Noise trading, transaction costs, and the relationship of stock returns and trading volume

Charles Kramer*

International Monetary Fund, Room 9-612, 700 19th St. NW, Washington, DC, 20431, USA
Received 13 May 1997; accepted 21 January 1998

Abstract

This article constructs an economic model of a rational trader who operates in a market with transaction costs and noise trading. The level of trading affects the rational trader's marginal cost of transacting; as a result, trading volume (through its effect on marginal cost) is a source of risk. This engenders an equilibrium relationship between returns and volume. The model also provides a simple way to scrutinize this relationship empirically. Empirical evidence supports the implications of the model. © 1999 Elsevier Science Inc. All rights reserved.

JEL classification: G12; G14

Keywords: Capital asset pricing model (CAPM); Noise trading; Trading volume

1. Introduction

The last few years have seen a resurgence of interest in the empirical relationship between stock prices and trading volume. Recent studies find significant statistical relationships between volume and returns, in terms of the level of returns, its volatility and its autocorrelation. Other studies explore intertemporal causality and feedback between returns and volume.1 The linkages between theory and empirics are rarely explored in a way that is most useful for applied researchers, however. I propose a model that links trading volume to equilibrium returns in a simple, intuitive way, in the context of a standard equilibrium model. Both direct and indirect evidence supports the model.

* Corresponding author. Tel.: 202-623-8491; fax: 202-623-6339.
E-mail address: ckramer@imf.org (C. Kramer)

1059-0560/99/$ – see front matter © 1999 Elsevier Science Inc. All rights reserved.
PII: S1059-0560(99)00019-2
Recent results raise a variety of questions about financial markets. Finance researchers and practitioners might wonder whether a relationship of returns and volume is at odds with equilibrium models such as the Capital Asset Pricing Model (CAPM). Regulators might be interested in the implications for market efficiency. For example, transactions taxes are sometimes proposed as a mechanism for reducing the effects of speculation on asset prices. Regulators would like to know whether such measures will work, and what side effects they might cause. One is unable to answer these questions from a purely empirical approach, of course; the analytical tools of financial economics must be applied.

These findings are confusing when analyzed from an economic perspective, though. If price reflects fundamental value, it is hard to say why rational agents should care about trading volume at all. Some authors motivate the relation of volume to returns by arguing that fundamentals traders act as market makers for liquidity traders, or that irrational traders have persistently biased expectations. Others examine the microstructure of asset markets with asymmetric information. It is difficult to link these models to classical theories of asset pricing, however, as they typically use a highly stylized environment and one or two assets. These features also limit the degree to which they can be implemented empirically.

One feature of markets that may engender a relationship between returns and volume is transaction costs. Suppose demand for risky shares by noise or liquidity traders evolves at random. For example, following Delong et al. (1990a), some investors may form their opinions randomly; or, following Campbell, Grossman, and Wang (1993), their risk aversion may fluctuate randomly, leading to perturbation in their optimal portfolio. Suppose also that the market has a fundamentals trader, who trades securities for the usual motive of hedging intertemporal shifts in marginal utility. This fundamental trader must pay transaction costs to trade shares. If the volume of noise trading influences the marginal cost of transacting, it can also influence equilibrium prices. Thus, a relationship of price and volume may exist that is consistent with rational pricing, albeit rational pricing that takes noise trading into account as a risk factor (through its influence on marginal costs). Moreover, there are already some arguments that transaction costs may figure in the relationship between returns and volume. In principle, this relationship could go either way: quiet markets could imply high transaction costs (through an absence of liquidity and high spreads) or low ones (if markets are quiet due to the absence of noise trading, there may be lower costs of gathering information because trades are more informative).

I use a dynamic model with an optimizing representative agent to explore the role that transaction costs might play in engendering a relationship between real stock returns and trading volume. When the marginal cost of transaction varies with the level of trading, an equilibrium relationship between returns and trading volume is implied. The model is amenable to Hansen’s (1982) Generalized Method of Moments (GMM) estimation technique. Estimates and tests from this model support the hypothesized role of volume in returns. Also, indirect evidence (using volume percentile as a conditioning variable in a linear pricing relation) shows a clear relation between equilibrium pricing and trading volume.
2. Model

The model is based on the Lucas (1978) model of intertemporal choice. An infinitely-lived representative agent is assumed to characterize the aggregate choices of a heterogeneous group. Assets are sources of real income, denominated in terms of a single, nonstorable consumption good. At each date, the agent plans future consumption and investment to maximize the expected discounted value of future utilities.

When the representative agent goes to the market to trade, he competes with noise traders. Such traders may buy and sell shares based on rumors, speculation, or bad investment advice; they may also trade to manage liquidity. The rational agent takes the actions of such traders into account, as the volume of trade influences the marginal cost of transaction.

This idea makes sense at a practical level. It is well-known among finance practitioners that market depth influences the cost of transaction. For example, the bid-asked spread is a function of the inventory cost of the market maker, which in turn relates to the level of activity in the market. More frequent trading implies lower carrying and order-processing costs per share for the market maker. On the other hand, if high trading volume reflects a preponderance of well-informed traders, the market maker’s adverse-selection costs will be high, and so must be the bid-asked spread. Market depth is also an important consideration in determining which assets to hold in a managed fund, and the optimal frequency of trading.

