How Well do Inventory Holding Components of Bid-Ask Spread Measure Inventory Holding Premium?

Birsel Tavukcu Pirim

The University of Mississippi

Current Version: August 28, 2005

Contact Information:

Birsel Tavukcu Pirim
Hwy 314 Apt # 179 C
Oxford, MS 38655
Phone # (662) 801 4630
btavukcu@bus.olemiss.edu
Essay I: How Well do Inventory Holding Components of Bid-Ask Spread Measure Inventory Holding Premium?

Abstract

Previous studies show that the quoted bid-ask spread generally has three components: order processing, adverse selection, and inventory-holding. In this essay, we focus on inventory holding cost component of spread. The inventory holding component of bid-ask spread compensates the dealer for holding a less diversified portfolio. Currently, we know little about how the models measure inventory holding costs and perform relative to each other. The empirical results of inventory holding component models display considerable variation in their estimates of inventory costs. If inventory holding models are to be more than just a theoretical abstract and have a use in empirical testing, such an analysis is essential. In this paper, we test the performance of four commonly used methods of computing the inventory holding premium. To determine the usefulness of the inventory holding models in measuring inventory costs, various proxies for the presence of inventory costs are examined. These variables include the number of shareholders, trading frequency, volatility, dollar trading volume, and options listings. As a benchmark for the analysis, we use Bessembinder’s (2002) and Chordia and Subramanyam’s (2002) estimates for order imbalances.

Key Words: Inventory Holding Cost, Bid-Ask Spread Components, Order Imbalances, Inventory Holding Models,
1. Introduction

Previous studies show that the quoted bid-ask spread generally has three components: order processing, adverse selection (or asymmetric information), and inventory-holding\(^1\). Order processing costs represent a fee charged by market makers for standing ready to match buy and sell orders. Adverse selection costs represent a reward to market makers for taking on the risk of dealing with traders who may possess superior information. The inventory holding component of bid-ask spread is compensation for a dealer holding a less diversified portfolio. In this study we focus on the inventory holding cost component of spread. Order-flow imbalances give rise to this component of the bid-ask spread (Stoll, 1978; Ho and Stoll, 1981). The process of equilibrating order imbalances may cause the market maker’s inventory position to deviate from optimal levels. When the deviation greater, the inventory holding costs are larger and the bid-ask spread is wider, ceteris paribus.

Bid-ask spread decompositions provide guidance on evaluating the merits of market structures and the fairness of market maker rents (Bollen et al. (2004)). Bid-ask spreads are also used to understand the role played by market type in determining the relative magnitudes of each cost component (Affleck-Graves, Hedge, and Miller, 1994 and Porter and Weaver, 1996). By investigating the components of the bid-ask spread, we gain further insight into variables underlying the costs of liquidity, the cost of capital, and the value of the firm. This line of research is useful to a wide range of market participants, including corporate managers, traders, securities exchanges, and securities regulators. While researchers develop models to measure the components of the bid-ask

spread, we know little about how these models measure inventory holding costs and how these models perform relative to each other. For instance, Stoll (1989) using a sample of NASDAQ stocks, finds that inventory holding costs account for 10% of the spread. Huang and Stoll (1997) using a sample of 20 of the largest and most active NYSE stocks, find inventory costs represent 28.7% of the spread. A recent study using NYSE stocks by Bollen, Smith, and Whaley (2004) finds that the inventory holding cost is the largest component of spread. They estimate inventory holding costs of 29.28%, 32.35%, and 44.69% of the spread for March 1996, April 1998, and December 2001, respectively. The empirical results show considerable variation. Currently, we know little about how the models measure inventory holding costs and perform relative to each other. If inventory holding models are to be more than just a theoretical abstract and have a use in empirical testing, such an analysis is essential. In this paper, we test the performance of four commonly used methods of computing the inventory holding premium. To determine the usefulness of the inventory holding models in measuring inventory costs, various proxies for the presence of inventory costs are used. These variables are the number of shareholders, trading frequency, volatility, dollar trading volume, and options listings. As a benchmark for the analysis, we use Bessembinder’s (2002) and Chordia and Subramanyam’s (2002) estimates for order imbalances.

