Dimensional portfolio reduction problems with asymptotic Markov processes

Enrico Angelelli
University of Brescia
Department Economics and Management
Brescia, Italy
angele@eco.unibs.it

Sergio Ortobelli Lozza
University of Bergamo
Department SAEMQ
24127, Bergamo (Italy) and VSB-TU of Ostrava
Department of Finance
70121 Ostrava, Czech Republic
sol@unibg.it

Gaetano Iaquinta
Private Research Analyst
20017- Rho (MI)- Italy
gaetano.iaquinta@unibg.it

Abstract: In this paper we deal how to reduce the dimensionality of the large scale portfolio selection problem under the assumption that financial returns follow homogenous Markovian chains and the underlying distributions present heavy tails. In this framework we first describe how to account the joint behavior of future wealth considering a bivariate Markov process. Secondly we suggest some techniques to reduce the large scale dimensional problem in a computationally tractable way. Finally we perform an ex-post wealth analysis to assess the profitability of a dimensional reduction technique.

Key–Words: portfolio selection, stable Paretian distributions, bivariate Markov process, reduction of dimensionality.

1 Introduction

In this paper we discuss how to reduce the large scale portfolio selection problem using the asymptotic approximation of future wealth estimated with Markov chains. Traditional financial theory is based on the assumption of financial returns normally distributed. Many empirical studies (fundamental works of Mandelbrot (1963) and Fama (1965) and more recent Rachev and Mittnik (2000), Rachev et al. (2007) and the reference therein) had rejected a normal shape for financial returns distribution. Several research works have been proposed to improve the realism of the underlying financial models. In this paper we assume assets returns follow Markov processes and thus their evolution is approximated with a Markov chain as suggested in empirical studies (see Cox et al. (1979), Angelelli and Ortobelli (2009)). In particular, we assume that each couple of assets follows a bivariate Markov process so we consider their joint distribution with a proper Markov chain. Then we approximate the future wealth taking into account their asymptotic approximation. Therefore we use the covariation of stable sub Gaussian distributions (see Samorodnitsky and Taqqu (1994)) to reduce the dimensionality of the large scale portfolio problem. Finally we propose an empirical comparison to show and evaluate the impact on portfolio problems of the reduction of preselection dimensionality technique. In this empirical analysis we consider all the stocks of the main US stock markets (NASDAQ, and NYSE).

The paper is organized as follows. In Section 2 we discuss how modeling return series and their asymptotic behavior. Section 3 introduces the technique used to manage the curse of dimensionality and an empirical analysis on the US stock market. Section 4 briefly summarizes the paper.

2 Returns dynamics model

In this section we introduce and discuss the returns dynamics modeling. Then, we analyze some possible consequences of that model. In particular, we show how to determine the distributions: of the future wealth, and of a couple of portfolios that follow a bivariate Markov process. Finally, we discuss how to take into account the asymptotic behavior of the future wealth process.
2.1 Bivariate Markov processes

Bivariate processes are used for different objectives in portfolio theory. Typically they are used either to evaluate the association between different financial factors or, as in this work, to reduce the dimensionality of large scale problems. We use a bivariate Markov process to value the joint behavior of the future wealth obtained by two different portfolios of returns $z(x)$, and $z(y)$.

Similarly to the univariate case, assume that an initial wealth $W_0 = (W_{0x}, W_{0y})' = (1, 1)'$ is invested at time $t = 0$ in two portfolios of weights $x = [x_1, \ldots, x_n]'$ and $y = [y_1, \ldots, y_m]'$, of $n$ and $m$ risky assets respectively. The vector $x$ and $y$ represent the percentage of the initial wealths (respectively $W_{0x}$ and $W_{0y}$) invested in each asset. Denote the prices of these assets at time $t$ by $P^{(x)}_t = [P^{(x)}_{1,t}, \ldots, P^{(x)}_{n,t}]'$ and $P^{(y)}_t = [P^{(y)}_{1,t}, \ldots, P^{(y)}_{m,t}]'$. The portfolios returns during the period $[t, t+1]$ are given by the vector $Z_{t+1} = (Z_{x,t+1}, Z_{y,t+1})'$ with components

$$Z_{x,t+1} = \sum_{i=1}^{n} x_i P^{(x)}_{i,t+1} / P^{(x)}_{i,t} \quad \text{and} \quad Z_{y,t+1} = \sum_{i=1}^{m} y_i P^{(y)}_{i,t+1} / P^{(y)}_{i,t}.$$  

