Stable Diffusions Interacting through Their Ranks, as Models of Large Equity Markets

IOANNIS KARATZAS

Department of Mathematics, Columbia University

INTECH Investment Management, Princeton

Joint work with Tomoyuki Ichiba, Adrian Banner, Vasileios Papathanakos and E. Robert Fernholz

Talk at USC Los Angeles, March 2013

Log-Log Capital Distribution Curves I

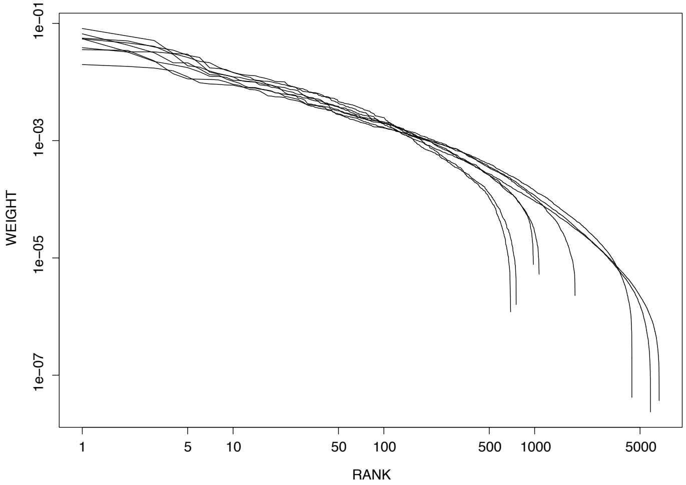


Figure: U.S. equity market, 1929-1999 (E.R. Fernholz (2002), p. 95)

Log-Log Capital Distribution Curves II

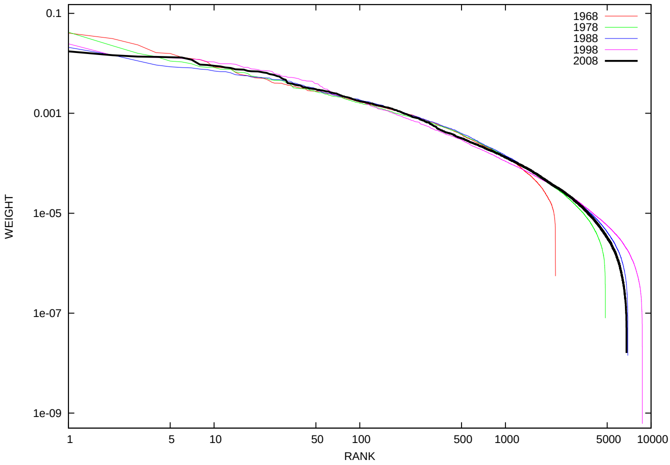


Figure: Capital distribution curves, U.S. equity market, 1968-2008

What kinds of models can describe this long-term stability?

Definition of Hybrid Atlas Model

- Capitalizations $\mathfrak{X} := \{(X_1(t), \dots, X_n(t)), \ \ 0 \le t < \infty\}$.
- Descending Order Statistics (lexicographic tie-breaks):

$$\max_{1 \le i \le n} X_i(t) =: X_{(1)}(t) \ge X_{(2)}(t) \ge \cdots \ge X_{(n)}(t) := \min_{1 \le i \le n} X_i(t) \,.$$

The curves of the previous slides are (smoothed) maps

$$\log k \longmapsto \frac{1}{T} \int_0^T \log \left(\frac{X_{(k)}(t)}{X_1(t) + \cdots + X_n(t)} \right) dt \,,$$

for $k = 1, 2, \cdots, n$ over different decades $[0, T]$
(for instance, Jan 1969 – Dec 1978; of course,
each decade has its own, associated market “size” n).

Log-Capitalizations

Log-capitalizations $Y_i(t) := \log X_i(t)$.
Descending Order Statistics: $Y_{(1)}(\cdot) \geq \cdots \geq Y_{(n)}(\cdot)$.

Postulated Dynamics for Log-Capitalizations (schematically):

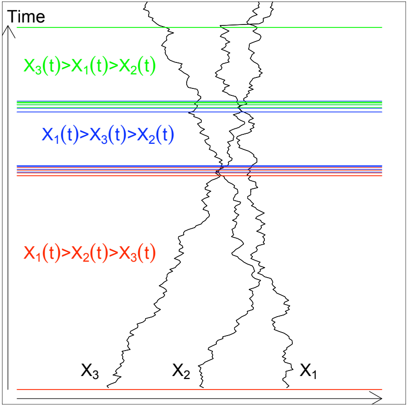
$$\mathrm{d}Y_i(t) = (\gamma + \gamma_i + g_k)\,\mathrm{d}t + \sigma_k\,\mathrm{d}W_i(t) \qquad \text{if} \quad Y_{(k)}(t) = Y_i(t)\,;$$

for $1 \leq i, k \leq n$, $0 \leq t < \infty$, where $W_1(\cdot), \cdots, W_n(\cdot)$ are independent standard Brownian Motions

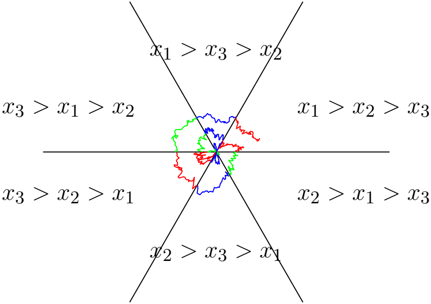
System of Brownian particles interacting through their ranks.
Unique weak solution (BASS & PARDOUX, PTRF '87).

	company name i	k^{th} ranked company
Drift ("mean")	γ_i	g_k
Diffusion ("variance")		$\sigma_k > 0$

Illustration ($n = 3$) of Interactions through Rank: Linear and Kaleidoscopic Views



Paths in $\mathbb{R}_+ \times$ time



A path in different wedges of $\mathbb{R}^n$

Permutations and Polyhedral Chambers

For $\mathbf{p} \in \Sigma_n$ (symmetric group on n elements), define wedge

$$\mathcal{R}_{\mathbf{p}} := \left\{ \xi \in \mathbb{R}^n : \xi_{\mathbf{p}(1)} > \xi_{\mathbf{p}(2)} > \cdots > \xi_{\mathbf{p}(n)} \right\},$$

the polyhedral Weyl chamber of all points $\xi \in \mathbb{R}^n$ such that $\xi_{\mathbf{p}(k)}$ is ranked k^{th} among $\xi_1, \cdots, \xi_n$.

To wit: $\mathbf{p}(k)$ is the "index" (name) in the permutation $\mathbf{p} \in \Sigma_n$ of the "particle" (coördinate) that occupies the k^{th} rank among $\xi_1, \cdots, \xi_n$.