The necessary notation is as follows: \( v_t \) denotes the current period’s amount of noise trading in \( K \) assets. I do not model the decision process of noise traders explicitly, but merely treat noise trading as a random process that is not controlled by the rational decision maker. The rational agent knows \( v_t \) at time \( t \), but not \( v_{t+1} \). This makes sense, as real-world traders can observe current market activity as they trade. However, this does not mean that noise trading is irrelevant. Noise trading risk affects decisions now, because those decisions take uncertainty about the future into account. Uncertainty about the future includes uncertainty about future noise trading, and hence future transaction costs and consumption.

I also do not model the trading process, the evolution of the supply of shares, or the equilibration of supply and demand for securities. These are not necessary given the setup of the model. One might picture, however, all agents trading through market makers who hold inventory. In this case asset demands by rational and noise traders are unlinked; demand by noise traders need not equal total supply less demand by rational traders. This also implies that noise traders’ demands are inelastic with respect to transaction costs; otherwise the level of fundamentals trading could influence noise trading.

One potential point of confusion merits a brief aside. In this model, noise trading affects prices only through the constraint on the optimization problem of the rational agent, in particular through the budget constraint. The fundamental value is pinned down not only by marginal utility, but also by marginal cost. Therefore, unlike in models where noise traders directly affect prices, the effects of noise trading are intrinsic to the optimization problem. This is not necessarily a good feature—depending on whether one thinks noise traders directly affect asset prices or not—but it does
mean that the optimization problem is well-posed. Of course, in some theory models, the role of noise traders is considerably different, and modeled in more detail. However, the simplicity of this model affords access to many standard tools, including empirical ones, which one could not use otherwise.

For the rational side, \( U(\cdot) \) denotes a one-period utility function, assumed concave, bounded, and twice differentiable. \( c_t \) denotes period-\( t \) consumption of the single good. \( \beta \in (0,1) \) denotes an impatience parameter, used to discount future utilities. \( x_t \) (Kx1) denotes the number of shares of the claim to output held by the rational agent at date \( t \). \( y_t \) (1xK) denotes this period’s output, distributed to shareholders as dividends. \( p_t \) denotes the (1xK) vector of share prices (in consumption units). \( \Delta x_t = x_t - x_{t-1} \) denotes the (Kx1) vector of changes in the number of shares held by the rational agent between periods \( t-1 \) and \( t \).

\( f(\cdot) \) is meant to capture all the costs of trading to the agent that vary with trading, including brokerage fees. While the bid-asked spread is not modeled explicitly, it can be though of heuristically as one component. With this notation, the agent’s optimization problem for date \( t \) can be written as follows [Eq. (1)]:

\[
\max_{\{c,x\}} V_t = \max_{\{c,x\}} E_t \sum_{\tau=0}^\infty \beta^\tau U(c_{t+\tau})
\]

subject to the sequence of budget constraints [Eq. (2)]:

\[
c_t + p_t x_t + f(\Delta x_t | \omega) \leq (y_t + p_t) x_{t-1}
\]

where \( E_t \) denotes conditional expectation. Because the budget space is convex and the program is discounted, the standard sufficient conditions for a unique equilibrium pricing function \( p(y) \) to exist are satisfied.\(^{13}\)

The first-order conditions for optimal investment \( x_t \), assuming an interior maximum, is

\[
-U'(c_t)(p_t + f_t') + E_t \beta U'(c_{t+1})(y_{t+1} + p_{t+1} + f'_{t+1}) = 0_{1xK}
\]

where \( f'(1xK) \) denotes the gradient of \( f(\cdot;\cdot) \) with respect to \( x_t \) (given \( x_{t-1} \) and \( \omega \)). Eq. (3) is the usual Euler equation of optimality with an adjustment for transaction costs. As usual, the representative agent chooses the best portfolio when the marginal cost of giving up one unit of consumption this period just counterbalances the marginal expected return from investing the unit and consuming it next period. The only difference from the usual case is that these marginal quantities include the cost of investing the unit of consumption this period and cashing it in next period.

### 3. Implications for equilibrium returns

The Euler Eq. (3) can be manipulated in standard ways to yield implications for equilibrium returns.\(^{14}\) For example, rearrangement of Eq. (3) shows that \( E_t(R_{t+1}) \) is linear in \( E(M_{t+1} f_{t+1}/p_t) \), where \( R_{t+1} = (p_{t+1} + y_{t+1})/p_t \) and \( M_{t+1} = U'(C_{t+1})/\).
The expression $E(M_{t+1} f_t + j / p_t)$ is identical to the Brock (1982) expression for the Ross (1976) Arbitrage Pricing Theory (APT) risk premium, where the systematic risk factor is transaction cost as a fraction of price ($f_t + j / p_t$). Hence, marginal cost is a priced risk factor; because marginal cost is a transformation of trading volume, trading volume (here, aggregate trading volume) also behaves like a priced risk factor.

This fact has both time-series and cross-sectional implications. In time-series, well-known features of trading volume (seasonality, for example) that are manifested in time-variation in the covariation between marginal utility and transaction costs could give rise to similar time-series behavior in expected returns. In the cross-section, if $f(\cdot)$ differs across assets, there will be concomitant implications for the cross-sectional distribution of returns. Consider two assets that are identical except for trading costs. Suppose that the first has zero trading costs and the second has trading costs that are a nondegenerate function of trading volume. The second asset will have a positive premium for the risk associated with transactions costs, and hence have a higher expected return than the first. Similar conclusions would follow from assuming that the argument of $f(\cdot)$ is the trading volume for stock $I$, rather than aggregate trading volume, and that the first asset has very high trading volume (so that $f'(\cdot)$ is negligible) and that the second asset has very low trading volume (so that $f'(\cdot)$ is large).

