THE JOURNAL OF FINANCE • VOL. LI, NO. 1 • MARCH 1996

Return Volatility and Trading Volume: An Information Flow Interpretation of Stochastic Volatility

TORBEN G. ANDERSEN*

ABSTRACT

The paper develops an empirical return volatility-trading volume model from a microstructure framework in which informational asymmetries and liquidity needs motivate trade in response to information arrivals. The resulting system modifies the so-called "Mixture of Distribution Hypothesis" (MDH). The dynamic features are governed by the information flow, modeled as a stochastic volatility process, and generalize standard ARCH specifications. Specification tests support the modified MDH representation and show that it vastly outperforms the standard MDH. The findings suggest that the model may be useful for analysis of the economic factors behind the observed volatility clustering in returns.

IT IS WIDELY DOCUMENTED that daily financial return series display strong conditional heteroskedasticity. This finding strikes at the heart of empirical financial research. The estimated return variance is routinely used as a simple, albeit crude, measure of risk, and the return variance enters directly into derivative pricing formulas such as the Black-Scholes formula. Moreover, tests of market efficiency based on asset returns must incorporate corrections for heteroskedasticity in order to produce the appropriate asymptotic distributions of the test statistics. And perhaps most importantly, empirically relevant asset pricing theories typically relate expected returns, i.e., risk premia, to the joint second order moments of returns and other stochastic processes. Again, heteroskedasticity must be accounted for in order to derive efficient estimation and testing procedures. Finally, a better characterization of return volatility sheds light on the virtues of alternative specifications for the return generating mechanism.

* Kellogg Graduate School of Management, Northwestern University. The article develops ideas originally put forth in my Ph.D. dissertation at Yale University. I am grateful to the members of my dissertation committee, Peter Phillips, Steve Ross, and Steve Heston for advice. In addition, conversations with Don Andrews, Tim Bollerslev, Mike Fishman, Bob Hodrick, Neil Shephard, Ken Singleton, Bent Sørensen, George Tauchen, and Stephen Taylor have been helpful. I also received valuable comments from seminar participants at the NBER Summer Institute, 1993, Northwestern University, the University of Illinois at Chicago, the EFA Meetings in Copenhagen, August 1993, University of Wisconsin at Madison, Ohio State University, the AFA Meetings in Boston, January 1994, and the Microstructure Workshop at The Aarhus School of Business, September 1994. Finally, I thank two anonymous referees and the editor, René Stulz, for numerous suggestions that have sharpened the focus of the article. Naturally, all errors remain my own responsibility.

Until recently, most empirical work on return volatility was devoted to univariate time series models for which the autoregressive conditional heteroskedasticity (ARCH) model of Engle (1982) and its extension into GARCH by Bollerslev (1986) have been very successful.[1] However, the research objectives have grown increasingly ambitious. Multivariate structural models focusing on the volatility comovements across assets and the interaction between volatility and other economic variables are now commonplace. An important motivation behind these studies is the attempt to capture and interpret the factors that are the source of ARCH effects in returns. Nonetheless, these studies have not yet established a common framework that has gained widespread acceptance.

From a market microstructure perspective, price movements are caused primarily by the arrival of new information and the process that incorporates this information into market prices. Theory suggests that variables such as the trading volume, the number of transactions, the bid-ask spread, or the market liquidity are related to the return volatility process. However, the focus of the market microstructure literature is on intraday patterns rather than interday dynamics, so there are typically no explicit predictions regarding the relation among these variables at the daily frequency.

On the empirical front, a sizeable literature has documented a strong positive contemporaneous correlation between daily trading volume and return volatility. Following the work of Clark (1973), the empirical specifications tend to follow the intuitively appealing, although somewhat ad hoc specification associated with the "Mixture of Distribution Hypothesis" (MDH) that posits a joint dependence of returns and volume on an underlying latent event or information flow variable, see, e.g., Epps and Epps (1976) and Tauchen and Pitts (1983). The emphasis on a latent driving process separates the approach from an ARCH modeling strategy and points instead towards a stochastic volatility representation. Lamoureux and Lastrapes (1990) insert volume directly in the ARCH variance process and find it to be strongly significant, while past return shocks become insignificant. This confirms that volume is driven by the identical factors that generate return volatility, but leaves the task of providing a model for the joint process unresolved. Early tests of the implications of the MDH were supportive of the model, e.g., Harris (1986, 1987), but more recent studies have produced largely negative evidence (Heimstra and Jones (1994), Lamoureux and Lastrapes (1994), Richardson and Smith (1994)). Finally, in a comprehensive empirical study of the joint distribution of returns and volume, Gallant, Rossi, and Tauchen (1992, 1993) apply a semi-nonparametric estimation technique to daily observations at the market wide level. They unearth a number of stylized facts that serve as a challenge for future theoretical work.

This article develops a model of the daily return-volume relationship by integrating the market microstructure setting of Glosten and Milgrom (1985) with the stochastic volatility, information flow perspective of the MDH. At first, the joint distribution is derived via weak conditions on the information

[1] See, e.g., Bollerslev, Chou, and Kroner (1992) for a survey of this literature.

arrival process. Subsequently, the model is expanded into a full dynamic representation by providing a specific stochastic volatility process for the information arrivals. Both representations are estimated and tested for five major individual common stocks on the New York Stock Exchange over the period 1973–1991. The main contributions of the article are as follows. First, we develop modifications to the standard MDH that arise naturally from the microstructure setting. Second, we reinforce the recent empirical findings by resoundingly rejecting the restrictions that the standard MDH imposes on contemporaneous return-volume observations, while controlling for the trend in volume and using a long sample. In contrast, our alternative version of the MDH provides an overall acceptable characterization of these features of the data, so the general framework of the MDH may yet provide a useful basis for structural modeling of the interaction of market variables in response to information flows and, ultimately, the sources of return volatility. Third, we demonstrate that a stochastic volatility representation of the information arrival process that generalizes the popular GARCH(1,1) results in a dynamic specification of the joint system that is consistent with the main contemporaneous as well as dynamic features of the data. Fourth, we document that, in spite of the overall satisfactory fit, the simultaneous incorporation of returns and volume data results in a significant reduction in the estimated volatility persistence relative to the usual results obtained from univariate return series. This points towards some new as well as promising directions for future research within the general framework proposed.

Our model differs in important respects from other models in the literature, but the papers by Foster and Viswanathan (1995), Brock and LeBaron (1993), Ghysels and Jasiak (1994), and Tauchen, Zhang and Liu (1994) share some of the basic motivations. The remainder of the article is organized as follows. Section I outlines the basic features of the Glosten and Milgrom model required for the exposition. Section II develops the testable implications of the model for the contemporaneous system of returns and trading volume, and contrasts the specification to the usual MDH. Section III provides a description of our data. Section IV conducts explicit tests of our modified version of the MDH as well as the standard specification. Section V presents our set of candidate stochastic volatility models for the dynamic representation of the system and discusses the implementation of the Generalized Method of Moment (GMM) estimation technique in this context. Next follows a presentation of the empirical results for the fully specified dynamic model. The section concludes with an exploration of the estimated volatility persistence relative to standard ARCH models and alternative univariate return models. Finally, section VI summarizes the article and provides suggestions for future research.

I. The Theoretical Framework

We develop a version of the MDH based on the theoretical framework of Glosten and Milgrom (1985), henceforth GM. This is a natural choice since GM is structured to explain the process of price discovery, or information assimilation, that occurs just after an event providing an informational advantage to informed traders. We outline the basic features of the model and impose sufficient structure on the intertemporal setting to allow for a rigorous exploration of the model's dynamic implications.

The model focuses on a single market for an asset with a random liquidation value of V at a (distant) point in the future. There are three distinct groups of risk-neutral traders, a specialist, and informed and uninformed investors. It is a competitive setting in which investors arrive sequentially to the market in random and anonymous fashion and then decide whether to trade one unit of the asset at the bid or ask price quoted by the specialist or not trade at all. During the trading day, informed investors obtain private signals regarding the value of the asset that may provide apparent profit opportunities at the quoted prices. An important finding is that over the course of a (short) period, the sequence of trades reveals the pricing implications of the private signals and subsequently—until new private information arrives—all market participants agree on the value of the traded asset. Thus, private information arrivals induce a dynamic learning process that results in prices fully revealing the content of the private information through the sequence of trades and transaction prices. We refer to this period as a price discovery or information assimilation phase. Subsequently, when all agents agree on the price we have a (temporary) market equilibrium characterized by uniform valuation and a low bid-ask spread. Hence, we assume that each information arrival induces a price discovery phase followed by an equilibrium phase. The analysis below is based on the bivariate series obtained by recording the price and the cumulative volume since last observation at an arbitrary point during each equilibrium phase.

We require a formal specification of the market dynamics in order to explore the intertemporal features of the model. The agents revise their estimates of the terminal asset value, V, over time. Direct information is obtained from either public signals that everybody observes simultaneously or from private signals that are only received by a subset of the informed traders. Additional information is gauged from transaction prices. The history of public information signals and trades constitutes the common information set, C_τ, at time τ. Each investors information set, Φ_τ, consists of C_τ plus any potential private information.[2] The specialist valuation of the asset at time τ is the expected value of the asset conditional on his current information, i.e., $P_\tau = \mathrm{E}[V|S_\tau]$, where the specialist information set, S_τ, is another refinement of C_τ. The specialist knows the probabilistic structure of investor and news arrivals and makes correct statistical inference from the observed data. Moreover, he works under a zero profit constraint enforced by regulation or by competition from floor traders and limit orders. Nonetheless, P_τ is not the quoted price since observing whether the next agent buys or sells provides additional informa-

[2] We follow GM and label by τ the trader who arrives at the market at time τ. Since investors arrive sequentially the notation is unique, but the trader may be either informed or uninformed. For the uninformed the refinement of C_τ is the empty set.

tion. If A_τ denotes the event that an agent purchases the asset at the ask and B_τ the event that an agent sells the asset at the bid at time τ, then "fair" ask and bid prices are $P^a_\tau = E[V|S_\tau \cup A_\tau]$ and $P^b_\tau = E[V|S_\tau \cup B_\tau]$. If a sell order is submitted, the transaction price, P^b_τ, represents the expected value of the asset to the market maker, since his information set now is $S_\tau \cup B_\tau$, rather than S_τ. This implies that the market maker never ex post regrets a trade: when the trade is effectuated it has an expected value of zero to the specialist. This fact explains the second key finding of GM, namely that the transaction prices follow a martingale w.r.t. S_τ and C_τ, i.e., the specialist as well as the public information set. Hence, observed prices represent fair assessments of future value conditional on the relevant information, and the bid-ask spread does not induce any bias or negative autocorrelation into the sequence of transaction prices.

We assume that uninformed investors arrive at the market according to a constant Poisson arrival process with intensity m_0 per day.[3] For simplicity, they have inelastic demand and supply schedules, i.e., a liquidity trader either buys or sells one unit of the asset, each with probability one half. In contrast, when an informed investor arrives at the market, the trading decision is based on the value, $E[V|\Phi_\tau]$, assigned to ownership of one unit of the asset, so these agents differ only in their information sets. The signals received by the informed are generally correlated but not identical. Thus, they may disagree on their assessment of asset value. However, during the assimilation phase, their valuation, as well as that of all other agents, converges. Consequently, informed trading tends to taper off at the end of the price discovery phase.[4] This implies that the arrival rate of informed traders within each such phase is endogenous as it depends on the informativeness of the private signals, the preceding trade history, the bid-ask spread and other factors.[5]

II. The Empirical Specification

A. The Distribution of Daily Returns and Trading Volume

This section develops an empirically testable version of the MDH. As described above, the market moves from one temporary equilibrium to the next

[3] This minimizes the impact of the liquidity traders on the dynamics of return volatility and trading volume. The presence of the liquidity traders help circumvent the no-trade theorem (Milgrom and Stokey (1982)).

[4] In GM each investor assigns a value, $\rho_\tau E[V|\Phi_\tau]$, to owning one unit of the asset, where ρ_τ, differs randomly from unity. Only the specialist has ρ_τ equal to one at all times. This allows for price-elastic noise trading and effectively introduces a liquidity motive into the trading decisions of informed agents. Neither feature changes the qualitative aspects of the model.

