TECHNICAL WORKING PAPER SERIES

THE BIAS OF THE RSR ESTIMATOR AND THE ACCURACY OF SOME ALTERNATIVES

William N. Goetzmann
Liang Peng

Technical Working Paper 270
http://www.nber.org/papers/T0270

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2001

This paper benefited significantly from the comments and suggestions of John Clapp. We also thank Donald Andrews, Stefano Athnasoulis, Bradford Case, David Geltner, Roger Ibbotson, Peter C B Phillips, Robert Shiller, Matthew Spiegel, Lu Zheng, and participants at the 2001 American Real Estate and Urban Economics meeting for helpful discussions and comments. We thank Karl Case and Robert Shiller for providing us historical data of house prices. All errors are ours alone. The views expressed in this paper are those of the authors and not necessarily those of the National Bureau of Economic Research.

The Bias of the RSR Estimator and the Accuracy of Some Alternatives
William N. Goetzmann and Liang Peng
NBER Technical Working Paper No. 270
April 2001
JEL No. R2

ABSTRACT

This paper analyzes the implications of cross-sectional heteroskedasticity in repeat sales regression (RSR). RSR estimators are essentially geometric averages of individual asset returns because of the logarithmic transformation of price relatives. We show that the cross sectional variance of asset returns affects the magnitude of bias in the average return estimate for that period, while reducing the bias for the surrounding periods. It is not easy to use an approximation method to correct the bias problem. We suggest a maximum-likelihood alternative to the RSR that directly estimates index returns that are analogous to the RSR estimators but are arithmetic averages of individual returns. Simulations show that these estimators are robust to time-varying cross-sectional variance and may be more accurate than RSR and some alternative methods of RSR.

William N. Goetzmann
Yale School of Management
Box 208200
New Haven, CT 06520-8200
Email: will@viking.som.yale.edu

Liang Peng
Yale Economics Department
Box 208268
New Haven, CT 06520-8268
Email: liang.peng@yale.edu

The Bias of the RSR Estimator and the Accuracy of Some Alternatives

I. Introduction

The repeat sales regression (RSR) and its variants are widely used to infer returns of equal-weighted portfolios of assets through time.[1] Most applications of RSR have been in the area of home price index estimation. Indeed, local home indices constructed with the RSR are becoming the benchmarks for home appraisal -- the RSR allows a rapid-web-based home price estimate that can be used for quick home mortgage assessment and approval. Although it is now becoming a pervasive tool for credit analysis, the RSR has some well-known econometric flaws.[2] One well known problem of the RSR estimators is that they are biased downwards from actual portfolio returns, which obviously is not desirable because the most common use of any index may be to estimate the current value of its underlying portfolio. While equal-weighted portfolios of assets have returns that are arithmetic averages of cross-sectional individual asset returns, the repeat sales estimators are essentially cross-sectional geometric averages. Because of Jensen's inequality, the logarithmic transformation of the price relatives used as a dependent variable in the repeat-sales regression results in a bias -- the RSR averages logs rather than takes a log of an average. Thus after getting rid of the log, the RSR estimators are geometric averages instead of arithmetic averages.

Three methods have been suggested to address the bias problem. Shiller (1991) proposes arithmetic-average price estimators for equal-weighted and value-weighted portfolios. Goetzmann (1992) proposes a method that approximates the arithmetic means given RSR estimators, under the assumption that asset returns in each period are lognormally distributed and the cross-sectional variance is constant over time. In another attempt toward unbiased estimators, Goetzmann and

Geltner propose a non-linear method that minimizes the sum of squared residuals directly without taking logs first.[3]

Though the bias problem of RSR is well known, its source and magnitude may not be well understood by many researchers and practitioners. In this paper, we interpret RSR estimators as *sample statistics*, and show how they are simultaneously determined in the regression and how they actually mimic cross-sectional geometric sample means. Specifically, we interpret each RSR estimator as a geometric average of *proxies* of individual single-period asset returns. As a result, we are able to explicitly decompose the bias of RSR estimators into two components and study them separately.

Our analysis shows that the two components of the bias are respectively determined by two different impacts of the logarithmic transformation of the price relatives: the direct impact and the serial impact. These two impacts push RSR coefficients toward opposite directions. Specifically, the direct impact makes RSR coefficients biased downwards, while the serial impact makes them biased upwards. The actual bias of a repeat sales estimator for one specific time period is jointly determined by the sum of these two impacts in that period.

We show that the magnitude of the actual bias of RSR estimators may not be uniform from period to period. In each time period, the magnitude of both direct impact and serial impact of logarithmic transformation is generally different. The magnitude of the direct impact is related to the cross-sectional sample variance of individual asset returns in that period, while the magnitude of the serial impact is related to the sample variances in surrounding periods. Therefore, the magnitude of the actual bias is generally different through time, since the sample variances of individual asset returns are usually different through time. Consequently, magnitude of the bias in the RSR estimator is predictable to some extent. For example, for time periods with larger cross-sectional variance of individual asset returns, the RSR estimator tends to be more downwards biased. At the same time, the performance of the approximation method proposed by Goetzmann (1992) is also predictable. This method would compensate for the bias, insufficiently for time periods with larger variances while more than enough for time periods with smaller variances. We use simulations to show such patterns for RSR and the approximation method.

We propose a new approach to mitigate the bias problem of the RSR estimators. The new arithmetic repeat sales estimators proposed here are unbiased, and have a natural interpretation as equal-weighted averages of individual single-period asset returns. With simulations, we examine the performance of this new method together with other alternative RSR approaches. The simulation results suggest that the arithmetic repeat sales estimators we propose may be more accurate than RSR and other alternatives.

The paper is organized as follows. Section 1 interprets RSR estimators as *sample statistics*, and shows that they are essentially geometric averages of individual returns or their proxies. Section 2 decomposes the bias into two components and investigates the determination of each. It shows that the magnitude of the bias of RSR estimators is not uniform from period to period. It also predicts patterns of RSR bias and performance of the approximation method. Section 3 proposes unbiased repeat sales estimators that are analogous with RSR estimators but are arithmetic averages. It also provides comparison between the unbiased estimators with the arithmetic-mean repeat-sale estimators by Shiller (1991), and shows the feasibility of the calculation of the unbiased estimators as well. Section 4 uses simulations to test our predictions of the behavior of RSR estimators and the performance of the unbiased estimators we propose and other alternatives. Section 5 concludes.

II. RSR Estimators as Geometric Means

II.1 RSR estimators

The repeat sales regression estimates the return of an equal-weighted portfolio of assets over each period in time. Assume in total there are N observations of repeat sales of individual assets numbered from $i=1$ to $i=N$. Each observation, say observation i, consists of the time of first sale b_i and the price B_i, the time of second sale s_i and the price S_i. Denote by H_i the holding interval of observation i, which consists of all time periods later than b_i and no later than s_i. Thus the length of holding interval for observation i, denoted by $\boldsymbol{t}_i$, equals $s_i - b_i$. We suppose that there are $T+1$ periods numbered from $t=0$ to $t=T$.

For each observation i, we define the compound return and the log compound return as

$$g_i \equiv \frac{S_i}{B_i}, \text{ and } y_i \equiv \log g_i = \log S_i - \log B_i .$$

Denote by $r_{i,t}$ the log gross return of the asset corresponding to observation i in period t.

$$r_{i,t} \equiv \log\left(\frac{P_{i,t}}{P_{i,t-1}}\right)$$

Thus

$$y_i = \sum_{t\in H_i} r_{i,t} . \tag{1}$$

Denote by $P_{m,t}$ the value of the portfolio (market) at the end of time period t. We define $\boldsymbol{b}_t$ as the gross return of index portfolio for time period t, and m_t as $\log(\boldsymbol{b}_t)$.

$$\boldsymbol{b}_t \equiv \frac{P_{m,t}}{P_{m,t-1}} \text{ and } m_t \equiv \log \boldsymbol{b}_t = \log P_{m,t} - \log P_{m,t-1} .$$

The RSR assumes that

$$r_{i,t} = m_t + e_{i,t}, \tag{2}$$

where the error term is assumed i.i.d. normally distributed.

From equation (1) and (2),

$$y_i = \sum_{t \in H_i} m_t + \sum_{t \in H_i} e_{i,t} . \tag{3}$$

Equation (3) provides conditions for identifying maximum likelihood estimators $\{\hat{m}_t\}_{t=1}^{T}$.

The RSR estimators are calculated according to

$$\hat{m} = \left(X'\Omega^{-1}X\right)^{-1} X'\Omega^{-1}Y . \tag{4}$$

where X, Y, and Ω are defined as follows. The X is a N by T dummy matrix whose ith row corresponds to the ith observation and tth column corresponds to time period t. In the ith row, the first nonzero dummy appears in the column that corresponds to the time period immediately after the buy period, and the last nonzero dummy appears in its sale period. Between are nonzero dummies. Other elements in this row are zero. For instance, if a asset was purchased at time 2 and sold at time 4, and $T = 5$, the its corresponding row is (0,0,1,1,0). The Ω is a N by N diagonal matrix with ith diagonal element is t_i, i.e. the length of the holding interval. The Y is a N by 1 matrix with whose ith element is y_i.

