A weak hidden Markov chain-modulated model for asset allocation

Rogemar S. Mamon[1,3]
University of Western Ontario
Canada

Joint work with **Matt Davison**[1,2,3] and **Jean Xi**[3]
[1] *Department of Statistical & Actuarial Sciences, UWO*

[2] *Richard Ivey School of Business, UWO*

[3] *Department of Applied Mathematics, UWO*

25 April 2012

OBJECTIVES

1. Develop a **stochastic model** that captures the most important features of the evolution of asset spot prices.

2. Estimate **efficiently** and **dynamically** the parameters of the proposed model.

3. Extend modelling framework and estimation algorithm to **multivariate data series**. Consider application to **asset allocation problem**.

Overview

- Financial model parameters change over time. Markov-switching models are deemed to be able to incorporate dynamics of business cycles and volatility regimes.

- Either the mean or variance level, or both may exhibit differences amongst regimes.

- Financial time series data possess some memories (or dependencies), i.e., current values are not necessarily independent from past observed values.

- The usual Markov assumption is inadequate. Weak Markov model or the so-called higher-order HMM offers more flexibility.

What are hidden Markov models (HMMs)?

These refer to models with states and measurements in a discrete set (state space) and discrete/continuous time.

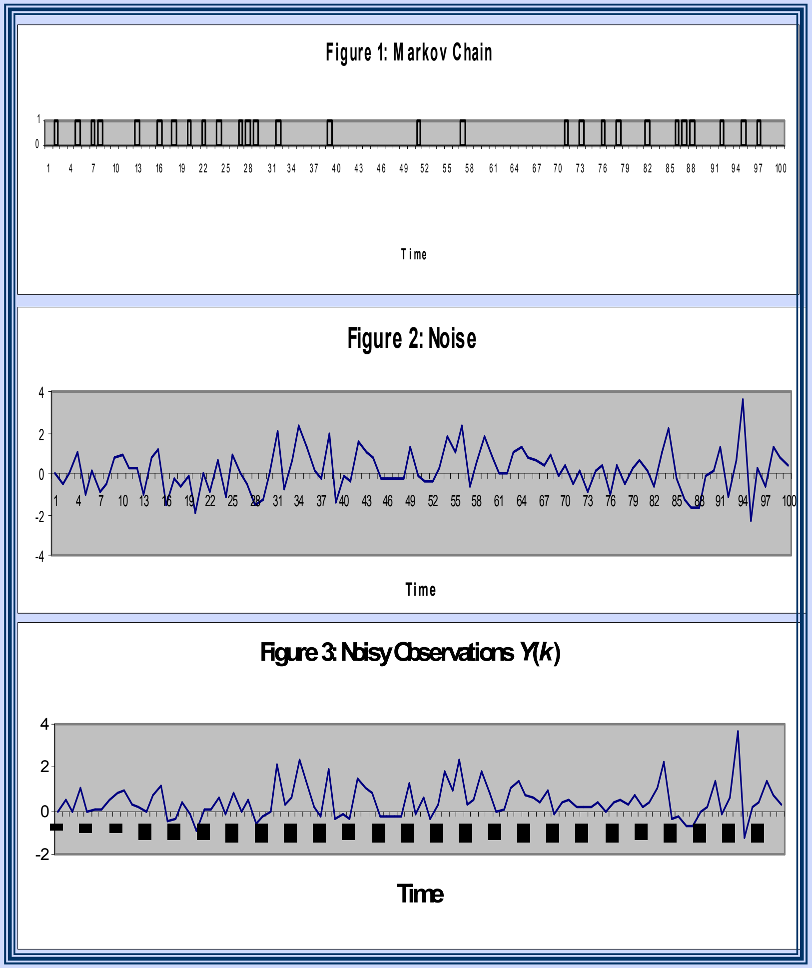
HMM theory addresses two important problems:
Estimation – signal filtering, model parameter identification, state estimation, signal smoothing and signal prediction.

Filtered estimates are estimates at time k of X_k based on processing past and present measurements $\{y_1, y_2, \ldots, y_k\}$. Smoothed estimates are estimates at time k of X_k based on processing past, present, and future measurements $\{y_1, \ldots, y_k, \ldots, y_m\}$ with $m > k$. Forecast/predicted estimates are estimates at time $k + h$ of X_k based on processing past and present measurements $\{y_1, \ldots, y_k\}$.

Control - selection of actions which effect the signal-generating system in such a way as to achieve certain control objectives.

ILLUSTRATION: X_k (k=1, 2, …) - a message sequence of 0's and 1's. Binary signal X (a Markov chain) is transmitted on a noisy communications channel. When signal is detected at receiver, we get some result Y_k.

Consequence: X_k is a Markov chain hidden in noise.



Aim: Develop optimal estimation algorithms for HMMs to remove the random noise in the "best" possible way.

HMM in financial modelling

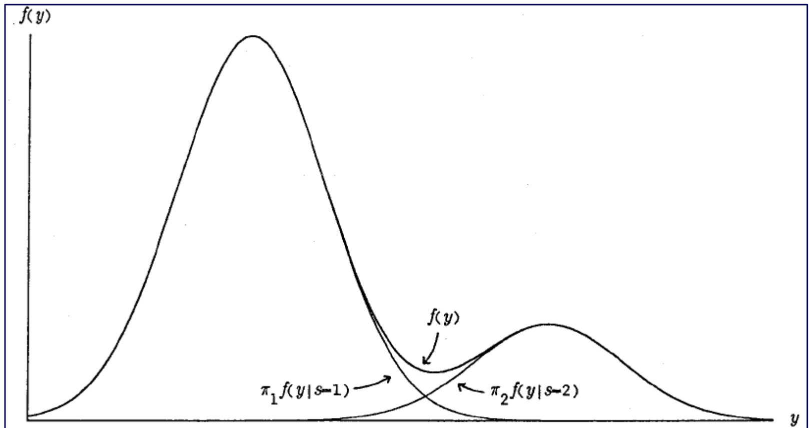
In Electrical Engineering: Charge $Q(t)$, at time t at a fixed point in an electrical circuit is of interest. However, due to error in measurement of $Q(s)\ \ (s < t)$, it cannot really be measured but rather just a noisy version of it.
The objective is to filter the noise out of our observations.

In Finance: Do financial data, interest rates, asset price processes, exchange rates, commodity prices, etc contain information about latent variables? *If so, how might their behaviour in general and in particular their dynamics be estimated?*

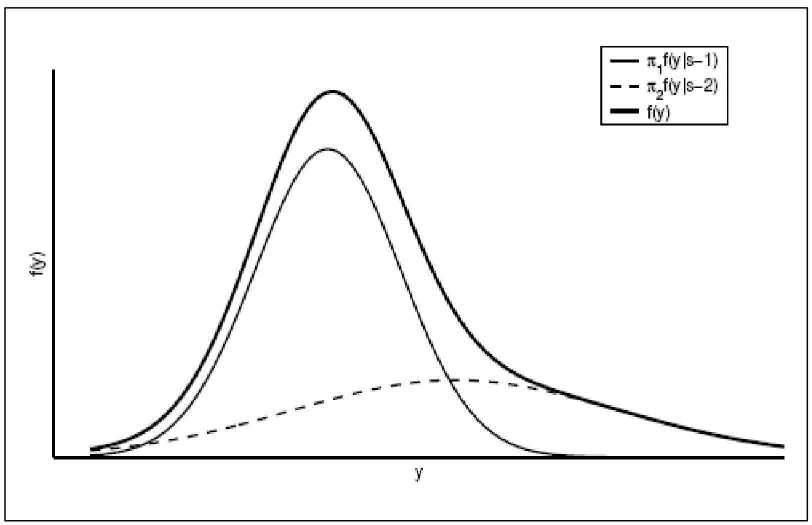
HMM and MARKOV SWITCHING MODELS (MSM)

RATIONALE FOR MSM:

- Market may switch between, say, a "quiet" (stable low volatility) and a "turbulent" (unstable high volatility).
  -incorporates **stochastic volatility** in a simple way
- Provides more flexibility to financial models
  -produces a mixture of distributions allowing kurtosis or skewness different from a single Gaussian distribution.