[5] The derivation of the distribution for daily returns and volume below avoids the complications associated with such short run dependencies in informed trading by defining the basic intraday time unit to be the full (variable length) price discovery and equilibrium phase associated with the latest information arrival. Thus, the dynamic theory of GM characterizes each short interval, but the properties of the return and volume series at the daily level are driven largely, as we shall see, by the (random) number of information events per day.

in response to a large number of information arrivals during each trading day. Thus, every day can be decomposed into a random number of small intervals, each consisting of an information assimilation phase and an equilibrium phase of varying length. A priori, the joint distribution of price changes and informed trading volume is identical over each interval. Thus, if we sample transaction prices and accumulated volume once during each equilibrium phase, then the joint distribution of price changes and informed trading volume constitutes an i.i.d. series.[6]

An important conclusion of GM is that the sequence of transaction prices follows a martingale with respect to public information. Thus, no matter how we sample prices the resulting series will remain a martingale. We denote the transaction price recorded during the j-th temporary equilibrium of day t by $P_{j,t}, j = 1, \ldots, J_t - 1$. The total number of information arrivals on day t, J_t, is random but large. Consequently, the return over the full trading day is given by

$$1 + R_t = \prod_{j=1}^{J_t} \frac{P_{j,t}}{P_{j-1,t}} = \frac{P_{J_t,t}}{P_{0,t}} \tag{1}$$

where $P_{0,t}$ denotes the first and $P_{J_t,t}$ the last transaction price of the day. From the martingale property, R_t has mean zero. It is convenient to work with continuously compounded return, so ignoring an inconsequential approximation error, we write $R_t = \ln(P_{J_t,t}/P_{0,t})$ and obtain

$$R_t = \sum_{j=1}^{J_t} \ln\left(\frac{P_{j,t}}{P_{j-1,t}}\right) = \sum_{j=1}^{J_t} \eta_{j,t}, \quad \eta_{j,t} \sim \text{i.i.d.}(0, \sigma_\eta^2) \tag{2}$$

The number of information arrivals, J_t, is assumed to be large, yet display a significant variation across the trading days. This feature is best captured by introducing the notion of a benchmark day with a fixed large number of arrivals, J. Then $J_t = K_t J$, where the positive scaling factor K_t denotes the intensity of information arrivals relative to the benchmark. Clearly, the benchmark day with J arrivals generates a random return with mean zero and variance $\sigma^2 = J\sigma_\eta^2$, and we may thus represent the intra-day return components as $\eta_{j,t} = \sigma\epsilon_{j,t}/J^{1/2}$, where $\epsilon_{j,t}$ is i.i.d., mean zero and has unit variance. The return specification for an arbitrary trading day now takes the form

$$R_t = \sigma K_t^{1/2} \frac{1}{(JK_t)^{1/2}} \sum_{j=1}^{JK_t} \epsilon_{j,t} \tag{3}$$

[6] In fact, our theory allows the series to display quite general forms of weak dependency—the pivotal point is that appropriate laws of large numbers and central limit theorems apply.

The result resembles that of Clark (1973). For large J, and under weak regularity conditions, we have[7]

$$R_t|K_t \sim N(0, \sigma^2 K_t) \tag{4}$$

Hence, daily returns are conditionally normal but have variances that reflect the intensity of information arrivals, K_t, as also noted by Ross (1989). This representation is closely related to the idea of time deformation as the return variance is driven by an event time scale (information arrivals) rather than a calendar time scale.[8] Moreover, we find that the systematic return volatility dynamics is governed strictly by the time series properties of the information flow into the market.

The daily trading volume, V_t, has informed and noise components, i.e., $V_t = IV_t + NV_t$. Noise trading is governed by a stochastic process with a constant arrival intensity of m_0 per day. Hence, NV_t is directed by a time-invariant Poisson process, $\mathrm{Po}(m_0)$. Consequently, the systematic variation in trading volume is due solely to fluctuations in the informed volume.

Each private information arrival tends to generate trading by informed agents over the information assimilation phase. Both the number of daily arrivals and the number of informed traders are large, but the daily informed volume remains moderate because the probability that a given (potentially) informed trader transacts as a result of a single information arrival is small. We denote this probability, associated with the j-th arrival on day t, by $p_{j,t}$. An informed agent may fail to trade on a given arrival for a number of reasons: (i) each insider only picks up information associated with a particular arrival with some (low) probability; (ii) public news that reveals the information may arrive to the market before the insider gets access to the news or the specialist; (iii) other informed traders may arrive at the market before him or her and through the intensity and unidirectional pattern of their trading reveal the pricing implication of the information arrival; (iv) as soon as the specialist infers that private information is present, the bid-ask spread increases, thus diminishing the chance that the private signal presents a profit opportunity. Of course, as the price discovery process unfolds the spread narrows again because the adverse selection problem facing the specialist is abating. But by then, the quoted prices reflect most of the information content of the private signals, and profitable trading opportunities are minimal.

Consequently, we envision a setting in which each informed trader, on average, makes only a few transactions per day. In addition, the likelihood of an informed agent trading on a given news arrival, $p_{j,t}$, may vary with the informational content of the signals. However, a number of factors tend to

[7] The textbook central limit theorem argument fails because the number of arrivals is stochastic. See Clark (1973), Theorem 3, for an appropriate generalization of the standard theorem. More general results that allow for weak dependency in the sequence of return components are provided by Billingsley (1968), Theorems 17.2, 20.1, and 20.3.

[8] See Stock (1988) for an introduction to time deformation and Ghysels and Jasiak (1994) for an application.

equalize this probability across the different types of information arrivals. For example, following an arrival that induces a large price revision, more insiders may be informed, and a higher fraction initially finds it profitable to trade. However, as the information is being incorporated into the quotes, the intensity and clustering of orders at one side of the market quickly induces a substantial change in the bid and ask prices as well as an increase in the spread. This lowers the probability of informed trading on the part of the remaining insiders. Conversely, an arrival with less information content tends to generate less concentrated insider trading, less rapid information assimilation, less of a change in the spread, and hence a longer period over which an insider may be able to exploit the superior information.[9] Exact predictions regarding the amount of informed trading associated with different types of events are not available. Nonetheless, the discussion suggests that the variation in $p_{j,t}$ across arrivals is rather limited, in which case the distributional approximations used below should be accurate.

First, consider the informed volume induced by a single news arrival. Each insider trades one unit of the asset with probability $p_{j,t}$ before the end of the assimilation period. Thus, the induced amount of trading is binomially distributed, $\text{Bin}(I, p_{j,t})$, with an expected value of $I \cdot p_{j,t}$, where I denotes the maximum number of insiders that might obtain a private signal associated with the event. For I large, $p_{j,t}$ small, and $I \cdot p_{j,t}$ moderate, the Poisson, $\text{Po}(I \cdot p_{j,t})$, approximates this binomial extremely well.[10]

Now, consider the informed volume on a day with J_t arrivals. We assume the variation in the probability of informed trading, $p_{j,t}$, across the arrivals is governed by a stationary process. Letting the expected number of trades by an insider be μ on a day with J arrivals, $p_{j,t}$ has mean μ/J. If the variation of $p_{j,t}$ around its mean is limited, as suggested above, then by using the Poisson approximation, one may show that the distribution of the daily informed volume is given by[11]

$$IV_t | K_t \sim \text{Po}(IK_t\mu) \tag{5}$$

and combining the expressions for the noise and informed components of trading, we obtain the following specification for daily trading volume

$$V_t | K_t \sim \text{Po}(m_0 + IK_t\mu) \tag{6}$$

Thus, the characterization of volume takes on a remarkably simple parametric form. The constant, m_0, reflects the noise component while the informed component is proportional to the information flow. The factor of proportional-

[9] This line of reasoning is formalized in Proposition 3 of GM, which states that the expected number of trades times the square of the average spread is bounded by a number independent of the pattern of trade.

[10] See Feller (1968), pp. 152–153, for specific illustrations regarding the closeness of this approximation. Alternatively, one may utilize a normal approximation to the binomial. This is, however, much less precise under the present circumstances.

[11] A derivation is provided in the appendix.

ity, $m_1 = I \cdot \mu$, determines how strongly volume fluctuates in response to news.

B. Parameter Identification and Non-Stationarity

This section deals with issues of parameter identification and the implications of nonstationarity in the trading volume series.

First, it is clear that $m_1 = I \cdot \mu$ is identifiable, while it is impossible to identify the parameters I and μ separately from observations on returns and volume alone. We ignore these parameters and concentrate on the estimation of m_1. Second, inspection of the equations (4) and (6) reveal that the scale of K_t itself is arbitrary. We normalize the system by choosing $\sigma \equiv 1$ in equation (4). This fixes the scale of m_0 and m_1 as well. Hence, our final specification for the return equation is

$$R_t | K_t \sim N(0, K_t) \tag{7}$$

An equivalent representation of the daily return is often convenient. If Z_t is i.i.d. $N(0, 1)$, then

$$R_t = K_t^{1/2} Z_t$$

In this specification it is transparent that the return volatility is identical to the information flow. Hence, the unobserved, or latent, process K_t is a genuine stochastic volatility process (Andersen (1992b)).

We now turn to the issue of nonstationarity. So far, we have assumed that volume is stationary. This ignores the fact that trading volume data typically have a strong trend component. We propose to detrend the volume data by extracting a time trend, π_t from the volume series, reflecting changes in the size of the basic unit traded or the number of active traders.[12] Thus, observable volume is given by $V_t^o = V_t \cdot \pi_t$, where V_t represent the stationary volume series from the previous sections. The detrending procedure used in the empirical work may depend on the characteristics of the data. However, even if the detrending procedure is perfect, we face yet another scaling problem. The reason is that our theory does not determine a specific value for the level, or unconditional mean, of the relevant stationary volume series. Hence, an estimated trend, $\hat{\pi}_t$, will at best reproduce the true underlying trend up to a factor of proportionality, i.e., $c \cdot \hat{\pi}_t = \pi_t$, where c is an unknown positive constant. In this case the detrended volume series becomes $\hat{V}_t = V_t^o / \hat{\pi}_t = cV_t$. The parameters in the volume expression are not invariant to this scaling of the series. Indeed, we have

$$\hat{V}_t | K_t \sim c \cdot \mathrm{Po}(m_0 + m_1 K_t) \tag{8}$$

[12] Implicit in this approach is the assumption that the relative size of informed and noise trading, i.e., the parameters m_0 and m_1, have been invariant over the sample period.

Hence, the scaling indeterminacy introduces an additional parameter that, in the absence of extraneous information, must be estimated.

C. Properties of the Implied Return-Volume System

The system (7)–(8) represents our empirical specification of the MDH as motivated by the stylized version of the GM model. However, the system differs from the usual characterization of the MDH, in particular with respect to the specification of the volume equation which typically takes the form

$$\hat{V}_t | K_t \sim N(\mu_v K_t, \sigma_v K_t) \tag{9}$$

(see, e.g., Harris (1986, 1987)). Consequently, we refer to the system (7)–(8) as the **Modified MDH**. The main differences in the volume specification are first, the presence of the constant term, m_0, which accounts for noise or liquidity components of trading; second, the imposition of a conditional Poisson rather than normal distribution. Besides reflecting a closer approximation to the underlying distribution of the model, the latter is appealing because it respects the nonnegativity constraint on trading volume. This is not a trivial issue. Empirical work on the standard MDH often produces point estimates, which imply that negative volume observations should occur with a disturbingly high frequency. Finally, notice that the scaling parameter, c, allows the conditional mean and variance of the trading volume to differ. Hence, this parameter, in conjunction with m_1, effectively serves the same function in the modified MDH as the two separate mean and variance parameters do in the standard MDH.[13]

The modified MDH determines the contemporaneous relation between return and volume. The series are related due to the common dependence on the information flow variable K_t. This is short of a full characterization of the bivariate system because the dynamics of the information flow remains unspecified—theory is simply moot on this point. Nonetheless, this system does impose testable restrictions on the bivariate return-volume series. This is exploited in Section IV as a preliminary diagnostic on the specification. For now, we simply notice that the specification, by application of the law of iterated expectations, implies:

$$\mathrm{cov}(R_t, V_t) = 0; \qquad \mathrm{cov}(R_t^2, V_t) = \sigma^2 m_1 \,\mathrm{var}(K_t) > 0.$$

This confirms that the model is consistent with well-known stylized facts. Moreover, it provides a formal basis for the assertion that it should be advantageous to utilize trading volume figures in conjunction with returns when constructing measures of daily return volatility. This follows because the information flow represents a stochastic volatility process that drives both returns and volume, so each series will provide information regarding the state

[13] If one applies a normal rather than a Poisson approximation during the derivation of the daily volume, one will detect similar theoretical constraints between the mean and variance parameters. However, the detrending procedure again destroys the linkage.

of the unobserved volatility process. This reflects an important difference vis-à-vis the ARCH modeling paradigm where return volatility, conditional on the parameter estimates, is observable ex post. In the present model we can never disentangle the impact of the two components of the product $K_t^{1/2} \cdot Z_t$ below equation (7), and the realization of the volatility process is thus subject to uncertainty which, in general, will be reduced by drawing on additional information. An alternative source of information is direct knowledge about the dynamics of the news arrival process. The choice of an appropriate representation for the K_t process is discussed in Section V.

D. Incorporating Public Information Arrivals

It is potentially important to allow for the arrival of public information that has an immediate impact on the valuation of the risky asset but does not induce any additional trading activity. This occurs in our model if informed traders receive no signals in advance of the information release, and the information content is readily interpretable to all market participants, i.e., there is no asymmetric information associated with the release. Assuming that this type of public information arrives randomly but at an expected rate that is proportional to the arrival rate of the other information processes, this will simply manifest itself in an additional random multiplicative term in the return equation,[14] i.e.,

$$R_t = K_t^{1/2} O_t \tilde{Z}_t = K_t^{1/2} Z_t \tag{10}$$

where O_t is a positive, nondegenerate i.i.d. random variable with (normalized) mean of unity and independent of $\tilde{Z}_t$, which is i.i.d. N(0, 1). It follows that Z_t is i.i.d. with mean zero and unit variance. However, the additional randomness associated with public information arrivals induces fat tails and excess kurtosis in the distribution of Z_t. Indeed, the decomposition of Z_t into O_t and $\tilde{Z}_t$ is not identifiable,[15] and the proper statistical representation of the process is $R_t = K_t^{1/2} Z_t$. Consequently, the returns process is no longer conditionally normal. Harvey, Ruiz, and Shephard (1994) stress that this feature may be important in the context of the lognormal stochastic volatility model.

III. Data Description

For brevity, this section describes only the series associated with IBM common stock. We find the same qualitative characteristics for the remaining four common stocks in our sample.

