Measuring Values of Illiquid Asset Portfolios: A General Model

Liang Peng[*]

First Draft: March 5, 2000
Current Draft: September 13, 2000

Abstract:

This paper proposes a model for the estimation of time series returns of illiquid asset portfolios. The model has four major advantages. First, the estimators are arithmetic averages of individual asset returns (or their proxies), so they strictly correspond to the portfolio returns. Second, the model is able to estimate returns of arbitrary-weighted portfolios, including equal-weighted, price-weighted, and value-weighted portfolios. Third, the model is very general and accommodates methods proposed in previous research. Fourth, the model is able to use both price data and data of asset characteristics to improve efficiency of estimators. It is also able to correct for the sample bias problem that transactions may take place more likely on over-valued assets. Simulations with actual data of Dow Jones Industrials show that this model supplies superior estimators than a simple benchmark method.

---

[*] Department of Economics, and International Center for Finance, Yale University, New Haven, CT 06520-8281 (liang.peng@yale.edu).

Measuring values of illiquid asset portfolios is important because many important assets transact infrequently. For example, the real estate, art, and bond markets are generally illiquid. In real estate and art markets, assets tend to be held for years or even decades between sales. In United States bond markets, less than 10% of bonds transact daily. The equity markets also have infrequent trading problem to researchers who work with high frequency transaction data. The global equity market, if considered as an integrated market, is actually illiquid. While the global equity market is considered open, equities in a regional market may not be tradable at all because this regional market is close. Not only many existing asset markets are illiquid; many to-be-established markets might be so. For example, for the new macro markets originally proposed by Shiller (1993a), such as national income and labor income markets, the underlying cash market prices may be observable only infrequently.

Repeat sales regression (RSR) is a broadly used method for estimating the returns of illiquid asset portfolios. It estimates the time series returns using the observed transaction prices for a subset of assets. First suggested by Bailey, Muth, and Nourse (1963), RSR has had a lot of modifications and variants since then. For example, Case and Shiller (1987, 1989), Goetzmann (1989, 1992), Goetzmann and Peng (2000), Goetzmann and Spiegel (1995, 1997), Shiller (1991, 1993b), Webb (1988). The original RSR model has two serious limitations. First, its estimators are geometric averages of cross-section individual asset returns, while the true returns for a portfolio, no matter an equal-weighted or a value-weighted one, are always arithmetic averages of individual asset returns. Goetzmann (1992) proposes a correction method that approximates the arithmetic average given the geometric average, under the assumption that the asset returns in each period are identically lognormally distributed. This method works well in simulations. However, this method needs to estimate unobserved cross-sectional variances, which may not be easy in some cases such as when time series data are heteroskedastic. As an alternative, Goetzmann and Peng (2000) propose a method that directly estimates the arithmetic equal-weighted portfolio returns. The second limitation of the original RSR method is that it actually supplies estimators for equal-weighted portfolio returns only. In many situations, researchers may be more interested in price-weighted, value-weighted, or other special-weighted portfolios. Shiller (1991) proposes estimators, either price-weighted or equal-weighted, that are analogous to the original RSR estimators but are arithmetic averages. However, more flexible method that could estimate returns of arbitrary-weighted portfolios would be desirable.

Researchers have proposed methods using both transaction data and data of asset characteristic to estimate returns of infrequent-traded asset portfolios. For example, Case and Quigley (1991), Case *et al* (1992), Clapp and Giaccotto (1992), "hedonic repeated measures" method (HRM) by Shiller (1993b), and "distance-weighted repeat-sales" procedure (DWRS) by Goetzmann and Spiegel (1997). The primary methodological advantage of using both transaction and characteristic data lies in its ability to explore stronger correlation between price paths of assets with more similar characteristics. Limitations of these methods are that they may not supply return estimators for arbitrary-weighted portfolios, and their estimators may not have natural interpretations, such as being arithmetic means of individual asset returns.

In this paper I propose a model for the estimation of time series returns of illiquid asset portfolios. My model has four major advantages. First, it is very powerful: it supplies estimators of any arbitrary-weighted portfolio returns, including equal-weighted, price-weighted, and value-weighted portfolios, which few methods claim to be able to do. Second, the model is meaningful: all estimators are cross-sectional arithmetic averages of individual asset returns, so they strictly correspond to portfolio returns and no correction is needed, which is an important improvement over the currently broadly used RSR method. Third, the model is general: it accommodates methods proposed in previous research, such as Shiller (1991) and Goetzmann and Peng (2000). Fourth, the model is very flexible and easy to extend. It can estimate the portfolio returns with or without asset characteristic data, while the estimators are more efficient if both price data and characteristic data are available. The model is able to correct for the sampling bias problem that transaction takes place more likely on over-valued assets, and supply more accurate estimators.

I present the model and discuss model estimation in three scenarios: when asset characteristics are not observable, when they are observable, and when transactions take place more likely on over-valued asset. For illustrative purposes, I use an extremely small data set to estimate the returns of the equal-weighted portfolio, the price-weighted portfolio and the value-weighted portfolio in the basic scenario that asset characteristics are not observable. To test the performance of the model, I use actual financial data to do simulations. I construct infrequent-traded data set by randomly drawing daily prices for the Dow Jones Industrials Index (DJII) stocks over a three-month period, and estimate the actual daily returns of the DJII over that period. The simulations show that the model supplies more accurate estimators than a simple benchmark method.

The paper is organized as follows. Next section presents the model. Then I discuss model estimation when asset characteristics are not observable and illustrate the estimators with the extremely small data set. I also show that my model accommodates estimators proposed in earlier research. After that, I discuss the model estimation in other two scenarios: when asset characteristics are observable, and when transactions take place more likely on over-valued assets. Then I use simulations to test the performance of the model. A final section concludes. In appendix, I present details of estimation algorithms for the three important types of portfolios.