FINE CHAMBERS $(n!)$

- Consider also the "coarser" chambers

$$Q_k^{(i)} := \{ \xi \in \mathbb{R}^n : \xi_i \text{ is ranked } k^{th} \text{ among } \xi_1, \ldots, \xi_n \} = \bigcup_{\{\mathbf{p} \in \Sigma_n : \mathbf{p}(k)=i\}} \mathcal{R}_\mathbf{p} \; ; \quad 1 \leq i, k \leq n .$$

We resolve ties "lexicographically", always in favor of the lowest index ("name") i .

This results in a partition of $\mathbb{R}^n$ into pairwise-disjoint chambers.

COARSE CHAMBERS (n^2)

Vector Representation as a System of Diffusions

$$\mathrm{d}Y(t) = \mathbf{C}\big(Y(t)\big)\,\mathrm{d}t + \mathbf{S}\big(Y(t)\big)\,\mathrm{d}W(t)\,; \quad 0 \le t < \infty$$

with Interactions of the Mean-Field-Type, but "rough":

$$\mathbf{C}(y) = \sum_{\mathbf{p}\in\Sigma_n} \left(g_{\mathbf{p}^{-1}(1)} + \gamma_1 + \gamma\,, \ldots, \,g_{\mathbf{p}^{-1}(n)} + \gamma_n + \gamma\right)' \cdot \mathbf{1}_{\mathcal{R}_\mathbf{p}}(y)$$

$$= \sum_{k=1}^{n} \Big(\big(g_k + \gamma_1 + \gamma\big) \cdot \mathbf{1}_{Q_k^{(1)}}(y)\,, \ldots\,, \big(g_k + \gamma_n + \gamma\big) \cdot \mathbf{1}_{Q_k^{(n)}}(y) \Big)'\,,$$

$$\mathbf{S}(y) = \sum_{\mathbf{p}\in\Sigma_n} \underbrace{\mathrm{diag}\left(\sigma_{\mathbf{p}^{-1}(1)}\,, \ldots, \,\sigma_{\mathbf{p}^{-1}(n)}\right)}_{\mathfrak{s}_\mathbf{p}} \cdot \mathbf{1}_{\mathcal{R}_\mathbf{p}}(y)\,; \qquad y \in \mathbb{R}^n$$

$$= \mathrm{diag}\left(\sum_{k=1}^{n} \sigma_k \cdot \mathbf{1}_{Q_k^{(1)}}(y)\,, \ldots\,, \sum_{k=1}^{n} \sigma_k \cdot \mathbf{1}_{Q_k^{(n)}}(y) \right).$$

SEMIMARTINGALE REPRESENTATION OF RANKED PROCESSES

Recall $Y_{(1)}(t) \geq \cdots \geq Y_{(n)}(t)$, and denote

$$\boxed{\Lambda^{k,\ell}(t) \;:=\; L^{Y_{(k)}-Y_{(\ell)}}(t)}$$

the local time accumulated at the origin by the semimartingale $Y_{(k)}(\cdot) - Y_{(\ell)}(\cdot) \geq 0$ up to time t, for $1 \leq k < \ell \leq n$.

These are the **collision local times** among particles, of order $\ell - k + 1$: double, if $\ell = k + 1$; triple, if $\ell = k + 1$; and so on.

Lemma: For $k = 1, \ldots, n$, $0 \le t \le T$, we have

$$\mathrm{d}Y_{(k)}(t) = \Big(\gamma + g_k + \sum_{i=1}^{n} \gamma_i \, \mathbf{1}_{Q_k^{(i)}}\big(Y(t)\big) \Big)\, \mathrm{d}t + \sigma_k \, \mathrm{d}B_k(t)$$
$$+ \frac{1}{2} \Big[\mathrm{d}\Lambda^{k,k+1}(t) - \mathrm{d}\Lambda^{k-1,k}(t) \Big],$$

with the independent Brownian Motions (P. LÉVY's theorem)

$$B_k(\cdot) := \sum_{i=1}^{n} \int_0^{\cdot} \mathbf{1}_{Q_k^{(i)}}(Y(t))\, \mathrm{d}W_i(t), \qquad k = 1, \cdots, n.$$

LOCAL TIME

Reminder: The “right" Local Time at the origin, accumulated on $[0, t]$ by a continuous semim'gale $Y(\cdot) = Y(0) + M(\cdot) + V(\cdot)$, is

$$L^Y(t) := Y^+(t) - Y^+(0) - \int_0^t \mathbf{1}_{\{Y(s)>0\}} \, \mathrm{d}Y(s)$$

$$= \lim_{\varepsilon \downarrow 0} \frac{1}{2\,\varepsilon} \int_0^t \mathbf{1}_{\{0 \le Y(s) < \varepsilon\}} \, \mathrm{d}\langle M \rangle(s) \, .$$

The resulting process $L^Y(\cdot)$ is increasing, continuous, flat off the set $\{t \ge 0 : Y(t) = 0\}$.

- If $Y(\cdot) \ge 0$, this becomes

$$L^Y(t) = \int_0^t \mathbf{1}_{\{Y(s)=0\}} \, \mathrm{d}Y(s) = \int_0^t \mathbf{1}_{\{Y(s)=0\}} \, \mathrm{d}V(s) \, .$$

ALGEBRAIC PROPERTIES OF LOCAL TIME

- For continuous semimartingales $Y_1(\cdot), \cdots, Y_n(\cdot)$ we have for the local times at the origin (Yan, Ouknine; mid-80's):

$$L^{Y_1 \wedge Y_2}(t) + L^{Y_1 \vee Y_2}(t) = L^{Y_1}(t) + L^{Y_2}(t), \qquad 0 \le t < \infty$$

Banner & Ghomrasni (2008): More generally,

$$\sum_{k=1}^{n} L^{Y_{(k)}}(t) = \sum_{i=1}^{n} L^{Y_i}(t), \qquad 0 \le t < \infty .$$

• They (B&G) also provide semimartingale representations for the ranked processes in terms of *Collision Local Times:*

$$\mathrm{d}Y_{(k)}(t) = \sum_{i=1}^{n} \mathbf{1}_{Q_k^{(i)}}(Y(t))\, \mathrm{d}Y_i(t)$$

$$+ \sum_{\ell=k+1}^{n} \frac{1}{\mathcal{N}_k(t)}\, \mathrm{d}\Lambda^{k,\ell}(t) - \sum_{\ell=1}^{k-1} \frac{1}{\mathcal{N}_k(t)}\, \mathrm{d}\Lambda^{k,\ell}(t)\,.$$

"Upward pressure" coming from the lower ranks, or "laggards" ($\ell = k+1, \cdots, n$), "downward pressure" from the upper ranks, or "leaders" ($\ell = 1, \cdots, k-1$).