A continuously compounded daily return series, corrected for dividends and stock splits, is constructed from closing prices on IBM common stock

[14] Bollerslev (1987) provides an explicit characterization of t-distributed return innovations along these lines.

[15] In the terminology of Andersen (1992b) the representation involving O_t is nongenuine, while the one involving Z_t is genuine.

Table I

IBM Common Stock Daily Percentage Returns: Summary Statistics

The statistics are based on continuously compounded percentage returns, corrected for dividends and stock splits, calculated from daily New York Stock Exchange closing prices on International Business Machines common stock over the period January 2, 1973–December 23, 1991. Observations between December 24 and January 1, inclusive, were deleted. The full sample consists of 4,693 observations. The subsamples cover 1973–1975, 1976–1978, 1979–1981, 1982–1984, 1985–1987, and 1988–1991, respectively. The first five have around 740 observations, while the sixth has 988 observations. The prices were obtained from Standard & Poor's Daily Stock Price Guide and checked against the returns indicated on tapes from the Center for Research in Security Prices. The Ljung-Box portmanteau test for 10th order autocorrelation in returns and squared returns are χ^2-distributed with 10 degrees of freedom. The *p*-values are provided in parentheses following the value of the statistics. The return autocorrelations for the full sample have standard errors of the order $(1/T)^{1/2} = 1/(4{,}693)^{1/2} = 0.0146$.

Sample	Full	1	2	3	4	5	6
100 × Mean	1.51	−4.61	4.52	−2.08	11.94	1.07	−0.94
St. Dev.	1.46	1.76	1.02	1.46	1.36	1.72	1.32
Skewness	−1.04	0.32	0.64	0.61	0.35	−4.56	−0.65
Kurtosis	27.8	5.61	5.49	5.34	3.86	75.41	8.76
Maximum	10.05	9.86	5.43	8.15	5.10	10.05	5.89
Minimum	−26.09	−9.13	−3.26	−4.74	−4.13	−26.09	−10.05

Additional Full Sample Statistics: First 8 Return Correlations: −0.031; −0.008; 0.005; −0.035; 0.025; 0.017; 0.017; 0.005; Ljung-Box Test: $\chi^2_{10}(R) = 19.7$ (0.033); $\chi^2_{10}(R^2) = 193$ (5.5×10^{-36}).

over January 1, 1973–December 31, 1991. The sample consists of 4693 observations. Summary statistics are provided in Table I. The sample mean is small and not significantly different from zero, but it is erratic in terms of both sign and size over subsamples. The sample standard deviation and variance exceed the sample mean by a factor of more than 95 and are fairly stable across subsamples. Finally, the returns display excess kurtosis and are slightly skewed to the left, although the skewness is positive in four of the six subsamples—the large negative value in the fifth, associated with October 1987, generates the overall negative value. The kurtosis is substantially larger than 3 in all subsamples, but again very much so in the fifth. Excluding October 1987 from this subsample yields a kurtosis consistent with the other subsamples. The extreme sensitivity to outliers illustrates the problem of obtaining reliable estimates of higher order return moments. They are invariably erratic and provide only limited information in the estimation phase.[16]

[16] Our estimation procedure reduces the impact of sample moments that have highly variable time series realizations. Outliers are very informative in other situations. For example, Phillips and Loretan (1994) assume returns are generated by a linear process with an underlying i.i.d. sequence of random variables whose tail behavior is of the Pareto-Levy form. They estimate the maximal moment exponent by fitting the slope of the tail of the (transformed) empirical cumulative distribution function of the return series.

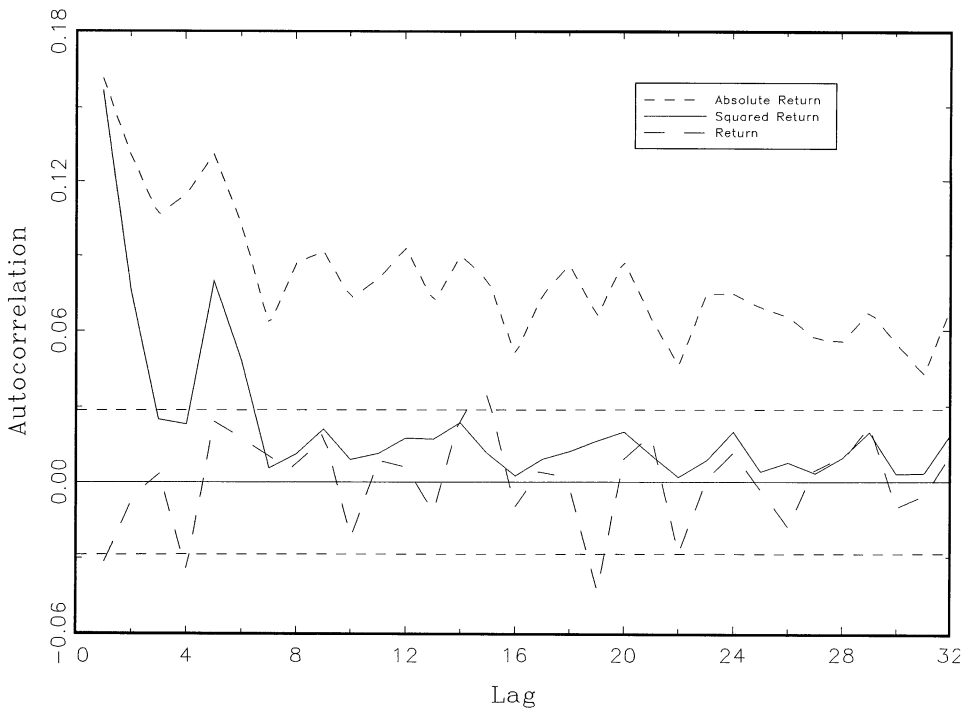

Figure 1. Autocorrelations for Raw, Absolute and Squared Daily Returns on International Business Machines Common Stock over 1973–1991. This figure shows the first 32 autocorrelations for the (transformed) return series. The returns are continuously compounded percentage returns, corrected for dividends and stock splits, calculated from daily New York Stock Exchange closing prices on International Business Machines common stock over the period January 2, 1973–December 23, 1991. Observations between December 24 and January 1, inclusive, were deleted. The sample consists of 4,693 observations. The stock prices were obtained from Standard & Poor's Daily Stock Price Guide and checked against the returns indicated on tapes from the Center for Research in Security Prices. The short dashed lines in the figure represent the 5 percent standard error band for the null hypothesis of independent returns.

In summary, the returns are clearly not drawn independently from a normal distribution. The empirical return distribution is more peaked around zero and has thicker tails than the corresponding normal with the same mean and variance. Figure 1 displays the autocorrelation coefficients up to lag 32 for the returns, absolute returns, and squared returns series. It shows not only that the returns are nonnormal, but also that they cannot be i.i.d. If they were, the transformed series $|R_t|$ and R_t^2 would be i.i.d. as well, but the significant low order auto-correlations of the latter strongly violate the indicated confidence bands. Thus, the IBM return series displays the usual dependency in higher order moments that traditionally are captured by ARCH formulations.

Preliminary justification for our approach is provided by Table II, where we report the cross-correlations between the return volatility and trading volume series within each year of our sample. The short annual samples mitigate the effect of the trend in the volume. There is a strong relationship between the

Table II

Cross-Correlations between Squared Returns and Trading Volume for IBM Common Stock

The cross-correlations are based on continuously compounded percentage returns, corrected for dividends and stock splits, and volume figures, corrected for stock splits, calculated from daily New York Stock Exchange closing prices and trading volume for International Business Machines common stock over the period January 2, 1973–December 23, 1991. Observations between December 24 and January 1, inclusive, were deleted. This leaves a sample of about 247 pairs of observations per year. The figures were obtained from Standard & Poor's Daily Stock Price Guide and checked against the returns and volume indicated on tapes from the Center for Research in Security Prices.

	$\mathrm{Corr}(R_t^2, \hat{V}_{t-\lvert j \rvert})$		
Year\Lag	$j = 0$	$j = 1$	$j = 2$
1973	0.54	0.51	0.19
1974	0.35	0.41	0.19
1975	0.55	0.33	0.25
1976	0.29	0.33	0.19
1977	0.48	0.21	−0.01
1978	0.25	0.31	0.14
1979	0.25	0.23	0.05
1980	0.29	0.22	0.01
1981	0.32	0.34	0.13
1982	0.34	0.30	0.21
1983	0.32	0.08	0.00
1984	0.44	0.33	0.08
1985	0.44	0.35	0.14
1986	0.39	0.30	0.09
1987	0.42	0.42	0.30
1988	0.34	0.21	0.14
1989	0.55	0.38	0.11
1990	0.33	0.18	0.07
1991	0.51	0.19	0.13

two series, which is consistent with evidence from previous work (Karpoff (1987)). The raw trading volume is displayed in Figure 2A. Most strikingly, the series has a strong but erratic trend. On average, the growth in the number of shares traded, corrected for stock splits, is about 13 percent a year. This characterization is reinforced by the volume growth rates reported in Table III. In addition, the series has a distinct seasonality at year's end as the trading intensity declines between Christmas and New Year.[17] Consequently, we exclude all observations from December 23 through the first trading day of the following year. Another distinctive feature is that the fluctuations around the trend grow more pronounced over time. The log-transformation of volume in Figure 2B accomplishes two things. It converts the long run trend, as a rough

[17] Due to the trend, year's end volume should on average exceed the mean daily volume over the preceding year. However, the average volume between Christmas and New Year is below the average volume in 14 of 17 cases, and often by a large margin.

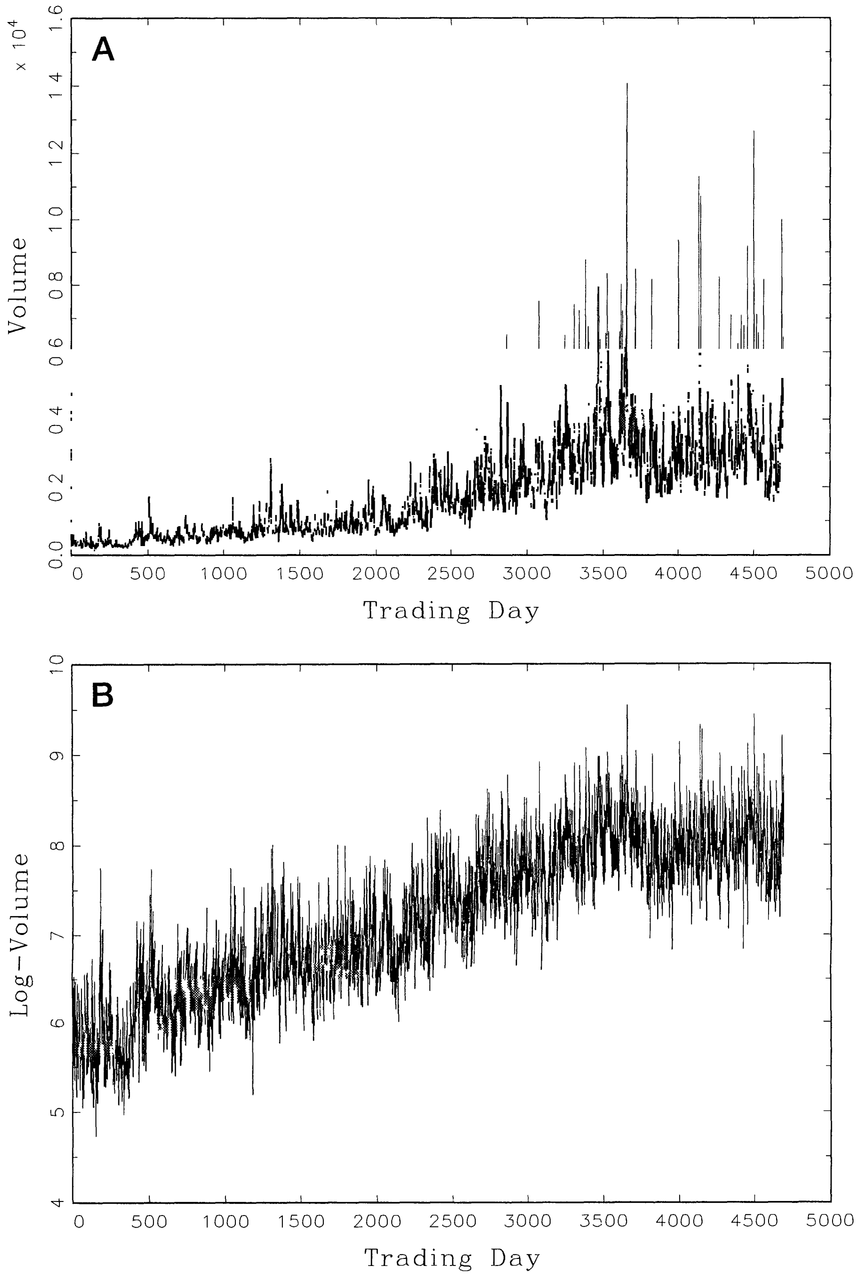

Figure 2A. Trading Volume for International Business Machines Common Stock over 1973–1991. This figure shows the daily trading volume, corrected for stock splits, for International Business Machines common stock over the period January 2, 1973–December 23, 1991. Observations between December 24 and January 1, inclusive, are omitted. The sample consists of 4,693 observations. The data were obtained from Standard and Poor's Stock Price Guide. **Figure 2B. Log-Trading Volume for International Business Machines Common Stock over 1973–1991.** This figure shows the daily log-trading volume, corrected for stock splits, for International Business Machines common stock over the period January 2, 1973–December 23, 1991. Observations between December 24 and January 1, inclusive, are omitted. The sample consists of 4,693 observations. The data were obtained from Standard and Poor's Stock Price Guide.