The Model of Return

Asset return

I define capital appreciation component of return of asset i (or a dollar invested in asset i) in time period t, $r_{i,t}$, as the ratio of the price of the asset at the end of time period t over its price at the end of period $t-1$.

$$r_{i,t} \equiv P_{i,t} / P_{i,t-1} .$$

I assume that the $r_{i,t}$ is determined as following:

$$r_{i,t} = E(r_{i,t} \mid m_t, c_{i,t}) \varepsilon_{i,t} . \tag{1}$$

Within equation (1), the term m_t is a vector of common factors that affect all assets' returns in time period t. The term $c_{i,t}$ is a vector of characteristics of asset i in time t. The error term $\varepsilon_{i,t}$ captures asset-specific events that are responsible for unexpected change of price. I assume that the conditional expectation of the error equals one, $E(\varepsilon_{i,t} | m_t, c_{i,t}) = 1$, and all $\varepsilon_{i,t}$ are independent from each other.

Based on these assumptions, common factors and asset characteristics jointly determine an asset's expected return in a time period. Common factors could be macroeconomic variables like the risk-free interest rate, inflation rate, unemployment rate, and so on and so forth. They could even be investor sentiment that has market-wide effects on asset returns. Four houses, asset characteristics could be hedonic variables such as location or square-feet of floor space. For equities, asset characteristics could be P/E ratio, B/M ratio, capitalization and so on. For bonds, they could be bond maturity, rating, coupon rate or other characteristics. The assumptions about asset return process are consistent with the fact that assets with different characteristics may have different return processes.

Portfolio return

The return of a portfolio in time period *t*, r_t, equals to the ratio of the value of the portfolio at the end of time period *t* over its value at the end of time period $t-1$.

$$r_t \equiv \frac{\sum_i (n_{i,t} P_{i,t})}{\sum_i (n_{i,t} P_{i,t-1})}.$$

The $n_{i,t}$ term is the number of units of asset i in the portfolio since the end of time period $t-1$. A more conventional but equivalent way to define the portfolio return is

$$r_t \equiv \sum_i (w_{i,t} \frac{P_{i,t}}{P_{i,t-1}}) = \sum_i (w_{i,t} r_{i,t}) .$$

The $w_{i,t}$ term is called the *weight* of asset i in the portfolio for time period *t*. Note that

$$w_{i,t} = \frac{n_{i,t} P_{i,t-1}}{\sum_j n_{j,t} P_{j,t-1}} \text{ and } \sum_i w_{i,t} = 1 .$$

Thus $w_{i,t}$ is the proportion of the portfolio value that is stored in asset i at the end of time period $t-1$. Most frequently observed portfolios are equal-weighted, price-weighted, and value-weighted. I denote their returns in time period *t* as r_t^e, r_t^p, and r_t^v respectively.

Return of a dollar and the expected return of a portfolio

A portfolio consists of units of value, say dollars, that are invested in different assets. The return of each dollar equals to return of the asset in which this dollar is invested. Thus one can say that each dollar in a portfolio has its characteristics. I define the probability distribution of the characteristics for a randomly selected dollar, say dollar l, from a portfolio in time period *t* as $f_t(c) = \Pr(c_{l,t} = c)$. The probability of the randomly selected dollar having specific characteristics equals to the weight of the asset that has these characteristics in the portfolio. The probability distribution is different if the weights of assets are different. For example, let us talk about a portfolio consisting of two bonds. One bond is risky and the other is risk-free. For an equal-weighted portfolio, the probability for a random dollar from the portfolio being risk-free is 0.5. For a price-weighted portfolio, if the risk-free bond is more expensive, the probability for a random dollar from the portfolio being risk-free is larger than 0.5.

I denote the expected return of a random dollar in a portfolio in time period *t* given common factors as γ_t. Thus

$$\gamma_t \equiv E(r_{l,t} | m_t) = \sum_c E(r_{l,t} | m_t, c) f(c) = \sum_i E(r_{i,t} | m_t, c) w_{i,t} . \tag{2}$$

It is clear that the γ_t equals the expected return of the portfolio in time period *t*. Since different portfolios give different weights to assets, the expected returns of different portfolios may be different.

One can always write the return of a dollar in the portfolio as

$$r_{l,t} = \gamma_t \eta_{l,t} \varepsilon_{l,t} \tag{3}$$

with $\eta_{l,t} \equiv E(r_{l,t} | m_t, c_{l,t}) / E(r_{l,t} | m_t)$, which is a function of the dollar's characteristics. It is obvious that $E(\eta_{l,t} | \gamma_t) = 1$.

The equation (3) is important because it connects the return of a dollar with the expected return of a specific-weighted portfolio. It also has an intuitive interpretation. The return of a dollar in a portfolio consists of three parts. The first part is the expected return of the portfolio. The second part is the expected deviation from the portfolio's expected return because of this dollar's characteristics. The third part is a random shock.

Model Estimation I: When Characteristics Are Not Observable

GMM estimators

In this scenario, the data consist of prices of transactions and the time periods when transactions took place. I assume that a transaction always take place at the end of a time period. I assume that there are N repeat-sale observations in total. Each observation consists of the first transaction price, the time period of the first transaction, the second transaction price, and the time period of second transaction. For observation n, I denote the first transaction price, the purchase or buy price, as B_n, the second transaction price, the sale price, as S_n, the time period of first transaction as b_n, that of the second transaction as s_n. The holding interval of observation n, denoted as H_n, consists of all time periods later than b_n and not later than s_n, i.e., $H_n \equiv \{t | b_n + 1 \le t \le s_n\}$. The length of H_n is denoted as T_n, so $T_n \equiv s_n - b_n$. I assume that there are $T+1$ time periods in the sample, numbered from 0 to T. For a time period t, I denote the set of all observations that have this time period in their holding intervals as $O_t \equiv \{n | t \in H_n\}$. I define the size of O_t, i.e. the number of observations that belong to O_t, as N_t.

