Does The "Spiders" Market Attract Uninformed Trading Volume?

Quentin C. Chu, University of Memphis, USA
Dan Zhou, California State University at Bakersfield, USA

ABSTRACT

The trading volume of Standard and Poor's Depository Receipts (SPDRs) - or Spiders - has grown consistently since the inception of trading in 1993. Theoretical models have predicted that the Spiders market would attract trading volume from uninformed traders because their losses due to adverse trades with informed traders would usually be lower in this market than in individual security markets. As an extension of the modified mixture distribution hypothesis model proposed by Andersen (1996), this study applies the estimated parameters from the generalized method of moments to derive the percentage trading volume of SPDRs attributable to uninformed trades. Using ninety securities selected from the S&P 500 index as benchmark stocks for comparison, we find that the Spiders market indeed attracts a relatively higher percentage of trading volume from uninformed traders.

Keywords: Standard and Poor's Depositary Receipts; Modified mixture distribution hypothesis; Generalized method of moments; Uninformed traders; Trading volume

1. INTRODUCTION

In the past decade, one of the most exciting financial innovations is the introduction of Exchange-Traded Funds (ETFs). The ETFs are traded just like shares of common stocks, but they are unit investment trusts and their share prices are directly linked to their respective stock indexes. The first ETF traded in the U.S. market, Standard and Poor's Depository Receipts (SPDRs) - or Spiders, was introduced by the American Stock Exchange (AMEX) on January 29, 1993.[1] It is designed to track the performance of the Standard & Poor's 500 index.

Immediately after the launching, SPDRs have attracted significant trading volume and have been deemed a great success of product innovation by the AMEX. Figure 1 presents the daily trading volume for SPDRs from February 1, 1993 through December 29, 2000. The trading volume series demonstrates a significant growth trend since the inception of trading in 1993. Currently, the SPDRs are one of most actively traded securities on the NYSE.

The creation of ETFs gives investors an alternative trading vehicle. An investor can choose to buy or sell either the ETFs or the underlying individual securities that compose the indexes. Poterba and Shoven (2002) point out that ETFs are more tax efficient than traditional mutual funds. The tax advantage is due to the "in-kind redemption" technique adopted by the ETFs. Ackert and Tian (2000) and Elton et al. (2002) examine the characteristics and performance of SPDRs. They show that SPDRs track the S&P 500 index quite precisely.

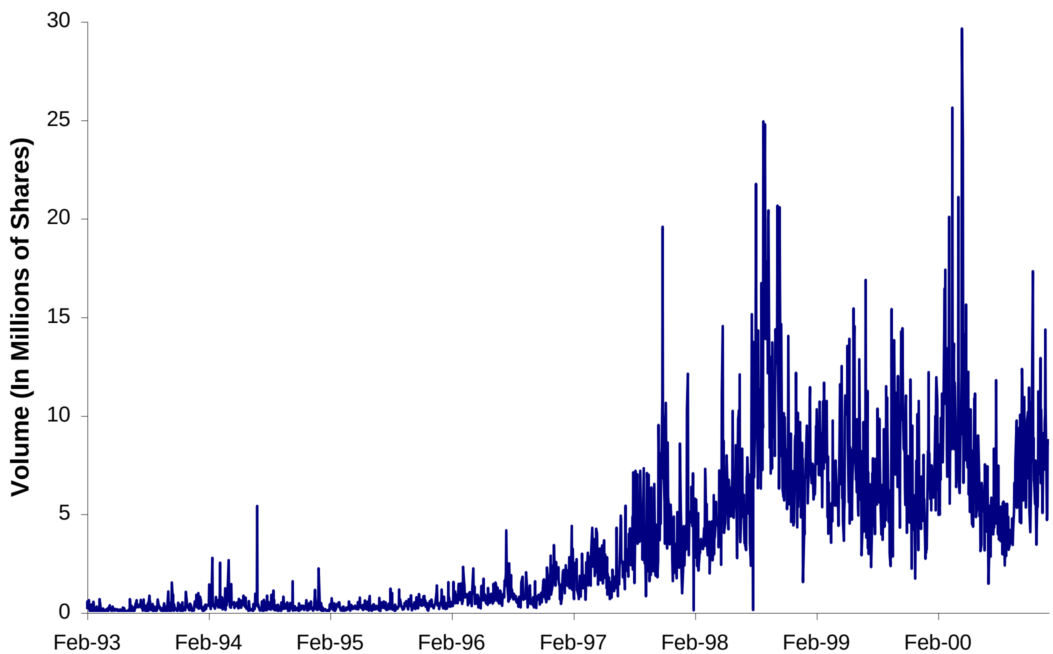
Given that SPDRs show excellent tracking record, Subrahmanyam (1991), Gorton and Pennacchi (1993) argue that composite securities, such as SPDRs, are not redundant because the return on these securities cannot be replicated by holding the individual underlying stocks when prices are not fully revealing or the market is not completely transparent.

[1] AMEX was acquired by the NYSE in 2008 and the SPDRs are currently traded on the NYSE

Subrahmanyam (1991) presents a model to characterize the trading strategy of discretionary liquidity (uninformed) traders. These uninformed traders can choose to execute their portfolio trades either in the market for the composite security or its underlying securities markets. The informed traders are another group of traders who possess firm-specific and/or systematic risk information. He finds that because of the "diversification" or "information offset" effect of the independent trades of the informed traders in the composite security, the total effect of informed trading is less damaging to the discretionary liquidity traders in the market of composite security than in its underlying individual securities markets.

Figure 1 displays the raw daily trading volume of SPDRs. The raw daily trading volume, as measured by the number of shares traded, is retrieved from the CRSP database over the period from February 1, 1993 to December 29, 2000. In total, there are 2,000 observations for SPDRs.

Figure 1: The Raw Daily Trading Volume for SPDRS over 1993-2000



Gorton and Pennacchi (1993) concentrate on characterizing the optimal trading strategy of uninformed traders. They illustrate that the creation of the composite security can reduce the informed traders' information advantages over the uninformed traders and minimize the uninformed traders' loss to the informed traders. Assuming that the investors' utility function depends only on the mean and variance of return from investing in securities, they prove that the existence of the informed traders in the markets can decrease uninformed traders' expected rate of return on any security and increase their return variance, thus reduces their expected utility. Due to the diversification effect, a composite security can always be created to increase the expected utility of uninformed traders.

Specifically, Gorton and Pennacchi (1993) show that for any set of individual security portfolio, an uninformed trader prefers to hold the counterpart composite security carrying the same portfolio weights. By holding this composite security, the uninformed trader would receive a higher expected return and face a lower variance.

Because there is less information asymmetry problem in the market of composite security than in the markets of its underlying individual securities, the uninformed traders would be subject to less adverse selection costs. The trading of the composite securities would attract trading volume from uninformed traders because their losses due to adverse trades with informed traders would usually be lower in this market than in individual securities markets.

To empirically confirm the hypothesis that composite securities would attract trading volume from uninformed traders, this study applies generalized method of moments (GMM) to estimate the relative trading volume of SPDRs attributable to uninformed trades. The estimation model is based on Andersen's (1996) modified mixture of distribution hypothesis (MDH) Model. Using ninety securities selected from the S&P 500 index as benchmark stocks for comparison, we find that the Spiders market indeed attracts a relatively higher percentage of trading volume from uninformed traders.

2. METHODOLOGY

The mixture of distribution hypothesis (MDH) posits that the joint distribution of daily return and volume can be modeled as a mixture of bivariate normal distributions. Specifically, they are contemporaneously dependent on an underlying mixing variable that represents the flow of information. Clark (1973) first develops MDH model to describe the distribution of speculative prices.

The assumption that the daily trading volume follows a normal distribution in the original model seems unreasonable because the normality assumption may result in a negative trading volume. Relying on the theoretical microstructure framework of Glosten and Milgrom (1985), Andersen (1996) proposes a modified MDH model, which assumes that daily trading volume follows a Poisson distribution. He divides trading volume into two components, informed and uninformed components. His model can be characterized as follows:

$$R_t \mid K_t \sim N(\bar{r}, K_t) \tag{1}$$

$$\hat{V}_t \mid K_t \sim c \cdot Po(m_0 + m_1 K_t) \tag{2}$$

where R_t is the stock return on day *t*; K_t is a mixing variable, usually interpreted as the unobserved flow of underlying information regarding the future dividends or the liquidation value of a particular stock; $\hat{V}_t$ is the detrended, stationary trading volume series; m_0 is the daily arrival intensity of uninformed trading, which is independent of the arrival of information; m_1 measures how strongly trading volume fluctuates in response to the news; and *c* is an unknown positive constant introduced due to a scaling indeterminacy that arises when detrended volume data are used in the estimation.

Equation (1) specifies that stock returns given information flow have a normal distribution with mean $\bar{r}$ and variance K_t. The conditional distribution for detrended volume specified in equation (2) follows a Poisson distribution with mean and variance parameter $m_0 + m_1 K_t$.

