The implied liquidity premium for equities*

ROBERT FERNHOLZ
INTECH
One Palmer Square
Princeton, NJ 08542
bob@enhanced.com

IOANNIS KARATZAS
Departments of Mathematics and Statistics
Columbia University
New York, NY 10027
ik@math.columbia.edu

May 17, 2005

Abstract

Over the long term, the returns on smaller stocks are likely to be higher than the returns on larger stocks. This phenomenon has been called the *size effect*, and a number of explanations have been proposed to account for it. Here we show that the difference in return between the larger and the smaller stocks is likely to be due to a liquidity premium for the smaller stocks, and we estimate the value of this premium using structural parameters for the capital distribution of the U.S. stock market during the 1990s.

Key words: Liquidity premium, ranked market weights, size effect, mean-variance efficiency.
JEL classification: G11, G35, L25, C62.
MSC (2000) classification: 91B28.

*Most of this work was completed in the spring semester of 2002, while the second author was on sabbatical leave at the Cowles Foundation for Research in Economics, Yale University. He wishes to thank the Foundation for its hospitality.

1 Introduction

Banz (1981) and Reinganum (1981) observed that in the U.S. stock market, smaller stocks tend to have higher returns on average than larger stocks have, even when an adjustment is made for risk. This phenomenon has been called the *size effect*, and there have been a number of attempts to justify it a number of ways (see, e.g., Roll (1981)). Here we are interested in an approach considered by Amihud and Mendelson (1986) and Vayanos (2003), who have suggested that the size effect is caused by a *liquidity premium* for smaller stocks. Since smaller stocks are more difficult and more costly to trade than are larger stocks, they must consequently offer higher returns than do larger stocks.

We show that in the context of large, diversified portfolios, the smaller stocks do not seem to exhibit higher risk than the larger stocks, at least over the long term. However, it is well known that smaller stocks are less liquid than larger stocks, so the size effect could be due to a *liquidity premium* on the smaller stocks. How large would this liquidity premium have to be in order to explain the size effect?

In this paper we attempt to answer this question. After introducing some basic definitions and notation in Section 2, we develop a mathematical model for the relative behavior of the larger and smaller stocks in Section 3. In Section 4 we use this model to estimate the liquidity premium that would account for the size effect. Section 5 is on the first-order model for an asymptotically stable market; we show that in the context of this model the liquidity premium of a stock amounts to the stock's entire rate of return.

2 The market model

In this section we introduce the general market model that we shall use in the rest of the paper. This model is consistent with the usual market models of continuous-time mathematical finance found in, e.g., Duffie (1992) or Karatzas and Shreve (1998). The preliminary material of this section is presented in greater detail in Fernholz (2002).

Consider a market consisting of n stocks represented by their price processes $X_1, \dots, X_n$. We assume that there is a single share of each stock, so $X_i(t)$ represents the total capitalization of the i-th company at time t. The price processes evolve according to

$$X_i(t) = X_0^i \exp\left(\int_0^t \gamma_i(s)\, ds + \int_0^t \sum_{\nu=1}^{d} \xi_{i\nu}(s)\, dW_\nu(s) \right), \quad t \in [0, \infty), \tag{2.1}$$

for $i = 1, \dots, n$, with $d \geq n$.. Here $X_0^1, \dots, X_0^n$, are positive constants and $W(t) = (W_1(t), \dots, W_d(t))$, $t \in [0, \infty)$, is a standard d-dimensional Brownian motion defined on a probability space $(\Omega, \mathcal{F}, P)$ and adapted to a given filtration $\{\mathcal{F}_t\}$. The *growth rate processes* $\gamma_i = \{(\gamma_i(t), \mathcal{F}_t), t \in [0, \infty)\}$, $i = 1, \dots, n$, are measurable, adapted, and satisfy $\int_0^T |\gamma_i(t)| dt < \infty$, a.s., for all $T > 0$. For $i = 1, \dots, n$ and $\nu = 1, \dots, d$, the *volatility processes* $\xi_{i\nu} = \{(\xi_{i\nu}(t), \mathcal{F}_t), t \in [0, \infty)\}$ are measurable, adapted, and satisfy:

i) $\int_0^T \xi_{i\nu}^2(t) dt < \infty$, a.s., for all $T > 0$;

ii) $\lim_{t\to\infty} t^{-1} \xi_{i\nu}^2(t) \log\log t = 0$, a.s.;

iii) $\xi_{i1}^2(t) + \cdots + \xi_{id}^2(t) > 0$, $t \in [0, \infty)$, a.s.

We shall assume the stock price processes satisfy:

iv) for all $i \neq j$, the set $\{t : X_i(t) = X_j(t)\}$ has Lebesgue measure zero, a.s.;

v) for all $i < j < k$, the set $\{t : X_i(t) = X_j(t) = X_k(t)\} = \varnothing$, a.s.

Note that condition (*ii*) is rather weak, and allows, e.g., for constant volatilities. This condition is necessary in order for growth rates to be meaningful. For an idea of the type of pathology that can occur in the absence of such a condition, see Fernholz and Karatzas (2005).

