A Second-Order Model of the Stock Market

Robert Fernholz
INTECH

Conference in honor of Ioannis Karatzas
Columbia University, New York
June 4–8, 2012

This is joint research with T. Ichiba, I. Karatzas.

Introduction

A *first-order model* for a stock market is a model that assigns to each stock in the market a return parameter and a variance parameter that depend only on the *rank* of the stock. A *second-order model* for a stock market is a model that assigns these parameters based on both the rank and the *name* of the stock. A second-order model is an example of a *hybrid Atlas model.*

Acknowledgements. The speaker thanks Adrian Banner, Daniel Fernholz, Vassilios Papathanakos, and Johannes Ruf for their many helpful discussions and suggestions, as well as for their participation and inspiration during the course of this research.

Reference

- R. Fernholz, T. Ichiba, and I. Karatzas (2012).
  A second-order stock market model. *Annals of Finance*, to appear.

Stock markets

A *market* is a family of *stocks* $X = (X_1, \ldots, X_n)$represented by positive absolutely continuous semimartingales defined on $[0, \infty)$ or on $\mathbb{R}$. The value $X_i(t)$ of the stock X_i at time t represents the total capitalization of the company at that time. Let Z represent the total capitalization of the market,

$$Z(t) \triangleq X_1(t) + \cdots + X_n(t).$$

Then the *market weights* μ_i, for $i = 1, \ldots, n$, given by

$$\mu_i(t) \triangleq \frac{X_i(t)}{Z(t)},$$

define the *market portfolio* μ.

Market stability

We shall assume that the market weight process $\mu = (\mu_1, \ldots, \mu_n)$ has a stable, or *steady-state*, distribution, and that the system is in that stable distribution. We shall be interested in the relative behavior of the log-capitalizations or log-weights. If $\mu(t)$ is in its steady-state distribution, then the *log-difference processes* defined by

$$\log X_i(t) - \log X_j(t) = \log \mu_i(t) - \log \mu_j(t),$$

for $i, j = 1, \ldots, n$, will also be in their steady-state distribution.

We shall also assume that there are almost surely no triple points for the market, i.e., there is no time t at which $X_i(t) = X_j(t) = X_k(t)$ for $i < j < k$, almost surely.

Ranked processes

Consider the *ranked capitalization processes*

$$X_{(1)}(t) \geq \cdots \geq X_{(n)}(t),$$

and the corresponding *ranked market weights*

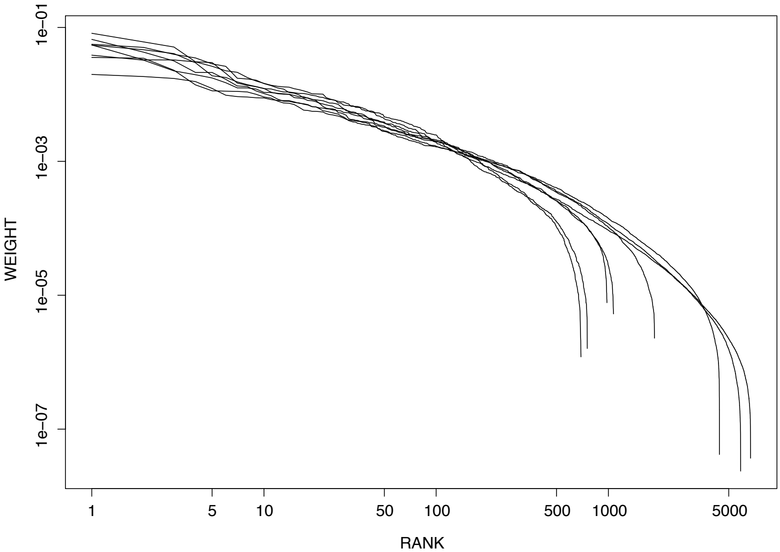
$$\mu_{(1)}(t) \geq \cdots \geq \mu_{(n)}(t).$$

Let $r_t(i)$ represent the rank of $X_i(t)$, and let p_t be the inverse permutation of r_t, so

$$X_i(t) = X_{(r_t(i))}(t) \quad \text{and} \quad X_{(k)}(t) = X_{p_t(k)}(t).$$

The ranked market weights $(\mu_{(1)}(t), \ldots, \mu_{(n)}(t))$ comprise the *capital distribution curve* of the market at time t.

Capital distribution curve



Capital distribution of the U.S. market: 1929–1999.