Andersen's (1996) modified MDH model allows us to estimate uninformed traders' average daily trading volume, which is cm_0. The average daily trading volume (detrended) is $E(\hat{V}_t) \equiv \bar{V} = cm_0 + cm_1\bar{K}$, so the other part $cm_1\bar{K}$ measures informed traders' average daily trading volume.[2] The percentage uninformed trading volume is

[2] Trading volume variables are based on the detrended time series specified in section 3.2.

measured by the ratio $\dfrac{cm_0}{cm_0 + cm_1\overline{K}}$.

To estimate percentage uninformed trading volume, we apply Hansen (1982) generalized method of moments (GMM) procedure to the moment conditions specified in Andersen (1996). Appendix I presents the derivation of the following twelve unconditional moment equations.[3, 4]

$$
\begin{aligned}
&(a) \quad E(R_t) = \bar{r} \\
&(b) \quad E|R_t - \bar{r}| = \sqrt{2/\pi}\,E\left[K_t^{1/2}\right] \\
&(c) \quad E[(R_t - \bar{r})^2] = E(K_t) \equiv \overline{K} \\
&(d) \quad E|R_t - \bar{r}|^3 = 2\sqrt{2/\pi}\,E\left[K_t^{3/2}\right] \\
&(e) \quad E\left[(R_t - \bar{r})^4\right] = 3\left[(\overline{K})^2 + Var(K_t)\right] \\
&(f) \quad E(\hat{V}_t) = c\cdot(m_0 + m_1\overline{K}) = \overline{V} \\
&(g) \quad E\left[(\hat{V}_t - \overline{V})^2\right] = c\overline{V} + c^2 m_1^2 Var(K_t) \\
&(h) \quad E\left[(\hat{V}_t - \overline{V})^3\right] = c^2\overline{V} + 3c^3 m_1^2 Var(K_t) + c^3 m_1^3 E[K_t - \overline{K}]^3 \\
&(i) \quad E[R_t \hat{V}_t] = \bar{r}\overline{V} \\
&(j) \quad E[|R_t - \bar{r}|(\hat{V}_t - \overline{V})] = c\sqrt{2/\pi}\cdot m_1\left\{E(K_t^{3/2}) - \overline{K}E(K_t^{1/2})\right\} \\
&(k) \quad E[(R_t - \bar{r})^2 \hat{V}_t] = m_1 Var(K_t) + \overline{K}\overline{V} \\
&(l) \quad E[(R_t - \bar{r})^2(\hat{V}_t - \overline{V})^2] = c\overline{K}\overline{V} + c^2 m_1 Var(K_t) + c^2 m_1^2[E(K_t - \overline{K})^3 + \overline{K}Var(K_t)]
\end{aligned}
\qquad (3)
$$

The parameter vector is $\theta = \left(\bar{r}, E\left[K_t^{1/2}\right], \overline{K}, E\left[K_t^{3/2}\right], Var(K_t), c, cm_0, cm_1, E\left(K_t - \overline{K}\right)^3\right)$. Thus, there are nine unknown parameters and twelve moment conditions, resulting in three over-identifying restrictions. The chi-square tests for goodness-of-fit have three degrees of freedom. The system is estimated by minimizing the distance between the sample and theoretical moments over the parameter space in a quadratic form in accordance with the Newey and West (1987) procedure.

The estimation procedure is as follows: θ is the (9×1) vector of unknown parameter. Let ω_t be a (2×1) vector of daily return and detrended volume observed at date *t*, $h(\theta, \omega_t)$ be a (12×1) vector-valued function. Since ω_t is a random variable, so is $h(\theta, \omega_t)$. Let θ_0 be the true parameter vector, then the twelve orthogonal conditions can be written as $E[h(\theta_0, \omega_t)] = 0$. Let $\eta_T \equiv (\omega_1', ..., \omega_T')$ be a ($T\times2$) matrix containing all the observation in a sample size of *T*. Then the vector of sample moments can be denoted as $g(\theta, \eta_T) = \left(\sum_{t=1}^{T} h(\theta, \omega_t)\right)/T$.

---

[3] In Andersen (1996), equation (j) was written as $E[|R_t - \bar{r}|(\hat{V}_t - \overline{V})] = c\sqrt{2/\pi}\,m_1\left\{E(K_t^{3/2}) - E(K_t^{1/2})\right\}$. The equation is corrected in errata posted at *The Journal of Finance* website.

[4] In Andersen (1996), equation (l) was written as
$E[(R_t - \bar{r})^2(\hat{V}_t - \overline{V})^2] = c\overline{K}\overline{V} + c^2 m_1 Var(K_t) + c^2 m_1^2[E(K_t - \overline{K})^3 - \overline{K}Var(K_t)]$.
The equation is corrected in errata posted at *The Journal of Finance* website.

The GMM estimator $\hat{\theta}_T$ is obtained by choosing θ to minimize the scalar:

$$Q(\theta,\eta_T) = [g(\theta,\eta_T)]' W_T [g(\theta,\eta_T)] \tag{4}$$

In this case, the sample moments are as close as possible to the population moments. The Newey-West estimator W_T is derived from a sequence of (12 × 12) weighting matrices.

$$W_T = S_T(\theta,\eta_T)^{-1} \tag{5}$$

$$S_T(\theta,\eta_T) = \Gamma_0(\theta,\eta_T) + \sum_{j=1}^{n}(1-\frac{j}{n+1})[\Gamma_j(\theta,\eta_T) + \Gamma_j(\theta,\eta_T)'] \tag{6}$$

$$\Gamma_j(\theta,\eta_T) = \frac{1}{T}\sum_{t=j+1}^{T} h(\theta,\omega_t)h(\theta,\omega_{t-j})' \tag{7}$$

where *n* is a parameter representing the maximal order of autocorrelation for ω_t. The estimator uses Bartlett kernel to smooth the sample autocovariance function and has been shown to be nonnegative definite in finite sample.[5] We choose n = 25 because there is little change in the estimated parameters value when we increase it from 25 to 30.

3. DATA AND SAMPLE SELECTION

3.1 Sample Selection

The sample time period for estimating modified MDH model is from February 1, 1993 through December 29, 2000. In total, there are two thousand daily returns and trading volume. In addition to the SPDRs, ninety securities from the S&P 500 index are selected as sample stocks.[6]

The S&P 500 index consists of 500 stocks chosen for market size, liquidity, and industry group representation. We exclude stocks that were added to or dropped from the S&P 500 index during the sample period.[7] Our initial sample from the S&P 500 index includes 266 common stocks that were active over the entire period from February 1, 1993 through December 29, 2000.

We further reduce the sample to 90 common stocks, selected as follows. For each common stock, we calculate market capitalization as of the end of each year (from 1992 through 2000) and average the market capitalization over the period. Market capitalization is calculated as the stock price times the number of shares outstanding at the end of each year. The stock price and the number of shares outstanding at the end of each year are retrieved from the CRSP daily database. The 266 stocks are ranked according to average market capitalization, and then divide them into three groups. In each group, we pick the median 30 stocks. Of these 90 stocks, 87 are traded on the New York Stock Exchange, and three are traded over-the-counter on the Nasdaq. To eliminate the effect of different trading mechanisms, we substitute NYSE listed stocks of similar size for the three Nasdaq listed stocks. Table 1 lists the sample of selected stocks, which are listed according to firm size ranging from Homestake Mining Co. ($1.97 billions) to Schering-Plough Corp. ($39.55 billions).

[5] Hayashi (2000) reviews and summarizes the properties of the Newey-West estimator.
[6] February 1, 1993 is the second trading day after SPDRs was introduced. Andersen (1996) covers a 19-year period from January 1, 1973 to December 31, 1991.
[7] Standard & Poor's 500 Index composition company lists are obtained from the Standard & Poor's Register of Corporations, Directors and Executives.

3.2 Data Sources

To estimate the modified MDH model, we use daily returns and trading volume data. The raw daily returns and trading volume (as measured by the number of shares traded, adjusted for stock splits) are directly obtained from the Center for Research in Security Prices (CRSP) daily database from February 1, 1993, through December 31, 2000. There are 2,000 observations in total for each common stock.[8]

There are 266 common stocks consistently listed on S&P 500 index over the 8-year sampling period from February 1, 1993 to December 31, 2000. These 266 common stocks are ranked according to average market capitalization, and then divided into three groups. In each group, the median 30 stocks are included in the final sample. Of these 90 stocks, 87 are traded on the NYSE and 3 are traded on the NASDAQ. To eliminate the effect of different trading mechanism, we exchange NASDAQ listed stocks with NYSE listed stocks of similar size. The market capitalization is measured as the average of the stock price times the number of shares outstanding at the end of each year. The stock price and the number of shares outstanding are retrieved from the CRSP database.

