The Impact of Security Concentration on Adverse Selection Costs and Liquidity: An Examination of Exchange Traded Funds

Kenneth Small[1]
Loyola College
Sellinger School of Business and Management
Baltimore, MD 21210
410-617-5210
ksmall@loyola.edu

James Wansley
Department of Finance
College of Business Administration
Knoxville, TN 37996
865-974-3216
jwansley@utk.edu

Abstract:
We examine the determinants of liquidity and adverse selection costs in a sample of basket securities. Using Exchange Traded Funds (ETFs), we find evidence that adverse selection costs are decreasing in the number of equities held in the underlying portfolio, but adverse selection costs do not increase as the concentration among the securities increases. We find no evidence that industry concentration increases basket security adverse selection costs or reduces liquidity. We also document significantly lower levels of adverse selection costs in ETFs versus a matched sample of equities. In addition, ETFs have quoted dollar depth that is 35 times larger than in a matched sample of equities, but ETFs also have higher effective and quoted spreads. However, when considering spreads and depth in a single metric, ETFs have significantly higher levels of liquidity.

Working Draft
Please Do Not Quote or Disseminate Without Permission

Current Draft: March 2006

[1] Corresponding author. We would like to thank Gary Bauer, Carry Collins, Philip Daves, John McDermott, Michael McKee, Stanly Gyoshev, and Laura Starks for helpful comments. In addition, we thank seminar participants at the 2004 Southern Finance Association, 2004 Eastern Finance Association annual conferences, and the 2004 Loyola College Research Symposium. We also thank participants at the 2004 Financial Management's Dissertation Consortium in Zurich, Switzerland for their valuable comments. Any remaining errors are the authors.

The Impact of Security Concentration on Adverse Selection Costs and Liquidity: An Examination of Exchange Traded Funds

I. Introduction:

In this work we explore the relationship between adverse selection costs, as a percent of the bid-ask spread, and liquidity for a sample of basket securities, namely Exchange Traded Funds, and we compare these relationships to a matched sample of equities. Prior theoretical work, especially that of Gorton and Pennacchi (1993), Kumar and Seppi (1994) and Subrahmanyam (1991) predict lower adverse selection costs in basket securities relative to individual equities.

While we present the first test comparing adverse selection costs for ETFs and equities, empirical tests comparing mutual funds and equities have generally borne out these predictions, although the differences in adverse selection component of the spread have typically been smaller than expected. Neal and Wheatley (1998) estimate the adverse selection component of the bid-ask spread for 17 mutual funds and a control sample of 17 common stocks, they find the Glosten and Harris (1988) model estimates averaged 19% for the funds and 34% for the control stocks. Neal and Wheatley find that estimates from the George, Kaul, and Nimalendran (1991) model average 52% and 65% for the mutual funds and control stocks, respectively. The small difference between estimates of the adverse selection component of the spread for equities and closed end mutual funds present a problem for Neal and Wheatley. They state: "Adverse selection arises primarily from factors *other* than a firm's current liquidation value" (p.123), and they also suggest that the adverse selection spread decomposition models may be misspecified. We test whether their findings may also be related to common factors in the portfolio of underlying securities, and we find evidence that adverse selection costs are decreasing in the number of equities held in the underlying ETF but adverse selection costs do not increase as the concentration (using a Herfindahl index) among the securities increases.

We estimate measures of liquidity and the adverse selection component of the bid-ask spread and test for determinants of the spread component for exchange traded funds. Given prior theoretical work and empirical findings comparing adverse selection components of mutual funds to individual equities, we expect to find lower adverse selection costs for ETFs than for matched equities. Following Gorton and Pennacchi (1993), one possible reason for these differences is that informed agents may prefer to trade industry concentrated basket securities to avoid detection by regulatory agents or uninformed traders who may monitor their trading activities in the underlying securities?

Finally, we explore the determinants of liquidity and the adverse selection component of the spread, focusing on differences in portfolio construction and concentration. Why do some basket securities rank as the most traded instruments in the U.S. market (e.g., QQQQ) and some are in the lowest quartile of trading volume (e.g., MTK), even though they may be concentrated in the same industry or hold the common securities[2]? Some ETFs hold as few as 11 securities and some hold over 2000. Does the addition of securities diversify away adverse selections costs?

---

[2] QQQQ is the ticker for the Nasdaq-100 Index Tracking Stock represents ownership in the Nasdaq-100 Trust, a unit investment trust established to accumulate and hold a portfolio of the equity securities that comprise the Nasdaq-100 Index; MTK is the ticker for the Morgan Stanley technology ETF. These two ETFs have considerable overlap in their holdings.

To explore these issues, we examine the adverse selection costs and liquidity of ETFs versus a matched sample of equity securities. We also explore factors that contribute to basket security liquidity and adverse selection costs. Several studies have examined adverse selection costs in closed end mutual funds (Chen, Jiang, Kim and McInish (2003), Clark and Shastri (2001), and Neal and Wheatley (1998)), but we focus on ETFs because of their unique structure. ETFs trade intra-day and earn returns that are very similar to those of their underlying portfolio of securities. Unlike many closed end mutual funds, which often trade at a discount or premium to net asset value, exchange traded funds are easily created and redeemed. This process reduces the difference between the price of the ETF and its net asset value. The elimination of the premium or discount also reduces investor uncertainty regarding the future value of the security. By focusing on ETFs we remove any noise that premiums and discounts introduced in previous studies.

Our results indicate that exchange traded funds have significantly lower adverse selection costs than a matched set of equities, regardless of the model used to estimate adverse selection costs. We find Lin, Sanger, and Booth (1995) ETF adverse selection costs, as a percent of the bid-ask spread, to be 19.7% for ETFs and 34.3% for a matched sample of equities; these percentages are 29.6% and 72.6% using the George, Kaul, and Nimalendran (1991) model, and they are 18.1% and 44.1%,. using Glosten and Harris (1988) In a multivariate framework, ETFs also have significantly lower adverse selection costs than do the sample of equities.

We also document significantly higher levels of quoted dollar depth for ETFs compared with matched equities. Actually, ETFs have quoted dollar depth that is 35 times as large as the quoted depth for the sample of equity securities. However, ETFs also have higher effective and quoted spreads than the matched sample of equities. When considering spreads and depth in a unified framework, ETF liquidity is significantly greater than that of a matched sample of securities.

An extended liquidity and adverse selection analysis indicates that sector concentrated ETFs do not exhibit lower levels of adverse selection costs or decreased liquidity. In addition, we find no evidence that the concentration among the equities held in the underlying portfolio of a basket security has an impact on adverse selection costs or liquidity. We do find evidence that the number of securities held in the basket has a significant impact on liquidity and adverse selection costs. As the number of securities held in the underlying portfolio increases, adverse selection costs decrease and liquidity increases.

The remainder of the work proceeds as follows: in Section II, we discuss the history, trading mechanics, and general characteristic of exchange traded funds. Section III contains a review of theories of basket security trading and explores past studies that have examined informed trading in closed end mutual funds. In Section IV, we discuss the methods employed to test these hypotheses, and in Section V, we discuss the results of the analysis. In the final section we conclude the work.

II. Exchange Traded Funds:

The popularity of Exchange Traded Funds (ETFs) has steadily increased since their introduction in the early 1990s. The first U.S. exchange-traded fund[3], was created as a result of action taken by Leland, O'Brien, Rubenstein Associates, who lobbied the SEC for the creation of an Standard and Poor 500 tracking instrument named the Index Trust

[3] The first exchange traded fund was listed on the Toronto Stock Exchange in 1989.

SuperUnit. The original SuperTrust was terminated in 1996, but The American Stock Exchange (AMEX) took advantage of the SuperTrust order to petition for, and receive, an SEC Order in 1992 to create a stand-alone Standard and Poor 500 index-based ETF. This unit is commonly known as the Standard and Poor Depository Receipt, or SPDR (Novakoff 2000). Some of the most popular ETFs track the Dow Jones Industrial Average index (Diamonds, DIA), the NASDAQ 100 index (Qubes, QQQ), and the Standard and Poor 500 index (SPDR, SPY)

As of January 2006 over 1800 ETFs are listed on the American Stock Exchange, and the Financial Research Corporation predicts that total assets held by ETFs will reach anywhere between $500 billion to $1 trillion by the year 2007. ETFs are popular investment vehicles because they offer investors continuous trading during exchange hours, low premium/discounts, tax efficiency, diversification benefits, and transparency. Unlike open-ended mutual funds, which are priced at the end of the day, ETFs trade continuously throughout the day. Annual expense ratios of exchange traded funds are often lower than those of mutual funds, because of decreased costs associated with marketing and distribution. Because index ETFs are passively managed, and on average produce lower levels of capital gains than actively managed funds, they offer an advantage to tax conscious investors.