The biases of RSR estimators resulting from the logarithmic transformation of the price relatives are discussed in Goetzmann (1992) who uses a one-period example to show how the logarithmic transformation makes RSR estimators biased downwards. I.e. the RSR estimator is expressed as:

$$\hat{m} = (1'1)^{-1}1'Y$$

or the simple average of the elements in the logged price relative vector Y. Because the log function is concave, Jensen's inequality implies that the average of the logs is less than the log of the average, when there is any variance in the data. Thus, if the elements in the Y vector differ at all, $\hat{m}$ is a biased estimate of the value $\log(1+m)$, where m is the simple return $(P_{m,1}-P_{m,0})/P_{m,0}$, and $P_{m,0}$ and $P_{m,1}$ are the initial and terminal values of the index over the single period. That is $\log\left[\frac{1}{N}\sum_{i=1}^{N}\left(1+\frac{S_i}{B_i}\right)\right] \geq \frac{1}{N}\sum_{i=1}^{N}\log\left(1+\frac{S_i}{B_i}\right)$.

Under the assumption that the property returns in each period are lognormally distributed, Goetzmann (1992) proposes a method to correct the bias of RSR and approximate the return of market index for each period.

$$\log(\hat{\boldsymbol{b}}_t) \approx \hat{m}_t + \frac{1}{2}\mathrm{var}(\boldsymbol{e}_t) \tag{5}$$

The $\mathrm{var}(\boldsymbol{e}_t)$ term is the cross-sectional variance of the gross returns of individual assets at time t. Although this method works well in simulations, the bias of RSR turns out to be more complex when there are more periods.

II.2 Illustration

To investigate the bias of RSR estimators more thoroughly, we interpret the RSR estimators sample statistics of repeat-sale coefficient observations. Consider a data set consisting of three repeat-sale observations and four time periods numbered from 0 to 2. Since period 0 is the base period, there are two index returns to estimate, corresponding to period 1 and 2. The first two observations respectively cover period 1 and 2. The third observation covers both period 1 and 2. Thus by assuming i.i.d. normally distributed errors, we have

$$X = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}, \; Y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} \log(S_1/B_1) \\ \log(S_2/B_2) \\ \log(S_3/B_3) \end{bmatrix}, \text{ and } \Omega = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}.$$

From equation (4), we have

$$\left(X'\Omega^{-1}X\right)\hat{m} = X'\Omega^{-1}Y\,, \tag{6}$$

In the example, we are able to write down the equations explicitly.

$$\begin{cases} (1+1/2)\hat{m}_1 = y_1 + 1/2\,(y_3 - \hat{m}_2) \\ (1+1/2)\hat{m}_2 = y_2 + 1/2\,(y_3 - \hat{m}_1) \end{cases} \tag{7}$$

From equation (7), we are easily able to interpret the RSR estimators of index returns as geometric averages of individual single period returns or their proxies.

$$\begin{cases} \hat{m}_1 = \dfrac{1}{(1+1/2)}\left[y_1 + 1/2(y_3 - \hat{m}_2)\right] \\ \hat{m}_2 = \dfrac{1}{(1+1/2)}\left[y_2 + 1/2(y_3 - \hat{m}_1)\right] \end{cases} \tag{8}$$

For example, the RSR estimator of log index return for the first time period $\hat{m}_1$ is

$$\hat{m}_1 = \frac{2}{3}y_1 + \frac{1}{3}(y_3 - \hat{m}_2)\,. \tag{9}$$

Obviously it is a weighted average with the weights inversely proportional to the assets' holding periods. This is the motivation for the GLS version of the RSR, which weights observations by the root inverse of the holding period. Therefore the estimator of actual (not-log) index return is

$$\begin{aligned} \hat{b}_1 &= \exp(\hat{m}_1) \\ &= \exp\left(\frac{2}{3}y_1\right)\exp\left(\frac{1}{3}(y_3 - \hat{m}_2)\right). \\ &= g_1^{\frac{2}{3}}\left(g_3 / \hat{b}_2\right)^{\frac{1}{3}} \end{aligned} \tag{10}$$

The g_i, as defined earlier, is $g_i = \exp(y_i) = S_i / B_i$, the not-log compound return for repeat sale observation i.

Clearly the RSR estimator of actual index return in period 1 is geometric average of two numbers. The first number, g_1, is an individual return in period 1. The second number, $g_3 / \hat{b}_2$, is a proxy of individual return in period 1. The third repeat sale observation covers all two periods and g_3 is a compound return. After subtracting the component of the compound return that corresponds to the second period, one can get a single period return in period 1. Though this component is unknown, the estimator for index return in the second period obviously is a proxy of it. Thus $g_3 / \hat{\boldsymbol{b}}_2$ is a proxy of an individual return in period 1.

Why is the RSR estimator of index return in period 1 an average of g_1 and $g_3 / \hat{\boldsymbol{b}}_2$? Clearly it is because the first and the third repeat sale observation, and no other observation, cover period 1. Thus, both of these two observations, and maybe only they, directly provide useful information about the index return in period 1. Another question is why RSR gives these two observations different weights? Notice that g_1 covers only period 1, while g_3 covers both periods and then contains both information and noise for all two periods. So intuitively g_3 contains much more noise, and the $g_3 / \hat{\boldsymbol{b}}_2$ term is not an actual individual return but just a proxy. Thus, it has smaller weight.

From equation (10), the RSR estimator of index return in period 1, $\hat{\boldsymbol{b}}_1$, is a cross-sectional sample geometric mean of all available individual returns in period 1 or their proxies. Actually, all RSR estimators can always be written as weighted geometric averages of individual single-period returns or proxies of them. It is also obvious that RSR estimators of index returns for different periods always depend on each other so that the logarithmic transformation at one period would have direct impact on that period's RSR estimator and serial impacts on other periods' estimators.

III. Bias Decomposition

III.1 Bias components

Interpreting the RSR estimator as a sample geometric average facilitates the investigation of its bias. Specifically, we are able to decompose the bias, defined as the difference between the RSR estimator and an unbiased estimation of index return, into two parts. Denote by $\left\{\hat{b}_t^*\right\}_{t=1}^T$ the unbiased estimators of index returns, which are analogous to the RSR estimators but are arithmetic averages, instead of geometric averages, of individual returns or their proxies. Therefore they directly correspond to actual index returns. We will talk about the estimation of the unbiased estimators in section four.

In the example, suppose the unbiased estimator for index return in time period 2 is $\hat{b}_2^*$, we are able to construct a unbiased proxy of a single period return corresponding to third observation as $g_3/\hat{b}_2^*$. Thus the unbiased estimator for index return in period 1, $\hat{b}_1^*$, would be the arithmetic average of g_1 and $g_3/\hat{b}_2^*$.

$$\hat{b}_1^* = \frac{2}{3}g_1 + \frac{1}{3}\left(g_3/\hat{b}_2^*\right) \tag{11}$$

Decompose the difference between the RSR estimator $\hat{b}_1$ and the unbiased estimator $\hat{b}_1^*$ into two parts.

$$\begin{aligned}\hat{b}_1^* - \hat{b}_1 &= \left[\frac{2}{3}g_1 + \frac{1}{3}\left(g_3/\hat{b}_2^*\right)\right] - g_1^{\frac{2}{3}}\left(g_3/\hat{b}_2\right)^{\frac{1}{3}} \\ &= \left[\frac{2}{3}g_1 + \frac{1}{3}\left(g_3/\hat{b}_2^*\right)\right] - g_1^{\frac{2}{3}}\left(g_3/\hat{b}_2^*\right)^{\frac{1}{3}} + g_1^{\frac{2}{3}}\left(g_3/\hat{b}_2^*\right)^{\frac{1}{3}} - g_1^{\frac{2}{3}}\left(g_3/\hat{b}_2\right)^{\frac{1}{3}}\end{aligned} \tag{12}$$

Jensen's inequality implies that

$$\left[\frac{2}{3}g_1+\frac{1}{3}\left(g_3/\hat{b}_2{}^*\right)\right]-g_1{}^{\frac{2}{3}}\left(g_3/\hat{b}_2{}^*\right)^{\frac{1}{3}}\geq 0\,. \tag{13}$$

This component of the bias is essentially the difference between a geometric average and an arithmetic average, which is always positive as long as all numbers are not the same. It can be corrected using equation (5). However, this is not the end of the story. The other component of the bias is

$$g_1{}^{\frac{2}{3}}\left(g_3/\hat{b}_2{}^*\right)^{\frac{1}{3}}-g_1{}^{\frac{2}{3}}\left(g_3/\hat{b}_2\right)^{\frac{1}{3}}. \tag{14}$$

It is the difference between the geometric mean constructed by subtracting the unbiased estimator for period 2 index return and the geometric mean constructed by subtracting the RSR estimator. If the unbiased estimator for period 2 is larger than the RSR estimator, this component of the bias is positive, otherwise it is negative. Clearly the direction and the magnitude of the bias of RSR estimator for period 2 helps to determine the direction and magnitude of the second component of the bias for RSR estimator for period 1. At the same time, for the same reason, the direction and magnitude of the bias for the RSR estimator for period 1 also helps to determine the direction and magnitude of the bias for RSR estimators in period 2.