• Here we keep track of the "size of the crowd" in rank k via

$$\mathcal{N}_k(t) := \#\left\{\, i \,:\, Y_i(t) = Y_{(k)}(t) \,\right\};$$

we also assume that all the semimatingales' bounded variation parts are absolutely continuous w.r.t. Lebesgue measure, and that for all (i,j) we have $\mathrm{Leb}\,(\{t \geq 0 : Y_i(t) = Y_j(t)\}) = 0$.

Lemma: For $k = 1, \ldots, n$, $0 \leq t \leq T$, we have

$$\mathrm{d}Y_{(k)}(t) = \Big(\gamma + g_k + \sum_{i=1}^{n} \gamma_i \, \mathbf{1}_{Q_k^{(i)}}\big(Y(t)\big)\Big) \, \mathrm{d}t + \sigma_k \, \mathrm{d}B_k(t)$$
$$+ \frac{1}{2} \Big[\mathrm{d}\Lambda^{k,k+1}(t) - \mathrm{d}\Lambda^{k-1,k}(t) \Big] .$$

Idea of Proof of Lemma: Why only the "nearest neighbor" (or "simple collision") local times $\Lambda^{k,k+1}(\cdot)$ and $\Lambda^{k-1,k}(\cdot)$?

Does this mean that triple (and higher-order) collisions do not occur ? **Far from it**; example of BASS & PARDOUX (1987), where all particles can collide at the origin at once, but can then extricate themselves from this massive collision.

Reason: For any three indices $1 \leq i, j, m \leq n$, the “rank-gap" process

$$\max_{\nu=i,j,m} Y_\nu(\cdot) - \min_{\nu=i,j,m} Y_\nu(\cdot)$$

turns out (some relatively hard work here...) to dominate a Bessel process

$$\mathrm{d}R(t) = \frac{\delta - 1}{2\,R(t)}\,\mathrm{d}t + \mathrm{d}\beta(t)$$

in dimension $\delta > 1$, and analysis of its local time shows

$$L^{Y_{(k)}-Y_{(\ell)}}(\cdot) \equiv \Lambda^{k,\ell}(\cdot) \equiv 0\,, \qquad |\,k-\ell\,| \geq 2\,.$$

Serendipity (and relief): *even if* triple (or higher-order) collisions occur, they just *do not matter* for the respective collision local times.

. Related results in REIMAN & WILLIAMS (1988), and in very recent work with M. SHKOLNIKOV & S. PAL.

Recent work with T. Ichiba & M. Shkolnikov on the absence of triple collisions (PTRF '12, to appear):

A *necessary* condition is the concavity of the graph of the variances

$$k \longmapsto \sigma_k^2 , \qquad k = 1, \cdots , n .$$

. A *sufficient* condition is the concavity of the graph of

$$k \longmapsto \sigma_k^2 , \qquad k = 0, 1, \cdots , n , n+1 ,$$

where we set

$$\sigma_0{}^2 = \sigma_{n+1}^2 = 0 .$$

• Pathwise uniqueness, thus also *strong solvability,* holds for the SDE for $Y(\cdot)$, up until the first time a triple collision occurs.

OPEN QUESTIONS: Is there a condition that is both necessary and sufficient for the absence of such triple collisions? Does the solution 'lose its strength' after the first triple collision?

These Local Times can be estimated...

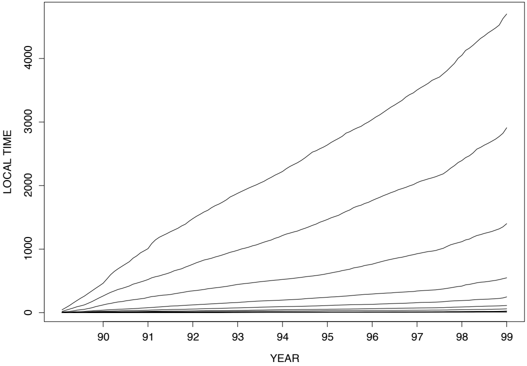


Figure: The estimated local time or "turnover" processes $\Lambda^{k,k+1}(\cdot)$ for $k = 10, 20, 40, \ldots, 5120$; U.S. CRSP data, Jan 1990 – Dec 1999. (From E.R. FERNHOLZ (2002) *Stochastic Portfolio Theory*, page 107.)

Local Times as Cumulative Turnover across Ranks

Discussion: Such estimation comes from the construction of rank-based portfolios that invest in an index-like fashion (according to relative capitalization) in, say, the top k stocks.

The performance of such a portfolio relative to the entire market, involves a **leakage** term proportional to the local time $\Lambda^{k,k+1}(\cdot)$. This leakage measures essentially the **"turnover"** between ranks k and $k+1$; it can then be estimated based on observable quantities.

Please note that this kind of turnover tends to increase, as one goes deeper down the ranks
(that is, with increasing k) , just as the picture suggests.

- The apparent linearity of the growth of local times is yet another indication of an underlying stability or ergodic behavior.

(Recall that for, say, Brownian motion, local time grows like $\sqrt{T}$; whereas for processes with an invariant distribution and stochastic stability, local time grows like T.)

What kinds of conditions can ensure such stochastic stability?

Very roughly speaking: *Assign big growth rates (and big variances) to the smallest stocks; then a stable capital distribution does indeed emerge.*

STABILITY CONDITIONS

In particular, we shall assume, for every $k = 1, \ldots, \mathbf{n-1}$ and $\mathbf{p} \in \Sigma_n$:

$$\boxed{\sum_{k=1}^{n} g_k + \sum_{i=1}^{n} \gamma_i = 0\,, \qquad \sum_{\ell=1}^{k} \left(g_\ell + \gamma_{\mathbf{p}(\ell)}\right) < 0\,.}$$

These conditions ensure that "the cloud of particles" will stick together: no sub-collection of particles can "form its own galaxy", as it were, and drift apart without ever again making contact with the rest.