From (2.1), we see that the stock price processes satisfy

$$d \log X_i(t) = \gamma_i(t)\, dt + \sum_{\nu=1}^{d} \xi_{i\nu}(t)\, dW_\nu(t), \quad t \in [0, \infty), \tag{2.2}$$

or equivalently,

$$dX_i(t) = X_i(t)\Big(b_i(t)\, dt + \sum_{\nu=1}^{d} \xi_{i\nu}(t)\, dW_\nu(t)\Big), \quad t \in [0, \infty), \tag{2.3}$$

for $i = 1, \ldots, n$, where

$$b_i(t) \triangleq \gamma_i(t) + \frac{1}{2}\sigma_{ii}(t) \tag{2.4}$$

is the *rate of return* of the i-th stock.. In the form (2.2) it is evident that these processes are continuous semimartingales, and we shall frequently refer to them simply as *stocks*. The growth rate of a stock determines its long-term behavior, since for $i = 1, \ldots, n$, conditions (*ii*) and (*iii*) guarantee that

$$\lim_{T\to\infty} \frac{1}{T}\Big(\log X_i(T) - \int_0^T \gamma_i(t)\, dt\Big) = 0, \quad \text{a.s.} \tag{2.5}$$

A proof of this can be found in Fernholz (2002).

The market *covariance process* is the matrix-valued process σ defined by

$$\sigma_{ij}(t) \triangleq \sum_{\nu=1}^{d} \xi_{i\nu}(t)\xi_{j\nu}(t) = \frac{d}{dt}\langle \log X_i, \log X_j \rangle_t, \quad t \in [0, \infty). \tag{2.6}$$

Definition 2.1. A *portfolio* of the stocks $X_1, \ldots, X_n$ in the market is a bounded, measurable, adapted process $\pi : [0, \infty) \times \Omega \to \mathbb{R}^n$ that satisfies $\pi_1(t) + \cdots + \pi_n(t) = 1$, for $t \in [0, \infty)$, a.s.

For each i, the process π_i represents the *proportion*, or *weight*, of X_i in the portfolio; a negative value for $\pi_i(t)$ indicates a short sale of the i-th stock. These portfolios are clearly self-financing. Suppose $Z_\pi(t)$ represents the value of an investment in the portfolio π at time t. Then $Z_\pi(t)$ satisfies

$$\begin{aligned} \frac{dZ_\pi(t)}{Z_\pi(t)} &= \sum_{i=1}^{n} \pi_i(t) \frac{dX_i(t)}{X_i(t)} \\ &= \sum_{i=1}^{n} \pi_i(t)\Big(b_i(t)\, dt + \sum_{\nu=1}^{d} \xi_{i\nu}(t)\, dW_\nu(t)\Big), \quad t \in [0, \infty). \end{aligned} \tag{2.7}$$

This equation and an initial value $Z_\pi(0) > 0$ determine the portfolio value through time, so we shall call the process Z_π the *portfolio value process* for π. Two applications of Itô's rule transform (2.7) into

$$d \log Z_\pi(t) = \sum_{i=1}^{n} \pi_i(t)\, d \log X_i(t) + \gamma_\pi^*(t)\, dt, \quad t \in [0, \infty), \quad \text{a.s.}, \tag{2.8}$$

where

$$\gamma_\pi^*(t) \triangleq \frac{1}{2}\Big(\sum_{i=1}^{n} \pi_i(t)\sigma_{ii}(t) - \sum_{i,j=1}^{n} \pi_i(t)\pi_j(t)\sigma_{ij}(t)\Big), \quad t \in [0, \infty), \tag{2.9}$$

is called the *excess growth rate process* of π. Equation (2.8) is equivalent to

$$d \log Z_\pi(t) = \gamma_\pi(t)\, dt + \sum_{i,\nu=1}^{d} \pi_i(t)\xi_{i\nu}(t)\, dW_\nu(t), \quad t \in [0,\infty), \quad \text{a.s.,} \tag{2.10}$$

where the *portfolio growth rate process* γ_π is defined by

$$\gamma_\pi(t) \triangleq \sum_{i=1}^{n} \pi_i(t)\gamma_i(t) + \gamma_\pi^*(t), \quad t \in [0,\infty).$$

The *portfolio variance process* for π is defined by

$$\sigma_{\pi\pi}(t) \triangleq \sum_{i,j=1}^{n} \pi_i(t)\pi_j(t)\sigma_{ij}(t), \quad t \in [0,\infty), \quad \text{a.s.,}$$

with

$$d\langle \log Z_\pi \rangle_t = \sigma_{\pi\pi}(t)\, dt, \quad \text{a.s.} \tag{2.11}$$

The portfolio μ defined by

$$\mu_i(t) \triangleq \frac{X_i(t)}{X_1(t) + \cdots + X_n(t)}, \quad t \in [0,\infty), \tag{2.12}$$

for $i = 1, \ldots, n$, is called the *market portfolio.* It is straightforward that the weights μ_i of (2.12) satisfy the requirements of Definition 2.1, and that they are continuous semimartingales. With an appropriate initial value, the value Z_μ of the market portfolio satisfies

$$Z_\mu(t) = X_1(t) + \cdots + X_n(t), \quad t \in [0,\infty), \quad \text{a.s.} \tag{2.13}$$

The processes $\mu_1, \ldots, \mu_n$ are called the *market weight processes*, or simply, *market weights.* A market will be called *coherent* if for $i = 1, \ldots, n$ we have

$$\lim_{t\to\infty} t^{-1} \log \mu_i(t) = 0, \quad \text{a.s.}$$

A necessary and sufficient condition for coherence is that

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T \big(\gamma_i(t) - \gamma_j(t)\big) dt = 0, \quad \text{a.s.,} \tag{2.14}$$

for all $1 \le i, j \le n$; see Fernholz (2002), Proposition 2.1.2. We shall assume henceforth that the market is coherent.