I define

$$y_n \equiv S_n / B_n .$$

Then,

$$y_n = \prod_{t\in H_n} r_{n,t} = \prod_{t\in H_n} \gamma_t \prod_{t\in H_n} (\eta_{n,t}\varepsilon_{n,t})$$

$$y_n / \prod_{t\in H_n} \gamma_t = \prod_{t\in H_n} (\eta_{n,t}\varepsilon_{n,t}) .$$

Since $E\left[\prod_{t\in H_n} (\eta_{n,t}\varepsilon_{n,t})\right] = 1$, the moment conditions $E(y_n / \prod_{t\in H_n} \gamma_t - 1) = 0$ for $n = 1,\ldots,N,$ yield a parameter defining mapping under suitable regularity conditions. I assume that regularity conditions are satisfied. I assume that the expectations exist and there is one and only one parameter vector $\gamma \equiv (\gamma_1,\ldots,\gamma_T)$ that satisfies the moment conditions and it is identified.

Sample counterparts to the moment conditions define GMM estimator of γ:

$$\sum_{n\in O_t} \tau_n \left[w_{n,t} \left(y_n / \prod_{s\in H_n} \hat{\gamma}_s - 1 \right) \right] = 0 \text{, for } t = 1,\ldots,T . \qquad (4)$$

Where $w_{n,t}$ indicates how many dollar samples the repeat-sale observation *n* supplies, so it equals the weight in the portfolio of the asset that corresponds to the repeat-sale observation *n*. A repeat-sale observation supplies more dollar samples of the portfolio if its asset has heavier weight in the portfolio. Thus by choosing different $w_{n,t}$, one could use the same data of repeat-sale observations to estimate returns of different portfolios. For example, each repeat-sale observation supplies the same amount of dollar sample for the equal-weighted portfolio, so one should let $w_{i,t}$ equal to each other to estimate the returns of the equal-weighted portfolio. At the same time, a repeat-sale observation supplies amount of dollar sample proportional to its asset price for the price-weighted portfolio, so one should let $w_{i,t}$ be proportional to its asset price to estimate the price-weighted portfolio.

The τ_n term in equation (4) is proportional to the reciprocal variance of observation *n*. Thus it down-weights the observation with higher variance to improve the efficiency of the estimator. The longer is an observation's holding interval; more noises tend to be contained in the observation. Then the variance of observation *n* is an increasing function of the length of its holding interval, and the τ_n term is a decreasing function of it. I define that $\tau_n = h(T_n)$ and scale the function by letting $h(1) = 1$, which gives observations held for one time period one unit of information-weight.

Rearranging the estimator-defining equations, I can express the estimator of portfolio return in time period *t* as

$$\hat{\gamma}_t = \sum_{n \in O_t} \varpi_{n,t} \left(y_n / \prod_{\{s / s \in H_n, s \neq t\}} \hat{\gamma}_s \right), \text{ with } \varpi_{n,t} = \frac{w_{n,t} \tau_n}{\sum_{n \in O_t} (w_{n,t} \tau_n)}. \tag{5}$$

Since the y_n term is a compounded return, the $y_n / \prod_{\{s / s \in H_n, s \neq t\}} \hat{\gamma}_s$ term is the compound return with all expected returns in time periods during its holding interval other than time period *t* subtracted. Thus it is a proxy of an asset's return in time period *t*. Thus, obviously the estimator $\hat{\gamma}_t$ is an arithmetic average of returns (or proxies of returns) of individual assets.

A very nice property of the GMM estimators is that: when all assets were frequently traded, all τ_n are the same for all observations, then

$$\hat{\gamma}_t = \sum_{n \in O_t} w_{n,t} y_n = \sum_{n \in O_t} w_{n,t} (S_n / B_n).$$

Thus, the estimators are exactly equal to the portfolio returns.

Equal-weighted, price-weighted, and value-weighted portfolios

The estimators of the equal-weighted portfolio returns can be easily obtained if one let $w_{n,t} = 1/N_t$, where N_t is defined earlier, as the number of observations that includes time period *t* in their holding intervals. The estimator-defining equations are

$$\hat{\gamma}_t^e = \frac{1}{\sum_{n\in O_t} \tau_n} \sum_{n\in O_t} \tau_n \left(y_n / \prod_{\{s/s\in H_n, s\neq t\}} \hat{\gamma}_s^e \right), \text{ for } t = 1,...,T . \tag{6}$$

For the price-weighted portfolio, an asset's weight in time period *t* is proportional to its price at the end of time period $t-1$. For illiquid assets, prices are not observable for all time periods, nor are corresponding weights. However, the model itself supplies estimators for all unobserved prices. For the asset corresponding to the repeat-sale observation *n*, an estimator for its price at the end of time period *t*-1 is $B_n \prod_{s=b_n+1}^{t-1} \hat{\gamma}_s^p$. Then the expected weight of asset *n* in time period *t* is

$$w_{n,t} = \frac{B_n \prod_{s=b_n+1}^{t-1} \hat{\gamma}_s^p}{\sum_{n\in O_t} \left(B_n \prod_{s=b_n+1}^{t-1} \hat{\gamma}_s^p \right)} .$$

With the estimated weights, the estimator-defining equations for the price-weighted portfolio is

$$\sum_{n\in O_t}\tau_n\left(B_n\prod_{s=b_n+1}^{t-1}\hat{\gamma}_s^p\right)=\sum_{n\in O_t}\tau_n\left(S_n/\prod_{s=t}^{s_n}\hat{\gamma}_s^p\right),\text{ for } t=1,\ldots,T\ . \tag{7}$$

The estimator of the price-weighted portfolio return in time period *t* is

$$\hat{\gamma}_t^p=\frac{\sum_{n\in O_t}\tau_n\left(S_n/\prod_{s=t+1}^{s_n}\hat{\gamma}_s^p\right)}{\sum_{n\in O_t}\tau_n\left(B_n\prod_{s=b_n}^{t-1}\hat{\gamma}_s^p\right)}. \tag{8}$$

If there is no infrequent-trading problem, the estimator will reduce to $\sum_{i=1}^{N}P_{i,t}/\sum_{i=1}^{N}P_{i,t-1}$, which is exactly the return of the price-weighted portfolio.

