, under our extended model specification. For computational ease, in the rest of this paper, we let  $\lambda \in (0, 1]$ .

Using the probabilities of different prices as just described, we have,

$$\begin{aligned} Cov(\Delta p_t, \Delta p_{t-1}) = & -s^2(1-\alpha)^3 - s^2\alpha^2(1-\alpha)[\lambda^2(1+4\theta-4\theta^2)+1] \\ & -s^2(1-\alpha)^2\alpha(\lambda^2+2) - 4s^2\alpha^3\lambda^2\theta(1-\theta), \end{aligned} \quad (13)$$

and

$$Var(\Delta p_t) = \sigma_\epsilon^2 + s^2[8\lambda^2(\theta-\theta^2)\alpha^2 + 2(\lambda^2+1)\alpha^2(1-\alpha) + 2(\lambda^2+2)(1-\alpha)^2\alpha + 2(1-\alpha)^3]. \quad (14)$$

Assuming that  $s = 1$  and  $\theta = 0.5$ , for  $\lambda \neq 1$  Roll's spread estimator will be biased as shown in Figure 3, and the bias increases while  $\alpha$  increases; for  $\lambda \neq 1$  Roll's variance estimator is also biased as shown in Figure 5, and the bias increases as  $\alpha$  increases.

Obviously, if we apply Roll's model to the kind of transaction data as shown in Figure 1, the results will be biased. This problem can be fixed by our extended model.

### 3 The model estimation and selection

The extended model described in the last section is estimated via a Bayesian approach proposed by Hasbrouck (2004). The latent variables in the extended model are the trade direction indicators  $q_t$  and the spread multiplier  $\lambda$ . Thus, there is no closed-form likelihood function. Moreover, it is difficult to compute the higher moments, using GMM estimation, which are needed in the extended model. We treat models with different specifications of  $\lambda$  as competitive models. The best model is determined via a Bayesian model selection method. It is also worth noticing that this approach shares some similarities with the Expectation-Maximization (EM) algorithm. First, we describe the parameter estimation procedure in detail, followed by the model selection method.

### 3.1 The Bayesian estimation procedure

The extended Roll model has two parameters  $s, \sigma_\epsilon$ , and  $T$  latent variables  $q = (q_1, q_2, \dots, q_T)$  given the value of  $\lambda$ . We estimate the model parameters via a Bayesian approach, using the Gibbs sampler. That is, given a known sequence of observations  $p = (p_1, p_2, \dots, p_T)$ , we estimate  $s$  and  $\sigma_\epsilon$  under mean squared error risk rule by taking the mean of the posterior distributions of  $s$  and  $\sigma_\epsilon$  for fixed  $\lambda$ .

Gibbs sampling is a technique for generating random variables from a marginal distribution indirectly, without having to calculate the joint density. As a Markov chain Monte Carlo method, the Gibbs sampler generates sample values from the distribution of each variable in turn, conditioned on the current values of the other variables.

Gibbs sampling is, essentially, an interactive procedure. An iteration is generally termed a “sweep”. Initially, i.e.,  $j = 0$ , for fixed  $\lambda$  the parameters and the latent variables are set to any values, where tick test results can be used as initial values for  $q$ . Denote these initial values  $(s^{(0)}, \sigma_\epsilon^{(0)}, q^{(0)})$ , where  $q^{(0)} = (q_1^{(0)}, q_2^{(0)}, \dots, q_T^{(0)})$ .

The steps in the first sweep ( $j = 1$ ) given  $\lambda$  and  $p$  are:

1. Draw  $s^{(1)}$  from  $f_s(s|\sigma_\epsilon^{(0)}, q^{(0)}, p, \lambda)$ ,
2. Draw  $\sigma_\epsilon^{(1)}$  from  $f_{\sigma_\epsilon}(\sigma_\epsilon|s^{(1)}, q^{(0)}, p, \lambda)$ ,
3. Draw  $q^{(1)}$  from  $f_q(q|s^{(1)}, \sigma_\epsilon^{(1)}, p, \lambda)$ ,

where  $f(\cdot|\cdot)$  is the complete conditional density.

Similarly, draws for  $q_t^{(1)}$  for  $t = 1, \dots, T$  are made sequentially using the Gibbs sampler. The steps in the first sweep ( $j = 1$ ) are then:

1. Draw  $q_1^{(1)}$  from  $f_q(q_1|s^{(1)}, \sigma_\epsilon^{(1)}, p, \lambda, q_2^{(0)}, q_3^{(0)}, \dots, q_T^{(0)})$ ,
2. Draw  $q_2^{(1)}$  from  $f_q(q_2|s^{(1)}, \sigma_\epsilon^{(1)}, p, \lambda, q_1^{(1)}, q_3^{(0)}, \dots, q_T^{(0)})$ ,

$\vdots \dots$

T. Draw  $q_T^{(1)}$  from  $f_q(q_T|s^{(1)}, \sigma_\epsilon^{(1)}, p, \lambda, q_1^{(1)}, q_2^{(1)}, \dots, q_{T-1}^{(1)})$ .

Note that, all parameters and latent variables except for the component being drawn are taken as given.

The next iteration starts with a draw of  $s^{(2)}$  conditional on  $\sigma_\epsilon^{(1)}, q^{(1)}, p$  and  $\lambda$ . Repeating this  $n$  times, we generate a sequence of draws  $(s^{(j)}, \sigma_\epsilon^{(j)}, q^{(j)})$  for  $j = 1, \dots, n$ , where  $q^{(j)} = (q_1^{(j)}, q_2^{(j)}, \dots, q_T^{(j)})$ . The Gibbs principle ensures that the limiting distribution of the  $n$ th draw (as  $n \rightarrow \infty$ ) is  $F(s, \sigma_\epsilon, q|p, \lambda)$ . The limiting draw for any parameter is distributed in accordance with the corresponding marginal posterior, i.e., the limiting density of  $s^{(n)}$  is  $f_s(s|p, \lambda)$ . Given some continuous function of the model parameters,  $h(s, \sigma_\epsilon, q)$  and for a set of parameter draws,  $(s^{(j)}, \sigma_\epsilon^{(j)}, q^{(j)} : j = 1, \dots, n)$ , the corresponding sequence  $(h(s^{(j)}, \sigma_\epsilon^{(j)}, q^{(j)} : j = 1, \dots, n))$  generally has its limiting distribution for  $h(s, \sigma_\epsilon, q)$ .

The consistent estimates of population parameters  $\theta = (s, \sigma_\epsilon)$  are given by the posterior mean  $\theta = E_{\theta|p, \lambda}(\theta)$ . If one is interested in some statistic  $g(\theta)$ ,  $E_{\theta|p, \lambda}g(\theta)$  is a consistent estimate of  $g(\theta)$ .

### 3.2 The Bayesian model selection

Suppose we have  $K$  models  $M_k$  ( $k = 1, 2, \dots, K$ ), given the model-specific parameter vector  $\theta_k$ . Our prior information on these models can be used to assign each of them a prior probability  $\pi(M_k)$ . A data set  $y = (y_1, y_2, \dots, y_n)$  is used to update these prior probabilities.

The Bayes factor for any two models  $k$  and  $l$  is defined as:

$$\hat{B}_{kl} = \frac{m(y|M_k)}{m(y|M_l)}, \quad (15)$$

where  $m(y|M_k)$  is the so-called marginal likelihood of the data  $y$  under model  $M_k$ . Therefore, evaluating the marginal likelihood is the key to Bayesian model selection.