Remark. It might be helpful to stress that coherence does not imply $\lim_{T\to\infty} T^{-1} \int_0^T \gamma_\mu^*(t) dt = 0$, a.s. (i.e., that there is no long-term-average gain from diversification). This latter property holds under the condition $\gamma_i(t) = \gamma_j(t)$, a.s., for all $t \ge 0$, $i \ne j$, which is clearly considerably stronger than (2.14); see Proposition 2.2.3 in Fernholz (2002). The Atlas model of section 5.3 of that book provides an example of a coherent market for which $\lim_{T\to\infty} T^{-1} \int_0^T \gamma_\mu^*(t) dt$ exists a.s. and is positive; see also Banner et al. (2005). For another such example, see Section 6 of Fernholz and Karatzas (2005).

The ranks of stocks in the market will be of interest to us, so let us consider the (reverse) order statistics for the stocks $X_1, \ldots, X_n$, represented by

$$X_{(1)}(t) = \max_{1\le i\le n} X_i(t) \ge X_{(2)}(t) \ge \cdots \ge X_{(n)}(t) = \min_{1\le i\le n} X_i(t), \quad t \in [0,\infty).$$

For $t \in [0,\infty)$, let p_t be the random permutation of $\{1,\dots,n\}$ such that for k in $\{1,\dots,n\}$,

$$X_{p_t(k)}(t) = X_{(k)}(t), \quad \text{and} \quad p_t(k) < p_t(k+1) \quad \text{if} \quad X_{(k)}(t) = X_{(k+1)}(t). \tag{2.15}$$

We shall consider not only the ranked stocks, but also the *ranked market weights*, $\mu_{(1)} \geq \dots \geq \mu_{(n)}$. For simplicity, we shall abuse the order statistics notation a bit, and for a general portfolio π, we shall let $\pi_{(k)}$ be the weight corresponding to the kth-ranked stock in the market.

In order to represent the behavior of the ranked weights as continuous semimartingales, it is necessary to recall the definition of a semimartingale local time. For a continuous semimartingale X, the *local time* (at 0) is the process Λ_X defined for $t \in [0,T]$ by

$$\Lambda_X(t) \triangleq \frac{1}{2}\Big(|X(t)| - |X(0)| - \int_0^t \operatorname{sgn}(X(s))\, dX(s)\Big),$$

where $\operatorname{sgn}(x) = 2I_{(0,\infty)}(x) - 1$, with $I_{(0,\infty)}$ the indicator function of $(0,\infty)$. The local time Λ_X is an increasing random function on $[0,\infty)$, with $\Lambda_X(0) = 0$ and flat off the set $\{t : X(t) = 0\}$. For more information about local times, see Karatzas and Shreve (1991). It was shown in Fernholz (2001, 2002) that the ranked weight processes $\mu_{(k)}$ satisfy

$$d\log\mu_{(k)}(t) = \sum_{i=1}^n I_{\{i\}}(p_t(k))\, d\log\mu_i(t) + \frac{1}{2}d\Lambda_{\log\mu_{(k)} - \log\mu_{(k+1)}}(t) - \frac{1}{2}d\Lambda_{\log\mu_{(k-1)} - \log\mu_{(k)}}(t), \tag{2.16}$$

for $t \in [0,\infty)$, a.s. By convention, $\Lambda_{\log\mu_{(0)} - \log\mu_{(1)}}(\cdot) \equiv 0 \equiv \Lambda_{\log\mu_{(n)} - \log\mu_{(n+1)}}(\cdot)$.

3 The size effect

In this section we recall an alternative, structural analysis of the size effect that was proposed in Fernholz (1998, 2001). This alternative analysis differs qualitatively from the empirical work of Banz (1981) and Reinganum (1981), and instead is based on a stochastic analysis of the relative behavior of small-stock portfolios and large-stock portfolios.