First-order models

A *first-order model* is a stock market model defined by

$$d \log \widehat{X}_i(t) = g_{\hat{r}_t(i)}\, dt + \sigma_{\hat{r}_t(i)}\, dW_i(t),$$

for $i = 1, \ldots, n$, where $g_1, \ldots, g_n$ are constants, $\sigma_1, \ldots, \sigma_n$ are positive constants, and $W = (W_1, \ldots, W_n)$ is a Brownian motion. We shall assume that the g_k satisfy

$$g_1 + \cdots + g_n = 0,$$

and

$$\sum_{k=1}^{m} g_k < 0,$$

for $m < n$. With these parameters, the $\widehat{X}_i$ form an asymptotically stable system.

Rank-based parameters

Suppose we have a market X that is in its steady-state distribution. We define the asymptotic *rank-based relative variances* for the market by

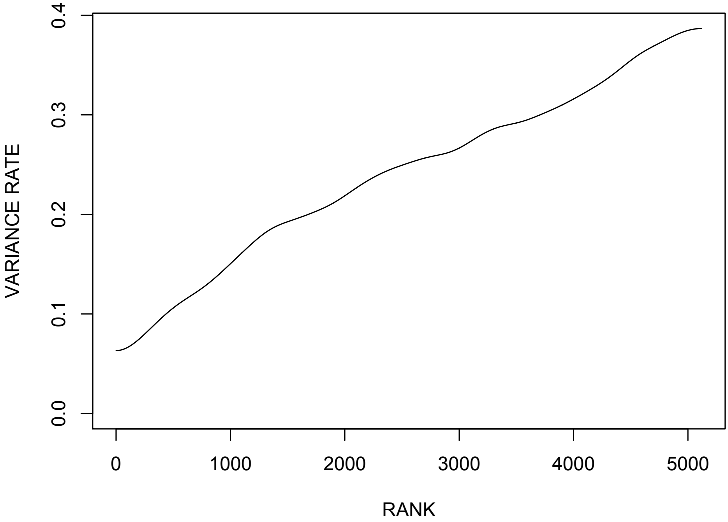
$$\boldsymbol{\sigma}_k^2 \triangleq \lim_{t \to \infty} \frac{\langle \mu_{p_t(k)} \rangle (t)}{t},$$

and the asymptotic *rank-based relative growth rates* by

$$\mathbf{g}_k \triangleq \lim_{T \to \infty} \frac{1}{T} \int_0^T d \log \mu_{p_t(k)}(t).$$

Since these parameters are based on the market weight processes μ_i, the parameters represent values measured relative to the market capitalization process X.

Rank-based variances



σ_k^2 for U.S. market: 1990–1999.

Local times for the rank processes

Let $\Lambda_{k,k+1}$ be the local time of $\log(\mu_{(k)}/\mu_{(k+1)}) \geq 0$ at the origin. Then

$$d \log \mu_{(k)}(t) = d \log \mu_{p_t(k)}(t) + \frac{1}{2}\, d\Lambda_{k,k+1}(t) - \frac{1}{2}\, d\Lambda_{k-1,k}(t).$$

For $k = 1, \ldots, n-1$, we can define the *asymptotic local time*

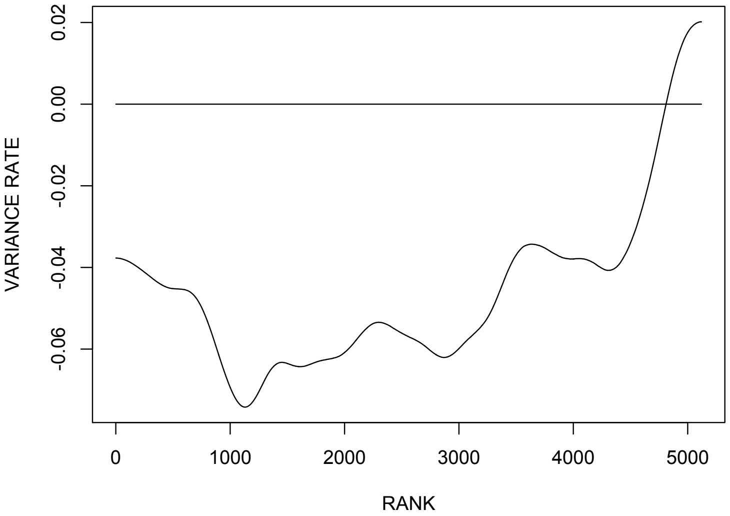
$$\boldsymbol{\lambda}_{k,k+1} \triangleq \lim_{t\to\infty} t^{-1}\Lambda_{k,k+1}(t),$$

and let $\boldsymbol{\lambda}_{0,1} = 0 = \boldsymbol{\lambda}_{n,n+1}$. It can be shown that

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

for $k = 1, \ldots, n$, and it follows that $\mathbf{g}_1 + \cdots + \mathbf{g}_n = 0$.