We drop  $M_k$  and subscripts  $k$ , and consider the situation where  $f(y|\theta)$  is the likelihood function for a given model and  $\pi(\theta)$  is the prior density. The marginal likelihood of the data  $y$  can be written as

$$m(y) = \frac{f(y|\theta)\pi(\theta)}{\pi(\theta|y)}, \quad (16)$$

where the numerator is the product of the likelihood and the prior, and the denominator is the posterior density. Conveniently, it can be expressed on logarithm scale as

$$\ln m(y) = \ln f(y|\theta) - \ln \frac{\pi(\theta|y)}{\pi(\theta)}, \quad (17)$$

where the first term measures that how well the model fits the data given the most probable parameter values, and the second term penalizes the model according to its complexity. Hence, the marginal likelihood automatically incorporates a tradeoff between model fit and its complexity.

The marginal likelihood can be easily evaluated from the Gibbs output, generated during the simulation stage, as suggested by Chibs (1995).

Let  $z$  denote latent data and suppose that for a given set of vector blocks  $\theta = (\theta_1, \theta_2, \dots, \theta_B)$ , the Gibbs sampling algorithm is applied to the set of  $(B + 1)$  complete conditional densities,  $\{\pi(\theta_r|y, \theta_s (s \neq r), z)\}_{r=1}^B$  and  $\pi(z|y, \theta)$ . The goal is to compute the marginal likelihood  $m(y)$  from the output  $\{\theta^{(j)}, z^{(j)}\}_{j=1}^n$  obtained from these full

conditional densities. In some situations, the density  $f(y|\theta) = \int f(y, z|\theta)dz$  is not available in closed form, however the likelihood  $f(y|\theta, z)$  is. The marginal likelihood  $m(y)$  in this situation has the form

$$m(y) = \frac{f(y|\theta, z)\pi(\theta, z)}{\pi(\theta, z|y)}, \quad (18)$$

where the likelihood  $f(y|\theta, z)$ , the prior  $\pi(\theta, z)$ , and the multivariate posterior density  $\pi(\theta, z|y)$  can be evaluated at the selected high density point  $(\theta^*, z^*)$ . When estimating  $\pi(\theta, z|y)$ ,  $z$  can be treated as an additional block, e.g.  $z \equiv \theta_{B+1}$ .

The posterior density at the selected high density point is given by

$$\pi(\theta^*|y) = \pi(\theta_1^*|y) \times \pi(\theta_2^*|y, \theta_1^*) \times \dots \times \pi(\theta_{B+1}^*|y, \theta_1^*, \dots, \theta_B^*), \quad (19)$$

where the first term on the right hand side is the marginal ordinate, which can be estimated from the draws of the initial Gibbs sampling, e.g.  $\hat{\pi}(\theta_1^*|y) = n^{-1} \sum_{j=1}^n \pi(\theta_1^*|y, \theta_s^{(j)}) (1 < s \leq B + 1)$ , and the rest of the terms are the reduced conditional ordinates with the typical form of  $\pi(\theta_r^*|y, \theta_1^*, \theta_2^*, \dots, \theta_{r-1}^*)$ , which are given by

$$\int \pi(\theta_r^*|y, \theta_1^*, \theta_2^*, \dots, \theta_{r-1}^*, \theta_{r+1}, \dots, \theta_B) d\pi(\theta_{r+1}, \dots, \theta_B|y, \theta_1^*, \theta_2^*, \dots, \theta_{r-1}^*). \quad (20)$$

To estimate the above integral, continue the sampling with the complete conditional densities of  $\{\theta_r, \theta_{r+1}, \dots, \theta_{B+1}\}$ , where in each of these full conditional densities,  $\theta_s$  is set equal to  $\theta_s^*$ , ( $s \leq r - 1$ ). If the draws from the reduced complete conditional Gibbs sampling are denoted by  $\{\theta_r^{(g)}, \theta_{r+1}^{(g)}, \dots, \theta_{B+1}^{(g)}\}$ , then an estimate of (20) is

$$\hat{\pi}(\theta_r^*|y, \theta_s^* (s < r)) = n^{-1} \sum_{g=1}^n \pi(\theta_r^*|y, \theta_1^*, \theta_2^*, \dots, \theta_{r-1}^*, \theta_l^{(g)} (l > r)). \quad (21)$$

The log of the marginal likelihood is

$$\begin{aligned} \ln \hat{m}(y) = & \ln f(y|\theta^*, z^*) + \ln \pi(\theta^*, z^*) - \ln \hat{\pi}(\theta_1^*|y) \\ & - \sum_{r=2}^B \ln \hat{\pi}(\theta_r^*|y, \theta_s^*(s < r)) - \ln \pi(z^*|y, \theta_s^*(s < B + 1)). \end{aligned} \quad (22)$$

In our situation, if we suppress  $\lambda$  for convenience, Gibbs sampling is applied to the complete conditional densities

$$\pi(\theta_1|y, \theta_2, z) = f_s(s|p, \sigma_\epsilon, q), \quad (23)$$

$$\pi(\theta_2|y, \theta_1, z) = f_{\sigma_\epsilon}(\sigma_\epsilon|p, s, q), \quad (24)$$

$$\pi(z|y, \theta) = f_q(q|p, s, \sigma_\epsilon), \quad (25)$$

where  $\theta = (\theta_1, \theta_2)$  and  $z \equiv \theta_3$ .

The performance of the Gibbs sampler and an example of the application of the Bayesian model estimation and selection techniques are illustrated in the next two sections.

## 4 Simulation analysis

The performance of the Bayesian estimators may be illustrated by considering simulated samples under three different specifications. The specifications correspond to typical situations in the marketplace. Assume trades can execute at bid-ask spreads with only two different magnitudes, e.g. inner bid-ask spreads and outer bid-ask spreads.

First scenario, trades execute at inner bid-ask spreads and outer bid-ask spreads with equal probabilities. That means  $p(q = \pm\lambda) = p(q = \pm 1)$ , say,  $p(q = \lambda) = p(q = -\lambda) =$

$p(q = 1) = p(q = -1) = 0.25$  in Figure 7, where  $p(\cdot)$  is the probability for trade direction indicator  $q$ .

Second scenario, most of trades are executed at outer bid-ask spread, while a few orders with better prices execute at inner bid-ask spread. In other words,  $p(q = \pm\lambda) > p(q = \pm 1)$ , say,  $p(q = \lambda) = p(q = -\lambda) = 0.40$  and  $p(q = 1) = p(q = -1) = 0.10$  in Figure 8.

Third scenario, most orders are traded at normal bid-ask spread. The market maker charges a few orders at a larger bid-ask spread. Therefore, in this case we have  $p(q = \pm\lambda) < p(q = \pm 1)$ , say,  $p(q = \lambda) = p(q = -\lambda) = 0.10$  and  $p(q = 1) = p(q = -1) = 0.40$  in Figure 9.

Four sample paths with different values of  $s$ ,  $\sigma^2$  and  $\lambda$  are simulated for each case. All simulated samples are of length 300. For each path the Gibbs sampler is run for 50,000 iterations, with first 20% discarded. The joint posteriors of  $s$  and  $\sigma^2$  for each sample path are characterised by the remaining 80% draws, and displayed in four sub-panels in Figure 7, 8 and 9, respectively.

There are commonalities among three cases. Firstly, in the left-hand side panels, the joint posteriors are neatly packed together, whereas in the right-hand side panels, the joint posteriors are more scattered and negatively sloped. Secondly, the joint posteriors are slightly more negatively correlated in the top panels than they are in the bottom panels. Finally, the joint posteriors in the bottom-left panels are slightly more dispersed than they are in the top-left panels, whereas the opposite is true for the right-hand side panels.

In Figure 7, the joint posteriors are generally centered around the true values. However, in Figure 8 the posterior for  $s$  is biased downwards, whereas in Figure 9 it is biased upwards, except that in the top-left panels.