Suppose that we fix some integer m in $\{2,\dots,n-1\}$ and define a *large-stock* portfolio ξ with weights

$$\xi_{(k)}(t) \triangleq \begin{cases} \dfrac{\mu_{(k)}(t)}{\mu_{(1)}(t) + \cdots + \mu_{(m)}(t)} & \text{for } k = 1,\dots,m, \\ 0 & \text{for } k = m+1,\dots,n, \end{cases} \tag{3.1}$$

for $t \in [0,\infty)$. Similarly, we define a *small-stock* portfolio η with weights

$$\eta_{(k)}(t) \triangleq \begin{cases} 0 & \text{for } k = 1,\dots,m, \\ \dfrac{\mu_{(k)}(t)}{\mu_{(m+1)}(t) + \cdots + \mu_{(n)}(t)} & \text{for } k = m+1,\dots,n. \end{cases} \tag{3.2}$$

With these two portfolios, it was shown in Fernholz (2002), p.87, that we have

$$\begin{aligned} \log\big(Z_\eta(T)/Z_\xi(T)\big) &= \log\Big(\frac{\mu_{(m+1)}(T) + \cdots + \mu_{(n)}(T)}{\mu_{(1)}(T) + \cdots + \mu_{(m)}(T)}\Big) \\ &\quad + \frac{1}{2}\int_0^T \big(\xi_{(m)}(t) + \eta_{(m+1)}(t)\big)\, d\Lambda_{\log\mu_{(m)} - \log\mu_{(m+1)}}(t), \quad T \in [0,\infty), \end{aligned} \tag{3.3}$$

a.s. If the ratio of the relative capitalizations of the large-stock and small-stock portfolios remains stable over time, as we might expect it would, then the logarithm on the right-hand side of (3.3) will remain bounded over time, but the local time integral in (3.3) is increasing, and hence will eventually dominate. As a result, *over the long term the return on the small-stock portfolio will be greater than the return on the large-stock portfolio.* This phenomenon is called the *size effect*, and (3.3) shows that it is a structural feature that will be present regardless of the relative riskiness of the two portfolios.

The size effect arises from the local-time terms in (2.16), so we are motivated to define the *size-effect process* for the kth-ranked stock to be

$$L_{(k)}(t) \triangleq \frac{1}{2}\big(\Lambda_{\log\mu_{(k-1)}-\log\mu_{(k)}}(t) - \Lambda_{\log\mu_{(k)}-\log\mu_{(k+1)}}(t)\big), \quad t \in [0,\infty), \tag{3.4}$$

for $k = 1, \ldots, n$ (cf. Fernholz (2002), Problem 4.3.3). Then the size-effect process L_ξ for the large-stock portfolio ξ of (3.1) can be defined by

$$\begin{aligned} dL_\xi(t) &\triangleq \sum_{k=1}^{m} \xi_{(k)}(t)\, dL_{(k)}(t) \\ &= \frac{1}{2}\sum_{k=1}^{m} \xi_{(k)}(t)\big(d\Lambda_{\log\mu_{(k-1)}-\log\mu_{(k)}}(t) - d\Lambda_{\log\mu_{(k)}-\log\mu_{(k+1)}}(t)\big) \\ &= \frac{1}{2}\sum_{k=1}^{m} \big(\xi_{(k-1)}(t)\, d\Lambda_{\log\mu_{(k-1)}-\log\mu_{(k)}}(t) - \xi_{(k)}(t)\, d\Lambda_{\log\mu_{(k)}-\log\mu_{(k+1)}}(t)\big) \\ &= -\frac{1}{2}\xi_{(m)}(t)\, d\Lambda_{\log\mu_{(m)}-\log\mu_{(m+1)}}(t), \quad t \in [0,\infty), \quad \text{a.s.} \end{aligned} \tag{3.5}$$

Note that (3.5) follows from the fact that the support of $\Lambda_{\log\mu_{(k)}-\log\mu_{(k+1)}}$ lies within the set $\{t : \mu_{(k-1)}(t) = \mu_{(k)}(t)\}$, which is the same set as $\{t : \xi_{(k-1)}(t) = \xi_{(k)}(t)\}$. In a similar manner, we can show that

$$dL_\eta(t) = \frac{1}{2}\eta_{(m+1)}(t)\, d\Lambda_{\log\mu_{(m)}-\log\mu_{(m+1)}}(t), \quad t \in [0,\infty) \tag{3.6}$$

holds almost surely for the small-stock portfolio η of (3.2). It is convenient to define the *size-corrected* portfolio value process $\widehat{Z}_\xi$ for ξ by

$$\log \widehat{Z}_\xi(t) \triangleq \log Z_\xi(t) - L_\xi(t), \tag{3.7}$$

and similarly for η and μ, in which case it follows from (3.3), (3.5), and (3.6) that a.s.,

$$\log\big(\widehat{Z}_\eta(t)/\widehat{Z}_\xi(t)\big) = \log\left(\frac{\mu_{(m+1)}(t) + \cdots + \mu_{(n)}(t)}{\mu_{(1)}(t) + \cdots + \mu_{(m)}(t)}\right), \quad t \in [0,\infty). \tag{3.8}$$

We see that the size effect *has been exactly neutralized by subtracting the size-effect processes of* (3.4). In particular, it is easily seen that coherence implies

$$\lim_{t\to\infty} \frac{1}{t} \log\big(\widehat{Z}_\eta(t)/\widehat{Z}_\xi(t)\big) = 0, \quad \text{a.s.} \tag{3.9}$$

Note also that, for $m = n$, a calculation similar to (3.5) implies

$$L_\mu(t) = 0, \quad t \in [0,\infty) \tag{3.10}$$

a.s., so the size-effect process for the market as a whole is zero, as we would have expected.