A process of creation and redemption works to limit the deviation of ETF prices from their underlying net asset value. An ETF unit is created when an investor deposits the underlying securities, and a creation unit is issued. The average creation unit multiple is 50,000, and share creation units range from 25,000 to 600,000 (AMEX 2002) (**We need a cite for this.)** . The deviations of exchange traded funds are much smaller than those found in closed end mutual funds. While many closed end mutual funds trade at a discount or premium to net asset value, exchange traded funds are easily created and redeemed. The ability to create and redeem ETFs essentially eliminates the difference between the price of the ETF and its net asset value. For instance, on June 1, 2004, the average deviation of the Standard and Poor Depository Receipt's (SPDR) price from its NAV since its inception was .0006%. The average deviation was .0004% for the DIA and .0006% for the QQQQ. The average discount of all closed–end mutual funds on June 30, 2001 was 4.8% and the average discount on equity closed-end funds was 11.1% (Lipper 2001**).**

Recent exchange traded funds research has included the work of Boehmer and Boehmer (2003), who study the liquidity impact of the cross-listing of several exchange traded funds on the NYSE. Elton, Gruber, Comer, and Li (2002) examine deviations in the SPDRs returns from the returns of an index fund. Hasbrouck (2000) studies price discovery in the SPDR as well as several sector exchange traded funds. Barari, Lucey, and Voronkova (2005) examine both short-term and long-term co-movements between the G7 exchange traded funds. Poterba and Shoven (2002) examine the tax effects of exchange traded funds. Small (2005) examines the deviations in the prices of ETF prices from their net asset values. Hedge and McDermott (2004) examine changes in liquidity of the component stocks of the NASDAQ 100 and the Dow Diamonds upon the introduction of the tracking ETFs. Van Ness, Van Ness, and Warr (2005a) also study the impact of the introduction of the Dow Jones Industrial Index tracking ETF (Diamonds) on the market quality of the underlying securities. Lipson and Mortal (2003) study the impact of SPDRs introduction on the underlying securities. Small and Wansley (2005) examine the impact of the sector SPDR funds introductions on the underlying securities. The studies of Small and Wansley (2005), Van Ness, Van Ness and Warr (2005a), Lipson and Mortal (2003), and Hedge and McDemott (2004) all examine the migration of informed and uninformed agents around the introduction of basket securities.

When uninformed agents migrate to the basket securities, adverse selection costs increase in the underlying securities. However, little research has examined the characteristic of basket securities that make them the preferred trading venue of uninformed agents. Is it possible that some basket securities are the preferred trading venue of informed agents or decrease their desirability to uninformed trading agents? In the next section, we discuss the theoretical and empirical evidence surrounding this question.

III. Adverse Selection Costs and Basket Securities: Theory and Evidence

Theoretical models predict that basket security traders will incur lower adverse selection costs relative to trading in equities. Subrahmanyam (1991) provides a model that demonstrates how markets in basket securities can provide a preferred trading medium for uninformed liquidity traders. A positive benefit accrues to uninformed liquidity traders because security-specific components of adverse selection are diversified away in basket securities. Consequently, market makers are exposed to lower levels of informed trading and as a result, adverse selection is decreased in basket securities. Subrahmanyam’s theory holds that liquidity traders are allowed to realize their trades more efficiently by trading in basket securities because their losses to informed trading are reduced.

Further theoretical research by Gorton and Pennacchi (1993) examines liquidity trading and informed agents. They present a model where liquidity traders form initial portfolios with knowledge of their future participation in markets where informed traders are present. The presence of informed traders places the liquidity trader at a disadvantage, and the liquidity traders' utilities are increased with the introduction of baskets securities. Liquidity traders can effectively reduce their expected losses of trading with informed agents if they choose to trade in basket securities.

The empirical evidence regarding adverse selection costs in basket securities has been mixed. Neal and Wheatley (1998) estimate the adverse selection component of the bid-ask spread for 17 mutual funds and a control sample of 17 common stocks. They find only small difference between these estimates of the adverse selection for the sample of equities and closed-end mutual funds. We argue later that these similarities may be explained by common factors in the portfolio of underlying securities.

Chen et al. (2002) find evidence of decreased levels of adverse selection in closed-end mutual finds. Using a sample of funds listed on the NYSE between 1994 and 1999, they find adverse selection costs are significantly lower for the closed end mutual funds than for the control sample of equities. Clark and Shastri (2001) examine the effects of ownership structure, the expense ratio, portfolio turnover, and discount to net asset value on information asymmetry in closed-end mutual funds. They find block ownership significantly impacts adverse selection costs in closed end mutual funds, while the other factors are not significant.

While it has generally been accepted that adverse selection costs of basket securities are lower than those of individual equities, previous empirical research has provided conflicting evidence. We expand on prior work in the area of basket securities by estimating and comparing measures of liquidity, including spreads and depth, between ETFs and a matched set of individual equity securities. Based on prior theoretical work and empirical work with mutual funds, we expect adverse selection costs to be lower for the ETF and liquidity to be higher. Offsetting this is the likely migration of informed trading to sector-specific exchange-traded funds. Informed traders, who posses firm-specific material nonpublic information, may choose to migrate from individual securities to industry specific baskets, because the return characteristics of the basket could, in some cases, be very similar to those of the individual security and may allow trading on material nonpublic information without detection[4]. This would provide informed traders a preferred venue for trading on material non-public information. Also as a result of possible legal scrutiny of trading the underlying equities, informed agents may prefer to trade in assets that mask their intentions. Furthermore, informed agents may choose to trade in securities that have the lowest probability of reveling the non-public information.

As the number of securities in a security basket increases, the costs associated with adverse selection should be diversified away. However, on the other hand, the costs associated with reconstitution increases. These costs could result in a divergence of the price of the ETF and its underlying basket's NAV (i.e tracking error). In addition, the concentration of the securities held in the basket could have a significant impact on adverse selection costs and liquidity. For example, a security basket may hold 100

[4] For example, suppose there exists a "basket security" that is focused in the pharmaceutical industry. Informed traders with firm level information from this industry may prefer to trade in the "basket security" to avoid detection of a regulatory body or to avoid detection by uninformed agents who monitor their trading for information signals. Informed agent may prefer these securities because the industry focused baskets are likely to exhibit return characteristics that are similar to those of the underlying securities, especially when the basked holds a small number of securities in the same industry.

securities, but two could comprise 90% of the NAV of the underlying portfolio. The characteristics of a concentrated security would be much different than one that held the 100 securities in equal proportions.

To explore this, we develop a cross-sectional regression model in which the adverse selection component of the bid-ask spread is related to several control factors including price, volatility and volume and several test variables that include a concentration measure, the number of securities in the ETF as well as dummy variables for international, sector and an indicator for broad market coverage.

In the next section, we discuss the liquidity, adverse selection and security "concentration" measures used to test the central hypotheses.

IV. Methods and Data:

IVa. Liquidity

We employ four commonly used liquidity measures, and we develop a measure that incorporates two dimensions of liquidity in a single metric. We evaluate the following measures:

1) Quoted Spread (*Quoted*) = $Ask_{i,t} - Bid_{i,t}$
2) Effective Spread (*Effective*) $= 2\left| p_{i,t} - MP_{i,t} \right|$
3) Depth (Depth Shares) = Number of Shares at Ask Price + Number of Shares at Bid Price
4) Dollar Depth (*Dollardepth*) = Number of Shares at Ask Price * Ask Price + Number of Shares at Bid Price * Bid Price

5) *Effective/Depth*[5] = The average effective spread averaged across the trading year divided by average dollar depth averaged across the year, where dollar depth is scaled by 100,000. This ratio captures increases in the bid-ask spread (width of the market) while also capturing changes in quoted depth (depth of the market).

Where $Ask_{i,t}$, $Bid_{i,t}$, $p_{i,t}$, $MP_{i,t}$, , are the best ask price, best bid price, price, and quoted midpoint, respectively, of firm *i* at time *t*. As in Chiyachanyana et al. (2005), the quoted spreads are time weighted and the effective spreads are trade weighted. We time weight the quoted spreads by the number of seconds the quote is outstanding weighted by the trading time in each trading day. The effective spread is weighted by the size of the trade. This weighting is calculated by dividing the size of the trade by the total trade volume for the trading day. These are summed over that trading day and then averaged over all trading days in the year. All measure of liquidity are averaged for each day and then averaged across the year to produce one observation per ETF and matched equity security for 2003.

IVb. Adverse Selection:

Kyle (1985) suggests market markers increase the bid-ask spread when trading with informed agents. In cases where informed agents enter a market and market makers are unable to differentiate informed and uninformed agents, the costs of informed trading (adverse selection costs) are often pooled across trading agents. When trading with informed agents, market makers could choose to increase the bid-ask spread to offset

---

[5] Effective/Depth = $\overline{Effective\ Spread}_i / (\overline{Dollardepth_i} / 100{,}000))$. This measure is similar in spirit to the *DepSpr* measure used in Kumar, Sarin, and Shastri (1998). *DepSpr* is dollar depth divided by the bid-ask spread. The *Depspr* measure is also employed in Hegde and McDermott (2004) and we thank John McDermott for suggesting its use.

losses associated with their information disadvantage, and in this work, we employ three commonly used bid-ask spread decomposition methodologies to measure adverse selection costs.

First, we follow the method of Lin, Sanger and Booth (1995), which decomposes the bid-ask spread into order processing and adverse selection components. Second, we employ a variant of the decomposition method of George, Kaul, and Nimalendran (1991), which also decomposes the spread into adverse selection and order processing components. Third, we use the model of Glosten and Harris (1988), which decomposes the bid-ask spread into order-processing/inventory-holding component and an adverse selection component. Van Ness, Van Ness, and Warr (2001), in an analysis of several adverse selection models, find that the adverse selection estimates from the Lin, Sanger, and Booth (1995) and Glosten and Harris (1988) models are highly correlated with accepted external measures of asymmetric information. We discuss each model in more detail below.