To estimate the value-weighted portfolio returns, I assume that all assets' share outstanding are observable for all time periods. I denote the share outstanding in time period *t* for the asset that corresponds to repeat-sale observation *n* as $e_{n,t}$. An estimator of the weight of the observation *n* in the portfolio in time period *t* is

$$w_{n,t}=\frac{e_{n,t-1}B_n\prod_{s=b_n+1}^{t-1}\hat{\gamma}_s^v}{\sum_{n\in O_t}\left(e_{n,t-1}B_n\prod_{s=b_n+1}^{t-1}\hat{\gamma}_s^v\right)}.$$

The estimator-defining equations are

$$\sum_{n\in O_t}\tau_n\left(e_{n,t-1}B_n\prod_{s=b_n+1}^{t-1}\hat{\gamma}_s\right)=\sum_{n\in O_t}\tau_n\left(e_{n,t-1}S_n/\prod_{s=t}^{s_n}\hat{\gamma}_s\right),\text{ for } t=1,\ldots,T\ .$$

Then the estimator of value-weighted portfolio return in time period *t* is

$$\hat{\gamma}_t = \frac{\sum_{n \in O_t} \tau_n \left(e_{n,t-1} S_n / \prod_{s=t+1}^{s_n} \hat{\gamma}_s \right)}{\sum_{n \in O_t} \tau_n \left(e_{n,t-1} B_n \prod_{s=b_n}^{t-1} \hat{\gamma}_s \right)}. \tag{9}$$

Without infrequent-trading problem, the estimator reduces to $\sum_{i=1}^{N} e_{i,t-1} P_{i,t} / \sum_{i=1}^{N} e_{i,t-1} P_{i,t-1}$, which is exactly the return of the value-weighted portfolio.

An Illustration of Estimators

For an example of the estimators, let us consider a very small data set consisting of two assets and three time periods numbered from 0 to 2. The first asset was sold at the end of each time period, while the second one was sold only at the end of the time period 0 and time period 2. I denote prices of the first asset as $P_{1,0}, P_{1,1}, P_{1,2}$, prices of the second asset as $P_{2,0}, P_{2,2}$. Thus there are three repeat-sale observations, the first two are for the first asset and the last one is for the second asset. Then

$$Y = \begin{bmatrix} P_{1,1} / P_{1,0} \\ P_{1,2} / P_{1,1} \\ P_{2,2} / P_{2,0} \end{bmatrix}.$$

Then in this example, the estimators of equal-weighted portfolio returns in time period 1 and 2 are

$$\hat{\gamma}_1^e = \frac{1}{2}\left(\frac{P_{1,1}}{P_{1,0}} + \tau_3 \frac{P_{2,2}}{P_{2,0}} \frac{1}{\hat{\gamma}_2}\right), \quad \hat{\gamma}_2^e = \frac{1}{2}\left(\frac{P_{1,2}}{P_{1,1}} + \tau_3 \frac{P_{2,2}}{P_{2,0}} \frac{1}{\hat{\gamma}_1}\right).$$

The estimators of price-weighted portfolio returns are

$$\hat{\gamma}_1^p = \frac{P_{1,1} + \tau_3 P_{2,2} / \hat{\gamma}_2}{P_{1,0} + \tau_3 P_{2,0}}, \ \hat{\gamma}_2^p = \frac{P_{1,2} + \tau_3 P_{2,2}}{P_{1,1} + \tau_3 P_{2,0} \hat{\gamma}_1}.$$

The estimators of value-weighted portfolio returns are

$$\hat{\gamma}_1^v = \frac{e_{1,0} P_{1,1} + \tau_3 e_{2,0} P_{2,2} / \hat{\gamma}_2}{e_{1,0} P_{1,0} + \tau_3 e_{2,0} P_{2,0}}, \ \hat{\gamma}_2^v = \frac{e_{1,1} P_{1,2} + \tau_3 e_{2,1} P_{2,2}}{e_{1,1} P_{1,1} + \tau_3 e_{2,1} P_{2,0} \hat{\gamma}_1}.$$

It is obvious that these estimators strictly correspond to returns of respective portfolios.

Accommodating Earlier Research

The GMM estimators have many variants because there are many possible ways to define the τ_n. One special way is let $\tau_n = 1/T_n$, the reciprocal holding length. In this case, the equal-weighted GMM estimator is equivalent to the arithmetic-average equal-weighted estimator by Goetzmann and Peng (2000); the price-weighted GMM estimator is equivalent to the instrumental-variables arithmetic estimators of price-weighted portfolios by Shiller (1991).

Estimators with information-weight $\tau_n = 1/T_n$ have a nice finite sample property. As an example, let us consider all assets were traded in time period t_1 and t_2, but were infrequently traded between these two periods. In this situation, the GMM method estimates the value change of the price-weighted or value-weighted portfolio from time period t_1 to t_2 without error, i.e. the estimated change of portfolio value is exactly equal to the actual one. Illustrating this with the small data set used in last sub-section, the estimated compound return of the price-weighted portfolio for two time periods is exactly the real return,

$$\hat{\gamma}_1\hat{\gamma}_2 = \frac{P_{1,2} + P_{2,2}}{P_{1,0} + P_{2,0}}.$$

Model Estimation II: When Characteristics Are Observable

Since I assume that the characteristics systematically affect a dollar's expected return, dollars invested in different assets generally have different return processes. However, more similar two dollars' characteristics are, the correlation between the expected return processes of these two dollars is higher. Thus using both transaction data and characteristic data may improve the efficiency of return estimators. For the purpose of simplicity, I assume asset characteristics remain the same within each repeat-sale observation's holding interval. Then for repeat-sale observation *n*, the systematic deviations of its returns from the expected portfolio returns, i.e. $\eta_{n,t}$ for $t \in T_n$, are constant over the whole holding interval. Then I simplify the notation to η_n.