Table 1: The Final Sample of the SPDRS Underlying Individual Securities

Group	Ticker	Average Size ($millions)	Group	Ticker	Average Size ($millions)	Group	Ticker	Average Size ($millions)
Small	HM	1,968.74	Middle	LNC	6,073.10	Large	CL	19,462.18
	RBK	1,972.43		SPC	6,094.23		MER	19,728.23
	BC	2,068.17		GD	6,214.49		DOW	20,504.27
	SFA	2,078.96		GP	6,385.05		EMR	21,098.57
	FMC	2,103.65		GT	6,394.04		HON	21,343.38
	U	2,134.31		CGP	6,403.44		FTU	22,527.31
	GR	2,189.81		CSC	6,404.84		KMB	22,940.75
	NMK	2,240.98		AL	6,435.90		BUD	22,961.30
	LIZ	2,287.94		CLX	6,457.77		FRE	23,904.05
	WEN	2,294.39		MHP	6,606.90		TX	24,090.91
	SVU	2,348.04		WWY	6,612.35		MDT	24,769.87
	X	2,460.30		OAT	6,714.82		SLB	24,842.11
	SWK	2,474.08		ETR	6,731.31		CPQ	26,507.12
	LPX	2,489.03		TRB	6,792.91		ONE	26,637.93
	CEN	2,525.91		IPG	6,820.05		TXN	27,473.68
	EC	2,546.89		MRO	6,900.95		TYC	28,635.17
	ASH	2,587.95		FDX	7,029.42		UN	29,393.26
	PLL	2,628.93		AVP	7,048.97		MMM	31,044.36
	DLX	2,650.25		OXY	7,055.25		BA	31,920.18
	BOL	2,672.19		LTD	7,073.01		G	32,393.97
	SUN	2,713.68		RAL	7,162.42		WFC	33,727.82
	MYG	2,779.76		APD	7,222.38		MCD	33,933.06
	W	2,792.34		MAS	7,301.74		CMB	34,642.72
	AMD	2,905.79		TOY	7,358.14		RD	35,161.95
	TIN	2,909.58		AMR	7,361.72		AXP	36,005.87
	ECL	3,018.85		TXT	7,483.83		TWX	36,323.47
	MEA	3,030.19		PEG	7,776.80		GM	37,818.47
	BDK	3,031.95		WMB	7,804.12		NT	38,432.36
	HUM	3,062.92		UCL	7,892.66		MOT	38,705.74
	DDS	3,155.20		AEP	8,253.80		SGP	39,548.24

[8] The trading volume for SPDRs on March 31, 1997 is missing in the CRSP database. A comparison of the CRSP and the Yahoo historical price database indicates that the two report the same daily trading volume, we collect the SPDRs daily trading volume for the missing day from the Yahoo historical price database.

Table 2 reports the summary statistics of the daily return for SPDRs. [9] The return series display excess kurtosis, meaning that extreme 1-day returns are frequently observed. The Ljung-Box Portmanteau statistic for autocorrelation in square daily returns up to 18th order is statistically significant at 1% level, which indicates that the return series display the usual dependency in higher order moments.

The daily returns with dividends are directly obtained from the CRSP database over the period from February 1, 1993 to December 29, 2000. In total, there are 2,000 observations. The Ljung-Box Portmanteau Statistic tests for serial correlation of squared daily returns up to order of 18.

Table 2: Summary Statistics for SPDRS Daily Returns

Median (%)	**Minimum (%)**	**First Quartile (%)**	**Third Quartile (%)**	**Maximum (%)**
0.067	-7.247	-0.418	0.594	5.808
Mean (%)	**Standard Deviation (%)**	**Skewness**	**Kurtosis**	**Ljung-Box Q(18)**
0.066	1.051	-0.137	7.809	433.5**

** significant at the 1% level.

Table 3 reports the summary statistics for raw daily trading volume. Figure 1 depicts that the raw daily trading volume series has a strong, but erratic, trend and has a distinct seasonality as well. The augmented Dickey-Fuller test rejects the null hypothesis of the presence of a unit root in the SRSPs raw daily trading volume. Because the modified MDH model is based on the intensity of information flow, the GMM estimation procedure uses the detrended trading volume.

The raw daily trading volume, as measured by the number of shares traded, is directly obtained from the CRSP database from February 1, 1993 to December 29, 2000. In total, there are 2,000 observations. The Augmented Dickey-Fuller unit-root test (ADF) is used to test the null hypothesis of difference-stationary in the time-series of square root of daily trading volume against the trend-stationary alternative hypothesis.

Table 3: Summary Statistics for SPDRS Raw Daily Trading Volume

Median ($\times 10^5$)	**Minimum ($\times 10^5$)**	**First Quartile ($\times 10^5$)**	**Third Quartile ($\times 10^5$)**	**Maximum ($\times 10^5$)**
15.106	0.052	3.293	57.266	296.038
Mean ($\times 10^5$)	**Standard Deviation ($\times 10^5$)**	**Skewness**	**Kurtosis**	**Unit Root Test Statistics**
34.336	40.410	1.650	6.710	−3.466**

** significance at the 1% level.

To detrend the trading volume time series data, we follow Gallant, Rossi and Tauchen (1992) and use a set of dummy and time-trend variables in the adjustment regression.

1. Day of the week dummy variables (one for each day, Tuesday through Friday). These variables are designed to capture the day of the week effect.
2. Dummy variables for the number of nontrading days preceding the current trading day (dummies for each of 1, 2, and 3 or more nontrading days preceding the current trading day). These dummy variables capture the systematic effects of weekends and holidays.
3. Dummy variables for months of March, April, May, June, July, August, September, October, and November (one for each month).

[9] Due to the space limitation, we do not report summary statistics for ninety stocks in the sample group. The summary statistics for individual ninety sample stocks are similar to those for SPDRs.

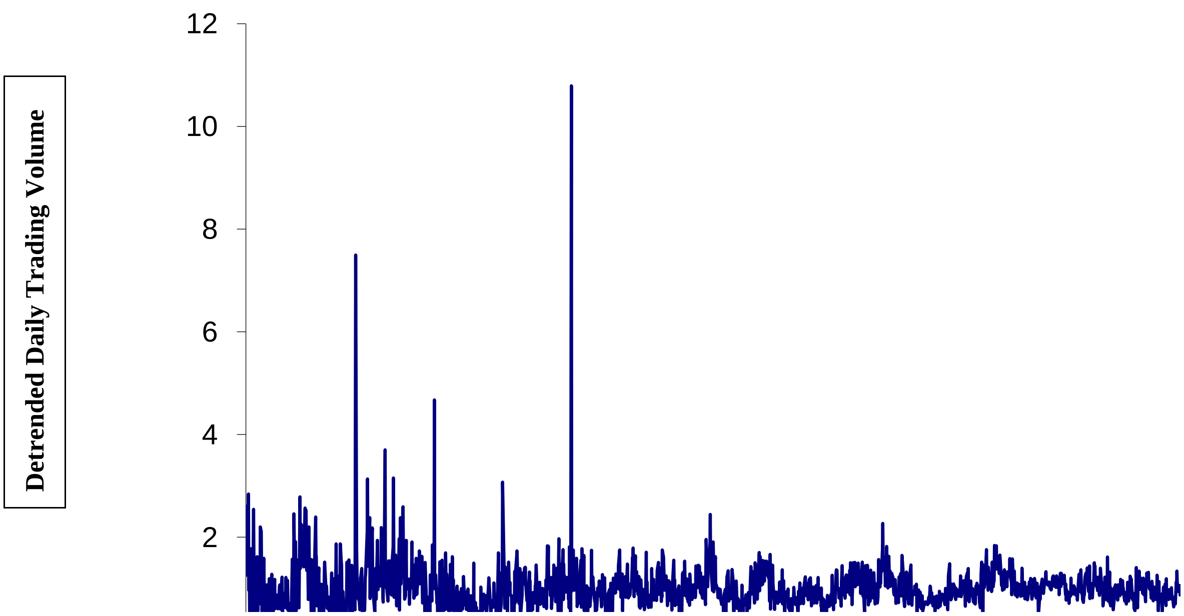
4. Dummy variables for each week of December and January. These variables are designed to accommodate the well-known January effect.
5. Dummy variables for each year (1994 to 2000).
6. A time trend variable (= 1, …, 2000) for all stocks.

We regress the square root of trading volume on this set of dummy and time-trend adjustment variables ($DUMMY_t'$) for SPDRs and the 90 underlying individual sample stocks. The regression model is specified as:

$$Y_t = \sqrt{V_t} = DUMMY_t'\beta + \varepsilon_t \quad (8)$$

Estimates of regression coefficients for a stock are used to produce time series of fitted values $\hat{Y}_t = DUMMY_t'\hat{\beta}$, which are assumed to be due to factors not systematically related to news or information arrival. The detrended volume time series $\hat{V}_t$ is defined as the ratio of square root of volume observation, Y_t over the corresponding fitted trading volume $\hat{Y}_t$. Figure 2 displays the SPDRs' detrended trading volume series, $\hat{V}_t$. There still exist several spikes in the detrended trading volume series. In comparison with the raw daily trading volume shown in Figure 1, the detrended trading volume series has removed the growth trend.