Relative growth rates



Normalized $\mathbf{g}_k$ for U.S. market: 1990–1999.

A first-order market model

The first-order model with

$$d \log \widehat{X}_i(t) = \mathbf{g}_{\hat{r}_t(i)}\, dt + \boldsymbol{\sigma}_{\hat{r}_t(i)}\, dW_i(t),$$

is called the *first-order model* for the market X. As we have seen, the growth and variance parameters for the $\widehat{X}_i$ are derived from the relative growth and variance parameters corresponding to the market weight processes μ_i, not directly from the price processes X_i.

The steady-state capital distribution curve for the first-order model of the stock market will be about the same as the capital distribution curve for the market itself.

Mathematical characteristics of first-order models

- A first-order model is asymptotically stable.
- A first-order model may have triple points where $\widehat{X}_i(t) = \widehat{X}_j(t) = \widehat{X}_k(t)$ for $i < j < k$, but the local time at the origin for $\log(\widehat{X}_{(k)}/\widehat{X}_{(\ell)}) \geq 0$ is zero if $\ell > k + 1$.
- A first-order model is ergodic in the sense that

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T \mathbf{1}_{\{\widehat{X}_i(t)=\widehat{X}_{(k)}(t)\}}\, dt = \frac{1}{n}.$$

This ergodicity does not seem reasonable for a real market, so we must extend our model.

Hybrid models

A *hybrid (Atlas) model* is a stock market model defined by

$$d \log \widehat{X}_i(t) = (\gamma_i + g_{\hat{r}_t(i)})dt + \sigma_{i,\hat{r}_t(i)}\, dW_i(t),$$

for $i = 1, \ldots, n$, with constants g_k, γ_i and $\sigma_{ik} > 0$, for $i, k = 1, \ldots, n$, and a Brownian motion W. These parameters satisfy

$$\sum_{k=1}^{n} g_k = \sum_{i=1}^{n} \gamma_i = 0,$$

and, for any permutation $\pi \in \Sigma_n$,

$$\sum_{k=1}^{m} (g_k + \gamma_{\pi(k)}) < 0, \quad \text{for } m < n.$$

We shall use first-order variances, so $\sigma_{ik}^2 = \sigma_k^2$ for all i and k.

Occupation rates

For a market X, the expected *occupation rate*

$$\theta_{ki} \triangleq \lim_{T\to\infty} \frac{1}{T} \int_0^T \mathbf{1}_{\{X_i(t)=X_{(k)}(t)\}}\, dt$$

is defined for all i and k. The matrix $\theta = (\theta_{ki})$ is bistochastic, and we shall assume that all the entries are positive. For a hybrid model $\widehat{X}$ with occupation-rate matrix $\hat{\theta}$,

$$\hat{\mathbf{g}}_k = g_k + \sum_{i=1}^{n} \hat{\theta}_{ki}\gamma_i$$

$$0 = \gamma_i + \sum_{k=1}^{n} \hat{\theta}_{ki} g_k,$$

where the $\hat{\mathbf{g}}_k$ are the rank-based relative growth rates.

Parameter estimates based on occupation rates

In matrix form, this can be expressed

$$\hat{\mathbf{g}} = g + \hat{\theta}\gamma$$
$$0 = \gamma + \hat{\theta}^T g,$$

where γ, g, and $\hat{\mathbf{g}}$ are column vectors. From this we see that

$$\gamma = -\hat{\theta}^T g, \tag{1}$$

so

$$\hat{\mathbf{g}} = \left(I_n - \hat{\theta}\hat{\theta}^T\right) g. \tag{2}$$

As we have seen, $\hat{\mathbf{g}}$ can be estimated directly, so we need to solve (2) for g, and then γ can be calculated using (1).

Second-order parameter estimation

Let θ be the occupation-rate matrix of the market X. Since θ is bistochastic with positive entries, so are θ^T and $\theta\theta^T$. By the Perron-Frobenius theorem, $\theta\theta^T$ will have a simple eigenvalue equal to 1 with eigenvector $e_1 = (1, 1, \ldots, 1)'$, and all the other eigenvalues will have absolute value less than 1 (see Perron, O. (1907) Zur Theorie der Matrices, *Math. Annalen 64*, 248–263).