One can estimate the model with a three-stage procedure. In the first stage, we estimate the model as if asset characteristics were not observable, which gives us first stage GMM estimators of portfolio returns $\left\{\hat{\gamma}_t^1\right\}_{t=1}^T$.

In the second stage, we use estimated residuals from stage 1 as estimation of deviations of assets' returns from expected portfolio returns,

$$\hat{\eta}_n = \left(y_n / \prod_{s \in H_n} \hat{\gamma}_s^1 \right)^{\frac{1}{T_n}} .$$

I assume $\eta_n = g(c_n)$, and we've gotten $\hat{\eta}_n$ and we observe c_n, we can estimate the functional form of $g(.)$. Both parametric and non-parametric approaches can be used to run regression

$$\hat{\eta}_n = g(c_n)\varepsilon_n \text{ , with } E[\varepsilon_n | g(c_n)] = 1 . \tag{10}$$

I denote the expected deviation from expected portfolio return as $\tilde{\eta}_n = \hat{g}(c_n)$.

In the third stage, we estimate the portfolio returns with $\tilde{\eta}_n$ plugged in.

$$\sum_{n \in O_t} w_{n,t} \left[\tau_n \left(y_n / (\prod_{s \in H_n} \hat{\gamma}_s \tilde{\eta}_n) - 1 \right) \right] = 0 \text{, for } t = 1,...,T .$$

This three-stage approach gives us estimators

$$\hat{\gamma}_t = \sum_{n \in O_t} \omega_{n,t} \left(\frac{y_n}{\tilde{\eta}_n^{T_n}} \frac{1}{\prod_{\{s / s \in H_n, s \neq t\}} \hat{\gamma}_s} \right) \text{, with } \varpi_{n,t} = \frac{w_{n,t}\tau_n}{\sum_{n \in O_t} (\varpi_{n,t}\tau_n)} . \tag{11}$$

The observable asset characteristics help us to subtract $\tilde{\mu}_n^{T_n}$, the expected deviations of individual asset returns from expected portfolio returns, from the observed compound return y_n, which improves the efficiency of portfolio return estimators.

This procedure also supplies estimators of expected returns for a sub-set of assets with specific characteristics. For example, in real estate research, this procedure is able to estimate housing index for not only a broad metropolitan but also any specific neighborhood within. The estimators would be the products of the expected portfolio returns and the expected deviations from it.

Model Estimation III: When Samples Are Biased

Sometimes the probability for a transaction to take place may be higher when the asset's price is higher than its expected market value, i.e. $\varepsilon_{i,t} > 0$. For example, when real estate prices fall, number of transactions may go down and the traded houses are more likely those that have lost unusual little value. Thus the samples are biased and over-represent transactions with positive price shocks. Ignoring this problem may result in upward-biased estimators for portfolio returns.

To deal with this problem when it exists, I define a new variable $O_{n,t}$ that is equal to 1 if asset *n* transact at the end of time period *t*, and 0 otherwise. As an example, for repeat-sale observation *n*, all $O_{n,t}$ for $b_n < t < s_n$ are equal to 0, and O_{n,s_n} is equal to 1. I assume that the probability for $O_{n,t}$ to be 1, i.e. the probability for a transaction to take place, is a monotonous increasing function of pricing error $\varepsilon_{n,t}$. Knowing this function and the probability density function of

$\varepsilon_{n,t}$, by Bayes rule, one can solve out $E(\varepsilon_{i,t} | O_{i,t})$, the expectation of pricing error $\varepsilon_{n,t}$ conditioned on observed $O_{n,t}$. I denote $\rho_{i,t} \equiv E(\varepsilon_{i,t} | O_{i,t})$. Then the pricing error is

$$\varepsilon_{i,t} = \rho_{i,t} \zeta_{i,t} \text{ with } E(\zeta_{i,t} | O_{i,t}) = 1.$$

As long as the probability for a transaction to take place is a increasing function of pricing error, one can easily tell that the conditional expectation of pricing error, $\rho_{i,t}$, is larger than 1 when transaction takes place, and smaller than 1 otherwise.

Now the return of a dollar in repeat-sale observation *n* is

$$r_{n,t} = \gamma_t \eta_{n,t} \varepsilon_{n,t} = \gamma_t \eta_{n,t} \rho_{n,t} \zeta_{n,t}. \tag{12}$$

Then for the repeat sale observation *n*,

$$y_n = \prod_{t \in H_n} r_{n,t} = \prod_{t \in H_n} \gamma_t \prod_{t \in H_n} (\eta_{n,t} \rho_{n,t} \zeta_{n,t}).$$

Thus the moment conditions are

$$E(y_n / \prod_{t \in H_n} (\gamma_t \eta_{n,t} \rho_{n,t}) - 1) = 0 \text{ for } n = 1,\ldots,N.$$

For simplicity, let us assume that asset characteristics are not observable and we already have estimators for the conditional expectations of pricing errors, then the estimator of portfolio return for time period *t* is

$$\hat{\gamma}_t = \sum_{n \in O_t} \varpi_{n,t} \left(\frac{y_n}{\prod_{s \in H_n} \hat{\rho}_{n,s} \prod_{\{s / s \in H_n, s \neq t\}} \hat{\gamma}_s} \right), \text{with } \varpi_{n,t} = \frac{w_{n,t} \tau_n}{\sum_{n \in O_t} (\varpi_{n,t} \tau_n)}.$$

Since all $\rho_{n,t}$ are smaller than 1 except for the last time period of holding interval, by divided by $\prod_{t \in H_n} \hat{\rho}_{n,t}$, the compound return y_n is discounted more for observations with shorter holding intervals. This is sensible because I assume that for assets that are traded more frequently, higher is the probability for their pricing errors to be positive. Thus discounting their returns more would correct the sample bias problem and improve the efficiency of estimation.