The Lin, Sanger and Booth (1995) adverse selection and persistence parameters are estimated from the following equations:

(adjust all equations to correct size and right justify the equation numbers)

$$\begin{aligned} M_{t+1} - M_t &= \lambda Z_t + \varepsilon_{t+1} \\ Z_{t+1} &= \theta Z_t + \eta_{t+1} \\ Z_t &= P_t - M_t \end{aligned} \quad , \qquad (1)$$

where M_t is the quote midpoint at time *t*, P_t is the transaction price at time *t*, ε_{t+1} and η_{t+1} are random error terms. The Lin, Sanger and Booth (1995) model estimate is bounded between 0 and 1, and is the proportion of the effective spread that is attributed to adverse selection.

George, Kaul, and Nimalendran (1991) (GKN) define transactions returns as:

$$R_t = E_t + \pi(\frac{S_q}{2})(Q_t - Q_{t-1}) + (1-\pi)(\frac{S_q}{2})Q_t + U_t, \qquad (2)$$

where E_t is the expected return from time t-1 to t, Q_t takes the value of 1 when the transaction is a purchase and –1 when the transaction is a sale[6], and U_t are unobservable public information innovations. Van Ness, Van Ness and Warr (2005b) employ a parameterization of the GKN model that is similar to that of Neal and Wheatley (1998), which allows the quoted spread to vary with each observation. We follow Van Ness, Van Ness and Warr (2005b) when we specify the GKN model as:

$$2RD_t = \pi_0 + \pi_1 S_t(Q_t - Q_{t-1}) + \eta_t, \qquad (3)$$

where $RD_t = R_t^T - R_t^Q$. R_t^T are returns derived from transactions, R_t^Q are returns derived from mid-point quotes, S_t is the quoted percentage bid-ask spread, Q_t takes the value of 1 when the transaction is a purchase and –1 when the transaction is a sale, and $\eta_t = 2(E_t - E_T) + 2(U_t - U_T)$. π_1 represents the order processing component of the bid-ask spread, and (1-π_1) is the adverse selection component of the bid-ask spread.

The last bid-ask spread decomposition model that we employ is the Glosten and Harris (1988) model. Glosten and Harris specify the adverse selection, and inventory-holding/order-processing costs, as a linear function of transaction volume. Their model can be expressed as:

$$\Delta P_t = c_0 \Delta Q_t + c_1 \Delta Q_t V_t + z_0 Q_t + z_1 Q_t V_t + e_t, \qquad (4)$$

[6] As suggested in Bessembinder (2003), we use the Ellis, Michaely, and O'Hara (2000) method to assign trades as buys or sells. Trades are classified buys (sells) when the trade occurs at the ask (bid), and trades not occurring at the bid or ask prices are classified using a tick test. We use the information from one quote before the reported trade time to perform the tick test.

where Q_t takes the value of 1 when the transaction is a purchase and –1 when the transaction is a sale, V_t is volume traded at time *t*, and e_t captures public information innovations. As in Jiang and Kim (2005), we use the average transaction volume to estimate the adverse selection component of the bid-ask spread as:

$$\frac{2(z_0 + z_1\overline{V})}{2(c_0 + c_1\overline{V}) + 2(z_0 + z_1\overline{V})} \tag{5}$$

We estimate all adverse selection costs measures across all transactions in 2003. We report the raw percentages and the dollar cost estimates. The dollar cost estimates are calculated by multiplying percentage adverse selection cost estimates times the quoted spreads for the Glosten and Harris (1988) and George, Kaul, and Nimalendran (1991) models and the Lin, Sanger and Booth (1995) model estimates times the effective spread.

IVc. Measures of Basket Security Concentration:

In this section we discuss the control variables and "concentration" measures used in the analysis. First, we proxy for industry concentration with the binary variable *Sector*. This variable takes the value of one when the AMEX classifies the security as a sector fund (i.e., when the basket holds securities primarily from one sector), and zero otherwise. This variable captures the impact of industry concentration. We also code ETFs that hold diversified portfolios of underlying securities. *Broad* is assigned the value of one when the security basket is classified by the AMEX as broad based basket and zero otherwise. To control for the impact that international ETFs (i.e., ETFs that hold portfolios of international securities) have on adverse selection costs and liquidity, we include the binary variable *International*. The information asymmetry between U.S. investors and foreign firms is well documented (Small, Flaherty, and Ionici (2005), Jiang and Kim (2005), Bacidore and Sofianos (2002)). *International* takes the value of one when the AMEX classifies the basket security as an international ETF.

We employ two measures of basket security concentration among the securities held in the basket. We use the natural log of the number *Ln(Number*) of equities that comprise the basket security. As the number of securities held in the basket increases, we expect the adverse selections costs of the basket security to decrease. We also measure the concentration among the equities held in the security by calculating the Herfindahl Index of the concentration of the top five holdings in the security. We specify the *Herfindahl* measure as:

$$Herfindahl : (\sum_{i=1}^{5} (\frac{VS_i}{NAV} x100)^2)$$

where VS_i is the value of underlying security *i*, and NAV is the net asset value of the security. We take the natural log of the Herfindahl Index to create the variable *LN(Herfindahl).* As the Herfindahl Index value for the ETFs increases, we expect adverse selection costs to also increase.

Empirical research suggests that quoted bid-ask spreads tend to increase in price and volatility, and spreads tend to decrease as trading volume increases (Demsetz (1968), Tinic (1972), Benston and Hagerman (1974), and Hamilton (1978)). To control for the impact of securities prices, we include the variables *Ln(Price).* *Ln(Price)* is the natural log of the average end of day price of the security. To control for volume, we include the variables *Ln(Volume),* which is the natural log of the average daily volume of the security. We also control for the volatility of security returns by including the variable *Ln(STD),* which is the natural log of the standard deviation of daily returns estimated over the year.

IVd. Data:

To conduct our tests, we identify all equity ETFs listed in the U.S. in 2003.[7] For this set of exchange traded funds and a matched sample of equities, we collect transactions data for all trading days in 2003 from The New York Stock Exchange Trade and Quote (TAQ) database. To estimate the spread and depth measures, we first calculate the NBBO (National best Bid-Offer) of each security at each time *t*. We exclude the following data points from the NBBO calculation:

- Non-positive prices and quotes
- All quotes with a time stamp before 9:30am (market opening) or after 4:00pm (market closing)
- Quoted with zero bid or offer sizes, and quoted that result in a negative spread
- Quoted and effective spreads that are more than 7.5 standard deviations away from the mean (McDermott, Hegde, and Ascioglu 2005)
- Quoted that were reported in error.

Price, volume, and return data are collected from the Center for Research in Security Prices (CRSP) database. Classification for industry, broad market, and international ETFs are taken from the American Stock Exchange's website, and ETF security holding information is obtained from Morningstar. We discuss the matching methodology in the next section. Our final sample consists of 113 ETFs and their matched equity firms

IVe. Matching Methodology:

[7] Bond holders are exposed to a different set of informed trader incentives than equityholders and the underlying portfolios for bond ETFs may exhibit microstructure characteristics, such as adverse selection costs, that are much different than those of equity ETFs.

To examine the levels of adverse selection between equities and ETFs, we first construct a matched sample of equity securities. Demsetz (1968) shows that bid-ask spreads are positively correlated with price and trading volume. The matching method is similar[8] to the one used in Huang and Stoll (1996), Van Ness Van Ness and Warr (2005a, 2005b) and Jiang and Kim (2005). Available matching equities, volume, return, and price data are obtained from CRSP. We remove all firms with fewer than 227 trading days and all non-ordinary common shares (ADRs, Certificates, Shares of Beneficial Interest, and other depository receipts). The data are averaged daily over all trading days in 2003, and the NYSE[9] or AMEX equity security that minimizes the following objective is selected as the matching equity:

resize this equation to make it normal and right hand justify the equation number

$$Score = \sum_{i=1}^{3}\left(\frac{X_i^{Non\text{-}ETF} - X_i^{ETF}}{X_i^{Non\text{-}ETF} + X_i^{ETF} / 2}\right)^2, \tag{6}$$

where X_i represent one of the three ETF matching attributes, which are the end of day price of the security averaged over the year, the standard deviation of daily returns estimated over the year, and the daily volume averaged over the all trading days in 2003. We select the stock with the lowest matching score, and this process provides one

[8] Unlike studies that match equities to equities, we do not match on market capitalization, because the market capitalization of ETFs and the market capitalization of equities do not capture the same factor.
[9] Because of a limitation with reported quoted depth in NASDAQ listed equities in the TAQ database, we limit our matching firms to NYSE and AMEX listed firms. TAQ only reports depth for one NASDAQ dealer, even if more than one dealer is at the best bid or offer. Because of this underreporting, depth for NASDAQ firms may be understated in the TAQ database. Our results when allowing the inclusion of the NASDAQ firms are quantitatively similar to those sample including NYSE firms only. Restricting the sample to NYSE and AMEX firms resulted in twenty-six firms being replaced by NYSE firms.

matched security to each ETF. The values obtained from the matching process are presented in Table I.