Figure 2: The Detrended Daily Trading Volume for SPDRs over 1993-2000



This figure displays the detrended daily trading volume of SPDRs. The raw daily trading volume (as measured by the number of shares traded, adjusted for stock splits) is directly obtained from the CRSP database over February 1, 1993 to December 29, 2000. The series are detrended by the following steps. First, we regress the square root of trading volume on a set of dummy and time-trend adjustment variables to get the fitted trading volume $\hat{Y}_t$, which is assumed to be due to factors not systematically related to news or information arrival. Then we divide each square root of observed trading volume, Y_t, by the corresponding fitted trading volume, $\hat{Y}_t$, for that day to obtain the detrended trading volume series, $\hat{V}_t = Y_t / \hat{Y}_t$.

4. EMPIRICAL RESULTS

It is hypothesized that the SPDRs market attracts trading volume from uninformed traders because they are subject less to adverse trades with informed traders in this market than in the individual securities markets. We expect more uninformed traders' trading volume in the SPDRs market than in the markets for the underlying individual securities. To estimate the uninformed traders' trading volume, we use Andersen (1996) modified MDH model and Hansen (1982) GMM estimation procedure. There are nine unknown parameters to be estimated and twelve orthogonal moment conditions. The chi-square test for goodness-of-fit has three degrees of freedom, which is based on:

$$\left[\sqrt{T}\cdot g(\hat{\theta}_T,\eta_T)\right]' W_T \left[\sqrt{T}\cdot g(\hat{\theta}_T,\eta_T)\right] \sim \chi^2_3 \tag{9}$$

The GMM estimator $\hat{\theta}_T$ is asymptotically distributed as a normal distribution with mean θ_0 and variance $\hat{\Omega}_T / T$, where $\hat{\Omega}_T = \left(\hat{D}_T \cdot W_T \cdot \hat{D}'_T\right)^{-1}$ and $\hat{D}'_T = \left.\frac{\partial g(\theta,\eta_T)}{\partial \theta'}\right|_{\theta=\hat{\theta}_T}$. The standard error and its corresponding t-statistic for each estimated parameters are derived from the asymptotic distribution.

Table 4 presents the estimation results of four selected parameters and their standard errors are reported in parentheses. The chi-square statistics test for goodness-of-fit of the GMM model. The *p*-value of a chi-square statistics is reported in brackets. The first column in Table 4 uses ticker symbol and CRSP five-digit permanent number to identify individual stocks and SPDRs. The last column in Table 4 reports the estimated percentage of uninformed traders' trading volume over the total daily trading volume.

All estimates of parameters, $\overline{K}$, cm_0 and cm_1, are positive as reported in Table 4. The detrended uninformed traders' daily trading volume, cm_0, is 0.7534 for SPDRs and 0.5762 for sample stock average. The percentage uninformed trading volume, $cm_0/(cm_0 + cm_1\overline{K})$, are 76.19% for SPDRs and 58.22% for the sample stock average.[10] To test the null hypothesis that SPDRs has the same percentage uninformed trading volume as the underlying sample of individual stocks, we compute student t statistic. Specifically, we compare SPDRs estimate of 76.19% with the sample mean of 58.22% from the 90 sample stocks scaled by the sample standard deviation.

The results are based on the daily percentage return and detrended daily volume, adjusted for stock splits, for SPDRs and its 90 underlying sample stocks over the period from February 1, 1993 to December 31, 2000. The following model involving the daily percentage returns, R_t, the detrended volume, $\hat{V}_t$, and the unobserved flow of underlying information arrivals, K_t, was estimated by the GMM methodology:

$R_t \mid K_t \sim N(\bar{r}, K_t)$

$\hat{V}_t \mid K_t \sim c \quad Po(m_0 + m_1 K_t)$.

The parameters vector to be estimated is $\theta = \left(\bar{r}, E\left[K_t^{1/2}\right], \overline{K}, E\left[K_t^{3/2}\right], Var(K_t), c, cm_0, cm_1, E\left(K_t - \overline{K}\right)^3\right)$, where $\bar{r}$ is the mean of the return; m_0 is the daily arrival intensity of noise (uninformed) trading , which is independent of the arrival of information; m_1 measures how strongly volume fluctuates in response to the news; and c is an unknown positive constant which is introduced due to a scaling indeterminacy that arises when detrended volume data are used in the estimation. Estimates are corrected for serially correlated and heteroskedastic errors by the Newey and West (1987) method with 25 lags. The five-digit number below ticker symbol is the CRSP

[10] We did the estimation by using SPDRs' detrended trading volume with outliers and without outliers and the results are very close.

permanent number. The standard errors are reported in parentheses and p-values in brackets. The χ^2-test for goodness-of-fit (Hansen, 1982) has three degrees of freedom.

Table 4: GMM Estimates of Selected Parameters in Modified MDH Model for SPDRS and its 90 Underlying Sample Stocks

Ticker [Permno]	**Parameters Estimate**				χ_3^2	$\frac{cm_0}{cm_0 + cm_1\bar{K}}$
	$\bar{r}$	$\bar{K}$	cm_0	cm_1		
HM	-0.1841	6.6281	0.4067	0.0859	21.8997	0.4167
[12319]	(0.0390)	(0.0247)	(0.0711)	(0.0110)	[0.0001]	
RBK	-0.0662	5.4189	0.7452	0.0472	4.9587	0.7444
[91380]	(0.0471)	(0.4507)	(0.1391)	(0.0287)	[0.1748]	
BC	-0.0161	3.8037	0.5421	0.1157	7.0237	0.5519
[10874]	(0.0438)	(0.2170)	(0.4179)	(0.1154)	[0.0711]	
SFA	-0.0174	11.1264	0.5582	0.0390	13.4409	0.5626
[45671]	(0.0574)	(0.5379)	(0.2803)	(0.0272)	[0.0038]	
FMC	-0.0054	1.7457	0.5673	0.2456	8.9542	0.5696
[19166]	(0.0325)	(0.0576)	(0.2753)	(0.1651)	[0.0299]	
U	0.0000	11.0287	0.6578	0.0306	80.3354	0.6608
[28847]	(0.0350)	(0.0839)	(0.0175)	(0.0015)	[0.0000]	
GR	0.0263	2.5871	0.7553	0.0895	17.7436	0.7654
[12140]	(0.0360)	(0.2223)	(0.1797)	(0.0754)	[0.0005]	
NMK	0.0299	2.5692	0.6177	0.1426	7.9423	0.6278
[24184]	(0.0295)	(0.1903)	(0.3726)	(0.1541)	[0.0472]	
LIZ	-0.0050	4.4581	0.6742	0.0712	8.1472	0.6798
[49905]	(0.0392)	(0.2363)	(0.1986)	(0.0478)	[0.0431]	
WEN	0.0043	3.1008	0.3837	0.1934	9.1583	0.3902
[63060]	(0.0289)	(0.0684)	(0.3554)	(0.1188)	[0.0273]	
SVU	-0.0050	2.0787	0.6740	0.1558	12.4754	0.6754
[44951]	(0.0313)	(0.0978)	(0.2917)	(0.1472)	[0.0059]	
X	-0.1482	4.0850	0.2188	0.1869	22.7549	0.2228
[76644]	(0.0397)	(0.0195)	(0.1600)	(0.0405)	[0.0000]	
SWK	-0.0058	2.9095	0.6881	0.1037	7.8311	0.6952
[43350]	(0.0343)	(0.1901)	(0.2254)	(0.0826)	[0.0496]	
LPX	-0.0982	4.9382	0.4218	0.1146	11.5219	0.4270
[56223]	(0.0409)	(0.0776)	(0.2487)	(0.0529)	[0.0092]	
CEN	0.0593	4.6842	0.7063	0.0600	9.5283	0.7154
[38914]	(0.0369)	(0.3059)	(0.2001)	(0.0463)	[0.0230]	
EC	0.0041	4.4608	0.7238	0.0612	2.8377	0.7260
[62834]	(0.0361)	(0.2576)	(0.1212)	(0.0294)	[0.4173]	
ASH	-0.0394	1.6552	0.4568	0.3141	20.7241	0.4677
[24272]	(0.0261)	(0.0083)	(0.0910)	(0.0551)	[0.0001]	
PLL	0.0249	3.4650	0.6063	0.1140	11.3025	0.6055
[35051]	(0.0299)	(0.05340	(0.1511)	(0.0454)	[0.0102]	
DLX	-0.0228	2.0808	0.6452	0.1687	12.1277	0.6477
[61743]	(0.0282)	(0.1787)	(0.4883)	(0.2483)	[0.0070]	
BOL	0.0497	3.7003	0.7847	0.0575	1.7259	0.7866
[26518]	(0.0348)	(0.2911)	(0.0900)	(0.0272)	[0.6312]	
SUN	-0.0134	2.8695	0.5060	0.1692	14.4304	0.5103
[14656]	(0.0350)	(0.1112)	(0.5186)	(0.1873)	[0.0024]	
MYG	0.0464	3.5482	0.7529	0.0654	11.3873	0.7645
[13119]	(0.0375)	(0.2467)	(0.1352)	(0.0424)	[0.0098]	
W	-0.0086	2.2489	0.5935	0.1757	21.1565	0.6003
[21186]	(0.0326)	(0.0639)	(0.3841)	(0.1756)	[0.0001]	

