Bootstrap Refinements in Tests of Microstructure Frictions

Thomas J. George
e-mail:*tom-george@uh.edu*
C.T. Bauer College of Business
University of Houston
Houston, TX 77204
713-743-4762

Chuan-Yang Hwang
e-mail:*cyhwang@uxmail.ntu.edu.sg*
Nanyang Business School
Nanyang Technological University
Singapore

and

Tavy Ronen
e-mail:*tronen2@nyc.rr.com*
Graduate School of Management
Rutgers University
Newark, NJ 07102

May 2009

We are grateful to two anonymous referees, seminar participants at the University of Iowa, and especially Joel Horowitz and Janis Zvingelis, for helpful discussions and suggestions. George acknowledges financial support from the C.T. Bauer Professorship.

Bootstrap Refinements in Tests of Microstructure Frictions

Abstract

Bootstrapping is often used as a substitute for asymptotic distributions when the latter are not available. Recent developments in the theory of the bootstrap show that combining the bootstrap with a known asymptotic distribution yields inferences that improve on those drawn from asymptotic distribution theory or bootstrapping alone. We review the key to obtaining the improvement and compare asymptotic and bootstrap inferences of three variance ratio tests used in microstructure research. The more precise bootstrap inferences lead to conclusions that differ from those found in extant research on transitory volatility. Asymptotic tests are biased toward rejection, and bootstrap and asymptotic critical values are not generally close to each other. These findings suggest that the more precise bootstrap inferences should be used in future applications of these tests, as well as in various other empirical applications where intradaily or other high frequency data are modeled using vector autoregressions.

1. Introduction

Since Efron (1979), the bootstrap has been applied with great success in finance research as a way to draw inferences when either the finite-sample or asymptotic distributions of test statistics are not known [Goetzmann and Jorion (1992), Goetzmann (1993), Jones, Kaul and Lipson (1994), Sullivan, Timmermanm and White (1999), and Kosowski, Timmermann, Wermers and White (2006) are prominent examples]. In these applications, bootstrapping is used as an *alternative* to drawing inferences from asymptotic distribution theory. More recently, Hall and Horowitz (1996) and Horowitz (2001) have shown that the bootstrap can be used to *improve* inferences in settings where the asymptotic distributions of test statistics are known. By incorporating features of the asymptotic distribution into the bootstrap procedure, the bootstrap distribution extracts more information from the sample than is captured by an asymptotic distribution alone, or by a bootstrap distribution that is constructed without incorporating information about the asymptotic distribution.

The nature of the improvement is in the closeness with which the bootstrap distribution approximates the true distribution of a test statistic under the null. In order to achieve this improvement the bootstrap must be structured in a manner different from procedures used when the asymptotic distribution is not known. In this paper, we review the key to attaining the improvement, or *refinement,* and explain how some popular approaches to bootstrapping contrast with it. We then turn our attention to a specific application in the microstructure of securities markets.

Microstructure research is a fruitful setting in which to exploit bootstrap refinements for three reasons. First, asymptotic distributions of many of the test statistics used in microstructure research are assumed known. In cases where the assumed asymptotic distribution is correct, the refinement we discuss is available. In cases where the assumption is incorrect, the procedure described here still produces valid inferences even though asymptotic distribution theory does not. Second, empirical microstructure models often employ vector autoregressions that reduce the data generating process to a sequence of disturbances that are assumed to be iid [see Hasbrouck (1991a, 1991b, 1993, 1995) and Huang (2002) among others]. Though refinements are available when dependence is present, the bootstrap algorithms are especially simple to implement in the iid case. Third, the vastness of many microstructure datasets suggests the information in samples is especially rich when compared to what can be summarized in an asymptotic distribution. This richness plays to the strength of the bootstrap, enabling it to achieve improvements over inferences based on asymptotic distribution theory that are potentially substantial.

The application we consider is a trio of variance ratio tests of whether market closures add noise to security prices at market reopenings. Existing studies that use these tests all draw inferences using asymptotic distribution theory and reach conflicting conclusions. We begin by documenting that the test statistics used in these studies do not conform to their asymptotic distributions, even when a very strong version of the null is true. We then reconsider these tests with the bootstrap refinement, using data that is similar to the data used in the earlier studies. The conclusions of the bootstrap tests are similar. Inferences from all three tests indicate that for a large majority of stocks, market closures do not add significant noise to security prices at reopenings.

We also use the resampling approach that generates bootstrap distributions to characterize properties of these variance ratio tests. This helps explain the conflicting inferences among the asymptotic tests, and provides some guidance for which test(s) will suffer the least distortion when used in future applications. Trading technologies, reporting requirements and the resulting transparency of equity and bond markets have changed a great deal over the last decade, and the richness in quality and frequency of available data has also increased. These developments provide new data and give rise to new questions regarding the quality of markets. Since there are many ways to estimate and draw inferences about market quality, it is important to know which metrics and econometric tests perform well, and how best to exploit the information in the data to draw conclusions. We believe the bootstrap methods discussed here can play a useful role in addressing these issues to facilitate the study of these rapidly evolving markets.

The next section explains how the bootstrap must be done to get an improvement over asymptotic tests. Section 3 describes the background on three variance ratio tests to which we apply these ideas, and spells out how we conduct the bootstrap and sensitivity tests. Section 4 presents the results. Section 5 concludes. Technical descriptions of the tests and their statistics are relegated to the Appendix.

2. Overview of the Bootstrap Refinement

There are several excellent articles that survey and critique the various approaches to bootstrapping used in econometrics and finance [Maddala and Li (1996), Berkowitz and Killian (2000), Ruiz and Pascual (2002) and MacKinnon (2006)]. These studies are mostly concerned with alternative approaches to bootsrapping and their relative performance in specific circumstances when used as a *substitute* for asymptotic distribution theory. In those applications, the goal is to obtain an alternative to an asymptotic distribution that might not be known.

Our focus is different. We are concerned with illustrating how knowledge of the asymptotic distribution can be used *in conjunction* with bootstrapping to yield an improvement over standard asymptotic inferences. In this sense the bootstrap provides a *refinement* of asymptotic inferences, rather than a substitute.[1] The approach is different from how bootstrapping is done in most finance research. In fact, we explain how common bootstrap procedures that do not incorporate information from the asymptotic distribution fit into this framework. We then apply the procedure to three hypothesis tests in market microstructure that yield conflicting conclusions when conducted using asymptotic distribution theory.

2.1 Approximating the Distribution of a Test Statistic

When the finite sample distribution of a test statistic is not known, inferences must be based on an approximation to the distribution of the statistic. Consider the standard normal approximation to the distribution of a t ratio as an example. Suppose we wish to test the null hypothesis that a population parameter θ is equal to a particular value θ_o, against the alternative that $\theta < \theta_o$. We would first compute a consistent estimate $\hat{\theta}$ and form the t ratio, $n^{1/2}\left(\frac{\hat{\theta}-\theta_o}{\hat{\sigma}}\right)$, where $\hat{\sigma}^2$ is a consistent estimate of the asymptotic variance of $n^{1/2}\hat{\theta}$. Second, we compare the t ratio to z_α, the α percentile of the standard normal distribution where α is our tolerance for type I error.

The justification for this procedure is that if $\theta = \theta_o$ then the true finite sample distribution function of $n^{1/2}(\hat{\theta}-\theta_o)/\hat{\sigma}$ can be written as

$$P\left\{n^{1/2}(\hat{\theta}-\theta_o)/\hat{\sigma} \le x\right\} = \Phi(x) + \hat{R}_n(x) \tag{1}$$

where $\Phi(x)$ is the standard normal distribution function, and $\hat{R}_n(x)$ is an approximation error that tends to zero as the sample size, n, gets large. This familiar approximation has an error of order $n^{-1/2}$ meaning that $\hat{R}_n(x)$ tends to zero at the same rate as $n^{-1/2}$. Bootstrapping is a way to construct an approximating distribution that can be used instead of the asymptotic distribution (i.e., instead of the standard normal in this case). When done in a particular way, the error associated with the bootstrap approximation converges to zero faster than does the error associated with the asymptotic distribution.

The principle behind the bootstrap is that the relationship between a population and a sample is also shared between the sample and a sample drawn from it. Accordingly, the relationship between

[1] Our discussion summarizes results derived in Hall (1992). An excellent exposition of these results and extensions is given in Horowitz (2001). In our applications, dependence is modeled using vector autoregressions (VARs), which means that the data generating process can be reduced to iid draws of VAR disturbances. Hall and Horowitz (1996), Zvingales (2001), and Hardle, Horowitz and Kreiss (2002) address situations with more general dependence. Our applications also satisfy the assumptions of their "smooth function model," so the rest of our discussion takes these as given. The main assumptions are that the test statistic can be written as a smooth function (i.e., with sufficiently many continuous derivatives) of sample moments of the data, that sufficiently many moments of the data are finite, and that the distribution function of the data is absolutely continuous.

a population parameter and the (unknown) distribution of an estimator can be approximated by the relationship between the sample estimate and the distribution of such estimates produced by resampling from the sample. This means the "bootstrap version" of a test statistic replaces population values with sample values, and sample values with estimates from resampling in order to construct the distribution.