The Bayesian estimation approach effectively needs to balance the composition of the price changes between the transient component and the permanent component in the security value as well as balance the composition of the transient component itself, which is the combination of small and large bid-ask spread changes. Hence, the concentration of the joint posterior generally indicates the certainty that the Bayesian trade classification procedure can assign a direction to a particular trade with. Intuitively, when the outer bid-ask spread  $s$  is relatively larger than the volatility of the efficient price changes  $\sigma$  and the inner bid-ask spread  $\lambda s$ , the bayesian procedure will be more certain about whether a spike is caused by a large bid-ask spread accompanied by a small efficient price increment or a small bid-ask spread accompanied by a large efficient price increment. The misclassification results in the spreading out and the negative sloping of the joint posterior. Moreover, when  $\lambda$  is close to 1, it is more difficult to distinguish the outer bids and asks from inner bids and asks due to the existence of non-zero variance.

When the data is not sufficient to determine a direction for a trade, the prior information will play a more influential role. In the asymmetric cases (Figure 8 and 9), when the volatility is large relative to  $s$  and  $\lambda s$ , the procedure may make inference more based on the prior. As a uniform prior is applied, the procedure under-estimates the proportion of large bid-ask spread in the second case. This results in downward bias for the estimated bid-ask spread in Figure 8. A similar analysis applies to the third case(Figure 9).

## 5 Empirical application

Next the extended model is applied to a time series of intra-day bond prices, as the data in Figure 10 suggests that the extended model might be an appropriate underlying model. Table 2 summarizes the basic information about the bond and the descriptive

statistics of the sample.

The Gibbs sampler is run for 10,000 iterations, with first 20% discarded. Our results for the data are presented in Table 3. The Method of Moments estimate for the half-spread  $s$  is lower than two standard deviations of the Bayesian estimate under the extended model, but bigger than two standard deviations of the Bayesian estimate under Roll's model. For the variance  $\sigma^2$ , on the other hand, the Method of Moments estimate is bigger than two standard deviations of the Bayesian estimate under the extended model, but smaller than two standard deviations of the Bayesian estimate under Roll's model. Generally speaking, Roll's model gives a weighted average estimate of the bid-ask spreads and the variance of the efficient price changes. This is because that Roll's model does not identify shocks in bid-ask spreads. The estimated log marginal likelihood is equal to -1182.8434 for Roll's model. By contrast with the extended model, the maximum of the estimated log marginal likelihood is equal to -326.5221 at  $\lambda = 0.13$ . Therefore, the extended model fits the data better than the original Roll's model.

Moreover, we take the posterior median  $\hat{m}_t$  as the estimate of the efficient price  $m_t$ . The estimated inner and outer bid-ask bounces defined as  $\hat{m}_t \pm s$  and  $\hat{m}_t \pm s\lambda$ , respectively, are also shown in Figure 10. It is clear from the figure that trades executed with different transaction costs are distributed between the corresponding estimated bid-ask bounces. The estimated efficient price changes  $\Delta\hat{m}$  under the extended model exhibits less serial correlation than they do under Roll's model, as shown in Table 4. This means that the transitory price components, which reflect the degree of illiquidity and lead to negatively serially correlated price changes, are better separated under the extended model.

## 6 Conclusions

In this paper we present an extended model for the estimation of effective bid-ask spread that allows trades to execute either inside or outside the quoted bid-ask spread. The extended model is estimated by a Bayesian Gibbs approach based on only transaction data. By applying the extended model to real transaction data, we show that the extended model fits the data better and reduces the serial correlation of the underlying efficient price process. The new model together with the model estimation and selection procedures offers an alternative way to more accurately estimate the effective bid-ask spread and the underlying return variance.

## Appendix A: Numerical standard error of marginal likelihood estimate

The numerical standard error of the marginal likelihood estimate gives the variation that can be expected in the estimate if the sampling were to be done again, but the selected point at which the ordinate is evaluated remains the same.

To calculate the numerical standard error, we first define the vector process

$$h = \begin{pmatrix} h_1(\sigma_\epsilon, q) \\ h_2(q) \end{pmatrix} = \begin{pmatrix} f(s^*|p, \sigma_\epsilon, q) \\ f(\sigma_\epsilon^*|p, s^*, q) \end{pmatrix}, \quad (26)$$

where in the first element  $(\sigma_\epsilon, q)$  follow  $[\cdot|p]$  whereas in the second element  $q$  follows  $[\cdot|p, s^*]$  as in the second stage of Gibbs sampling.

Then the estimate of  $h$  is

$$\hat{h} \equiv n^{-1} \sum_{g=1}^n h^{(g)} = \begin{pmatrix} \hat{f}(s^*|p) \\ \hat{f}(\sigma_\epsilon^*|p, s^*) \end{pmatrix}. \quad (27)$$

The goal is to find the variance of two functions of  $\hat{h}$ , namely  $\Psi \equiv \ln \hat{f}(s^*|p) + \ln \hat{f}(\sigma_\epsilon^*|p, s^*)$ . The variance of  $\hat{h}$  is determined as

$$Var(\hat{h}) = \frac{1}{n} \left[ \Omega_0 + \sum_{s=1}^q \left( 1 - \frac{s}{q+1} \right) (\Omega_s + \Omega'_s) \right], \quad (28)$$

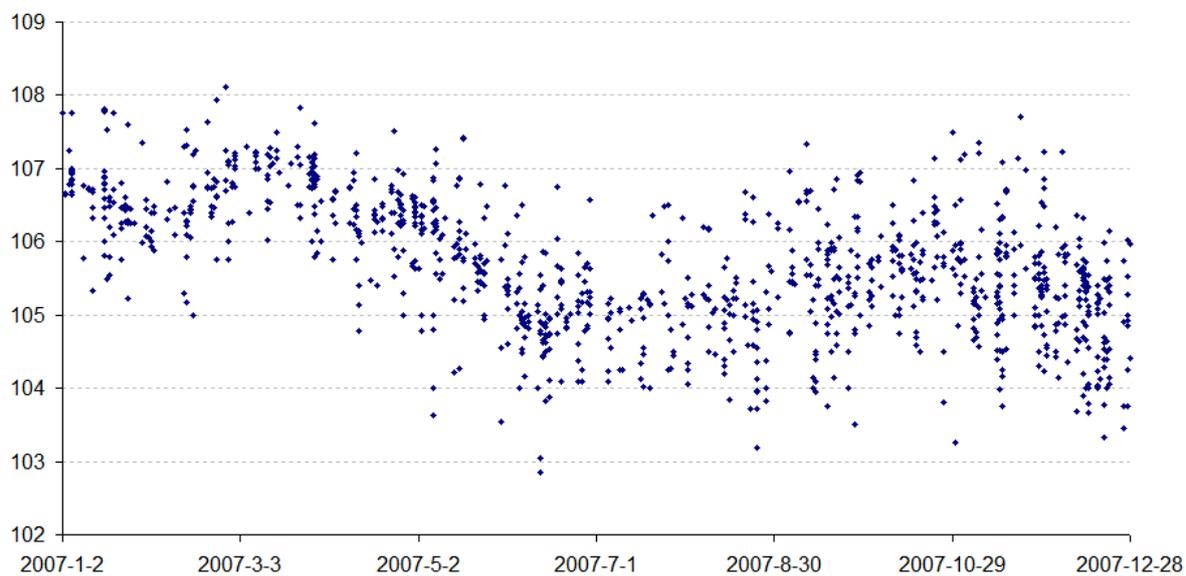
where  $\Omega_s = \frac{1}{n} \sum_{g=s+1}^n (h^{(g)} - \hat{h})(h^{(g)} - \hat{h})'$  and  $q$  takes some value such that the autocorrelation function is negligible. In our application,  $q$  is set equal to 10.