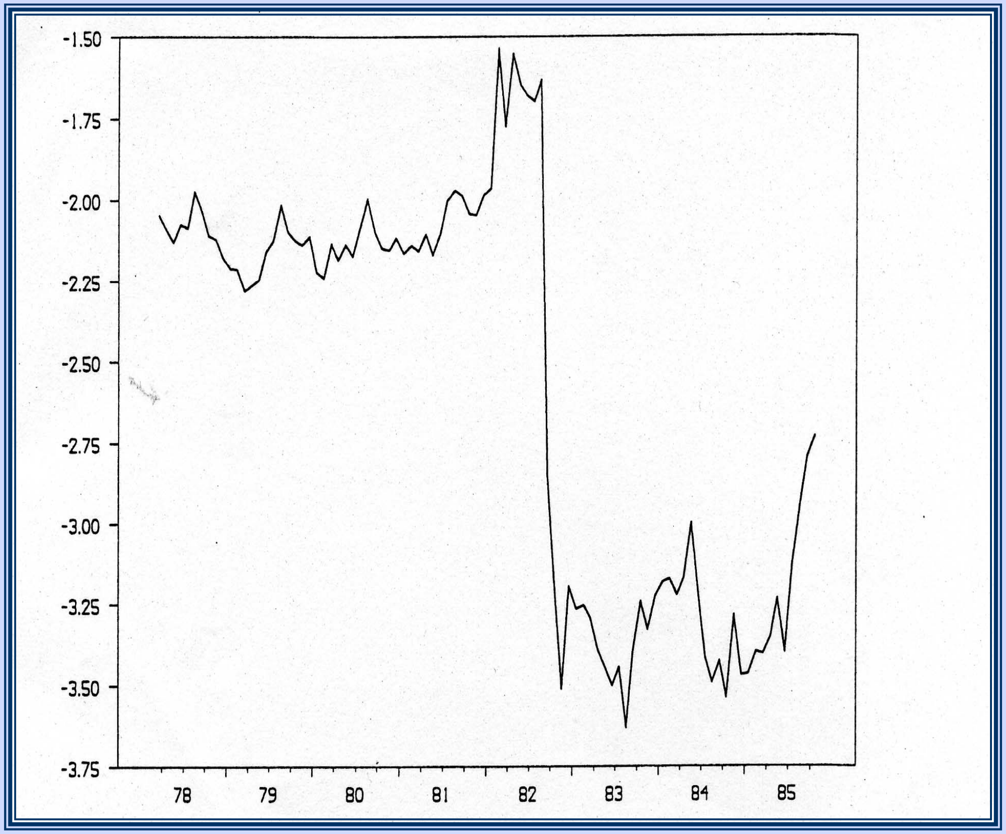


Density of mixtures of 2 Gaussian distributions with $y_t|s_t$=1~N(0,1), $y_t|s_t$ =2~N(4,2), and P(s_t = 1)=0.8



Density of mixtures of 2 Gaussian distributions with $y_t|s_t$=1~N(0,1), $y_t|s_t$ =2~N(2,8), and P(s_t = 1)=0.6

A process with **regime-switching behaviour.**



The monthly log ratio of peso value of **dollar-denominated bank accounts** in Mexico to the peso value of the **peso-denominated bank accounts, 1978-85**

Source: Rogers, J. (1992) The currency substitution hypothesis and relatively money demand in Mexico and Canada, ***Journal of Money, Credit and Banking*** 24(3), 300-318.

Some previous works on (usual) Markov-switching models

■ Quandt (1958) and Goldfield & Quandt (1973)
Early works showing idea of regime-switching model

■ Hamilton (1989)
Markov-switching methods in modelling nonstationary time series

■ Hundreds of papers have been written highlighting regime-switching models in recent years. Applications to option pricing, FX rate market, commodity markets, bond pricing, etc. See Nieh, et al. (2010)

Dynamic optimal estimation of parameters

■ Continual, recursively self-updating estimates for various types of models with HMM-driven parameters. These are advanced in Elliott, et al (1995), Elliott & van der Hoek (2003), Elliott, et al. (2003), Mamon, et al. (2007), Erlwein & Mamon (2009), Erlwein, et al. (2011), amongst others

The weak hidden Markov model (WHMM)

▶ WHMM enhances HMM's power and sophistication.

▶ *Drawbacks:* (i) Enlargement of the number of states (ii) Parameter estimation becomes more involved.

▶ Some works on WHMM: Siu, et al. (2005), Ching, et al. (2007), Siu, et al. (2009)

− research developing modelling frameworks for risk management, computation of credit ratings and valuation of derivatives. They do not, however, deal with estimation applied to data.

This work

▶ proposes a model that captures regime-switching and memory of asset price data (both univariate and multivariate);

▶ develops recursive estimates for model parameters via adaptive filters of MC state and auxiliary processes; and

▶ includes an empirical implementation using Russell 3000 data focusing on asset allocation problem.

REMARKS:

⊙ The state of MC represents the state of the economy.

⊙ The state of MC is *hidden* in the observation process (log-returns of spot price process).

MODEL FORMULATION & DESCRIPTION

Underlying processes are defined on $(\Omega, \mathcal{F}, P)$.

The usual HMM case

$\mathbf{x}_k$ is a homogeneous MC with finite state in discrete time ($k = 0,\ 1,\ 2,\ \ldots$).

State space of $\boldsymbol{x}$ is $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n\} \in \Re^N$,
where $\boxed{\boldsymbol{e}_i = (0, \ldots, 0, 1, 0, \ldots, 0)^\top}$, $\top$ is the transpose of a vector.

Regular/usual MC satisfies

$$\begin{aligned}&P(\mathbf{x}_{k+1} = x_{k+1} | \mathbf{x}_0 = x_0, \mathbf{x}_1 = x_1, \ldots, \mathbf{x}_{k-1} = x_{k-1}, \mathbf{x}_k = x_k)\\&= P(\mathbf{x}_{k+1} = x_{k+1} | \mathbf{x}_k = x_k).\end{aligned}$$

It has semi-martingale representation

$$\boxed{\mathbf{x}_{k+1} = \Pi \mathbf{x}_k + \mathbf{v}_{k+1}},$$

where $\mathbf{\Pi}$ is a transition matrix and $\mathbf{v}_{k+1}$ is a martingale increment.

Description of WHMM

The process $\mathbf{x}$ is a *weak (hidden) MC* of order n, $n \geq 1$, if its value at present time k depends on its value in the previous n times steps, i.e.,

$$P(\mathbf{x}_{k+1} = x_{k+1} | \mathbf{x}_0 = x_0, \mathbf{x}_1 = x_1, \ldots, \mathbf{x}_{k-1} = x_{k-1}, \mathbf{x}_k = x_k)$$
$$= P(\mathbf{x}_{k+1} = x_{k+1} | \mathbf{x}_{k-n+1} = x_{k-n+1}, \ldots, \mathbf{x}_{k-1} = x_{k-1}, \mathbf{x}_k = x_k).$$

Note: When $n = 1$, the regular MC is recovered.

Remark: To simplify the discussion, we focus on a weak MC of order 2 so that

$$P(\mathbf{x}_{k+1} = x_{k+1} | \mathbf{x}_0 = x_0, \mathbf{x}_1 = x_1, \ldots, \mathbf{x}_{k-1} = x_{k-1}, \mathbf{x}_k = x_k)$$
$$= P(\mathbf{x}_{k+1} = x_{k+1} | \mathbf{x}_{k-1} = x_{k-1}, \mathbf{x}_k = x_k).$$

Write $\boxed{a_{lmv} := P(\mathbf{x}_{k+1} = \mathbf{e}_l | \mathbf{x}_k = \mathbf{e}_m, \mathbf{x}_{k-1} = \mathbf{e}_v)}$,

where $v, m, l \in \{1, 2, \ldots, N\}$. The associated $N \times N^2$ transition matrix is

$$\mathbf{A} = \begin{pmatrix} a_{111} & a_{112} & \cdots & a_{11N} & \cdots & a_{1N1} & \cdots & a_{1NN} \\ a_{211} & a_{212} & \cdots & a_{21N} & \cdots & a_{2N1} & \cdots & a_{2NN} \\ & & \cdots & & & & \cdots & \\ a_{N11} & a_{N12} & \cdots & a_{N1N} & \cdots & a_{NN1} & \cdots & a_{NNN} \end{pmatrix}.$$

$y_k :=$ log-return of asset prices.

<u>$\mathbf{x}$ is not observed directly from financial market</u>.