Example 1 – Atlas model:

$$g_1 = \cdots = g_{n-1} = -\mathfrak{g} < 0;$$
$$g_n = (n-1)\mathfrak{g} > 0;$$
$$\gamma_1 = \cdots = \gamma_n = 0.$$

The company with the lowest capitalization provides all the growth – or support, as with the Titan of mythical lore – for the entire structure. *(Here, companies are totally “anonymous" as far as their growth rates are concerned.)*

Example 2 – Atlas model with stock-specific drifts:

$$g_1, \ldots, g_n \quad \text{as above;} \qquad \sum_{i=1}^{n} \gamma_i = 0\,, \quad \max_{1 \le i \le n} \gamma_i < \mathfrak{g}\,.$$

Now companies can have "eponymous" growth rates; e.g.

$$\gamma_i = \mathfrak{g}\left(1 - \frac{2i}{n+1}\right), \quad 1 \le i \le n\,.$$

STOCHASTIC STABILITY

The average (center of gravity)

$$\overline{Y}(\cdot) := \frac{1}{n} \sum_{i=1}^{n} Y_i(\cdot)$$

of the log-capitalizations

$$\overline{Y}(t) = \overline{Y}(0) + \gamma\, t + \frac{1}{n} \sum_{k=1}^{n} \sigma_k\, B_k(t)$$

is Brownian motion with variance $\sum_{k=1}^{n} (\sigma_k / n)^2$, drift γ .

Recall here the independent Brownian Motions

$$B_k(\cdot) = \sum_{i=1}^{n} \int_0^{\cdot} \mathbf{1}_{Q_k^{(i)}} (Y(t)) \, \mathrm{d} W_i(t) \,, \qquad k = 1, \cdots , n \,.$$

Role of the stability conditions

$$\boxed{\sum_{k=1}^{n} g_k + \sum_{i=1}^{n} \gamma_i = 0\,, \qquad \sum_{\ell=1}^{k} \left(g_\ell + \gamma_{\mathbf{p}(\ell)}\right) < 0\ :}$$

There to guarantee that the process of deviations from the center of gravity

$$\widetilde{Y}(\cdot) := \left(Y_1(\cdot) - \overline{Y}(\cdot), \ldots, Y_n(\cdot) - \overline{Y}(\cdot)\right)$$

is positive recurrent, uniformly over compact sets.

From the theory of R.Z. Khas'minskii (1960, 1980) we have then the following stochastic stability result:

Proposition: The process $\widetilde{Y}(\cdot)$ is stable in distribution; to wit, there is a unique invariant probability measure $\mu(\cdot)$ such that for every bounded, measurable $f : \Pi \to \mathbb{R}$ we have, with $\Pi := \{y \in \mathbb{R}^n : y_1 + \cdots + y_n = 0\}$, the Strong Law of Large Numbers

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T f\left(\widetilde{Y}(t)\right) \mathrm{d}t = \int_\Pi f(y)\,\mu(\mathrm{d}y)\,, \quad a.s.$$

Can this invariant measure be described ?

Average Occupation Times

Setting $f(\cdot) = \mathbf{1}_{\mathcal{R}_\mathbf{p}}(\cdot)$ (respectively, $\mathbf{1}_{Q_k^{(i)}}(\cdot)$), we define the average occupation times of $X(\cdot)$ in the polyhedral chambers $\mathcal{R}_\mathbf{p}$ (respectively, $Q_k^{(i)}$):

$$\theta_\mathbf{p} := \mu(\mathcal{R}_\mathbf{p}) = \lim_{T\to\infty} \frac{1}{T} \int_0^T \mathbf{1}_{\mathcal{R}_\mathbf{p}}(X(t))\,\mathrm{d}t\,, \qquad \mathbf{p} \in \Sigma_n\,,$$

$$\vartheta_{k,i} := \mu(Q_k^{(i)}) = \lim_{T\to\infty} \frac{1}{T} \int_0^T \mathbf{1}_{Q_k^{(i)}}(X(t))\,\mathrm{d}t\,, \quad 1 \le k, i \le n\,.$$

Equilibrium Identity:

$$\boxed{\gamma_i + \sum_{k=1}^{n} g_k\,\vartheta_{k,i} = 0\,; \quad i = 1, \ldots, n\,.}$$

Example 2 – Atlas model with stock-specific drifts:

$$g_1 = \cdots = g_{n-1} = -\mathfrak{g} < 0; \quad g_n = (n-1)\mathfrak{g} > 0\,;$$

$$\sum_{i=1}^{n} \gamma_i = 0\,, \quad \max_{1 \le i \le n} \gamma_i < \mathfrak{g}\,.$$

• In this case, the proportions of time the various stocks occupy the lowest (“Atlas") rank are given by

$$\vartheta_{n,i} \;=\; \frac{1}{n}\left(1 - \frac{\gamma_i}{\mathfrak{g}}\right)\,, \qquad i = 1, \cdots, n\,.$$

We shall obtain more general formulas for these quantities in a short while ... .

Strong Laws of Large Numbers

Stability implies a *SLLN* for Local Times: $\forall\, k = 1, \cdots, n-1$:

$$\lim_{T\to\infty} \frac{1}{T}\, \Lambda^{k,k+1}(T) \;=\; -2\sum_{\ell=1}^{k}\Big(g_\ell + \sum_{i=1}^{n} \vartheta_{\ell,i}\,\gamma_i \Big)$$

$$= -2\sum_{\mathbf{p}\in\Sigma_n} \theta_{\mathbf{p}} \sum_{\ell=1}^{k}\Big(g_\ell + \gamma_{\mathbf{p}(\ell)} \Big) \;>\; 0\,, \qquad \text{a.s.}$$

• Typically, this quantity increases with rank k, much like the picture we saw a moment ago: the higher the rank (to wit: the bigger the k, the smaller the stock in terms of capitalization), the bigger the intensity of “market turnover" around it.

- This will be the case, for instance, under the condition (satisfied in Examples 1, 2):

$$g_k + \gamma_i < 0 \, ; \qquad \forall \;\; 1 \leq k \leq n-1 \, , \;\; 1 \leq i \leq n \, .$$

Together with

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

this condition implies stochastic stability.

What can be said about $\vartheta_{k,i}$ and μ ?