The variance of  $\Psi$  is found by the delta method to be

$$\left(\frac{\partial\Psi}{\partial\hat{h}}\right)' Var(\hat{h}) \left(\frac{\partial\Psi}{\partial\hat{h}}\right). \quad (29)$$

The numerical standard error is obtained by taking the square root of this variance.

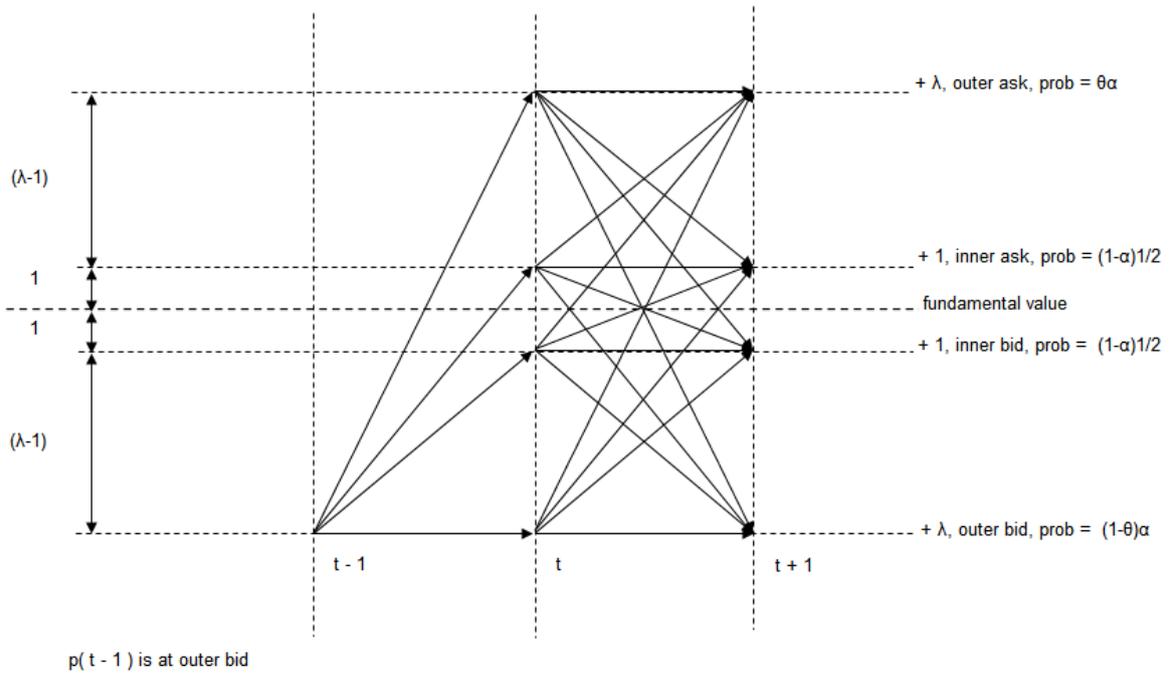
## Appendix B: Figures and Tables



**Figure 1:** Time series plot of corporate bond transaction prices

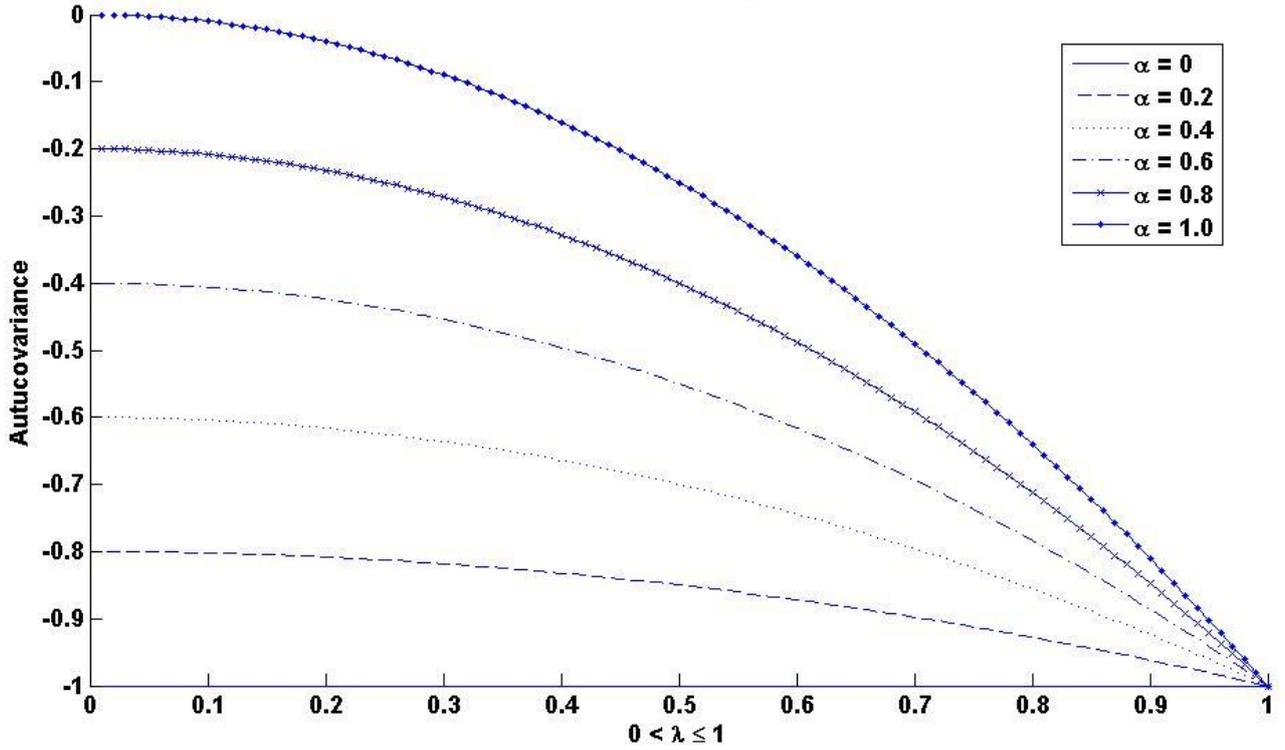
**Table 1:** Joint probabilities of consecutive price changes

		$\Delta q_{t-1}$		
		-2	0	+2
$\Delta q_t$	-2	0	$\frac{1}{8}$	$\frac{1}{8}$
	0	$\frac{1}{8}$	$\frac{4}{8}$	$\frac{1}{8}$
	+2	$\frac{1}{8}$	$\frac{1}{8}$	0