There exists a function h with values in a finite set s.t.

$$\boxed{y_{k+1} = h(\mathbf{x}_k, z_{k+1}) = f(\mathbf{x}_k) + \sigma(\mathbf{x}_k) z_{k+1}, \;\; k \geq 1},$$

where $\{z_k\}_{k \geq 1}$ is a sequence of IID standard normal RVs independent of $\mathbf{x}$.

Assume there are vectors $\mathbf{f} = (f_1, f_2, \ldots, f_N)^\top$ and $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \ldots, \sigma_N)^\top$.

Mean of $y_k = f(\mathbf{x}_k) = \langle \mathbf{f}, \mathbf{x}_k \rangle$
Volatility of $y_k = \sigma(\mathbf{x}_k) = \langle \boldsymbol{\sigma}, \mathbf{x}_k \rangle$ and $\sigma_i > 0$ for every $1 \leq i \leq N$.

$\mathcal{F}_k :=$ complete filtration generated by $\mathbf{x}$
$\mathcal{Y}_k :=$ complete filtration generated by y and
$\mathcal{H}_k := \mathcal{F}_k \vee \mathcal{Y}_k$ complete filtration generated by $\mathbf{x}$ and y.

Main idea of constructing filtering equations for WHMM:
Embed the 2nd-order MC into a 1st-order MC and then apply ordinary methods.

Define a mapping α by: $\alpha\binom{\mathbf{e}_r}{\mathbf{e}_s} = \mathbf{e}_{rs}$, for $1 \leq r, s \leq N$, where $\mathbf{e}_{rs}$ is a unit vector in $\mathbb{R}^{N^2}$ with 1 in its $((r-1)N + s)$th position.

Transforming WHMM into a regular HMM

The mapping α groups 2 time steps of $\mathbf{x}$ to form a new 1st-order MC.

$\left\langle \alpha\binom{\mathbf{x}_k}{\mathbf{x}_{k-1}}, \mathbf{e}_{rs} \right\rangle = \langle \mathbf{x}_k, \mathbf{e}_r \rangle \langle \mathbf{x}_{k-1}, \mathbf{e}_s \rangle \equiv$ identification of $\mathbf{x}$ at current and previous time steps with the canonical basis of $\mathbb{R}^{N^2}$.

Under P, the weak Markov chain $\mathbf{x}$ has semi-martingale representation

$$\boxed{\alpha\binom{\mathbf{x}_k}{\mathbf{x}_{k-1}} = \Pi \alpha \binom{\mathbf{x}_{k-1}}{\mathbf{x}_{k-2}} + \mathbf{v}_k},$$

where $\{\mathbf{v}_k\}_{k \geq 1}$ is a sequence of $\mathbb{R}^{N^2}$-martingale increments and Π is an $N^2 \times N^2$ matrix that can be constructed from $\mathbf{A}$.

Converting WHMM into HMM: 2-state case

Consider a 2-state, 2nd-order MC.

- $\alpha\binom{\mathbf{x}_k}{\mathbf{x}_{k-1}}$ corresponds to a 4-state, 1st-order MC.

- 4 states: (1,1), (1,2), (2,1) and (2,2).

Mapping produces $\alpha(1,1)$ =state 1, $\alpha(1,2)$ =state 2, $\alpha(2,1)$ =state 3 and $\alpha(2,2)$ =state 4.

- $\mathbf{\Pi}$ can be constructed from matrix $\mathbf{A}$.

$$\mathbf{\Pi} = \begin{bmatrix} \pi_{11} & \pi_{12} & \pi_{13} & \pi_{14} \\ \pi_{21} & \pi_{22} & \pi_{23} & \pi_{24} \\ \pi_{31} & \pi_{32} & \pi_{33} & \pi_{34} \\ \pi_{41} & \pi_{42} & \pi_{43} & \pi_{44} \end{bmatrix} = \begin{bmatrix} a_{111} & a_{112} & \boxed{0} & 0 \\ 0 & 0 & a_{121} & a_{122} \\ a_{211} & a_{212} & 0 & 0 \\ 0 & 0 & a_{221} & a_{222} \end{bmatrix},$$

where $\boxed{a_{rst} = P(\mathbf{x}_k = r, \mathbf{x}_{k-1} = \boxed{s} | \mathbf{x}_{k-1} = \boxed{s}, \mathbf{x}_{k-2} = t)}$.

Note: It is **not possible**, for instance, to make the transition (1,1) $\Leftarrow$ (2,1) [i.e., going to state 1 from state 3] in 2 steps; hence $\pi_{13} = 0$. Thus, *there are structural zeros for some entries of $\mathbf{\Pi}$ matrix.*

Change of Reference Probability Measure

By Kolmogorov's Extension Theorem, $\exists\ \bar{P}$, an ideal measure, under which observed data are independent (calculations are easier to perform).

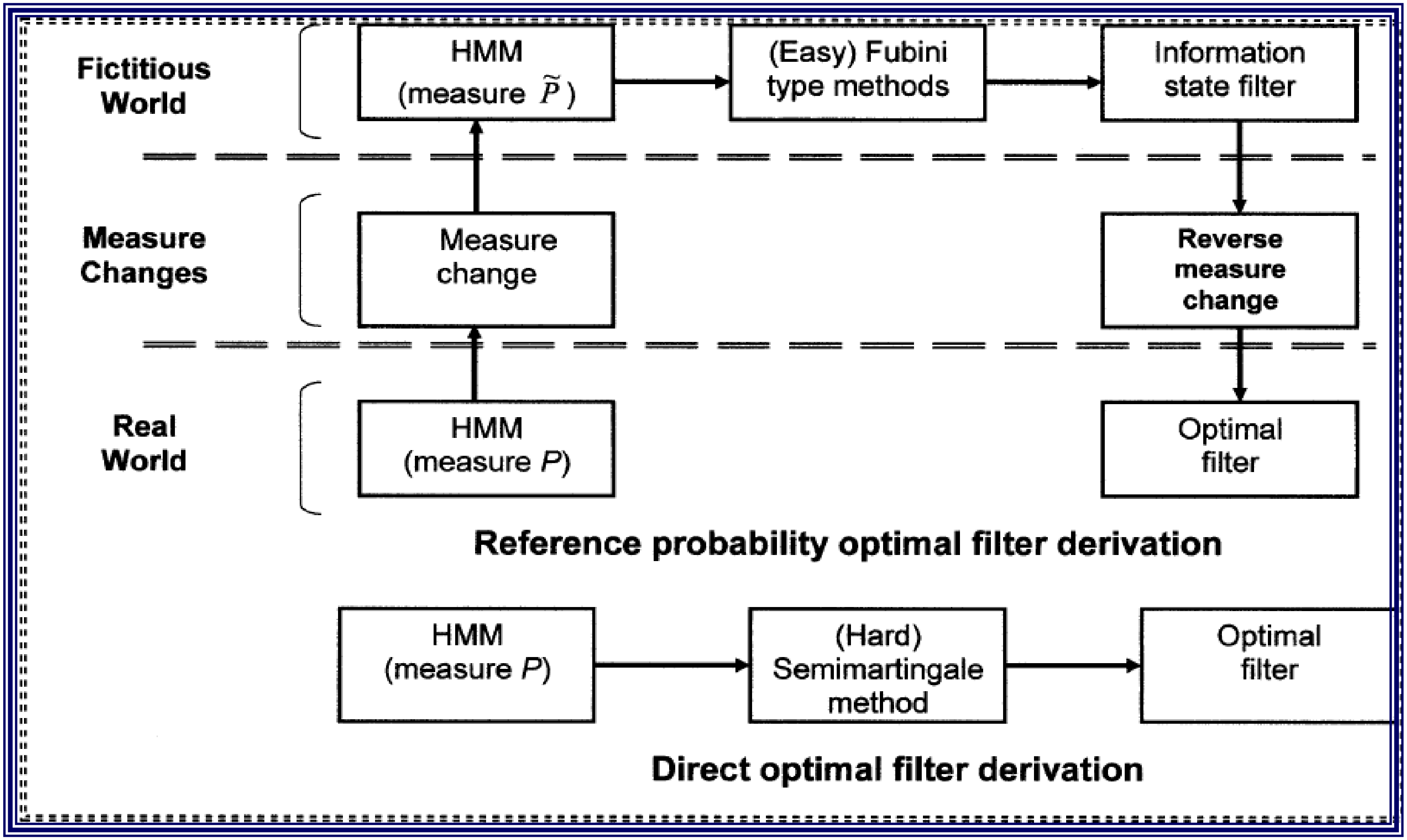
Under $\bar{P}$,
(i) $\{y_k\}_{k\geq 1}$:= sequence of $N(0,1)$ IID RVs and independent from $\mathbf{x}_k$, and
(ii) $\{\mathbf{x}_k\}_{k\geq 0}$:= weak MC s.t. semi-martingale representation holds & $\bar{E}[\mathbf{v}_k|\mathcal{F}_k] = \mathbf{0}$.