We call the first component of the bias for the RSR estimator in each period the *direct* impact of the logarithmic transformation for that period. We call the second component of the bias in each period as the *serial* impact of logarithmic transformation for other periods because it is determined by the bias of RSR estimators for other periods.

These two impacts tend to offset each other since RSR estimators are simultaneously determined and depend on each other. For instance, suppose the direct impact in period 2 is strong enough that the RSR estimator in that period is biased downward: $\hat{b}_2 < \hat{b}_2{}^*$, then the second component of the RSR bias is negative.

$$g_1^{\frac{2}{3}}\left(g_3 / \hat{b}_2^{\,*}\right)^{\frac{1}{3}} - g_1^{\frac{2}{3}}\left(g_3 / \hat{b}_2\right)^{\frac{1}{3}} < 0 .$$

At the same time, the first component is always positive. Therefore it is no longer clear if the RSR estimator for period 1 is lower than the unbiased arithmetic mean for this period or not.

In general, the logarithmic transformation's direct impact in one period tends to make this period's RSR estimator lower than the unbiased arithmetic mean; at the same time, its serial impact on other periods' RSR estimators tend to push them upward. Thus the direction and the magnitude of the bias of RSR estimators becomes ambiguous in the multi-period case, which is obviously different from the one-period case where RSR estimator is simply biased down. We expect on average that the bias is negative, but it may not be straightforward to correct.

III.2 Determination of the magnitude of bias

The first component of the bias of RSR is the difference between a sample arithmetic mean and a sample geometric mean. Jensen's inequality tells us that the difference of $E(\log(r))$ and $\log(E(r))$ depends on the population variance of the random variable r. However, when we estimate the market index, we always work with finite samples and are not able to observe the population variance. We want to investigate what determines the difference between the sample arithmetic mean and the sample geometric mean. Specifically, we want to make sure which one actually determines this difference: the population variance of the underlying data generating distribution or just the sample variance. Suppose we have different samples of cross-sectional returns generated from the same distribution. Are the differences between the log equal-weighted index returns (log of arithmetic averages of individual returns) and the RSR estimators (averages of the log individual returns or geometric averages of not-log individual returns) uniform for all these samples, or they depend on the variance of the sample?

This question is important because if the difference between arithmetic mean and geometric mean is determined by the sample variance, the magnitude of the direct impact component of the RSR bias is potentially predictable from the data no matter what the actual underlying process may be. For periods the variances of sample returns are larger, the magnitude of the first component of the bias for RSR estimators tend to be larger. Thus the RSR estimators for those periods tend to be biased downward. Goetzmann (1992) uses a single number to correct the bias for all periods. Our previous example shows the correction would be insufficient for the periods with larger sample variances, and probably would be too much for other periods with smaller sample variances.

We use a simple experiment to show that sample variance, not the population variance, determines the difference between an arithmetic average and a geometric average of sample returns. We randomly sample from a lognormal distribution 100 times. Each sample consists of 20 observations, and all observations (say, individual gross returns) are logged terms. The mean of the lognormal distribution we use is 0.0414 (1.10 before taking log). The population standard deviation is arbitrarily chosen as 0.18 (1.2 before taking log). We calculate the RSR estimator, which is just the sample mean of these 20 individual observations (logged). We can also easily get the arithmetic average of these 20 individual returns (not logged). Thus we can get the difference between the logged value of the arithmetic average and the RSR estimator for each sample. This difference is what Goetzmann (1992) intended to adjust with approximation.

Figure 1 clearly shows that the difference is almost perfectly related to sample variance of these 20 observations, and the slope of the straight line is just 0.5, which confirms the formula in Goetzmann (1992). Clearly it is the sample variance, not the population variance, that determines the magnitude of the first component of the bias of RSR estimators. Thus, for each time period, the larger is the cross-sectional sample variance, the bigger is the difference between geometric mean and arithmetic mean, i.e., the larger the magnitude of the bias for RSR estimator for that period tend to be. For the approximation method, since estimators for all periods are adjusted with one single number, for periods with larger sample variances, adjustments may be insufficient; for periods with smaller sample variances, adjustments tend to be too much. This phenomenon has been found in our simulation results.

IV. Unbiased RSR Estimators

IV.1 Unbiased RSR Estimators

This section proposes a method to calculate the unbiased estimators that we use to compare with the RSR estimators in earlier section. The unbiased estimators have natural interpretation as arithmetic means of the cross-sectional returns of assets or their proxies. We still denote by $\left\{\hat{\boldsymbol{b}}_t^*\right\}_{t=1}^T$ the unbiased arithmetic mean estimators. The unbiased estimator of index return in time period t, $\hat{\boldsymbol{b}}_t^*$, may be expressed as

$$\hat{\boldsymbol{b}}_t^* = \frac{1}{\sum_{i \in N_t} w_i} \sum_{i \in N_t} w_i \left(g_i \Big/ \prod_{s \in H_i, s \neq t} \hat{\boldsymbol{b}}_s^* \right). \tag{15}$$

The H_i is defined as the holding interval of observation i, which consists of all time periods later than the purchase time and no later than the sale time. Thus $s \in H_i, s \neq t$ are all time periods, except period t, that belong to the holding interval of repeat sale observation i. The N_t is the set of all repeat sale observations that contain period t in their holding intervals. The n_t is the number of observations that belong to N_t. The w_i term is the weight of the repeat observation i. We could follow the RSR and let w_i equal to $\frac{1}{\boldsymbol{t}_i}$.

Obviously the unbiased estimators are analogous to the RSR estimators. First of all, the unbiased estimators are cross-sectional averages of all available individual returns or their proxies in corresponding time periods, just like RSR estimators. Second, they are determined simultaneously and thus depend on each other, also like RSR estimators. However, the unbiased estimators are arithmetic averages of individual asset returns thus strictly correspond to actual index returns, while the RSR estimators are geometric averages. One special advantage of the unbiased estimators is that when all assets trade in all periods, i.e., there is no data missing, the unbiased estimators exactly equal the actual equal-weighted index returns.

Rearrange expression (15) and let w_i equal to $\frac{1}{\boldsymbol{t}_i}$, we get

$$\sum_{i\in N_t}\frac{1}{\boldsymbol{t}_i}=\sum_{i\in N_t}\frac{1}{\boldsymbol{t}_i}\left(g_i\Big/\prod_{s\in H_i}\hat{\boldsymbol{b}}_s^{\,*}\right). \tag{16}$$

We define $u_i\left(\hat{\boldsymbol{b}}^*\right)=g_i\Big/\prod_{s\in H_i}\hat{\boldsymbol{b}}_s^{\,*}$, and a N by 1 vector $U\left(\hat{\boldsymbol{b}}^*\right)$ and a N by 1 vector I.

$$U\left(\hat{\boldsymbol{b}}^*\right)=\begin{bmatrix} u_1\left(\hat{\boldsymbol{b}}^*\right) \\ u_2\left(\hat{\boldsymbol{b}}^*\right) \\ . \\ u_{N-1}\left(\hat{\boldsymbol{b}}^*\right) \\ u_N\left(\hat{\boldsymbol{b}}^*\right) \end{bmatrix}, \text{ and } I=\begin{bmatrix} 1 \\ 1 \\ . \\ 1 \\ 1 \end{bmatrix}.$$

Equation (16) is equivalent to

$$X_{t,\cdot}'\Omega^{-1}I=X_{t,\cdot}'\Omega^{-1}U\left(\hat{\boldsymbol{b}}^*\right),$$

where the $X'_{t,\cdot}$ term is the row t of the matrix X'. Thus the unbiased estimators for index returns for all periods are determined by

$$X'\Omega^{-1}I = X'\Omega^{-1}U\left(\hat{b}^*\right), \text{ or } X'\Omega^{-1}\left(I - U\left(\hat{b}^*\right)\right) = 0. \tag{17}$$

In equation (17), there are T equations and T unknown $\hat{b}_t$. The solution of these equations is the unbiased estimators. Though the equations are not linear, it is feasible to search for the solution via maximum-likelihood. At the same time, the unbiased estimators are solutions to the optimization problem:

$$\min_{\{\hat{b}^*\}} \left(I - U\left(\hat{b}^*\right)\right)' \Omega^{-1} X X' \Omega^{-1} \left(I - U\left(\hat{b}^*\right)\right). \tag{18}$$

Thus quadratic-searching technology can help to calculate the unbiased estimator $\hat{b}^*$.

The RSR estimators are shown to have other problems. For example, there are the sample selectivity (Clapp and Giaccotto [1992]) and unobserved fix-ups between sales (Goetzmann and Spiegel [1995] and Clapp and Giaccotto [1999]). The RSR has many variants to address these problems. Though the unbiased estimators we propose intend to address the bias problem only, in principle they could have variants that address not only the bias problem but also other problems at the same time.

IV.2 Comparison with ARS by Shiller

We use the same small data set to compare the unbiased estimator with the instrumental variable arithmetic-mean repeat sale estimator (ARS) proposed by Shiller (1991). The ARS estimators by Shiller (1991) are for reciprocal index levels instead of returns. Here we translate the reciprocal index level estimators into return estimators to facilitate the comparison.