In the example above, the bootstrap version of the test statistic is $n^{1/2}(\theta^* - \hat{\theta})/\sigma_*$ where θ^* and σ_* are estimates from a resampling of the original sample. The sample estimate $\hat{\theta}$ replaces θ_o because θ_o is the population parameter value if the null is true, and $\hat{\theta}$ is the parameter value for the "population" from which resampling is conducted. By repeatedly resampling and reestimating θ^* and σ_*, the distribution of the bootstrap version of the test statistic can be tabulated. Its distribution function is conditional on the sample, $\mathcal{X}$, which we denote by

$$P\left\{n^{1/2}(\theta^* - \hat{\theta})/\sigma_* \leq x \mid \mathcal{X}\right\}. \tag{2}$$

Moments of this distribution are not constrained to match those of the standard normal distribution, which is why bootstrapping can yield a better approximation. Resampling captures bias and skewness in the true (unknown) distribution of the test statistic that the normal approximation necessarily ignores.

If the distribution of a statistic can be represented as in equation (1), then it turns out that a representation of the same form exists for the bootstrap version of the statistic:

$$P\left\{n^{1/2}(\theta^* - \hat{\theta})/\sigma_* \leq x \mid \mathcal{X}\right\} = \Phi(x) + R_n^*(x). \tag{3}$$

As with equation (1), the error $R_n^*(x)$ converges to zero at the rate of $n^{-1/2}$. In other words, the (unknown) finite sample distribution of the test statistic, and the bootstrap distribution of the test statistic, both converge to the standard normal distribution at the rate of $n^{-1/2}$.

The substance of the boostrap refinement is that the distributions *converge to each other faster* than they converge to the standard normal when the null is true. Subtracting (3) from (1) yields

$$P\left\{n^{1/2}(\hat{\theta} - \theta_o)/\hat{\sigma} \leq x\right\} = P\left\{n^{1/2}(\theta^* - \hat{\theta})/\sigma_* \leq x \mid \mathcal{X}\right\} + \hat{R}_n(x) - R_n^*(x), \tag{4}$$

noting that the $\Phi(x)$ terms disappear. The bootstrap refinement is based on the fact that the *difference* between the errors converges to zero faster than the individual errors converge to zero. In this example, one can show that $\hat{R}_n(x) - R_n^*(x)$ converges to zero at the rate of n^{-1} when the null is true. The bootstrap approximation to the left-hand-side of (4) is *an order of magnitude better* than the asymptotic standard normal approximation.

The crucial element to attaining the refinement is that the leading Φ terms cancel. In other words, the asymptotic distributions of the sample and bootstrap versions of the statistics must be identical. This will only happen if the asymptotic distribution does not depend on any unknown parameters. Statistics with this property are said to be *asymptotically pivotal*.

One must know the asymptotic distribution of the parameter of interest in order to construct an asymptotically pivotal statistic. In the particulars of this example, one must know that under the null, the asymptotic distribution of $n^{1/2}\hat{\theta}$ is normal with a mean of θ_o and a variance that is consistently estimated by $\hat{\sigma}^2$. Without this knowlege, it is impossible to know how to construct a statistic that is asymptotically pivotal. It is by bootstrapping an asymptotically pivotal statistic that one incorporates information about the asymptotic distribution into the bootstrap procedure, which in turn delivers the refinement.

What if the asymptotic distribution of $\hat{\theta}$ is not known? In that case, it is common to approximate the distribution of the non-studentized statistic $n^{1/2}(\hat{\theta} - \theta_o)$ using the distribution of its bootstrap counterpart. This results in a bootstrap distribution of the *estimator* $\hat{\theta}$ rather than the (studentized) test statistic. In this case, the Φ terms do not disappear from equation (4). Instead, the difference $\Phi(x/\sigma) - \Phi(x/\hat{\sigma})$ appears, which converges to zero at the same rate as $n^{-1/2}$. In this case, the bootstrap approximation is no better than that of asymptotic distribution theory. However, note that this formulation of the bootstrap delivers an approximation with an error whose order is no larger than the error associated with asymptotic distribution theory. Thus, bootstrapping the distribution of an estimator that satisfies the assumptions of the smooth function model provides an alternative to deriving an asymptotic distribution that is analytically intractable.

Another approach that is sometimes used is that of *bootstrap standard errors*. This involves forming the test statistic $n^{1/2}(\hat{\theta} - \theta_o)/\sigma_*$ and comparing the result to the standard normal distribution. Assuming the asymptotic distribution is indeed normal, this is simply an asymptotic test where σ_* is used in place of $\hat{\sigma}$ as a consistent estimator of σ—it clearly does not incorporate bias and skewness of the distribution of the test statistic. As an asymptotic test, its approximation error is of the order $n^{-1/2}$.

A somewhat different situation arises when a statistic is *improperly studentized;* in particular, when $n^{1/2}(\hat{\theta} - \theta_o)$ is divided by an estimator, say $\hat{\kappa}$, that converges in probability to κ, which is something other than σ. This is a case where the incorrect asymptotic distribution is assumed, and arises in one of the tests we examine below. Since $\kappa \neq \sigma$, the asymptotic distribution of $n^{1/2}(\hat{\theta} - \theta_o)/\hat{\kappa}$ is obviously not standard normal, so an asymptotic test based on the standard normal is not even valid. However, a test based on the bootstrap distribution of this statistic *is* asymptotically valid. The Φ terms do not disappear in equation (5). The difference $\Phi(x\kappa/\sigma) - \Phi(x\hat{\kappa}/\hat{\sigma})$ appears, which is of the order $n^{-1/2}$. In this case, the bootstrap delivers the same degree of accuracy that could have been attained by properly studentizing and using an asymptotic test, or by using a bootstrap test on a non-studentized statistic such as the estimator $\hat{\theta}$. Since the proper information about the asymptotic distribution has not been incorporated into the bootstrap procedure, the bootstrap refinement is not achieved.

2.2 Extensions

The analysis above relates to a one sided distribution function. For two sided distribution functions where the critical values are symmetric about zero, the bootstrap provides an even greater improvement over asymptotics. The standard normal approximation error is of the order n^{-1}, while that of the bootstrap is of the order $n^{-3/2}$.

The results described so far indicate that when approximating the *distribution function* of an asymptotically pivotal test statistic, the error is of a smaller order when its bootstrap distribution is used than when its asymptotic distribution is used. Analogous results hold for approximating the probability α of rejecting a correct null hypothesis using bootstrap and asymptotic critical values. The error in the true rejection probability as an approximation of α is of a smaller order when bootstrap critical values of an asymptotically pivotal test statistic are used than when asymptotic critical values are used. Specifically, for a two sided test, using asymptotic critical values leaves an error in the rejection probability of the order n^{-1}, and using symmetrical bootstrap critical values leaves an error of the order n^{-2} [see Horowitz (2001) section 3.3 for details]. This means that in large samples, the probability of rejecting a correct null using the bootstrap refinement is closer to the nominal size of the test than is the rejection probability of an asymptotic test.[2]

These results, of course, do not say that an arbitrary bootstrap approach to inference will perform better than using an asymptotic distribution. If the bootstrap is merely a substitute for an asymptotic inference, there is no a-priori reason to believe one will outperform the other. However, bootstrapping an asymptotically pivotal statistic will perform better in situations where both approaches are feasible and justifiable (i.e., the sample is large, an asymptotically pivotal test statistic is available, and the assumptions of the smooth function model hold). This means that if the nature of the experiment is such that the researcher has a choice between basing inferences on the bootstrap refinement or asymptotic distribution theory, *combining* information in the

[2] Asymptotic critical values are the numbers $-z_a$ and $+z_a$ nearest to zero between which is $1-\alpha$ of the density of the standard normal distribution. Symmetric bootstrap critical values are the numbers $-z_b$ and $+z_b$ nearest to zero between which is $1-\alpha$ of the mass of the bootstrap distribution of the test statistic.

asymptotic distribution with the bootstrap will perform better. This situation characterizes many problems addressed in microstructure research. In these cases, the only reason not to bootstrap an asymptotically pivotal statistic would be that the potential benefit is small or the programming is difficult.[3]

3. Empirical Application of the Bootstrap Algorithm

3.1 Background on Pricing Error Tests

The pricing error tests we compare assess the impact of trading suspensions on the degree to which security prices deviate from true values at market openings. Such deviations indicate an impairment to market quality and imply greater trading costs to demanders of liquidity. Existing studies compare the volatility of transitory components of returns at the open and the close. Most evidence from the NYSE and other equity markets indicates that 24-hour returns computed from opening prices are, on average, about 20% more volatile than those based on closing prices. Despite the large magnitude of this difference, its interpretation has not been settled.

Initially, the difference in variances was attributed to the trading mechanism employed at the open [Amihud and Mendelson (1987), and Stoll and Whaley (1990)]. Since trading on the NYSE opens with a call auction, the interpretation of the evidence was that call auctions produce noisier prices than continuous trading. Subsequent evidence that volatility is not greater for call auctions that occur *during* the day in Japan led to the conclusion that the greater volatility of opening prices is attributable to uncertainty about the security's value at the open resulting from an overnight non-trading period [Amihud and Mendelson (1991)]. Consistent with this interpretation, Gerety and Mulherin (1994) find that the volatility of 24-hour returns computed from prices sampled later and later within the day exhibit a declining pattern. However, the evidence in Forster and George (1996) is inconsistent with this interpretation. They document that returns to New York stocks that are traded (overnight) in London and Tokyo have volatility patterns that are similar to New York stocks that do not trade in overseas markets. Their interpretation is that larger order imbalances cause greater price concessions to liquidity providers at the open than at the close.