Simulations

The principal goal of the simulations is to compare the performance of estimators proposed in this paper with alternative methods. Since Goetzmann and Peng (2000) have tested the accuracy of the equal-weighted GMM estimators, and the price-weighted and value-weighted estimators are extremely similar, here I test the accuracy of the price-weighted GMM estimators and assume characteristics are not observable.

Simulation Procedure

The basic approach is to select N prices randomly (according to uniform distribution) and without replacement from the Dow Jones Industrial daily stock prices over a three-month interval to construct an infrequent-traded data set. I then estimate the time series of Dow Jones Industrial Index (DJII) daily returns over the interval with different methods. Next, I calculate some statistics that measure the accuracy of each method, such as the mean squared difference

(MSE) between the actual and estimated series and R^2 resulting from a regression of actual Dow Jones Industrial Index daily returns upon the estimated series. I repeat this procedure many times for the *N*. Since the accuracy of a method may be related to the density of observable transactions, I do simulations with *N* as a relatively small number and a relatively large number.

I arbitrarily choose the three months as September to December 1999. There are 85 trading days (so 84 daily returns for DJII since the first day is the base period) and totally 2550 daily prices for 30 Dow Jones Industrial companies. I choose *N*, the number of prices I randomly draw from all 2550 prices, as 800 and 1500 respectively to represent scenarios when observable prices are sparse and intensive. I repeat the procedure of price drawing and return estimating 1,000 times for each scenario. I estimate time series returns of the price-weighted index, i.e. the actual DJII, with two methods. One is the GMM method proposed in this paper with information-weight defined as $1/T_n$. The second one is a simple benchmark method, which estimates a daily DJII return by averaging all available individual daily returns for that day (weighted by prices).

Simulation Results

Table 1 reports the geometric mean, standard deviation, and autocorrelation of actual and estimated DJII daily returns for *N*=800. For estimated returns, reported statistics are averages of results of 1000 simulations. The geometric mean of the actual DJII daily returns is 0.0604%, the average geometric mean of the GMM estimators are 0.0612%, while that of the simple method is 0.044%. The GMM method obviously supplies more accurate estimate for the average long-run performance. The average standard deviation of the GMM estimators is closer to the actual one than that of the simple estimators. However, the GMM estimators have much higher negative autocorrelation. Table 1 also reports the averaged MSE and R^2 for different methods. The average MSE for the GMM method is about 30% lower than that of the simple method. The average R^2 of the GMM method is higher than that of the simple method by about 9%.

Figure 1 reveals the dynamics of MSE of different methods through all 84 days. It is clear that for almost all days, the MSE of the GMM is lower than that of the simple method. So the GMM method in principal supplies better estimator for each single time period's return than the simple method.

Table 2 reports the same statistics for N=1500, the scenario that observable prices are more intensive. Since more transactions are observed, estimators of both methods are more accurate, in terms of average geometric mean, average standard deviation, average autocorrelation, MSE, and R^2. The GMM estimators are still significantly more accurate than the simple estimators. For example, the MSE of the GMM is still about 30% lower than the MSE of the simple method. The R^2 of the GMM is higher than that of the simple method by about 7.5%. The negative autocorrelation introduced by the GMM method is much lower now. It decreases from –28.29% to –11.85%. Figure 2 shows that GMM estimators are superior to the simple estimators for almost all time periods.

In conclusion, the simulations confirm that the GMM method is superior to the simple method in estimating both the average long run performance and returns in any single time period. Though the GMM method introduces negative autocorrelation, the magnitude is much less when price data become more intensive.

Conclusions

This paper proposes a model to measure values of illiquid asset portfolios. This model is meaningful and its estimators have natural interpretations. All the estimators are arithmetic averages of individual asset returns (or their proxies) and strictly correspond to portfolio returns, which is an important improvement over the currently broadly used RSR method. This model is very powerful: it supplies estimators for returns of any arbitrary-weighted portfolio, including equal-weighted, price-weighted, and value-weighted portfolio, which few models claim to be able to do. This model is general: it accommodates some important estimators from earlier research such as the arithmetic-RSR by Shiller (1991) and Goetzmann and Peng (2000). Also, this model is flexible and very easy to extend. It can estimate the portfolio returns with or without asset characteristic data, while the estimators are more efficient if both price data and characteristic data are available. The model may be able to supply more accurate estimators by correcting the sample bias problem that transactions may take place more likely on over-valued assets.

I use simulations to test the accuracy of the estimators proposed in this paper, specifically the estimator of price-weighted portfolio returns when characteristic data not available. The data I use are actual financial data: 2,550 daily prices of Dow Jones Industrial Index stocks over September 1999 to December 1999. I randomly draw prices from them to create artificial infrequent transaction data set and estimate the actual DJII daily returns. I use a simple intuitive method as the benchmark to test the performance of my model. I run 1,000 simulations for two scenarios: one with sparse data, in which I pick 800 prices out from 2,550, and the other with intensive data, in which I pick out 1,500 prices out from 2,550. Simulations show that the estimators proposed in this paper are more accurate than the benchmark estimators are. My estimators have much smaller MSE and larger R^2 than the benchmark method. One drawback of my estimators is that they introduce in negative autocorrelation, but this problem is eased when more transactions are observed.