EXAMPLE: Equal Variances, $\gamma = \gamma_1 = \cdots = \gamma_n = 0$

Just a bunch of Brownian motions with drifts determined by their ranks. In this case the equations become

$$\mathrm{d}Y_i(t) = \Big(\sum_{k=1}^{n} g_k \mathbf{1}_{Q_k^{(i)}}(Y(t)) \Big) \mathrm{d}t + \mathrm{d}W_i(t) \; = \; D_i \Phi(Y(t))\, \mathrm{d}t + \mathrm{d}W_i(t)\,.$$

A *conservative diffusion,* with drift given by a conservative vector field and continuous, piecewise smooth potential

$$\Phi(y) \; := \sum_{k=1}^{n} g_k \, y_{(k)}\,, \qquad y \in \mathbb{R}^n\,.$$

The stability conditions imply that $\Phi(\cdot)$ vanishes on the axis $\mathcal{A} := \{ y \in \mathbb{R}^n : y_1 = \cdots = y_n \}$, and

$$\Phi(y) \; = \sum_{k=1}^{n-1} \big(y_{(k)} - y_{(k+1)} \big) \Big(\sum_{\ell=1}^{k} g_\ell \Big) < 0\,, \qquad y \in \mathbb{R}^n \setminus \mathcal{A}\,.$$

Now standard theory shows the existence of invariant measure for the process $Y_{(k)}(\cdot) - Y_{(k+1)}(\cdot)$, $k = 1, \cdots, n-1$ of successive gaps, with unnormalized probability density function in the form of a **product-of-exponentials**

$$e^{2\Phi(y)} = \exp\left\{ -\sum_{k=1}^{n-1} \lambda_k \left(y_{(k)} - y_{(k+1)} \right) \right\},$$

with (the stability conditions once again!)

$$\lambda_k := -2 \sum_{\ell=1}^{k} g_\ell > \mathbf{0}, \qquad k = 1, \cdots, n-1 \ .$$

(Independence of successive gaps. Reversibility.)

In reality: Variances are not equal, but rather grow with rank (the smaller the stock, the more volatile it tends to be). And of course, growth rates should depend on name as well as rank...

.

Linearly Growing Variances

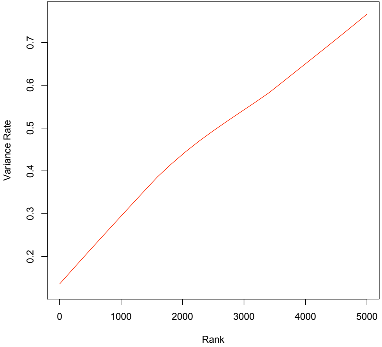


Figure: Smoothed variance by rank, U.S. Equity market, 1990-1999.

We shall assume that variances grow linearly with rank:

$$\sigma_2^2 - \sigma_1^2 = \sigma_3^2 - \sigma_2^2 = \cdots = \sigma_n^2 - \sigma_{n-1}^2 \geq 0\,.$$

SEMIMARTINGALE REFLECTED BROWNIAN MOTIONS

Recall the ranked semimartingale decomposition

$$\mathrm{d}Y_{(k)}(t) = \sum_{i=1}^{n} \mathbf{1}_{Q_k^{(i)}}(Y(t))\,\mathrm{d}Y_i(t) + \frac{1}{2}\left[\mathrm{d}\Lambda^{k,k+1}(t) - \mathrm{d}\Lambda^{k-1,k}(t)\right]$$

of BANNER & GHOMRASNI (2008). Equivalently:

$$\mathrm{d}Y_{(k)}(t) = \left(\gamma + g_k + \sum_{i=1}^{n} \gamma_i\,\mathbf{1}_{Q_k^{(i)}}\big(Y(t)\big)\right)\mathrm{d}t + \sigma_k\,\mathrm{d}B_k(t)$$
$$+ \frac{1}{2}\left[\mathrm{d}\Lambda^{k,k+1}(t) - \mathrm{d}\Lambda^{k-1,k}(t)\right].$$

The vector $\Xi(\cdot)$ of "**Successive Gaps**"

$$\Xi_k(\cdot) := Y_{(k)}(\cdot) - Y_{(k+1)}(\cdot) \geq 0\,, \qquad k = 1, \cdots, n-1$$

then satisfies

$$\mathrm{d}\,\Xi_k(t) \;=\; \Big(g_k + \sum_{i=1}^{n} \gamma_i \, \mathbf{1}_{Q_k^{(i)}}\big(Y(t)\big)\Big)\,\mathrm{d}t + \sigma_k\,\mathrm{d}B_k(t)$$

$$-\Big(g_{k+1} + \sum_{i=1}^{n} \gamma_i \, \mathbf{1}_{Q_{k+1}^{(i)}}\big(Y(t)\big)\Big)\,\mathrm{d}t - \sigma_{k+1}\,\mathrm{d}B_{k+1}(t)$$

$$-\frac{1}{2}\Big[\mathrm{d}\Lambda^{k-1,k}(\cdot) + \mathrm{d}\Lambda^{k+1,k+2}(\cdot)\Big] + \mathrm{d}\Lambda^{k,k+1}(\cdot)\,.$$

It is a **semimartingale reflected Brownian motion** in the nonnegative orthant $\mathbb{R}_+^{n-1}$ (HARRISON, REIMAN, WILLIAMS).

- Finally, we define the *indicator map* $\boxed{\mathbb{R}^n \ni \xi \mapsto \mathfrak{p}^{\xi} \in \Sigma_n}$

$$\xi_{\mathfrak{p}^{\xi}(1)} \geq \xi_{\mathfrak{p}^{\xi}(2)} \geq \cdots \geq \xi_{\mathfrak{p}^{\xi}(n)}\,, \quad \text{so that} \quad \mathfrak{p}^{\xi} = \mathbf{p} \iff \xi \in \mathcal{R}_{\mathbf{p}}\,,$$

where $\mathfrak{p}^{\xi}(k)$ is the name (index) of the coördinate that occupies the k^{th} rank among $\xi_1, \cdots, \xi_n$.

. We introduce also the **Index Process**

$$\boxed{\mathfrak{P}_t := \mathfrak{p}^{Y(t)}} \qquad 0 \leq t < \infty\,,$$

with values in the symmetric group Σ_n. The definition implies

$$Y_{\mathfrak{P}_t(1)} = Y_{(1)}(t) \geq \cdots \geq Y_{(n)}(t) = Y_{\mathfrak{P}_t(n)}\,, \qquad 0 \leq t < \infty\,.$$

Keeps track of "who is sitting in a particular rank k at any given time".

Invariant Distribution for Adjacent Gaps and Indices

Proposition: Under the stability and linearly-growing-variance conditions, the invariant distribution $\nu(\cdot)$ of $(\Xi(\cdot), \mathfrak{P}.)$ is

$$\nu(A \times B) = \Big(\sum_{\mathbf{p} \in \Sigma_n} \prod_{k=1}^{n-1} \lambda_{\mathbf{p},k}^{-1} \Big)^{-1} \cdot \sum_{\mathbf{p} \in B} \int_A \exp\big(- \langle \lambda_{\mathbf{p}}, z \rangle \big) \, \mathrm{d}z$$

for every measurable set $A \times B \in (\mathbb{R}_+)^{n-1} \times \Sigma_n$.