The Euler equation can also be used to derive a consumption CAPM in the presence of transaction costs when there are many assets. Denote by $R^c = (P_{i,t+1} + Y_{i,t+1} + f_{i,t+1})/P_{i,t}$ the rate of return net of transaction costs. Suppose there exist a risk free or `zero-beta' asset and a reference or market portfolio, with cost-adjusted returns denoted $R$ and $R^c$ respectively. Then, manipulation of Eq. (3) yields

$$E_t[R^c_{i,t+1} - R^c_{f,t+1} = \gamma_{i,t}E_t[R^c_{m,t+1} - R^c_{f,t+1}]$$

where

$$\gamma_{i,t} = \frac{\text{Cov}_t(M_{i,t+1}, R^c_{i,t+1})}{\text{Cov}_t(M_{i,t+1}, R^c_{m,t+1})}$$

That is, a conditional consumption CAPM relationship holds for cost-adjusted returns, where $\gamma_{i,t}$ is similar to Breeden’s (1979) conditional consumption risk coefficient. Hence, the relation of the marginal rate of substitution (or fundamentals more generally) to unadjusted returns may fail to explain cross sectional differences in returns (as in for example the January effect or the equity premium). It is also easy to see how volume dynamics could figure in returns dynamics through effects on $\gamma_{i,t}$, and indeed because the effect of volume on $\gamma_{i,t}$ may be asymmetric across stocks (due to asymmetry in costs), how volume dynamics could affect different stocks differently.

These relationships are suggestive of alternative approaches to exploring the empirical relationship of returns and trading volume. However, estimation of a model based on the above would take us far afield from the main goal (to link equilibrium returns to aggregate trading volume). I examine the model Eq. (3) instead. Also, linear pricing relations conditioned on volume are examined in section 6 in order to shed some light on the above ideas.
4. Estimation

4.1. Generalized Method of Moments (GMM) estimation

Because Eq. (3) is a first order condition under rational expectations, $\beta$, parameters in $U(\cdot)$, and parameters in $f(\cdot)$ can be estimated by GMM. A brief exposition of the method is given here; see Ogaki (1993) for a more thorough one.

In Hansen and Singleton’s (1982) notation, Eq. (3) corresponds to $E_t h(w_t + n, b_0) = 0$ where $w_{t+n}$ denotes a $k$-vector of variables observed by the agent and the econometrician as of date $t+n$, $b_0$ is an $l$-vector of parameters unobservable by the econometrician, $h(\cdot, \cdot)$ maps $\mathbb{R}^k \times \mathbb{R}^l$ into $\mathbb{R}^m$, and $E_t$ is the expectations operator conditioned on the agent’s period-$t$ information set. Under rational expectations $h(\cdot, \cdot)$ should not be correlated with any information in the representative agent’s time-$t$ information set (e.g. the agent should not be able to forecast his own optimization errors). This orthogonality condition is then used to motivate taking the products of the sample analog of $h(\cdot, \cdot)$ with various instruments, then minimizing the weighted sum of the time averages of these products over choice of the parameters $b_0$.

Estimation was done in two steps as in Hansen and Singleton (1982). The identity matrix was arbitrarily chosen for the weighting matrix in the first pass. The resulting parameter estimates (which are consistent but inefficient) were then used to estimate the weighting matrix for the second step. The tables below show the results for this second stage. In addition, the covariance matrix of the estimator used in the tests is robust to heteroskedasticity and serial correlation in the errors.16

The estimation procedure also yields a test of the joint restrictions implied by the orthogonality conditions. A quadratic form in the orthogonality conditions (proportional to the objective evaluated at the minimizer) is distributed $\chi^2(r - a)$, where $r$ is the number of orthogonality conditions (restrictions) and $a$ the number of parameters.

This test assesses whether the orthogonality restrictions implied by rational expectations are consistent with the data. A rejection ($\chi^2$ large) implies that they are not.

4.2. Specification of functional forms

For estimation, I assume that utility is of the constant relative risk aversion form,

$$U(c) = \frac{Cy^{\gamma+1}}{\gamma+1}.$$

Because aggregate data are used in estimation, it is implicitly assumed that the consumption growth of the rational agent is the same as aggregate consumption growth. However, the return to noise trading may be lower than returns to fundamentals trading, if noise traders tend to lose to arbitrageurs; it also may be higher than returns to fundamentals trading, if noise traders bear more of the risk they create.17 Hence, the consumption growth of fundamentals traders may actually be higher or lower than aggregate consumption growth. Because there is no way to identify the type of consumption in the data, there seems no alternative to assuming that consumption growth is the same for both types.

Also, because the trading volume of noise traders cannot be distinguished from
that of fundamentals traders in the data, marginal transactions costs are expressed as a function of total volume $v_t = \delta x_t + \omega_t$. This is relatively innocuous; estimation requires the specification of the marginal cost of transaction, rather than the total cost.

A neural network function is used to approximate the unknown marginal cost function. This function takes the form

$$h(v) = \sum_{j=1}^{J} \delta_j g(\alpha_0 + \alpha_j v),$$

where $g(*)$ is the logistic function \(g(x) = e^x/(1+e^x)\).\(^{18}\) There are a number of applications of these functions in finance; Bansal and Viswanathan (1993) find this functional form useful in seminonparametric estimation of a nonlinear pricing kernel. Numerous applications have shown that neural networks serve as good approximations to complicated functions, including the derivatives of unknown functions.