Denote by $\hat{b}_t^{\,S}$ the ARS estimator of equal-weighted index return in time period t. The estimator-determining equations of the ARS for the simple data set are

$$\begin{cases} (g_1+1)\dfrac{1}{\hat{b}_1^{\,S}} - g_2\dfrac{1}{\hat{b}_1^{\,S}\hat{b}_2^{\,S}} = 1 \\ -\dfrac{1}{\hat{b}_1^{\,S}} + (g_2+g_3)\dfrac{1}{\hat{b}_1^{\,S}\hat{b}_2^{\,S}} = 1 \end{cases}.$$

Rearrange the equations, we have

$$\begin{cases} g_1 - \hat{b}_1^{\,S} = \dfrac{g_2}{\hat{b}_2^{\,S}} - 1 \\ \dfrac{g_2}{\hat{b}_2^{\,S}} + \dfrac{g_3}{\hat{b}_2^{\,S}} = \hat{b}_1^{\,S} + 1 \end{cases}.$$

We express the ARS estimators in an economically sensible way as follows:

$$\hat{b}_1^{\,S} = \frac{1}{2}g_1 + \frac{1}{2}\left(g_3/\hat{b}_2^{\,S}\right),\ \hat{b}_2^{\,S} = \frac{2g_2+g_3}{2+g_1} = \frac{2}{2+g_1}g_2 + \frac{g_1}{2+g_1}(g_3/g_1).$$

Notice both g_1 and g_2 are single period individual returns, in period 1 and 2 respectively, while the g_3 is a two-period compound return. Obviously the $g_3/\hat{b}_2^{\,S}$ is a proxy of single-period return in period 1 and the g_3/g_1 is a proxy of single-period return in period 2. The ARS estimators are arithmetic averages of individual single-period returns. Specifically, the index return estimator in period 1 is average of g_1 and $g_3/\hat{b}_2^{\,S}$, and the index return estimator in period 2 is average of g_2 and g_3/g_1. As mentioned earlier, the unbiased estimators proposed in this article for the data set are

$$\hat{b}_1^{*} = \frac{2}{3}g_1 + \frac{1}{3}\left(g_3/\hat{b}_2^{*}\right),\ \hat{b}_2^{*} = \frac{2}{3}g_2 + \frac{1}{3}\left(g_3/\hat{b}_1^{*}\right).$$

Obviously they are also arithmetic averages.

Though both the ARS estimators and the unbiased estimators are arithmetic averages of individual single-period returns or their proxies, the example shows two interesting differences between them. First, in the ARS estimators the proxy for period 2 return is g_3/g_1, while in the unbiased estimators it is $g_3 \big/ \hat{b}_1^*$. If $\hat{b}_1^*$ is a more accurate estimator of index return in period 1 than g_1, then we expect $g_3 \big/ \hat{b}_1^*$ to be a better proxy than g_3/g_1. Second, the weights of assets are consistent in the unbiased estimators. The first asset, which provides two single-period returns g_1 and g_2, receives two third weight in both periods; while the second asset, which provides one compound return g_3, receives one third weight in both periods. The weights of assets in the ARS estimators are different in two periods: in period 1 two assets receive equal weights, while in period 2 the weight of the first asset is almost two times heavier than that of the second asset. In short, based on the example the unbiased estimators seem to be more sensible than the ARS estimators are in constructing the proxies of individual returns and weighting assets. However, the ARS estimators have a great advantage over the unbiased estimators: they are easy to calculate.

IV.3 Feasibility of Calculation

Though the calculation of the unbiased estimators is not as easy as that of the RSR or the ARS estimators, it is still feasible. To show this, we use the same data used by Case and Shiller (1987, 1989) and Shiller (1991) to estimate the equal-weighted quarterly price index for Dallas from 1970-1 to 1986-2. The data have 6,669 repeat sale observations. Let the index value at 1970-1 be 1, there are 65 index returns to estimate. The software we use is S-plus, and the computer is a public-shared Unix server in Yale International Center for Finance. The calculation of the unbiased estimators is equivalent to searching for $\{\hat{b}^*\}$ that minimizes the objective function:

$$\left(I - U\left(\hat{b}^*\right)\right)' \Omega^{-1} X X' \Omega^{-1} \left(I - U\left(\hat{b}^*\right)\right).$$

We use an extremely simple but obviously not-efficient search procedure. First, we calculate the RSR estimators. Second, we use the RSR estimators as the starting point, randomly draw a new point within a small region around the starting point. We do so until the new point is better than the starting point, i.e. it reduces the value of the objective function, then use the new point as the starting point for next run of searching. This procedure is repeated until the value of the objective function is small enough. After several hours' searching, the value of the objective function is reduced from 238.5, which corresponds to the RSR estimators, to 2.5. Taking account that more efficient search procedure and more powerful computer could be dedicated to the estimation, the feasibility of the calculation of the unbiased estimators is obvious.

Figure 2 shows the RSR, the ARS, and the unbiased estimators for the quarterly equal-weighted price index for Dallas. Obviously these indices are different. However, little can be said about the accuracy of each method, which is investigated in next section.

V. Simulation Test

V.1 Methods

The goal of the simulation is twofold. First, we want to test our predictions about the RSR estimators and the approximation method by Goetzmann (1992). We predict that the RSR estimators tend to be biased downward more for periods with larger cross-sectional variances of asset returns. We also predict that the adjustment according to Goetzmann (1992) may be insufficient for periods with larger variances but too much for periods with smaller variances. Second, we want to test the performance of the unbiased estimators proposed here, together with some other alternative estimators for RSR. Specifically, we test performance of five different estimators. They are RSR estimators, adjusted RSR estimators according to Goetzmann (1992), ARS suggested by Shiller (1991), nonlinear direct estimators suggested by Goetzmann and Geltner (1999), and the unbiased arithmetic mean RSR proposed in this article.

The RSR estimators, the adjusted RSR estimator by approximation, and the ARS estimators are well known. The direct estimators suggested by Goetzmann and Geltner come from solving following problem:

$$\min_{\{\hat{b}_t\}_{t=1}^{T}} \sum_{i=1}^{N} w_i \left(\frac{g_i}{\prod_{s \in H_i} \hat{\boldsymbol{b}}_s} - 1 \right)^2 . \tag{19}$$

Equation (19) can be rewritten as

$$\min_{\{\hat{b}_t\}_{t=1}^{T}} \left(I - U\left(\hat{\boldsymbol{b}}\right)\right)' W \left(I - U\left(\hat{\boldsymbol{b}}\right)\right). \tag{20}$$

It is interesting to notice that the optimization problem is very similar to that in equation (18), which our arithmetic repeat sales estimators solve. The only difference is the weight matrix in the middle. While our method uses $\Omega^{-1} X X' \Omega^{-1}$, Goetzmann and Geltner use W. A nice property of the direct method is that, if let the weight being constant for all observations, the direct estimators would minimize the mean-squared-error (MSE). In the simulations, we let the weight for repeat-sale observation i be $1/\boldsymbol{t}_i$ for the direct method.

V.2 Simulation steps

In each simulation, we construct the "actual" market first. The following steps are performed:

1. Specify the number of assets N , and length of time horizon T .
2. Randomly draw the underlying marker returns (log term) for all periods from a normal distribution. The distribution has mean corresponding to 110% gross return and standard

deviation corresponding to 17%, which are reasonable numbers for financial market annual returns.

3. For each and every time period, randomly draw N individual asset returns from a normal distribution with mean equal to that period's underlying market return and variance assigned by us. After that, we have a N by T panel data set, which is treated as a perfect T-period sample data set from a N-asset market.
4. From this complete data set, construct the actual equal-weighted market index, which is the benchmark used to test estimators' accuracy.

After constructing the actual market, we are able to construct repeat sale data from the complete data set. Following steps are repeated for 100 times for each actual market data we create:

1. Randomly draw two different dates for each asset, and calculate the compound returns between them. Then, instead of having the perfect panel data, we now have only one repeat sale observation for each asset.
2. Use all five methods to estimate the index returns based on this repeat sale data set. For two nonlinear estimators, direct estimators and the unbiased RSR estimators, we use RSR estimators as starting point for search.
3. For each method, calculate the estimators' percent deviations from the actual index return in each period. For example, if an estimated return is 1.1 but the real market return is 1.0, the percent deviation is 10%. We also calculate all methods' mean squared errors over all periods.

We choose $N = 30$, $T = 3$. We use the short time horizon $T = 3$ because in each round of simulation we need to search for two kinds of nonlinear estimators 100 times. The cross-sectional variances for three periods are (0.02, 0.02, 0.02), (0.02, 0.08, 0.02), (0.08, 0.02, 0.08), and (0.02, 0.04, 0.08). We purposely control the variances so that the actual market will exhibit specific patterns of time-varying cross-sectional variance. For example, by setting the variances as 0.02, 0.08, and 0.02, we are able to test if RSR estimators are more biased downward in the second period. We are also interested in the performance of five different methods in each scenario. For each setting of variances, we run 3 rounds of simulations. Each round has its own actual market, and consists of 100 different repeat sale data sets generated from the same actual market. Thus there are 12 rounds of simulations overall.