An example of the size effect is presented in Fernholz (2002), pp. 133–136; it shows that over the period from 1939 to 1998, the stocks ranked 101 to 1000 in the U.S. market had average logarithmic return more that 1% a year greater than the stocks ranked 1 to 100. In this example, the relative capitalization of the small-stock portfolio versus the large-stock portfolio is mean-reverting over the long term. This implies that the variances of the large-stock and the small-stock portfolios will be about the same when measured using a sampling interval with length equal to or greater than the relaxation time of the relative capitalization process. Hence, over periods of this length, the two portfolios will have about the same risk.

The process on the right-hand side of (3.8) is the logarithm of the ratio of the capitalizations of the small-stock portfolio and the large-stock portfolio. In a stable market, it is reasonable to expect that this process will be stationary, and hence the size-corrected processes $\log \widehat{Z}_\eta$ and $\log \widehat{Z}_\xi$ will be *cointegrated* (see Engle and Granger (1987)). If the size-corrected processes are cointegrated, then they will have about the same variance when sampled over intervals of the order of the relaxation time of the process in (3.8). Again we have evidence that the long-term risk of the small-stock portfolio and the large-stock portfolio will be about the same. Since risk cannot explain the difference in return, then, by default, *the difference must be due to a liquidity premium for the smaller stocks.*

4 The implied liquidity premium for an asymptotically stable market

If a liquidity premium is the justification for the size effect, then we should be able to estimate this liquidity premium as a function of stock rank. We have seen that the size-effect processes $L_{(k)}$ represent the contribution of each ranked stock to the size effect, so our estimate of the liquidity premium will be based on these processes.

The implied liquidity premium of a stock is the rate at which the size-effect process $L_{(k)}$ contributes to a portfolio holding that stock while the stock resides at rank k. In this section we shall formalize this notion of liquidity premium, and estimate its value by rank for the stocks in the market.

We say that the market is *asymptotically stable* if the limits

$$\boldsymbol{\lambda}_{k,k+1} = \lim_{t\to\infty} t^{-1}\Lambda_{\log\mu_{(k)}-\log\mu_{(k+1)}}(t) \tag{4.1}$$

and

$$\boldsymbol{\sigma}^2_{k:k+1} = \lim_{t\to\infty} t^{-1}\langle \log\mu_{(k)} - \log\mu_{(k+1)}\rangle_t \tag{4.2}$$

exist a.s. for $k = 1, \dots, n-1$, and are positive, real constants. (If the tail σ-algebra of the filtration $\{\mathcal{F}_t\}$ is trivial, as is the case where $\{\mathcal{F}_t\}$ is generated by a Brownian motion, then the limits in (4.1) and (4.2), whenever they exist, will be non-random and will take values in $[0,\infty]$.) In an asymptotically stable market we have

$$\lim_{T\to\infty}\frac{1}{T}\int_0^T\big(\log\mu_{(k)}(t) - \log\mu_{(k+1)}(t)\big)dt = \frac{\boldsymbol{\sigma}^2_{k:k+1}}{2\boldsymbol{\lambda}_{k,k+1}} \quad \text{a.s.,} \tag{4.3}$$

for $k = 1, \dots, n-1$ (see Fernholz (2002), p.102).

It was shown in Fernholz (2002), Section 5.4, that the U.S. equity market over the period from 1990 to 1999 appears to be consistent with this asymptotically stable structure. Hence, if we assume the market continues to behave as it did in the 1990s, it will be asymptotically stable, and we shall be able to calculate its structural parameters by means of (4.1), (4.2), and (4.3). Accordingly, we shall henceforth assume that the market we consider is asymptotically stable. Following Fernholz (2002), Section 5.3, let us define the parameters

$$\boldsymbol{g}_k \triangleq \frac{1}{2}\boldsymbol{\lambda}_{k-1,k} - \frac{1}{2}\boldsymbol{\lambda}_{k,k+1}, \tag{4.4}$$

for $k = 1, \ldots, n$. The $2n-2$ parameters $\boldsymbol{g}_1, \ldots, \boldsymbol{g}_{n-1}, \boldsymbol{\sigma}^2_{1:2}, \ldots, \boldsymbol{\sigma}^2_{n-1:n}$ are called the *characteristic parameters* of the market.

From (3.4) and (4.1), it would seem that we could estimate the liquidity premium for the kth-ranked stock to be simply

$$\frac{1}{2}\big(\boldsymbol{\lambda}_{k-1,k} - \boldsymbol{\lambda}_{k,k+1}\big) = \boldsymbol{g}_k, \tag{4.5}$$

for $k = 1, \ldots, n$. This estimate might be adequate for individual stocks, but it will not be accurate in a portfolio setting, as with ξ, η, and μ. This is because (4.5) does not capture the dynamics of the process $L_{(k)}$ of (3.4), which increases when $\mu_{(k)}(t)$ is near $\mu_{(k-1)}(t)$, and decreases when $\mu_{(k)}(t)$ is near $\mu_{(k+1)}(t)$. One manifestation of this problem is that with the estimate in (4.5), the total liquidity premium for the market will not be zero. We shall try, then, to estimate $L_{(k)}$ in a manner that is consistent with the dynamics of $L_{(k)}$, and which preserves the property that the total market liquidity premium vanishes, as it does in (3.10). In particular, if $\ell_{(k)}(t)$ represents the liquidity premium of the kth-ranked stock at time t, then we would like to have

$$\int_0^T \mu_{(k)}(t)\ell_{(k)}(t)\,dt \cong \int_0^T \mu_{(k)}(t)\,dL_{(k)}(t), \tag{4.6}$$

for large enough values of T, at least if the ranked weight $\mu_{(k)}$ remains reasonably stable over time.