Write $\phi(z)$:= pdf of a standard normal RV z. To construct P from $\bar{P}$, define processes λ_l and Λ_k by

$$\lambda_l = \frac{\phi(\sigma(\mathbf{x}_{l-1})^{-1}(y_l - f(\mathbf{x}_{l-1})))}{\sigma(\mathbf{x}_{l-1})\phi(y_l)}, \quad \text{and} \quad \boxed{\Lambda_k = \prod_{l=1}^{k} \lambda_l,\ k \geq 1,\ \Lambda_0 = 1}.$$

Set $\left.\frac{dP}{d\bar{P}}\right|_{\mathcal{H}_k} = \Lambda_k$. Calculations are performed under $\bar{P}$ and use Bayes' thm for conditional expectations to relate $\bar{P}$ and P.

Implementation of HMM filtering technique



$\mathbf{p}_k :=$ conditional distribution of $\alpha\binom{\mathbf{x}_k}{\mathbf{x}_{k-1}}$ given $\mathcal{Y}_k$ under P.

$\mathbf{p}_k = \left(p_k^{11}, \ldots, p_k^{ij}, \ \ldots, p_k^{NN}\right)^\top$, $1 \le i, j \le N$ is a $\mathbb{R}^{N^2}$ vector and

$$p_k^{ij} = P(\mathbf{x}_k = \mathbf{e}_i, \mathbf{x}_{k-1} = \mathbf{e}_j | \mathcal{Y}_k) = E\left[\left\langle \alpha\binom{\mathbf{x}_k}{\mathbf{x}_{k-1}}, \mathbf{e}_{ij} \right\rangle \middle| \mathcal{Y}_k\right].$$

Write $\boxed{\mathbf{q}_k := \bar{E}\left[\Lambda_k \alpha\binom{\mathbf{x}_k}{\mathbf{x}_{k-1}} \middle| \mathcal{Y}_k\right]}$.

Since

$$\sum_{i,j=1}^{N} \left\langle \alpha\binom{\mathbf{x}_k}{\mathbf{x}_{k-1}}, \mathbf{e}_{ij} \right\rangle = 1,$$

$$\boxed{\langle \mathbf{q}_k, \mathbf{1} \rangle = \bar{E}\left[\Lambda_k \left\langle \alpha\binom{\mathbf{x}_k}{\mathbf{x}_{k-1}}, \mathbf{1} \right\rangle \middle| \mathcal{Y}_k\right] = \bar{E}[\Lambda_k | \mathcal{Y}_k]}.$$

Note that by Bayes' theorem $\boxed{\mathbf{p}_k = E\left[\alpha\binom{\mathbf{x}_k}{\mathbf{x}_{k-1}} \middle| \mathcal{Y}_k\right] = \dfrac{\bar{E}\left[\Lambda_k \alpha\binom{\mathbf{x}_k}{\mathbf{x}_{k-1}} \middle| \mathcal{Y}_k\right]}{\bar{E}[\Lambda_k | \mathcal{Y}_k]}}$.

Consequently, $\boxed{\mathbf{p}_k = \dfrac{\mathbf{q}_k}{\langle \mathbf{q}_k, \mathbf{1} \rangle}}$.

Recursive filter for the process $\mathbf{q}_k$: Let $\mathbf{B}_k$ be the $N^2 \times N^2$ diagonal matrix at time k formulated as

$$\mathbf{B}_k = \begin{pmatrix} b_k^1 & & & & & & \\ & \ddots & & & & & \\ & & b_k^1 & & & & \\ & & & \ddots & & & \\ & & & & b_k^N & & \\ & & & & & \ddots & \\ & & & & & & b_k^N \end{pmatrix},$$

where

$$b_k^i = \frac{\phi(\sigma_i^{-1}(y_k - f_i))}{\sigma_i \phi(y_k)}.$$

A recursive expression for $\mathbf{q}_k$ is

$$\boxed{\mathbf{q}_k = \mathbf{B}_k \mathbf{\Pi} \mathbf{q}_{k-1}}.$$

Recursive filters for 4 other processes related to the MC are also derived.

(i) $J_k^{rst} \equiv$ no. of jumps from state $(\mathbf{e}_t, \mathbf{e}_s)$ to $(\mathbf{e}_s, \mathbf{e}_r)$ up to time k,

$$J_k^{rst} = \sum_{l=2}^{k} \langle \mathbf{x}_l, \mathbf{e}_r \rangle \langle \mathbf{x}_{l-1}, \mathbf{e}_s \rangle \langle \mathbf{x}_{l-2}, \mathbf{e}_t \rangle.$$

(ii) $O_k^{rs} \equiv$, the occupation, or amt. of time spent up to time k, of the weak Markov chain $\mathbf{x}$ in state $(\mathbf{e}_r, \mathbf{e}_s)$,

$$O_k^{rs} = \sum_{l=2}^{k} \langle \mathbf{x}_{l-1}, \mathbf{e}_r \rangle \langle \mathbf{x}_{l-2}, \mathbf{e}_s \rangle.$$

(iii) $O_k^{r} \equiv$ occupation, or amt. of time spent up to time k, of the weak Markov chain $\mathbf{x}$ in state $\mathbf{e}_r$,

$$O_k^{r} = \sum_{l=1}^{k} \langle \mathbf{x}_{l-1}, \mathbf{e}_r \rangle.$$

(iv) $T_k^{r}(g) \equiv$ level sum for state $\mathbf{e}_r$,

$$T_k^{r}(g) = \sum_{l=1}^{k} g(y_l) \langle \mathbf{x}_{l-1}, \mathbf{e}_r \rangle.$$

and g is a function of form $g(y) = y$ or $g(y) = y^2$.

For any $\mathcal{Y}$-adapted process X, write $\widehat{X}_k := E[X|\mathcal{Y}_k]$ and $\gamma(X)_k := \bar{E}[\Lambda_k X|\mathcal{Y}_k]$ for the filter of X. From Bayes' theorem,

$$\boxed{\widehat{J}_k^{rst} = \frac{\bar{E}[\Lambda_k J_k^{rst}|\mathcal{Y}_k]}{\bar{E}[\Lambda_k|\mathcal{Y}_k]} = \frac{\bar{E}[\Lambda_k J_k^{rst}|\mathcal{Y}_k]}{\langle \mathbf{q}_k, \mathbf{1}\rangle}}.$$

Rather than estimating quantities J_k^{rst}, O_k^{rs}, O_k^{r} and $T_k^r(g)$ directly, recursive forms can be found for related product-quantities:

$$J_k^{rst}\alpha\binom{\mathbf{x}_k}{\mathbf{x}_{k-1}},\ O_k^{rs}\alpha\binom{\mathbf{x}_k}{\mathbf{x}_{k-1}},\ O_k^{r}\alpha\binom{\mathbf{x}_k}{\mathbf{x}_{k-1}} \text{ and } T_k^r(g)\alpha\binom{\mathbf{x}_k}{\mathbf{x}_{k-1}}.$$

Relating vector and scalar quantities of interest

$$\left\langle \gamma\left(J^{rst}\alpha\binom{\mathbf{x}_k}{\mathbf{x}_{k-1}}\right)_k, \mathbf{1}\right\rangle = \bar{E}\left[\Lambda_k J_k^{rst}\left\langle \alpha\binom{\mathbf{x}_k}{\mathbf{x}_{k-1}}, \mathbf{1}\right\rangle \middle| \mathcal{Y}_k\right] \text{ and so}$$

$$\boxed{\left\langle \gamma\left(J^{rst}\alpha\binom{\mathbf{x}_k}{\mathbf{x}_{k-1}}\right)_k, \mathbf{1}\right\rangle = \bar{E}\left[\Lambda_k J_k^{rst}\middle|\mathcal{Y}_k\right] = \gamma(J^{rst})_k}.$$