Table 4 (continued)

Ticker [Permno]	**Parameters Estimate**				χ_3^2	$\frac{cm_0}{cm_0 + cm_1\overline{K}}$
	$\bar{r}$	$\overline{K}$	cm_0	cm_1		
AMD	-0.0570	11.7662	0.5812	0.0327	18.3386	0.6017
[61241]	(0.0614)	(0.1806)	(0.1103)	(0.6952)	[0.0004]	
TIN	-0.0195	2.6601	0.4884	0.1879	15.0961	0.4943
[66114]	(0.0348)	(0.0984)	(0.6659)	(0.2570)	[0.0017]	
ECL	0.0335	1.8453	0.6065	0.2015	17.7474	0.6199
[70578]	(0.0265)	(0.0584)	(0.2681)	(0.1500)	[0.0005]	
MEA	-0.0494	2.9382	0.5048	0.1664	16.5736	0.5079
[19895]	(0.0344)	(0.0142)	(0.1172)	(0.0109)	[0.0009]	
BDK	-0.0085	3.4890	0.6065	0.1066	17.5255	0.6200
[20220]	(0.0364)	(0.28550	(0.3847)	(0.1193)	[0.0006]	
HUM	0.0081	7.4510	0.4798	0.0668	19.0810	0.4907
[48653]	(0.0438)	(0.2378)	(0.0219)	(0.0382)	[0.0003]	
DDS	-0.0496	4.2139	0.7122	0.0662	8.6465	0.7185
[49429]	(0.0354)	(0.34190	(0.2427)	(0.0627)	[0.0344]	
LNC	0.0182	2.5631	0.7624	0.0894	10.1484	0.7689
[49015]	(0.0332)	(0.15950	(0.1717)	(0.0719)	[0.0173]	
SPC	-0.0060	2.3312	0.6712	0.1369	9.9204	0.6778
[59459]	(0.0306)	(0.1644)	(0.3357)	(0.1530)	[0.0193]	
GD	0.0408	1.7356	0.5466	0.2531	19.7726	0.5544
[12052]	(0.0273)	(0.1067)	(0.5811)	(0.3509)	[0.0002]	
GP	-0.0578	3.2371	0.5022	0.1497	20.2078	0.5089
[23915]	(0.0381)	(0.02150	(0.0806)	(0.0267)	[0.0002]	
GT	-0.0281	3.1344	0.6392	0.1145	8.7258	0.6404
[16432]	(0.0329)	(0.0644)	(0.1308)	(0.0441)	[0.0332]	
CGP	0.0829	2.4079	0.3074	0.2848	17.4570	0.3095
[38893]	(0.0314)	(0.0061)	(0.0661)	(0.0278)	[0.0006]	
CSC	0.0709	3.6260	0.4019	0.1624	18.9338	0.4056
[40125]	(0.0358)	(0.0644)	(0.3413)	(0.0980)	[0.0003]	
AL	0.0019	2.5787	0.7935	0.0681	23.2267	0.8188
[24264]	(0.0342)	(0.0444)	(0.0609)	(0.0235)	[0.0000]	
CLX	0.0451	2.5529	0.7567	0.0912	10.3186	0.7648
[46578]	(0.0295)	(0.3164)	(0.2556)	(0.1107)	[0.0160]	
MHP	0.0576	1.7740	0.6958	0.1636	23.2009	0.7057
[17478]	(0.0249)	(0.0360)	(0.1304)	(0.0762)	[0.0000]	
WWY	0.0214	2.1174	0.6084	0.1776	9.7852	0.6180
[15472]	(0.0280)	(0.0442)	(0.1835)	(0.0903)	[0.0205]	
OAT	-0.0014	2.5143	0.6205	0.1430	14.9554	0.6331
[24539]	(0.0260)	(0.1389)	(0.2365)	(0.1014)	[0.0019]	
ETR	0.0390	1.9922	0.8414	0.0787	5.1383	0.8429
[24010]	(0.0313)	(0.1235)	(0.0651)	(0.0371)	[0.1619]	
TRB	0.0671	2.4125	0.6588	0.1407	3.8520	0.6600
[65787]	(0.0309)	(0.0532)	(0.1460)	(0.0642)	[0.2779]	
IPG	0.0554	2.6346	0.6110	0.1457	17.3530	0.6142
[53065]	(0.0303)	(0.0564)	(0.2005)	(0.0789)	[0.0006]	
MRO	-0.0571	3.3479	0.3060	0.2019	18.2787	0.3117
[15069]	(0.0290)	(0.0085)	(0.0479)	(0.0147)	[0.0004]	
FDX	0.0161	3.9547	0.5813	0.1047	13.4346	0.5840
[60628]	(0.0355)	(0.0818)	(0.1658)	(0.0446)	[0.0038]	
AVP	0.0729	3.6676	0.7415	0.0693	4.8470	0.7447
[40416]	(0.0381)	(0.4485)	(0.2617)	(0.0788)	[0.1833]	
OXY	-0.0201	2.3032	0.3915	0.2616	17.6539	0.3938
[34833]	(0.0313)	(0.0159)	(0.1620)	(0.0726)	[0.0005]	

Table 4 (continued)

Ticker [Permno]	Parameters Estimate				χ^2_3	$\frac{cm_0}{cm_0 + cm_1\bar{K}}$
	$\bar{r}$	$\bar{K}$	cm_0	cm_1		
LTD	-0.0468	4.1879	0.3582	0.1501	15.7161	0.3630
[64282]	(0.0385)	(0.0227)	(0.1458)	(0.0354)	[0.0013]	
RAL	0.0270	1.9902	0.5376	0.2253	17.4189	0.5453
[28353]	(0.0258)	(0.0251)	(0.1890)	(0.0980)	[0.0006]	
APD	0.0255	2.7092	0.6836	0.1142	9.7018	0.6883
[28222]	(0.0326)	(0.1014)	(0.2521)	(0.0971)	[0.0213]	
MAS	-0.0261	3.1206	0.6810	0.0979	19.4579	0.6904
[34032]	(0.0345)	(0.1134)	(0.2604)	(0.0871)	[0.0002]	
TOY	-0.0677	3.5503	0.5433	0.1271	15. 8310	0.5462
[61065]	(0.0396)	(0.0754)	(0.2480)	(0.0729)	[0.0012]	
AMR	0.0019	3.9830	0.2832	0.1774	15.3561	0.2861
[21020]	(0.0337)	(0.0163)	(0.1001)	(0.0260)	[0.0015]	
TXT	-0.0007	1.7915	0.5947	0.2178	18.3850	0.6038
[23579]	(0.0298)	(0.0119)	(0.13740	(0.0780)	[0.0004]	
PEG	0.0459	1.2781	0.4677	0.4149	5.8934	0.4687
[23712]	(0.0269)	(0.0930)	(1.0874)	(0.8790)	[0.1169]	
WMB	0.0440	3.3621	0.6818	0.0915	16.8571	0.6891
[38156]	(0.0309)	(0.1837)	(0.2436)	(0.0770)	[0.0008]	
UCL	-0.0258	3.0734	0.4924	0.1616	14.4513	0.4979
[14891]	(0.0304)	(0.0086)	(0.0599)	(0.0195)	[0.0024]	
AEP	0.0173	1.1828	0.7256	0.2321	9.4424	0.7255
[24109]	(0.0247)	(0.0627)	(0.2152)	(0.1910)	[0.0240]	
CL	0.1127	2.9323	0.6812	0.1070	2.9214	0.6847
[18729]	(0.0309)	(0.1966)	(0.2786)	(0.1011)	[0.4039]	
MER	0.0839	5.2908	0.1461	0.1591	11.4766	0.1478
[52919]	(0.0370)	(0.0127)	(0.0694)	(0.0137)	[0.0094]	
DOW	0.0126	2.1218	0.6757	0.1514	9.3192	0.6777
[20626]	(0.0274)	(0.1552)	(0.3936)	(0.1962)	[0.0253]	
EMR	0.0028	2.0802	0.7157	0.1311	15.7630	0.7241
[22103]	(0.0248)	(0.0431)	(0.1596)	(0.0791)	[0.0013]	
HON	0.0845	3.6371	0.8182	0.0487	1.9647	0.8221
[10145]	(0.0272)	(0.3044)	(0.1274)	(0.0388)	[0.5798]	
FTU	0.0464	2.0405	0.4803	0.2510	14.2394	0.4840
[36469]	(0.0309)	(0.0359)	(0.3838)	(0.1921)	[0.0026]	
KMB	0.0622	2.5866	0.4994	0.1894	8.0385	0.5048
[17750]	(0.0288)	(0.0669)	(0.3387)	(0.1368)	[0.0452]	
BUD	0.0706	1.8084	0.5706	0.2336	10.2257	0.5746
[59184]	(0.0194)	(0.0088)	(0.0635)	(0.0363)	[0.0167]	
FRE	0.0497	3.3606	0.4920	0.1514	17.4743	0.4916
[75789]	(0.0310)	(0.0149)	(0.0763)	(0.0235)	[0.0006]	
TX	0.0108	1.9288	0.5628	0.2225	14.6740	0.5674
[14736]	(0.0210)	(0.0147)	(0.1046)	(0.0561)	[0.0021]	
MDT	0.1338	4.1945	0.3778	0.1461	11.7979	0.3814
[60097]	(0.0349)	(0.0180)	(0.1189)	(0.0294)	[0.0081]	
SLB	-0.0078	3.6264	0.4704	0.1437	20.7126	0.4745
[14277]	(0.0318)	(0.0094)	(0.0435)	(0.0121)	[0.0001]	
CPQ	0.0599	8.8211	0.4076	0.0661	14.5610	0.4114
[68347]	(0.0444)	(0.0639)	(0.1188)	(0.0142)	[0.0022]	
ONE	0.0088	2.9123	0.4807	0.1724	11.6933	0.4892
[65138]	(0.0320)	(0.0169)	(0.0590)	(0.0183)	[0.0085]	