Let us consider the contribution of the kth-ranked stock to L_μ, which we can express as

$$\begin{aligned}
\mu_{(k)}(t)\,dL_{(k)}(t) &= \mu_{(k)}(t)\frac{1}{2}\Big(d\Lambda_{\log\mu_{(k-1)}-\log\mu_{(k)}}(t) - d\Lambda_{\log\mu_{(k)}-\log\mu_{(k+1)}}(t)\Big)\\
&= \frac{\mu_{(k-1)}(t)+\mu_{(k)}(t)}{4}\,d\Lambda_{\log\mu_{(k-1)}-\log\mu_{(k)}}(t)\\
&\quad - \frac{\mu_{(k)}(t)+\mu_{(k+1)}(t)}{4}\,d\Lambda_{\log\mu_{(k)}-\log\mu_{(k+1)}}(t),
\end{aligned} \tag{4.7}$$

for $k = 1, \ldots, n$, by the same reasoning that was used for (3.5). Besides being formally correct, this representation recalls the property that the local time $\Lambda_{\log\mu_{(k)}-\log\mu_{(k+1)}}$ contributes to the market size effect when $\mu_{(k)}(t)$ and $\mu_{(k+1)}(t)$ are close in value, for all k. Following (4.7), we are motivated to define

$$\ell_{(k)}(t) \triangleq \left(\frac{\mu_{(k)}(t)+\mu_{(k-1)}(t)}{4\mu_{(k)}(t)}\right)\boldsymbol{\lambda}_{k-1,k} - \left(\frac{\mu_{(k)}(t)+\mu_{(k+1)}(t)}{4\mu_{(k)}(t)}\right)\boldsymbol{\lambda}_{k,k+1}, \tag{4.8}$$

for all $t \in [0,\infty)$. With this definition, if the ranked weights are reasonably stable over time, then the approximate equality (4.6) should follow from (4.1).

With the definition (4.8), it is not difficult to verify that the liquidity premium ℓ_μ for the entire market satisfies almost surely

$$\ell_\mu(t) \triangleq \sum_{k=1}^n \mu_{(k)}(t)\ell_{(k)}(t) = 0, \quad t \in [0,\infty). \tag{4.9}$$

This equality implies that the liquidity premium for some of the stocks will be negative, so perhaps we should call it a liquidity "penalty" in those cases. But, nevertheless, we shall continue to call it a premium even in the negative cases.

For the portfolios η and ξ, it is not difficult to see that reasoning similar to (3.5) will hold, so the liquidity premium for the large-stock portfolio ξ will be given by

$$\ell_\xi(t) = -\left(\frac{\xi_{(m)}(t)+\xi_{(m+1)}(t)}{4}\right)\boldsymbol{\lambda}_{m,m+1}, \quad t \in [0,\infty), \quad \text{a.s.}, \tag{4.10}$$

and the liquidity premium for the small-stock portfolio η will be given by

$$\ell_\eta(t) = \frac{1}{2}\bigg(\frac{\eta_{(m)}(t) + \eta_{(m+1)}(t)}{4}\bigg)\boldsymbol{\lambda}_{m,m+1}, \quad t \in [0,\infty), \quad \text{a.s.} \tag{4.11}$$

Since $\mu_{(m)}$ and $\mu_{(m+1)}$ are equal on the support of $\Lambda_{\log\mu_{(m)}-\log\mu_{(m+1)}}$, it follows that the integrals of (3.5) and (4.10) are likely to be close in value, at least over the long term, and that the same is true for the integrals of (3.6) and (4.11). Hence, for these portfolios, the liquidity premium given by (4.8) approximates the size-effect process of (3.4), and should approximately neutralize the size effect, as in (3.8) and (3.9).

With $\boldsymbol{g}_k$ from (4.4), the liquidity premium in (4.8) becomes

$$\begin{aligned}\ell_{(k)}(t) &= \boldsymbol{g}_k + \bigg(\frac{\mu_{(k-1)}(t) + \mu_{(k)}(t)}{4\mu_{(k)}(t)} - \frac{1}{2}\bigg)\boldsymbol{\lambda}_{k-1,k} - \bigg(\frac{\mu_{(k)}(t) + \mu_{(k+1)}(t)}{4\mu_{(k)}(t)} - \frac{1}{2}\bigg)\boldsymbol{\lambda}_{k,k+1}\\ &= \boldsymbol{g}_k + \frac{1}{4}\bigg(\frac{\mu_{(k-1)}(t)}{\mu_{(k)}(t)} - 1\bigg)\boldsymbol{\lambda}_{k-1,k} - \frac{1}{4}\bigg(\frac{\mu_{(k+1)}(t)}{\mu_{(k)}(t)} - 1\bigg)\boldsymbol{\lambda}_{k,k+1},\end{aligned}$$