Recursive filters for vector processes

Let $\mathbf{V}_r, 1 \le r \le N$ be an $N^2 \times N^2$ matrix such that the $((i-1)N+r)$th column of $\mathbf{V}_r$ is $\mathbf{e}_{ir}$ for $i = 1 \dots N$ and zero elsewhere. If $\mathbf{B}$ is the diagonal matrix defined in Eq.(), then

$$
\begin{aligned}
\gamma\Big(J^{rst}\alpha\tbinom{\mathbf{x}_{k+1}}{\mathbf{x}_k}\Big)_{k+1} &= \mathbf{B}_{k+1}\mathbf{\Pi}\gamma\Big(J^{rst}\alpha\tbinom{\mathbf{x}_k}{\mathbf{x}_{k-1}}\Big)_k + b^r_{k+1}\pi_{rst}\langle \mathbf{q}_k, \mathbf{e}_{st}\rangle \mathbf{e}_{rs}, \\
\gamma\Big(O^{rs}\alpha\tbinom{\mathbf{x}_{k+1}}{\mathbf{x}_k}\Big)_{k+1} &= \mathbf{B}_{k+1}\mathbf{\Pi}\gamma\Big(O^{rs}\alpha\tbinom{\mathbf{x}_k}{\mathbf{x}_{k-1}}\Big)_k + \mathbf{B}_{k+1}\mathbf{\Pi}\mathbf{e}_{rs}\langle \mathbf{q}_k, \mathbf{e}_{rs}\rangle, \\
\gamma\Big(O^{r}\alpha\tbinom{\mathbf{x}_{k+1}}{\mathbf{x}_k}\Big)_{k+1} &= \mathbf{B}_{k+1}\mathbf{\Pi}\gamma\Big(O^{r}\alpha\tbinom{\mathbf{x}_k}{\mathbf{x}_{k-1}}\Big)_k + \mathbf{V}_r\mathbf{B}_{k+1}\mathbf{\Pi}\mathbf{q}_k, \\
\gamma\Big(T^{r}(g)\alpha\tbinom{\mathbf{x}_{k+1}}{\mathbf{x}_k}\Big)_{k+1} &= \mathbf{B}_{k+1}\mathbf{\Pi}\gamma\Big(T^{r}(g)\alpha\tbinom{\mathbf{x}_k}{\mathbf{x}_{k-1}}\Big)_k + g(y_{k+1})\mathbf{V}_r\mathbf{B}_{k+1}\mathbf{\Pi}\mathbf{q}_k.
\end{aligned}
$$

We employ the Expectation-Maximisation (EM) algorithm (Dempster, Laird & Rubin, 1977) to find the parameters.

If $\{\widehat{a}_{rst},\ \widehat{f}_r,\ \widehat{\sigma}_r\}$ determines the model then EM estimates are

$$\boxed{\widehat{a}_{rst} = \frac{\widehat{J}_k^{rst}}{\widehat{O}_k^{st}} = \frac{\gamma(J^{rst})_k}{\gamma(O^{st})_k},\ \forall \text{ pairs } (r,s), r \neq s},$$

$$\boxed{\widehat{f}_r = \frac{\widehat{T}_k^r}{\widehat{O}_k^r} = \frac{\gamma(T^r(y))_k}{\gamma(O^r)_k}},$$

$$\boxed{\widehat{\sigma}_r^2 = \frac{\widehat{T}_k^r(y^2) - 2\widehat{f}_r\widehat{T}_k^r(y) + \widehat{f}_r^2\widehat{O}_k^r}{\widehat{O}_k^r},\ \text{and } \widehat{\sigma}_r = \sqrt{\widehat{\sigma}_r^2}}.$$

Implementation procedure

▶ Initialise f_r and σ_r, $r = 1, \ldots, N$. Initial value for non-zero entries in $\mathbf{\Pi}$ is $1/N$.

▶ Data is processed in batches of 10 observation points constituting one complete algorithm step or pass. At the end of each step, new estimates for $\mathbf{f}$, $\boldsymbol{\sigma}$ and $\mathbf{A}$ are computed. We reconstruct $\mathbf{A}$ to get $\mathbf{\Pi}$.

▶ New estimates are used iteratively to give updated values of the parameters. Frequency of parameter updating depends on nature of observation data. A moving window of 10 weekly data points in this multi-pass procedure appears sufficient for our asset allocation application.

WHMM: The Multivariate Case

Let $\mathbf{y}_k = (y_k^1, y_k^2, \ldots, y_k^d)$ be a $d-$dim process. Each y_k^g, $1 \le g \le d$ has dynamics

$$\boxed{y_{k+1}^g = f^g(\mathbf{x}_k) + \sigma^g(\mathbf{x}_k) z_{k+1}^g.}$$

$\{z_k^g\}$ is a sequence of N(0,1) IID RVs and independent from $\boldsymbol{x}$. Functions f^g and σ^g are determined by
$\mathbf{f}^g = (f_1^g, f_2^g, \ldots, f_N^g)^\top$ and
$\boldsymbol{\sigma}^g = (\sigma_1^g, \sigma_2^g, \ldots, \sigma_N^g)\top$, respectively.
Specifically, $f^g(\mathbf{x}_k) = \langle \mathbf{f}^g, \mathbf{x}_k \rangle$ and $\sigma_k^g(\boldsymbol{x}_k) = \langle \boldsymbol{\sigma}^g, \mathbf{x}_k \rangle$.

The relevant Girsanov density in changing from P to $\bar{P}$ is

$$\boxed{\Lambda_k = \prod_{g=1}^{d} \prod_{l=1}^{k} \lambda_l^g, \quad k \ge 1, \quad \Lambda_0 = 1,}$$

$$\text{where} \lambda_l^g = \frac{\phi(\sigma^g(\mathbf{x}_{l-1})^{-1}(y_l^g - f^g(\boldsymbol{x}_{l-1})))}{\sigma^g(\mathbf{x}_{l-1})\phi(y_l^g)}.$$

The EM estimates in the multivariate case

$$\boxed{\hat{a}_{rst} = \frac{\hat{J}_k^{rst}}{\hat{O}_k^{st}} = \frac{\gamma(J^{rst})_k}{\gamma(O^{st})_k}, \ \forall \text{ pairs } (r,s), r \neq s},$$

$$\boxed{\hat{f}_r^g = \frac{\hat{T}_k^r}{\hat{O}_k^r} = \frac{\gamma(T^r(y^g))_k}{\gamma(O^r)_k}},$$

$$\boxed{\hat{\sigma}_r^g = \sqrt{\frac{\hat{T}^r((y^g)^2)_k - 2\hat{f}_r^g\hat{T}^r((y^g))_k + \hat{f}_r^2\hat{O}_k^r}{\hat{O}_k^r}}}.$$

Standard errors for parameter estimates can be obtained from *Fisher Information matrix* $\mathbf{I}$ (negative expectation of the second derivative of log-likelihood functions for each parameter). MLEs are asymptotically normally distributed with variance $\mathbf{I}^{-1}$.

Forecasting indices

▶ We apply our filtering results to weekly datasets of stock indices: Russell 3000 growth and value indices [*June 1995 to December 2010 (783 data points each dataset)*]. Russell 3000 represents 98% of the investable US equity market.

▶ h-step ahead forecast of the price/index level:

$$\boxed{E[S_{k+h}|\mathcal{Y}_k] = S_k \sum_{i,j=1}^{N} \langle \mathbf{\Pi}^{h-1}\mathbf{p}_k, \mathbf{e}_{ij}\rangle \exp\left(f_i + \frac{\sigma_i^2}{2}\right)}$$

Goodness of fit: Error measures for one-step ahead forecasts under 1-, 2- and 3-state WHMM set-ups

	1-state WHMM	2-state WHMM	3-state WHMM
APE value	0.020	0.016	0.028
MAE value	38.610	36.958	50.556
RMSE value	54.394	51.480	66.867
APE growth	0.023	0.018	0.030
MAE growth	42.889	39.209	54.477
RMSE growth	64.280	61.486	76.577

Error analysis (along with the Akaike information criterion, AIC $= -2\log(\theta) + 2s$, $\log(\theta)$=likelihood function and s =no. of parameters) confirms support for the two-state WHMM.

A switching investment strategy (2-state WHMM)

▶ An investor *chooses between two investments* to diversify his risk. Growth and value stocks tend to perform well at different times of economic cycle. Switching between two classes at appropriate times may add value. This strategy *utilises forecasted risk-adjusted returns of the indices as signals* to switch investments between Russell 3000 growth and Russell 3000 value indices.

▶ Dataset is *divided into 15 subintervals* (containing roughly one-year data). Assume a *starting investment of $100.* Full amount is *invested in the index with higher forecasted risk-adjusted return* at the beginning of each interval. *Transaction cost* (fixed percentage of total investment) is subtracted from total investment.