Here $\lambda_{\mathbf{p}} := (\lambda_{\mathbf{p},1}, \ldots, \lambda_{\mathbf{p},n-1})'$ is the vector with components

$$\lambda_{\mathbf{p},k} := \frac{-2 \sum_{\ell=1}^{k} \big(g_\ell + \gamma_{\mathbf{p}(\ell)} \big)}{(\sigma_k^2 + \sigma_{k+1}^2)/2} > 0 \, ; \quad \mathbf{p} \in \Sigma_n \, , \; 1 \le k \le n-1 \, .$$

Please compare with expression on slide 31.

Discussion: The invariant measure $\nu(\cdot\,,\cdot)$ of $(\Xi(\cdot), \mathfrak{P}_\cdot)$ satisfies the **"Basic Adjoint Relationship"** (BAR) of HARRISON & WILLIAMS (1987) (chamber-by chamber, then globally thanks to the linearly-growing-variance condition).

The particular form of $\nu(\cdot\,,\cdot)$ leads to the density

$$\mathbb{P}\big(\Xi(t) \in A\big) = \Big(\sum_{\mathbf{p} \in \Sigma_n} \prod_{k=1}^{n-1} \lambda_{\mathbf{p},k}^{-1} \Big)^{-1} \cdot \sum_{\mathbf{p} \in \Sigma_n} \int_A \exp(-\langle \lambda_{\mathbf{p}}, z \rangle)\, \mathrm{d}z$$

of sums-of-products-of-exponentials type, for the distribution (under the invariant measure $\nu(\cdot\,,\cdot)$) of the semimartingale reflected Brownian motion process

$$\Xi(\cdot) := (\Xi_1(\cdot), \ldots, \Xi_{n-1}(\cdot))'$$

of adjacent gaps

$$\Xi_k(\cdot) := Y_{(k)}(\cdot) - Y_{(k+1)}(\cdot) \geq 0\,, \qquad k = 1, \ldots, n-1\,.$$

Discussion (cont'd): The assumption of *linearly growing variances* is crucial in the Proposition.

It guarantees that the structural **"Skew-Symmetry Condition"** (SSC) is satisfied, and that the process of adjacent gaps

$$\Xi(\cdot) = (\Xi_1(\cdot), \ldots, \Xi_{n-1}(\cdot))'$$

actually never visits the nonsmooth part of the boundary of the positive orthant (R. WILLIAMS (1987)).

This condition also implies the **absence of triple collisions** *for the components of the original process* $Y(\cdot)$. Special case of a theory developed by T. ICHIBA (2009) in his dissertation, concerning the absence of triple collisions.

Comment: With $\mathfrak{D}$ the diagonal matrix of the covariance matrix $\mathfrak{A} = \{a_{k\ell}\}_{1 \le k,\ell \le n-1}$ with

$$a_{k\ell} := \left(\sigma_k^2 + \sigma_{k+1}^2\right) \mathbf{1}_{\{\ell=k\}} - \sigma_k^2 \, \mathbf{1}_{\{\ell=k-1\}} - \sigma_{k+1}^2 \, \mathbf{1}_{\{\ell=k+1\}} ,$$

and with the $(n-1) \times (n-1)$ "reflection matrix" (slide 34)

$$\mathbf{R} := \begin{pmatrix} 1 & -1/2 & & & \\ -1/2 & 1 & -1/2 & & \\ & \ddots & \ddots & \ddots & \\ & & -1/2 & 1 & -1/2 \\ & & & -1/2 & 1 \end{pmatrix} ,$$

the **Skew-Symmetry Condition** (SSC) mandates

$$\boxed{2\left(\mathfrak{D} - \mathfrak{A}\right) = \left(\mathbf{I} - \mathbf{R}\right)\mathfrak{D} + \mathfrak{D}\left(\mathbf{I} - \mathbf{R}\right) .}$$

A compatibility condition between the covariance and the reflection matrix - which ordinarily ‘fight each other’. When it prevails, peace is restored; it is satisfied in the case of linearly growing variances.

The components of the column $\varrho_k \in \mathbb{R}^{n-1}$ of the reflection matrix

$$\mathbf{R} = \begin{pmatrix} 1 & -1/2 & & & \\ -1/2 & 1 & -1/2 & & \\ & \ddots & \ddots & \ddots & \\ & & -1/2 & 1 & -1/2 \\ & & & -1/2 & 1 \end{pmatrix},$$

provide the *(non-tangential!)* directions of reflection, when the face of the boundary

$$\mathfrak{F}_k := \left\{ \left(z_1, \cdots, z_{n-1} \right)' \,|\, z_k = 0 \right\}, \qquad k = 1, \ldots, n-1$$

of the state-space $\mathfrak{S} = (\mathbb{R}_+)^{n-1}$ is hit and the k^{th} component of $\Lambda(\cdot)$ increases.

The **BAR (Basic Adjoint Relationship)** is

$$\int_{\mathfrak{S}\times\Sigma_n} \big[\mathcal{A}(\mathbf{p})\, f\,\big](z)\,\mathrm{d}\nu(z,\mathbf{p}) + \frac{1}{2}\sum_{k=1}^{n-1}\int_{\mathfrak{F}_k} \langle \varrho_k\,, \nabla f(z)\rangle(z)\,\mathrm{d}\nu_{0k}(z) = 0$$

for $f \in \mathcal{C}^2(\mathfrak{S})$, where

$$\big[\mathcal{A}(\mathbf{p})\, f\,\big](z) \;:=\; \frac{1}{2}\sum_{k=1}^{n-1}\sum_{\ell=1}^{n-1} \mathrm{a}_{k,\ell}\, \frac{\partial^2 f(z)}{\partial z_k \partial z_\ell} + \sum_{k=1}^{n-1} b_k(\mathbf{p})\, \frac{\partial f(z)}{\partial z_k}\,,$$

$$\mathrm{a}_{k\ell} \;:=\; \big(\sigma_k^2 + \sigma_{k+1}^2\big)\,\mathbf{1}_{\{\ell=k\}} \;-\; \sigma_k^2\,\mathbf{1}_{\{\ell=k-1\}} \;-\; \sigma_{k+1}^2\,\mathbf{1}_{\{\ell=k+1\}}\,,$$

$$b_k(\mathbf{p}) \;:=\; \big(g_k + \gamma_{\mathbf{p}^{-1}(k)}\big) - \big(g_{k+1} + \gamma_{\mathbf{p}^{-1}(k+1)}\big)\,.$$