We assume that the portfolios returns $Z_{x,t}$ and $Z_{y,t}$ follow two homogeneous Markov processes. We introduce the multi-index $i = (i_x, i_y)$ and denote by $z(i) = (z_{i_x}(i), z_{i_y}(i))'$, $i \in I := \{(i_x, i_y) : 1 \leq i_x \leq N, 1 \leq i_y \leq M\}$ the states of the Markov chain. First we discretize the support of the Markov process $\{Z_t\}$. Given a set of past observations $\{z_{-K}, \ldots, z_0\}$, we consider the range of the portfolios returns

$$\left( \min_{k=-K,\ldots,0} z_{x,k}, \max_{k=-K,\ldots,0} z_{x,k} \right) \times \left( \min_{k=-K,\ldots,0} z_{y,k}, \max_{k=-K,\ldots,0} z_{y,k} \right)$$

and divide it into $N \times M$ bi-dimensional intervals $(a_i, a_{i-1}) \times (b_j, b_{j-1})$, where $\{a_i\}$ and $\{b_j\}$ are two decreasing sequences given by

$$a_i := \left( \frac{\min_k z_{x,k}}{\max_k z_{x,k}} \right)^{i/N} \max_k z_{x,k}, \quad i = 0, \ldots, N$$

$$b_j := \left( \frac{\min_k z_{y,k}}{\max_k z_{y,k}} \right)^{j/M} \max_k z_{y,k}, \quad j = 0, \ldots, M.$$  

The idea is to approximate the returns associated to values of the Markov process in $(a_i, a_{i-1}) \times (b_j, b_{j-1})$ by the state $(z_{i_x}(i), z_{i_y}(i))$ of the Markov chain defined by

$$z_{i_x}(i) = \sqrt{a_{i_x} a_{i_x-1}} = \max_k z_{x,k} \left( \frac{\max_k z_{x,k}}{\min_k z_{x,k}} \right)^{1-2i_x / \min_k z_{x,k}},$$

$$i_x = 1, \ldots, N$$

$$z_{i_y}(i) = \sqrt{b_{i_y} b_{i_y-1}} = \max_k z_{y,k} \left( \frac{\max_k z_{y,k}}{\min_k z_{y,k}} \right)^{1-2i_y / \min_k z_{y,k}},$$

$$i_y = 1, \ldots, M.$$  

Introducing,

$$u_x := \left( \frac{\max_k z_{x,k}}{\min_k z_{x,k}} \right)^{1/N} \quad \text{and} \quad u_y := \left( \frac{\max_k z_{y,k}}{\min_k z_{y,k}} \right)^{1/M},$$

we may write $z_{i_x}(i) = z_{x}^{(1)} u_x^{i_x-1}$ and $z_{i_y}(i) = z_{y}^{(1)} u_y^{i_y-1}$. Assuming the Markov chain $\{Z_t\}$ homogeneous, we denote its transition matrix by $Q = \{q(i,j)\}_{i,j \in I}$, where

$$q(i, j) = P(Z_{t+1} = z(j) | Z_t = z(i)), \quad i, j \in I,$$

represents the probability of observing the returns $z(j)$ in $t + 1$ being in $z(i)$ at time $t$. These probabilities are estimated by the maximum likelihood estimates:

$$\hat{q}(i, j) = \frac{\pi_{ij}}{\pi_i},$$

where $\pi_{ij}$ is the number of observations that transit from $z(i)$ to $z(j)$ and $\pi_i$ the number of observations in $z(i)$. Let us now consider the bivariate wealth process generated by the gross returns. The wealth $W_t = (W_{tx}, W_{ty})'$ at time $t$ is a bivariate random variable with $N \times M$ possible values:

$$W_t = z^{(t)} \otimes W_{t-1} = \left( z_{x}^{(t)} W_{(t-1)x}, z_{y}^{(t)} W_{(t-1)y} \right)'$$

$$i \in I,$$

where $W_{t-1}$ is the wealth at time $t - 1$. Denoting $i_s = (i_{x,s}, i_{y,s})$ the realized state of the Markov chain at time $s$, the value of $W_t$ is given by

$$W_t = \left( W_{0x} z_{x}^{(i_{x,s})} z_{x}^{(i_{x,s}+1)} \cdots z_{x}^{(i_{x,s}+i_{x,t})}, W_{0y} z_{y}^{(i_{y,s})} z_{y}^{(i_{y,s}+1)} \cdots z_{y}^{(i_{y,s}+i_{y,t})} \right).$$

It is clear that the sequence $(i_0, i_1, \ldots, i_t)$ identifies uniquely the path followed by the bivariate wealth process up to time $t$. Thus, using formulas (1), the wealth obtained along the path $(i_0, i_1, \ldots, i_t)$ is given by

$$W_t = \left( W_{0x} z_{x}^{(1)} u_x^{i_{x,1}+i_{x,2}+\cdots+i_{x,t}}, W_{0y} z_{y}^{(1)} u_y^{i_{y,1}+i_{y,2}+\cdots+i_{y,t}} \right).$$
Notice that vector $x$ and $y$ represent the percentages of the initial wealths. Thus, if we want to evaluate the sample path of the ex-post wealths, we have to recalibrate each portfolio in order to maintain these percentages constant over time.

Moreover, describing the gross returns by a general bivariate Markov chain with $N \cdot M$ possible states implies that the number of possible values for $W_t$ grows exponentially with the time. However, $W_t$ can take only $[1 + t(N - 1)] \cdot [1 + t(M - 1)]$ values. In particular, in this way the final wealth $W_t$ does not depend on the specific path followed by the process, but only on the sums of the indices of the states traversed by the Markov chain in the first $t$ steps (recombining effect of the Markov chain on the wealth process).

Let us denote the $[1 + t(N - 1)] \times [1 + t(M - 1)]$ possible values of $W_t$ at time $t$ by

$$w^{(l,t)} = \begin{pmatrix} w^{(l_x,t)}_x \\ w^{(l_y,t)}_y \end{pmatrix} = \begin{pmatrix} (z^{(1)})^t u^ {l-1 - l_x} \\ (z^{(1)})^t w^{1-l_y} \end{pmatrix},$$

where $l = (l_x, l_y) \in L_t := \{(l_x, l_y) : 1 \leq l_x \leq 1 + t(N - 1), 1 \leq l_y \leq 1 + t(M - 1)\}$. The possible values of $W_t$ up to time $T$ can be stored in $T$ matrices of dimension $[1 + (N - 1)T] \times [1 + (M - 1)T]$ or in a mono-dimensional vector of size $\sum_{t=1}^{T} [1 + (N - 1)t][1 + (M - 1)t] = O(NMT^3)$.

The wealth $W_t$ can be represented by a three-dimensional Markovian tree, starting with a single node $w^{(l,t),0} = (1, 1)'$ and presenting at each time instant $t$ the $[1 + t(N - 1)] \times [1 + t(M - 1)]$ nodes given by $w^{(l,t)}$, $l \in L_t$.

We are interested in the evolution of such a process $\{W_t\}$, which is clearly connected to the evolution of $\{Z_t\}$. Consider the matrix

$$P(W_t, Z_t) = \{p(W_t, Z_t)(l, i)\}_{l \in L_t, i \in I}$$

with components

$$p(W_t, Z_t)(l, i) = P(W_t = w^{(l,t)} \cap Z_t = z^{(i)}),$$

which represents the probability of obtaining the wealth $w^{(l,t)}$ and to be in state $z^{(i)}$ at time $t$, and the vector $P_{W_t} = \{p_{W_t}(l)\}_{l \in L_t}$ with components

$$p_{W_t}(l) = P(W_t = w^{(l,t)}), \quad l \in L_t.$$ 

The probabilities $p(W_t, Z_t)(l, i)$ and $p_{W_t}(l)$ can be computed recursively by the formulas