Chordia, Roll, and Subrahmanyam (2001) state that prices and liquidity should be strongly affected by extreme order imbalances, regardless of volume, for two reasons. First, order imbalances sometime signal private information, which may reduce liquidity, at least temporarily, and also may move the market price permanently, as also suggested by Kyle’s (1985) theory of price formation. Second, even a random large order
imbalance exacerbates the inventory problem faced by the market maker, who can be expected to respond by changing the bid-ask spread. Hence, order imbalances are one of the important factors on market maker’s inventory holding premium. In fact, the inventory models of Stoll (1978a) and Ho and Stoll (1981, 1983) suggest that the market makers’ quote-placement strategies, accommodating buying and selling by outside investors, are influenced by inventory concerns.

We estimate order imbalances using the techniques developed by Bessembinder (2002) and Chordia, Roll, and Subrahmanyam (2001). Bessembinder (2002) constructs relative order imbalance for each market based on the accumulated difference between customer buy and customer sell trades since the open on a particular market. He signs trades as customer buys or sales using the Ellis, Michaely, and O’Hara (2000) algorithm. On the other hand, Chordia, Roll, and Subrahmanyam (2001) sign trades using the Lee and Ready (1991) algorithm. Then, they calculate their measure of the daily order imbalance in each stock using both the number of buys and sells, as well the quantity bought and sold.

The paper proceeds as follows: Section II discusses the background on inventory holdings costs. Section III describes the inventory holding components models used in this paper, while Section IV lays out the variables related to inventory-holding costs. Section V discusses the data and methodology. Section VI presents the results and analysis, and Section VI concludes.

---

2 He also uses the Lee and Ready (1991) algorithm, with no significant deviations in reported results.
2. Background

In inventory based models where the market maker is risk averse, holding inventory introduces risks for the market maker and his pricing strategy, at least partially, reflects his efforts to minimize this risk. There are two aspects to inventory risk: the risk of being unable to trade the stock and the risk that prices will adversely change while long in the stock. Amihud and Mendelson (1980) and Ho and Stoll (1980) state that the higher the risk of being unable to trade the stock, the more difficult it is for the market maker to return to his desired inventory level.

The component models that estimate an inventory cost rely on the assumption that the uncertainties in order flow can result in inventory problems for the market makers. Since order flow is unpredictability, demand and supply are not always balanced. Market makers must often carry inventory in the course of supplying liquidity and, hence, bear risk. The size of the spread, therefore, includes compensation for bearing this risk.

Ho and Macris (1985) state that multiple dealer markets are characterized by higher aggregate inventory levels, which increase in proportion to the number of market makers. However, Affleck-Graves, Hedge, and Miller (1994) state that multiple dealer markets have several advantages over specialist markets with regard to the ability to manage inventory risk. First, Ho and Macris (1985) and Vijd (1990) show that multiple dealer markets have to absorb large imbalances in order flow than specialist markets because of the large inventory dealer’s carry collectively. Second, Ho and Stoll (1983) propose that the potential for interdealer trading reduces the inventory risk exposure of competitive dealers by allowing for quick reallocation of inventory imbalances across market makers. Third, Grossman and Miller (1988) argue that, since dealers assume the
price risk of traders who place market orders, the price risk borne by any single dealer can be reduced by spreading (or diversifying) this risk across several market makers. Affleck-Graves et al. (1994) suggest that other things equal, the larger the group of dealers, the greater the potential for diversification gains and the lower the expected inventory cost per unit traded by a market maker. These arguments imply that it is more expensive for the specialist to absorb a given amount of order imbalances than for a group of competing market makers with heterogeneously distributed inventory positions.