(Insert Table I about here)

As shown in Table I, the average price of the ETFs is $48.30 and the average price of the matching securities is $43.16. The mean daily return standard deviation of the ETFs is 1.31% and the mean daily return standard deviation for the matched sample is 1.39%. The average volume for the ETFs was 1,368,408 shares and the average volume of the matched sample was 717,355. Note, however, that the median volume for the matched sample and the ETFs are very similar. The average matching score from equation 6 is also shown in Table I. The mean/median matching score of 0.113/0.043 suggests that the ETFs and matched equities are similar along the the pre-specified attributes, and the matched portfolio acts as a benchmark for drawing conclusions regarding spread and adverse selection costs in the cross-security analysis. As a comparison, Van Ness, Van Ness and Warr (2005b) in a similar scoring exercise, find a mean matching score of 0.386.[10]

(Insert Tables II here)

V. Results

V.a Univariate Analysis

Table II presents univariate characteristics for the 113 ETFs in our sample. We categorize the ETFs in our sample using various binary variables. *Broad* is a binary variable that takes the value of one when the ETF primarily holds securities from many diverse industry groups. *Sector* is a binary variable that takes the value of one when the ETF primarily holds securities in the same industry sector. *International* is a binary

[10] Van Ness et al. (2005b) are comparing NYSE and NASDAQ stocks and they match on price, trades, trade size, and volatility.

variable that takes the value of one when the ETF primarily holds non-U.S. denominated securities, and *Equity* takes the value of one when the security is an equity security, *Number* is the average number of underlying equity securities in the ETFs. Approximately 23 percent of the ETFs are classified as sector funds, while 15 percent are broadly based and 11 percent are international funds.

(Insert Table III here)

We split our sample into ETFs and equities, and we examine the dimensions of liquidity and adverse selection costs. Mean differences and significance tests are reported in Table III for the quoted, effective, dollar depth, effective/depth ratio, and dollar and percentage estimates for the George, Kaul, and Nimalendran (GKN) (1991), Glosten and Harris (GH) (1988), and Lin, Sanger, and Booth (LSB) (1995) adverse selection costs estimates.

As seen in Table III, the adverse selection percentage cost estimates are consistently larger for equities than for ETFs, regardless of the model used. LSB adverse selection costs percentage bid-ask spread estimates are 19.7% for ETFs and 34.3% for the matched sample of equities. GKN adverse selection bid-ask spread component percentages are 29.6% for ETFs and 72.6% for equities. Finally, GH adverse selection estimates are 18.1% for ETFs and 44.1% for equities. In addition, the dollar cost estimates for equities are also significantly greater for the GH and GKN adverse selection models, but not for LSB. The univariate results clearly support the conjecture that basket securities have significantly lower levels of adverse selection costs than a matched sample of equities.

Table III also reports mean liquidity measures for the ETFs and matched equities. Lower adverse selection costs may or may not lead to an increase in liquidity. Higher levels of liquidity would be associated with lower bid-ask spreads and greater depth. We find that quoted and effective spreads are statistically and economically larger for ETFs than the sample of equities while dollar depth is also statistically and economically larger in ETFs versus the equities. Average quoted dollar depth for the ETFs is 2,531,219, while the quoted dollar depth for the sample of equities is only 71,189. Thus, ETFs possess one characteristic of higher liquidity, namely greater dollar depth, but their wider average and quoted spreads suggests lower liquidity.

To capture both the spread and depth dimensions of liquidity, we compute the *Effective/Depth* ratio, defined as the average effective spread over the trading year divided by the average dollar depth, where dollar depth is scaled by 100,000. This ratio captures increases in the bid-ask spread (width of the market) while also capturing changes in quoted depth (depth of the market). Liquidity decreases as the ratio increases (decreased dollar depth or increased effective spreads) and liquidity increases as the ratio decreases (increased dollar depth or decreased effective spreads).

The *Effective/Depth* ratio for equities is 0.192 but only 0.014 for ETFs. Using this metric as a broad proxy for liquidity suggests that ETFs are much more liquid than equity securities. Table IV presents a simple example of spreads and depth in two markets. Using the spreads and depth in markets A and B and a market buy order for 500 shares, the buy order in market A would move the price to $61.50 with an average price of $60.90, but the same 500 share order in Market B would move the price to $61 with an average price of $60.50. **(maybe another sentence or two on Table IV here).**

(Insert Table IV here)

Univariate analysis cannot not control for other factors affecting liquidity such as volume, risk, and price differences across the securities used in the analysis. In the next section we discuss the multivariate framework that we employ to examine liquidity and adverse selection differences between the sample of equities and the sample of exchange traded funds.

V.b Multivariate Analysis of Industry Concentration

To control for effects of price, volume, and standard deviation in security returns across the securities included in the sample, we estimate the following model:

$$Liquidity_i = \alpha_i + \beta_1 Ln(\mathrm{Pr}\,ice)_{\mathrm{i}} + \beta_2 Ln(STD)_{\mathrm{i}} + \beta_3 Ln(Vol)_{\mathrm{i}} + \sum_{j=4}^{n} \beta_j X_i + e_i \text{ , (7)}$$

where in separate regressions ***Liquidity*** takes the value of the variables *Quoted*, *Effective*, *Dollardepth*, and *Effective/Depth*. *Quoted* is the quoted spread, *Effective* is the effective spread, *Dollardepth* is the dollar value of the shares quoted at the bid and ask prices, *Effective/Depth* is the average effective spread averaged across the trading year divided by average dollar depth averaged across the year, where dollar depth is scaled by 100,000. Ln(*Price)* is the natural log of the average end of day price of the security, Ln(*STD)* is the natural log of the standard deviation of daily returns estimated over the year, Ln(*Vol)* is the natural log of trading volume averaged daily over the year. $\boldsymbol{X_i}$ is a vector of security specific characteristics that includes *ETF*, *Broad*, *Sector,* and *International*. *ETF* takes the value of one when the security is an exchange traded fund, *Broad* is a binary variable that takes the value of one when the ETF primarily holds securities from many diverse industry groups, *Sector* is a binary variable that takes the value of one when the ETF primarily holds securities in the same industry sector, and

International is a binary variable that takes the value of one when the security primarily holds non-U.S. denominated securities.

(Insert Table V here)

The results of the multivariate liquidity analysis can be found in Table V. The binary variable *ETF* is positive and significant in the quoted spread and effective spread regressions. Quoted spreads for ETFs are 5 cents and effective spreads are 4.9 cents higher than those in a matched sample of equities, while controlling for price, volume and standard deviation. ETFs have, on average, $2,475,679 more quoted dollar depth than the matched sample of equity securities. The positive and significant coefficient estimate on the *ETF* binary variable in the dollar depth specification and the positive and significant coefficient estimate on the *ETF* binary variable in the quoted and effective spread analyses lead to an ambiguous result.

To reconcile this difference, we turn to the *Effective/Depth* ratio to aid in determining the liquidity difference between ETFs and equities. Recall that liquidity decreases as the ratio increases (decreased dollar depth or increased effective spreads) and liquidity increases as the ratio decreases (increased dollar depth or decreased effective spreads). As seen in Table V, the coefficient on *Effective/Depth* (-0.185) is negative and significant. Thus, when spreads and depth are considered in a unified framework, exchange traded funds have greater liquidity than the matched sample of equities.

We now extend the previous model to allow for industry concentration effects by including the *International, Sector, and Broad* indicator variables. *Sector* is a binary variable that takes the value of one when the ETF primarily holds securities in the same industry sector, and *Broad* is a binary variable that takes the value of one when the ETF primarily holds securities from many diverse industry groups. We include the binary variable *International* to control for the impact that international ETFs have on the analysis. *International* takes the value of one when the ETF primarily holds non-U.S. denominated securities. In this specification, *Equity* is the reference group for comparing the coefficient estimates of *Sector, Broad*, and *International.* The results of the analysis can be found in the bottom panel of Table V.

The liquidity analysis indicates that broad market, sector and internationally-concentrated baskets all have significantly higher effective spreads than do the matched set of equities. Quoted spreads are also significantly larger for broad market and sector ETFs, although the coefficient on *International* is not significant. The broad-based baskets have the highest spreads relative to equities, followed by the sector concentrated baskets, and then international baskets. Further, the parameter estimates on the variables in the depth specifications, indicate that all forms of ETFs have significantly higher dollar depth than the matched set of equity securities. Again, we turn to the *Effective/Depth* ratio to examine the relationship the liquidity of the ETFs and the matched sample of equities using a single metric.

In the *Effective/Depth* regression framework, the parameter estimates on *International, Sector,* and *Broad* are all negative and significant. This suggests that when spreads and depth are taken together, liquidity is greater in International, Sector and broad-based exchange traded funds than in the matched sample of equity securities.

Recall that the concentration hypothesis suggests that as industry basket concentration increases, the liquidity of the basket should decrease. Based on this hypothesis, we would expect to find the smallest coefficient on *Sector*, given the inverse relationship between the variable and liquidity. However, we find ETF sector funds to have the highest liquidity (most negative coefficient), when using the *Effective/Depth* as a proxy for liquidity, although the differences in coefficients across *International, Sector* and *Broad* are not significant. Thus, we fail to find support to the industry concentration liquidity hypothesis, and our results show that, although sector ETFs have greater quoted and effective spreads than matched equities, they also have substantially greater depth, which dominates the relationship between concentration and liquidity in our analysis. We now turn our attention to examining the adverse selection costs of equity exchange traded funds versus those the matched sample of equities.