In addition to the debate about the economic explanation of greater volatility at the open, two studies question the validity of the measurement techniques used in the earlier studies, and propose distinct alternatives [Ronen (1997) and George and Hwang (2001)]. When these alternatives are

[3] One important example is a modestly sized sample of *dependent* data where the dependence cannot be modeled down to iid disturbances. Even when implemented optimally (i.e., with asymptotically pivotal statistics), the bootstrap refinement for dependent data is small. The bootstrap error is of the order $n^{-5/4}$ versus n^{-1} for the asymptotic test [see Zvingales (2001)].

used to analyze data for NYSE stocks, the hypothesis that pricing errors are the same at the open and the close is *not* rejected. Hypothesis testing in both the initial studies and these critiques uses asymptotic distribution theory. The technique used in the earlier studies and those proposed by Ronen (1997) and George and Hwang (2001) are reexamined in this paper to illustrate the benefits of the bootstrap refinement.

3.2 Properties of Pricing Error Tests

The first technique draws inferences by aggregating a cross section of return variance ratios that are estimated separately from each other. We refer to it as the *Standard* approach since it, or something similar, has been used in so many prior studies [Amihud and Mendelson (1987) and (1991), Stoll and Whaley (1990), Gerety and Mulherin (1994), Forster and George (1996) and others]. With the second technique, inferences are drawn from a Wald test, where return variance ratios are estimated jointly for the securities in a sample. This approach was devised by Ronen (1997), and was used in George and Hwang (1995). We refer to it as the *Joint Estimation* approach. Both the Standard and Joint Estimation approaches are based on ratios of the variances of open-to-open and close-to-close returns.

The test statistic of the Standard approach is a t-statistic constructed for the null that the cross sectional average variance ratio in a sample of stocks is equal to one. The construction of this t-statistic implicitly assumes that the variance ratios are independent draws from the same distribution, and attempts to test the null that the mean of this distribution is one. However, since the draws are neither independent nor identically distributed, this statistic is improperly studentized (see the Appendix for details). The Joint Estimation approach is a Wald test for whether the vector of variance ratios of a sample of stocks is equal to the vector of ones. Since this statistic allows for cross sectional dependence and differences in second moments of the distributions from which variance ratios are drawn, it is properly studentized. In fact, this statistic is the properly studentized version of the statistic used in the Standard approach.

The third technique draws inferences based on ratios of pricing error variances that are estimated using time series variance decomposition techniques as in Hasbrouck (1993). The details of this approach are contained in George and Hwang (2001); we refer to it as the *Variance Decomposition* approach. This test is based on the ratio of variances of the transitory components of returns at the open and the close, estimated from a vector auto-regression (VAR). With this approach, hypothesis testing is done for each individual stock. For each stock, the test statistic is a t-statistic that tests whether the variance ratio for that stock equals one.

The relation between the bootstrap ideas described above and these tests can be summarized as follows (see the Appendix for details). The Variance Decomposition approach is based on asymptotically standard normal properly studentized statistics like the one described in section 2 above. The bootstrap version of this test achieves the refinement. The Joint Estimation method is based on a statistic that is asymptotically chi-square. It turns out that the asymptotic and bootstrap approximation errors of such tests behave like those of a two sided symmetric asymptotically standard normal test—i.e., the rejection probability of the chi-square approximation has an error on the order of n^{-1} while the rejection probability of the bootstrap test has an error on the order of n^{-2} [see Horowitz (2001)]. The bootstrap version of this test also achieves the refinement. The Standard approach is based on a statistic that is improperly studentized, so the asymptotic test is invalid and the bootstrap version of this test does not achieve the refinement—i.e., its error is of order n^{-1}. Taken together, this means that the bootstrap inferences of the Variance Decomposition and Joint Estimation approaches are more reliable than those of the Standard approach; and that the asymptotic inferences of the Standard approach are *meaningless.* Nevertheless, we report the results of both versions of all three tests to assess the validity of conclusions reached in prior research.

To quantify how distorted asymptotic tests might be, we generate simulated samples under the restriction that the dynamics of daytime and overnight returns are the same.[4] This restriction is even stronger than the null hypothesis that pricing errors are the same at the open and the close. For each simulated sample, we construct test statistics for the two tests that employ t-ratios as test statistics. The empirical distribution of each test statistic under this strong version of the null is tabulated from the collection of simulated samples. Moments of these distributions are given in Table 1. Two observations are noteworthy. First, there is bias, skewness and excess kurtosis in these distributions; meaning they depart from the standard normal asymptotic distributions they are usually assumed to follow. Second, as can be seen from the Variance Decomposition approach, the severity of these departures is quite different across individual securities. In one instance, the Kolmogorov test is unable to reject the null that the shape of the distribution is standard normal, though the null is rejected in the other five cases. It is, therefore, impossible to know ex-ante whether distortions in asymptotic tests are sufficiently tame for a particular sample to justify their

[4]The simulated samples are generated as follows. First, actual daytime and overnight returns data for each individual stock are fit to a bivariate VAR (equation (10) in the Appendix). The VAR parameter estimates describe the true return dynamics for each stock. In order to generate simulated data in which the daytime and overnight returns have the same dynamics, we (i) replace overnight parameter estimates with daytime parameter estimates and (ii) re-scale the VAR residuals to have equal variances. The simulated samples are then generated from the modified VARs by selecting a random starting point and randomly resampling the rescaled residuals.

use. These results suggest that potentially large distortions are associated with inferences based on asymptotic distribution theory. These distortions are what bootstrap refinements address.

3.3 Bootstrap Algorithm

Table 1 deals with fictitious data created to satisfy the null. This was simply a simulation, not bootstrapping. In order to generate bootstrap distributions of the statistics used in the three tests described above, we resample randomly and with replacement from the original sample. The object that is resampled can be the sample analog of either a random vector that is iid, or a block of serially correlated random vectors. In the Variance Decomposition test, the disturbances to the VAR are assumed to be iid, so the sample analog of the disturbance vectors (i.e., the residual vectors) are resampled in order to bootstrap the distributions of the statistics used by that approach [see MacKinnon (2006)]. For simplicity, and to ease comparison across methods, this same procedure is used to bootstrap the statistics used by the Standard and Joint Estimation approaches. Specifically, the data are assumed to follow a VAR (equation (10) in the Appendix), and bootstrap distributions are generated by resampling the residuals from this VAR. Following George and Hwang (2001), we use a VAR of order two. This specification models serial dependence in returns to a lag of 48 hours, which is more than sufficient to capture the 24 hours of dependence reflected in the covariance terms of the variance ratios used by the Standard and Joint Estimation approaches (see equation (10) in the Appendix). Parameters of the VAR are estimated from the original sample.[5]

To generate a bootstrap resample, a starting date is randomly selected from the original sample. Daytime and overnight returns from that date and the prior date are used in the VAR as initial observations to start the recursion. These returns are the first two observations in the bootstrap sample. A vector of VAR residuals is selected randomly from the set of residuals obtained when the VAR was fit to the original sample. Substituting this residual vector and the starting date returns into the fitted VAR generates the third observation in the bootstrap sample. This observation is then substituted into the fitted VAR along with another randomly chosen residual vector to generate the fourth observation, etc. This process is continued until the bootstrap sample contains the same number of observations as the original sample.

The resampling process is carried out jointly for all securities in a quartile group (quartile groupings are described in section 4.1 below). In particular, if the original date t residual for stock i is the random draw determining the kth observation in the bootstrap sample for stock i, the date

[5] See Hasbrouck (1991a) and George and Hwang (2001) for structural models that justify the VAR specification.

t residual for stock j is used to generate the kth observation in the bootstrap sample for stock j as well. This is done to preserve whatever cross sectional correlation exists among the returns. Preserving this correlation is necessary because its impact on inferences could differ across the three approaches.[6]

Once a bootstrap sample is generated jointly for all 20 stocks in a quartile group, one of the estimation and inference procedures described in section 3.2 above is applied, and the bootstrap version of the corresponding test statistic(s) is computed.[7] This produces the first draw from the bootstrap distribution of the test statistic(s) for that procedure. The entire process is then repeated. A new starting date is selected randomly from the original sample, residuals are randomly drawn and substituted into the fitted VARs to generate a new bootstrap resample jointly for all 20 stocks. The estimation procedure is applied to this resample and the bootstrap version of the test statistic(s) is computed, producing the second draw from the bootstrap distribution of the test statistic(s). This continues until the bootstrap distribution stabilizes.[8] The Standard and Joint Estimation procedures generate a single test statistic for the entire cross section of stocks in a quartile group; so this process will produce one bootstrap distribution per quartile group. The Variance Decomposition procedure generates a test statistic for each stock. Since we include 20 stocks in each quartile group, this procedure produces 20 bootstrap distributions per quartile group.

3.4 Comparison of Rejection Probabilities Between Asymptotic and Bootstrap Tests

The procedure described in the previous subsection produces a set of asymptotic and bootstrap inferences for the variance ratio tests studied here. We can also examine *properties* of these tests because our sample is representative of the true data generating process.

We examine the nature of biases in the asymptotic tests by comparing rejection frequencies of the asymptotic and bootstrap versions of each test. We know from the discussion above that the

[6] Some observations are missing from the original sample. If missing data are chosen as the random starting point, or if a missing residual is randomly chosen, missing observations would propagate through the remainder of the bootstrap resample. To prevent this, we sample starting points and residuals randomly and with replacement from those that are non-missing in the original sample. After generating each bootstrap resample, we overwrite missing values in the same places they appeared in the original sample. Our treatment of missing data attempts to make the impact of the data collection technology on the bootstrap distribution as similar as possible to its impact on the distribution of the test statistic from the original sample.

[7] We work with one procedure at a time. We do not generate a single bootstrap resample and apply all three procedures to it.