**Figure 2:** Price paths.

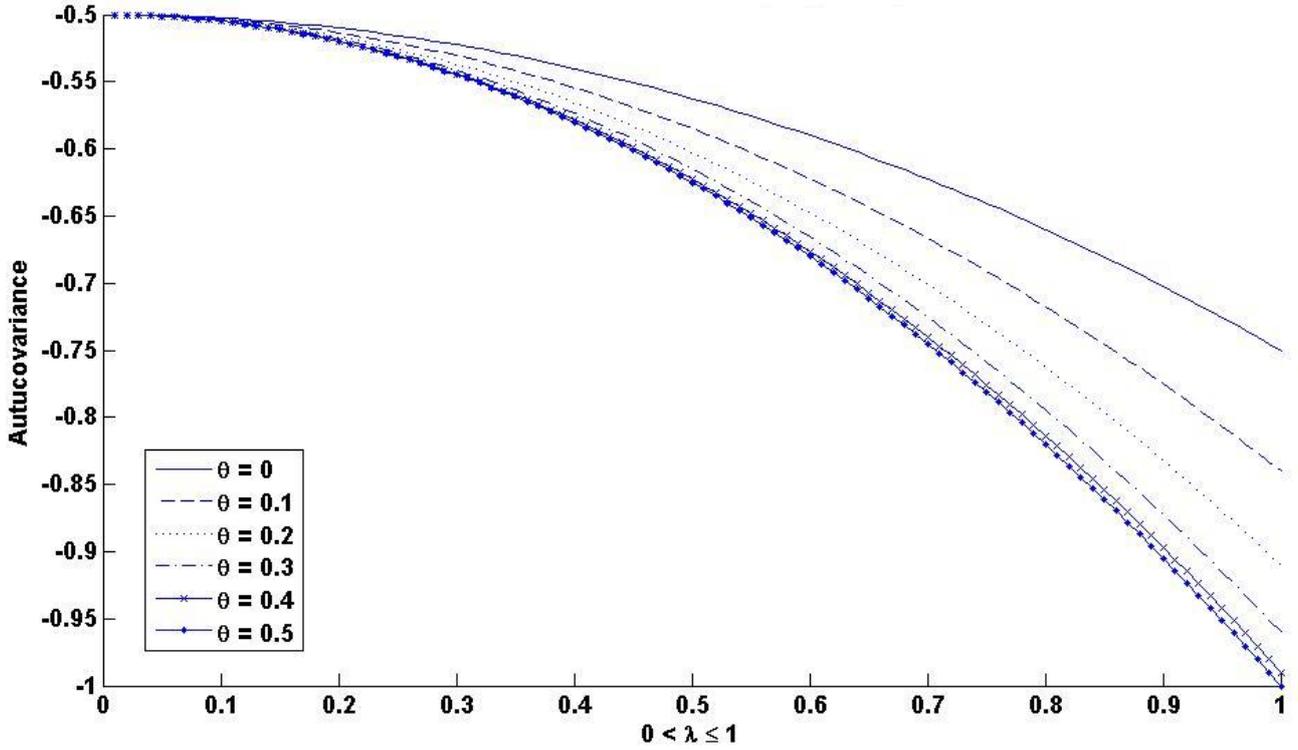
This figure shows the possible paths of observable market price between successive time periods, given that the price at time  $t - 1$  is at outer bid, and given that no new information arrives in the market.  $\lambda$  is the ‘spread multiplier’ which is used to distinguish different spreads.  $\lambda \neq 1$  means that there exist two different sizes of spreads. At any point in time, there exist four positions for the security price to be located. They are: Outer bid, outer ask, inner bid and inner ask, for which the trade direction indicator  $q_t$  assign corresponding values, e.g.  $-1, +1, -\lambda$  and  $+\lambda$ , respectively.  $\alpha$  controls the possibility of the occurrence of outer bid and ask.  $\theta$  determines whether the model has a symmetric structure for outer bid and ask. Given the position of the price at time  $t - 1$ , there are sixteen possible price paths in total between  $t - 1$  and  $t + 1$ .



**Figure 3:** Plot of autocovariance

This figure provides the autocovariances calculated under the extended model.  $s$  is the half-spread.  $\lambda$  is the ‘spread multiplier’ which is used to distinguish different spreads.  $\lambda \neq 1$  means that there exist two different sizes of spreads.  $\alpha$  controls the possibility of the occurrence of outer bid and ask.  $\theta$  determines whether the model has a symmetric structure for outer bid and ask. Assume  $s = 1$  and  $\theta = 0.5$ . Different lines in this figure represent autocovariances as a function of  $\lambda$ , corresponding to cases where  $\alpha$  is assigned to different values ranging from 0 to 1. Autocovariance under the extended model is computed as

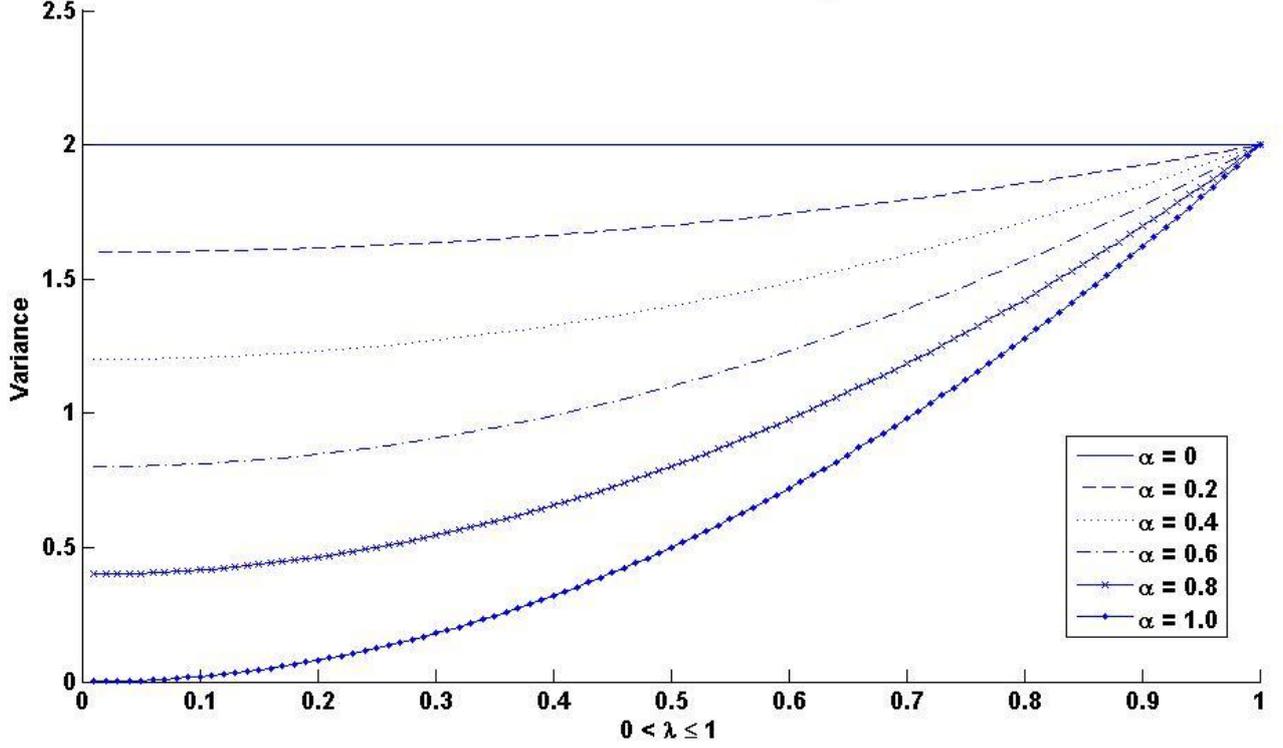
$$Cov(\Delta p_t, \Delta p_{t-1}) = -s^2(1 - \alpha)^3 - s^2\alpha^2(1 - \alpha)[\lambda^2(1 + 4\theta - 4\theta^2) + 1] - s^2(1 - \alpha)^2\alpha(\lambda^2 + 2) - 4s^2\alpha^3\lambda^2\theta(1 - \theta).$$



**Figure 4:** Plot of autocovariance

This figure provides the autocovariances calculated under the extended model.  $s$  is the half-spread.  $\lambda$  is the ‘spread multiplier’ which is used to distinguish different spreads.  $\lambda \neq 1$  means that there exist two different sizes of spreads.  $\alpha$  controls the possibility of the occurrence of outer bid and ask.  $\theta$  determines whether the model has a symmetric structure for outer bid and ask. Assume  $s = 1$  and  $\alpha = 0.5$ . Different lines in this figure represent autocovariances as a function of  $\lambda$ , corresponding to cases where  $\theta$  is assigned to different values ranging from 0 to 0.5. Autocovariance under the extended model is computed as