V.3 Simulation results

Table 1 presents simulation results for four different variances specifications. In each setting, the table presents the percentage deviation of each method's estimator in each period (median of 100 simulations), and each method's MSE for all three periods (median of 100 simulations). In each time period, the two smallest percentage deviations are in bold. Figure 3 and figure 4 plot the percentage deviations of RSR estimators and adjusted RSR estimators; figure 5 and figure 6 plot the percentage deviation of the ARS and the unbiased RSR estimators, all under the settings that cross-sectional variances are 0.02, 0.08, and 0.02.

We have four major findings from the simulation results:

1. As we predict, the RSR estimators tend to be more biased down in periods with larger variances. After adjusted according to Goetzmann (1992), the estimators tend to be biased down for periods with larger variances, but tend to be higher than the actual index returns for periods with smaller variances. These are shown in Table 1, figure 3, and figure 4. For example, in figure 3, the second period has larger variance. Clearly the RSR estimators are biased down much more in the second period than in other two periods. In figure 4, after adjusted according to Goetzmann (1992), the RSR estimators are biased down in the second

period, but generally not biased down in other periods. This confirms that the adjustment may be insufficient for periods with larger variances, while may be too much for periods with smaller variances.

2. The unbiased estimators and the ARS estimators seem to be robust to time-varying cross-sectional variance. There is no obvious bias-pattern for these two methods, as can be seen in figure 5 and figure 6. At the same time, the unbiased estimators seem to be more accurate than the RSR estimators, the adjusted estimators, and the direct estimators, and may be more the ARS estimators too. The unbiased estimators generally have smaller percentage deviations from actual index returns from period to period, which is shown in Table 1 and the figure 6.
3. The direct estimators of Goetzmann and Geltner tend to have larger percent deviations from the actual index returns. This may be partially caused by the difficulty of searching for the global minimum value. However, this method tends to have small MSE.

VI. Conclusion

We interpret the RSR estimators as sample statistics and show that they are essentially geometric averages of individual returns or their proxies. At the same time, it is clear that the RSR estimators are jointly determined and depend on each other. We decompose the bias of a RSR estimator into two parts. The two components of the bias are respectively determined by two different impacts of the logarithmic transformation of the price relatives. One we term the direct impact and the other we term the serial impact. These two impacts push the bias toward opposite directions. Specifically, the direct impact pushes RSR estimators downwards, while the serial impact pushes them upwards. The actual bias of a repeat sales estimator for one specific time period is jointly determined by the summary of these two impacts in that period.

We show that the magnitude of the direct impact is positively related to the cross-sectional sample variance of individual returns in the prevailing period, while the magnitude of the serial impact is negatively related to the sample variances in surrounding periods. Therefore, the magnitude of the actual bias generally varies through time since the sample variances of individual asset returns are usually different through time. At the same time, the bias magnitude of the RSR estimators and the accuracy of the approximation method by Goetzmann (1992) are predictable to some extent. The RSR estimator tends to be biased down more in periods with larger cross-sectional variances of individual asset returns and less in periods with smaller cross-sectional variances. The approximation method would insufficiently compensate for the bias in periods with larger variances while more than enough for time periods with smaller variances.

We propose unbiased repeat sales estimators that are analogous to the RSR estimators but are arithmetic averages of individual returns instead of geometric averages of them. The unbiased estimators strictly correspond to the actual index returns and there is no bias caused by the difference between geometric means and arithmetic means. When there is no data missing, i.e. all assets trade in all periods, the unbiased repeat sale estimators would exactly equal the actual index returns.

We use simulations to test our predictions of the behavior of RSR estimators and the adjusted RSR estimators, and the performance of the unbiased estimators proposed in this paper together with other alternative methods. We construct artificial "actual" market data, and create repeat sale observations from them. We estimate the "actual" index returns with different methods and calculate the deviations of different estimators from the "actual" index returns. The simulation results confirm our predictions. They show that RSR estimators tend to be more biased down in periods with larger sample variances. After adjusted according to Goetzmann (1992), the estimators tend to be biased down for periods with larger variance, but tend to be higher than the actual index returns for periods with smaller variances. The results also show that the unbiased estimators are robust to time-varying cross-sectional variance and may be more accurate than the RSR estimators as well as some other alternatives.

References:

Abraham, J.M. 1990. Statistical Biases in Transaction-Based Indices. *Mimeo, Federal Home Loan Mortgage Corporation*.

Abraham, J.M. and W.S. Schauman. 1991. New Evidence on Home Prices from Freddie Mac Repeat Sales. *AREUEA Journal* 19(3): 333-352.

Bailey, M.J., R.F. Muth and H.O. Nourse. 1963. A Regression Method for Real Estimate Price Index Construction. *Journal of the American Statistical Association* 58: 933-942.

Case, B., H.O. Pollakowski and S.M. Wachter. 1991. On Choosing among House Price Index Methodologies. *AREUEA Journal* 19(3): 286-307.

Case, B. and J.M. Quigley, J. M. 1991. The dynamics of Real Estate Prices. *The Review of Economics and Statistics* 73(1): 50-58.

Case, K.E. and R.J. Shiller. 1987. Prices of Single Family Homes since 1970: New Indexes for Four Cities. *New England Economic Review*: 46-56.

Case, K.E. and R.J. Shiller. 1989. The Efficiency of the Market for Single Family Homes. *American Economic Review* 79: 125-137.

Case, K.E., R.J. Shiller and A.N. Weiss. 1993. Index-based Futures and Options Markets in Real Estate. *Journal of Portfolio Management* 19: 83-92.

Clapp, J.M. and C. Giaccotto. 1992. Estimating Price Trends for Residential Property: A Comparison of Repeat Sales and Assessed Value Methods. *Journal of Real Estate Finance and Economics* 5(4): 357-74.

Clapp, J.M. and C. Giaccotto. 1999. Revisions in Repeat-Sales Indexes: Here Today, Gone Tomorrow? *Real Estate Economics* 27(1): 79-104.

Dombrow, J., J.R. Knight and C.F. Sirmans. 1997. Aggregation Bias in Repeat-Sales Indices. *Journal of Real Estate Finance and Economics* 14(1): 75-88.

Geltner, D. 1997. Bias and Precision of Estimates of Housing Investment Risk Based on Repeat-Sales Indices: A Simulation Analysis. *Journal of Real Estate Finance and Economics* 14(1): 155-171.

Goetzmann, W.N. 1992. The Accuracy of Real Estate Indices: Repeat Sale Estimators. *Journal of Real Estate Finance and Economics* 5(1): 5-53.

Goetzmann, W.N. 1993. Accounting for Taste: Art and the Financial Markets over Three Centuries. *American Economic Review* 83(5): 1370-1376.

Goetzmann, W.N. and M. Spiegel. 1995. Non-temporal Components of Residential Real Estate Appreciation. *Review of Economics and Statistics* 77(1): 199-206.

Goetzmann, W.N. and M. Spiegel. 1997. A Spatial Model of Housing Returns and Neighborhood Substitutability. *Journal of Real Estate Finance and Economics* 14(1): 11-31.

Kuo, C. 1996. Housing Price Dynamics and the Valuation of Mortgage Default Options. *Journal of Housing Economics* 5(1): 18-40.

Mark, J.H. and M.A. Goldberg. 1984. Alternative Housing Price Index: An Evaluation. *AREUEA Journal* 12(1): 30-49.

Palmquist, R.B. 1982. Measuring Environmental Effects on Property Values without Hedonic Prices. *Journal of Urban Economics* 11(3): 333-347.

Pollakowski, H.O. and S.M. Wachter. 1990. Effects of Land Use Constraints on Housing Prices. *Land Economics* 66(3): 315-324.

Shiller, R.J. 1991 Arithmetic Repeat Sales Price Estimators. *Journal of Housing Economics* 1(1): 110-126.

Shiller, R.J. 1993a. *Macro Markets*. Clarendon Press: Oxford.