Ho and Stoll (1983) suggest that market-makers’ quote placements strategies are affected by accumulated inventory. Naik and Yadav (2002) state, that if the inventory of dealer goes above (below) her target inventory level, she will potentially change her quotes and/or offer better prices during negotiation to attract order flow in a direction that will bring her inventory toward her target level. When a dealer has an extreme inventory position, she can post competitive quotes on one side and have a better chance of executing order flow in that direction, resulting in relatively rapid reduction in the inventory imbalance. On the other hand, when a dealer’s inventory is close to the target inventory, she is less likely to post competitive quotes and therefore has a reduced chance of executing public order flow, resulting in a relatively slow reduction in the inventory imbalance. An excess of customer buy (sell) orders can lead to a reduction (increase) in the market-makers inventory.

Chordia, Roll, and Subrahmanyam (2001) and Bessembinder (2002) suggest that if market makers perceive that competitive quotations will attract orders, then reductions (increases) in inventory should lead to the posting of more aggressive quotations at the bid (ask) to attract customer sell (buy) orders and restore inventory. Bessembinder finds
a statistically significant relationship between inventory effects and quote placement for
US equities. He states that the evidence supports inventory control as a determinant of
quote placement both on and off the NYSE, and the estimated effects are substantially
stronger for the NYSE.

3. Inventory Holding Components Models

Several techniques are developed to estimate the components of the bid–ask spread. In one group of models, inferences about the bid-ask spread is made from the
serial covariance of the time series of transaction prices (George, Kaul, and Nimalendran,
1991; Roll, 1984; and Stoll, 1989). In the second group, the components are inferred by
relating the change in price to transaction size and to direction of the trade (buyer or
seller initiated) (Glosten and Harris, 1988; Hasbrouck, 1988, 1991; Huang and Stoll,
1997; and Madhavan, Richardson, and Roomans, 1997).

Empirical estimates of the components of the spread vary considerably across
models. For instance, Stoll (1989), using a sample of NASDAQ stocks, finds that order
processing costs account for 47% of the spread, asymmetric information costs account for
43% of the spread, with the remaining 10% attributable to inventory costs. Huang and
Stoll (1997) report similar results for a sample of 20 of the largest and most active NYSE
stocks. They estimate asymmetric information costs of 9.6% of the spread and order
processing costs and inventory costs of 61.8% and 28.7% of the spread, respectively.
Madhavan and Smidt (1991, 1993) initiate a line of research examining specialist
behavior. Using NYSE specialist inventory positions and their quotes, they find that

---

3 Hansch, Naik, and Viswanathan (1998) also document that inventory affects the quote behavior of
London Stock Exchange dealers.
specialist inventories are slow to adjust to mean levels. They estimate that it takes over 49 trading days for an inventory imbalance to be reduced by 50%. Unlike previous studies, Bollen et al. (2004) find that the inventory holding cost is the largest component of spread. They estimate inventory holding costs of 29.28%, 32.35%, and 44.69% of the spread for March 1996, April 1998, and December 2001, respectively, and order processing costs of 8.92%, 15.79%, and 15.27% of the spread for each sub-sample.

We calculate the inventory holding component of the spread for each stock in our sample using the estimation procedures of Stoll (1989), Madhavan and Smidt (1991), Huang and Stoll (1997), and Bollen, Smith, and Whaley (2004). For all models, we initially compute the inventory holding component as a percentage of spread. To compute the inventory holding cost of transactions, we express the dollar inventory holding component as a percentage of the stock price. Thus, we control for stock price and measure the inventory holding cost of trading for a given dollar.

A. The Stoll (1989) Model (SM)

Stoll (1989) decomposes the spread into three components: order processing, inventory holding, and adverse selection. Under the assumption of market efficiency, the serial covariance of price changes due to the spread may be inferred from observed price changes. The serial covariance depends on the two-period (three-date) sequence of prices.

Stoll calculates the three spread components from slope coefficients of regressions of the serial covariance of the percentage price change series on the bid-ask spread:
\[
\text{cov}_T = a_0 + a_1 S^2 + u \quad (1)
\]
\[
\text{cov}_Q = b_0 + b_1 S^2 + v \quad (2)
\]

where \( S \) denotes the quoted proportional spread (i.e., the difference between the ask and bid quotes divided by the average of these quotes); \( \text{cov}_T \) is the serial covariance of transaction price changes; \( \text{cov}_Q \) is the serial covariance of changes in bid (or ask) quotes; and \( u \) and \( v \) are random error terms.