$$p(W_t, Z_t)(l, i) = \begin{cases} p_i & t = 0, l = 1 \\ \sum_{h \in I} \{p(W_{t-1}, Z_{t-1}) \cdot (l - (i - 1), h) \mid l_x - (i_x - 1) > 0, \\ \mid l_y - (i_y - 1) > 0 \} & t > 0, \\ \mid \text{otherwise} \end{cases}$$

and

$$p_{W_t}(l) = \begin{cases} 1 & t = 0, l = 1 \\ \sum_{h \in I} p_{W_t}(l, h) & t > 0, \\ \mid \text{otherwise} \end{cases}$$

(2)

where $p_i = P(Z_0 = z^{(i)})$ is the probability that the return at time zero is $z^{(i)}$. We assume these probabilities to be known from past observations.

Even if for computing the distribution of the bivariate process we need more time than the univariate case, the computational complexity of this algorithm is still of polynomial order.

### 2.2 Modeling the asymptotic behavior of the log returns

The fact that Log returns present a distribution with heavier tail than distributions with finite variance is documented in several empirical research works. The empirical investigation (see, among others, Rachev and Mittnik (2000) and the references therein) shows that

$$\Pr(|\ln(z(x))| > u) \sim u^{-\alpha} L(u) \text{ as } u \to \infty$$

(3)

where $0 < \alpha < 2$ and $L(u)$ is a slowly varying function at infinity, i.e.,

$$\lim_{u \to \infty} \frac{L(cu)}{L(u)} \to 1 \text{ for all } c > 0,$$

Our dataset satisfies the relation (3) for values $1 < \alpha < 2$. This tail condition implies that the log returns $r(x) = \ln(z(x))$ distribution admits finite mean and not finite variance and belongs to the domain of attraction of an $\alpha$-stable law. This asymptotic behavior of data can be model assuming that for each portfolio $x \in S$, the forecasted log wealth $(\tilde{W}_T(x) = \sum_{t=1}^{T} \ln(z(x), t))$ at a given future time $T$ is in the domain of attraction of an $\alpha(x)$ stable distribution, i.e.,

$$\tilde{W}_T(x) \overset{d}{=} S_{\alpha(x)}(\sigma(x), \beta(x), \mu(x)),$$

(4)
where $x \in (0, 2]$ is the index of stability, $\sigma(x)$ is the scale parameter, $\mu(x)$ is the location parameter and $\beta(x)$ is the skewness parameter. McCulloch’s method (see McCulloch (1986)) provides an efficient technique to derive stable Pareto parameters estimation. In particular, this method requires the knowledge of 5%, 25%, 50%, 75%, 95% quantiles of the log wealth $W_T(x)$ for any portfolio. Optimal portfolio strategies that account the Markovian and asymptotic behavior of the final wealth can be derived by computing reward and risk measures with stable distributions.

Alternatively, we can consider the asymptotic behavior of the future wealth, assuming that the vector of the forecasted log wealths (obtained investing in each assets) is in the domain of attraction of a particular stable law. Typically we can assume that the vector of the future log wealths, denoted by

$$\bar{W}_T(x) = [\bar{W}_{T,1}, ..., \bar{W}_{T,n}]',$$

where $\bar{W}_{T,i}$ is the log wealth at time $T$ obtained investing in the $i$-th asset, is a stable sub Gaussian distributed. That is, the characteristic function of $\bar{W}_T$ has the following form:

$$\Phi_{\bar{W}_T}(u) = E(\exp(\text{iu}'r)) = \exp(-\text{iu}'V\mu + \text{iug}'),$$

where $V = [v_{ij}]$ is a positive definite dispersion matrix, $\mu$ is the mean vector (when $\alpha > 1$) and $i$ is the imaginary unit. Since $\mu_i$ and $\sigma_i^2$ are respectively the location parameter and the square scale parameter of the stable distribution at time $T$ of $\bar{W}_{T,i}$, we can estimate the parameters $\hat{\mu}_i$ and $\hat{\sigma}_i^2$ ($i = 1, ..., n$) using the McCulloch’s quantile estimator, fixing the skewness parameter $\beta = 0$ and imposing a common stability parameter $\alpha$ for all the components. Generally, as stability parameter $\alpha$ we use either the empirical mean of the stability parameters of the assets (i.e., $\alpha = \frac{1}{n} \sum_{i=1}^{n} \alpha_i$) or the stability parameter of the market index, if it exists. As remarked by Kring et al. (2008), the covariation parameter can be seen as the difference of square scale parameters, i.e.