(Table VI here)

To explore the relationship between security industry concentration and adverse selection costs, we estimate the following OLS specification:

$$Adverse_i = \alpha_i + \beta_1 Ln(Price)_i + \beta_2 Ln(STD)_i + \beta_3 Ln(Vol)_i + \sum_{j=4}^{n} \beta_j X_i + e_i, \quad (9)$$

In separate regressions, ***Adverse*** takes the value of the adverse selection estimates summarized in Table III, and the explanatory variables are as defined earlier. *Price, STD* and *Vol* are control variables, and we are primarily interested in the relationships between adverse selection and the indicator variable ETF and, within ETFs, the relationships between adverse selection and the categories of ETFs, namely *International*, *Sector* and *Broad.*

The results of the adverse selection estimation are shown in Table VI. Coefficient estimates on the *ETF* binary variable are negative and significant in all three adverse selection models. Regardless of the method used to determine adverse selection, we find lower levels of adverse selection in ETFs than for the matched equities. In the *LSB* model, the percentage of the spread that is attributed to adverse selection costs is on average 14.3% smaller than for the matched sample of equities, and is 42.5% and 26.4% smaller for the GKN and GH models, respectively. We next expand the specifications for ETFs to allow for industry concentration effects by including the binary variables *Sector* and *Broad.*

The coefficient estimates on *Sector, Broad* and *International* are negative and significant in all percentage adverse selection specifications. We find no consistent pattern in the coefficients across the different specifications, and the differences in coefficients across these three indicator variables are not significant. This evidence suggests that industry concentration does not increase adverse selection costs, compared with broad-based or Internationally-focused ETFs.

We do find that equities have significantly higher adverse selection costs than the sample of basket securities (ETFs), providing provides support for the adverse selection basket security hypothesis and strengthening prior research on this topic. The industry concentration hypothesis cannot be used to explain the result of Neal and Wheatley (1998). Recall, we conjectured that commonalities in the underlying mutual funds in their sample might have led to the similar parameter estimates on adverse selection models from their equity and mutual fund samples. Investor uncertainty associated with NAV price deviations would certainly create adverse selections costs. Since we use ETFs, which have essentially no premium or discount, we eliminate noise associated with deviations from NAV. So it is possible that the similarities found by Neal and Wheatley between mutual fund and equity adverse selection costs were driven by premiums and discounts and not by commonalities in the underlying portfolio of securities held by the funds themselves.

We note that equities have higher adverse selection costs than even the internationally-concentrated basket securities, which is surprising given the amount of informational asymmetry between U.S. investors and foreign firms. However, given the diversification of adverse selection costs across the securities held in the baskets, lower levels of total basket security adverse selection costs appear to be achievable. In the next section, we examine this conjecture more closely.

V.c Multivariate Analysis of Security Concentration

We next expand our analysis to determine whether security concentration in ETFs significantly affects liquidity and adverse selection costs. We construct the following model:

$$Liquidity_i = \alpha_i + \beta_1 Ln(Price)_i + \beta_2 Ln(STD)_i + \beta_3 Ln(Vol)_i + \sum_{j=4}^{n} \beta_j X_i + e_i \quad , (11)$$

Liquidity, Price, STD and *Volume* are as defined earlier, and *Herfindahl* and *Number* are included to test for concentration effects. *Ln(Number)* is the natural log of the number of underlying equities that comprise the sample security, and *Ln(Herfindahl)* is the natural log of the Herfindahl Index concentration value of the security. These results are reported in Table VII.

(Table VII)

In this analysis, we are interested in the coefficient estimates on the concentration proxies, *Ln(Herfindahl) and Ln(Number)*. The coefficient estimate on *Ln(Herfindahl)* is insignificant in all specifications except the dollar depth model in which our measure of the Herfindahl index is positively associated with dollar depth. The coefficient on *Ln(Herfindahl)* in the *Effective/Depth* model, however, suggests that it has no impact on security liquidity. Since the concentration hypothesis posits a positive and significant parameter estimate on this variable, our evidence rejects the conjecture that concentration among the equities in the security leads to a decrease in liquidity.[11]

Next we turn our attention to the coefficient estimate on *Ln(Number). Ln(Number)* has no significant impact on quoted or effective spreads, but it significantly increases dollar depth and significantly decreases the *Effective/Depth* ratio. Thus, as the number of securities in a basket security increases, the liquidity of the security increases. In the next section, we examine the impact that these factors have on adverse selection costs.

(Table VIII about here)

Table VIII presents the results of the adverse selection model for ETFs to which we add *Ln(Herfindahl)* and *Ln(Number).* Unlike Table VI, which tests for differences in adverse selection between ETFs and equities, Table VIII presents the adverse selection results for our sample of ETFs. The variables in Table VIII are as defined earlier and *Adverse* is determined separately using the LSB, GKN and GH models.

(Table VII Here)

Our focus in this analysis is on the coefficient estimates on the concentration proxies, *Ln(Herfindahl) and Ln(Number).* The coefficient estimate on *Ln(Herfindahl)* is insignificant in all specifications. Similar to the results of the liquidity analysis, the

[11] It may also be that there is inadequate cross-sectional variation in Herfindahl for the relationship to be significant. However, in Table II, the minimum and maximum values and the standard deviation of Herfindahl suggest considerable cross-sectional variation.

concentration among the equities held in the underlying portfolio of the basket security appears to have no impact on the adverse selection costs. The coefficient estimate on *Ln(Number)* is negative and significant in the partial adverse selection costs specifications for the LSB and GKN models but insignificant in the GH model. These results provide evidence that as the number of securities in an ETF increases, adverse selection is diversified away, and general liquidity is improved.

Table VIII also contains full model estimates that include all concentration (*Sector, Broad, Ln(Number), and Ln(Herfindahl)* and control variables. Note that *Ln(Number), and Ln(Herfindahl)* have no significant impact on dollar adverse selection costs in the full model. However, we caution the interpretation of the economic significance of these results because we use the log of the number of securities and the Herfindahl Index in the specification. In addition, we also point out that the *Sector, International,* and *Broad* retain their negative and significant coefficient estimates in the GKN dollar and GKN percentage estimate models, although the results are generally mixed in the other models. We interpret these results as providing modest support for the negative relationship between adverse selection costs of ETFs and specific ETFs characteristics, namely industry concentration, measured by *Sector* and *Broad* and security concentration, measured by *Ln(Number)* and *Ln(Herfindahl).* . **KEN, IF WE SUGGEST MULTICOLLINEARITY, WE SHOULD TEST FOR IT.**

VI. Conclusion:

We examine liquidity and adverse selection costs in a sample of Exchange Traded Funds (ETFs) and a matched sample of equities. These relationships depend, to a large extent, on the definition of liquidity. When liquidity is viewed separately as a spread or depth measure, then we find ETFs have larger quoted and effective spreads but substantially larger dollar depth. We reconcile these two dimensions of liquidity by computing *Effective/Depth*, which is defined as the average effective spread divided by average dollar depth, and we focus primarily on the relationship between this measure of liquidity and adverse selection for ETFs and sample equities.

We document significantly lower levels of adverse selection costs in the sample of equity ETFs versus the matched sample of equities. In addition, we present evidence that adverse selection costs are decreasing in the number of equities held in the underlying portfolio of the ETF. We show that adverse selection costs do not increase as the concentration among the securities increases, and we find no evidence that industry concentration increases ETF adverse selection costs or reduces liquidity. We also show that when considering the ratio of the effective spreads to scaled dollar depth, ETFs have significantly higher levels of liquidity than a matched sample of equity securities. As a whole, ETFs, and security baskets in general, provide a beneficial trading medium for uninformed traders.

References

Bacidore, J. M., and Sofianos, G., 2002, Liquidity Provisions and Specialist Trading in NYSE-Listed non-U.S. stocks, *Journal of Financial Economics* 63, 133-58.

Barari, M., Lucey, B., and Voronkova, S., 2005, Integration among G7 Equity Markets: Evidence from ishares, *Trinity College Dublin Working Paper*.

Benston, G., and Hagerman, R., 1974, Determinants of the Bid-Ask Spread in the Over-the-Counter Market, *Journal of Financial Economics* 1, 353-364.

Bessembinder, H., 2003, Issues in Assessing Trade Execution Costs, *Journal of Financial Markets* 6, 233-57.

Boehmer, B., and Boehmer, E., 2003, Trading your neighbor's ETFs: Competition or Fragmentation?, *Journal of Banking and Finance* 27, 1667-1704.

Chen, J., Jiang, C., Kim, J., and McInish, T., 2002, Bid-ask spreads, information asymmetry, and investor sentiment: Evidence from closed-end funds, *Review of Quantitative Finance and Accounting* 21, 303-321.

Chiyachantana, C., Jiang, C., Taechapiroontong, N., and Wood, R., 2004, The Impact of Regulation Fair Disclosure on Information Asymmetry and Trading: An Intraday Analysis, *The Financial Review* 39, 549-577.