Table 4 (continued)

Ticker [Permno]	**Parameters Estimate** $\bar{r}$	$\bar{K}$	cm_0	cm_1	χ_3^2	$\frac{cm_0}{cm_0 + cm_1\bar{K}}$
TXN	0.1591	9.3952	0.4507	0.0573	14.4776	0.4556
[15579]	(0.0498)	(0.0368)	(0.0662)	(0.0074)	[0.0023]	
TYC	0.0935	3.4289	0.7464	0.0680	12.9882	0.7620
[45356]	(0.0284)	(0.3425)	(0.2149)	(0.0679)	[0.0047]	
UN	0.0625	2.1796	0.8039	0.0825	10.0528	0.8173
[28310]	(0.0279)	(0.2114)	(0.1912)	(0.0941)	[0.0181]	
MMM	0.0616	2.0906	0.6573	0.1611	10.0347	0.6612
[22592]	(0.0227)	(0.1278)	(0.3576)	(0.1811)	[0.0183]	
BA	0.0187	2.9318	0.6332	0.1184	19.3103	0.6460
[19561]	(0.0286)	(0.0467)	(0.0882)	(0.0327)	[0.0002]	
G	0.0789	3.1317	0.6651	0.1051	7.4413	0.6690
[16424]	(0.0298)	(0.1370)	(0.2011)	(0.0684)	[0.0591]	
WFC	0.0632	2.9968	0.4248	0.1893	9.5175	0.4282
[38703]	(0.0306)	(0.0092)	(0.0853)	(0.0285)	[0.0231]	
MCD	0.0703	2.4032	0.5421	0.1878	11.6897	0.5456
[43449]	(0.0255)	(0.0129)	(0.0857)	(0.0371)	[0.0085]	
CMB	0.0714	3.6754	0.3525	0.1736	59.6976	0.3558
[47896]	(0.0304)	(0.0255)	(0.2753)	(0.0761)	[0.0000]	
RD	0.0705	1.8074	0.6860	0.1662	11.6324	0.6955
[25267]	(0.0239)	(0.1669)	(0.5552)	(0.3201)	[0.0088]	
AXP	0.0543	3.8483	0.4324	0.1407	17.0747	0.4440
[59176]	(0.0258)	(0.0102)	(0.0468)	(0.0130)	[0.0007]	
TWX	0.0289	4.9219	0.8009	0.0395	5.5498	0.8046
[40483]	(0.0318)	(0.4285)	(0.0099)	(0.0226)	[0.1357]	
GM	-0.0132	3.3157	0.2206	0.0095	17.6719	0.2247
[12079]	(0.0315)	(0.0081)	(0.0972)	(0.0305)	[0.0005]	
NT	0.1711	6.3696	0.7984	0.0314	3.3666	0.7995
[58640]	(0.0450)	(0.3818)	(0.0757)	(0.0134)	[0.3385]	
MOT	0.0964	5.9468	0.5650	0.0721	10.6849	0.5684
[22779]	(0.0407)	(0.0685)	(0.1208)	(0.0209)	[0.0136]	
SGP	0.0984	3.1523	0.5511	0.1395	51.1323	0.5561
[25013]	(0.0405)	(0.0371)	(0.1072)	(0.0348)	[0.0000]	
SPY	0.0869	0.8797	0.7534	0.2676	9.3860	0.7619
[84398]	(0.0158)	(0.0802)	(0.3114)	(0.3726)	[0.0246]	

Table 5, Panel A, reports a student t statistic of −11.25 indicating that SPDRs have a higher percentage uninformed trading volume. The average daily return volatility $\bar{K}$ representing the unobserved flow of underlying information is much higher for individual sample stocks than the SPDRs. The result is consistent with the observation that the determinants of SPDRs market are more related to public macroeconomic news than private firm valuation information.

The 90 underlying stocks are sampled according to the size of the 266 companies in the S&P 500 index. Table 5, Panel B, groups the uninformed trading volume into three subgroups based on size. There is no obvious relationship between firm size and the percentage uninformed trading volume. We also run a least-square regression model using the estimated percentage uninformed trading volume as the dependent variable and the logarithm of firm size as the independent variable. The coefficient on firm size is −0.2204 and the t-value is –1.02, which is not significantly different from zero. Overall, we find no evidence to indicate that the size of a firm has an effect on the estimated percentage uninformed trading volume.

Uninformed Trading Volume are estimated by using GMM methodology and Andersen (1996)'s modified MDH model:

$R_t \mid K_t \sim N(\bar{r}, K_t)$ and $\hat{V}_t \mid K_t \sim c \cdot Po(m_0 + m_1 K_t)$

where R_t is daily percentage returns, $\hat{V}_t$ is the detrended trading volume, K_t is the unobserved flow of underlying information arrivals. The parameter vector to be estimated is $\theta = \left(\bar{r}, E\left[K_t^{1/2}\right], \bar{K}, E\left[K_t^{3/2}\right], Var(K_t), c, cm_0, cm_1, E\left(K_t - \bar{K}\right)^3\right)$, where $\bar{r}$ is the mean of the return; m_0 is the daily arrival intensity of uninformed trading, which is independent of the arrival of information; m_1 measures how strongly volume fluctuates in response to the news; and *c* is an unknown positive scaling constant. The percentage uninformed trading volume is measured by the ratio $cm_0 / (cm_0 + cm_1 \bar{K})$ and $\bar{K}$ is the population mean of the daily return volatility.

Table 5

Panel A: Percentage Uninformed Trading Volume for SPDRS and 90 Underlying Sample Stocks

	Mean	Minimum	First Quartile	Median	Third Quartile	Maximum
Percentage Uninformed Trading Volume						
Underlying Sample Stocks	0.5822	0.148	0.490	0.603	0.690	0.843
SPDRs	0.7619					
Difference	−0.1797					
Student t statistic	−11.25**					
Daily Return Volatility ($\bar{K}$)						
Underlying Sample Stocks	3.55	1.183	2.310	3.087	3.837	11.766
SPDRs	0.88					

** significant at the 1% level.

Panel B: Underlying Sample Stocks' Percentage Uninformed Trading Volume in Size Subgroup

Security	*Average Size ($millions)*	Percentage Uninformed Trading Volume					
		Mean	Maximum	3rd Quartile	Median	1st Quartile	Minimum
SPDRs	3,507.47	0.7619					
Underlying Stocks:							
Small Group	2,537.44	0.5956	0.7866	0.6914	0.6127	0.5085	0.2228
Median Group	6,928.88	0.5887	0.8429	0.6901	0.6161	0.5006	0.2861
Large Group	28,882.61	0.5623	0.8221	0.6830	0.5617	0.4603	0.1478
Full Sample	12,782.98	0.5741	0.8429	0.6901	0.6028	0.4896	0.1478

5. CONCLUSION

The successful introduction and impressive growth of SPDRs in the last decade raises the question why a composite security like SPDRs attracts so many investors and trading volume. Individual investors obviously can choose buying and selling their underlying securities that compose the indexes in the same proportions to get the same cash flow. Subrahmanyam (1991) and Gorton and Pennacchi (1993) propose theories to justify the presence of composite securities and explain why they are popular. Due to the "diversification" or "information offset" effect, the introduction of ETFs product, like SPDRs, reduces the informed traders' information advantage and the uninformed traders would face less adverse selection problem in this market than in the market of individual securities.