[8] The criterion we use for determining whether the distribution has stabilized is to start with 100 draws, then increase the number of draws (for all three methods) in steps of 25 until the inferences (for all three methods) are unchanged through two such increases. The results in the tables that follow reflect 175 draws. Inferences for at least one method changed when stepping from 100 to 125, but inferences for none of the methods changed when stepping from 125 to 150 and again from 150 to 175.

bootstrap version of each test has a smaller size distortion than the asymptotic version. Comparing the frequency with which the two versions of a test reject the null gives an indication of the direction of the distortion in asymptotic inferences. For example, if the asymptotic test rejects more frequently than the bootstrap test, we conclude that the size distortion in the asymptotic test leads to over rejection of the null. Knowing the nature of the distortion helps us interpret the direction of biases in inferences drawn in existing studies that use asymptotic tests.

To estimate the rejection probability of an asymptotic test, we generate a simulated sample from the true return generating process (i.e., resample from the original sample in the manner described in section 3.3 above, and *not* as done to produce Table 1), compute the test statistic, check whether its value is extreme relative to critical values of the asymptotic distribution under the null, and record whether this comparison indicates a rejection of the null. This process is repeated, and the proportion of times the null is rejected is an estimate of the rejection probability of the asymptotic test in samples that are consistent with the true return generating process.[9] In our applications this involves estimating the VAR parameters, then generating simulated samples by sampling from the VAR residuals with replacement.

For inferences conducted using the bootstrap, the procedure is the same except that an entire bootstrap *distribution* of the test statistic is generated for each simulated sample. If the value of the test statistic is extreme compared to critical values from this bootstrap distribution, the null is rejected in this simulated sample. Our hypothesis tests on the original sample employ bootstrap distributions constructed from 175 resamplings. If rejection frequencies are computed by simulating 100 samples that are consistent with the true return generating process, this procedure would involve 175×100 iterations of a resampling of some sort—one bootstrap *distribution* for each simulated sample. This makes computing the rejection frequencies of bootstrap tests generally very resource intensive, and nearly impossible for tests such as the Variance Decomposition approach where estimation and inference are conducted security by security for a moderately sized cross section of securities.

To reduce the problem to a manageable size, we simulate only 50 samples, with 175 bootstrap replications per sample. We then compute rejection frequencies for the two extreme quartile groups—those that contain the most active and least active stocks. The entire quartile group is used for the Standard and Joint Estimation tests so that cross sectional correlation affects their

[9] Horowitz and Savin (2000) point out that the common procedure of simulating samples that are not consistent with the true data generating process (e.g., by constraining the resampling so that the null is necessarily true) is a waste of time because it serves only to compute rejection probabilities in cases that are empirically irrelevant.

rejection frequencies in the same manner as it affects the asymptotic and bootstrap inferences. Since the Variance Decomposition tests are done stock by stock, we reduce the number of stocks per quartile group to ten for just these tests.

4. Data and Results

4.1 Data

All NYSE listed stocks on CRSP are assigned to quartiles by average daily dollar volume for the period 1990 - 1992.[10] The top 50 stocks in each quartile are then ranked alphabetically by ticker symbol. Our sample consists of the first 20 stocks from this alphabetical ordering that have data available on the ISSM files during that period. This results in four quartile groups containing 20 stocks each.

Continuously compounded returns are computed from the transaction prices, cash distributions, stock dividends, and stock splits reported in the ISSM files. For consistency with the earlier studies, the Standard and Joint Estimation approaches are implemented using raw returns. The Variance Decomposition approach uses residuals from a regression of raw returns on day-of-week and turn-of-year dummies. Variance ratio estimates, test statistics, bootstrap distributions of test statistics, and rejection frequencies are computed in the manner described in section 3 above.

Table 2 reports information on sample sizes employed by the various approaches. For the Standard and Variance Decomposition approaches, estimation is done security by security, so the table reports attributes of the cross sectional distribution of the sample sizes for the individual securities. Most of the stocks have samples exceeding 730 observations. For the Joint Estimation approach, estimation is done jointly for all securities. The sample sizes for this test are considerably smaller, consisting of the number of time series observations that are non-missing for *all* securities. Quartile 1 has a sample size of over 730 observations, but the other quartiles have between 167 and 479 observations.

4.2 Bootstrap Refinement Results

In this section, we report the results of drawing inferences using asymptotic distribution theory and the bootstrap refinement for the three methods described above. The two issues we want

[10] We selected this time period because it is close to those used in the earlier studies, which mostly use data from the 1980s. We want to isolate the impact of the choice of inferential technique, and did not want differences between our findings and those reported in earlier studies to be driven by possible regime changes in the data generating process that might have occurred later as a result of liberalized opportunities for after hours trading and decimalization. Indeed, the asymptotic inferences in our sample match those of the studies that use earlier data. In section 4.4 we report results for a more recent sample.

to address are: (i) whether tests using the bootstrap refinement indicate that pricing errors are statistically different at the open and the close, and (ii) whether inferences drawn from bootstrap tests are different from those of asymptotic tests.

Table 3 reports asymptotic and bootstrap critical values for all three approaches. The Standard and Joint Estimation approaches result in one test statistic per quartile group; the Variance Decomposition approach results in one test statistic per stock. The panels report the test statistics, and their asymptotic and bootstrap critical values. The results indicate that: (i) the bootstrap inferences of all three tests lead to the conclusion that for a large majority of stocks, pricing errors are not different at the open and close, and (ii) the bootstrap and asymptotic inferences differ for the Standard approach but not for the Joint Estimation or Variance Decomposition approaches.

The results for the Standard approach indicate that the "asymptotic" and bootstrap inferences are different for stocks in the most and least actively traded quartile groups.[11] For the most active stocks, the "asymptotic" test rejects the null, but the bootstrap test does not. The "asymptotic" rejection is consistent with the findings in other studies that focus primarily on active stocks [e.g., Amihud and Mendelson (1987) for the Dow 30 stocks, Amihud and Mendelson (1991) for the most active stocks on the Tokyo Stock Exchange, and Forster and George (1996) for NYSE stocks with international cross listings]. However, the "asymptotic" test is improperly studentized and is therefore not valid. The fact that the valid bootstrap test does not reject indicates that the inferences drawn in earlier studies that use this technique are probably in error. For the least active stocks, the "asymptotic" test does not reject the null, but the bootstrap test does reject. Even the bootstrap results should be regarded as tentative. The fact that the test statistic is improperly studentized means that the bootstrap Standard test does not achieve the refinement and, therefore, is not as reliable as the bootstrap versions of the Joint Estimation and Variance Decomposition tests.

Inferences using the Joint Estimation approach are the same for asymptotic and bootstrap based tests. All four quartiles fail to reject the hypothesis that the vector of return variance ratios is different from the vector of ones; even the least active quartile group.[12] These inferences are

[11] We write "asymptotic" in quotations because this test is not valid; the test statistic is improperly studentized.

[12] The bootstrap critical values are very wide for the least active group, indicating that the bootstrap approximation to the true distribution of the test statistic has much greater dispersion than that of the chi-square approximation of asymptotic distribution theory. This suggests that the asymptotic test will tend to reject the null too frequently, which is confirmed in the next subsection. However, this tendency depends on the sample. The bootstrap test has narrower critical values than the asymptotic test when applied to data from the 1996 - 1999 period examined below. This illustrates the strength of the bootstrap over asymptotic distribution theory. The bootstrap critical values adapt to differences in the sampling distributions of parameters across different populations.

the same as Ronen's (1997) asymptotic test on the Dow 30 stocks for the period 1982 to 1986. Asymptotic and bootstrap inferences using the Variance Decomposition approach are very similar also. Using the asymptotic test, five stocks reject the null that the pricing error variance ratio is equal to one, and the rejections favor the alternative that the ratio is less than one. This is qualitatively the same inference drawn by George and Hwang (2001) using a sample of 100 stocks during 1986 - 1989. Using the bootstrap test, three stocks reject, and these three happen to be a subset of the five that reject using the asymptotic test.

It is useful to compare asymptotic and bootstrap critical values. The greater is the difference between them, the stronger is the case for adopting a bootstrap approach to inference when using these test statistics. This issue is not germane to the Standard approach. Since its "asymptotic" test is not valid, the researcher does not have a choice but to bootstrap when using this test.

For the Joint Estimation approach, asymptotic and bootstrap critical values differ by orders of magnitude for two of the four quartiles, suggesting that whether one uses asymptotic or bootstrap based tests could dramatically affect the inferences drawn. For our samples, it turns out that bootstrap critical values lie outside those of the asymptotic chi-square approximation. This suggests that asymptotic inferences are likely to reject too often. We will have more to say about this when we estimate rejection probabilities below.

For the Variance Decomposition approach, asymptotic and bootstrap critical values are very different for many of the stocks. As a measure of this, 44 of the 80 bootstrap critical values are greater than three in absolute value, and seven are less than 1.5, compared to the critical values of -/+ 1.96 for the asymptotic test. Though these differences did not affect the conclusion drawn with respect to our sample, they do suggest that whether one uses asymptotic or bootstrap methods could affect inferences in other samples enough that bootstrapping is worth the extra trouble.

To summarize, the bootstrap inferences convey a consistent message—pricing errors are not significantly different at the open than at the close for most stocks. For the Standard approach, "asymptotic" and bootstrap inferences differ. For the Joint Estimation and Variance Decomposition approaches, the asymptotic and bootstrap inferences are almost identical but the critical values are not. In many cases, the asymptotic and bootstrap critical values are different by orders of magnitude, indicating that in applications of these tests to other samples, the asymptotic inferences could be very different from their more precise bootstrap refinement counterparts.