Given \( a_1 \) and \( b_1 \), Stoll solves for intermediate values, \( \partial \), the size of a price continuation as a fraction of the spread, \( \pi \), the probability of a price reversal, and \((1-\pi)\), the probability of a price continuation, from two auxiliary equations\(^4\):

\[
a_i = \partial^2 (1 - 2\pi) - \pi^2 (1 - 2\partial) \quad (3)
\]
and

\[
b_i = \partial^2 (1 - 2\pi) \quad (4)
\]

Components of bid-ask spreads are then determined as follows:

\text{Adverse Selection Cost} = \left[ 1 - 2(\pi - \partial) \right] \quad (5)

\text{Inventory Holding Cost} = 2(\pi - 0.5) \quad (6)

\text{Order Processing Cost} = (1 - 2\partial) \quad (7)

---

\(^4\) If a transaction at the bid is followed by another transaction at the bid (a continuation), the price change is \(-\partial S\), where \(0 < \partial < 1\). If a transaction at the bid is followed by transaction at the ask, the price change is \((1 - \partial)S\), a reversal.
B. The Madhavan and Smidt (1991) Model (MS)

The Madhavan and Smidt (1991) model uses a combination of inventory and information to estimate spread components. They assume that the past value of an asset makes no contribution to the current formation of beliefs about underlying value. In other words, they assume that, on arrival into a new transaction period, agents forget all past information shocks and use the current public information as their prior belief about the ‘true’ value of the asset being traded. Their data compromises NYSE specialist inventory positions and their quotes. They find that specialist inventories are slow to adjust to mean levels.

There are two key equations. First, the pricing equation is consistent with inventory models. The price set by Dealer $i$ ($P_{it}$) is linearly related to the dealer’s conditional expectation about the true value ($\mu_{it}$) and current inventory measured at the beginning of the period ($I_{it}$):

$$P_{it} = \mu_{it} + \alpha (I_{it} - I^*_i) + \gamma D_i$$  \hspace{1cm} (8)

where $I^*_i$ is Dealer $i$’s desired inventory position. The inventory response effect ($\alpha$) is negative to capture “quote shading.” The $D_i$ term is a direction dummy that takes the value of 1 if Dealer $i$ sells (trades at the ask) and -1 if Dealer $i$ buys (trades at the bid). $\gamma D_i$ is the order processing costs such as labor and equipment costs and rents.

The second key equation is the demand ($Q_{jt}$) of the informed Dealer $j$, who has exponential utility over terminal wealth. The demand equation enables the market maker to extract information from Dealer $j$’s trade using Bayes rule, hence private information effects enter through the conditional expectation term $\mu_{it}$ in Eq. (8).
Dealer i’s price schedule is a function of his conditional expectations, $\mu_i$, at the time of quoting. The expectation is based on a public signal, and the noisy signal that he can extract from Dealer j’s trading behavior.

After allowing for conditional expectation and unobservable priors the empirical model is as follows:

$$
\Delta P_i = \left( \frac{\alpha}{\phi} - \alpha \right) I^*_i + \left( \frac{1 - \phi}{\phi^*} \right) \Theta_{jt} - \left( \frac{\alpha}{\phi} \right) I_t + \frac{\alpha}{\phi} D_{t-1} + \left( \frac{\gamma}{\phi} \right) D_{t-1} + \varepsilon_t
$$

where $\Delta P_i$ is the price change between two incoming trades. The coefficient $\beta_1$ and $\beta_3$ measure the information effect and inventory effect, respectively, while $\beta_5$ measures order processing costs and rents. The regression constant, $\beta_0$, is close to zero if desired inventory $I^*_i$ is close to zero. The model predicts that:

$$
\{\beta_1, \beta_3, \beta_4\} > 0; \{\beta_2, \beta_5\} < 0; |\beta_2| > \beta_3; |\beta_4| > |\beta_5|
$$

### C. The Huang and Stoll (1997) model (HS)

The Huang and Stoll (1997) model estimates spread components using two-way and three-way decomposition approaches. They develop a basic trade indicator model that provides a flexible framework for examining a variety of microstructure issues. A distinguishing characteristic of trade indicator models is that these models are driven solely by the direction of trades. Huang and Stoll’s two-way decomposition of the spread does not distinguish between adverse selection and inventory holding costs, the model separates order processing costs from other sources of the spread, and it provides
estimates of the spread. The three-way decomposition of the spread is based on serial
correlation in trade flows and distinguishes inventory and adverse information effects.