▶ *Overall performance* of switching strategy is *compared with pure investment strategy* on basis of log-return of terminal wealth. We consider differences of log-returns:

$$\boxed{X_i^{RG} = \log \frac{SW_i}{100} - \log \frac{RG_i}{100} \quad \text{and} \quad X_i^{RV} = \log \frac{SW_i}{100} - \log \frac{RV_i}{100}}.$$

Performance comparison between WHMM- and HMM-based switching strategies with varying transaction costs (1bp=0.01%)

Transaction Cost	5 bps		20 bps		50 bps		70 bps	
	WHMM	HMM	WHMM	HMM	WHMM	HMM	WHMM	HMM
Mean (X^{RG}) %	3.911	-0.690	3.516	-0.928	2.726	-1.406	2.1980	-1.726
Std (X^{RG}) %	2.410	6.714	2.470	6.588	2.709	6.360	2.940	6.226
Mean (X^{RV}) %	11.239	-9.529	10.845	-9.768	10.055	-10.246	9.527	-10.565
Std (X^{RV}) %	18.470	15.855	18.286	16.005	17.929	16.312	17.700	16.523

▶ On average, log-returns from switching strategy are higher than those from pure index investment. However, high SDs indicate high risk in switching strategy.

A mixed investment strategy (2-state WHMM)

▶ Investor determines optimal weights of each asset to allocate between growth and value stocks based on estimated parameters and state of WHMM.

▶ Mean-variance problem: This involves solving the optimal $\mathbf{w} = (w_g,\ w_v)$ which maximises the function

$$MV(\mathbf{w}) = \nu E[w_g y_{k+1}^{RG} + w_v y_{k+1}^{RV} | \mathscr{Y}_k] - \mathrm{Var}[w_g y_{k+1}^{RG} + w_v y_{k+1}^{RV} | \mathscr{Y}_k],$$

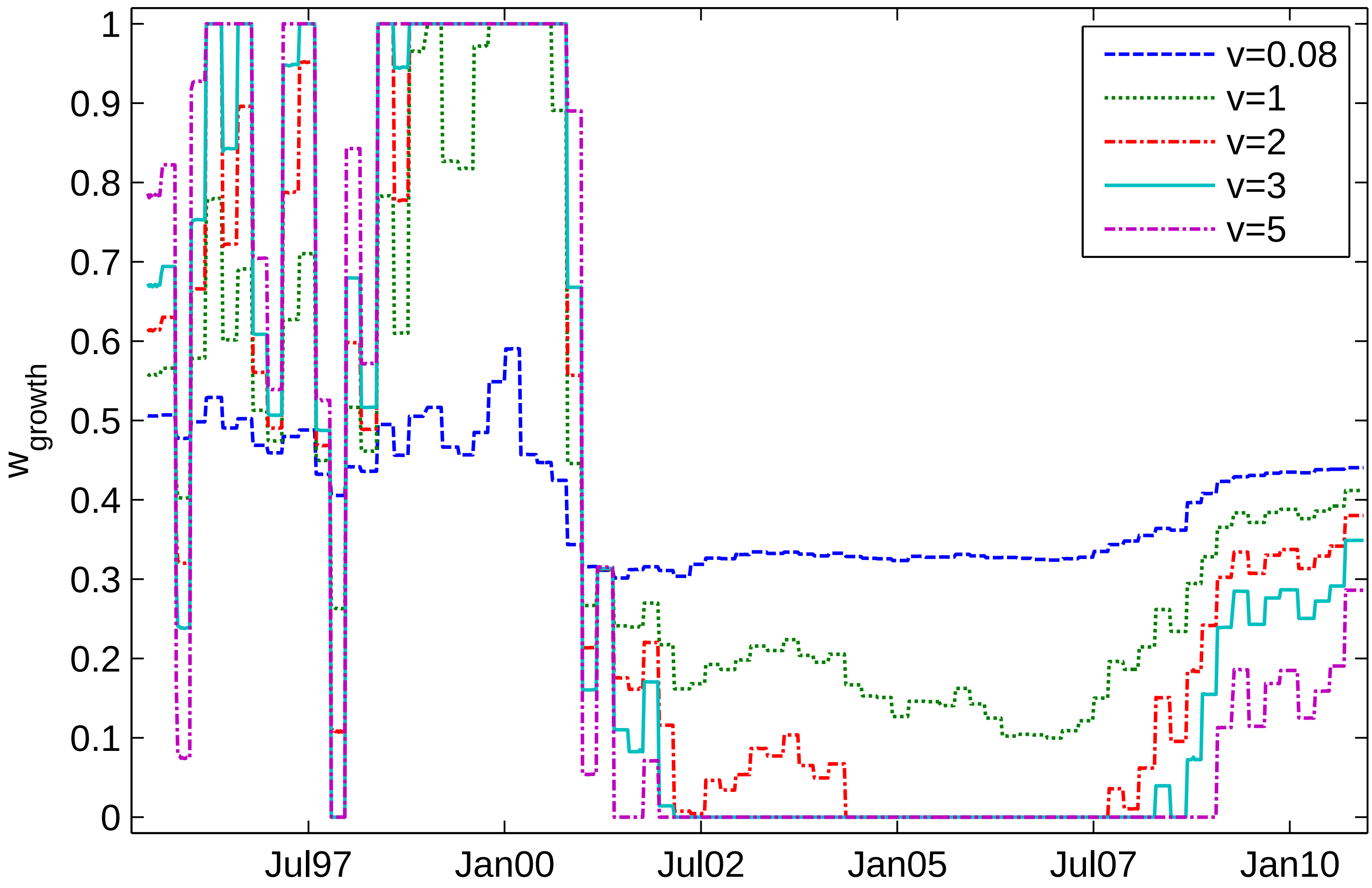
and ν is a nonnegative risk aversion factor.

▶ The optimal weight w_g is given by

$$w_g = \begin{cases} \frac{v(\langle \mathbf{f}^{RG}, \hat{\mathbf{x}}_k\rangle - \langle \mathbf{f}^{RV}, \hat{\mathbf{x}}_k\rangle) + 2\langle \boldsymbol{\sigma}^{RV}, \hat{\mathbf{x}}_k\rangle^2}{2(\langle \boldsymbol{\sigma}^{RG}, \hat{\mathbf{x}}_k\rangle^2 + \langle \boldsymbol{\sigma}^{RV}, \hat{\mathbf{x}}_k\rangle^2)} & \text{when } -2\langle \boldsymbol{\sigma}^{RV}, \hat{\mathbf{x}}_k\rangle^2 < v(\langle \mathbf{f}^{RG}, \hat{\mathbf{x}}_k\rangle - \langle \mathbf{f}^{RV}, \hat{\mathbf{x}}_k\rangle) \\ & < 2\langle \boldsymbol{\sigma}^{RG}, \hat{\mathbf{x}}_k\rangle^2 \\ 1 & \text{when } v(\langle \mathbf{f}^{RG}, \hat{\mathbf{x}}_k\rangle - \langle \mathbf{f}^{RV}, \hat{\mathbf{x}}_k\rangle) > 2\langle \boldsymbol{\sigma}^{RG}, \hat{\mathbf{x}}_k\rangle^2 \\ 0 & \text{when } v(\langle \mathbf{f}^{RG}, \hat{\mathbf{x}}_k\rangle - \langle \mathbf{f}^{RV}, \hat{\mathbf{x}}_k\rangle) < -2\langle \boldsymbol{\sigma}^{RV}, \hat{\mathbf{x}}_k\rangle^2 \end{cases},$$

and $w_v = 1 - w_g$. Proof is similar to the HMM case given in Erlwein, et al. (2011).