Some experimentation with the form of the transaction-cost function reveals that $J = 1$, $\delta = 1$, and $\alpha_0 = 0$ produces the best results, possibly due to a failure to identify the other components in the data.\(^{19}\) The resulting function (a logistic in $\alpha V_t$) ranges between 0 and 1. It proves useful to scale the function by multiplying it by price, so that marginal cost is expressed as a fraction of the current price. Also, because $g(x) = 1/2$ at $x = 0$, $1/2$ is subtracted from the function so that the case where volume does not enter the Euler equation corresponds to $\alpha = 0$.\(^{20}\) The marginal transaction-cost function is then written

$$f'(v_t) = \left(\frac{e^{\alpha_0 V_t}}{1 + e^{\alpha_0 V_t}} - \frac{1}{2}\right) P_t.$$

With this modification, the Euler Eq. (3) is

$$-U'(c_t)[p_t(1 + f')] + E_t U'(c_{t+1})[y_{t+1} + p_{t+1}(1 + f_{t+1})] = 0 \quad (4)$$

Scaling by price also makes it possible that the model will detect price-affecting noise traders, rather than the effects of transactions costs. Given the setup of the model, these two effects are difficult to distinguish in any event.

4.3. Data

As in Hansen and Singleton (1983), a constant, lagged consumption, and lagged returns are used as instruments. Volume is also used as an instrument, because if the model explains the equilibrium relationship of volume and returns, volume should be uncorrelated with agents' forecast errors.

I use monthly data on consumption, asset returns, and trading volume for estimation. The consumption series is consumption of nondurable goods, seasonally adjusted (unadjusted data are available only at quarterly frequency). These data are divided by an implicit price deflator and population to yield per-capita figures in constant dollars.

Volume is the number of shares traded on the New York Stock Exchange.\(^{21}\) The original data are daily in frequency; the realization for the end of the month is taken as the month’s value, to coincide with the end-of-month prices used in the Center for Research in Security Prices (CRSP) data employed later.
It is well-known that trading volume is nonstationary. The GMM instruments must be stationary, so when volume is used as an instrument, its ratio to its 100-day moving average is used. This daily series is sampled at the end of the month to yield a monthly series. An augmented Dickey-Fuller test (Davidson & MacKinnon, 1993, pp.710–712) rejects the null hypothesis of a unit root in the data at the 1% significance level, and the trend term is insignificant. Raw volume is used in estimation of the Euler equation, however, both because it has better explanatory power and because it makes more sense economically than detrended volume. Because volume appears in both the numerator and denominator of cost-adjusted returns, the nonstationarity of this series is apt to wash out in estimation (the same way that the trends in prices and consumption wash out when returns and growth rates are computed in typical Euler equations). This assertion is supported by augmented Dickey-Fuller tests applied to returns adjusted using estimated parameters for transactions costs (e.g., $R_{c}$). These tests reject the null hypothesis of a unit root in the data.

A single return series, the Standard & Poor’s (S&P) 500, is used for estimation, because price, dividend, and trading volume must be separately identified for each return series. Prices and dividends are imputed from the S&P income and capital appreciation series in Ibbotson and Sinquefeld (1990) by arbitrarily fixing a price number for the first month of the sample and calculating implied prices and dividends recursively. The resulting series are then divided by the implicit price deflator for nondurable goods to yield prices and dividends in constant dollars.

5. Direct tests: estimates of structural parameters

5.1. Model estimates

Estimates were calculated using zero, two, four, and six moving-average (MA) lags to calculate the weighting matrix and two, four, and six instrument lags. When transaction costs were included, I used a grid of starting values for $\beta$ and $\gamma$ around the initial zero-cost estimates. Because the estimates are similar for the various instrument lags, only the estimates from six instrument lags are shown.

Estimates of the model with no transactions costs are displayed in Table 1. Estimated parameters are generally statistically different from zero, and have the appropriate sign. Estimates of the impatience parameter $\beta$ are greater than 1.0, but are all within two standard errors of 1.0. The $\chi^2$ test of the overidentifying restrictions is not rejected at usual significance levels in any of the tests.

Estimates of the model with transaction costs ($\alpha$ estimated) are displayed in Table 2. As with the estimates of the model with no costs, the estimated parameters are generally significant and have the correct sign. The problem with the impatience parameter remains, however. The $\chi^2$ statistic for the overidentifying restrictions implies that the orthogonality conditions are still not rejected by the data.

It is interesting to note that the estimate of the parameter governing marginal transactions costs is negative, and significantly different from zero (recall that $\alpha = 0$ implies that volume plays no role). This implies that volume plays a statistically
No transaction costs.

<table>
<thead>
<tr>
<th>MA lags</th>
<th>( \beta )</th>
<th>( \gamma )</th>
<th>( \chi^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1.0055</td>
<td>-4.3299</td>
<td>10.2795</td>
</tr>
<tr>
<td></td>
<td>(208.867)</td>
<td>(-2.5531)</td>
<td>17</td>
</tr>
<tr>
<td>2</td>
<td>1.0056</td>
<td>-4.2572</td>
<td>11.7402</td>
</tr>
<tr>
<td></td>
<td>(201.638)</td>
<td>(-2.4027)</td>
<td>17</td>
</tr>
<tr>
<td>4</td>
<td>1.0047</td>
<td>-3.8814</td>
<td>11.9204</td>
</tr>
<tr>
<td></td>
<td>(194.434)</td>
<td>(-2.1805)</td>
<td>17</td>
</tr>
<tr>
<td>6</td>
<td>1.0039</td>
<td>-3.6380</td>
<td>12.2628</td>
</tr>
<tr>
<td></td>
<td>(197.342)</td>
<td>(-2.1218)</td>
<td>17</td>
</tr>
</tbody>
</table>

GMM estimates of the parameters describing impatience (\( \beta \)) and risk aversion (\( \gamma \)) for the Euler Eq (3). \( t \)-statistics are in parentheses. \( \chi^2 \) denotes the value of the statistic for the test of the overidentifying restrictions; degrees of freedom are below.