Huang and Stoll (1997) model serial correlation in trade flows given by the
conditional expectation of the trade indicator at time t-1, given Q_{t-2}, as

$$E(Q_{t-1}|Q_{t-2}) = (1 - 2\pi)Q_{t-2}$$  \hspace{1cm} (10)

where \(\pi\) is the probability that the trade at t is opposite in sign to the trade at t-1. \(\pi\) is
different from one-half, and the market knows equation above, the change in fundamental
value will be given by

$$\Delta V_t = \alpha \frac{S}{2} Q_{t-1} - \alpha \frac{S}{2} (1 - 2\pi) Q_{t-2} + \epsilon_t,$$

where the \(S_t\) is the posted spread just prior to the transaction, \(\alpha \frac{S}{2} Q_{t-1}\) reflects the private
information revealed by the last trade, as in Copeland and Galai (1983) and Glosten and
Milgrom (1985). The public information component is captured by \(\epsilon_t\).

They estimate the components of the spread directly from the following equation:

$$\Delta M_t = (\alpha + \beta) \frac{S_{t-1}}{2} Q_{t-1} - \alpha \frac{S_{t-2}}{2} (1 - 2\pi) Q_{t-2} + \epsilon_t$$  \hspace{1cm} (15)

where \(Q_t\) is the buy-sell indicator for the transaction price, \(P_t\). \(M_t\) is the midpoint of the
quote that prevails just before the transaction at time t in above equation. \(\alpha\) and \(\beta\) are
the percentage of the half-spread attributable to adverse selection and inventory holding
costs, respectively. Since \(\alpha\) and \(\beta\) are stated as proportions, the order processing
component is equal to \((1 - \alpha - \beta)\).
D. The Bollen, Smith, and Whaley (2004) model (BSW)

The market maker’s spread needs to include a premium to cover expected inventory-holding costs, independent of whether the trade is initiated by an informed or an uninformed trader. In the BSW model, they assume that the length of time a stock is held in inventory is known and short. They also assume that the risk-free rate and the expected change in the true price of the stock are equal to zero for this short holding period.

In the absence of a viable hedging instrument, the market maker faces inventory–holding price risk for which he will demand compensation. They assume that the market maker wants to set his inventory holding premium (IHP) such that he minimizes the risk of losing money should the market move against him; that is

\[
\text{Min } \mathbb{E} \left[ \left( \Delta S + IHP \Delta S < 0 \right) \right]^2
\]

The risk measure (Equation 16) is called the lower semi-variance. Setting the first-order condition to zero shows that the minimum inventory-holding premium that the market maker is willing to charge is,

\[
IHP = - \mathbb{E} \left( \Delta S \right) \mathbb{P} \left( \Delta S < 0 \right)
\]

According the equation (17), the minimum IHP equals the expected loss on the trade conditional on an adverse stock price movement times the probability of an adverse stock price movement.

Suppose the market maker buys a call option. Using the Black and Scholes (1973) and Merton (1973) option valuation formulas, the expected inventory-holding premium can be written
\[ IHP = SN \left( \ln \left( \frac{S}{X} + 0.5\sigma\sqrt{t} \right) \right) - XN \left( \ln \left( \frac{S}{X} \right) - 0.5\sigma\sqrt{t} \right). \]  

(18)

where \( S \) is the true stock price at the time that the market maker opens his position, \( X \) is the exercise price of the option, \( \sigma \) is the standard deviation of security return, \( t \) is the time until the offsetting order, and \( N(.) \) is the cumulative unit normal density function.