4.3 Rejection Frequencies

Table 4 reports the results of comparing rejection frequencies of asymptotic and bootstrap refinement tests. For a given test, rejection frequencies indicate the nature of the bias in the asymptotic test relative to the more precise bootstrap test. If the rejection frequencies are not similar, it means that the asymptotic test is likely to produce a distorted inference.

With the Standard approach, the "asymptotic" test over rejects for active stocks relative to the more precise bootstrap test. This is the basis for our earlier assertion that "asymptotic" inferences in the existing literature, which reject for samples that are primarily active stocks, are likely to be in error. For inactive stocks, the "asymptotic" test under rejects. This is consistent with our earlier finding that for inactive stocks, the bootstrap test rejects, but the "asymptotic" test does not.

With the Joint Estimation approach, the asymptotic test strongly over rejects, especially for active stocks. Although the asymptotic test is valid, the bootstrap test rejects far less frequently. This suggests that the refinement achieved by bootstrapping is very *likely* to affect the conclusion one draws from samples such as ours, even though the inferences from our particular sample turned out to be the same for asymptotic and bootstrap tests. It is advisable to use a bootstrap test for drawing inferences with this approach in order to avoid this distortion.

For the Variance Decomposition approach, the average (across ten stocks) rejection frequencies for the asymptotic and bootstrap tests are similar, with a slight tendency for the asymptotic test to over reject. A similar picture emerges from the *maximum* differences between asymptotic and bootstrap rejection frequencies. The maximum is more relevant than an average if the test is used on a single stock or to describe cross sectional differences among stocks because it illustrates how far off an inference can be for a given stock. The maximum difference is only 4% for active stocks, but is 40% for inactive stocks. This suggests that if this test were used on active stocks (either individually or a sample), the conclusions drawn would probably not depend on whether asymptotic or bootstrap distributions were used for inference. However, if used on individual *inactive* stocks, bootstrapping to achieve the refinement would be advisable to avoid a potentially large distortion.

Taken together, these results indicate that the "asymptotic" version of the Standard test should never be used. The bootstrap version of this test is not as precise as the bootstrap version of the Joint Estimation test, so if one goes to the trouble of bootstrapping, the Joint Estimation test is preferred over the Standard test. The asymptotic Variance Decomposition test seems to be better behaved than the asymptotic version of the Joint Estimation test, especially for active stocks. So in future applications, the Variance Decomposition test is preferable to the Joint Estimation test, especially if the researcher does not go to the trouble of bootstrapping.

4.4 Bootstrap Results for a More Recent Sample

The tests above use data from 1990-1992 to ensure comparability of our asymptotic inferences with many existing studies and to show that using the bootstrap makes a difference in those inferences. We now examine a more recent sample to address whether the conclusions about market closures and pricing errors drawn from the earlier sample also apply to a more recent time period.

We construct a sample using data from 1996 - 1999 that has the same structure as the sample described above. NYSE stocks on CRSP were ranked by quartile of CRSP dollar volume, then the top 20 stocks alphabetically by ticker symbol were included in each quartile group. Dividends and split information are obtained from CRSP and used to adjust prices in the TAQ database. The adjusted TAQ prices are then used to compute the returns used in the analysis.

Table 5 presents results for the Joint Estimation and Variance Decomposition approaches.[13] The results for the Variance Decomposition approach are very similar to those in Table 3 using the older sample: just a few securities reject the hypothesis that pricing errors are the same at the open and the close in favor of the alternative that pricing errors are *smaller* at the opening. There are four such securities in total—two in quartile 3 and two in quartile 4.

The results for the Joint Estimation approach are different from Table 3 because variance ratios are estimated more precisely in this sample than in the sample analyzed in Table 3. The bootstrap critical values in this table uniformly lie inside the asymptotic critical values, whereas the opposite is true in Table 3. This illustrates a strength of the bootstrap refinement over asymptotic distribution theory. Bootstrap critical values adjust to differences in the sampling distributions of test statistics across populations and asymptotic critical values do not.

Here, the bootstrap tests rejects the null that all variance ratios in the quartile group are equal to one for quartile groups 1, 3 and 4. There are no rejections in Table 3. The rejections for quartiles 3 and 4 reflect the rejections the Variance Decomposition bootstrap test detects for those quartiles, where pricing errors are smaller at the opening. The rejection for quartile 1 indicates that although no single variance ratio departs reliably from one in the Variance Decomposition test, the group jointly does reject the null (similar to a regression F test rejecting when none of the t tests reject).[14] To understand this, note that the *narrowest* bootstrap critical values correspond to *negative* t statistics for quartile 1 of the Variance Decomposition results. The cross-sectional

[13] We exclude the Standard approach at this point because it can be ruled out on a-priori grounds. The test statistic is improperly studentized and its bootstrap version is not as precise as the others.

[14] The Joint Estimation test statistic is based on a precision-weighted average of the individual estimates.

correlation between the t statistics and absolute critical values is 0.45—i.e., narrower critical values are associated with more negative t statistics suggesting that the most precisely estimated ratios in quartile 1 are those that are less than one. Thus, the rejection by the Joint Estimation approach for quartile 1 appears to reflect the impact of pricing errors that are *less* at the open than the close similar to quartiles 3 and 4. The policy implications of these results are the same as the conclusions drawn using the older data—regular market closures do not induce large errors into reopening prices. In fact, reopening auctions might enhance market quality in the more recent sample.

5. Conclusions

The bootstrap has been used for almost three decades as a way to construct a substitute for the finite sample or asymptotic distribution of a test statistic. More recently, Hall and Horowitz (1996) and Horowitz (2001) illustrate how bootstrapping techniques can be used to improve or refine inference when asymptotic distributions are known. This way of using the bootstrap has not been emphasized in Finance research. In this paper we argue that microstructure applications are especially well suited to this application of the bootstrap for three reasons. First, asymptotic distributions of many of the test statistics used in microstructure research are known. Second, structural models estimated in microstructure typically reduce the sample to a sequence of iid disturbances, which makes bootstrap algorithms easy to implement. Finally, microstructure data is very rich in terms of the breadth of samples and the frequency with which variables are observed. Since the bootstrap improves on asymptotics by capturing features of the data that asymptotic distributions ignore, the richness of microstructure data suggests that the improvement associated with bootstrapping is potentially large and well worth the trouble.

Even in fictitious data sets in which the null is necessarily true, test statistics may not conform to their asymptotic distributions. Indeed, in the empirical application we consider here, we show that variance ratio test statistics typically used in microstructure studies do not conform to their asymptotic distributions. Moreover, the severity of the departures depends on the data. We then illustrate how a bootstrap approach can be used to draw more precise inferences, and compare asymptotic and bootstrap inferences of three variance ratio tests of the magnitudes of microstructure induced errors in security prices.

The bootstrap based inferences indicate that pricing errors do not affect opening and closing prices differently. This is contrary to the asymptotic inference drawn in many existing studies. The implication for market design is that, for most stocks, trading suspensions do not result in reopening prices that are especially noisy or especially clear as signals of security value. This suggests, for example, that decisions about whether to adopt 'round-the-clock trading can be made without regard to whether regular trading suspensions lead to reopening prices that reflect information more or less accurately than prices that follow periods of ongoing trading.

For all three tests, bootstrap and asymptotic critical values are not generally close to each other. We also document that in many cases, the size distortions associated with *not* bootstrapping are large. These findings suggest that the more precise bootstrap inferences yield meaningful improvements over those based on asymptotic distribution theory in applications of all three methods. Thus, in applying these tests to other samples, bootstrapping merits serious consideration.

It is useful to contrast the work in this paper with that of Andersen, Bollerslev and Das (2001), who also encounter test statistics that deviate from their asymptotic distributions. Their application is quite different from ours, as they consider variance ratio tests of rates of information flow using high frequency (5-minute) returns data. Nevertheless, like us, they find that the test statistics commonly used do not conform to their asymptotic distributions even when the null is true. Andersen, et.al. address the problem by building more structure into the estimation. Our approach calls for drawing inferences relative to distributions that better approximate the true distribution of the test statistics under the null. Both approaches improve the quality of inferences, and complement each other in applications. Good research exploits the richness of theory by tailoring the specifications of empirical models to incorporate fully theory's predictions. In a similar way, bootstrap refinements exploit the richness of information in the sample by basing inferences on distributions that incorporate important features of the data more fully than is possible using asymptotic distributions.

APPENDIX

Details of the Estimation Techniques

With the exception of George and Hwang (2001), existing studies compare pricing errors at the open and close using a ratio of overlapping *return* variances:

$$\Upsilon = \frac{V^o}{V^c} = \frac{\mathrm{Var}\left[R_t^o\right]}{\mathrm{Var}\left[R_t^c\right]}, \tag{5}$$

where R_t^o is the 24-hour continuously compounded return based on opening prices, and R_t^c is the 24-hour continuously compounded return based on closing prices. The justification for using this ratio is a simple model in which the log cum-dividend prices generating these returns are the sum of two components:

$$\begin{aligned} p_t^o &= m_t^o + s_t^o \\ p_t^c &= m_t^c + s_t^c, \end{aligned} \tag{6}$$

where p_t^o and p_t^c are the log of the observed opening and closing prices on day t, and m_t^o and m_t^c are the log of the security's true value (defined as the security's price in a frictionless market). The pricing errors, s_t^o and s_t^c, exist because markets are not frictionless. Inventory control, the non-information based component of the bid-ask spread, price discreteness, and transient liquidity effects all contribute to pricing errors. Pricing errors are assumed to have unconditional means of zero, so variances measure their average (squared) magnitudes at the open and close $\mathrm{Var}\left[s_t^o\right] = E[(s_t^o)^2]$, and $\mathrm{Var}\left[s_t^c\right] = E[(s_t^c)^2]$. These variances are inverse measures of market quality.