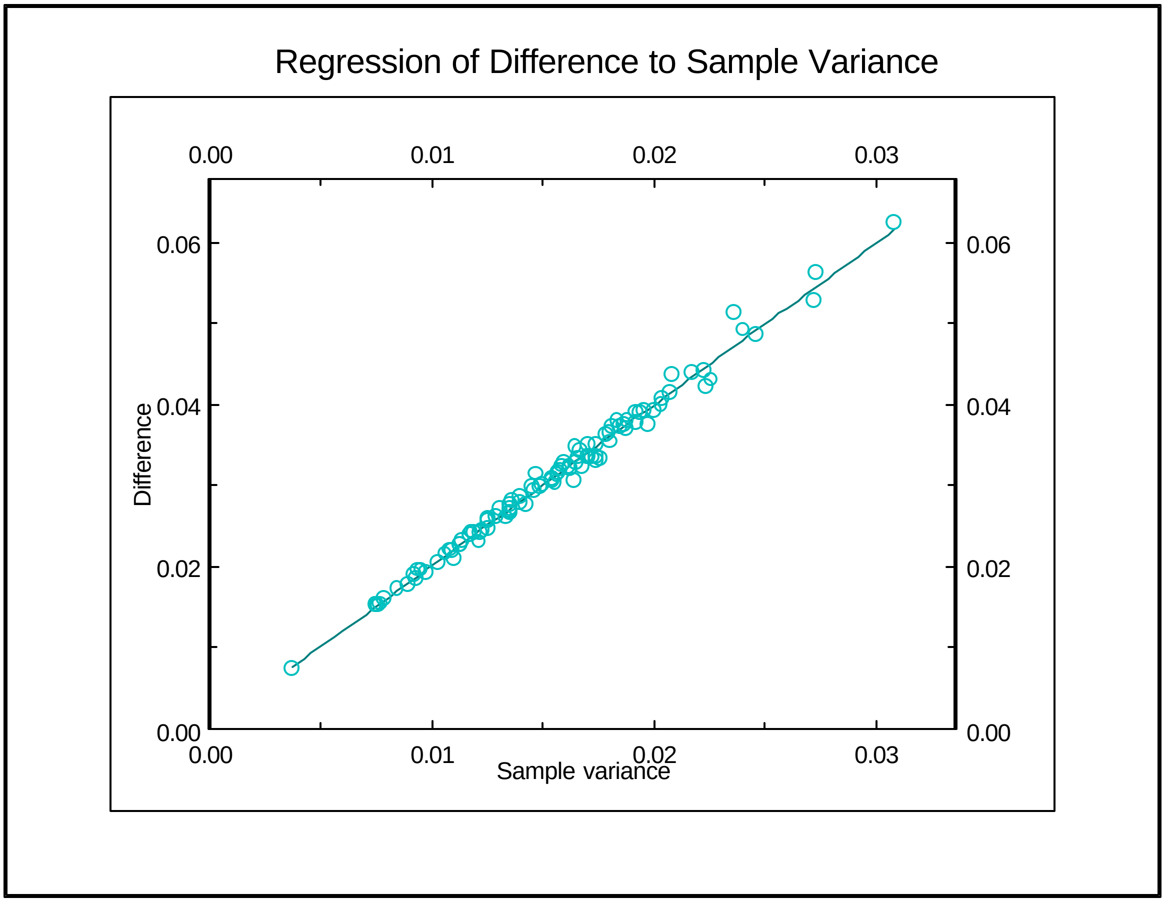
Shiller, R.J. 1993b. Measuring Asset Values for Cash Settlement in Derivative Markets: Hedonic Repeated Measures Indices and Perpetual Futures. *Journal of Finance* 48 (3): 911-931.

Table 1

The variances of individual assets for three periods are 0.02, 0.02, and 0.02												
	Simulation round 1				Simulation round 2				Simulation round 3			
	Median of deviation for each period (in percentage)			MSE	Median of deviation for each period (in percentage)			MSE	Median of deviation for each period (in percentage)			MSE
Methods	1	2	3		1	2	3		1	2	3	
RSR	-0.61	-1.13	-0.59	0.21	**0.38**	-1.88	**-0.34**	0.32	-2.79	**0.54**	-2.04	0.64
Adjusted RSR	**0.58**	**-0.28**	**0.46**	0.22	1.11	-1.17	0.50	0.32	-2.34	**0.85**	-1.63	0.62
ARS	0.60	-1.18	0.69	0.28	**0.98**	**-0.81**	0.73	0.37	-3.17	1.82	-2.57	0.73
Direct Method	1.89	0.96	3.55	0.21	2.71	1.21	1.42	0.25	**-1.48**	4.30	**0.55**	0.41
Unbiased	**0.26**	**-0.26**	**0.30**	0.23	1.29	**-0.90**	**0.04**	0.31	**-2.21**	1.83	**-0.98**	0.61
The variances of individual assets for three periods are 0.02, 0.08, and 0.02												
	Simulation round 1				Simulation round 2				Simulation round 3			
	Median of deviation for each period (in percentage)			MSE	Median of deviation for each period (in percentage)			MSE	Median of deviation for each period (in percentage)			MSE
Methods	1	2	3		1	2	3		1	2	3	
RSR	-0.91	-3.67	-1.76	0.89	**-0.30**	-3.92	-1.26	0.68	-0.60	-2.71	**0.29**	0.42
Adjusted RSR	0.58	-2.38	**-0.24**	0.86	0.84	-2.06	0.68	0.69	**-0.17**	-1.64	1.43	0.44
ARS	**0.13**	**-0.01**	-2.79	1.26	1.13	**-1.92**	**0.66**	0.87	-0.19	**-0.57**	**0.09**	0.46
Direct Method	1.90	6.67	-2.89	1.54	2.55	4.96	3.02	1.04	0.92	2.88	1.05	0.38
Unbiased	**0.11**	**-0.43**	**-1.66**	0.98	**0.24**	**-0.90**	**0.50**	0.68	**-0.11**	**-0.95**	0.72	0.42
The variances of individual assets for three periods are 0.08, 0.02, and 0.08												
	Simulation round 1				Simulation round 2				Simulation round 3			
	Median of deviation for each period (in percentage)			MSE	Median of deviation for each period (in percentage)			MSE	Median of deviation for each period (in percentage)			MSE
Methods	1	2	3		1	2	3		1	2	3	
RSR	-3.42	-1.44	-4.12	1.05	-2.20	-2.71	-3.04	0.78	-5.41	**-0.19**	-4.40	1.09
Adjusted RSR	**-0.33**	1.40	-1.75	1.04	**1.43**	**0.33**	**0.68**	0.82	-3.45	2.34	**-1.09**	1.06
ARS	-0.46	**0.38**	**0.88**	0.98	2.71	-2.04	0.98	0.83	**-1.27**	0.49	-2.22	0.85
Direct Method	4.50	3.04	9.63	1.18	9.90	**0.49**	8.84	1.10	4.51	1.33	3.91	0.90
Unbiased	**-0.30**	**-0.42**	**0.43**	1.00	**1.42**	-1.45	**0.67**	0.75	**-1.68**	**0.30**	**-1.62**	0.94
The variances of individual assets for three periods are 0.02, 0.04, and 0.08												
	Simulation round 1				Simulation round 2				Simulation round 3			
	Median of deviation for each period (in percentage)			MSE	Median of deviation for each period (in percentage)			MSE	Median of deviation for each period (in percentage)			MSE
Methods	1	2	3		1	2	3		1	2	3	
RSR	-0.92	-1.07	-4.16	0.54	-1.67	-2.06	-2.35	0.87	**-0.08**	-2.90	-4.04	0.70
Adjusted RSR	1.01	1.20	-2.31	0.57	**-0.98**	-0.84	-1.66	0.85	1.25	**-1.00**	-2.20	0.72
ARS	**-0.56**	**-0.33**	**0.63**	0.67	-1.27	**-0.58**	**-0.39**	0.94	1.09	**-0.20**	**-0.39**	0.84
Direct Method	2.82	2.31	5.21	0.51	1.79	3.15	4.75	1.02	1.59	2.54	6.68	0.86
Unbiased	**0.14**	**0.50**	-1.15	0.53	**-0.23**	**-0.50**	**0.14**	0.86	**0.52**	-1.19	**-1.61**	0.69

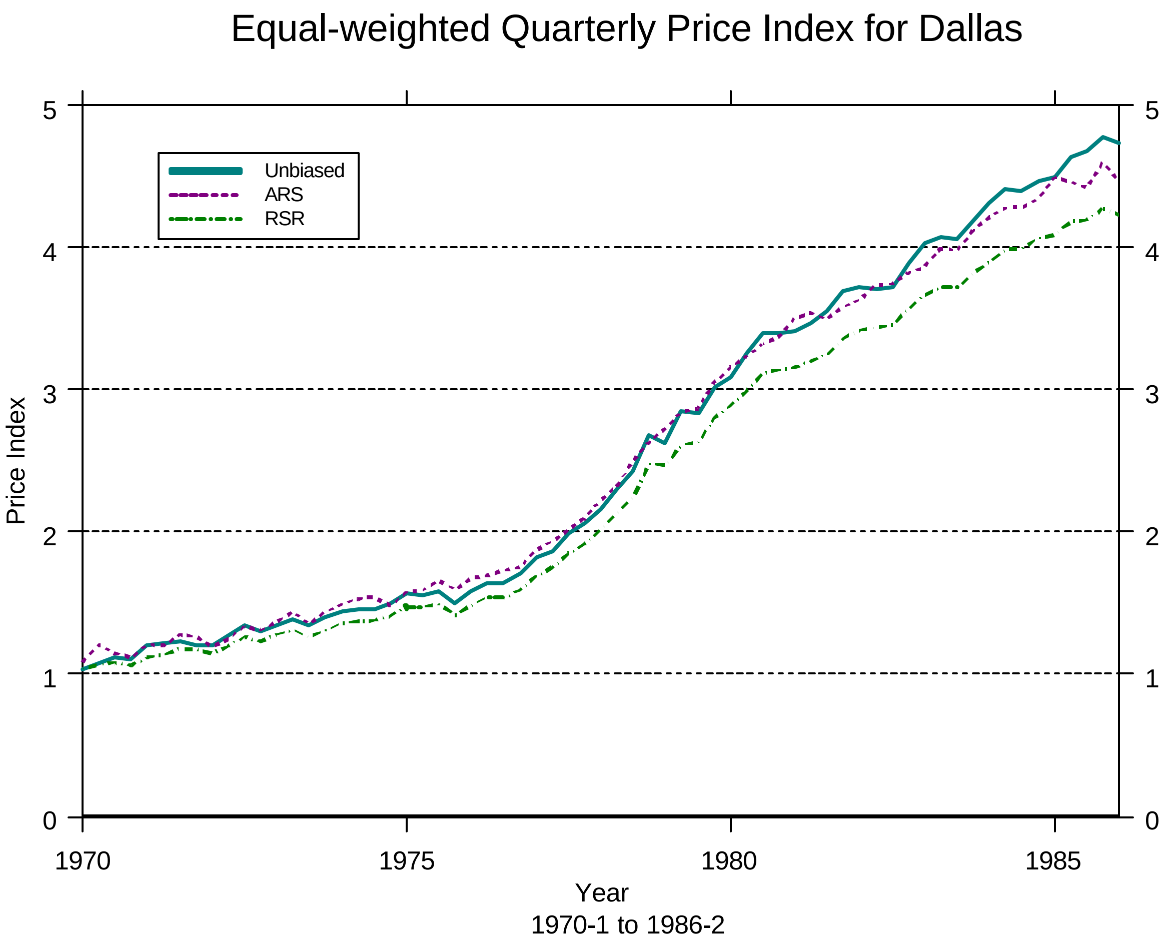
In each round of simulation, complete data of individual returns are generated by drawing 100 individual asset returns each period from a distribution with mean equal to the index return for that period, which is randomly generated too, and with variances specified by us. A round consists of 100 simulations. In each simulation, first we randomly generate one repeat sale observation for each "asset" and construct a repeat sale data set. Then we estimate the "actual " index returns with different methods. We calculate the percentage deviation from the "actual" index return in each period for each method. We also calculate the mean squared errors (MSE) for each method. The numbers in the table are medians over 100 simulations. In each round of simulation, the two smallest percentage deviation numbers are in bold.