for $t \in [0,\infty)$, a.s. Since $\log x \cong x - 1$ for x sufficiently close to 1, this becomes

$$\ell_{(k)}(t) \cong \boldsymbol{g}_k + \frac{1}{4}\log\bigg(\frac{\mu_{(k-1)}(t)}{\mu_{(k)}(t)}\bigg)\boldsymbol{\lambda}_{k-1,k} - \frac{1}{4}\log\bigg(\frac{\mu_{(k+1)}(t)}{\mu_{(k)}(t)}\bigg)\boldsymbol{\lambda}_{k,k+1}, \quad t \in [0,\infty), \quad \text{a.s.}$$

From this and (4.3), we obtain the long-term approximation

$$\ell_{(k)}(t) \cong \boldsymbol{g}_k + \frac{1}{8}\boldsymbol{\sigma}^2_{k-1:k} + \frac{1}{8}\boldsymbol{\sigma}^2_{k:k+1}, \quad t \in [0,\infty), \quad \text{a.s.} \tag{4.12}$$

The values in (4.12) can be estimated for actual equity markets. In Fernholz (2002), p. 109, estimates were given for the terms on the right-hand side of (4.12) for the U.S. equity market under the assumption that the market behaves asymptotically as it did during the period from 1990 to 1999. From those estimates, we have calculated the liquidity premium rates $\ell_{(k)}(t)$, and the results are shown in Figure 1. The values are smoothed by convolution with a Gaussian kernel with $\pm 3.16\sigma$ spanning 1000 units on the horizontal axis, with reflection at the ends of the data. Dividend payments could partially offset the liquidity premium for a stock, so, to correct for this, we added the dividend yield for each ranked stock to the stock's liquidity premium. This dividend-corrected liquidity premium was then renormalized to satisfy (4.9). We can see from Figure 1 that the liquidity premium is essentially nil for the largest 1000 stocks, but then grows to nearly 20% annually for the stock ranked 5000. The values used in Figure 1 are based on the values from Figures 5.4 and 5.5 in Fernholz (2002).

5 The liquidity premium in the first-order model

The *first-order model* for the market, proposed in Fernholz (2002), Section 5.5, is a model of the asymptotic structure of the market, and hence gives a representation of the market's long-term behavior. Let us now recall the structure of the first-order model.

Consider the quantities $\boldsymbol{\sigma}^2_1, \dots, \boldsymbol{\sigma}^2_n$ defined by

$$\begin{gathered}\boldsymbol{\sigma}^2_k \triangleq \frac{1}{4}(\boldsymbol{\sigma}^2_{k-1:k} + \boldsymbol{\sigma}^2_{k:k+1}), \quad k = 2, \dots, n-1,\\ \boldsymbol{\sigma}^2_1 \triangleq \frac{1}{2}\boldsymbol{\sigma}^2_{1:2}, \quad \text{and} \quad \boldsymbol{\sigma}^2_n \triangleq \frac{1}{2}\boldsymbol{\sigma}^2_{n-1:n}.\end{gathered} \tag{5.1}$$

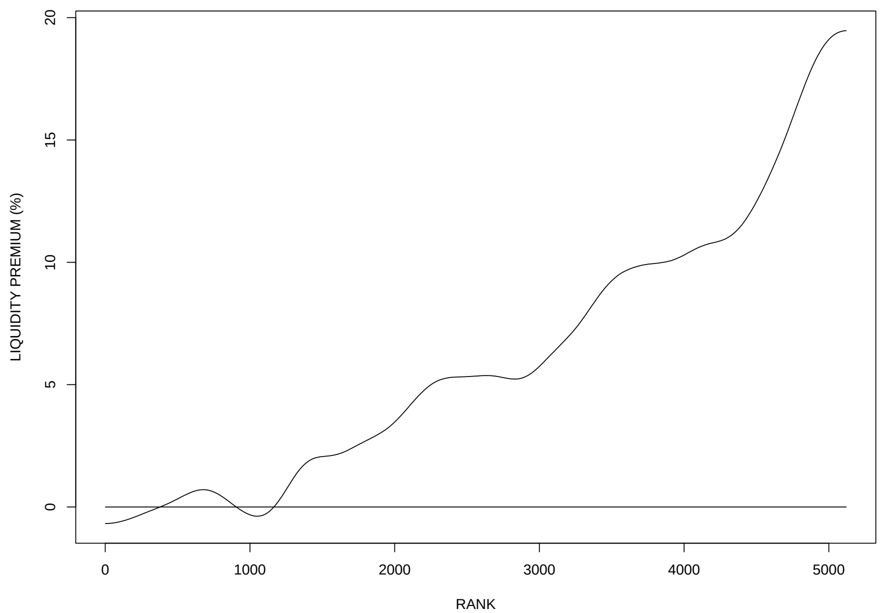

Figure 1: Implied annual liquidity premium for the U.S. equity market, 1990–1999, corrected for dividends and renormalized to satisfy (4.9).