Instruments: Constant, lagged returns, lagged nondurables consumption growth, and lagged volume detrended by its moving average.

significant role in explaining the relationship between real returns and intertemporal substitution. The way in which volume enters the intertemporal substitution problem strongly suggests that this role relates to transactions costs. The negative estimate also implies that marginal costs decrease in volume, which is sensible given the actual structure of brokerage commission schedules, for example. Also, it implies that fundamentals (the relationship of consumption growth and returns) will be more in evidence in deeper markets, so that classical asset pricing models should fit better in more active markets, all else equal.

It is also interesting that the estimates of risk aversion (\( \gamma \)) are larger for the model with transactions costs. Perhaps this is a reflection of the fact that, as argued above, marginal costs serve as an additional risk factor; this implies a larger market risk premium, which is consistent with greater risk aversion. Alternatively, the adjusted returns, which are more variable than unadjusted returns, require more variable marginal rates of substitution to fit the consumption data.

To investigate the influence of October 1987 on these estimates, the estimation was re-done with data from September 1959 to August 1987 (with six MA lags and six instrument lags). The estimates are different from those estimated through 1989, with estimated \( \beta \) around .99, estimated \( \gamma \) about \(-2.64\), and estimated \( \alpha \) about \(-0.19\). All but the estimate of \( \gamma \) are significant at the 1% level. The larger estimate of \( \alpha \) in this sample demonstrates that the significance of volume in returns does not stem solely from the October 1987 crash.

5.2. Specification tests

Finally, the model is subjected to some specification tests. The time-additive setup of the model implies that the residuals from the Euler equation should form a martingale
difference sequence (MDS), as they have a conditional mean of zero. This implies that their level should be unpredictable from lagged variables. Bansal and Viswanathan (1993) exploit this fact to motivate some simple diagnostics for Euler equation residuals. Following their example, I examine regressions of the Euler equation residuals on lagged residuals, lagged absolute residuals, and lagged volume, none of which should help predict the residual.22 I also include dummy variables for October and November 1987 (the market crash and its aftermath) and the uniform reduction of brokerage commissions on the New York Stock Exchange as of January 1972, as a specification test on the model.

These tests require picking a particular specification to test. I use the model with six MA lags and six instrument lags. Results for the MDS specification test are in Table 3. Only the dummy variables for the 1987 stock market crash and for the reduction of brokerage fees are significant at usual levels.23 This means that the model cannot account for these structural breaks. October 1987 contains by far the largest negative real return (about −21%), so it is not surprising that the model does not fit well to this observation.

Another specification issue is the constancy of parameters over the sample. The stability of taste parameters in consumption-based asset pricing equations has been the focus of some recent interest.24 Likewise, the trading environment implicitly modeled in \( f(\cdot) \) has evolved substantially since the late 1950s. Hence, the stability of these parameters ought to be tested. I use the Wald test of Andrews and Fair (1988).25

Table 4 displays the subsample estimates and tests for constant parameters. The impatience parameter remains nearly the same, but the risk aversion parameter

<table>
<thead>
<tr>
<th>Table 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GMM estimates of the model</strong></td>
</tr>
<tr>
<td>[ \beta E [(c_{t+1}^{(1)} - c_t^{(1)}) \gamma R_{t+1}^{(1)}] - 1 = 0. ]</td>
</tr>
<tr>
<td><strong>Transaction costs.</strong></td>
</tr>
<tr>
<td>MA Lags</td>
</tr>
<tr>
<td>---------</td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td>(−3.8619)</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>(−4.3232)</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>(−4.6137)</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>(−4.9628)</td>
</tr>
</tbody>
</table>

GMM estimates of the parameters describing impatience (\( \beta \)), risk aversion (\( \gamma \)), and marginal transactions costs (\( \alpha \)) for the Euler Eq. (4). \( t \)-statistics are in parentheses. \( \chi^2 \) denotes the value of the statistic for the test of the overidentifying restrictions; degrees of freedom are below.