$$v_{ij} = \sigma^2(\bar{W}_{T,i} - \bar{W}_{T,j})/2 - \sigma^2(\bar{W}_{T,i} + \bar{W}_{T,j})/2,$$

where $\sigma^2(\bar{W}_{T,i} \pm \bar{W}_{T,j})/2$ are the square scale parameter of the random variables

$$\left(\bar{W}_{T,i} \pm \bar{W}_{T,j}\right)/2$$

whose distributions can be evaluated using the bivariate Markovian approximation. Thus, to estimate the covariation parameters $v_{ij}$ (with $i \neq j$) of the stable vector we use the estimator

$$\hat{v}_{ij} = \frac{\sigma^2(\bar{W}_{T,i} + \bar{W}_{T,j})/2 - \sigma^2(\bar{W}_{T,i} - \bar{W}_{T,j})/2}{\sigma^2(\bar{W}_{T,i} + \bar{W}_{T,j})/2}$$

based on the estimates of the scale parameters of $\left(\bar{W}_{T,i} + \bar{W}_{T,j}\right)/2$ and $\left(\bar{W}_{T,i} - \bar{W}_{T,j}\right)/2$.

### 3 Dimensional reduction of the large scale portfolio selection problem

The typical functionals which are defined under the assumption that the gross return of each portfolio follows a Markov chain with $N$ states are called OA performance (utility) functionals or OA performance measures ([1]). OA performance measures can be used either to optimize the choices or to reduce the dimensionality of the portfolio problem. In particular we use the following two OA performance functionals in the next empirical analysis:

**OA-Sharpe ratio (OA-SR).** The classic version of the Sharpe ratio (see Sharpe (1994)) values the expected excess return for unity of risk (standard deviation). With the OA-Sharpe ratio we value the expected excess final wealth at time $T$ for unity of risk, i.e.,

$$\text{OA-SR}(W_T(x)) = \frac{E(W_T(x) - 1)}{\sqrt{\text{var}(Q_T)}}$$

where $Q_T = [q_{ij,T}]$ is the variance covariance matrix of the final wealth obtained with each asset at time $T$. Thus we consider the Markov joint distribution at time $T$ of $i$-th and $j$-th assets and then we compute their covariance $q_{ij,T}$. However, using Sharpe type measures we generally don’t take into account the asymptotic behavior of the wealth (except in the case the optimal portfolios are in the domain of attraction of the Gaussian law).

**OA-stable Sharpe ratio (OA-SSR)** This performance functional is defined as

$$\text{OA-SSR}(W_T(x)) = \frac{\mu_{\text{in}}(W_T(x))}{\sqrt{\text{var}(V_x)}}$$

where $V = [v_{ij}]$ is the dispersion matrix computed with formula (6), $\mu_{\text{in}}(W_T(x)) = \mu(x)$ is the mean of the stable distribution.

that better approximates the log final wealth
\[ \ln(W_T(x)) = S_0(\sigma(x), \beta(x), \mu(\ln(W_T(x))) \] As stability parameter \( \alpha \) we use either the empirical mean of the stability parameters of the assets (i.e., \( \alpha = \frac{1}{n} \sum_{i=1}^{n} \alpha_i \)). As for the Sharpe ratio, this ratio is isotonic with the preferences of non satiable risk averse investors (see Rachev et al. 2007). Using this performance measure we also account the asymptotic behavior of the future wealth.