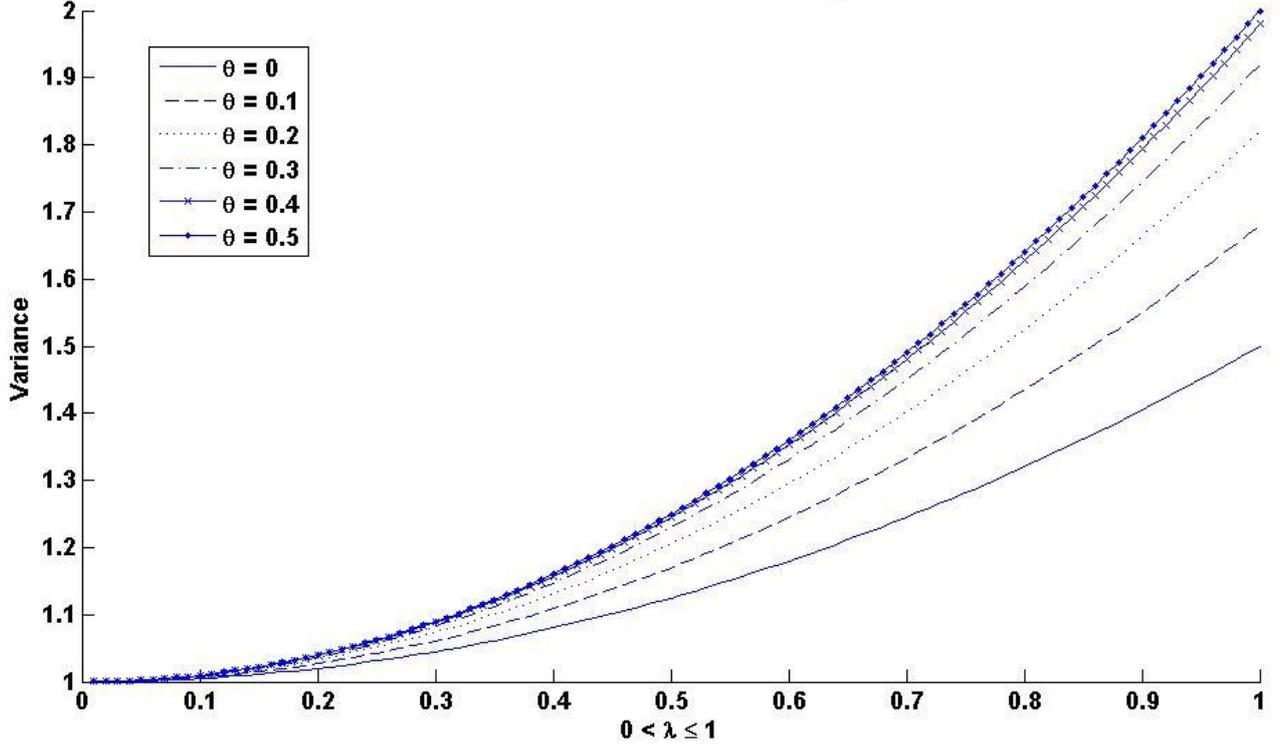
$$Cov(\Delta p_t, \Delta p_{t-1}) = -s^2(1 - \alpha)^3 - s^2\alpha^2(1 - \alpha)[\lambda^2(1 + 4\theta - 4\theta^2) + 1] - s^2(1 - \alpha)^2\alpha(\lambda^2 + 2) - 4s^2\alpha^3\lambda^2\theta(1 - \theta).$$



**Figure 5:** Plot of variance

This figure provides the variances calculated under the extended model.  $s$  is the half-spread.  $\sigma_\epsilon$  is the variance of the efficient price changes.  $\lambda$  is the ‘spread multiplier’ which is used to distinguish different spreads.  $\lambda \neq 1$  means that there exist two different sizes of spreads.  $\alpha$  controls the possibility of the occurrence of outer bid and ask.  $\theta$  determines whether the model has a symmetric structure for outer bid and ask. Assume  $s = 1$ ,  $\sigma_\epsilon = 0$  and  $\theta = 0.5$ . Different lines in this figure represent variances as a function of  $\lambda$ , corresponding to cases where  $\alpha$  is assigned to different values ranging from 0 to 1. Variance under the extended model is computed as

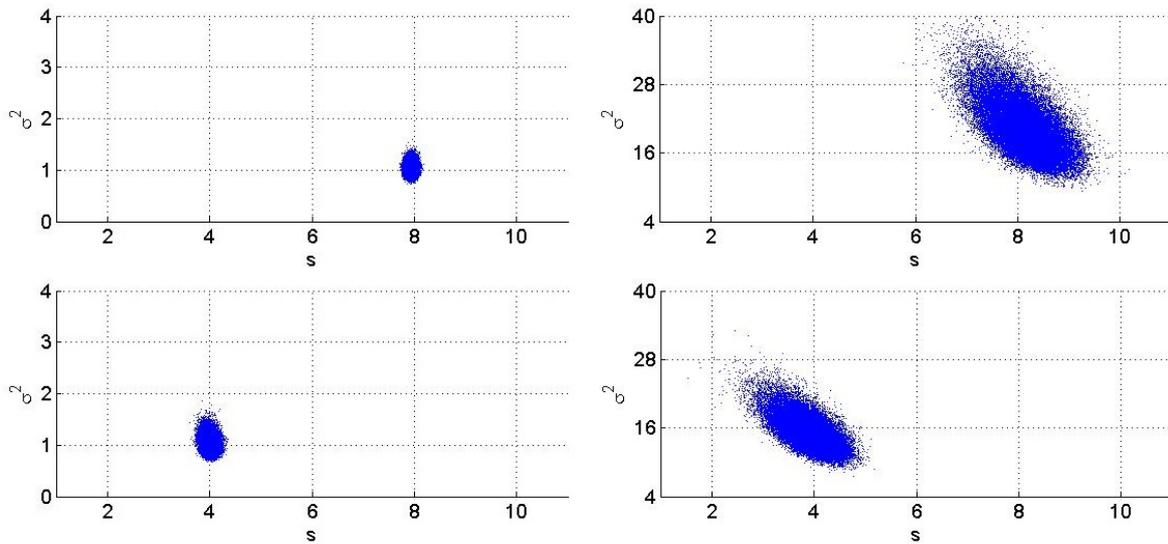
$$Var(\Delta p_t) = \sigma_\epsilon^2 + s^2[8\lambda^2(\theta - \theta^2)\alpha^2 + 2(\lambda^2 + 1)\alpha^2(1 - \alpha) + 2(\lambda^2 + 2)(1 - \alpha)^2\alpha + 2(1 - \alpha)^3].$$