Evolution of optimal weights for Russell 3000 growth index in the WHMM-based mixed strategy with varying ν's



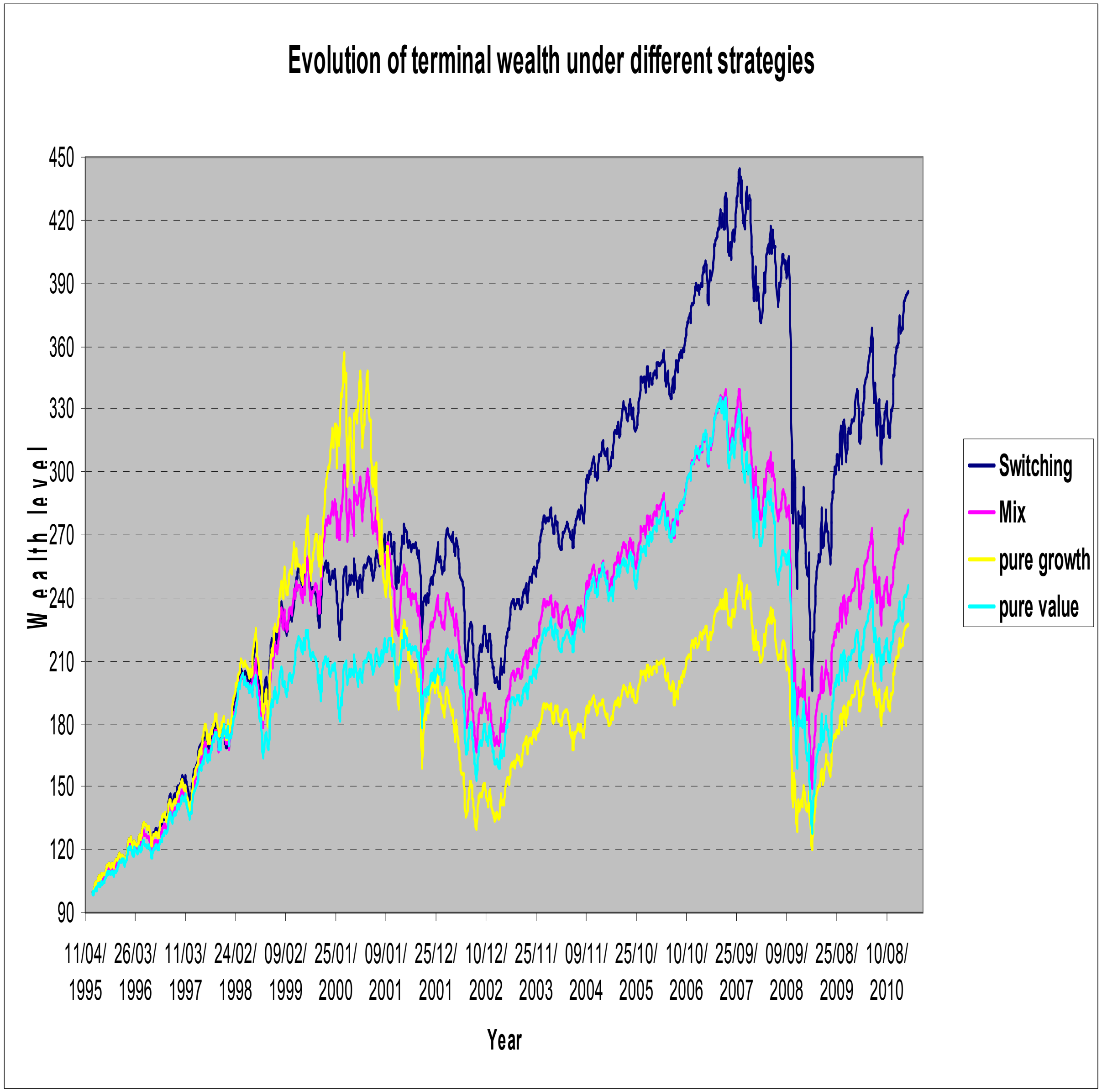
Performance comparison between WHMM- and HMM-based mixed strategies with varying transaction costs and $v = 0.08$

Transaction Cost	5 bps		20 bps		50 bps		70 bps	
	WHMM	HMM	WHMM	HMM	WHMM	HMM	WHMM	HMM
Mean (X^{RG}) %	4.561	2.929	3.575	1.869	1.599	-0.256	0.279	-1.677
Std (X^{RG}) %	10.664	11.887	10.309	11.540	9.709	10.965	9.404	10.680
Mean (X^{RV}) %	-2.768	-5.911	-3.754	-6.971	-5.729	-9.096	-7.050	-10.516
Std (X^{RV}) %	8.510	7.608	8.985	8.191	10.017	9.430	10.754	10.295

▶ SDs of return differences, std(X^{RG}) and std(X^{RV}), are lower than those from switching strategy.

▶ Mean(X^{RG}) and Mean(X^{RV}) decrease as transaction cost increases.

▶ WHMM-based mixed strategy produces higher values in both mean and SD compared to HMM-based mixed strategy. WHMM setting certainly carries more opportunities to explore tradeoff between expected return and risk, which means higher risk may lead to higher return.



Evolution of terminal wealth under different strategies
450
420
390
360
330
300
270
240
210
180
150
120
90
Wealth level
11/04/ 1995
26/03/ 1996
11/03/ 1997
24/02/ 1998
09/02/ 1999
25/01/ 2000
09/01/ 2001
25/12/ 2001
10/12/ 2002
25/11/ 2003
09/11/ 2004
25/10/ 2005
10/10/ 2006
25/09/ 2007
09/09/ 2008
25/08/ 2009
10/08/ 2010
Year
Switching
Mix
pure growth
pure value

EVALUATING PORTFOLIO PERFORMANCE

▶ We evaluate portfolio performance via a benchmark. Entire Russell 3000 index is a natural benchmark. Comparison of 4 portfolios with the benchmark is made using 3 measures on returns.

▶ Sharpe ratio (SR) - characterises how well asset return compensates investors for risk taken. The higher the SR the higher the return with same level of risk.

$$\mathsf{SR} = \frac{E[R_{\text{portfolio}} - R_{\text{riskfree}}]}{\sqrt{\mathsf{Var}(R_{\text{portfolio}} - R_{\text{riskfree}})}},$$

▶ Jensen's alpha (α_J) - measures abnormal return of a portfolio over expected return.

$$\alpha_J = R_{\text{portfolio}} - [R_{\text{riskfree}} - \beta_{\text{portfolio}}(R_{\text{benchmark}} - R_{\text{riskfree}})].$$

▶ Treynor and Black's appraisal ratio (AR) - examines returns relative to a risky benchmark.

$$\mathsf{AR} = \frac{E[R_{\text{portfolio}} - R_{\text{benchmark}}]}{\sqrt{\mathsf{Var}(R_{\text{portfolio}} - R_{\text{benchmark}})}}.$$

SUMMARY OF RESULTS

Period	Switching strategy	Mixed strategy	Pure Russell 3000 value	Pure Russell 3000 growth	Pure Russell 3000 index
Mean	0.0741	0.0495	0.0435	0.0521	0.0499
Std	0.1976	0.1890	0.1577	0.2168	0.1906
Mean/Std	0.3751	0.2622	0.2755	0.2405	0.2618

Sharpe ratio statistics for five investment strategies using 15 intervals. Switching & mixed strategies outperform benchmark in 11 and 6 intervals, respectively out of 15.

Period	Switching Strategy ($\times 10^{-4}$)	Mixed Strategy ($\times 10^{-4}$)	Pure Russell 3000 Value ($\times 10^{-4}$)	Pure Russell 3000 Growth ($\times 10^{-4}$)
Mean	5.8977	-0.0722	0.5987	-1.4289
Std	7.3348	1.0969	9.6750	10.3850
Mean/Std	0.8041	-0.0658	0.0619	-0.1376

Jensen's alpha statistics for four investment strategies using 15 intervals. 11 and 5 positive α's out of 15 for switching and mixed strategies, respectively.

Period	Switching Strategy	Mixed Strategy	Pure Russell 3000 Value	Pure Russell 3000 Growth
Mean	0.1023	-0.0445	-0.0235	0.0207
Std	0.1633	0.1869	0.1926	0.1953
Mean/Std	0.6263	-0.2381	-0.1220	0.1060

AR statistics for four investment strategies using 15 intervals. Switching strategy outperforms mixed strategy in 11 intervals.

p-values for a one-tailed significance test on the performance results based on Jensen's α and AR ratios

	Switching vs Mixed	Switching vs Pure growth	Switching vs Pure value	Mixed vs Pure growth	Mixed vs Pure value
Jensen's alpha	0.0035	0.0515	0.0174	0.3967	0.3113
Appraisal ratio	0.0149	0.0321	0.1127	0.3821	0.1791

▶ We note that the switching strategy has the best performance for the period considered based on the 3 performance metrics.

▶ Difference in means under the performance measures between switching and mixed strategies is highly significantly. The same can be said for the comparison between switching and pure growth strategies. The $t-$test assumption was validated by the Jarque-Bera test for normality. *The SR ratio data was left out in the above hypothesis testing exercise since the t-test assumption was not met*.