Clark, J., and Shastri, K., 2001, Adverse selection costs and closed-end funds, *University of Pittsburgh Working Paper*.

Demsetz, H., 1968, The Cost of Transacting, *Quarterly Journal of Economics* 82, 33-53.

Ellis, K., Michaely, R., O'Hara, M., 2000. The Accuracy of Trade Classification Rules: Evidence from NASDAQ, *Journal of Financial and Quantitative Analysis* 35, 529–552.

Elton, E., Gruber, M., Comer, G., and Li, K., 2002, Spiders: Where are the Bugs?, *Journal of Business* 3, 453-472.

George T., Kaul, G., and Nimalendran, M., 1991, Estimation of the Bid-Ask Spread and its Components: A New Approach, *Review of Financial Studies* 4, 623-656.

Glosten, L. R., and Harris, L. E., 1988, Estimating the Components of the Bid-Ask Spread, *Journal of Financial Economics*, 21, 123-42.

Gorton, G., and Pennacchi, G., 1993, Security Baskets and Index-Linked Securities, *Journal of Business* 66, 1-27.

Hamilton, J., 1978, Marketplace organization and marketability: NASDAQ, the stock exchange, and the national market system, *Journal of Finance* 33, 487-503.
Hasbrouck, J., 2000, Intraday Price Formation in U.S. Equity Index Markets, *New York University Working Paper*.

Hegde, S. and McDermott, J., 2004, Market Liquidity of Diamonds, Q's and Their Underlying Stocks, *Journal of Banking and Finance* 28, 1043-1067.

Huang, R., and Stoll, H., 1996, Dealer Versus Auction Markets: A Paired Comparison of Execution Costs on NASDAQ and the NYSE, *Journal of Financial Economics* 41, 313-358.

Jiang, C. X., and Kim, J.C., 2005, Trading Costs in Non-U.S. Stocks on the New York Stock Exchange: The Effect of Institutional Ownership, Analyst Following, and Market Regulation, *The Journal of Financial Research* 3, 439-459.

Kumar, P., and Seppi, D. J., 1994, Information and Index Arbitrage, *Journal of Business* 67, 481-509.

Kumar, R., Sarin, A., and Shastri, K., 1998, The Impact of Options Trading on the Quality of the Market for the Underlying Security: An Empirical Analysis, *Journal of Finance* 53, 717-732.

Kyle, A., 1985, Continuous Auctions and Insider Trading, *Econometrica*, 53, 1315-1335.

Lin J., Sanger, G., and Booth, G., 1995, Trade Size and Components of the Bid-Ask Spread, *Review of Financial Studies* 8, 1153-1183.

Lipper Investments, 2001, http://www.closed-endfunds.com.

McDermott, J. B., Hegde, S., and Ascioglu, A., 2005, Bid-Ask Spread, Informed Trade, and Investment-Cash Flow Sensitivity, *Fairfield University Working Paper.*

Mortal, S., 2003, The Impact of Diversification on Trading Environment, *University of Georgia Dissertation.*

Neal, R. and Wheatley, S., 1998, Adverse Selection and Bid-Ask Spreads: Evidence from Closed-End Mutual Funds, *Journal of Financial Markets* 1, 121-149.

Novakoff, J., 2001, Exchange traded funds, White Paper, 1-3.

Poterba, J., and Shoven, J., 2002, Exchange Traded Funds: A New Investment Option for Taxable Investors, *American Economic Review* 92, 422-427.

Small, K., 2005, The limits of Arbitrage: The Case of Exchange Traded Funds, *Loyola College Working Paper.*

Small, K., Flaherty, S., and Ionici, I., 2005, U.S. Cross-listing and Market Liquidity: A Test of the Bonding Hypothesis, *Loyola College Working Paper*.
Small, K., and Wansley, J., 2005, An Examination of Informed Trader Migration, *Loyola College Working Paper.*

Subrahmanyam, A., 1991, A Theory of Trading in Stock Index Futures, *Review Of Financial Studies* 4, 17-51.

Tinic, S., 1972, The Economics of Liquidity Services, *Quarterly Journal of Economics*, 86, 79-93.

Van Ness, B., Van Ness, R. and Warr, R., 2001, How Well Adverse Selection Components Measure Adverse Selection?, *Financial Management* 30, 77-98.

Van Ness, B., Van Ness, R. and Warr, R., 2005a, The Impact of the Introduction of Index Securities on the Underlying Stocks: The Case of Diamonds and the Dow 30, *Advances in Quantitative Analysis of Finance and Accounting* 2, 105-129.

Van Ness, B., Van Ness, R. and Warr, R., 2005b, The Impact of Market Maker Concentration on Adverse-Selection Costs for NASDAQ Stocks 3, *The Journal of Financial Research* 3, 461-485.

Table I
ETF Matching Scores

This table contains the ETF and control sample matching attributes. The matching score is calculated as:

$$Score = \sum_{i=1}^{3}\left(\frac{X_i^{Non\text{-}ETF} - X_i^{ETF}}{X_i^{Non\text{-}ETF} + X_i^{ETF} / 2}\right)^2,$$

X_i takes the value of the end of day price of the security averaged over the year, the daily standard deviation of daily returns estimated over the year, and the daily volume of security averaged daily over all trading days in 2003. Firms with the minimum matching score are chosen as matching firms.

		Mean	Median	Standard Deviation	Min	Max
Price	**ETF**	48.30	42.61	32.10	2.58	132.92
	Non-ETF	43.16	39.70	29.01	3.24	188.39
Volume	**ETF**	1,368,408	85,392	8,271,034	2,662	77,513,826
	Non-ETF	717,355	86,885	3,999,864	2,351	20,554,232
STD	**ETF**	.0131	.0117	.0038	.0076	.0268
	Non-ETF	.0139	.0131	.0041	.007	.0325
Matching Score		.113	.043	.218	.00061	1.64

Table II
ETF and sample Equity Univariate Characteristics

The table contains the mean, standard deviation, minimum, and maximum values of the variables used in the analysis. *Broad* is a binary variable that takes the value of one when the ETF primarily holds securities from many diverse industry groups, *Sector* is a binary variable that takes the value of one when the ETF primarily holds securities in the same industry sector, *International* is a binary variable that takes the value of one when the security primarily holds non-U.S. denominated securities, Equity takes the value of one when the security is an equity security, *Number* is the number of underlying equities that comprise the sample security, portfolio, *Herfindahl* is the Herfindahl Index concentration value of the security, *Price* is the end of day price of the security averaged over the year, *STD* is the standard deviation of daily returns estimated over the year, *Vol* is the daily volume of security averaged over the year, *Quoted* is the quoted spread, *Effective* is the effective spread, *Depthshares* is the number of shares quoted at the best bid and offer, *Dollardepth* is the dollar value of the shares quoted at the bid and ask prices, *Effective/Depth* is the average effective spread averaged across the trading year divided by average dollar depth averaged across the year, where dollar depth is scaled by 100,000, *GKN* is the George, Kaul, and Nimalendran (1991) percentage bid-ask spread decomposition adverse selection estimate, *GKNdollar* is the percentage *GKN* adverse selection cost estimate times the quoted spread, *GH* is the Glosten and Harris (1988) percentage bid-ask spread decomposition adverse selection estimate, *GHdollar* is the percentage *GH* estimate times the quoted spread, *LSB* is the Lin, Sanger, and Booth (1995) percentage bid-ask spread decomposition adverse selection estimate, and *LSBdollar* is the *LSB* percentage adverse selection cost estimate time the effective spread.

	Mean	***STD***	***Min***	***Max***
Broad	.15	.36	.00	1.00
Sector	.23	.42	.00	1.00
International	.11	.31	.00	1.00
Equity	.50	.50	.00	.00
Number	126	349	1.00	2,891
Herfindahl	5,308	4,806	0.41	10,000
Price	45.80	30.66	2.58	188.40
Vol	894,721	5,958,915	2,351	77,513,826
Std	.013	.00	.01	.03
Quoted	.12	.09	.01	.69
Effective	.11	.08	.01	.47
DepthShares	27,558	39,506	419.70	286,085
DollarDepth	1,290,122	2,223,629	9,438	16,334,232
Effective/Depth	.10	.19	.00	1.17
GKN	.51	.26	.00	1.00
GKNDollar	.05	.06	.00	.55
GH	.31	.19	.00	1.00
GHDollar	.03	.04	.00	0.33
LSB	.27	.16	.00	1.00
LSBDollar	.03	.03	.00	.18

Table III
Univariate Comparisons between ETFs and matched Equity Securities

This table contains the mean values of the liquidity and adverse selection model estimates for the sample of ETFs and equities. *Quoted* is the quoted spread, *Effective* is the effective spread, *Depthshares* is the number of shares quoted at the best bid and offer, Dollardepth is the dollar value of the shares quoted at the bid and ask prices, *Effective/Depth* is the average effective spread averaged across the trading year divided by average dollar depth averaged across the year, where dollar depth is scaled by 100,000, *LSB* is the Lin, Sanger, and Booth (1995) percentage bid-ask spread decomposition adverse selection estimate, *LSBdollar* is the *LSB* percentage adverse selection cost estimate time the effective spread, GH is the Glosten and Harris (1988) percentage bid-ask spread decomposition adverse selection estimate, *GHdollar* is the percentage *GH* estimate times the quoted spread, *GKN* is the George, Kaul, and Nimalendran (1991) percentage bid-ask spread decomposition adverse selection estimate, and *GKNdollar* the percentage *GKN* adverse selection cost estimate times the quoted spread.