Appendix: Estimation Algorithms

To estimate the returns of equal-weighted portfolios, I define matrix *X*, *Y*, *W* and *I* as following. The *X* is a *N* by *T* dummy matrix. Its rows correspond to repeat-sale observations, and columns correspond to time periods. For row *n*, the first nonzero dummy appears in the position that corresponds to the time period $b_n + 1$, the time period immediately after the first sale of *n*th observation, and the last nonzero dummy appears in the position corresponding to s_n, the time period of the second sale. Between these two nonzero dummies are all nonzero dummies, while other elements in this row are zero. As an example, if an asset was purchased at the end of time period 2 and sold at the end of time period 4, and *T*=*5*, its corresponding row in *X* is $(0,0,1,1,0)$. The *Y* is defined as a *N* by 1 vector whose *n*th element is y_n. The *W* is a *N* by *N* diagonal matrix whose *n*th element is τ_n. The *I* is a *N* by 1 vector of 1.

Now, the estimator-defining equations for the equal-weighted portfolio can be written in matrix form as

$$X'WI = X'W \exp[\log(Y) - X \log(\hat{\gamma})],$$

$$\text{or } X'W\{\exp[\log(Y) - X \log(\hat{\gamma})] - I\} = 0.$$

It is clear that there are *T* equations for *T* estimators. Though these equations are not linear, solving them with searching techniques may not be very difficult.

To estimate the returns of price-weighted and value-weighted portfolios, I define a T by 1 vector β whose tth element is a reciprocal price index for time t,

$$\beta_t \equiv 1/\prod_{s=}^{t} \hat{\gamma}_s \ .$$

It is obvious that knowing β is equivalent to knowing $\hat{\gamma}$. I also define a N by $T+1$ matrix Z. Its rows correspond to repeat-sale observations, and columns correspond to time periods but start with time period 0. For row n, its b_nth element is equal to $-B_n$, and its s_nth element is equal to S_n, all other elements are 0. For example, for the data set used in earlier section to illustrate the estimators, the Z is

$$Z = \begin{bmatrix} -P_{1,0} & P_{1,1} & 0 \\ 0 & -P_{1,1} & P_{1,2} \\ -P_{2,0} & 0 & P_{2,2} \end{bmatrix}.$$

With matrix X, Z, and vector β, the estimator-defining equations for price-weighted portfolio returns can be written in matrix form as

$$X'WZ\begin{bmatrix} 1 \\ \beta \end{bmatrix} = 0 \ .$$

They are linear equations and it is trivial to solve out β. Once β is known, $\hat{\gamma}$ is known. For example, $\hat{\gamma}_1 = 1/\beta_1$, $\hat{\gamma}_t = \beta_{t-1}/\beta_t$ for $t > 1$.

If the number of shares for each asset remained constant for all time periods, i.e. $e_{n,t} = e_n$ for all t, it is easy to write the estimator-defining equations in matrix form. Let us define a N by N

diagonal matrix *E* whose *n*th diagonal element is equal to e_n and other elements are zero. The estimator-defining equations, in terms of β, are defined by

$$X'EWZ\begin{bmatrix}1\\ \beta\end{bmatrix}=0\,.$$

If the numbers of shares for each asset are generally not the same for different time periods, we need to stack matrixes as following. Let us define a *T* by *NT* matrix

$$A\equiv\begin{bmatrix}X'_{(1,.)} & 0 & . & 0\\ 0 & X'_{(2,.)} & . & 0\\ . & . & . & .\\ 0 & 0 & . & X'_{(T,.)}\end{bmatrix}$$

where $X'_{(t,.)}$ is equal to the tth row of matrix X'; a *NT* by *NT* diagonal matrix

$$B\equiv\begin{bmatrix}E_0W & 0 & . & 0\\ 0 & E_1W & . & 0\\ . & . & . & .\\ 0 & 0 & . & E_{T-1}W\end{bmatrix}\text{ where }E_t\equiv\begin{bmatrix}e_{1,t} & 0 & . & 0\\ 0 & e_{2,t} & . & 0\\ . & . & . & .\\ 0 & 0 & . & e_{N,t}\end{bmatrix};$$

a *NT* by $N(T+1)$ block-diagonal matrix

$$C\equiv\begin{bmatrix}Z & 0 & . & 0\\ 0 & Z & . & 0\\ . & . & . & .\\ 0 & 0 & . & Z\end{bmatrix}$$ where *Z* is the *N* by $T+1$ matrix defined earlier;

and a $N(T+1)$ by 1 vector

$$D\equiv\begin{bmatrix}[1 \quad \beta'] & [1 \quad \beta'] & . & [1 \quad \beta']\end{bmatrix}'\,.$$

Then the estimator-defining equations, in terms of β, can be written as $ABCD=0$.

I would like to thank Donald Andrews, Stefano Athanasoulis, William N. Goetzmann, Roger Ibbotson, Peter C B Phillips, Rainer Schulz, Robert J. Shiller, and Matthew Spiegel for numerous helpful comments and discussions. I thank participants in the workshops at Yale University for useful suggestions. I also thank Yale Economics department and Yale International Center for Finance for research support. All errors are mine alone.

References

Bailey, M.J., Muth, R.F., and H.O. Nourse. 1963. A Regression Method for Real Estimate Price Index Construction. *Journal of the American Statistical Association* 58: 33-942.

Case, B., Pollakowsi, H. O., and S. M. Wachter. 1992. On Choosing among house price index methodologies. *AREUEA journal* February.

Case, B., and J. M. Quigley. 1991. The Dynamics of Real Estate Prices. Review of Economics and Statistics 73: 50-58.

Case, K. E., and R. J. Shiller. 1987. Prices of Single Family Homes since 1970: New Indices for Four Cities. New England Economic Review (September/October): 45-56.

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

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: 357-374.

Goetzmann, W. N. 1992. The Accuracy of Real Estate Indices: Repeat Sale Estimators.