Instruments: Constant, lagged returns, lagged nondurables consumption growth, and lagged volume detrended by its moving average.
Table 3
Diagnostic regressions for Euler equation residuals from Eq. (4) using six instrument lags and six moving-average lags. Sample period: November 1959 to June 1989.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>t-statistic</th>
<th>Estimate</th>
<th>t-statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-0.0126</td>
<td>-1.9172*</td>
<td>-0.0131</td>
<td>-0.7501</td>
</tr>
<tr>
<td></td>
<td>-0.1069</td>
<td>-1.3249</td>
<td>-0.1083</td>
<td>-1.3290</td>
</tr>
<tr>
<td></td>
<td>0.1297</td>
<td>1.6659*</td>
<td>0.1287</td>
<td>1.6188</td>
</tr>
<tr>
<td></td>
<td>-0.0402</td>
<td>-0.7549</td>
<td>-0.0392</td>
<td>-0.7264</td>
</tr>
<tr>
<td></td>
<td>-0.0505</td>
<td>-0.9678</td>
<td>-0.0516</td>
<td>-0.9729</td>
</tr>
<tr>
<td>$\epsilon_{t-1}$</td>
<td>0.00016</td>
<td>0.1280</td>
<td>-0.0011</td>
<td>-0.0892</td>
</tr>
<tr>
<td>$\epsilon_{t-2}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oct. 1987</td>
<td>-0.2176</td>
<td>-3.7631**</td>
<td>-0.2175</td>
<td>-3.7504**</td>
</tr>
<tr>
<td>Nov. 1987</td>
<td>-0.1280</td>
<td>-2.1010**</td>
<td>-0.1275</td>
<td>-2.0831**</td>
</tr>
<tr>
<td>1972±</td>
<td>0.0226</td>
<td>3.5335**</td>
<td>0.0226</td>
<td>3.5094**</td>
</tr>
<tr>
<td>$F$</td>
<td>5.29**</td>
<td></td>
<td>4.098**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(7.348)</td>
<td></td>
<td>(9.346)</td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.0781</td>
<td></td>
<td>0.0728</td>
<td></td>
</tr>
</tbody>
</table>

Estimates for a regression of the GMM residuals on lagged residuals ($\epsilon_t$), lagged absolute value of residuals ($|\epsilon_t|$), lagged trading volume, and dummy variables for October 1987, November 1987, and the reduction of brokerage commissions in 1972. $F$ denotes the test statistic for the null hypothesis that the regression coefficients are jointly zero, with degrees of freedom below.

* Significance at the 5% level.
** Significance at the 1% level.

changes sign (and is small and insignificant in both subsamples). Both parameters are smaller than their full-sample values. The marginal cost parameter is significant in both subsamples. It is smaller than the full-sample value (of about 0.1) after the break and larger before. The tests confirm that the joint change in parameters is significant, and the change in the transaction-cost parameter by itself is also significant.

It is interesting that the transaction-costs parameter decreases in absolute value after January 1972. This implies a smaller role of volume in returns, since the function $f'(\bullet)$ is closer to zero for any given level of volume. This is consistent with the decrease in transactions costs along with the lowering of NYSE brokerage commissions in January 1972.

A more serious problem is the inability to distinguish changes in trading within a fixed cost structure from changes in the cost structure itself (outside known events such as the one mentioned above). One possibility to proxy for changes in cost structure over time is expressing volume as the deviation from long-run trend. Transaction costs could then be considered as depending on volume’s relationship to market capacity (measured as the long-run trend), for example, as in Tsibouris (1993). Experiments with detrended volume and various marginal-cost functions in Eq. (3) were unsuccessful; the resulting parameters were often significant but took on implausible values. Also, in the instances when a significant marginal-cost function could be identified, the parameters of that function still displayed substantial instability.
6. Indirect tests

6.1. Volume as a conditioning variable in the risk-return relationship

Estimation of the Euler equation requires strong assumptions about the functional forms for utility and transactions costs. Other questions of specification, such as stability and time-separability, may obscure the results as well. For these reasons, it is useful to compare results from a simpler (and, I hope, more robust) procedure. The next set of tests posit trading volume as a conditioning variable in linear pricing relations. In particular, the relation of risk—here, market risk and consumption risk—to average return across stocks is examined over high-volume and low-volume subsamples. If volume is irrelevant to this relationship, it should look the same in both subsamples.

The exercise is well motivated by the discussion of equilibrium pricing in section 3. There, it was shown that there is a risk premium associated with the marginal cost of transaction. If the marginal cost of transaction were always zero, the model would collapse to the standard (classical) asset-pricing model. It follows (from the continuous differentiability of the key functions) that when marginal cost is small, the pricing relationship should be close to the standard case. Likewise, if (as the estimates in the previous section suggest) the marginal cost of transaction falls with trading volume, then marginal cost will be low in high-volume months. In those months, then, classical models should fit well. In particular, the risk premium for transaction costs should be small. Conversely, in months when volume is low, the marginal cost of transaction should be high, and classical models should fit poorly. This notion is explored below.
by sorting returns into high-volume and low-volume months, and fitting separate cross-sectional asset-pricing models to each sample.

The classical tests of the linear asset pricing theory employ a two-stage approach to estimate cross-sectional models of the form

$$E(r_i) = \lambda_0 \beta_i$$

(5)

where $\beta_i$ is a measure of the riskiness of stock $I$, $\lambda_1$ is the premium for risk-bearing over the riskless rate ($\lambda_0$), and $r_i$ is the return on stock $I$ for period $t$. $\beta_i$ and $\lambda_1$ may be scalars or vectors, depending on whether one or many risk factors are priced. For example, in the CAPM, $\beta_i$ is the covariance of the return on stock $I$ with the market return divided by the variance of the market return, or the slope coefficient in the time-series regression

$$r_{it} = \alpha_i + \beta_i r_{mt} + \epsilon_{it}$$

(6)

where $r_{mt}$ is the return on a stock market index. This suggests the following estimation strategy: estimate $\beta_i$ for a group of stocks $I = 1, \ldots, N$ using time series data as in Eq. (6), then regress average returns on these estimates to yield estimates of $\lambda_{1i}$, the market risk premium. That is, the cross-sectional regression

$$\bar{r}_i = \lambda_0 + \lambda_1 \hat{\beta}_i + \eta_i$$

(7)

yields estimates $\lambda_0$ and $\lambda_1$. The parameter $\lambda_1$ characterizes the tradeoff between risk and expected return. If increased risk is compensated by increased expected return, then the estimate $\lambda_1$ should be positive and significantly different from zero.