Variable	*ETFs*	*Equities*	*Difference*
Quoted	.142	.089	.0527*** (4.39)
Effective	.125	.072	.053*** (5.23)
Dollar Depth	2,531,219	71,189	2,460,030*** (9.88)
Depth Shares	53,545	2,035	51.510*** (12.8)
Effective/Depth	.014	.192	-.179*** (8.09)
LSB	.197	.343	0.146*** (7.66)
LSBDollar	.027	.033	0.005 (-1.14)
GH	.181	.441	0.261*** (-13.72)
GHDollar	.027	.042	0.016*** (-2.69)
GKN	.296	.726	0.43*** (-21.96)
GKNDollar	.038	.066	0.027*** (-3.43)

t-statistic in Parentheses
** indicates Significance 10% level*
*** indicates Significance 5% level*
**** indicates significance 1% level*

Table IV Example of Spreads and Depth in Two Markets

Market A

	Offer (Ask)	Depth (Ask)	Bid	Depth (Bid)
	$62	100		
	$61.50	200		
	$61	100		
	$60.50	100		
	$60	100		
			$59.50	100
			$59	200
			$48.50	100
			$48	100
			$47.50	100
Spread = 50 cents Depth = 200 shares Dollar Depth = $11,050				

Market B

	Offer (Ask)	Depth (Ask)	Bid	Depth (Bid)
	$62.50	100		
	$62	200		
	$61.50	100		
	$61	100		
	$60.50	500		
			$59	500
			$48.50	200
			$48	100
			$47.50	100
			$47.00	100
Spread = $1.50 Depth = 1000 shares Dollar Depth = $59,750				

In market A, spreads (width) and quoted depth (depth) are smaller than in market B. In market B, spreads are larger but depth is much larger than in market A. If market A receives a buy order for 500 shares the price of the security will move to $61.50 and the buyer will pay an average price of $60.90. A buy order submitted to market B will move the price to $61 and the buyer will pay and average price of $60.50. As an example, Market A would have a *Quoted/DollarDepth* ratio of 4.51 and B has a ratio of 2.51. The example illustrates that liquidity is more than just signed spreads and depth.

Table V
Exchange Traded Fund and Equity Liquidity Analysis
(Industry Concentration Analysis)

This table contains regression analysis coefficient estimates for the following model:

$$Liquidity_i = \alpha_i + \beta_1 Ln(\text{Price})_i + \beta_2 Ln(STD)_i + \beta_3 Ln(Vol)_i + \sum_{j=4}^{n} \beta_j X_i + e_i$$

Liquidity takes the value of the variables *Quoted, Effective, Dollardepth*, and *Effective/Depth*. Where *Quoted* is the quoted spread, *Effective* is the effective spread, *Dollardepth* is the dollar value of the shares quoted at the bid and ask prices, *Effective/Depth* is the average effective spread averaged across the trading year divided by average dollar depth averaged across the year, where dollar depth is scaled by 100,000. Ln(*Price)* is the natural log of the end of day price of the security averaged over the year, Ln(*STD)* is the natural log of the daily standard deviation of daily returns estimated over the year, ln(*Vol)* is the natural log of the daily volume of security averaged daily over the year. X_i is a vector of security specific characteristics that includes *Broad*, *ETF*, *Sector*, and *International*. Where *Broad* is a binary variable that takes the value of one when the ETF primarily holds securities from many diverse industry groups, *Sector* is a binary variable that takes the value of one when the ETF primarily holds securities in the same industry sector, *International* is a binary variable that takes the value of one when the security primarily holds non-U.S. denominated securities, and *ETF* takes the value of one when the security is an exchange traded fund.

	Liquidity Measures							
	Quoted Spread	**Quoted Spread**	**Effective Spread**	**Effective Spread**	**Dollar Depth**	**Dollar Depth**	**Effective/Depth**	**Effective/Depth**
Intercept	.045	0.029	.047	.031	1,262,656	-371,275	.151	.149
	(.486)	(.311)	(.589)	(.399)	(.667)	(-.237)	(.391)	(.832)
Ln(Price)	-.014	-.014	-.011	-.010	-232,341	-136,019	-.028	-.028
	(-1.45)	(-1.37)	(-1.31)	(-1.20)	(1.19)	(-1.11)	(-1.24)	(-1.22)
Ln(STD)	-.025	-.030	-.021	-.026	229,652	-301,326	-.058	-.058
	(-1.00)	(-1.23)	(-1.04)	(-1.29)	(.517)	(-.804)	(-1.15)	(-1.13)
Ln(Volume)	-0.001	-.002	-.001	-.002	54,822	-32,578	-.009*	-.0095*
	(-.351)	(-.68)	(-.622)	(-.935)	(.941)	(-.743)	(-1.77)	(-1.69)
ETF	0.050***		.049***		2,475,679***		-0.185***	
	(4.02)		(4.83)		(9.94)		(-7.78)	
International		.019		.033**		347,807***		-.166***
		(1.21)		(2.24)		(4.60)		(-6.63)
Sector		.0477***		.043***		1,932,351***		-.192***
		(3.07)		(3.30)		(9.52)		(-7.51)
Broad		.077***		.071***		4,883,989***		-.189***
		(4.54)		(5.16)		(8.66)		(-8.45)
Adjusted R^2	.07	.09	.10	.11	.30	.58	.232	.23
F Value	5.60	4.78	7.55	5.76	25.64	52.31	18.03	12.02
N	226	226	226	226	226	226	226	226

t-statistic in Parentheses
** indicates Significance 10% level*
*** indicates Significance 5% level*
**** indicates significance 1% level*

Table VI
Exchange Traded Fund and Equity Adverse Selection Cost Analysis
(Industry Concentration Analysis)

This table contains regression analysis coefficient estimates for the following model:

$$Adverse_i = \alpha_i + \beta_1 Ln(\text{Price})_i + \beta_2 Ln(STD)_i + \beta_3 Ln(Vol)_i + \sum_{j=4}^{n} \beta_j X_i + e_i$$

Adverse takes the value of the variables *LSB*, *LSBdollar*, *GH GHdollar*, GKN, and *GKNdollar*. Where *GKN* is the George, Kaul, and Nimalendran (1991) percentage bid-ask spread decomposition adverse selection estimate, *GKNdollar* the percentage *GKN* adverse selection cost estimate times the quoted spread, *GH* is the Glosten and Harris (1988) percentage bid-ask spread decomposition adverse selection estimate, *GHdollar* is the percentage *GH* estimate times the quoted spread, *LSB* is the Lin, Sanger, and Booth (1995) percentage bid-ask spread decomposition adverse selection estimate, and *LSBdollar* is the *LSB* percentage adverse selection cost estimate time the effective spread, Ln(*Price)* is the natural log of the end of day price of the security averaged over the year, Ln(*STD)* is the natural log of the daily standard deviation of daily returns estimated over the year, ln(*Vol)* is the natural log of the daily volume of security averaged daily over the year. X_i is a vector of security specific characteristics that includes *Broad*, *ETF*, *Sector*, and *International*. Where *Broad* is a binary variable that takes the value of one when the ETF primarily holds securities from many diverse industry groups, *Sector* is a binary variable that takes the value of one when the ETF primarily holds securities in the same industry sector, *International* is a binary variable that takes the value of one when the security primarily holds non-U.S. denominated securities, and *ETF* takes the value of one when the security is an exchange traded fund.

	Adverse Selection Measures								
	LSB (Cents)	**LBS (Percentage)**	**LSB (Percentage)**	**GKN (Cents)**	**GKN (Percentage)**	**GKN (Percentage)**	**GH (Cents)**	**GH (Percentage)**	**GH (Percentage)**
Intercept	.0106	.298*	.269	.009	.456**	.503***	.025	.509***	.471***
	(.279)	(1.71)	(1.62)	(.157)	(2.35)	(2.65)	(.552)	(2.82)	(2.66)
Ln(Price)	-.003	-.0038	-.002	-.011	-.003	-.006	-.006	-.011	-.009
	(-1.04)	(-.312)	(-.169)	(1.41)	(-.238)	(-.451)	(-1.35)	(-.960)	(-0.79)
Ln(STD)	-.010	.004	-.003	-.021	-.032	-.017	-.017	-.015	-.027
	(-1.27)	(.142)	(-.118)	(1.08)	(-.801)	(-.420)	(-1.43)	(-.426)	(-0.77)
Ln(Volume)	-.0008	.006	.005	.0002	.012	.014	-.0028*	-.007	-.009*
	(-.477)	(.736)	(.558)	(.108)	(1.49)	(1.82)	(-1.85)	(-1.40)	(-1.73)
ETF		-.143***			-.425***			-.264***	
		(-6.88)			(-20.26)			(-14.1)	
International	-.008		-0.121***	-.028***		-.346***	-.029***		-.314***
	(-1.48)		(-4.69)	(2.97)		(-10.7)	(-4.69)		(-14.6)
Sector	-.010*		-.170***	-.027***		-.415***	-.021***		-.277***
	(-1.86)		(-5.94)	(-2.73)		(-15.3)	(-3.25)		(-13.6)
Broad	.003		-.117***	-.030***		-.500***	-.002		-.208***
	(.464)		(-3.32)	(-3.36)		(-18.0)	(-.242)		(-5.96)
Adjusted R^2	.002	.202	.207	.042	.688	.70	.055	.458	.47
F Value	1.09	15.05***	10.67***	2.63**	122.88***	90.27***	3.15***	47.7***	34.09***
N	226	226	226	226	226	226	226	226	226

t-statistic in Parentheses
** indicates Significance 10% level*
*** indicates Significance 5% level*
**** indicates significance 1% level*