**Figure 6:** Plot of variance

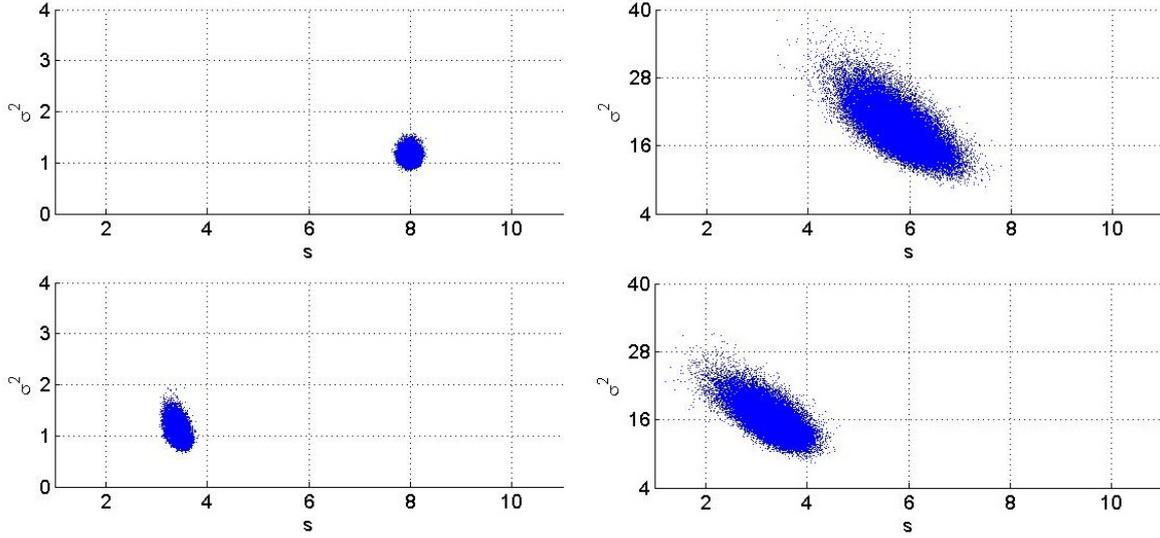
This figure provides the variances calculated under the extended model.  $s$  is the half-spread.  $\sigma_\epsilon$  is the variance of the efficient price changes.  $\lambda$  is the ‘spread multiplier’ which is used to distinguish different spreads.  $\lambda \neq 1$  means that there exist two different sizes of spreads.  $\alpha$  controls the possibility of the occurrence of outer bid and ask.  $\theta$  determines whether the model has a symmetric structure for outer bid and ask. Assume  $s = 1$ ,  $\sigma_\epsilon = 0$  and  $\alpha = 0.5$ . Different lines in this figure represent variances as a function of  $\lambda$ , corresponding to cases where  $\theta$  is assigned to different values ranging from 0 to 0.5. Variance under the extended model is computed as

$$Var(\Delta p_t) = \sigma_\epsilon^2 + s^2[8\lambda^2(\theta - \theta^2)\alpha^2 + 2(\lambda^2 + 1)\alpha^2(1 - \alpha) + 2(\lambda^2 + 2)(1 - \alpha)^2\alpha + 2(1 - \alpha)^3].$$



**Figure 7:** Joint posteriors of half spread  $s$  and variance  $\sigma^2$

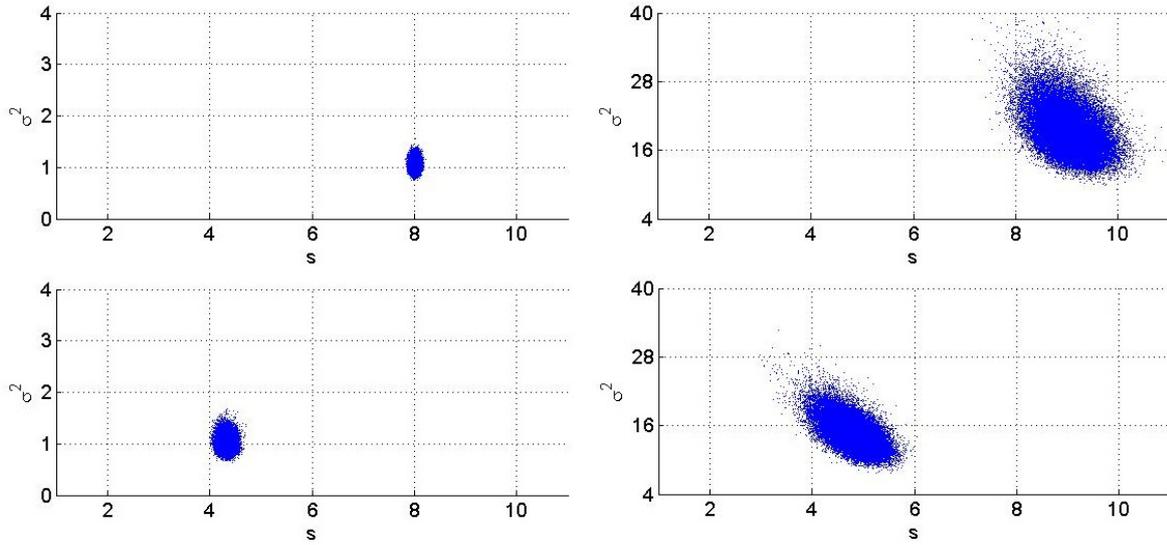
This figure provides the joint posteriors of half spread  $s$  and variance  $\sigma^2$ , given  $\lambda$  for four simulated series of length 300 with probabilities  $p(q = +\lambda) = 0.25$ ,  $p(q = +1) = 0.25$ ,  $p(q = -1) = 0.25$ ,  $p(q = -\lambda) = 0.25$ : Top-left panel:  $s = 8$ ,  $\sigma^2 = 1$ ,  $\lambda = 0.25$ ; Top-right panel:  $s = 8$ ,  $\sigma^2 = 16$ ,  $\lambda = 0.25$ ; Bottom-left panel:  $s = 4$ ,  $\sigma^2 = 1$ ,  $\lambda = 0.5$ ; Bottom-right panel:  $s = 4$ ,  $\sigma^2 = 16$ ,  $\lambda = 0.5$ .



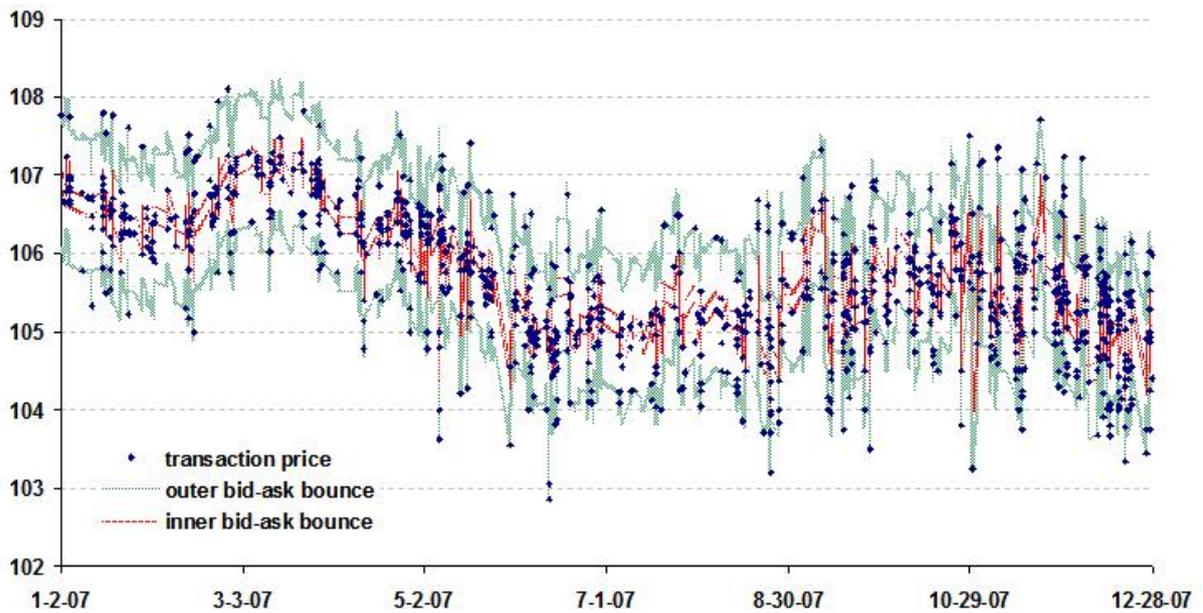
**Figure 8:** Joint posteriors of the half spread  $s$  and the variance  $\sigma^2$  given  $\lambda$  for four simulated series of length 300 with probabilities  $p(q = +\lambda) = 0.40$ ,  $p(q = +1) = 0.10$ ,  $p(q = -1) = 0.10$ ,  $p(q = -\lambda) = 0.40$ : Top-left panel:  $s = 8$ ,  $\sigma^2 = 1$ ,  $\lambda = 0.25$ ; Top-right panel:  $s = 8$ ,  $\sigma^2 = 16$ ,  $\lambda = 0.25$ ; Bottom-left panel:  $s = 4$ ,  $\sigma^2 = 1$ ,  $\lambda = 0.5$ ; Bottom-right panel:  $s = 4$ ,  $\sigma^2 = 16$ ,  $\lambda = 0.5$ .