Relying on a modified mixture distribution hypothesis model proposed by Andersen (1996), this study applies generalized method of moments to estimate the unobservable percentage of trading volume attributable to uninformed trades. Using ninety securities selected from the S&P 500 index as sample stocks for comparison, we find that the Spiders market indeed attracts a relatively higher percentage of trading volume from uninformed traders.

AUTHOR INFORMATION

Quentin C. Chu earned his Ph.D. at the University of Illinois at Urbana-Champaign. Currently he is a Professor of Finance at the University of Memphis. He has a long-standing interest in both applied and theoretical issues in finance. His current research interests include Treasury inflation-indexed securities and exchange traded funds. He has numerical journal publications including the *Journal of Finance,* the *Journal of Futures Markets,* the *Journal of Fixed Income*, etc.

Dan Zhou earned her Ph.D. at University of Memphis. Currently she is an assistant professor of finance at California State University-Bakersfield. Her research interests include market microstructure, investment and empirical corporate finance. She has published in peer review journals and made presentations at international/national/regional conferences.

REFERENCES

1. Ackert, L. F., and Tian, Y. S., "Arbitrage and valuation in the market for Standard and Poor's depository receipts", *Financial Management*, Vol.29, 2000, 71-88.
2. Andersen, T. G., "Return volatility and trading volume: An information flow interpretation of stochastic volatility", *Journal of Finance*, Vol. 51, 1996, 169-204.
3. Clark, P. K., "A subordinated stochastic process model with finite variance for speculative prices", *Econometrica*, Vol.41, 1973, 135-155.
4. Elton, E. J., Gruber, M. J., Comer, G., and Li, K., "Spiders: Where are the bugs?" *Journal of Business*, Vol. 75, 2002, 453-472.
5. Gallant, A., Rossi, P. E., and Tauchen, G., "Stock prices and volume", *Review of Financial Studies*, Vol. 5, 1992, 199-242.
6. Glosten, L. R. and Milgrom, P. R., "Bid, ask and transaction prices in a specialist market with heterogeneously informed traders", *Journal of Financial Economics*, Vol.14, 1985, 71-100.
7. Gorton, G. B. and Pennacchi, G. G., "Security baskets and index-linked securities", *Journal of Business*, Vol.66, 1993, 1-27.
8. Hansen, L. P., "Large sample properties of generalized methods of moments estimators", *Econometrica*, Vol. 50, 1982, 1029-1054.
9. Hayashi, F., *Econometrics* (Princeton University Press, Princeton, NJ), 2000.
10. Newey, W. K. and West, K. D., "A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix", *Econometrica*, Vol.55, 1987, 703-708.
11. Poterba, J. M. and Shoven, J. B., "Exchange traded funds: A new investment option for taxable investors", *American Economic Review*, Vol.92, 2002, 422-427.
12. Subrahmanyam, A., "A theory of trading in stock index futures", *Review of Financial Studies*, Vol.4, 1991, 17-51.

APPENDIX I

Derivation of Unconditional Moment Equations Used in GMM Procedure

Andersen's (1996) "modified mixture distribution hypothesis" (MMDH) model can be represented as:

$$R_t \vert K_t \sim N(\bar{r},\ K_t)$$

$$\hat{V}_t \vert K_t \sim cPo(m_0 + m_1 K_t)$$

and

$$Cov\left(R_t,\ \hat{V}_t\right) = 0$$

$$Cov\left(R_t^2,\ \hat{V}_t\right) = m_1 \operatorname{var}(K_t) > 0$$

The twelve unconditional moment equations used in the GMM procedure are derived as follows:

(a) $$E(R_t) = \underset{K_t}{E}\left\{E\left[R_t \vert K_t\right]\right\} = E(\bar{r}) = \bar{r}$$

(b) If $x \sim N(0,1)$, then

$$E|x| = \int_{-\infty}^{\infty} |x| \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx = 2\int_{0}^{\infty} x \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx = \sqrt{2/\pi}\int_{0}^{\infty} x e^{-\frac{1}{2}x^2} dx$$

Let $u = \frac{1}{2}x^2$, then $du = xdx$

$$E|x| = \sqrt{2/\pi}\int_{0}^{\infty} e^{-u} du = \sqrt{2/\pi}\left(-e^{-u}\right)\Big|_{0}^{\infty} = \sqrt{2/\pi}$$

Because $R_t \vert K_t \sim N(\bar{r}, K_t)$, then $\left.\frac{R_t - \bar{r}}{\sqrt{K_t}}\right| K_t \sim N(0,1)$

$$E|R_t - \bar{r}| = \underset{K_t}{E}\left\{E\left[\,|R_t - \bar{r}|\,\big| K_t\right]\right\} = E\left(\sqrt{2/\pi}\ \sqrt{K_t}\right) = \sqrt{2/\pi}\, E\left[K_t^{1/2}\right]$$

(c) Because $\left.\frac{R_t - \bar{r}}{\sqrt{K_t}}\right| K_t \sim N(0,1)$, then

$$\left(\left.\frac{R_t - \bar{r}}{\sqrt{K_t}}\right| K_t\right)^2 = \left.\frac{(R_t - \bar{r})^2}{K_t}\right| K_t \sim \chi_1^2, \text{ and } E\left\{\left.\frac{(R_t - \bar{r})^2}{K_t}\right| K_t\right\} = 1$$

$$E[(R_t - \bar{r})^2] = \underset{K_t}{E}\left\{E[(R_t - \bar{r})^2 \vert K_t]\right\} = E(K_t) = \bar{K}$$

(d) Because $R_t \left| K_t \right. \sim N(\bar{r}, K_t)$, then, $\left. \frac{R_t - \bar{r}}{\sqrt{K_t}} \right| K_t \sim N(0,1)$

$$x = \left. \frac{R_t - \bar{r}}{\sqrt{K_t}} \right| K_t \sim N(0,1)$$

$$E|x|^3 = \int_{-\infty}^{\infty} \left|x^3\right| \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx = 2\int_0^{\infty} x^3 \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx = \sqrt{2/\pi}\int_0^{\infty} x^3 e^{-\frac{1}{2}x^2} dx$$

Let $u = \frac{1}{2}x^2$, then $du = xdx$

$$E|x|^3 = \sqrt{2/\pi}\int_0^{\infty} 2ue^{-u} du = 2\sqrt{2/\pi}\,\Gamma(2) = 2\sqrt{2/\pi}$$

where $\Gamma(2)$ is Γ function.[11]

$$E|R_t - \bar{r}|^3 = \underset{K_t}{E}\left\{E\left[\left|R_t - \bar{r}\right|^3 \middle| K_t\right]\right\} = E\left(2\sqrt{2/\pi} \cdot K_t^{3/2}\right) = 2\sqrt{2/\pi}E\left[K_t^{3/2}\right]$$

(e) $\left. \frac{R_t - \bar{r}}{\sqrt{K_t}} \right| K_t \sim N(0,1) \quad \Rightarrow \quad x = \left(\left. \frac{R_t - \bar{r}}{\sqrt{K_t}} \right| K_t \right)^2 = \left. \frac{(R_t - \bar{r})^2}{K_t} \right| K_t \sim \chi_1^2$

$$E\left\{ \left. \frac{(R_t - \bar{r})^4}{K_t^2} \right| K_t \right\} = E(x^2) = \left(E(x)\right)^2 + Var(x) = 1^2 + 2 = 3$$

$$E[(R_t - \bar{r})^4] = \underset{K_t}{E}\left\{E[(R_t - \bar{r})^4] \mid K_t\ ]\right\} = 3E\left(K_t^2\right) = 3\left[(\bar{K})^2 + Var(K_t)\right]$$

(f) $E(\hat{V}_t) = E\{E(\hat{V}_t \left| K_t \right.)\}_{K_t} = E\left[c(m_0 + m_1 K_t)\right] = c(m_0 + m_1 \bar{K}) = \bar{V}$

(g)
$$\begin{aligned} E\left[(\hat{V}_t - \bar{V})^2\right] &= Var\left(\hat{V}_t\right) = E\left\{Var(\hat{V}_t \left| K_t \right.)\right\} + Var\left\{E(\hat{V}_t \left| K_t \right.)\right\} \\ &= E\left\{c^2(m_0 + m_1 K_t)\right\} + Var\left\{c(m_0 + m_1 K_t)\right\} \\ &= c^2(m_0 + m_1\bar{K}) + c^2 m_1^2 Var(K_t) \\ &= c \cdot c(m_0 + m_1\bar{K}) + c^2 m_1^2 Var(K_t) \\ &= c\bar{V} + c^2 m_1^2 Var(K_t) \end{aligned}$$

[11] Γ function: $\Gamma(r) = \int_0^{\infty} u^{r-1} e^{-u} du$, $\Gamma(r+1) = r\Gamma(r)(r > 0)$, $\Gamma(1) = \int_0^{\infty} e^{-u} du = 1$.