Journal of Real Estate Finance and Economics 5: 5-53.

Goetzmann, W. N., and L. Peng. 2000. The Accuracy of Some Alternative Methods to Repeat Sales Regression: A Simulation. Working Paper, International Center for Finance, Yale University.

Goetzmann, W.N., and M. Spiegel. 1995. Non-temporal Components of Residential Real Estate Appreciation. *Review of Economics and Statistics* (September): 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: 11-31.

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

Shiller, R. J. 1993a. *Macro Markets: Creating Institution for Managing Society's Largest Economic Risks*. Oxford University Press: New York.

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

Webb, C. 1988. A probabilistic Model for Price Levels in Discontinuous Markets. W. Eichhorn, editor. Measurement in Econometrics: Physica-Verla, (1988).

***Table 1.* Comparisons of accuracy of different estimation methods**

This table reports results for *N*=800. Panel 1 reports geometric means, standard deviations, and autocorrelations of daily returns of actual and estimated (averaged over 1,000 simulations for each method) Dow Jones Industrial Index from September 1st 1999 to December 31st 1999. Panel 2 reports the mean squared errors between actual and estimated time series returns, and R^2 resulting from regression of actual Dow Jones Industrial Index daily returns upon estimated series. The two estimation methods are the GMM method proposed in this paper, and a simple method that averages all available one-period individual returns for each time period (weighted by prices at the beginning of the two days).

Statistics of actual and estimated daily return series of the Dow Jones Industrial Index

	Geometric Mean	Std. Deviation	Autocorrelation
Actual DJII	0.0604%	1.0558%	-6.7977%
GMM	0.0612%	1.5838%	-28.2914%
Simple	0.0440%	1.7536%	-4.3354%

MSE and R^2 for each method

	MSE	R^2
GMM	1.1890%	43.65%
Simple	1.6860%	34.61%

Table 2. Comparisons of accuracy of different estimation methods

This table reports results for N=1500. Panel 1 reports geometric means, standard deviations, and autocorrelations of daily returns of actual and estimated (averaged over 1,000 simulations for each method) Dow Jones Industrial Index from September 1st 1999 to December 31st 1999. Panel 2 reports the mean squared errors between actual and estimated time series returns, and R^2 resulting from regression of actual Dow Jones Industrial Index daily returns upon estimated series. The two estimation methods are the GMM method proposed in this paper, and a simple method that averages all available one-period individual returns for each time period (weighted by prices at the beginning of the two days).

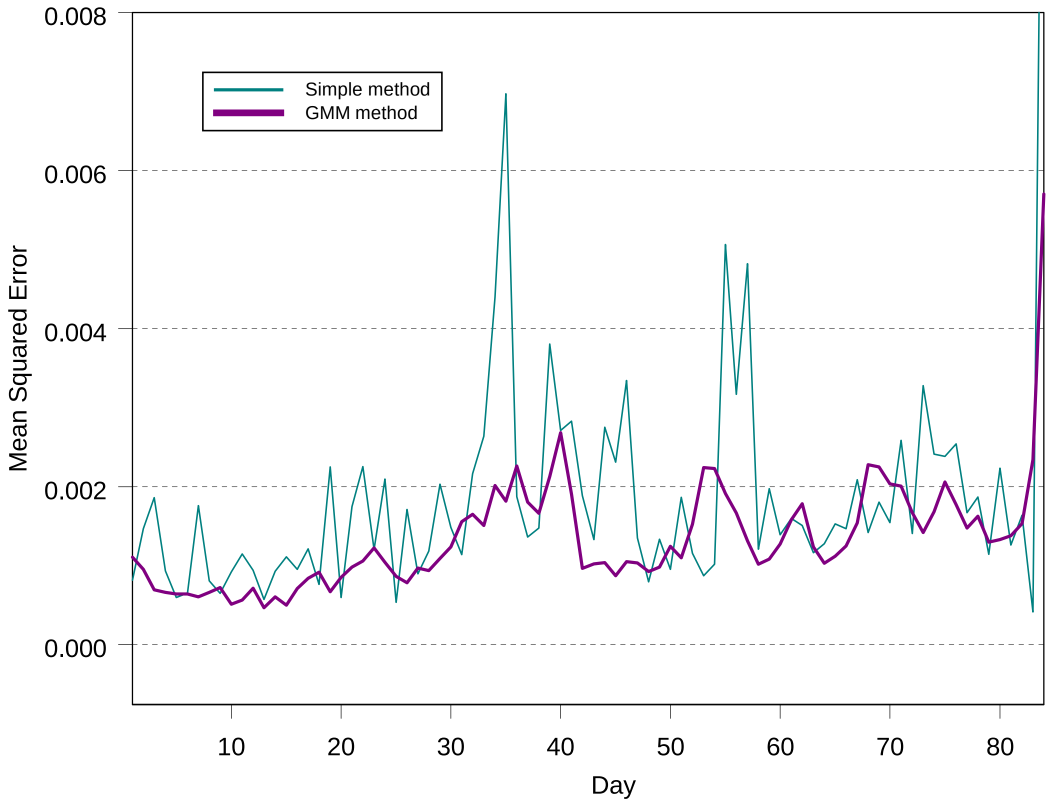
Statistics of actual and estimated daily return series of the Dow Jones Industrial Index

	Geometric Mean	Std. Deviation	Autocorrelation
Actual DJII	0.0604%	1.0558%	-6.7977%
GMM	0.0598%	1.1721%	-11.8534%
Simple	0.0627%	1.1901%	-7.8771%

MSE and R^2 for each method

	MSE	R^2
GMM	0.2253%	80.33%
Simple	0.3250%	72.85%

Figure 1. Mean Squared Errors of GMM and Simple Methods



Based on simulations with N=800

Figure 2. Mean Squared Errors of GMM and Simple Methods