The true value of the security is assumed to evolve as a random walk:

$$\begin{aligned} m_t^o &= m_{t-1}^c + \epsilon_t^{co} \\ m_t^c &= m_t^o + \epsilon_t^{oc}, \end{aligned} \tag{7}$$

where ϵ_t^{co}, and ϵ_t^{oc} are the innovations in the true value overnight and during the trading day, respectively. As information innovations, they have expectation zero, are serially independent, and independent of each other. The variances of these innovations will differ if the rates at which information flows into prices are different during the day and overnight.

Open-to-open and close-to-close returns are given by

$$\begin{aligned} R_t^o &= p_t^o - p_{t-1}^o = \epsilon_{t-1}^{oc} + \epsilon_t^{co} + s_t^o - s_{t-1}^o, \\ R_t^c &= p_t^c - p_{t-1}^c = \epsilon_t^{co} + \epsilon_t^{oc} + s_t^c - s_{t-1}^c. \end{aligned} \tag{8}$$

Thus, 24-hour returns are composed of a 24-hour true value innovation, and a pricing error component. The variance ratio in equation (5) can be written as:

$$\Upsilon = \frac{V^o}{V^c} = \frac{\mathrm{Var}\left[\epsilon_{t-1}^{oc} + \epsilon_t^{co}\right] + \mathrm{Var}\left[s_t^o - s_{t-1}^o\right] + 2\,\mathrm{Cov}\left[\epsilon_{t-1}^{oc} + \epsilon_t^{co}, s_t^o - s_{t-1}^o\right]}{\mathrm{Var}\left[\epsilon_t^{co} + \epsilon_t^{oc}\right] + \mathrm{Var}\left[s_t^c - s_{t-1}^c\right] + 2\,\mathrm{Cov}\left[\epsilon_t^{co} + \epsilon_t^{oc}, s_t^c - s_{t-1}^c\right]}. \tag{9}$$

The first term in the numerator and denominator are equal. Assuming that the covariance terms are equal, empirical evidence that the variance ratio does not equal one implies that the average magnitude of pricing errors at the open and close differ from each other. Variance ratios that exceed one are typically interpreted to imply that price discovery is less precise (more noisy) at the open than at the close.

Most of the studies in this area employ a method of inference that is based on Amihud and Mendelson (1987), which we refer to as the *Standard* approach [see also Amihud and Mendelson (1989), Amihud and Mendelson (1991), Stoll and Whaley (1990) and Forster and George (1996)]. The variance of open-to-open returns, and the variance of close-to-close returns, are estimated for each security individually as the average squared deviation of the return from its sample mean. The variance ratio estimate for stock i is the ratio of these two individual return variance estimates: $\hat{\Upsilon}_i = \frac{\widehat{\mathrm{Var}}[R^o_{it}]}{\widehat{\mathrm{Var}}[R^c_{it}]}$. Inference is based on a t-statistic whose numerator is the cross sectional average of the variance ratio estimates, and whose denominator is the standard deviation of the cross section of variance ratio estimates. This t-ratio is compared to critical values from a standard normal distribution. Formally, the null hypothesis that $\frac{1}{n}\sum_{i=1}^{n}\Upsilon_i = 1$ is tested with a cross sectional t-statistic specified as:

$$n^{1/2}\left(\bar{\Upsilon}-1\right)/\hat{\sigma}(\hat{\Upsilon}) \sim^{\text{asy}} N(0,1),$$

where $\bar{\Upsilon} = \frac{1}{n}\sum_{i=1}^{n}\hat{\Upsilon}_i$, and $\hat{\sigma}(\hat{\Upsilon})$ is the sample (cross sectional) standard deviation of the $\hat{\Upsilon}_i$s. These studies document that $\bar{\Upsilon}$ is approximately 1.2, and reject the null.

Ronen (1997) questions this approach. She points out that drawing inferences from a cross sectional distribution of variance ratios implicitly assumes that the estimates are independent draws from an identical distribution. This is unlikely because variance ratios are estimated using returns from the same time period for all stocks in the sample—the draws are not independent. Moreover, there is no reason to believe that the sampling distribution of $\hat{\Upsilon}_i$ is the same for all stocks in the sample.[15] Consequently, $\hat{\sigma}(\hat{\Upsilon})^2$ is not a consistent estimator of the asymptotic variance of $n^{1/2}\bar{\Upsilon}$. From the perspective of the bootstrap discussed in section 2, this means that the test statistic of the Standard approach is improperly studentized.

To overcome the deficiencies of the asymptotic Standard test, Ronen advocates an approach in which all variance ratios are jointly estimated using Hansen's (1982) generalized method of

[15] Without these strong assumptions

$$\mathrm{Var}\left[n^{1/2}\bar{\Upsilon}\right]=\frac{1}{n}\sum_{i=1}^{n}\mathrm{Var}\left[\hat{\Upsilon}_i\right]+\frac{1}{n}\sum_{i}\sum_{j\neq i}\mathrm{Cov}\left[\hat{\Upsilon}_i,\hat{\Upsilon}_j\right].$$

moments (GMM), and inference conducted by way of a joint test that the vector of variance ratios equals the vector of ones. This procedure accounts for cross sectional correlations, and differences in the sampling distributions of the individual variance ratio estimates. We refer to this as the *Joint Estimation* approach.

Ronen's variance ratios are estimated by setting to zero the following (exactly identified) system of moment restrictions jointly for all n stocks in the sample:

$$g(\Upsilon_1,\ldots,\Upsilon_n,V_1^c,\ldots,V_n^c)=\frac{1}{T}\sum_{t=1}^{T}\begin{pmatrix}(R_{1,t}^o)^2-\Upsilon_1 V_1^c\\ \vdots\\ (R_{n,t}^o)^2-\Upsilon_n V_n^c\\ (R_{1,t}^c)^2-V_1^c\\ \vdots\\ (R_{n,t}^c)^2-V_n^c\end{pmatrix},$$

where $R_{i,t}^o$ and $R_{i,t}^c$ are open-to-open and close-to-close returns for stock i, V_i^c is the variance of the 24-hour return based on closing prices for stock i, and $\Upsilon_i=\frac{V_i^o}{V_i^c}$ is the ratio of the opening variance to the closing variance for stock i. Ronen tests the null hypothesis, $H_0:[\Upsilon_1,\cdots,\Upsilon_n]'=[1,\cdots,1]'$ with a Wald statistic specified as:

$$T\times\underbrace{(\hat{\Upsilon}-\iota)'}_{1\times n}\underbrace{\hat{\Sigma}_{\Upsilon}^{-1}}_{n\times n}\underbrace{(\hat{\Upsilon}-\iota)}_{n\times 1}\sim^{\text{asy}}\chi^2(n),$$

where ι is an $n\times 1$ vector of ones, and $\hat{\Sigma}_{\Upsilon}$ is a consistent estimator of the asymptotic covariance matrix of $T^{1/2}\hat{\Upsilon}$. Using this procedure, Ronen does not reject the null for a sample of NYSE stocks that is almost identical to that of Amihud and Mendelson's (1987) sample. She attributes the discrepancy to the fact that inferences in the earlier studies are invalid.[16]

Finally, George and Hwang (2001) offer a procedure that isolates the variances of pricing errors from their covariances with information innovations to estimate the ratio $\Gamma\equiv\frac{\text{Var}[s_t^o]}{\text{Var}[s_t^c]}$ for each individual security.[17] They estimate parameters for each security using GMM, which provides an asymptotic sampling error structure of the parameters for each stock. Inferences are then conducted at the individual security level.

---

[16]Ronen's (1997) results are not likely to be a consequence of a low power test. George and Hwang (1995) estimate variance ratios for Tokyo stocks using the Joint Estimation approach, and reject the hypothesis that the ratios are equal to the vector of ones for three of the four quartiles of trading activity examined. Their most active quartile rejects in favor of the alternative that the variance ratios are greater than one, while the two quartiles of least active trading reject in favor of the alternative that the variance ratios are less than one. They attribute this difference to the impact of price limit rules on the Tokyo Exchange.

[17]Their approach actually provides joint estimates of the variances of the temporary *and* permanent components of returns (i.e., s_t^o,s_t^c,ϵ_t^{oc}, and ϵ_t^{co} in equations (6) and (7)) for each security.

George and Hwang (2001) model the daytime and overnight returns to an individual stock, $\mathbf{x}_t \equiv (r_{dt}, r_{nt})'$, as a vector auto-regression of order two:

$$A\mathbf{x}_t = B_1\mathbf{x}_{t-1} + B_2\mathbf{x}_{t-2} + \mathbf{u}_t \qquad \mathbf{u}_t \sim iid(0\,,\,\Omega). \tag{10}$$

(Though we have not included an i subscript, the parameters of the VAR are indeed different for each stock.) Theorem 1 in their paper gives closed form expressions for the variances of $\epsilon_t^{co}, \epsilon_t^{oc}, s_t^o$ and s_t^c in terms of the auto-regressive parameters A, B_1, B_2 and Ω. These expressions provide the functional form $\gamma(\cdot)$, of the relationship between the ratio of pricing error variances and the auto-regressive parameters $\Gamma \equiv \frac{\mathrm{Var}[s_t^o]}{\mathrm{Var}[s_t^c]} = \gamma(A, B_1, B_2, \Omega)$. Their variance ratio estimate for an individual stock is $\hat{\Gamma} = \gamma(\hat{A}, \hat{B}_1, \hat{B}_2, \hat{\Omega})$, where the parameters $\hat{A}, \ldots, \hat{\Omega}$ are estimated separately for each stock using GMM.[18] Hypothesis tests are based on asymptotic normality of functions of the GMM estimators. George and Hwang's test statistic for $H_o : \Gamma = 1$ is

$$T^{1/2}(\hat{\Gamma} - 1)/\hat{\sigma}(\hat{\Gamma}) \sim^{\text{asy}} N(0, 1)$$

where $\hat{\sigma}(\hat{\Gamma})^2$ is a consistent estimator of the asymptotic variance of $T^{1/2}\hat{\Gamma}$. We refer to this as the *Variance Decomposition* approach.