Two types of circumstances can give rise to a model such as Eq. (5). Under certain restrictions on preferences or the distribution of returns, for example, $\beta_i$ is the slope coefficient from a regression of stock $I$'s return on the return on the market portfolio (e.g., the CAPM). Under restrictions on the economy, $\beta_i$ is the slope coefficient from a regression of stock $I$'s return on real consumption growth (e.g., a variant of the Consumption CAPM or CCAPM).

The fundamental question is whether aggregate volume plays any role in the equilibrium pricing relationship Eq. (5). This question is posed by dividing a sample of stock returns into three parts, corresponding to the months of the 1/3 highest, 1/3 lowest, and 1/3 median aggregate volume on the New York Stock Exchange. The ratio of volume to its 100-day moving average is used to form this ranking, rather than its level. Otherwise, all the high-volume months would come toward the end of the sample, and all the low-volume months toward the beginning; this might confuse shifts due to volume with shifts due to institutional phenomena.

Relationships as in Eq. (7) are then examined over the full sample and the high- and low-volume subsamples. If volume is irrelevant to stock returns, the relationship of risk and return, or $\lambda_1$, should be the same in all three samples.

The estimation procedure is contaminated by errors-in-variables bias, as estimated (rather than population) $\beta_i$ are used in the second pass regression. However, because this estimation error shrinks as the time-series sample increases, the estimators $\lambda_0$ and
\( \lambda_1 \) are consistent (as the time-series sample size increases) under moderate restrictions on the processes governing the time-series behavior of returns. They are also asymptotically normal; the usual ordinary least squares standard errors are inconsistent, though, and here are replaced by standard errors adjusted for estimation error (Shanken, 1992).

6.2. Data

Monthly data on individual stock returns for the sample period May 1959 to June 1987 are taken from the CRSP data tape. This sample consists of both NYSE and American Exchange stocks. The period of the market crash of 1987 and months thereafter is excluded from the sample, as the sample is large enough to produce good estimates from this technique without the last few years of data. Individual stock returns are grouped into 50 equally-weighted portfolios on the basis of the previous month’s size (price times number of shares outstanding). Grouping returns into portfolios reduces the estimation error of the first-pass estimates by decreasing firm-specific variation (which tends to cancel out in portfolios). Grouping by size increases the power of tests based on the cross-sectional regression by assuring that the sample will have a good dispersion of risk and return across portfolios.

Returns on the S&P 500, the equally-weighted CRSP index, and the value-weighted CRSP index are also taken from CRSP. The return on the one-month Treasury bill with maturity closest to one month (Ibbotson & Sinquefeld, 1990) was used as the risk-free rate of return. This was subtracted from all returns series (including the market return \( r_{mt} \)) before estimation, as net-risk-free returns are indistinguishable from real returns at one-month intervals. This also implies that \( \lambda_0 = 0 \). Consumption of nondurables in 1982 dollars was used for the consumption series.

6.3. Results

Table 5 shows the results for the various linear pricing models, with consumption, the S&P 500, the CRSP equally-weighted index, and the CRSP value-weighted index as risk factors \((r_{mf})\). I find, as do Mankiw and Shapiro (1986) and Chen, Roll, and Ross (1986), that consumption risk does not explain average return. It is interesting to note that both by itself and combined with the market return, the estimated premium for consumption risk is largest in high-volume months (though still not significant at usual levels).

In contrast, market risk explains average return for high-volume months, with a risk premium significantly different from zero. This is true regardless of how the market portfolio is defined. In low-volume months, however, market risk is no longer significant, and point estimates fall by a factor of 10. The results for all months together are similar to those for high-volume months.

Figs. 1 and 2 Highlight the differences in risk and return in high- and low-volume months. In these figures, the squares show pairs of \( \beta_I \) and \( \bar{r}_I \) for the 50 portfolios and the straight line shows the fitted values \( \bar{E}r_I = (\lambda_0 + \lambda_1 \beta_I) \). Fig. 1 shows market beta...
Table 5
Estimates of linear pricing models

\[ E(r_t) = \lambda_0 + \lambda_1 \beta_t \]


<table>
<thead>
<tr>
<th></th>
<th>All</th>
<th>High</th>
<th>Low</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCAPM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \lambda_0 )</td>
<td>0.0063</td>
<td>0.0062</td>
<td>0.0108</td>
</tr>
<tr>
<td>( t )</td>
<td>5.1606</td>
<td>4.0269</td>
<td>2.9290</td>
</tr>
<tr>
<td>( \lambda_m )</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>( t )</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>( \lambda_c )</td>
<td>0.0014</td>
<td>0.0039</td>
<td>-0.0013</td>
</tr>
<tr>
<td>( t )</td>
<td>0.5148</td>
<td>1.0067</td>
<td>-0.5361</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CAPM (Equal-weighted CRSP)</th>
<th>CAPM (S&amp;P 500)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \lambda_0 )</td>
<td>-0.0090</td>
</tr>
<tr>
<td>( t )</td>
<td>-2.4165</td>
</tr>
<tr>
<td>( \lambda_m )</td>
<td>0.0181</td>
</tr>
<tr>
<td>( t )</td>
<td>3.5432</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CAPM (VW)/CCAPM</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \lambda_0 )</td>
</tr>
<tr>
<td>( t )</td>
</tr>
<tr>
<td>( \lambda_m )</td>
</tr>
<tr>
<td>( t )</td>
</tr>
<tr>
<td>( \lambda_c )</td>
</tr>
<tr>
<td>( t )</td>
</tr>
</tbody>
</table>

Estimates of CAPM and Consumption CAPM models over different volume regimes. \( \lambda_m \) denotes the estimated market risk premium and \( \lambda_c \) the estimated consumption risk premium. \( t \) denotes the test statistic (asymptotically distributed \( N(0,1) \)) for the hypothesis that the corresponding parameter is zero, based on the Shanken (1992) standard error (adjusted for measurement error in \( \beta \)).