Figure 1



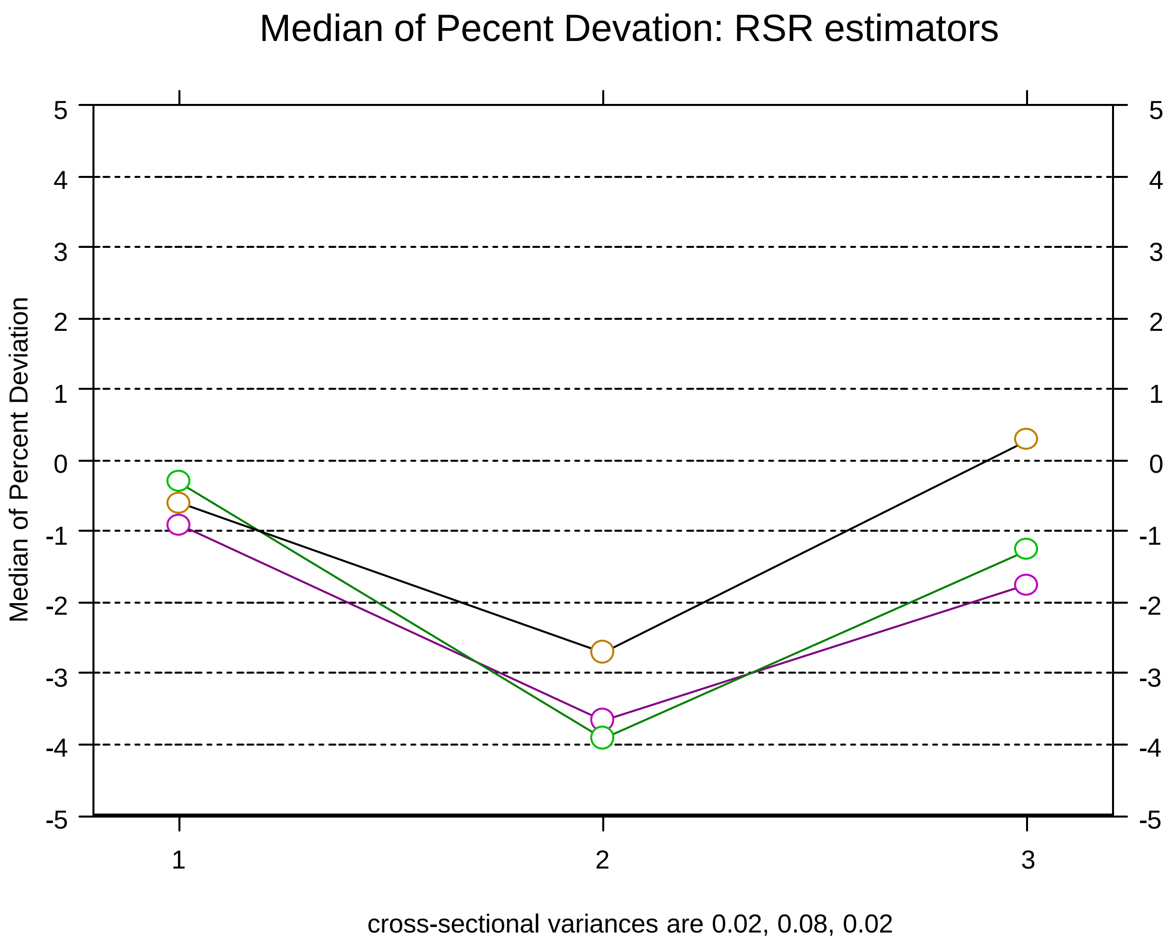
We randomly get 100 samples from a lognormal distribution. Each sample consists of 20 observations randomly generated from the distribution. The mean of the lognormal distribution is 0.0414 (1.10 before taking log). The population standard deviation is 0.18. The "difference" for each sample is the difference between the average of these 20 log values and the log of the average of 20 not-log values. The "sample variance" is the sample variance of these 20 observations.

Figure 2



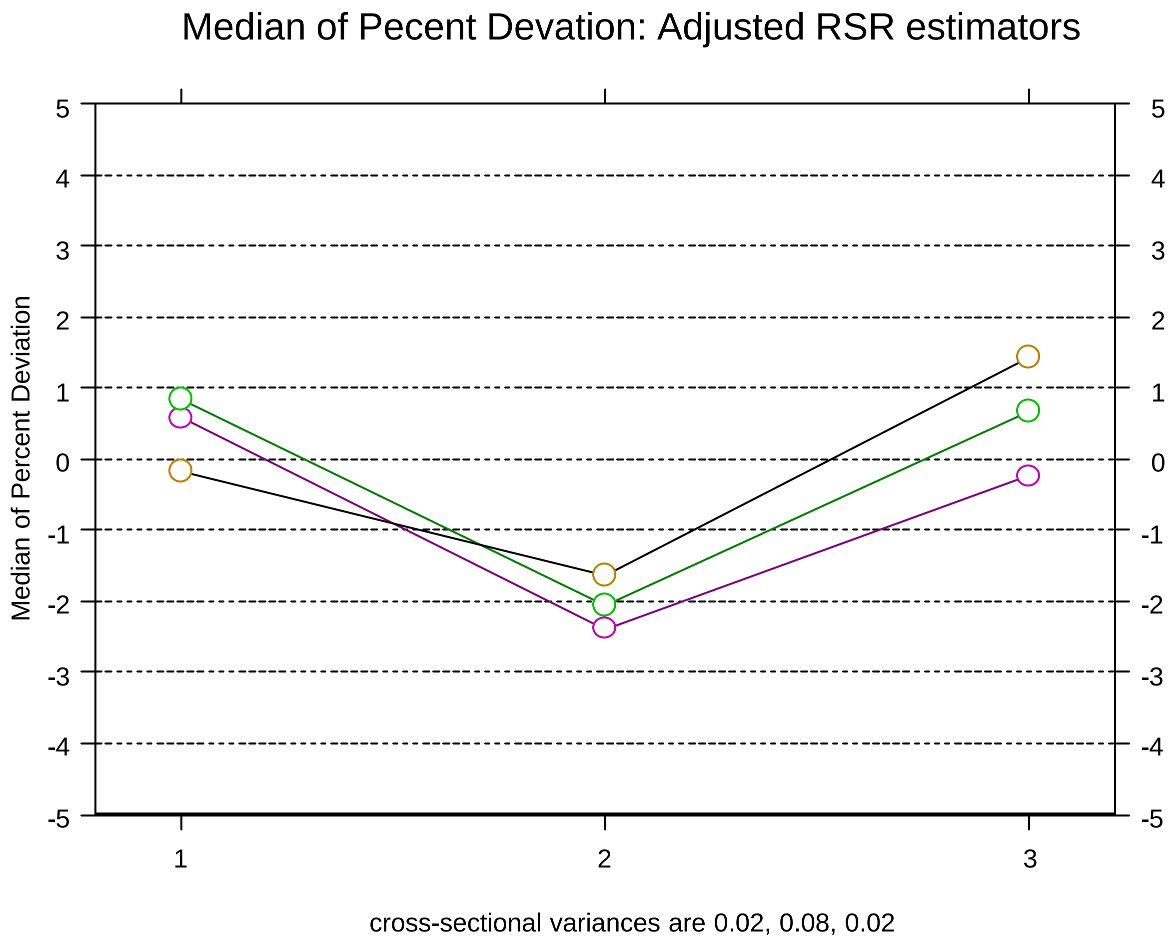
This figure shows the equal-weighted quarterly price indices for Dallas from 1970-1 to 1986-2 calculated with the RSR, the ARS, and the unbiased method.