Table III

Annual Percentage Growth Rates in Daily Trading Volume for IBM Common Stock

The growth rates in trading volume are based on volume figures, corrected for stock splits, and calculated from the daily New York Stock Exchange trading volume for International Business Machines common stock over the period January 2, 1973–December 23, 1991. Observations between December 24 and January 1, inclusive, were deleted. This leaves a sample of about 247 pairs of observations per year. The Growth rates were calculated as the difference between the average trading volume in the two indicated periods and converted into annual percentage rates. The subsamples cover 1973–1975, 1976–1978, 1979–1981, 1982–1984, 1985–1987, and 1988–1991, respectively. The first five have around 740 observations, while the sixth has 988 observations. The figures were obtained from Standard & Poor's Daily Stock Price Guide and checked against the tapes from the Center for Research in Security Prices.

Panel A: Measured from One Subsample to the Next

Sample	1–2	2–3	3–4	4–5	5–6
Growth Rate	17.8	7.8	22.3	17.8	−0.0

Panel B: Measured Biannually

Sample	73–75	75–77	77–79	79–81	81–83	83–85	85–87	87–89	89–91
Growth Rate	18.4	11.7	7.4	11.6	25.9	16.1	26.9	−13.3	3.9

approximation, to a straight line, lending some support to the hypothesis of a constant long run growth rate, and it stabilizes the variability of the series.[18] Nonetheless, the assumption of a constant growth rate is unreasonably strict. Instead, we employ detrending procedures that allow for a stochastic trend component in volume as well as an autocorrelated disturbance term. The task is to filter out the trend, i.e., the fluctuations in the expected or "normal" trading volume, while retaining a measure of the correlated deviations around this trend associated with the periods of unusual informational intensity. The exact method chosen is necessarily governed by the data since we have very little guidance from theory.

We briefly describe our detrending procedures. Each method estimates a trend component that produces a "normal" or expected volume series, and the detrended series is then obtained by dividing each trading figure with the corresponding "normal" volume for that trading day. The approach is motivated by the indication of approximate stationarity of the log-differences of the volume series, which suggests that percentage deviations from trend are stationary. Our first method uses a nonparametric kernel regression procedure based on the normal kernel. This corresponds to a two-sided moving average with weights that decline as the distance from the trading day increases. The key choice is the relative weighting of remote and nearby obser-

[18] It is difficult to test these conjectures due to the complex nature of the generating process of trading. However, informal analysis of the sum of squared residuals over subsamples obtained from fitting a linear trend in the log-model, using the two-step Cochrane-Orcutt to account for autocorrelation in the error term, suggests that the variance process has been stabilized.

Table IV

Detrended Daily Trading Volume for IBM Common Stock: Summary Statistics

The summary statistics are based on detrended volume figures, corrected for stock splits, calculated from the daily New York Stock Exchange trading volume for International Business Machines common stock over the period January 2, 1973–December 23, 1991. Observations between December 24 and January 1, inclusive, were deleted. The subsamples cover 1973–1975, 1976–1978, 1979–1981, 1982–1984, 1985–1987, and 1988–1991, respectively. The first five have around 740 observations, while the sixth has 988 observations. The figures were obtained from Standard & Poor's Daily Stock Price Guide and checked against the tapes from the Center for Research in Security Prices. The detrending was performed by dividing the actual trading volume for a given day by the expected value calculated using either a nonparametric kernel regression (Panel A) or a centered rolling two-year mean (Panel B).

Sample	Full	1	2	3	4	5	6
Panel A: Detrending using a Nonparametric Regression with a Normal Kernel							
Mean	0.993	0.974	1.02	0.948	1.02	1.03	0.977
St. Dev.	0.413	0.486	0.428	0.366	0.397	0.402	0.391
Skewness	2.10	3.28	1.60	1.60	1.26	1.74	2.06
Kurtosis	13.0	23.9	7.00	7.18	5.05	9.49	11.2
Maximum	5.93	5.93	3.40	3.13	2.88	3.86	3.73
Minimum	0.222	0.292	0.222	0.331	0.332	0.285	0.276
Panel B: Detrending using a Centered Two Year Rolling Sample Mean							
Mean	0.995	0.984	1.03	0.946	1.02	1.03	0.972
St. Dev.	0.409	0.458	0.433	0.370	0.407	0.404	0.377
Skewness	1.81	2.25	1.63	1.62	1.29	1.79	1.99
Kurtosis	9.31	12.7	7.21	7.25	5.10	9.93	10.6
Maximum	4.49	4.49	3.46	3.14	2.92	3.93	3.64
Minimum	0.216	0.318	0.216	0.325	0.332	0.284	0.279

vations.[19] Our second method is an equally weighted moving average of length two years centered on the estimated trend component.[20] The "window" of one year in each direction renders this method compatible to the first.[21] An alternative detrending procedure was implemented by replacing the moving average by the moving median, which is less sensitive to outliers. However, the two series were quite similar, and we report only results based on the moving average.

Summary statistics of the derived volume series, reported in Table IV,

[19] Our leading scheme allocates about 5.2 percent weight to the one month of observations centered on the given trading day. Similarly, 15.0 percent weight is assigned to the 3 month period, 55.7 percent on the one year period, 87.5 percent on the two years, and 98 percent on the 3 years around the trading day.

[20] Standard one-sided (weighted) averages are used for the end of the sample where two-sided averages are inapplicable. See Brockwell & Davis (1987) for a discussion of this approach.

[21] The kernel-based estimates track positively correlated volume observations more closely than the estimates from alternative methods, so deviations from trend tend to be smaller for the former. We compensate by limiting the window to one year in each direction for the moving mean, while the kernel method assigns weight to observations outside of that horizon.

Table V

Autocorrelations for the Daily Detrended Trading Volume for IBM Common Stock

The correlations are based on detrended volume figures, corrected for stock splits, calculated from the daily New York Stock Exchange trading volume for International Business Machines common stock over the period January 2, 1973–December 23, 1991. Observations between December 24 and January 1, inclusive, were deleted. The figures were obtained from Standard & Poor's Daily Stock Price Guide and checked against the tapes from the Center for Research in Security Prices. The detrending was performed by dividing the actual trading volume for a given day by the expected value calculated using a nonparametric kernel regression with a normal kernel (corresponding to the series in Panel A of Table IV).

	Corr($\hat{V}_t$, $\hat{V}_{t-j}$)											
j	1	2	3	4	5	6	7	8	9	10	11	12
	0.52	0.32	0.25	0.23	0.22	0.18	0.14	0.14	0.13	0.13	0.12	0.10
j	13	14	15	16	17	18	19	20	21	22	23	24
	0.09	0.09	0.10	0.06	0.05	0.03	0.05	0.05	0.04	0.02	0.05	0.05
j	25	26	27	28	29	30	31	32	33	34	35	36
	0.03	0.02	0.02	0.02	0.03	0.03	−0.00	−0.00	−0.00	0.03	0.03	0.03

demonstrate the similarity of the basic characteristics of the two volume series. In both cases the mean is near unity, which is to be expected given the normalization rule. Furthermore, the standard deviations and skewness measures are quite stable across subsamples and do not indicate overly erratic higher order moments. The skewness is positive in all subsamples. This is a feature that must be accounted for by any model that purports to capture the stylized facts of the volume series. In addition, the kurtosis exceeds three but is smaller overall than for the returns series. However, this finding is not robust. In the majority of the subsamples the volume series has a higher kurtosis than returns. Finally, Table V lists the lower order autocorrelations for the derived volume series. The autocorrelations display a regular and smooth decline from significantly positive values at small lags to about zero at lags above thirty.[22] Moreover, a simple Goldfeld-Quandt test detected no signs of a shift in the unconditional variance. In sum, the derived volume series appears stationary and displays the type of properties we would expect from theory. We elect the method based on the nonparametric normal kernel regression as our basic "normalized" volume series and present empirical results for this series only. However, most of the empirical work on the IBM data was conducted for both volume series and for various weighting schemes in the kernel regression scheme. The results display some variation across the procedures, but the qualitative conclusions appear robust. The volume series based on the nonparametric kernel regression is displayed in Figure 3. This is the series that underlies the empirical work summarized in the following sections.

[22] This is consistent with Harris' (1987) observation that trading volume should display a higher degree of autocorrelation than the squared return series in the normal-mixture model.

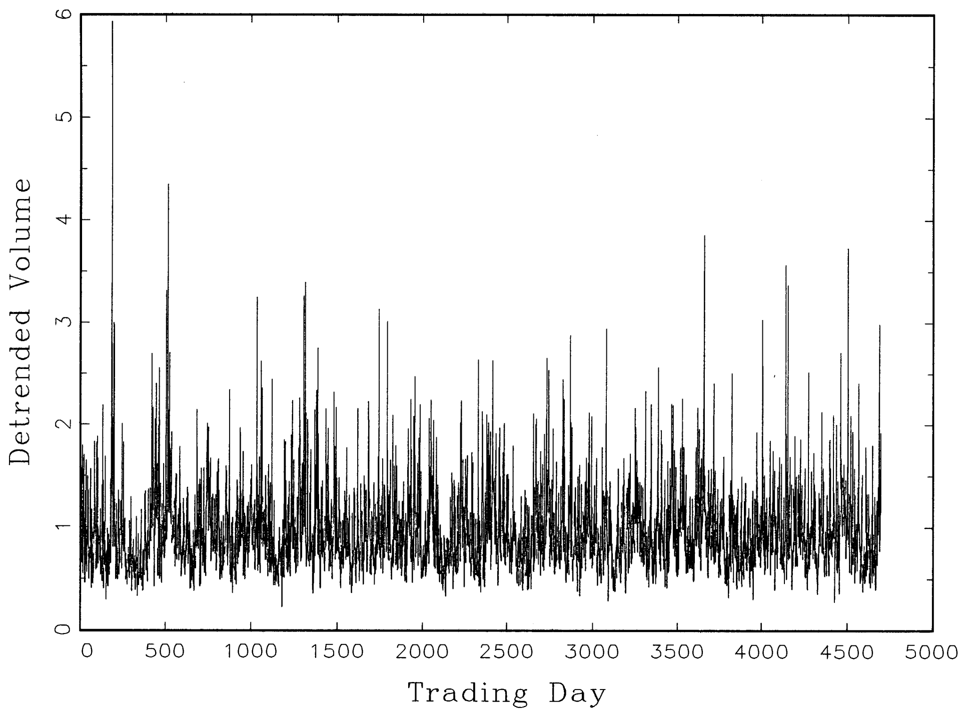

Figure 3. Detrended Trading Volume for International Business Machines Common Stock over 1973–1991. This figure shows the daily trading volume, corrected for stock splits, for International Business Machines common stock over the period January 2, 1973–December 23, 1991. Observations between December 24 and January 1, inclusive, are omitted. The sample consists of 4,693 observations. The data were obtained from Standard and Poor's Stock Price Guide. The series was detrended by dividing the actual trading volume for a given day by the expected value calculated using a nonparametric kernel regression with a normal kernel (corresponding to the series in Panel A of Table IV).

IV. Testing the Modified Mixture of Distributions Hypothesis

This section tests the Modified MDH under minimal restrictions on the information flow. This may be viewed as a preliminary diagnostic designed to assess whether the contemporaneous characterization of the return-volume system we have developed is consistent with the basic features of data. However, it is also of interest in its own right, as the exploration of the empirical merits of the MDH has attracted considerable attention in the recent literature, and our findings shed new light on these results. Since most studies of the standard MDH allow for a nonzero mean in the return equation, we include a constant, $\bar{r}$, in equation (7), but otherwise the test is based on the specification (7)–(8). The test investigates the restrictions on the joint unconditional moments of returns and volume implied by the system. No dynamic moments, i.e., moments incorporating lagged relationships, are included since the theory in its pure form puts no restrictions on the intertemporal behavior of the information flow variable. In fact, beyond the existence of some lower unconditional moments of K_t, the test is void of distribution assumptions on the information flow.

We apply the GMM procedure of Hansen (1982) which exploits the convergence of selected sample moments to their unconditionally expected values.[23] It provides a simultaneous test of the cross-restrictions on various lower order moments of the return-volume system. The following set of unconditional moment conditions is used:

$$
\begin{aligned}
\mathrm{E}[R_t] &= \bar{r} \\
\mathrm{E}|R_t - \bar{r}| &= (2/\pi)^{1/2}\mathrm{E}[K_t^{1/2}] \\
\mathrm{E}[(R_t - \bar{r})^2] &= \mathrm{E}[K_t] = \bar{K} \\
\mathrm{E}|R_t - \bar{r}|^3 &= 2(2/\pi)^{1/2}\mathrm{E}[K_t^{3/2}] \\
\mathrm{E}[(R_t - \bar{r})^4] &= 3\mathrm{E}[\bar{K}^2 + \mathrm{var}(K_t)] \\
\mathrm{E}[\hat{V}_t] &= c(m_0 + m_1\bar{K}) = \bar{V} \\
\mathrm{E}[(\hat{V}_t - \bar{V})^2] &= c\bar{V} + c^2 m_1^2 \,\mathrm{var}(K_t) \\
\mathrm{E}[(\hat{V}_t - \bar{V})^3] &= c^2\bar{V} + 3c^3 m_1^2 \,\mathrm{var}(K_t) + c^3 m_1^3 \mathrm{E}[K_t - \bar{K}]^3 \\
\mathrm{E}[R_t\hat{V}_t] &= \bar{r}\bar{V} \\
\mathrm{E}[|R_t - \bar{r}|(\hat{V}_t - \bar{V})]) &= c(2/\pi)^{1/2} m_1(\mathrm{E}[K_t^{3/2}] - \mathrm{E}[K_t^{1/2}]) \\
\mathrm{E}[(R_t - \bar{r})^2\hat{V}_t] &= \bar{V}\bar{K} + m_1 \,\mathrm{var}(K_t) \\
\mathrm{E}[(R_t - \bar{r})^2(\hat{V}_t - \bar{V})^2] &= c\bar{K}\bar{V} + c^2 m_1 \,\mathrm{var}(K_t) \\
&\quad + c^2 m_1^2[\mathrm{E}[K_t - \bar{K}]^3 - \bar{K}\,\mathrm{var}(K_t)]
\end{aligned}
$$

The parameter vector is ($\bar{r}$, $\mathrm{E}[K_t^{1/2}]$, $\bar{K}$, $\mathrm{E}[K_t^{3/2}]$, $\mathrm{var}(K_t)$, $\mathrm{E}[K_t - \bar{K}]^3$, m_0, m_1, c). Thus, there are nine free parameters and twelve moment conditions and hence three over-identifying restrictions. 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 & West (1987a) (NW) procedure.