Then the model given by

$$d\log X_i(t) = \boldsymbol{g}_{q_t(i)}\,dt + \boldsymbol{\sigma}_{q_t(i)}\,dV_i(t), \quad t \in [0,\infty), \tag{5.2}$$

for $i = 1,\ldots,n$, where q_t is the inverse permutation of p_t and $(V_1,\ldots,V_n)$ is n-dimensional Brownian motion, is called the *first-order model* for the market. (Here, there are exactly as many driving processes as there are stocks; $d = n$.) It was shown in Fernholz (2002) that the first-order model defines an asymptotically stable market with characteristic parameters $\boldsymbol{g}_1,\ldots,\boldsymbol{g}_{n-1}$, $\big(3\boldsymbol{\sigma}^2_{1:2} + \boldsymbol{\sigma}^2_{2:3}\big)/4$, $\big(\boldsymbol{\sigma}^2_{k-1:k} + 2\boldsymbol{\sigma}^2_{k:k+1} + \boldsymbol{\sigma}^2_{k+1:k+2}\big)/4$ for $k = 2,\ldots,n-2$, and $\big(\boldsymbol{\sigma}^2_{n-2:n-1} + 3\boldsymbol{\sigma}^2_{n-1:n}\big)/4$. Since this is a smoothed version of the original variance parameters, if the original parameters are sufficiently smooth, the first-order parameters will match them closely. A more complete discussion of this relationship can be found in Fernholz (2002), Section 5.5, and a study of the probabilistic structure of the system (5.2) of stochastic differential equations appears in Banner et al. (2005).

If we compare (4.12) with (5.1) and recall the rate of return processes b_i from (2.4), we see that for the first-order model, the "ranked" rate of return processes are given by

$$b_{(k)}(t) = \ell_{(k)}(t), \quad t \in [0,\infty), \quad \text{a.s.}, \tag{5.3}$$

for $k = 2,\ldots,n-1$, with $b_{(1)}(t) = \ell_{(1)}(t) + \boldsymbol{\sigma}^2_1/4$ and $b_{(n)}(t) = \ell_{(n)}(t) + \boldsymbol{\sigma}^2_n/4$. The discrepancies for $k = 1$ and n are a consequence of the approximation (4.8), and have little effect on the liquidity premium for the portfolios η, ξ, and μ.

We see from (5.3) that in the first-order model, *the asymptotic rate of return of a stock investment is equal to the stock's liquidity premium.* While it may be reasonable for long-term investors to benefit from holding stocks that pose short-term trading difficulties, it seems surprising that the entire asymptotic rate of return should be due to the liquidity premium alone.

6 Conclusion

In a stable market, a small-stock portfolio is likely to outperform a large-stock portfolio, even though the two portfolios have about the same intermediate-term risk. This phenomenon, known as the *size effect*, can be characterized in terms of the semimartingale local times of the ranked market weights. In an asymptotically stable market, the contributions of these local times, and, hence, of the size effect, can be calculated analytically. This contribution, since it is not due to risk, is attributed to a liquidity premium for the smaller stocks. This liquidity premium is approximately equal to the rate of return for the first-order model, a model of the asymptotic behavior of the market.

Acknowledgement: The authors wish to express their gratitude to an anonymous referee for a very thorough and incisive reading, as well as for many constructive suggestions that have significantly improved this paper.

References

Amihud, Y. and H. Mendelson (1986). Asset pricing and the bid-ask spread. *Journal of Financial Economics 17*, 223–249.

Banner, A., R. Fernholz, and I. Karatzas (2005). On Atlas models of equity markets. *Annals of Applied Probability*. To appear.

Banz, R. (1981). The relationship between return and market value of common stocks. *Journal of Financial Economics 9*, 3–18.

Duffie, D. (1992). *Dynamic Asset Pricing Theory.* Princeton, NJ: Princeton University Press.

Engle, R. F. and C. W. J. Granger (1987). Co-integration and error correction: representation, estimation, and testing. *Econometrica 55*, 251–276.

Fernholz, R. (1998, May/June). Crossovers, dividends, and the size effect. *Financial Analysts Journal 54*(3), 73–78.

Fernholz, R. (2001). Equity portfolios generated by functions of ranked market weights. *Finance and Stochastics 5*, 469–486.

Fernholz, R. (2002). *Stochastic Portfolio Theory.* New York: Springer-Verlag.

Fernholz, R. and I. Karatzas (2005). Relative arbitrage in volatility-stabilized markets. *Annals of Finance 1*, 149–177.

Karatzas, I. and S. E. Shreve (1991). *Brownian Motion and Stochastic Calculus.* New York: Springer-Verlag.

Karatzas, I. and S. E. Shreve (1998). *Methods of Mathematical Finance.* New York: Springer-Verlag.

Reinganum, M. (1981). Misspecification of capital asset pricing: empirical anomalies based on earnings yields and market values. *Journal of Financial Economics 9*, 19–46.

Roll, R. (1981). A possible explanation of the small firm effect. *Journal of Finance 36*, 879–888.

Vayanos, D. (2003). Flight to quality, flight to liquidity, and the pricing of risk. Technical report, MIT, Cambridge, MA.