Table VII

Exchange Traded Funds and Equity Liquidity Analysis (Number and Full Model Concentration Analysis)

This table contains regression analysis coefficient estimates for the following model:

$$Liquidity_i = \alpha_i + \beta_1 Ln(\text{Price})_i + \beta_2 Ln(STD)_i + \beta_3 Ln(Vol)_i + \sum_{j=4}^{n} \beta_j X_i + e_i$$

Liquidity takes the value of the variables *Quoted*, *Effective*, *Dollardepth*, and *Effective/Depth*. Where *Quoted* is the quoted spread, *Effective* is the effective spread, *Dollardepth* is the dollar value of the shares quoted at the bid and ask prices, *Effective/Depth* is the average effective spread averaged across the trading year divided by average dollar depth averaged across the year, where dollar depth is scaled by 100,000. Ln(*Price)* is the natural log of the end of day price of the security averaged over the year, Ln(*STD)* is the natural log of the daily standard deviation of daily returns estimated over the year, ln(*Vol)* is the natural log of the daily volume of security averaged daily over the year. X_i is a vector vector of security specific characteristics that includes *Broad, ETF, Sector, International, Ln(Herfindahl), and Ln(Number)*. Where *Broad* is a binary variable that takes the value of one when the ETF primarily holds securities from many diverse industry groups, *Sector* is a binary variable that takes the value of one when the ETF primarily holds securities in the same industry sector, *International* is a binary variable that takes the value of one when the security primarily holds non-U.S. denominated securities, Ln(*Number)* is the natural log of the number of underlying equities that comprise the sample security, Ln(*Herfindahl)* is the natural log of the Herfindahl Index concentration value of the security.

	Quoted Spread	Quoted Spread	Effective Spread	Effective Spread	Dollar Depth	Dollar Depth	Effective/ Depth	Effective/ Depth
Intercept	.100	.134	.103	.128	-1,103,488	-3,692,916	.205	.174
	(.987)	(1.18)	(1.20)	(1.34)	(-.424)	(1.49)	(1.15)	(.993)
Ln(Price)	-.012	-.011	-.008	-.008	-51,114	-49,538	-.035	-.029
	(1.19)	(-1.15)	(-1.03)	(-1.01)	(-.361)	(-.479)	(1.46)	(-1.23)
Ln(STD)	-.031	-.029	-.028	-.026	247,577	-68,079	-.050	-.061
	(1.23)	(-1.14)	(-1.32)	(-1.21)	(.731)	(-.217)	(.958)	(-1.15)
Ln(Volume)	-.002	-.002	-.002	-.002	47,922	-6,755	-.007	-.009*
	(.761)	(-.751)	(1.13)	(-1.03)	(1.00)	(-.166)	(-1.28)	(1.71)
Ln(Herfindahl)	-.009	-.012	-.009	-.011	90,326	316,001**	.0002	-.002
	(1.22)	(1.35)	(-1.44)	(-1.46)	(.631)	(2.86)	(.071)	(-.991)
Ln(Number)	.0045	.009	.003	.0061	886,970***	813,869***	-.041***	-.007*
	(.494)	(1.02)	(.47)	(.731)	(4.60)	(2.86)	(-5.88)	(-1.81)
International		-.053*		-0.024		1,044,284		-.149***
		(-1.86)		(-.96)		(1.36)		(5.70)
Sector		-.018		-.007		476,508		-.175***
		(-.642)		(-.311)		(.66)		(-6.59)
Broad		-.051		-.031		2,865,816**		-.164***
		(-1.01)		(-.711)		(2.34)		(-6.22)
Adjusted R^2	.113	.115	.142	.136	.50	.63	.19	.22
F Value	6.66***	4.60***	8.36***	5.35***	45.66***	49.70***	11.37***	8.97***
N	226	226	226	226	226	226	226	226

t-statistic in Parentheses
** indicates Significance 10% level*
*** indicates Significance 5% level*
**** indicates significance 1% level*

Table VIII
Exchange Traded Fund and Equity Adverse Selection Cost Analysis (Number and Full Model Concentration Analysis)

This table contains regression analysis coefficient estimates for the following model:

$$Adverse_i = \alpha_i + \beta_1 Ln(\text{Price})_i + \beta_2 Ln(STD)_i + \beta_3 Ln(Vol)_i + \sum_{j=4}^{n} \beta_j X_i + e_i$$

Adverse takes the value of the variables *LSB, LSBdollar, GH GHdollar*, GKN, and *GKNdollar.* Where *GKN* is the George, Kaul, and Nimalendran (1991) percentage bid-ask spread decomposition adverse selection estimate, *GKNdollar* the percentage *GKN* adverse selection cost estimate times the quoted spread, *GH* is the Glosten and Harris (1988) percentage bid-ask spread decomposition adverse selection estimate, *GHdollar* is the percentage *GH* estimate times the quoted spread, *LSB* is the Lin, Sanger, and Booth (1995) percentage bid-ask spread decomposition adverse selection estimate, and *LSBdollar* is the *LSB* percentage adverse selection cost estimate time the effective spread.. Ln(*Price)* is the natural log of the end of day price of the security averaged over the year, Ln(*STD)* is the natural log of the daily standard deviation of daily returns estimated over the year, ln(*Vol)* is the natural log of the daily volume of security averaged daily over the year. X_i is a vector of security specific characteristics that includes *Broad, ETF, Sector, International, Ln(Herfindahl), and Ln(Number)*. Where *Broad* is a binary variable that takes the value of one when the ETF primarily holds securities from many diverse industry groups, *Sector* is a binary variable that takes the value of one when the ETF primarily holds securities in the same industry sector, *International* is a binary variable that takes the value of one when the security primarily holds non-U.S. denominated securities, Ln(*Number)* is the natural log of the number of underlying equities that comprise the sample security, Ln(*Herfindahl)* is the natural log of the Herfindahl Index concentration value of the security.

	LSB (Percentage)	**LBS (Cents)**	**LSB (Percentage)**	**GKN (Percentage)**	**GKN (Cents)**	**GKN (Percentage)**	**GH (Percentage)**	**GH (Cents)**	**GH (Percentage)**
Intercept	.311	.012	.174	.027	.032	.469**	.067	.040	.476**
	(1.51)	(.299)	(.942)	(.467)	(.519)	(2.26)	(1.35)	(.782)	(2.38)
Ln(Price)	-.008	-.004	-.0045	-.012	-.011	-.0102	-.007	-.007	-.013
	(-.692)	(-1.08)	(-.348)	(-1.49)	(-1.38)	(-.710)	(-1.50)	(-1.39)	(-1.05)
Ln(STD)	.012	-.0112	-.0057	-.021	-.022	-.023	-.014	-.018	-.034
	(.383)	(-1.32)	(-.174)	(-1.09)	(-1.10)	(-.567)	(-1.20)	(-1.50)	(-.973)
Ln(Volume)	.008	-.0008	.005	.0004	.0001	.014*	-.002	-.003	-.010*
	(1.00)	(-.502)	(.553)	(.219)	(.060)	(1.75)	(-1.49)	(-1.90)	(-1.79)
Ln(Herfindahl)	.003	-.00004	.011	-.001	-.002	.005	-.003	-.001	.0015
	(.498)	(-.023)	(1.52)	(-.627)	(1.05)	(.756)	(-1.22)	(-.489)	(.178)
Ln(Number)	-.028***	-.0026	-.011	-.007**	-.001	-.025**	-.006**	-.004	-.029*
	(-3.10)	(-.743)	(-.638)	(.007)	(-.59)	(-2.12)	(-1.22)	(-1.10)	(-1.69)
International		-.0003	-.046		-.031**	-.248***		-.020	-.220***
		(-.025)	(-.655)		(2.46)	(-4.25)		(-1.32)	(-3.40)
Sector		-.003	-.103*		-.029**	-.323***		-.011	-.188***
		(-225)	(1.71)		(-2.34)	(-6.31)		(-.806)	(-3.11)
Broad		.016	.012		-.037**	-.338***		.012	-.054
		(.592)	(.090)		(-2.01)	(-3.83)		(.395)	(-.433)
Adjusted R^2	.158	.000	.21	.038	.034	.711	.022	.051	.479
F Value	9.30***	.892	8.37***	2.76**	2.00**	69.01***	2.01*	2.48**	26.46***
N	226	226	226	226	226	226	226	226	226

t-statistic in Parentheses
** indicates Significance 10% level*
*** indicates Significance 5% level*
**** indicates significance 1% level*