The estimation is carried out for five separate stocks. Table VI presents the findings, including the χ_3^2-test for goodness-of-fit based on the over-identifying restrictions. All point estimates are of the expected sign and the magnitudes appear reasonable, although the higher order moments of the information arrival process, not unexpectedly, are imprecisely estimated. Moreover, the

[23] The GMM procedure appears relatively more efficient than the alternative of Quasi Maximum Likelihood (QML) Kalman filter estimation at the moderate levels of volatility persistence that we encounter in these bivariate systems (Andersen and Sørensen (1994)).

Table VI

Estimation Results for the "Modified Mixture of Distributions Hypothesis"

The results are based on continuously compounded percentage returns, corrected for dividends and stock splits, and detrended volume figures, corrected for stock splits, calculated from the daily New York Stock Exchange closing prices and trading volume for a set of five common stocks over the period January 2, 1973–December 23, 1991. Observations between December 24 and January 1, inclusive, were deleted. The figures were obtained from Standard & Poor's Daily Stock Price Guide and checked against the returns and volume indicated on tapes from the Center for Research in Security Prices. The volume series was detrended by dividing the actual trading volume for a given day by the expected value calculated using a nonparametric kernel regression with a normal kernel (corresponding to the series in Panel A of Table IV). The following system involving the daily returns, R_t, the detrended volume, $\hat{V}_t$, and the (unobserved) number of information arrivals, K_t, was estimated by the Generalized Method of Moments:

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

The estimated parameters include the mean return, $\bar{r}$, various unconditional moments of the K_t-process, including the mean $\bar{K} = \mathrm{E}[K_t]$, a (nuisance) scaling parameter for volume, c, and the approximate average fraction of daily volume independent of (associated with) information arrivals, $c \cdot m_0 (c \cdot m_1)$. Estimated standard errors are provided below the point estimates, while p-values are indicated below the test statistics. The weighting matrix used in the objective function was calculated according to Newey and West (1987a) using 25 lags, except that the weighting matrix and parameter estimates were iterated until convergence, so the weighting matrix reflects the final parameter estimates. The χ^2-test for Goodness-of-Fit (Hansen (1982)) has three degrees of freedom since there are 12 moment restrictions and 9 free parameters. The "likelihood-ratio" test statistic for the restriction $m_0 = 0$ is χ^2-distributed with one degree of freedom (Newey and West (1987b)).

	$\bar{r}$	$\mathrm{E}[K_t^{1/2}]$	$\bar{K}K$	$\mathrm{E}[K_t^{3/2}]$	$\mathrm{Var}(K_t)$	$\mathrm{E}[K_t - \bar{K}]^3$	$c \cdot m_0$	$c \cdot m_1$	c	χ^2_3 (p-val.)	$m_0 = 0$ (p-val.)
Alcoa											
	0.019	1.66	3.28	7.70	9.82	68.7	0.607	0.110	0.071	6.93	46.2
	(0.022)	(0.039)	(0.173)	(0.763)	(3.04)	(106)	(0.065)	(0.024)	(0.008)	(0.074)	(0.000)
Amoco											
	0.026	1.47	2.58	5.35	6.47	114	0.622	0.144	0.057	8.01	37.0
	(0.019)	(0.039)	(0.142)	(0.460)	(1.35)	(26.9)	(0.058)	(0.027)	(0.008)	(0.046)	(0.000)
Coca-Cola											
	0.034	1.46	2.60	5.51	7.34	190	0.559	0.163	0.057	7.20	0.935
	(0.020)	(0.044)	(0.294)	(2.60)	(11.9)	(392)	(0.422)	(0.178)	(0.034)	(0.066)	(0.232)
IBM											
	0.007	1.29	1.85	3.06	2.08	24.3	0.564	0.226	0.026	4.31	7.06
	(0.016)	(0.029)	(0.090)	(0.258)	(0.488)	(6.29)	(0.073)	(0.044)	(0.010)	(0.229)	(0.008)
Kodak											
	−0.016	1.55	2.79	6.10	7.87	162	0.553	0.153	0.027	4.90	7.03
	(0.019)	(0.037)	(0.164)	(0.733)	(2.37)	(60.4)	(0.082)	(0.035)	(0.009)	(0.179)	(0.008)

χ^2-statistics, while overall rather low, do not reject the model at standard significance levels.[24] Given our large sample size and the possibility of structural breaks in the series, these findings are perhaps more supportive of the

[24] Construction of an overall test for the joint restrictions on all five stocks simultaneously is complicated by the fact that the return and volume processes are correlated across the securities.

Modified MDH than one would expect. It does, however, raise a couple of questions. First, the power of the goodness-of-fit test may be so low against interesting alternatives as to render the test uninformative. Second, our findings seem at odds with results reported by Richardson & Smith (1994) and Lamoureux & Lastrapes (1994) who present mixed, but overall negative, evidence relating to the standard MDH specification. We address these questions in two different ways. Both are motivated by considering the standard MDH the natural alternative hypothesis in this context. As discussed in section II.C, the main difference between the models is the volume equation, where the Modified MDH posits a conditional Poisson rather than conditional normal distribution, and it includes a constant term to accommodate the noise trading component of volume. The significance of the latter feature can be assessed directly within the Modified MDH framework by testing the restriction $m_0 = 0$. The last column of Table VI reports the χ_1^2-statistics and associated p-values for a "likelihood ratio" type test based on the difference in the objective function with and without imposing this restriction. In four of five cases it is soundly rejected. Moreover, in all cases of rejection, some of the estimated parameter values (not reported) are negative, which is incompatible with the theory.[25]

A second way of assessing the modified MDH is by direct comparison to a test of the standard MDH on the same data set. Table VII provides the results from a GMM test on the standard MDH. This is similar to the estimation strategy pursued by Richardson and Smith (1994). Our results indicate an overwhelming rejection of the standard specification of the MDH. Even for Coca-Cola, where the constant term was not deemed significant, the specification fares poorly with exceptionally low p-values and parameter estimates that violate the theoretical restrictions.[26] Thus, our evidence is much more conclusive than that presented by previous authors. This is probably due to our larger sample size and our explicit handling of the trend in trading volume.[27] In addition, it illustrates that the proposed test is powerful enough to reject the standard MDH without resorting to dynamic features of the system as in Lamoureux and Lastrapes (1994).

Our findings have the following implications. One, the standard version of the MDH is soundly rejected. But, second, this does not imply that the reasoning behind the MDH is useless as a basis for empirical asset pricing and market microstructure research because fairly minor modifications result in specifications that are roughly consistent with the joint behavior of stock returns and appropriately detrended trading volume over long sample periods.

[25] Only for Coca-Cola do we fail to reject the restriction. This is likely due to the fact that Coca-Cola is unusual in other respects as well. Its parameters in the Modified MDH system are much less precisely estimated than those for the other stocks, as evidenced by the relative standard errors. Interestingly, Tauchen, Zhang, and Liu (1994) also find this series hard to fit.

[26] Restricting the parameter estimates to stay within the proper parameter space results in estimates at the boundary of the parameter space and even lower p-values.

[27] For example, the average daily trading volume for IBM rises by about 40 percent over the period 1982–1986 that is investigated by Richardson and Smith (1994).

Table VII

Estimation Results for the Standard Mixture of Distributions Hypothesis

The results are based on continuously compounded percentage returns, corrected for dividends and stock splits, and detrended volume figures, corrected for stock splits, calculated from the daily New York Stock Exchange closing prices and trading volume for a set of five common stocks over the period January 2, 1973–December 23, 1991. Observations between December 24 and January 1, inclusive, were deleted. The figures were obtained from Standard & Poor's Daily Stock Price Guide and checked against the returns and volume indicated on tapes from the Center for Research in Security Prices. The volume series was detrended by dividing the actual trading volume for a given day by the expected value calculated using a nonparametric kernel regression with a normal kernel (corresponding to the series in Panel A of Table IV). The following system involving the daily returns, R_t, the detrended volume, $\hat{V}_t$, and the (unobserved) number of information arrivals, K_t, was estimated by the Generalized Method of Moments:

$$R_t | K_t \sim N(\bar{r}, K_t) \qquad \hat{V}_t | K_t \sim N(\mu_v, \sigma_v K_t)$$

The estimated parameters include the mean return, $\bar{r}$, various unconditional moments of the K_t-process, including the mean $\bar{K} = \mathrm{E}[K_t]$, and the volume parameters, μ_v and σ_v. Estimated standard errors are provided below the point estimates, while p-values are indicated below the test statistics. The weighting matrix used in the objective function was calculated according to Newey and West (1987a) using 25 lags, except that the weighting matrix and parameter estimates were iterated until convergence, so the weighting matrix reflects the final parameter estimates. The χ^2-test for Goodness-of-Fit (Hansen (1982)) has three degrees of freedom since there are 12 moment restrictions and 9 free parameters.

	$\bar{r}$	$\mathrm{E}[K_t^{1/2}]$	$\bar{K}$	$\mathrm{E}[K_t^{3/2}]$	$\mathrm{E}[K_t^2]$	μ_v	σ_v	χ_3^2 (p-val.)
Alcoa	−0.015	1.89	4.34	9.92	27.2	0.236	0.013	96.8
	(0.021)	(0.032)	(0.131)	(0.494)	(1.65)	(0.008)	(0.011)	($5 \cdot 10^{-20}$)
Amoco	−0.036	1.72	3.43	7.14	15.7	0.302	−0.019	168
	(0.019)	(0.028)	(0.096)	(0.283)	(0.875)	(0.009)	(0.007)	($4 \cdot 10^{-35}$)
Coca-Cola	−0.153	2.59	13.0	57.2	436	0.086	−0.110	937
	(0.020)	(0.026)	(0.098)	(0.651)	(7.77)	(0.001)	(0.003)	($2 \cdot 10^{-201}$)
IBM	−0.048	1.73	3.34	6.95	16.0	0.331	−0.090	4.31
	(0.016)	(0.019)	(0.065)	(0.279)	(0.695)	(0.007)	(0.008)	($4 \cdot 10^{-83}$)
Kodak	−0.042	1.93	4.41	11.7	29.1	0.250	−0.056	4.90
	(0.019)	(0.029)	(0.131)	(0.501)	(1.79)	(0.007)	(0.006)	($3 \cdot 10^{-65}$)

In light of the dramatic failure of the standard specification, this improvement in empirical performance is remarkable. Third, the proposed testing methodology is indeed powerful enough to reject interesting alternative hypotheses in the present context.

V. Representation and Estimation of the Full Dynamic Model

A. The Information Arrival Process

The theory developed in previous sections has little bearing on the process generating information arrivals, except for the observation that the information flow constitutes a stochastic volatility process for returns in equation (6). A couple of considerations are relevant when selecting a dynamic representation for the information variable. First, casual empiricism suggests that news arrivals are positively correlated. When unanticipated news breaks on a given day, more detailed disclosures tend to follow over the next few days or weeks, and different interpretations of the circumstances leading to the event surface. This tends to keep "the story" in the headlines for an extended period of time. Moreover, important changes in the tactical orientation of a company do typically play themselves out over longer periods (take-over battles and proxy fights), and these developments are revealed through periodic news updates. Second, and more importantly, judging from the success of modeling return volatility dynamics by means of ARCH processes, and in particular models closely related to the GARCH(1, 1), it is clear that an information arrival process governing the dynamic features of return volatility must display a similar type of positive conditional dependency. Andersen (1994) develops a class of stochastic volatility models that are closely related to GARCH(1, 1), termed Stochastic AutoRegressive Volatility (SARV) models. They satisfy the above criteria in that they allow for positive autocorrelation in news arrivals, and they serve as natural stochastic volatility generalizations of relevant ARCH models. We have investigated two promising specializations, the first being a standard SARV model:

$$K_t^{1/2} = \omega + \beta K_{t-1}^{1/2} + \alpha K_{t-1}^{1/2} u_t \tag{11}$$

where $\alpha, \beta \geq 0$, $\omega > 0$, u_t is i.i.d. with $u_t > 0$, $\mathrm{E}[u_t] = 1$ and $\mathrm{Var}(u_t) = \sigma_u^2$. In general, specific distributional assumptions for u_t are not required. SARV generalizes a GARCH(1, 1) model for the conditional return standard deviation, and, in particular, the volatility persistence is closely related to the sum, $\alpha + \beta$, as is the case for GARCH(1, 1).[28]

The second specification is a lognormal stochastic volatility process (exponential SARV):

$$\ln(K_t) = \omega + \beta \ln(K_{t-1}) + \sigma_u u_t, \qquad \sigma_u > 0 \tag{12}$$

where $u_t \sim$ i.i.d. N(0, 1). The process is closely related to the EGARCH(1, 1) (Nelson (1991)) and is popular in the options pricing literature. The volatility persistence is governed by β. Given certain parameter restrictions, both processes are strictly stationary and have finite lower order moments.[29]

B. The GMM—Framework for the Dynamic Model

GMM estimation of the system (7)–(8) plus either equation (11) or (12) exploits the convergence of selected sample moments to the corresponding

[28] Let $u_t = c|Z_{t-1}|$ with $c = \mathrm{E}(|Z_t|)^{-1}$ to obtain a GARCH(1, 1) specification for $K_t^{1/2}$.