Average Occupation Times

Corollary: The long-term-average occupation times are

$$\boxed{\theta_{\mathbf{p}} = \mu(\mathcal{R}_{\mathbf{p}}) = \nu(\mathfrak{S}, \{\mathbf{p}\}) = \Big(\sum_{\mathbf{q}\in\Sigma_n} \prod_{k=1}^{n-1} \lambda_{\mathbf{q},k}^{-1} \Big)^{-1} \cdot \prod_{k=1}^{n-1} \lambda_{\mathbf{p},k}^{-1}}$$

for each chamber $\mathcal{R}_{\mathbf{p}}$ $(\mathbf{p} \in \Sigma_n)$, and

$$\underbrace{\vartheta_{k,i} = \sum_{\{\mathbf{p}\in\Sigma_n : \mathbf{p}(k)=i\}} \theta_{\mathbf{p}}}, \qquad i = 1, \ldots, n.$$

Please recall

$$\lambda_{\mathbf{p},k} := \frac{-2\sum_{\ell=1}^{k} \big(g_\ell + \gamma_{\mathbf{p}(\ell)}\big)}{(\sigma_k^2 + \sigma_{k+1}^2)/2} > 0\,; \quad \mathbf{p} \in \Sigma_n\,,\ 1 \le k \le n-1\,.$$

These DO satisfy (sanity check) the equilibrium identities

$$\gamma_i + \sum_{k=1}^{n} g_k \, \vartheta_{k,i} \; = \; 0 \, ; \quad i = 1, \ldots, n \, .$$

- If all $\gamma_i = 0$, then

$$\vartheta_{k,i} = \frac{1}{n} \quad \text{for} \quad 1 \leq k, i \leq n$$

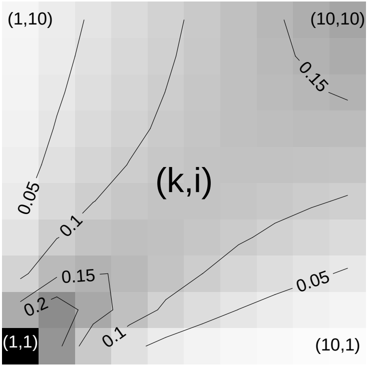
(first-order model of BFK (2005), includes the simple Atlas model as a special case).

- Heat map of $\vartheta_{k,i}$ when $n = 10$, $\sigma_k^2 = 1 + k$,

  $g_k = -1$ for $1 \leq k \leq 9$, $g_{10} = 9$, and

  $\gamma_i = 1 - (2i)/(n+1)$

  for $i = 1, \ldots, 10$.



Distribution of Ranked Market Weights

Corollary: The invariant distribution of ranked market weights

$$M_{(k)}(\cdot) := \frac{X_{(k)}(\cdot)}{X_1(\cdot) + \cdots + X_n(\cdot)} \; ; \quad k = 1, \ldots, n$$

has probability density function $\wp(m_1, \ldots, m_{n-1})$ given by

$$\boxed{\sum_{\mathbf{p} \in \Sigma_n} \theta_{\mathbf{p}} \frac{\lambda_{\mathbf{p},1} \cdots \lambda_{\mathbf{p},n-1}}{m_1^{\lambda_{\mathbf{p},1}+1} \cdot m_2^{\lambda_{\mathbf{p},2}-\lambda_{\mathbf{p},1}+1} \cdots m_{n-1}^{\lambda_{\mathbf{p},n-1}-\lambda_{\mathbf{p},n-2}+1} \, m_n^{-\lambda_{\mathbf{p},n-1}+1}}} \, ,$$

$$0 < m_n \leq m_{n-1} \leq \ldots \leq m_1 < 1 \, , \;\; m_n := 1 - (m_1 + \cdots + m_{n-1}) \, .$$

- This is a distribution of ratios of Pareto type.

In the special case

$$\gamma_1 = \cdots = \gamma_n = 0$$

it takes the far more appealing form

$$\wp(m_1, \ldots, m_{n-1}) = \prod_{k=1}^{n-1} \lambda_k \, m_k^{1+\lambda_k - \lambda_{k-1}} \big(1 - m_1 - \cdots - m_{n-1}\big)^{1-\lambda_{n-1}}$$

with

$$\lambda_k := \frac{(-4)\sum_{\ell=1}^{k} g_\ell}{\sigma_k^2 + \sigma_{k+1}^2}\,, \qquad \lambda_0 = \lambda_n = 0\,.$$

CAPITAL DISTRIBUTION CURVES

$$\wp(m_1, \ldots, m_{n-1}) =$$

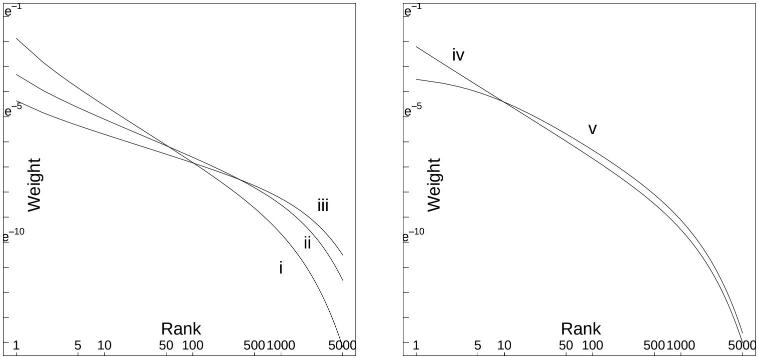
$$= \sum_{\mathbf{p} \in \Sigma_n} \theta_{\mathbf{p}} \frac{\lambda_{\mathbf{p},1} \cdots \lambda_{\mathbf{p},n-1}}{m_1^{\lambda_{\mathbf{p},1}+1} \cdot m_2^{\lambda_{\mathbf{p},2}-\lambda_{\mathbf{p},1}+1} \cdots m_{n-1}^{\lambda_{\mathbf{p},n-1}-\lambda_{\mathbf{p},n-2}+1} m_n^{-\lambda_{\mathbf{p},n-1}+1}}$$

The invariant probability density for the ranked market weights from the previous slide, allows us to describe the long term average (and "expected") **slope** of the capital distribution curve at the various ranks k , thus also its shape:

$$\lim_{T \to \infty} \frac{1}{T} \int_0^T \frac{\log M_{(k+1)}(t) - \log M_{(k)}(t)}{\log(k+1) - \log k} \, \mathrm{d}t =$$

$$\mathbb{E}^\nu \left(\frac{\log M_{(k+1)} - \log M_{(k)}}{\log(k+1) - \log k} \right) = \frac{-\mathbb{E}^\nu(\Xi_k)}{\log(1+k^{-1})} = -\frac{\sum_{\mathbf{p} \in \Sigma_n} \theta_{\mathbf{p}} \lambda_{\mathbf{p},k}^{-1}}{\log(1+k^{-1})} .$$

Illustrations



- $n = 5000$, $g_n = c_*(2n-1)$, $g_k = 0$, $1 \le k \le n-1$, $\gamma_1 = -c_*$, $\gamma_i = -2c_*$, $2 \le i \le n$, $\sigma_k^2 = 0.075 + 6k \times 10^{-5}$, $1 \le k \le n$. (i) $c_* = 0.02$, (ii) $c_* = 0.03$, (iii) $c_* = 0.04$.
- (iv) $c_* = 0.02$, $g_1 = -0.016$, $g_k = 0$, $2 \le k \le n-1$, $g_n = (0.02)(2n-1) + 0.016$,
- (v) $g_1 = \cdots = g_{50} = -0.016$, $g_k = 0$, $51 \le k \le n-1$, $g_n = (0.02)(2n-1) + 0.8$.