George and Hwang (2001) study four quartile groups, each containing 25 stocks, for the period 1986 - 1989. They find that pricing errors at the open and close are not significantly different for most stocks. However, when differences do exist, pricing errors at the close tend to be *greater* than at the open.

[18] Mean effects associated with day of the week and turn of the year are accounted for by estimating regressions of each return variable on an intercept, indicators for each of Tuesday through Friday, and an indicator for the first fourteen days of January. The residuals from these regressions (estimated separately for each security) are then used to estimate the vector auto-regressions.

Table 1
Simulated Distributions of Test Statistics Under the Null

This table reports moments of distributions of test statistics for Traditional and Variance Decomposition approach, and significance levels of a Kolmogorov-Smirnov test of whether the distribution is standard normal. Test statistics are computed for samples simulated under the constraint that the dynamics of overnight returns matches that of daytime returns. Each distribution below is tabulated from 175 simulated samples. From a given simulated sample, the Traditional Approach computes a single test statistic from a cross-section of 20 stocks, whereas the Variance Decompositon Approach computes a test statistic for each stock. For the Variance Decomposition Approach only five stocks were included chosen by alphabetical order of their ticker symbol from quartile 1 (most actives) of our sample.

Traditional Approach

	Simulated t-ratios
Median	0.287
Mean	0.196
Standard Deviation	0.813
Skewness	-0.293
Excess Kurtosis	0.184
Kolmogorov-Smirnov significance level	< 1%

Variance Decomposition Approach

	Simulated t-ratios				
	Stock 1	Stock 2	Stock 3	Stock 4	Stock 5
Median	-0.184	-0.099	-0.625	-0.277	-0.217
Mean	-0.440	-0.243	-0.699	-0.414	-0.560
Standard Deviation	1.270	1.048	1.175	0.949	1.475
Skewness	-1.599	-0.517	-0.189	-0.611	-1.911
Excess Kurtosis	6.711	-0.012	-0.322	0.252	5.959
Kolmogorov-Smirnov significance level	< 1%	> 10%	< 1%	< 1%	< 1%

Table 2
Sample Sizes

Entries in the top panel report the cross-sectional distribution of the number of non-missing daily time-series observations from the ISSM data during the sample period 1990 - 1992. Each quartile group contains the top 20 NYSE stocks alphabetically by ticker symbol from quartile groups of average dollar trading volume reported on CRSP. Each quartile group contains the top 50 stocks by trading volume in that quartile. Entries in the lower panel are the number of days for which all 20 stocks have non-missing ISSM data.

Standard & Variance Decomposition Approaches

Attributes of the cross-sectional distribution of sample sizes for each security	Quartile 1 (most active)	Quartile 2	Quartile 3	Quartile 4 (least active)
Minimum	739	168	499	498
25th Percentile	749	739	739	739
Median	752	749	744	752
75th Percentile	752	749	752	752
Maximum	752	752	752	752

Joint Estimation Approach

	Quartile 1 (most active)	Quartile 2	Quartile 3	Quartile 4 (least active)
Sample size	732	167	479	238

Table 3
Bootstrap Results
Variance Ratios

Variance ratios and the associated test statistics are computed as described in section 3 of the text using data from 1990-1992. Each quartile group contains 20 NYSE stocks selected from quartiles of average dollar volume. With each test statistic is reported its bootstrap and asymptotic critical values. For the Standard Approach, the null hypothesis is that the cross-sectional average ratio of open-to-open and close-to-close return variances is equal to one. For the Joint Estimation approach, the null hypothesis is that the vector of such variance ratios is (jointly) equal to one. For the Variance Decomposition approach, the null hypothesis is that the ratio of variances of pricing errors at the open and close is equal to one. Bootstrap critical values are based on 175 replications.

Standard Approach

	Quartile 1 (most active)	Quartile 2	Quartile 3	Quartile 4 (least active)
Average Variance Ratio – 1	0.245	0.238	0.044	0.387
Median Variance Ratio – 1	0.106	0.034	0.022	0.043
Cross-sectional t-ratio	2.344	1.640	1.301	1.226
5% symmetric bootstrap critical values	–/+ 2.610	–/+ 8.222	–/+ 1.520	–/+ 0.833
5% symmetric standard normal critical values	–/+ 1.96	–/+ 1.96	–/+ 1.96	–/+ 1.96

Joint Estimation Approach

	Quartile 1 (most active)	Quartile 2	Quartile 3	Quartile 4 (least active)
GMM Wald Statistic	30.932	12.016	14.370	11.589
5% bootstrap critical value	675.73	30.389	88.644	5920.9
5% Chi-Square (20 d.f.) critical value	31.410	31.410	31.410	31.410

Table 3 (cont.)
Bootstrap Results
Variance Ratios

Variance Decomposition Approach								
	Quartile 1		Quartile 2		Quartile 3		Quartile 4	
	GMM t-statistic	5% symmetric bootstrap critical values	GMM t-statistic	5% symmetric bootstrap critical values	GMM t-statistic	5% symmetric bootstrap critical values	GMM t-statistic	5% symmetric bootstrap critical values
Stock 1	1.329	–/+ 5.559	0.319	–/+ 12.57	0.534	–/+ 5.708	0.620	–/+ 1.272
Stock 2	0.915	–/+ 75.23	-1.056	–/+ 1.661	-1.625	–/+ 1.960	-0.124	–/+ 1.562
Stock 3	0.503	–/+ 9.966	0.978	–/+ 6.255	0.295	–/+ 103.5	0.692	–/+ 7.826
Stock 4	-2.034	–/+ 3.167	-1.518	–/+ 3.400	0.448	–/+ 4.640	0.692	–/+ 6.579
Stock 5	0.346	–/+ 30.77	0.002	–/+ 1.242	0.448	–/+ 1.620	-0.528	–/+ 1.373
Stock 6	1.179	–/+ 62.28	-2.578	–/+ 4.273	0.621	–/+ 2.741	0.435	–/+ 85.80
Stock 7	1.011	–/+ 2.871	-0.610	–/+ 1.690	0.479	–/+ 2.090	0.035	–/+ 2.886
Stock 8	0.574	–/+ 102.8	0.859	–/+ 1.316	0.250	–/+ 7.194	1.132	–/+ 10.14
Stock 9	-0.238	–/+ 4.149	0.169	–/+ 1.256	-0.150	–/+ 1.377	-0.976	–/+ 1.674
Stock 10	0.914	–/+ 14.23	-0.542	–/+ 3.165	0.247	–/+ 6.822	0.219	–/+ 1.606
Stock 11	-1.592	–/+ 2.546	-3.155	–/+ 1.767	1.055	–/+ 4.677	0.500	–/+ 2.349
Stock 12	0.538	–/+ 3.266	-0.052	–/+ 1.339	0.148	–/+ 59.78	0.234	–/+ 3.010
Stock 13	0.626	–/+ 2.517	0.369	–/+ 10.41	-3.416	–/+ 2.051	0.154	–/+ 5.919
Stock 14	0.687	–/+ 3.105	-0.270	–/+ 1.708	-0.804	–/+ 1.744	0.503	–/+ 4.468
Stock 15	-0.743	–/+ 2.890	0.212	–/+ 3.305	0.535	–/+ 10.80	0.586	–/+ 1.523
Stock 16	0.473	–/+ 4.015	0.679	–/+ 57.47	0.476	–/+ 3.620	0.148	–/+ 9.800
Stock 17	-0.264	–/+ 2.852	0.438	–/+ 2778	-0.378	–/+ 1.361	-0.254	–/+ 1.594
Stock 18	-0.292	–/+ 1.779	-1.067	–/+ 1.507	0.618	–/+ 3.037	0.302	–/+ 25.11
Stock 19	0.250	–/+ 1.975	-1.957	–/+ 1.535	0.690	–/+ 14.60	-0.593	–/+ 2.105
Stock 20	1.040	–/+ 2.191	0.810	–/+ 3.354	0.761	–/+ 19.02	0.429	–/+ 20.73

Note: The 5% two-sided critical values for the asymptotic standard normal approximation are –/+ 1.96.

Table 4
Rejection Probabilities

Numbers in the table report the proportion of times the null hypothesis is rejected in 50 samples drawn from the original sample as described in section 3 of the text. For each of the 50 samples, bootstrap tests are based on critical values from 175 replications. For the Standard and Joint Estimation Approaches, each quartile group contains 20 stocks. For the Variance Decomposition Approaches, each quartile group contains 10 stocks.