[29] For the SARV process the proof relies on a generalization of arguments used for GARCH(1, 1) by Nelson (1990). Closed form solutions for the unconditional moments of SARV are required for our GMM procedure, and are provided in Andersen (1994).

unconditionally expected values. Typically, moments such as $\mathrm{E}(R_t^m R_{t-k}^n)$, $\mathrm{E}(R_t^m \hat{V}_{t-k}^n)$, and $\mathrm{E}(\hat{V}_t^m \hat{V}_{t-k}^n)$ are utilized for low nonnegative integer values of m, n, and k. Let the $q \times 1$ vector of unknown parameters be denoted by θ, the true parameter vector by θ_0, and the $Q \times 1$ vector of selected sample moments by $M_r(\theta)$, where $Q \geq q$, and T equals the sample size. Finally, let the vector of corresponding analytical moments be denoted $A(\theta)$. The GMM-estimator, $\hat{\theta}_T$, minimizes the distance between $A(\theta)$ and $M_T(\theta)$ over the parameter space Ξ in the following quadratic form

$$\theta_T = \arg\min_{\theta \in \Xi} (A(\theta) - M_T(\theta))' \Gamma_T^{-1} (A(\theta) - M_T(\theta)) \tag{13}$$

where the specific metric is determined by the choice of the positive definite and possibly random weighting matrix, Γ_T. Under suitable regularity conditions $\hat{\theta}$ is consistent and asymptotically normal

$$T^{1/2}(\hat{\theta}_T - \theta_0) \sim N(0, \Omega)$$

Our procedure is based on consistent estimates of the optimal weighting matrix that minimizes the asymptotic covariance matrix of the parameter estimates, Ω.

Several problems arise in the estimation phase. First, although distributional assumptions are not required for SARV, the particular parameterization and the erratic behavior of higher order return and volume moments combine to make the individual higher order moments of u_t hard to identify. Estimation efficiency is greatly enhanced by imposing reasonable constraints on these moments. A number of candidate distributions for u_t were utilized in this capacity, each generating a slightly different set of moment conditions for the model. Second, as noted, the estimation of the full model requires the inclusion of lagged variables among the moments. We rely on lower order dynamic moments of order up to 20. This leads to the following choice of moments: $\mathrm{E}(R_t)$, $\mathrm{E}(R_t^2)$, $\mathrm{E}|R_t - r|$, $\mathrm{E}(R_t^4)$, $\mathrm{E}(\hat{V}_t)$, $\mathrm{E}(\hat{V}_t^2)$, $\mathrm{E}(\hat{V}_t R_t^2)$, $\mathrm{E}|R_t - r|\,|R_{t-i} - r|$, $\mathrm{E}(R_t^2 R_{t-i}^2)$, $\mathrm{E}(\hat{V}_t \hat{V}_{t-i})$, $\mathrm{E}(\hat{V}_{t-|i|}|R_t - r|)$, $i > 0$.[30] Simulation evidence (Andersen and Sørensen (1995), henceforth AS) indicates, however, that it is inappropriate to include all twenty lags for each moment, since it results in an excessive number of moments relative to sample size. Hence, we include an arbitrary selection of lags for each of the dynamic moments, except for the following criteria adapted from AS: (i) a majority are lower order moments (more precisely estimated than higher order moments); (ii) we avoid having all dynamic moments be included at the same lag (to reduce the multicollinearity in the weighting matrix); (iii) the total number of included moments is kept at 30 for the present sample size. Third, estimation of the weighting matrix involves a choice between a host of alternative procedures. AS find that the single most important feature for relatively large samples is to include a sufficiently large number of lags in the

[30] Andersen (1994) verifies that the parameters are identified from these moments.

construction of the weighting matrix. Thus, our reported results rely on the NW procedure with 75 lags.[31]

C. The Modeling of the IBM Series

The model has numerous variants depending on the distributional assumptions imposed on u_t. Given the limited prior evidence with estimation of these models, a preliminary specification search is inevitable. Unfortunately, there are many alternative ways in which to implement the GMM procedure, even if we adhere to the guidelines of AS. This flexibility in the choice of distributional assumptions and estimation procedure raises serious questions regarding the potential for overfitting or "data snooping" (see Lo and MacKinlay (1990)). We try to mitigate these concerns by selecting the model specification and estimation procedure from experimentation with the IBM data alone. Only thereafter do we proceed to the analysis of the remaining series. Hence, for each stock we rely on the identical model specification and moment conditions, and we apply the same estimation procedure for the GMM weighting matrix. Since the other series are only weakly correlated with the IBM series, this "cross-validation" should alleviate the concerns relating to overfitting. Moreover, we explore the robustness of the IBM results by investigating the behavior of the system over various subsamples.

After rather extensive experimentation with models fit to the IBM data series, we settled on the SARV specification for $K_t^{1/2}$ and u_t generated by the absolute value of a Generalized Error distributed variable, $\mathrm{GED}_w(0, 1)$.[32] This choice provided one of the more satisfactory fits judged from the standard χ^2-tests, and the parameter estimates are quite robust to the selections of moments and to estimation over subsamples. At the same time it is not unrepresentative of the findings for alternative specifications.[33]

Below, we provide the exact specification of the system that was selected from the investigation of the IBM series. The results reported in the following

[31] The NW procedure is nonparametric and uses the Bartlett kernel (Newey and West (1987a)). We also implemented the Quadratic Spectral (QS) kernel estimator, and for both procedures we experimented with data-dependent choice of bandwidth (Andrews (1991)). However, the AS findings appear robust in this regard as well.

[32] The class of GED_w-distribution ($w > 0$), also known as power exponential distributions, is a natural choice because it encompasses the normal within the interior of the parameter space ($w = 2$). For $w < 2$, the tails are fatter than for a normal distribution, while for $w > 2$ they are thinner than for a normal. The imposition of a unit variance results in only one remaining free parameter, w. The density and the unconditional lower order moments for this normalized GED_w-variable are provided in Andersen (1992a). For further discussion of this class of distributions, see, e.g., Nelson (1991).

[33] We considered specifications with u_t distributed according to a Lognormal, Gamma, or χ_1^2 random variable or the absolute value of a standard normal or student t variable. Further, we explored the lognormal representation given by equation (12).

sections utilize the same basic framework.

$$R_t = \bar{r} + K_t^{1/2} Z_t \tag{7'}$$

$$\hat{V}_t | K_t \sim c \cdot \mathrm{Po}(m_0 + m_1 K_t) \tag{8}$$

$$K_t^{1/2} = \omega + \beta K_{t-1}^{1/2} + \alpha K_{t-1}^{1/2} u_t \tag{11}$$

where $u_t = |v_t|/\mathrm{E}|v_t|$, $v_t \sim \mathrm{GED}_w(0, 1)$. Letting $\bar{\sigma} = \mathrm{E}(K_t^{1/2}) = \omega/(1 - (\alpha + \beta))$, the parameterization of the system involves eight unknown coefficients, namely $\theta = (\bar{r}, \bar{\sigma}, \beta + \alpha, \alpha, c, cm_0, cm_1, w)$.

Table VIII summarizes results obtained for the IBM data over the full sample and three subsamples, 1973–78, 1979–84, and 1985–91. Overall, the χ^2-statistics for goodness-of-fit are supportive of the model at standard significance levels. Moreover, the point estimates appear reasonable and fall within the theoretical bounds that are consistent with the assumptions underlying the estimation procedure. The estimated daily mean returns are small and vary substantially across subsamples. This reflects the usual difficulty of obtaining precise estimates of expected returns (see, e.g., Merton (1980)). The return standard deviation is governed by $\bar{\sigma}$, which is fairly uniform across subsamples with estimates ranging from 1.24 to 1.31. More interestingly, the estimates of the persistence measure, $\alpha + \beta$, are lower than the ones usually obtained from GARCH models of daily returns. On the other hand, this finding parallels results obtained by Foster & Viswanathan (1995), who use half-hourly return and volume observations on IBM.[34] In conjunction, these studies seem to indicate that the extension of univariate models of return volatility into bivariate systems via the incorporation of volume observations results in a significant reduction in the estimated volatility persistence. At this point, such a conclusion is premature since the direct comparison of volatility persistence measures between the ARCH and stochastic volatility models is tenuous. However, we present further results and formal statistical tests that strongly support this conclusion in the following sections.

The parameter $c \cdot m_0$ measures the fraction of the average daily volume that is independent of the information flow. From this perspective, between 37 percent and 75 percent of the daily volume is unrelated to the arrival of news, while the information sensitive components of trade, reflected in the coefficient $c \cdot m_1$, accounts for the remainder. Point estimates of the scaling constant c are small but strictly positive. Finally, the estimates of the tail parameter indicate that the error processes have fatter tails than the normal, which corresponds to $w = 2$. This is consistent with the evidence from the GARCH models.[35]

[34] In fact, given the high frequency observations used by Foster and Viswanathan, we find much longer half-lives of shocks to volatility than they do.

[35] The large standard errors on w reflect the fact that the identification of this parameter hinges on higher order sample moments, which invariably are erratic.

Table VIII

Estimation Results for the IBM Return-Volume System Based on the SARV-GED$_w$ Model

The results are based on continuously compounded percentage returns, corrected for dividends and stock splits, and detrended volume figures, corrected for stock splits, calculated from the daily New York Stock Exchange closing prices and trading volume for International Business Machines common stock over the period January 2, 1973–December 23, 1991. Observations between December 24 and January 1, inclusive, were deleted. The figures were obtained from Standard & Poor's Daily Stock Price Guide and checked against the returns and volume indicated on tapes from the Center for Research in Security Prices. The volume series was detrended by dividing the actual trading volume for a given day by the expected value calculated using a nonparametric kernel regression with a normal kernel (corresponding to the series in Panel A of Table IV). The three subsamples are nonoverlapping and contain 1,564 observations each. The following system involving the returns, R_t, the detrended volume, $\hat{V}_t$, and the (unobserved) number of information arrivals, K_t, was estimated by the Generalized Method of Moments:

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

$$K_t^{1/2} = \omega + \beta K_{t-1}^{1/2} + \alpha\, K_{t-1}^{1/2} u_t \qquad u_t = |v_t| / \mathrm{E}|v_t|, \quad v_t \sim \mathrm{GED}_w(0,\, 1)$$

The estimated parameters include the mean return, $\bar{r}$, a (nuisance) scaling parameter for the volume process, c, and the approximate average fraction of daily volume independent of (associated with) information arrivals, $c \cdot m_0 (c \cdot m_1)$, and the parameters of the Stochastic Auto-Regressive Volatility (SARV) specification for $K_t^{1/2}$, $(\bar{\sigma}, \alpha, \alpha + \beta, w)$. Among the latter, $\bar{\sigma} = \mathrm{E}[K_t^{1/2}] = \omega/(1 - \alpha - \beta)$ measures the return standard deviation, $\alpha + \beta\,(<1)$ indicates the degree of persistence in shocks to volatility and volume, and w denotes the tail parameter of the standardized Generalized Error Distribution (Nelson (1991)) and reflects the propensity for outliers in the innovations to the information arrival process. Estimated standard errors are provided below the point estimates, while p-values are reported for the χ^2-test. The test has 22 degrees of freedom for the full sample and 16 degrees of freedom for the subsamples. The weighting matrix used in the objective function was calculated according to Newey and West (1987a), except that the weighting matrix and parameter estimates were iterated until convergence, so the weighting matrix reflects the final parameter estimates, and a data-dependent bandwidth based on univariate AR(1) approximations to the time series moments was employed (Andrews (1991)).