From this it follows that $I_n - \theta\theta^T$ has rank $n - 1$, and its kernel is generated by e_1, so the condition that the g_k sum to zero ensures a unique solution to

$$\mathbf{g} = \left(I_n - \theta\theta^T\right)g.$$

Exploratory second-order parameter estimation

Unfortunately, it seems to be impossible to estimate θ, so although we can use it to prove the existence and uniqueness of g, we cannot actually solve the equations. Instead, we plan to work with

$$\mathbf{g}_k = g_k + \sum_{i=1}^{n} \theta_{ki}\gamma_i$$

in such a way that we can solve for the g_k recursively. Once we have found $\mathbf{g}$ and g, we then can estimate γ directly by using

$$\gamma_i = \lim_{T\to\infty} \frac{1}{T} \int_0^T \big(d \log \mu_i(t) - g_{r_t(i)}\, dt\big).$$

Time reversal and parameter estimation

Let $\mathbf{X}$ be a stable market defined for $t \in \mathbb{R}$. We can define the *time-reversed* market $\widetilde{X}$ with price processes $\widetilde{X}_i(t) \triangleq X_i(-t)$ and weight processes $\widetilde{\mu}_i(t) = \mu_i(-t)$. Then:

- The forward and backward occupation rates θ_{ki} are equal.
- The forward and backward asymptotic local times $\boldsymbol{\lambda}_k$ are equal (see, e.g., Bertoin, J. (1987) Temps locaux et intégration stochastique pour les processus de Dirichlet, *Séminaire de Probabilités (Strasbourg) 21*, 191–205).
- Hence, the forward and backward $\mathbf{g}_k$ are equal.
- Hence, the forward and backward g_k are equal.
- Hence, the forward and backward γ_i are equal.
- Quadratic-variation is invariant under time reversal.

Market flow

The *forward flow* φ_k of the market X at rank k in the market is defined for $\tau \in [0, \infty)$ by

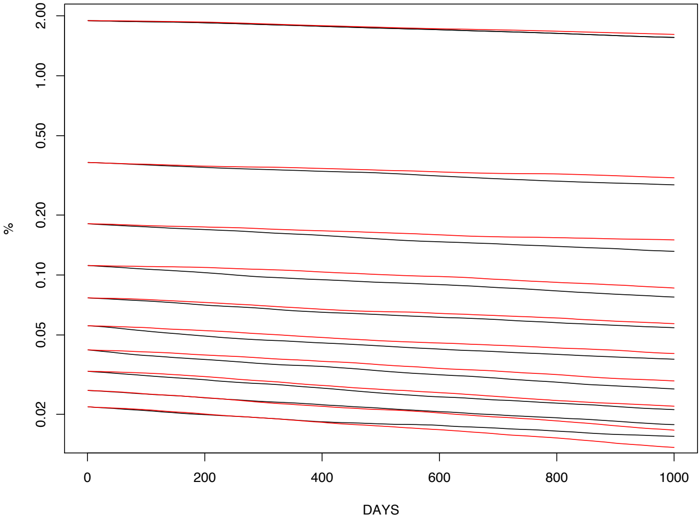
$$\varphi_k(\tau) \triangleq \lim_{T\to\infty} \frac{1}{T} \int_0^T \log\left(\frac{\mu_{p_t(k)}(t+\tau)}{\mu_{(k)}(t)}\right) dt.$$

The *backward flow* $\widetilde{\varphi}_k$ of the market is defined by

$$\widetilde{\varphi}_k(\tau) \triangleq \lim_{T\to\infty} \frac{1}{T} \int_0^T \log\left(\frac{\widetilde{\mu}_{p_t(k)}(t+\tau)}{\widetilde{\mu}_{(k)}(t)}\right) dt.$$

The forward and backward flows need not be equal.

Forward and backward flow for top 250 stocks



$\mu_{(k)}(0)e^{\varphi_k(\tau)}$ (black), $\mu_{(k)}(0)e^{\widetilde{\varphi}_k(\tau)}$ (red): 1990–1999.