(measured relative to the value-weighted CRSP index) versus average return for high-volume months. Beta varies from about .9 to about 1.4, while average return runs from about 0 to almost 2% per month. Moreover, there is a significant relationship between the two. Fig. 2 shows market beta versus average return in low-volume months. Here, the spread in both risk and return is smaller; beta varies from about .9 to about 1.15, while average return varies from about 0.4% per month to about 1.5% per month. Moreover, the relationship between the two is insignificant (the reader should note that the two plots are not strictly comparable since the x-axes are scaled differently).

These results are thought-provoking in light of the discussion of the previous section. Consider for example the case where the marginal cost of transaction decreases in volume. During periods of high trading volume, the marginal cost of transaction will be relatively insignificant, and the pricing relation will look similar to one without transaction costs (the traditional models). When trading volume is low, fundamentals
are more likely to be obscured by noise trading (through transactions costs). This is consistent with the estimates of declining marginal transaction costs from the model of the previous section.

7. Conclusions

I examine a dynamic economy where a rational representative agent makes intertemporal choices about consumption and asset holdings. The rational agent operates in
a market with noise traders, whose activities affect the marginal cost of transaction. This implies that trading volume will play a role in determining the real equilibrium prices of assets. The model is amenable to estimation using GMM. Both direct and indirect estimates lend support to a role for volume in equilibrium asset returns. In particular, both imply that variations in volume when average volume is high affect equilibrium returns less than when average volume is low.

In this model, frictions do not eliminate the effects of noise trading on prices; rather, they cause the (marginal-cost) risk associated with noise trading to be priced by rational agents. It is not hard to see why: rational agents, who create the link between fundamentals and asset prices, are subject to transactions costs as well. The resulting equilibrium is not inefficient per se; risk is still correctly priced. However, there is an additional source of risk that will confound analysis that uses traditional tools (such as the CAPM) to assess market equilibrium and the cost of capital. This assertion is consistent with the evidence that the CAPM relationship of return to risk is different in high-volume months than in low-volume months.

Acknowledgments

I am grateful to Erik Benrud, Craig Hiemstra, Jonathan Jones, Robert Flood, Lee Redding, and an anonymous referee for helpful comments.

Notes

2. See Niehans (1994), for example.
4. Recent examples are Blume, Easley, and O'Hara (1994) and Easley and O'Hara (1992).
5. See Demsetz (1968) and Epps (1976).
6. These mirror similar considerations of the rate of information transmission as a function of market depth. See Easley and O'Hara (1992).
7. See Black (1986) for a discussion and motivation of the economics of noise trading.
8. See Demsetz (1968) and Cohen et al. (1980).
9. Some costs of trading relevant to this decision—so-called “market impact” costs—are not really costs per se but represent the effect of large trades on
prices. These are different from the transactions costs implied by the model, which more closely resemble brokerage fees and the bid-asked spread.

10. Similar assumptions are made in the microstructure literature (e.g., Easley & O’Hara, 1992). Such assumptions mean that the issue of no-trade equilibria can be ignored (Milgrom & Stokey, 1982).

11. One exception might be if agents place market orders that brokers execute with a lag (for example, in a call auction market). The markets examined here (the NYSE and AMEX) are more like continuous auctions than call auctions, though.

12. That is, for a given \( \omega \), \( f(|\Delta x|\omega) \) is the cost of trading |\( \Delta x \) shares. Quasiconvexity ensures that the budget set is convex.


14. For brevity, the algebra is omitted.


17. See Delong et al. (1990a).


19. While the parameters might be locally identified, Granger and Teräsvirta (1993, p. 125) point out that neural network functions are not globally identified.

20. Omitting this transformation produced qualitatively similar results (e.g. negative and significant estimates of \( \alpha \)).

21. I am grateful to Craig Hiemstra and Jonathan Jones for providing this data.

22. Bansal and Viswanathan (1993) recommend including lagged absolute values as a check for neglected nonlinearity. Also, strictly speaking, lagged volume should be orthogonal to the residual because it is used as an instrument. Including it serves as a second check that the model explains the relationship of volume and returns (e.g., that the failure to reject the overidentifying restrictions is not due to a power problem).

23. A dummy variable for the SEC’s deregulation of minimum brokerage commissions on the NYSE (May 1975) was insignificant when added to these regressions.


26. Shanken (1992) reviews these tests.

27. Mankiw and Shapiro (1986) discuss these two models.

28. The dates associated with high- and low-volume months are scattered throughout the sample. This scheme destroys any serial correlation in the residual, but because such correlation is rarely exploited in this type of test, no real harm is done.


30. The current month’s size was not used as this would induce a spurious correlation between average return and portfolio rank.

References