3.1 The large scale portfolio dimensional problem

From a statistical point of view the number of observations should increase proportionally with the number of assets (see Papp et al. (2005), Kondor et al. (2007)) to achieve a reasonable statistical approximation of the historical series. Thus when we have a considerable number of assets we need to reduce the dimensionality of the portfolio problem. In the empirical analysis we adopt a methodology to reduce the dimensionality of the large scale portfolio problems that is based on the preselection of a asset universe subset relevant for some optimality criteria.

3.1.1 Preselection

The preselection methodology is performed to identify some relevant assets involved in the portfolio choice. In our dataset the choice is among more than 1500 US stocks. We suggest to preselect no more than 170 assets following two steps:

Step 1: Order the assets with the two performance measures (8) and (9).

Step 2: Select the "best" 170 assets satisfying the same common criteria (one group of 170 assets obtained by the intersection of the ordered groups of assets).

3.2 Portfolio preselection in practice

In this section, we evaluate the impact of the proposed model on the US stock market. In particular, we consider the stocks traded on the NYSE and on the NASDAQ. Since we want to propose as much as possible a realistic empirical analysis, we have developed a dynamic dataset that uses all the useful financial data from DataStream.

Using this dynamic dataset we propose two different ex-post comparisons during a period of about two years (500 daily observations) from 15-Sep-2008 till 31-Aug-2010. In all the empirical analyses we assume:

a) that investors have a temporal horizon of \( T = 20 \) working days (thus, for each portfolio strategy we should optimize the portfolio every 20 working days for a total of 25 optimizations);
b) that investors cannot invest more than 10% in a single asset, and thus the portfolio weights should belong to the interval \([0,0.1]\) (i.e.: \( x_i \in [0,0.1]\));
c) Markov chains have \( N = 9 \) states;
d) the initial wealth \( W_0 \) is equal to 1 at the date 15-Sep-2008.

The first comparison applies the portfolio selection to the approximated preselected returns using all the active assets and the daily observations during the previous ten years (2600 working days). Thus, with this analysis we use 3100 daily observations overall from 14-May-1998 till 31-Aug-2010. The second comparison applies the portfolio selection to the approximated preselected returns using all the active assets and the daily observations during the previous six months (125 working days). Thus with this analysis we use 625 daily observations overall from 17-Mar-2008 till 31-Aug-2010. For both comparisons we value the empirical evidence from the preselected assets.

3.2.1 Empirical evidence from the preselected assets

In order to understand if preselection gives some benefits we compare the ex-post wealth of two portfolio strategies with the behavior of two market indexes: NASDAQ Composite and NYSE Composite. In the first strategy the investor uses a completely diversified portfolio on all active assets either in the last six months or in the last ten years (i.e., he/she invests \( 1/n \) in each asset where \( n \) is the number of available assets at the optimization date). With the second strategy the investor uses a completely diversified portfolio only on the preselected assets. For both strategies the investors recalibrate the portfolio every month (every 20 working days).

Figure 1 reports the sample paths of the ex-post wealth of the two strategies and of the market indexes when we consider all active assets either in the last six months, or in the last ten years.

4 Concluding remarks

From the proposed analysis we deduce that the diversification and preselection could play a crucial role in portfolio selection, since in both cases
we get more wealth than the indexes. This appears much more evident when we compare the final wealths of the two strategies. Moreover, considering the limited transaction costs we obtain when we use preselection, we also deduce that it makes sense considering a preselected number of assets. The difference between the strategies based on preselected assets among all those active either in the last ten years or in the last six months suggests that:

1) the set of the preselected assets is completely different in the two cases;
2) preselection works better if it is applied to more assets;
3) the recent entries in the market could have an important impact in the portfolio choices.

Acknowledgement. The paper has been supported by the Italian funds ex MURST 60% 2013 and MIUR PRIN MISURA Project, 2013–2015. The research was also supported through the Czech Science Foundation (GAČR) under project 13-13142S and SP2013/3, an SGS research project of VSB-TU Ostrava, and furthermore by the European Regional Development Fund in the IT4Innovations Centre of Excellence project (CZ.1.05/1.1.00/02.0070), including the access to the supercomputing capacity, and the European Social Fund in the framework of CZ.1.07/2.3.00/20.0296 (second author).

References