Forward and backward expected rank

If we follow the flow of a stock that is in rank k at time 0, then we can estimate its expected rank at time $\tau \in \mathbb{R}$ by

$$\mathbf{R}_k(\tau) \triangleq \lim_{T\to\infty} \frac{1}{T} \int_0^T r_{s+\tau}(p_s(k))\, ds,$$

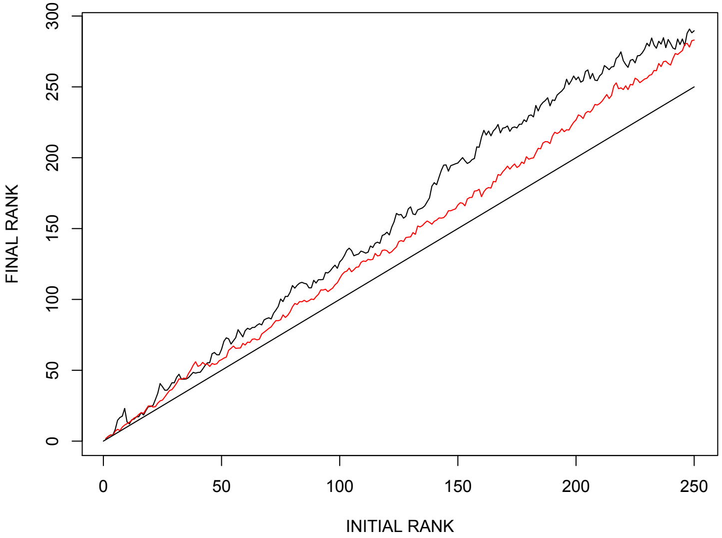
so $\mathbf{R}_k(0) = k$.

We shall use the $\mathbf{R}_k$ to estimate the g_k. Although $\mathbf{R}_k(\tau)$ need not equal $\mathbf{R}_k(-\tau)$, the g_k generated using either one provide estimates for the solution of

$$\mathbf{g} = \left(I_n - \theta\theta^T\right) g,$$

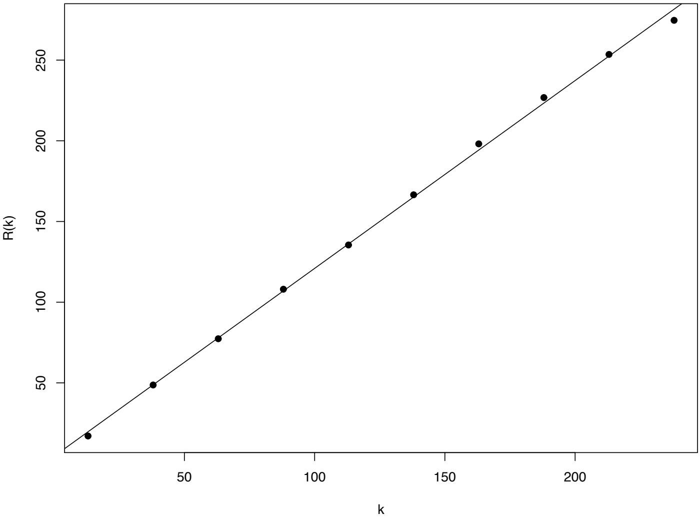
so we shall use them both.

Forward and backward expected rank



$\mathbf{R}_k(4)$ (black) and $\mathbf{R}_k(-4)$ (red): 1990–1999.

Change in rank for $\tau = 4$ years



$$\overline{\mathbf{R}}_k(4) \triangleq \frac{1}{2}\big(\mathbf{R}_k(4) + \mathbf{R}_k(-4)\big) \cong 4.6 + 1.16k.$$

Expected growth rate

The expected growth rate at time τ of the stock which is at rank k at time 0 will be

$$\overline{\mathbf{G}}_k(\tau) \triangleq g_{\overline{\mathbf{R}}_k(\tau)} + \sum_{i=1}^{n} \theta_{ki}\gamma_i,$$