The expected loss described by Equation (17) is an at-the-money option, so the valuation formula Equation (18) simplifies to

\[ IHP = S \left[ 2N(0.5\sigma)\sqrt{t} - 1 \right]. \]  

(19)

4. Variables

All inventory holding cost models assume that the market maker will prefer to sell if he is long in inventory and prefer to buy if he is short in inventory, whether or not the spread changes depends on the model used. Inventory-based models show that the spread is influenced by the initial wealth and risk preferences of the market maker and the variance and covariance of the stock. In this section, we identify the variables that are related to inventory-holding.

a) Volatility

Inventory explanations of spreads predict a positive relationship between spreads and volatility [Stoll (1978b) and Ho and Stoll (1981, 1983)]. Ho and Stoll (1981) demonstrate that the more volatile the stock price, the more the market maker is exposed to the risk of adverse price movements, and consequently the wider the bid-ask spread necessary to compensate the market maker. Hence, there is a positive correlation between return volatility and the spread. Tinic (1972) includes a direct measure of
volatility, the standard deviation of price, as a measure of inventory price risk. We include the standard deviation of the quote midpoint as a measure of intraday volatility to capture volatility in true price of the stock.

b) Volume

Inventory control models predict that as the liquidity of a stock increases, the compensation required by the market maker through the spread declines, resulting in a negative relationship between trading volume and spreads\(^5\). If trading volume is generally low, market makers will find it difficult to adjust their inventory levels and will increase their spreads to compensate. Stoll (1978b) uses dollar trading volume as a proxy for the length of the market maker’s holding period (one of the factors influencing inventory-holding costs). Stoll (1989) uses trading volume as a proxy to measure the risk of holding non-optimal inventory levels. Hence, we use trade size, daily dollar volume, and daily total volume which is average trade size times the number of trades as a proxy for inventory holding costs since volume, dollar or total, is one of the factors that influence inventory holding cost components of spread.

c) Option Listings

Option listings have beneficial impacts on the underlying market for several reasons. First, as suggested by Ross (1976) and Hakansson (1982), options can improve the efficiency of incomplete asset markets by expanding the opportunity set facing investors. Second, as suggested by Bollen et al (2004), option listings reduce spreads and improve liquidity in the underlying market by reducing the inventory costs of the market maker since options provide a mechanism for hedging their inventory positions. Since

\(^{5}\) Kyle (1985) notes that the term market liquidity encompasses a number of transactional properties of markets use trading volume as a measure of liquidity.
there is a relationship between option listing and inventory holding costs, this variable is included.

d) Trade Frequency

Theoretical and empirical evidence suggests that order flow affects transactions costs by changing the dealer’s cost of holding inventory, and by providing signals about security value. Lin, Sanger, and Booth (1995) suggest that inventory of order flow may be influenced by the dimensions of order flow, such as the frequency of trades, trade size, and the timing of trades.

Demsetz (1968) argues that price concessions charged for immediate transactions will be inversely related to the particular security’s trading activity. In an actively traded market the dealer can correct sub-optimal inventory positions with greater speed. Demsetz uses trade frequency as proxy for inventory-holdings costs. For frequently traded stocks the market maker is certain that he can balance the orders in the long term. For thinly traded stocks the market maker is less certain, therefore quoting involves more risk and the spreads quoted on thinly traded stocks will be larger. Hence, we use the number of trades as a proxy for inventory holding.

e) Number of shareholders

Demsetz (1968) uses the number of shareholders as a proxy for inventory-holdings costs. He argues that the number of shareholders is direct proxy for the transaction rate. The higher the transaction rate, the lower the cost of waiting and hence the lower the bid-ask spread. Similar to Demsetz, Benston and Hargeman (1974) use the number of shareholders as a proxy for the long-term trading activity of the stock. Hence, we use number of shareholders as a proxy for inventory-holding cost.
References


Chordia, Roll, and Subrahmanyam (2001, working paper), Order Imbalance, Liquidity, and Market Returns, Emory University, and UCLA.

Chordia, T. and A., Subrahmanyam, (2002 working paper), Order Imbalance and Individual Stock Return, Emory University and UCLA.