Figure 3



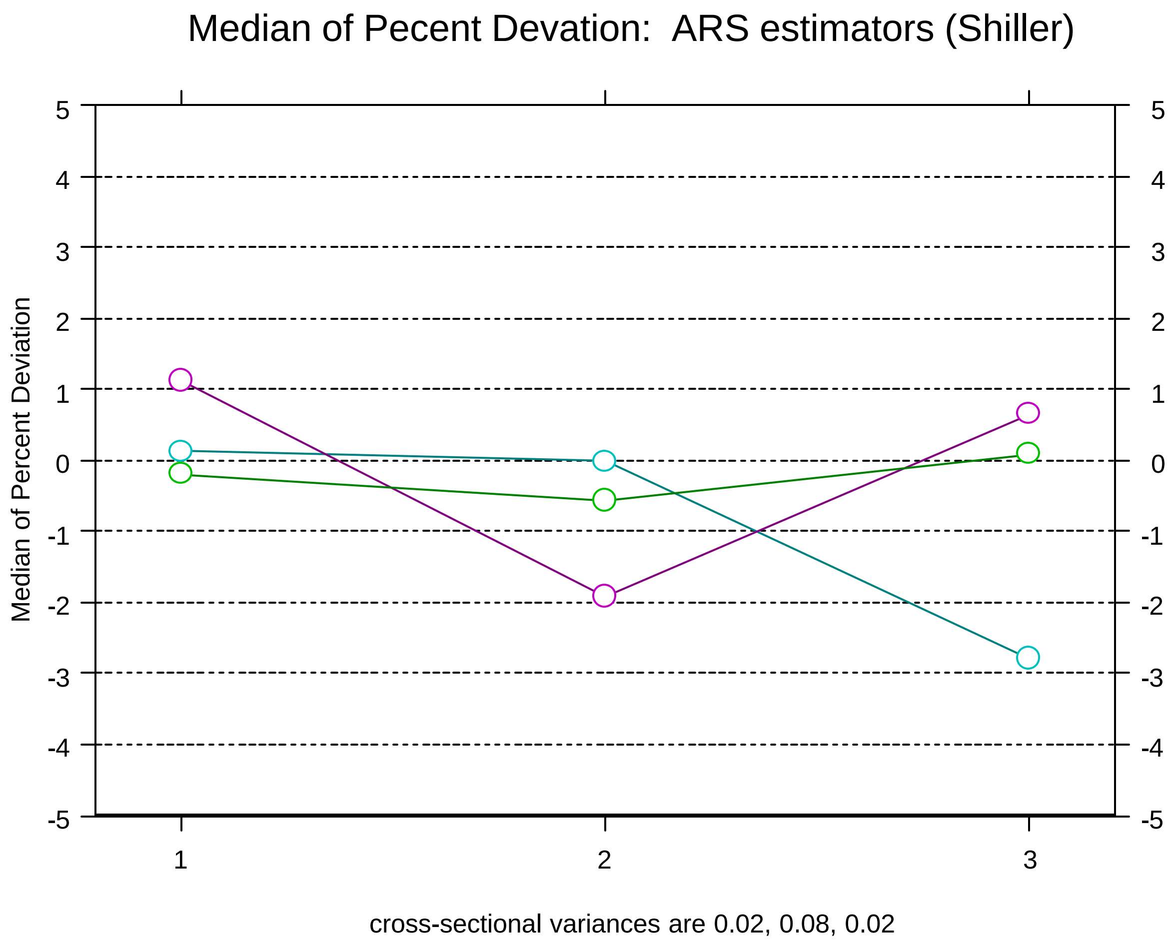
This figure shows the accuracy of the RSR estimators measured by the deviation (in percentage) from the "real equal-weighted market index". Results (three lines) are for three rounds of simulations. Three points in each line correspond to the medians of percent deviation of estimators in three periods.

Figure 4



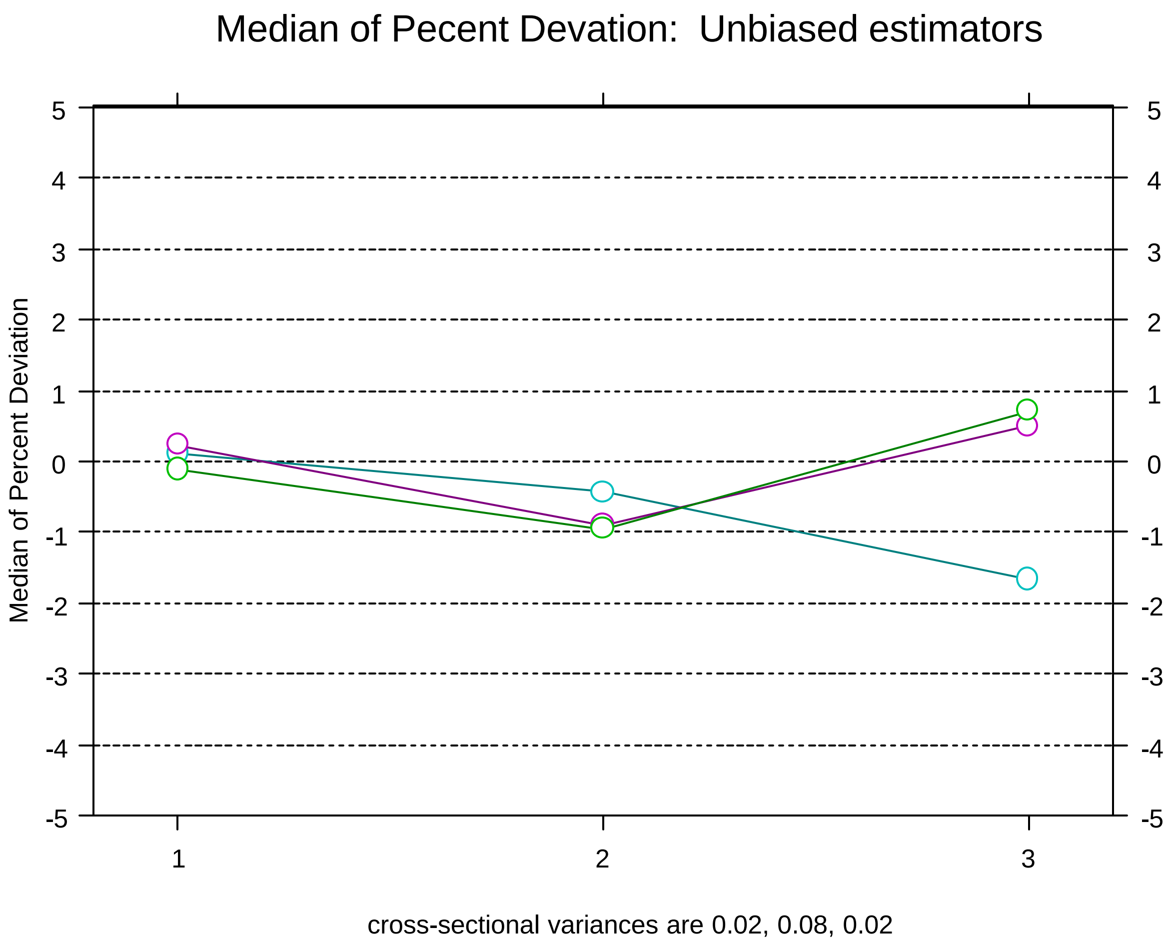
This figure shows the accuracy of the adjusted RSR estimators, which correct for the bias caused by Jensen's inequality with approximation proposed by Goetzmann (1992), measured by the deviation (in percentage) from the "real equal-weighted market index". Results (three lines) are for three rounds of simulations. Three points in each line correspond to the medians of percent deviation of estimators in three periods.

Figure 5



This figure shows the accuracy of the ARS estimators by Shiller (1991) measured by the deviation (in percentage) from the "real equal-weighted market index". Results (three lines) are for three rounds of simulations. Three points in each line correspond to the medians of percent deviation of estimators in three periods.

Figure 6



This figure shows the accuracy of the unbiased RSR estimators that we propose measured by the deviation (in percentage) from the "real equal-weighted market index". Results (three lines) are for three rounds of simulations. Three points in each line correspond to the medians of percent deviation of estimators in three periods.

[1] For example, Abraham and Schauman (1990), Case and Quigley (1991), Case and Shiller (1987, 1989), Case, Shiller, and Weiss (1993), Goetzmann (1993), Goetzmann and Spiegel (1997), Mark and Goldberg (1984), Palmquist (1982), Pollakowski and Wachter (1990).

[2] There has been a great deal of research about problems of RSR. For example, Abraham (1990), Case, Pollakowsi, and Wachter (1992), Clapp and Giaccotto (1992, 1999), Dombrow, Knight, and Sirmans (1997), Goetzmann and Spiegel (1995), Geltner (1997), Kuo (1996), and Shiller (1993a, 1993b).

[3] No paper has been written to propose this method yet.