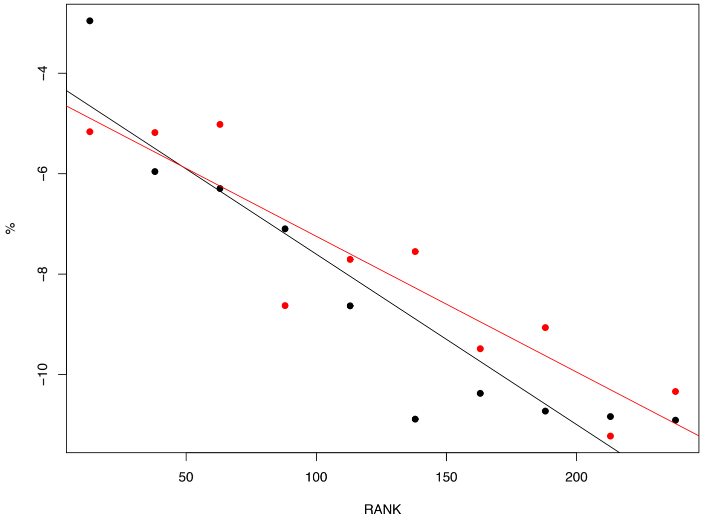
for $\tau \in \mathbb{R}$, with

$$g_{\overline{\mathbf{R}}_k(\tau)} \triangleq \left(\ell + 1 - \overline{\mathbf{R}}_k(\tau)\right) g_\ell + \left(\overline{\mathbf{R}}_k(\tau) - \ell\right) g_{\ell+1},$$

where ℓ the largest integer such that $\ell \leq \overline{\mathbf{R}}_k(\tau)$. We shall estimate $\overline{\mathbf{G}}_k(\tau)$ from the slope of the estimated flow, with

$$\overline{\mathbf{G}}_k(\tau) = \frac{1}{2}\left(D_\tau \varphi_k(\tau) + D_\tau \widetilde{\varphi}_k(\tau)\right).$$

Dependence of $\overline{\mathbf{G}}_k(\tau)$ on k



$\overline{\mathbf{G}}_k(0) \cong -4.2 - .034k$ (black), $\overline{\mathbf{G}}_k(4) \cong -4.5 - .027k$ (red).

Recursive calculation of g_k

Recall that we have

$$\overline{\mathbf{G}}_k(\tau) = g_{\overline{\mathbf{R}}_k(\tau)} + \sum_{i=1}^{n} \hat{\theta}_{ki}\gamma_i,$$

and we can combine this with

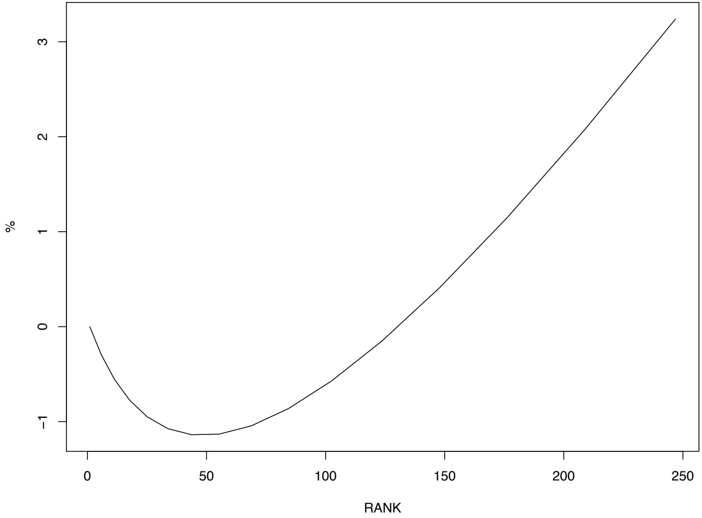
$$\mathbf{g}_k = g_k + \sum_{i=1}^{n} \hat{\theta}_{ki}\gamma_i,$$

so

$$g_{\overline{\mathbf{R}}_k(\tau)} = g_k + \overline{\mathbf{G}}_k(\tau) - \mathbf{g}_k. \tag{3}$$

We can calculate $\mathbf{g}_k$, $\overline{\mathbf{G}}_k(\tau)$, and $\overline{\mathbf{R}}_k(\tau)$ as above, and then use (3) recursively to calculate the g_k.

Recursive calculation of g_k



Non-normalized values of g_k for ranks 1 to 250.

An estimate of the γ_i

With these values of the g_k, we can estimate the γ_i directly by

$$\gamma_i = \frac{1}{2}\Big(\lim_{T\to\infty} \frac{1}{T}\int_0^T \big(d\log\mu_i(t) - g_{r_t(i)}\,dt\big) + \lim_{T\to\infty}\frac{1}{T}\int_0^T \big(d\log\widetilde{\mu}_i(t) - g_{r_t(i)}\,dt\big)\Big).$$

Our second-order model then will be

$$d\log\widehat{X}_i(t) = (\gamma_i + g_{\hat{r}_t(i)})dt + \sigma_{\hat{r}_t(i)}\,dW_i(t).$$

Here are some of the values for γ_i for 1990–1999:

AAPL (93)	=	−1.67%	GE (1)	=	0.14%
IBM (6)	=	−.10%	KO (4)	=	0.26%
MSFT (5)	=	−.12%	XON (3)	=	0.11%