	$\bar{r}$	$\bar{\sigma}$	$\beta + \alpha$	α	$c \cdot m_0$	$c \cdot m_1$	c	w	χ^2
Full	0.012	1.31	0.756	0.239	0.650	0.171	0.041	1.13	0.66
	(0.020)	(0.017)	(0.050)	(0.060)	(0.044)	(0.025)	(0.010)	(0.377)	
1973–1978	−0.037	1.31	0.787	0.327	0.650	0.157	0.054	2.26	0.38
	(0.035)	(0.044)	(0.044)	(0.089)	(0.029)	(0.017)	(0.011)	(1.35)	
1979–1984	0.032	1.30	0.710	0.156	0.371	0.329	0.021	0.932	0.77
	(0.031)	(0.027)	(0.079)	(0.080)	(0.097)	(0.058)	(0.015)	(0.693)	
1985–1991	0.013	1.24	0.834	0.135	0.747	0.139	0.050	0.575	0.99
	(0.032)	(0.024)	(0.068)	(0.062)	(0.051)	(0.033)	(0.011)	(0.186)	

D. Analysis of the Five Individual Stocks

This section summarizes our findings for the five individual stocks over the full sample. Table IX reports point estimates for the full dynamic system allowing for fat-tailed conditional return innovations as outlined in Section II D. Consequently, Z_1 is distributed i.i.d. $\mathrm{GED}_{w(r)}(0, 1)$ rather than standard

Table IX

Estimation Results for the Return-Volume System Based on the SARV-GED$_w$-GED$_{w(r)}$ Model

The results are based on continuously compounded percentage returns, corrected for dividends and stock splits, and detrended trading volume, corrected for stock splits, calculated from the daily New York Stock Exchange closing prices and trading volume for five common stocks over the period January 2, 1973–December 23, 1991. Observations between December 24 and January 1, inclusive, were deleted. The figures were obtained from Standard & Poor's Daily Stock Price Guide and checked against the returns and volume indicated on tapes from the Center for Research in Security Prices. The volume series was detrended by dividing the actual trading volume for a given day by the expected value calculated using a nonparametric kernel regression with a normal kernel (corresponding to the series in Panel A of Table IV). The following system involving the returns, R_t, the detrended volume, $\hat{V}_t$, and the (unobserved) number of information arrivals, K_t, was estimated by the Generalized Method of Moments:

$$R_t = K_t^{1/2} Z_t \qquad \hat{V}_t | K_t \sim c \cdot \mathrm{Po}(m_0 + m_1) \qquad K_t^{1/2} = \omega + \beta K_{t-1}^{1/2} + \alpha K_{t-1}^{1/2} u_t$$

$$Z_t \sim \mathrm{GED}_{w(r)}(0, 1) \qquad u_t = |v_t| / \mathrm{E}|v_t|, \quad v_t \sim \mathrm{GED}_w(0, 1),$$

where Z_t and v_t are i.i.d. and mutually independently distributed according to the standardized Generalized Error Distribution (GED), whose outlier behavior are determined by the tail parameters $w(r)$ and w respectively (Nelson (1991)). In addition, the estimated parameters include the mean return, $\bar{r}$, a (nuisance) scaling parameter for the volume process, c, and the approximate average fraction of daily volume independent of (associated with) information arrivals, $c \cdot m_0 (c \cdot m_1)$, and the parameters of the Stochastic Auto-Regressive Volatility (SARV) specification for $K_t^{1/2}$, $(\bar{\sigma}, \alpha, \alpha + \beta, w, w(r))$. Among the latter, $\bar{\sigma} = \mathrm{E}[K_t^{1/2}] = \omega/(1 - \alpha - \beta)$ measures the return standard deviation, $\alpha + \beta$ (<1) indicates the degree of persistence in shocks to volatility and volume. Estimated standard errors are provided below the parameters, while p-values are reported for the χ^2-tests. The overall (joint) goodness-of-fit test has 22 degrees of freedom. The tests of parameter restrictions in the third and second to last column are $\chi^2(1)$-distributed, while the joint test of goodness-of-fit of the subset of moments involving volume data reported in the last column is $\chi^2(9)$-distributed (Eichenbaum, Hansen, and Singleton (1988)). The weighting matrix used in the objective function was calculated according to Newey and West (1987a) with 75 lags, except that the weighting matrix and parameter estimates were iterated until convergence, so the weighting matrix reflects the final parameter estimates.

$\bar{r}$	$\bar{\sigma}$	$\beta + \alpha$	α	$c \cdot m_0$	$c \cdot m_1$	c	w	$w(r)$	Joint χ^2-Test	Test: $w(r) = 2$	Test: $\alpha + \beta = 0.885$	Joint Volume Moments
Alcoa												
0.060	1.65	0.707	0.143	0.454	0.172	0.290	0.599	1.76	19.2	3.38	18.3	11.0
(0.023)	(0.041)	(0.049)	(0.047)	(0.063)	(0.023)	(0.034)	(0.186)	(0.099)	(0.571)	(0.066)	(0.000)	(0.615)
Amoco												
0.054	1.23	0.803	0.138	0.535	0.274	0.135	0.688	1.86	23.3	1.23	12.8	12.9
(0.018)	(0.033)	(0.029)	(0.043)	(0.057)	(0.038)	(0.011)	(0.246)	(0.115)	(0.328)	(0.267)	(0.000)	(0.459)
Coca-Cola												
0.070	1.34	0.759	0.098	0.528	0.222	0.153	0.435	1.59	20.4	16.3	9.99	11.0
(0.020)	(0.037)	(0.046)	(0.028)	(0.083)	(0.047)	(0.030)	(0.098)	(0.068)	(0.496)	(0.000)	(0.002)	(0.610)
IBM												
0.003	1.24	0.694	0.243	0.621	0.196	0.039	1.07	1.99	21.9	0.004	55.2	18.9
(0.018)	(0.034)	(0.034)	(0.063)	(0.035)	(0.024)	(0.005)	(0.415)	(0.118)	(0.406)	(0.948)	(0.000)	(0.127)
Kodak												
0.005	1.45	0.789	0.124	0.460	0.210	0.088	0.611	1.91	20.9	0.373	8.87	18.3
(0.023)	(0.037)	(0.037)	(0.041)	(0.086)	(0.045)	(0.009)	(0.241)	(0.136)	(0.465)	(0.541)	(0.003)	(0.145)

Normal which adds $w(r)$ to the parameter vector θ. Otherwise, the system is identical to the one selected above on the basis of the IBM data.

The estimation results for IBM are virtually identical to those presented in Table VIII because $w(r)$ at 1.99 is almost identical to the conditionally normal case.[36] The findings for the remainder of the stocks are generally consistent with those for IBM. The most noteworthy features are as follows: the estimated mean returns are all positive, and significantly nonzero in three of five cases; the volatility persistence measure, $\alpha + \beta$, fluctuates between 0.70 and 0.80; across the stocks the information-insensitive component of trading lies between 45 percent and 62 percent of the daily volume; for all stocks the estimates indicate the presence of fat tails in the volatility innovations as well as the conditional return innovations. However, the former are quite imprecisely estimated, while the latter only deviate significantly from normality in the case of Coca-Cola and, marginally, Alcoa. An alternative test of the hypothesis, $w(r) = 2$, is available in the form of a likelihood-ratio type test as described in section IV, and the associated test statistics are included in Table IX. It is reassuring that they point to the same conclusion as before: the restriction is only strongly rejected for Cola-Cola (at the 1 percent level), while it is rejected at the 10 percent level for Alcoa. For the remainder we fail to reject the conditional normality of the return innovations. This points to yet another theoretically motivated modification of the MDH, which appears to improve the performance of the model; the release of (unanticipated) public information induces return variability but may have little concurrent effect on trading volume. This can induce conditionally fat tails and excess kurtosis in the return distribution.

In conclusion, the full dynamic specification fares well, as it provides an overall reasonable fit to the joint return and volume moments of the individual stocks. The results support the hypothesis that an appropriately specified structural system of the return-volume relation inspired by the reasoning underlying the MDH is likely to constitute a fruitful avenue for future research. Nonetheless, it is also evident that the current specification has some shortcomings. This shows up primarily in the reduction of the measure of return volatility persistence. The next section explores the statistical robustness of this finding and verifies that a truly significant reduction takes place. The concluding section of the article discusses some potential explanation for this and points to modeling strategies that may help accommodate these features of the joint return-volume system.

E. Exploring the Decay in the Estimated Volatility Persistence

This section investigates the decay in the estimated volatility persistence as the return system is expanded into a joint return-volume system. For brevity, the analysis focuses on the IBM series. Initially, we consider the following

[36] The drop in the p-value associated with the χ^2-test for goodness-of-fit is due to the additional parameter, which reduces the degrees of freedom by one without improving the fit of the model.

explanations. First, it is possible that our selected series happen to have a lower volatility persistence than that typical of high frequency financial return series. Second, our volatility process corresponds to the conditional return standard deviation rather than the conditional return variance typically modeled in the ARCH setting, and this may explain the difference in persistence characteristics. Third, there may be sizeable finite sample biases inherent in the GMM estimation procedure, which can account for the lower estimated volatility persistence. Fourth, it is possible that the SARV specification has fundamentally different volatility persistence characteristics from GARCH(1, 1) so that we, in fact, should expect to detect different volatility properties. The evidence reported below rejects all these explanations for the estimated low volatility persistence in the bivariate system.

Table X presents estimation results for seven different univariate models fit to the IBM return series. The first two rows of Panel A demonstrate that standard GARCH(1, 1) models, estimated by maximum likelihood, produce the usual high volatility persistence measures. Indeed, $\alpha + \beta$ is in excess of 0.985 for both the normally distributed and the thick tailed ($\text{GED}_{w(r)}$) return innovations. Likewise, maximum likelihood estimation of a GARCH(1, 1) model for the conditional return standard deviation, termed SD-GARCH, produces a point estimate of $\alpha + \beta = 0.989$. The GMM estimation results reported in Panel B further reinforces the conclusion that return volatility persistence is extremely high. The first row refers to the identical SD-GARCH model and indicates no significant decay in $\alpha + \beta$. The next two rows show that corresponding SARV models produce almost identical estimates for the relevant parameters. The SARV model in the second row restricts the tail parameter of the return and volatility innovations to be identical, so that the only difference to the SD-GARCH setting is the independence of the two processes. In contrast, the model in the third row of Panel B allows these tail parameters to differ. Both innovation processes appear fat-tailed, but the parameter for the volatility innovation process is extremely imprecisely estimated and, in fact, close to being underidentified by the data. In conclusion, none of the results indicate that the significant reduction in the estimated volatility persistence can be explained by differences between the estimation methods or by differences in the fundamental volatility persistence characteristics across the models.

Finally, the last SARV model restricts the return innovation to be normal. This is typical of most proposed stochastic volatility representations and specifically the standard MDH. Interestingly, this restriction appears to induce a substantial thickening of the tails in the estimated volatility innovation process and, simultaneously, a significant drop in the volatility persistence. Thus, the imposition of normality in the return innovations seems to have the same qualitative implications for the volatility persistence as the incorporation of the volume processes, albeit quantitatively less significant. Indeed, in Table IX we find that the normality restriction on the return innovation is strongly rejected only for one of the stocks. Moreover, while a joint test for the goodness-of-fit of the subset of the moments that involve volume data fails to reject the model, and thus confirms the overall satisfactory fit of the bivariate specification,

Table X

Univariate Models for the IBM Return Series

The results are based on continuously compounded percentage returns, corrected for dividends and stock splits, calculated from daily New York Stock Exchange closing prices on International Business Machines common stock over the period January 2, 1973–December 23, 1991. Observations between December 24 and January 1, inclusive, were deleted. The full sample consists of 4,693 observations. The prices were obtained from Standard & Poor's Daily Stock Price Guide and checked against the returns indicated on tapes from the Center for Research in Security Prices. The following models of the return series, R_t and the conditional return variance process, K_t, were estimated by either Maximum Likelihood (Panel A) or Generalized Method of Moments (Panel B):

$$R_t = \bar{r} + K_t^{1/2} Z_t \qquad K_t^q = \omega + \beta K_{t-1}^q + \alpha K_{t-1}^q u_t, \quad q = 1 \text{ or } 2.$$

The error processes Z_t and u_t are independent and both standardized to have mean zero and unit variance, but may be normally distributed or Generalized Error Distributed (GED_w) with outlier behavior governed by the tail parameter w (Nelson (1991)). In addition, the estimated parameters include the mean return, $\bar{r}$, and the parameters of the (for $q = ½$, Standard Deviation, or SD-) GARCH or Stochastic Auto-Regressive Volatility (SARV) specifications for K_t^q, (ω, α, $\alpha + \beta$), where $\alpha + \beta$ indicates the degree of persistence in shocks to the return volatility (variance) process. The table displays the point estimates for the parameter and provides standard errors in parentheses below the estimates. The LB_{10}-statistics are Ljung-Box tests for the joint significance of the first ten autocorrelations in returns and squared returns respectively, with P-values given below the statistics. The following 14 moments were used for the Generalized Method of Moments estimation: $\mathrm{E}(R_t)$; $\mathrm{E}(R_t^2)$; $\mathrm{E}(|R_t - \bar{r}|)$; $\mathrm{E}(R_t^4)$; $\mathrm{E}(|R_t - \bar{r}|\ |R_{t-j} - \bar{r}|)$, $j = 1, 3, 6, 10, 13, 16, 19$; $\mathrm{E}(R_t^2\ R_{t-i}^2)$, $i = 2, 4, 7$. The χ^2-test of overall goodness-of-fit for the overidentifying restrictions are χ^2-distributed with degrees of freedom equal to 14 minus the number of estimated parameters. For rows 1, 2 and 4 in Panel B this implies 9 degrees of freedom, while the model in column 3 of Panel B has 8 degrees of freedom, and the associated p-values are provided below the statistics. The weighting matrices for the GMM objective functions were calculated by the Newey and West (1987a) procedure using 75 lags, except that the weighting matrix and the parameter estimates were iterated to convergence, so that the weighting matrix reflects the parameter estimates.