CONCLUDING REMARKS

▶ A weak MC modulated model for asset prices is proposed. A WHMM is transformed into a regular HMM. General recursive filters were developed. The EM algorithm in conjunction with the change of probability measure was applied to re-estimate model parameters.

▶ The h-step ahead predictions were analysed and results compared with regular HMM. Empirical results suggest that, by involving memories, there is a gain in using a 2-state WHMM.

▶ We explored applications to asset allocation problem. Future directions include other applications and addressing WHMM statistical inference problems of determining (i) *optimal no. of states* and (ii) *optimal order* as implied by the dataset.

Let $\mathbf{\Pi}$ be an $N^2 \times N^2$ matrix, the transition probability matrix of *new MC* $\alpha\binom{\mathbf{x}_k}{\mathbf{x}_{k-1}}$. It can be reconstructed from $\mathbf{A}$:

$$\mathbf{\Pi} = \begin{pmatrix} a_{111} & \dots & a_{11N} & 0 & \dots & 0 & \dots & 0 & \dots & 0 \\ 0 & \dots & 0 & a_{121} & \dots & a_{12N} & \dots & 0 & \dots & 0 \\ & \dots & & & \dots & & \dots & & \dots & \\ 0 & \dots & 0 & 0 & \dots & 0 & \dots & a_{1N1} & \dots & a_{1NN} \\ & \dots & & & \dots & & \dots & & \dots & \\ a_{N11} & \dots & a_{N1N} & 0 & \dots & 0 & \dots & 0 & \dots & 0 \\ 0 & \dots & 0 & a_{N21} & \dots & a_{N2N} & \dots & 0 & \dots & 0 \\ & \dots & & & \dots & & \dots & & \dots & \\ 0 & \dots & 0 & 0 & \dots & 0 & \dots & a_{NN1} & \dots & a_{NNN} \end{pmatrix},$$

where

$$\pi_{ij} = \begin{cases} a_{lmv} & \text{if } i = (l-1)N + m, \quad j = (m-1)N + v \\ 0 & \text{otherwise} \end{cases}$$

Each nonzero element in $\mathbf{\Pi}$ represents the probability:

$$\pi_{ij} = a_{lmv} = P(\mathbf{x}_k = \mathbf{e}_l | \mathbf{x}_{k-1} = \mathbf{e}_m,\ \mathbf{x}_{k-2} = \mathbf{e}_v),$$

and each zero represents an impossible transition.

Derivation of filters can be found in the paper downloadable at:

doi:10.1016/j.econmod.2010.10.002

Recalling the EM Algorithm

The vector of parameters $\widehat{\theta} = \{(\widehat{\pi}_{ji}),\ \widehat{\sigma}_i,\ \widehat{f}_i,\ 1 \le i,\ j \le N\}$ determines the proposed model.

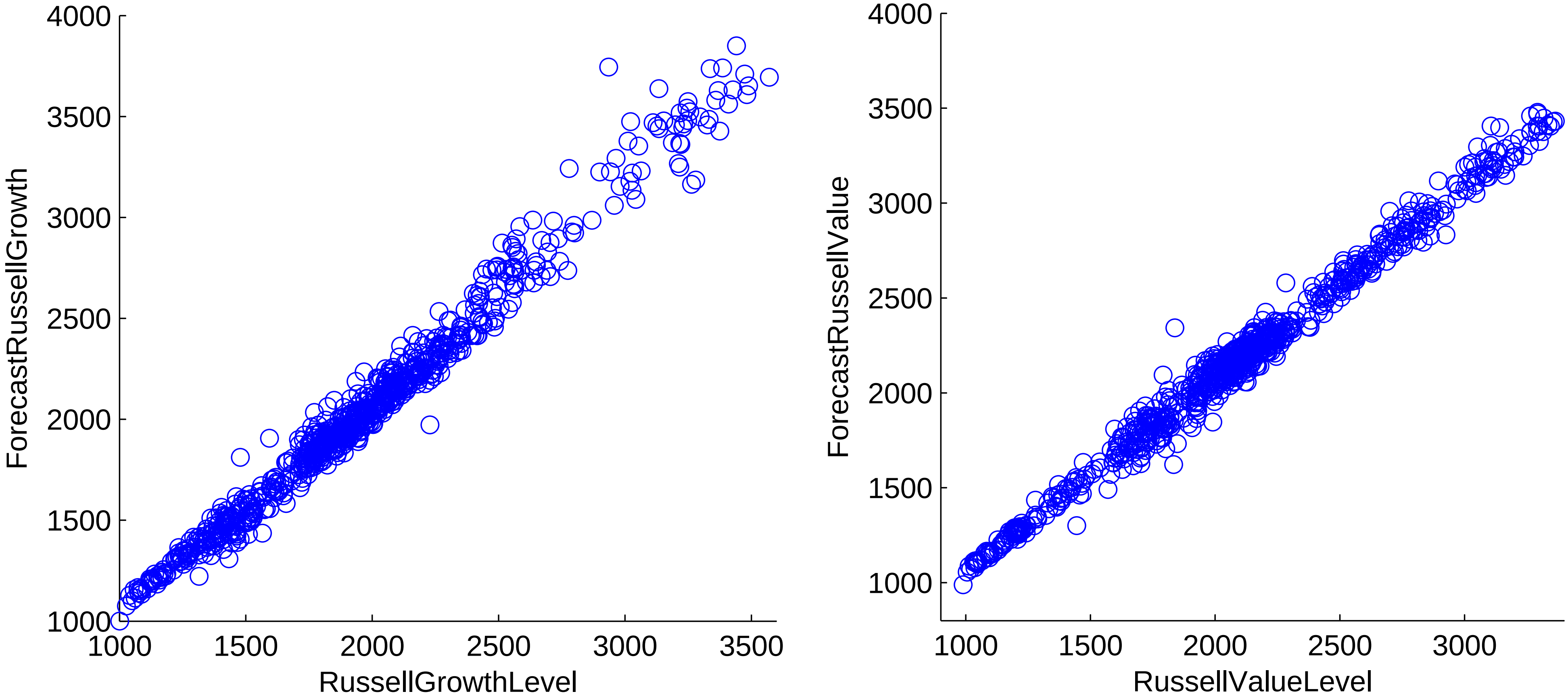
Given: A family of probability measures $\{P^\theta, \theta \in \Theta\}$ on some measurable space $(\Omega, \mathcal{G})$ and $\mathcal{Y} \subset \mathcal{G}$.

Aim: Calculate the parameter θ.

1. Set $m = 0$ and choose $\widehat{\theta}_0$.
2. **(E-Step)** Set $\theta^* = \widehat{\theta}_m$ and determine $L(\theta, \theta^*) = E^{\theta^*}\left[\log \frac{dP^\theta}{dP^{\theta^*}}\middle| \mathcal{Y}\right]$.
3. **(M-Step)** Find $\widehat{\theta}_{m+1} \in \arg_{\theta\Theta} \max L(\theta, \theta^*)$.
4. Replace m by $m+1$ and repeat procedures (**E** and **M** steps) until some stopping criterion is satisfied (e.g., $|\widehat{\theta}_{m+1} - \theta_m| < \epsilon$ for some $\epsilon > 0$).

$\{\widehat{\theta}_m\}$ produces non-decreasing likelihood values that converge to a local maximum of the likelihood function; see, eg., Wu (1983). EM algorithm involves changing P_θ to $P_{\widehat{\theta}}$ *via likelihood* $dP_{\widehat{\theta}}/dP_\theta$. Under P_θ, $\mathbf{x}$ is a weak MC with transition matrix $\mathbf{A}$. Under $P_{\widehat{\theta}}$, $\mathbf{x}$ remains a weak MC with transition $\widehat{\mathbf{A}} = (\widehat{a}_{rst})$.

$P_{\widehat{\theta}}(\mathbf{x}_{k+1} = \mathbf{e}_r | \mathbf{x}_k = \mathbf{e}_s, \mathbf{x}_{k-1} = \mathbf{e}_t) = \widehat{a}_{rst}$; $\widehat{a}_{rst} \ge 0$; and $\sum_{r=1}^N \widehat{a}_{rst} = 1$.



Scatter plots of actual data and one-step ahead forecasts under 2-state WHMM for Russell 3000 growth index level (left) and Russell 3000 value index level (right)