Standard Approach		
Frequency of Rejections	Quartile 1 (most active)	Quartile 4 (least active)
Bootstrap Test	0.42	0.26
"Asymptotic" Test	0.64	0.04

Joint Estimation Approach		
Frequency of Rejections	Quartile 1 (most active)	Quartile 4 (least active)
Bootstrap Test	0.10	0.06
Asymptotic Test	1.00	0.12

Variance Decomposition Approach		
Frequency of Rejections	Quartile 1 (most active)	Quartile 4 (least active)
Average using Bootstrap Test	0.06	0.14
Average using Asymptotic Test	0.08	0.16
Maximum Difference in Rejection Frequency (asy – b/s)	0.04	0.40

Table 5
Bootstrap Results
Sample Period 1996-1999

Variance ratios and the associated test statistics are computed as described in section 3 of the text using data from 1996-1999. Each quartile group contains 20 NYSE stocks selected from quartiles of average dollar volume. With each test statistic is reported its bootstrap and asymptotic critical values. For the Joint Estimation approach, the null hypothesis is that the vector of such variance ratios is (jointly) equal to one. For the Variance Decomposition approach, the null hypothesis is that the ratio of variances of pricing errors at the open and close is equal to one. Bootstrap critical values are based on 175 replications.

Joint Estimation Approach				
	Quartile 1 (most active)	Quartile 2	Quartile 3	Quartile 4 (least active)
GMM Wald Statistic	22.072	9.363	15.183	19.860
5% bootstrap critical value	13.420	12.970	14.833	17.200
5% Chi-Square (20 d.f.) critical value	31.410	31.410	31.410	31.410

Table 5 (cont.)
Bootstrap Results
Sample Period 1996-1999

Variance Decomposition Approach								
	Quartile 1		Quartile 2		Quartile 3		Quartile 4	
	GMM t-statistic	5% symmetric bootstrap critical values	GMM t-statistic	5% symmetric bootstrap critical values	GMM t-statistic	5% symmetric bootstrap critical values	GMM t-statistic	5% symmetric bootstrap critical values
Stock 1	0.796	–/+ 10.04	0.190	–/+ 36.71	0.860	–/+ 49.45	-1.310	–/+ 1.211
Stock 2	-0.016	–/+ 1.710	0.331	–/+ 22.84	0.289	–/+ 3.827	0.266	–/+ 3.415
Stock 3	-0.431	–/+ 1.600	-0.549	–/+ 1.139	-1.514	–/+ 1.120	0.776	–/+ 3.008
Stock 4	0.562	–/+ 26.52	-0.102	–/+ 1.167	-0.031	–/+ 5.123	0.619	–/+ 2.300
Stock 5	0.569	–/+ 1.520	1.145	–/+ 60.73	-1.660	–/+ 3.146	0.248	–/+ 3.845
Stock 6	0.443	–/+ 11.09	0.168	–/+ 2.374	-0.884	–/+ 4.803	-1.144	–/+ 1.370
Stock 7	-0.296	–/+ 1.382	0.234	–/+ 1.726	0.446	–/+ 31.10	0.850	–/+ 2.041
Stock 8	-0.362	–/+ 1.700	-0.385	–/+ 1.268	-8.183	–/+ 1.780	0.335	–/+ 3.858
Stock 9	0.256	–/+ 6.918	0.320	–/+ 11.64	0.678	–/+ 28.17	0.114	–/+ 2.789
Stock 10	0.947	–/+ 3.690	0.513	–/+ 35.52	-0.825	–/+ 2.264	-6.803	–/+ 1.569
Stock 11	0.357	–/+ 1.639	0.828	–/+ 10.63	0.507	–/+ 11.00	1.039	–/+ 1.904
Stock 12	-1.246	–/+ 1.892	-0.079	–/+ 1.295	0.552	–/+ 4.550	1.105	–/+ 2.546
Stock 13	0.269	–/+ 5.437	0.684	–/+ 3.170	0.187	–/+ 43.30	-0.529	–/+ 1.750
Stock 14	-1.330	–/+ 2.392	1.003	–/+ 17.38	0.418	–/+ 1.195	1.042	–/+ 1.987
Stock 15	0.685	–/+ 22.29	-0.941	–/+ 2.753	-0.274	–/+ 1.780	0.398	–/+ 17.80
Stock 16	0.770	–/+ 2.997	0.662	–/+ 8.850	0.475	–/+ 9.772	0.931	–/+ 2.605
Stock 17	-1.437	–/+ 2.091	0.334	–/+ 11.70	0.140	–/+ 1.564	0.139	–/+ 1.558
Stock 18	0.058	–/+ 4.613	0.583	–/+ 8.914	0.321	–/+ 35.45	-0.470	–/+ 3.160
Stock 19	-0.992	–/+ 2.482	0.358	–/+ 10.53	-0.318	–/+ 1.280	-0.244	–/+ 1.570
Stock 20	0.457	–/+ 7.879	-1.562	–/+ 2.029	0.318	–/+ 8.620	1.158	–/+ 2.400

Note: The 5% two-sided critical values for the asymptotic standard normal approximation are –/+ 1.96.

REFERENCES

Amihud, Y., and Mendelson, H., 1987, Trading Mechanisms and Stock Returns: An Empirical Investigation, *Journal of Finance*, 42, 533-553.

Amihud, Y., and H. Mendelson, 1991, "Volatility and Trading: Evidence from the Japanese Stock Market," *Journal of Finance,* 46, 1765-1789.

Andersen, T., T. Bollerslev, and A. Das, 2001, "Variance-Ratio Statistics and High-Frequency Data: Testing for Changes in Intraday Volatility Patterns," *Journal of Finance,* 56, 305-327.

Berkowitz, J. and L. Kilian, 2000, "Recent Developments in Bootstrapping Time Series," *Econometric Reviews,* 19, 1-48.

Efron, B., 1979, "Bootstrap Methods: Another Look at the Jackknife," *Annals of Statistics,* 7, 1-26.

Forster, M. and T. George, 1996, "Pricing Errors at the NYSE Open and Close: Evidence from Internationally Cross-Listed Stocks," *Journal of Financial Intermediation,* 5, 95-126.

George, T., and C. Hwang, 1995, "Transitory Price Changes and Price Limit Rules: Evidence from the Tokyo Stock Exchange," *Journal of Financial and Quantitative Analysis,* 30, 313-327.

George, T. and C. Hwang, 2001, "Information Flow and Pricing Errors: A Unified Approach to Estimation and Testing," *Review of Financial Studies,* 14,979-1020.

Gerety, M. and H. Mulherin, 1994, "Price Formation on Stock Exchanges: The Evolution of Trading Within the Day," *Review of Financial Studies,* 7, 609-629.

Goetzmann, W., 1993, "Patterns in Three Centuries of Stock Market Prices," *Journal of Business,* 66, 249-270.

Goetzmann, W. and P. Jorion, 1992, "Testing the Predictive Power of Dividend Yields," *Journal of Finance,* 48, 663-679.

Hall, P., 1992, *The Bootstrap and Edgeworth Expansion,* Springer.

Hall, P., and J. Horowitz, 1996, "Bootstrap Critical Values for Tests Based on Generalized Method of Moments Estimators," *Econometrica,* 64, 891-916.

Hardle, W., J. Horowitz, and J. Kreiss, 2003, "Bootstrap Methods for Time Series," *International Statistics Review,* 71, 435-459.

Hasbrouck, J., 1991a, "Measuring the Information Content of Stock Trades," *Journal of Finance,* 46, 179-207.

Hasbrouck, J., 1991b, "The Summary Informativeness of Stock Trades: An Econometric Analysis," *Review of Financial Studies,* 4, 571-595.

Hasbrouck, J., 1993, "Assessing the Quality of a Security Market: A New Approach to Transaction-Cost Measurement," *The Review of Financial Studies,* 6, 191-212.

Hasbrouck, J., 1995, "One Security, Many Markets: Determining the Contributions to Price Discovery," *Journal of Finance,* 50, 1175-1199.

Horowitz, J., 2001, "The Bootstrap," *Handbook of Econometrics,* North Holland, Vol. 5.

Horowitz, J. and N. E. Savin, 2000, "Empirically Relevant Critical Values for Hypothesis Tests: A Bootstrap Approach," *Journal of Econometrics,* 95, 375-389.

Huang, R., 2002, "The Quality of ECN and Nasdaq Market Maker Quotes," *Journal of Finance,* 57, 1285-1319.

Jones, C., G. Kaul, and M. Lipson, 1994, "Information, Trading and Volatility," *Journal of Financial Economics,* 36, 127-154.

Kosowski, R., A. Timmerman, H. White and R. Wermers, 2006, "Can Mutual Fund "Stars" Really Pick Stocks? New Evidence from a Bootstrap Analysis," *Journal of Finance,* 2551-2595.

MacKinnon, J., 2001, "Bootstrap Methods in Econometrics," *Economic Record,* 82, S2-S18.

Maddala, G.S. and H. Li, 1996, "Bootstrap Based Tests in Financial Models," In G.S. Maddala and C.R. Rao (eds.) *Handbook of Statistics,* vol. 14, Elsevier, Amsterdam, 463-488.

Ronen, T., 1997, "Tests and Properties of Variance Ratios in Microstructure Studies," *Journal of Financial and Quantitative Analysis,* 32, 183-204.

Ruiz, E. and L. Pascual, 2002, "Bootstrapping Financial Time Series," *Journal of Economic Surveys,* 16, 271-300.

Stoll, H. and Whaley, R., 1990, "Stock Market Structure and Volatility," *Review of Financial Studies,* 3, 37-71.

Sullivan, R., A. Timmermann, and H. White, 1999, "Data-Snooping, Technical Trading Rule Performance, and the Bootstrap," *Journal of Finance,* 54, 1647-1691.

Zvingales, J., 2001, "On Bootstrap Coverage Probability with Dependent Data," University of Iowa working paper.