Panel A: Maximum Likelihood Estimation Results

	$\bar{r}$	ω	β	α	$w(r)$	$\mathrm{LB}_{10}(R)$	$\mathrm{LB}_{10}(R^2)$
GARCH; $Z_t \sim N(0, 1)$;	0.035	0.036	0.918	0.068		8.90	3.68
$u_t = Z_{t-1}^2$; $q = 1$	(0.012)	(0.007)	(0.008)	(.006)		(0.54)	(0.96)
GARCH; $Z_t \sim \mathrm{GED}_{w(r)}$;	0.000	0.023	0.943	0.046	1.37	9.47	4.97
$u_t = Z_{t-1}^2$; $q = 1$	(0.007)	(0.005)	(0.007)	(0.006)	(0.026)	(0.49)	(0.89)
SD-GARCH; $Z_t \sim \mathrm{GED}_{w(r)}$;	0.002	0.014	0.948	0.041	1.37	10.6	17.5
$u_t = \|Z_{t-1}\|/\mathrm{E}\|Z_t\|$; $q = ½$	(0.011)	(0.009)	(0.009)	(0.008)	(0.011)	(0.393)	(0.064)

Panel B: Generalized Method of Moments Estimation Results

	$\bar{r}$	$\bar{\sigma}$	$\beta + \alpha$	α	w	$w(r)$	χ^2 (p-val)
SD-GARCH; $Z_t \sim \mathrm{GED}_{w(r)}$;	0.027	1.34	0.958	0.055	[$w(r)$]	1.45	3.06
$u_t = \|Z_{t-1}\|/\mathrm{E}\|Z_t\|$; $q = ½$	(0.019)	(0.041)	(0.037)	(0.018)		(0.060)	(0.96)
SARV; $Z_t \sim \mathrm{GED}_{w(r)}$; $q = ½$	0.027	1.31	0.960	0.087	[$w(r)$]	1.65	2.96
$v_t \sim \mathrm{GED}_{w(r)}$; $u_t = \|v_t\|/\mathrm{E}\|v_t\|$	(0.019)	(0.041)	(0.037)	(0.045)		(0.086)	(0.97)
SARV; $Z_t \sim \mathrm{GED}_{w(r)}$; $q = ½$	0.027	1.31	0.960	0.083	1.45	1.65	2.96
$v_t \sim \mathrm{GED}_w$; $u_t = \|v_t\|/\mathrm{E}\|v_t\|$	(0.019)	(0.041)	(0.036)	(0.190)	(7.53)	(0.087)	(0.94)
SARV; $Z_t \sim N(0, 1)$; $q = ½$	0.014	1.25	0.783	0.164	0.699	[2]	8.47
$v_t \sim \mathrm{GED}_w$; $u_t = \|v_t\|/\mathrm{E}\|v_t\|$	(0.020)	(0.039)	(0.062)	(0.076)	(.540)		(0.49)

the test of $\alpha + \beta = 0.885$, which is high from the perspective of the joint return-volume model but low for the univariate returns model, is strongly rejected for all stocks. These significantly lower estimates of volatility persistence are again accompanied by extremely fat tails in the volatility innovations processes.

VI. Conclusion

This article studies the joint distribution for return volatility and trading volume at the daily level. The contemporaneous relation is derived from a stylized microstructure framework in which informational asymmetries and liquidity needs motivate trade in response to the arrival of new information. The specification is generally consistent with the "Mixture of Distributions Hypothesis" for asset returns, although the volume equation differs from standard specifications. This is due to an accommodation of microstructure features as well as a Poisson, rather than normal, approximation to the limiting distribution of the binomial process that drives trading volume. The test of the modified MDH indicates that the specification is broadly consistent with the data and performs vastly better than the standard formulation. The dynamic features of the joint system is governed by a random mixing variable representing an information flow—or stochastic volatility—variable. The latter is modeled as a SARV process which allows us to estimate the system by GMM under a variety of alternative distributional assumptions regarding the innovations to the information arrival process.

The full dynamic model is estimated and tested for daily data on five common stocks over 1973–1991. Although the standard specification tests generally support the full fledged version of the model, a significant reduction in the estimated measure of volatility persistence takes place when the univariate returns model is expanded to encompass the volume data. Moreover, interesting implications arising from standard restrictions on stochastic volatility models are noted. This points to promising avenues for future research. For example, it is natural to hypothesize that there are two or more types of information arrival processes that have different implications for volume and return volatility persistence. Prominent candidates for news releases and events, that have little effect on volatility persistence but induce a relatively heavy trading volume, are periodic macroeconomic announcements (Ederington and Lee (1993)) and so-called triple witching days where futures and options on stock indices and individual stocks expire simultaneously (Sofianos (1992)). These incidents typically generate only short lived bursts of volatility in returns, but more significantly induce very heavy trading volumes. Consequently, the volume figures will indicate a high information flow activity and a subsequent fast decay, whereas the incident may be barely noticeable in the returns process. Failing to control for such events will bias our volatility persistence measures downward. An exhaustive exploration of this hypothesis requires explicit use of intraday data. However, in addition to the above mentioned phenomena, the intraday return volatility and volume processes contain predictable, deterministic components of a form that is quite distinct from the ones captured by standard volatility models. In fact, Andersen and Bollerslev (1994) demonstrate that failure to account for these effects may result in dramatic downward biases in the estimated volatility persistence measure. The impact is sufficient to explain the very low persistence measures in half-hourly observations found by Foster and Viswanathan (1995). Thus, further progress in this area is likely to stem from theoretical and empirical work that simultaneously deals with the diversity of information arrival processes, the prominent microstructure features of the data, and the interplay between the intraday and the daily volatility processes.

Appendix

Derivation of Equation (5)

From the main text we have that the informed volume associated with a given information arrival j on day t is given by $\mathrm{Po}(Ip_{j,t})$. Since the informed volume is independent across information arrivals, we obtain the following expression for the daily informed volume conditional on the number of arrivals:

$$IV_t|K_t \sim \sum_{j=1}^{JK_t} \mathrm{Po}(I \cdot p_{j,t}) \equiv \mathrm{Po}\left(I \cdot \sum_{j=1}^{JK_t} p_{j,t}\right) \tag{A-1}$$

Next, for $J_t = JK_t$ large, we find

$$\sum_{j=1}^{JK_t} p_{j,t} = JK_t\left[\frac{1}{JK_t}\sum_{j=1}^{JK_t} p_{j,t}\right] \simeq JK_t \mathrm{E}(p_{j,t}) = JK_t\,\frac{\mu}{J} = K_t\mu \tag{A-2}$$

Thus, for large J_t, also relative to I, we apply the following distributional approximation

$$IV_t|K_t \sim \mathrm{Po}\left(I \cdot \sum_{j=1}^{JK_t} p_{j,t}\right) \simeq \mathrm{Po}(I \cdot K_t\mu) \tag{A-3}$$

which corresponds to equation (5).

REFERENCES

Andersen, Torben G., 1992a, Return volatility and trading volume in financial markets: An information flow interpretation of stochastic volatility, Unpublished Ph.D. Dissertation, Yale University.

Andersen, Torben G., 1992b, Volatility, Working Paper, Northwestern University.

Andersen, Torben G., 1994, Stochastic autoregressive volatility: A framework for volatility modeling, *Mathematical Finance* 4, 75–102.

Andersen, Torben G., and Tim Bollerslev, 1994, Intraday seasonality and volatility persistence in financial markets, Working Paper, Northwestern University.

Andersen, Torben G., and Bent E. Sørensen, 1994, GMM and QML asymptotic standard deviations in stochastic volatility models: A response to Ruiz (1994), *Journal of Econometrics*, forthcoming.

Andersen, Torben G., and Bent E. Sørensen, 1995, Estimation of a stochastic volatility model: A Monte Carlo study, *Journal of Business and Economic Statistics*, forthcoming.

Andrews, Donald W.K., 1991, Heteroskedasticity and autocorrelation consistent covariance matrix estimation, *Econometrica* 59, 817–858.

Billingsley, Patrick, 1968, *Convergence of Probability Measures*, New York: John Wiley and Sons.

Bollerslev, Tim, 1986, Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics* 31, 307–327.

Bollerslev, Tim, 1987, A conditional heteroskedastic time series model for speculative prices and rates of return, *The Review of Economics and Statistics* 69, 542–547.

Bollerslev, Tim, Ray Y. Chou, and Kenneth F. Kroner, 1992, ARCH modeling in finance, *Journal of Econometrics* 52, 5–59.

Brock, William A., and Blake LeBaron, 1993, Using structural modeling in building statistical models of volatility and volume of stock market returns: Unabridged version, Social Science Research Center Working Paper, University of Wisconsin, Madison.

Brockwell, Peter, J., and Richard A. Davis, 1987, *Time Series: Theory and Methods*, New York: Springer-Verlag.

Clark, Peter K., 1973, A subordinated stochastic process model with finite variance for speculative prices, *Econometrica* 41, 135–155.

Ederington, Louis H., and Jae Ha Lee, 1993, How markets process information: News releases and volatility, *Journal of Finance* 48, 1161–1191.

Eichenbaum, Martin S., Lars Peter Hansen, and Kenneth J. Singleton, 1988, A time series of analysis of representative agent models of consumption and leisure choice under uncertainty, *Quarterly Journal of Economics* 103, 51–78.

Engle, Robert F., 1982, Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation, *Econometrica* 50, 987–1007.

Epps, Thomas W., and Mary L. Epps, 1976, The stochastic dependence of security price changes and transaction volumes: Implications for the mixture-of-distributions hypothesis, *Econometrica* 44, 305–321.

Feller, William, 1968, *An Introduction to Probability Theory and Its Applications*, vol. 3rd ed., I, New York: John Wiley and Sons.

Foster, Doug, and S. Viswanathan, 1995, Can speculative trading explain the volume-volatility relation?, *Journal of Business and Economic Statistics* 13, 379–396.

Gallant, A. Ronald, Peter E. Rossi, and George E. Tauchen, 1992, Stock prices and volume, *Review of Financial Studies* 5, 199–242.

Gallant, A. Ronald, Peter E. Rossi, and George E. Tauchen, 1993, Nonlinear dynamic structures, *Econometrica* 61, 871–908.

Ghysels, Eric, and Joanna Jasiak, 1994, Stochastic volatility and time deformation: An application of trading volume and leverage effects, Working Paper, University of Montreal.

Glosten, Larry R., and Paul R. Milgrom, 1985, Bid, ask, and transaction prices in a specialist market with heterogeneously informed traders, *Journal of Financial Economics* 14, 71–100.

Hansen, Lars Peter, 1982, Large sample properties of generalized methods of moments estimators, *Econometrica* 50, 1029–1054.

Harris, Larry, 1986, Cross-security tests of the mixture of distributions hypothesis, *Journal of Financial and Quantitative Analysis* 21, 39–46.

Harris, Larry, 1987, Transaction data tests of the mixture of distributions hypothesis, *Journal of Financial and Quantitative Analysis* 22, 127–141.

Harvey, Andrew C., Esther Ruiz, and Neil Shephard, 1994, Multivariate stochastic variance models, *Review of Economic Studies* 61, 247–264.

Heimstra, Craig, and Jonathan D. Jones, 1994, Testing for linear and nonlinear Granger causality in the stock price-volume relation, *Journal of Finance* 49, 1639–1664.

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

Lamoureux, Chris G., and William D. Lastrapes, 1990, Heteroskedasticity in stock return data: Volume versus GARCH effects, *Journal of Finance* 45, 221–229.

Lamoureux, Chris G., and William D. Lastrapes, 1994, Endogenous trading volume and momentum in stock-return volatility, *Journal of Business and Economic Statistics* 12, 253–260.

Lo, Andrew W., and A. Craig MacKinlay, 1990, Data-snooping biases in tests of financial asset pricing models, *Review of Financial Studies* 3, 431–468.

Merton, Robert C., 1980, On estimating the expected return on the market. An exploratory investigation, *Journal of Financial Economics* 8, 323–361.

Milgrom, Paul, and Nancy Stokey, 1982, Information, trade, and common knowledge, *Journal of Economic Theory* 26, 17–27.

Nelson, Daniel B., 1990, Stationarity and persistence in the GARCH(1,1) model, *Econometric Theory* 6, 318–334.

Nelson, Daniel B., 1991, Conditional heteroskedasticity in asset returns: A new approach, *Econometrica* 59, 347–370.

Newey, Whitney K., and Kenneth D. West, 1987a, A simple, positive semi-definite, heteroskedasticity consistent covariance matrix, *Econometrica* 55, 703–708.

Newey, Whitney K., and Kenneth D. West, 1987b, Hypothesis testing with efficient method of moments estimation, *International Economic Review* 28, 777–787.

Phillips, Peter C.B., and Mico Loretan, 1994, Testing the covariance stationarity of heavy-tailed time series: An overview of the theory with applications to several financial datasets, *Journal of Empirical Finance* 1, 211–248.

Richardson, Matthew, and Tom Smith, 1994, A direct test of the mixture of distributions hypothesis: Measuring the daily flow of information, *Journal of Financial and Quantitative Analysis* 29, 101–116.

Ross, Stephen A., 1989, Information and volatility: The no-arbitrage martingale approach to timing and resolution irrelevancy, *Journal of Finance* 44, 1–18.

Sofianos, George, 1992, Expirations and stock price volatility, NYSE Working Paper 92–04.

Stock, Jim H., 1988, Estimating continuous-time processes subject to time deformation, *Journal of the American Statistical Association* 83, 77–85.

Tauchen, George E., and Mark Pitts, 1983, The price variability-volume relationship on speculative markets, *Econometrica* 51, 485–505.

Tauchen, George E., Harold Zhang, and Ming Liu, 1994, Volume, volatility, and leverage: A dynamic analysis, Working Paper, Duke University.