Parameter Estimation

in the context of these models: massive... .
Some preliminary results can be found in this paper

FERNHOLZ, E.R., ICHIBA, T. & KARATZAS, I. (2012) A second-order stock market model. *Annals of Finance,* in press,

but a lot more work remains to be done.

These kinds of model have very good micro- and macro-scopic behavior, but lousy meso-scopic behavior. Models that try to amend this drawback are being developed by
• V. Papathanakos (using interesting SPDE theory); also by
• J. Teichmann and Ph.D. students in his group at ETH (using the infinite-dimensional “Wasserstein diffusions"); also in
• Joint work with S. Pal and M. Shkolnikov (asymmetric collisions).

Connections

This theory allows us to compute growth-optimal and universal portfolios, for long-term money management (almost tailor-made for the "Empirical Bayes" theory of COVER (1991), JAMSHIDIAN (1992); this is another story, and talk...).

The theory we presented relies heavily on

- The "semimartingale reflecting Brownian motion" analysis of Queueing Networks in their heavy traffic limit approximation (J.M. HARRISON, M. REIMAN, R. WILLIAMS),

and has strong connections with

• The combinatorial analysis of interacting diffusions based on COXETER groups (e.g., groups of orthogonal matrices generated by reflections);

• The theory of POISSON-DIRICHLET distributions for the market weights (CHATTERJEE AND PAL);

• Propagation of chaos results of SZNITMAN, JOURDAIN, MALRIEU (all in the case of equal variances); and with

• Discrete-time models of competing particle systems in Statistical Mechanics, such as SHERRINGTON-KIRKPATRICK models of spin glasses, with similar invariant distributions (M. AIZENMAN, A. RUZMAIKINA, P.L. ARGUIN).

Refinements

Now let the number $n \to \infty$ of particles increase to infinity in the model with $\gamma_i \equiv 0$.

• The market weights exhibit phase transitions: depending on the size of a parameter, they either: (i) all converge to zero in probability; (ii) all but one converge to zero in probability; (iii) the point process generated by them converges weakly to a POISSON-DIRICHLET distribution (CHATTERJEE AND PAL);

• The empirical measure of the "configuration of particles" is characterized by evolution equations of the MCKEAN-VLASOV type (but very non-smooth), and by partial differential equations of the **porous medium** form with convection (recent work of M. SHKOLNIKOV at Stanford/Berkeley/INTECH).

• Concentration of Measure results (T. ICHIBA, S. PAL, M. SHKOLNIKOV, to appear in PTRF).

• Large Deviation results (A. DEMBO, M. SHKOLNIKOV, S. VARADHAN, O. ZEITOUNI,very recent).

Some References

BANNER, A., FERNHOLZ, E.R., & KARATZAS, I. (2005) Atlas models of equity markets. *Annals of Applied Probability* **15**, 2296-2330.

ICHIBA, T. & KARATZAS, I. (2010) On collisions of Brownian particles. *Annals of Applied Probability* **20**, 951-977.

ICHIBA, T., PAPATHANAKOS, V., BANNER, A., KARATZAS, I. & FERNHOLZ, E.R. (2011) Hybrid Atlas Models. *Annals of Applied Probability* **11**, 609-644.

ICHIBA, T., KARATZAS, I. & SHKOLNIKOV, M. (2011) Strong solutions of stochastic equations with rank-based coëfficients. *Probability Theory and Related Fields,* to appear. Available at *arxiv.org/pdf/1109.3823*.

HARRISON, J.M. & WILLIAMS, R. (1987) Brownian models of open queueing networks with homogeneous customer populations. *Stochastics* **22**, 77-115.

HARRISON, J.M. & WILLIAMS, R. (1987) Multidimensional reflected Brownian motions having exponential stationary distributions. *Annals of Probability* **15**, 115-137.

WILLIAMS, R. (1987) Reflected Brownian motion with skew symmetric data in a polyhedral domain. *Probability Theory and Related Fields* **75**, 459-485.

KHAS'MINSKII, R. (1980) *Stochastic Stability of Differential Equations.* Sijthoff and Noordhoff, The Netherlands.

YAN, J.A. (1985) A formula for local times of semimartingales. *Northeast Mathematical Journal* **1**, 138-110.

OUKNINE, Y. (1990) Local times of the product and the maximum of two semimartingales. *Lecture Notes in Mathematics* **1426**, 477-479.

BANNER, A. & GHOMRASNI, R. (2008) Local times of ranked continuous semimartingales. *Stochastic Processes and Their Applications* **118**, 1244-1253.

BASS, R. & PARDOUX, E. (1987) Uniqueness for diffusions with piecewise constant coëfficients. *Probability Theory and Related Fields* **76**, 557-572.

PAL, S. & PITMAN, J. (2008) One-dimensional Brownian particle systems with rank-dependent drifts. *Annals of Applied Probability* **18**, 2179-2207.

CHATTERJEE, S. & PAL, S. (2010) A phase transition behavior for Brownian motions interacting through their ranks. *Probability Theory and Related Fields* **147**, 123-159.

CHATTERJEE, S. & PAL, S. (2011) A combinatorial analysis of interacting diffusions. *Journal of Theoretical Probability,* to appear.

SHKOLNIKOV, M. (2012) Large systems of diffusions interacting through their ranks. *Stochastic Processes and Their Applications* **122**, 1730-1747.

ICHIBA, T., PAL, S. & SHKOLNIKOV, M. (2011) Concentration of measure for Brownian particle systems interacting through their ranks. *Probability Theory and Related Fields,* to appear.

COVER, T. (1991) Universal Portfolios. *Mathematical Finance* **1**, 1-29.

JAMSHIDIAN, F. (1992) Asymptotically Optimal Portfolios. *Mathematical Finance* **2**, 131-150.

KARATZAS, I., PAL, S. & SHKOLNIKOV, M. (2012) Brownian particles with asymmetric collisions. *Submitted for publication.*