(h) $E\left[(\hat{V}_t - \overline{V})^3\right] = E\left(\hat{V}_t - E(\hat{V}_t \middle| K_t) + E(\hat{V}_t \middle| K_t) - \overline{V}\right)^3$

$$= E\left[\hat{V}_t - E(\hat{V}_t \middle| K_t)\right]^3 + 3E\left\{\left[\hat{V}_t - E(\hat{V}_t \middle| K_t)\right]^2 \left[E(\hat{V}_t \middle| K_t) - \overline{V}\right]\right\}$$

$$+ 3E\left\{\left[\hat{V}_t - E(\hat{V}_t \middle| K_t)\right]\left[E(\hat{V}_t \middle| K_t) - \overline{V}\right]^2\right\} + E\left[E(\hat{V}_t \middle| K_t) - \overline{V}\right]^3$$

$$= E[c^3(m_0 + m_1 K_t)] + 3E\left\{c^2(m_0 + m_1 K_t)[c(m_0 + m_1 K_t) - c(m_0 + m_1 \overline{K})]\right\}$$

$$+ 3\{[E(\hat{V}_t) - E[E(\hat{V}_t \mid K_t)]]E[(E(\hat{V}_t \mid K_t) - \overline{V})^2]\}$$

$$+ E\left\{[c(m_0 + m_1 K_t) - c(m_0 + m_1 \overline{K})]^3\right\}$$

$$= E[c^3(m_0 + m_1 K_t)] + 3E\left\{c^2(m_0 + m_1 K_t)\cdot(cm_1 K_t - cm_1 \overline{K})\right\}$$

$$+ 3\{[E(\hat{V}_t) - E(\hat{V}_t)]E[(E(\hat{V}_t \mid K_t) - \overline{V})^2]\} + E\left\{[cm_1 K_t - cm_1 \overline{K}]^3\right\}$$

$$= E[c^3(m_0 + m_1 K_t)] + 3E\left\{c^3 m_1 m_0 (K_t - \overline{K}) + c^3 m_1^2 K_t \cdot (K_t - \overline{K})\right\}$$

$$+ 0 + E\left\{[cm_1 K_t - cm_1 \overline{K}]^3\right\}$$

$$= c^3(m_0 + m_1 \overline{K}) + 3c^3 m_1^2 E\left\{K_t (K_t - \overline{K})\right\} + c^3 m_1^3 E[K_t - \overline{K}]^3$$

$$= c^3(m_0 + m_1 \overline{K}) + 3c^3 m_1^2 [E(K_t^2) - \overline{K}^2] + c^3 m_1^3 E[K_t - \overline{K}]^3$$

$$= c^2 \overline{V} + 3c^3 m_1^2 Var(K_t) + c^3 m_1^3 E[K_t - \overline{K}]^3$$

(i) Because $Cov(X,Y) = E(XY) - E(X)E(Y)$, then

$$E[R_t \hat{V}_t] = Cov(R_t, \hat{V}_t) + E(R_t)E(\hat{V}_t) = 0 + \bar{r}\overline{V}$$

$$= \bar{r}\overline{V}$$

(j) $E[\mid R_t - \bar{r} \mid (\hat{V}_t - \overline{V})]$

$$= E[\mid R_t - \bar{r} \mid \hat{V}_t] - \overline{V}\, E \mid R_t - \bar{r} \mid$$

$$= E\{E[\mid R_t - \bar{r} \mid \hat{V}_t \mid K_t]\} - \overline{V}\sqrt{2/\pi}E(K_t^{1/2})$$

$$= E\{\sqrt{2/\pi}\; K_t^{1/2}\; c(m_0 + m_1 K_t)\} - c(m_0 + m_1 \overline{K})\sqrt{2/\pi}\; E(K_t^{1/2})$$

$$= c\sqrt{2/\pi}\; E\{K_t^{1/2}(m_0 + m_1 K_t)\} - c(m_0 + m_1 \overline{K})\sqrt{2/\pi}\; E(K_t^{1/2})$$

$$= c\sqrt{2/\pi}\left\{m_0 E(K_t^{1/2}) + m_1 E(K_t^{3/2}) - m_0 E(K_t^{1/2}) - m_1 \overline{K} E(K_t^{1/2})\right\}$$

$$= c\sqrt{2/\pi}\; m_1\left\{E(K_t^{3/2}) - \overline{K}E(K_t^{1/2})\right\}$$

(k) Because $Cov\left(R_t^2, \hat{V}_t\right) = m_1 \, \text{var}(K_t) > 0$, and $Cov\left(R_t^2, \hat{V}_t\right) = E(R_t^2 \hat{V}_t) - E(R_t^2) E(\hat{V}_t)$

$$E(R_t^2 \hat{V}_t) = Cov\left(R_t^2, \hat{V}_t\right) + E(R_t^2) E(\hat{V}_t) = m_1 Var(K_t) + E(R_t^2) E(\hat{V}_t)$$

$$\begin{aligned} E[(R_t - \bar{r})^2 \hat{V}_t] &= E[(R_t^2 - 2R_t \bar{r} + \bar{r}^2)\hat{V}_t] \\ &= E(R_t{}^2 \hat{V}_t) - 2\bar{r} E(R_t \hat{V}_t) + \bar{r}^2 E(\hat{V}_t) \\ &= m_1 Var(K_t) + E(R_t^2) E(\hat{V}_t) - 2\bar{r} \; \bar{r}\bar{V} + \bar{r}^2 \; \bar{V} \\ &= m_1 Var(K_t) + [Var(R_t) + (E(R_t))^2] \, \bar{V} - \bar{r}^2 \; \bar{V} \\ &= m_1 Var(K_t) + [\bar{K} + \bar{r}^2] \, \bar{V} - \bar{r}^2 \; \bar{V} \\ &= m_1 Var(K_t) + \bar{K}\bar{V} \end{aligned}$$

(l)

$$\begin{aligned} & E[(R_t - \bar{r})^2 (\hat{V}_t - \bar{V})^2] \\ &= \underset{K_t}{E}\left\{ E[(R_t - \bar{r})^2 \mid K_t] \, E[(\hat{V}_t - \bar{V})^2 \mid K_t] \right\} \\ &= \underset{K_t}{E}\left\{ K_t \;\; E[(\hat{V}_t - E(\hat{V}_t \mid K_t) + E(\hat{V}_t \mid K_t) - \bar{V})^2 \mid K_t] \right\} \\ &= \underset{K_t}{E}\{ K_t \;\; [E[(\hat{V}_t - E(\hat{V}_t \mid K_t))^2 \mid K_t] + 2E\left[(\hat{V}_t - E(\hat{V}_t \mid K_t)) \cdot (E(\hat{V}_t \mid K_t) - \bar{V}) \mid K_t \right] \\ &\quad + E[(E(\hat{V}_t \mid K_t) - \bar{V})^2 \mid K_t]]\} \\ &= \underset{K_t}{E}\left\{ K_t \; [c^2 (m_0 + m_1 K_t) + 0 + [c(m_0 + m_1 K_t) - c(m_0 + m_1 \bar{K})]^2 \right\} \\ &= \underset{K_t}{E}\left\{ K_t \; [c^2 (m_0 + m_1 K_t) + c^2 m_1^2 (K_t - \bar{K})^2] \right\} \\ &= c^2 m_0 E(K_t) + c^2 m_1 E(K_t^2) + c^2 m_1^2 E[K_t (K_t - \bar{K})^2] \\ &= c^2 m_0 \bar{K} + c^2 m_1 E(K_t^2) + c^2 m_1^2 [E(K_t - \bar{K})^3 + \bar{K} E(K_t - \bar{K})^2] \\ &= c^2 m_0 \bar{K} + c^2 m_1 [Var(K_t) + \bar{K}^2] + c^2 m_1^2 [E(K_t - \bar{K})^3 + \bar{K} Var(K_t)] \\ &= c^2 m_0 \bar{K} + c^2 m_1 \bar{K}^2 + c^2 m_1 Var(K_t) + c^2 m_1^2 [E(K_t - \bar{K})^3 + \bar{K} Var(K_t)] \\ &= c\bar{K}\bar{V} + c^2 m_1 Var(K_t) + c^2 m_1^2 [E(K_t - \bar{K})^3 + \bar{K} Var(K_t)] \end{aligned}$$

NOTES