**Table 2:** Description and Summary Statistics

CUSIP ID	172967AZ4
Company name	Citigroup Inc.
Data source	TRACE
Data beginning	02/01/2007
Data ending	28/12/2007
# trading days in samples	197
Total # of trades	1339
Price units	US Dollars
Average price	105.72
Standard deviation	1.00
Maximum price	108.106
Minimum price	102.85
Average daily trades	6.80
Average time between trades	06:27:35



**Figure 9:** Joint posteriors of half spread  $s$  and variance  $\sigma^2$  given  $\lambda$  for four simulated series of length 300 with probabilities  $p(q = +\lambda) = 0.10$ ,  $p(q = +1) = 0.40$ ,  $p(q = -1) = 0.40$ ,  $p(q = -\lambda) = 0.10$ : Top-left panel:  $s = 8$ ,  $\sigma^2 = 1$ ,  $\lambda = 0.25$ ; Top-right panel:  $s = 8$ ,  $\sigma^2 = 16$ ,  $\lambda = 0.25$ ; Bottom-left panel:  $s = 4$ ,  $\sigma^2 = 1$ ,  $\lambda = 0.5$ ; Bottom-right panel:  $s = 4$ ,  $\sigma^2 = 16$ ,  $\lambda = 0.5$ .



**Figure 10:** Time series plot of transaction prices with estimated bid-ask bounces

**Table 3:** Summary of estimates for bond 172967AZ4.

Parameter	MM, Roll	Post.mean, Roll	Post.mean, Extended
$s$	0.4062	0.3302(0.0225)	0.8428(0.0227)
$\sigma^2$	0.1982	0.3231(0.0235)	0.0898(0.0066)
$\lambda$		1	0.13
$\lambda s$		0.3299	0.1096
Log marginal likelihood		-1182.8434(0.0286)	-326.5221(0.0333)

**Table 4:** Serial correlation of  $\Delta\hat{m}$ .

Lag	Roll	Extended
1	-0.1646	-0.0623
2	-0.1186	-0.0331
3	-0.1323	-0.0893
4	-0.0786	-0.0814
5	0.0132	-0.0689
6	0.0295	0.0059
7	-0.0153	-0.013
8	0.0074	-0.0424
9	0.0046	-0.0311
10	-0.0056	0.0269

## References

- [1] C.A. Ball and T. Chordia, *True spreads and equilibrium prices*. The Journal of Finance, 56: 1801–1835,2001.
- [2] M.J. Barclay and J.B. Warner, *Stealth trading and volatility: Which trades moves prices?* The Journal of Financial Economics, 34(3): 282–305,1993.
- [3] M.j. Brennan and A. Subrahmanyam, *Market microstructure and asset pricing: on the compensation for illiquidity in stock returns*. The Journal of Financial Economics, 41(3): 441–464,1996.
- [4] J.Y. Campbell, A.W. Lo and A.C. Mackinlay, *The Econometrics of Financial Markets*. Princeton University Press, Princeton, NJ, 1997.
- [5] S. Chib, *Marginal likelihood from the Gibbs output*. The Journal of the American Statistical Association, 90(432): 1313–1321,1995.
- [6] S. Chib and E. Greenberg, *Markov Chain Monte Carlo Simulation Methods in Econometrics*. The Econometric Theory, 12: 409–431,1996.
- [7] J.Y. Choi, D. Salandro and K. Shastri, *On the estimation of bid-ask spread: Theory and evidence*. The Journal of Financial and Quantitative Analysis, 23: 219–229,1988.
- [8] J.H. Cochrane, *Asset Pricing*. Princeton University Press, Princeton, NJ, 2009.
- [9] P.J. Deschamps, *Bayesian estimation of an extended local scale stochastic volatility model*. The Journal of econometrics, 162: 369–382,2011.
- [10] D. Easley and M. O’Hara, *Price, Trade size, and information in securities markets*. The Journal of Financial Economics, 19(1): 69–90,1987.

- [11] L.R. Glosten and L.E. Harris, *Estimating the components of the bid-ask spread*. The Journal of Financial Economics, 21(1): 123–142,1988.
- [12] L.R. Glosten and P. Milgrom, *Bid, ask, and transaction prices in a specialist market with heterogeneously informed traders*. The Journal of Financial Economics, 14:71–100,1985.
- [13] T.J. George, G. Kaul and M. Nimalendran, *Estimating components of the bid-ask spread and its components: A new approach*. The Review of Financial Studies, 4: 622–656,1991.
- [14] L.E. Harris, *Statistical Properties of the Roll Serial Covariance Bid-Ask Spread Estimator*. The Journal of Finance, 45(2): 579–590,1990.
- [15] L.E. Harris, *Trading and Exchanges: Market microstructure for practitioners*. Oxford University Press, New York, 2003.
- [16] J. Hasbrouck, *Trades, quotes, inventories, and information*. The Journal of Financial Economics, 22(2): 229–252,1988.
- [17] J. Hasbrouck, *Liquidity in the futures pits: Inferring market dynamics from incomplete data*. The Journal of Financial and Quantitative Analysis, 39: 305–326,2004.
- [18] J. Hasbrouck, *Empirical Market Microstructure*. Oxford University Press, New York, 2007.
- [19] J. Hasbrouck, *Trading costs and returns for U.S. equities: Estimating effective costs from daily data*. The Journal of Finance, LXIV(2): 1445–1477,2009.
- [20] R.D. Huang and H.R. Stoll, *The components of the bid-ask spread: a general approach*. The Review of Financial Studies, 10(4): 955–1034,1997.

- [21] A. Madhavan, *Market microstructure: A survey*. The Journal of Financial Markets, 3(3): 208–258,2000.
- [22] A. Madhavan, M. Richardson and M. Roomans, *Why do security prices fluctuate? A transaction-level analysis of NYSE stocks*. The Review of Financial Studies, 10: 1035–1064,1997.
- [23] J. Osiewalski and M. Pipien, *Bayesian comparison of bivariate ARCH-type models for the main exchange rates in Poland*. The Journal of econometrics, 123: 371–391,2004.
- [24] R. Roll, *A simple implicit measure of the effective bid-ask spread in an efficient market*. The Journal of Finance, 39(4): 1127–1139,1984.
- [25] T.H. Rydberg and N. Shephard, *Dynamics of trade-by-trade price movements: Decomposition and models*. The Journal of Financial econometrics, 1(1): 2–25, 2003.
- [26] P. Schultz, *Corporate bond trading costs: A peek behind the curtain*. The Journal of Finance, 56(2): 677–698,2001.
- [27] H.R. Stoll, *Inferring the components of the bid-ask spread: Theory and empirical tests*. The Journal of Finance, 44(1): 115–134,1989.