Time Varying Sensitivities on a GRID architecture ⋆

Mattia Ciprian [c] Stefano d'Addona [a,b,∗],

[a] *University of Rome III*

[b] *Graduate School of Business, Columbia University*

[c] *University of Trieste, Italy*

Abstract

We estimate time varying risk sensitivities on a wide range of stocks' portfolios of the US market. We empirically test, on a 1926-2004 Monthly CRSP database, a classic one factor model augmented with a time varying specification of betas. Using a Kalman filter based on a genetic algorithm, we show that the model is able to explain a large part of the variability of stock returns. Furthermore we run a Risk Management application on a GRID computing architecture. By estimating a parametric Value at Risk, we show how GRID computing offers an opportunity to enhance the solution of computational demanding problems with decentralized data retrieval.

Key words: Time Varying Beta, Kalman Filter, GRID Computing, Value at Risk
JEL: G11, G12, C32, C63, C88

⋆ This work is partly funded by MIUR as part of the EGRID project.

∗ Corresponding author phone:+1-646-257-3803; fax:+1-360-251-6475.
E-mail addresses: sd2123@columbia.edu (S. d'Addona), mciprian@units.it (M. Ciprian)

1 Introduction

The estimation of systematic risk has been one of the most studied topic in empirical finance. Standard OLS estimation of market model sensitivities produces a beta that is supposed to be constant, but there is no evidence of such statistical property. Indeed, time varying betas were studied by many authors. In one of the earliest paper investigating the time series properties of risk sensitivities, Blume (1971) found some evidence of mean reversion in the beta. In a following empirical work, Blume (1975) shows stationarity for portfolios' betas and unstable behavior for single stock's betas. For explaining these findings, the author claims that firms may tend to undertake riskier projects at the beginning of their life, leading to the founded mean reversion nature of beta for single stocks. Following these papers, Brenner and Smidt (1977) proposed a non stationary model, where the risk sensitivity of a stock is related with the value of the stock itself, showing further evidence of the time varying nature of betas. Furthermore, in an empirical work, Francis (1979) provided an analysis, on a decade of CRSP data, confirming his findings. Further evidence on the US market is presented in Sunder (1980), where the null hypothesis of market risk stationarity is tested against a random walk specification, and in Ohlson and Rosenberg (1982), where an ARMR(1,1) model is proposed and tested on an equally weighted portfolio based on 50 years of CRSP data. In a following empirical work, Collins et al. (1987) confirmed the autoregressive nature of risk sensitivities found in Ohlson and Rosenberg (1982), with a detailed comparison of four different model specifications. On the other hand some authors (e.g Fabozzi and Francis (1978) and Bos and Newbold (1984)) show compelling evidence of time varying systematic risk due to micro and macro factors.

The introduction of more sophisticated econometric techniques in the financial literature, influenced also the empirical research on risk sensitivities. In particular, following the seminal contributions by Engle (1982) and Bollerslev (1986) on modeling heteroskedasticity in time series, GARCH techniques are applied for modeling time varying risk sensitivities. In this strand of literature Bollerslev et al. (1988) apply a GARCH model for estimating a conditional CAPM model with the assumption of heteroskedaticity in the covariance between risky assets and market portfolio. By testing their assumption on the US market, the authors find a strong support for their hypothesis of time varying covariance matrix for assets' returns. In the same fashion, a Multivariate GARCH application to model time varying betas is developed in Braun et al. (1995). Furthermore the time varying nature of systematic risk is confirmed on several international markets in Giannopoulos (1995). Alternatively, Schwert and Seguin (1990) propose and estimate a single factor model of portfolio returns heteroskedaticity: to estimate time-varying monthly variances for size-ranked portfolios, they use predictions of aggregate stock return variances from daily data.

Most of the above studies have focused on the empirical test of stochastic nature of betas regardless the "type" of the stock/portfolio investigated. The first step in this direction is in Ghysels (1998), where the time varying nature of the systematic risk for several industry portfolios is investigated. Following this paper Groenewold and Fraser (1999) applied a Kalman filter estimation to Australian industry portfolios, and argued that industrial sectors are divided in two classes, one with time varying risk sensitivities and the other one with relatively stable behavior. Interestingly enough Groenewold and Fraser (1999) run also a recursive regression and a rolling regression on the same data, finding inconsistencies in the obtained results. For investigating these results Brooks et al. (1998) performed a horse race amongst three different model specifications on the same market. Based on both in-sample and out-of-sample forecast errors, they found overwhelming support for the Kalman filter approach. A Kalman filter estimation is also performed by Black et al. (1992) for analyzing the performance of UK Unit trusts in the '80s.

Our paper is closely related with the presented empirical literature on estimation of time varying risk sensitivities. Our contribution is twofold. First we provide an up to date and detailed analysis of time varying nature of risk sensitivities on the US market. By using a a Kalman filter approach augmented with a genetic algorithm for the log-likelihood optimization, we investigate the risk sensitivity for a broad class of portfolios as well as for a wide range of stocks with different characteristics. [1] Second, we propose and estimate a Value at Risk application on several stock portfolios based on the estimation on a GRID computing environment, showing its potential for enhancing the solution of computational demanding problems with decentralized data retrieval.

The remainder of the paper is organized as follows. In Section 2 we present the market model framework as a theoretical background to the empirical investigation. Section 3 introduces the data set used in the empirical part and provides descriptive statistics of the analyzed stock portfolios. In Section 4 we describe the estimation procedure and discuss the results of the empirical investigation on the US stock market. In Section 5 we implement the risk management application and Section 6 concludes.

2 Theoretical Background

In this section we review the theoretical framework for our empirical estimation. Starting from the Arbitrage Pricing Theory (APT) (cfr. Ross (1976), Roll and Ross (1980) and Chen et al. (1986)), which models the statistical evidence that asset payoff tends to move together, we derive a simple market

[1] To economize space and keep the paper readable, results on sigle stocks are available upon request.

model for stock returns. Standard assumptions of APT are that markets are competitive and frictionless, and that returns are generated according to

$$R = a + Bf + \epsilon \tag{1}$$

with $\epsilon \sim N(0, \Sigma)$ Σ diagonal, where R is an (Nx1) vector of returns, a is the (Nx1) vector of intercepts of the factor model, B is the (NxN) matrix of factor sensitivities, f is the (Nx1) vector of factors and ϵ is the (Nx1) vectors of disturbances.
If a risk free asset exists and adopted factors are traded portfolios, exact factor pricing holds. Throughout the paper we assume that a risk free asset is traded and the market portfolio is the pricing factor. Therefore the pricing model can be expressed using a market portfolio as a factor:

$$R^e_{it} = \beta_i R^e_{mt} + \epsilon_{it}, \tag{2}$$

where the superscript e indicates excess returns.
As a departure from the classical APT models we consider time varying factor sensitivities. More specifically we assume a mean reverting process for the beta:

$$\beta_{it} = \overline{\beta_i} + \alpha_i \left(\beta_{it-1} - \overline{\beta_i} \right) + \sigma_i \varepsilon^i_t, \tag{3}$$

where $\overline{\beta_i}$ is the unconditional mean of the sensitivity relative to the asset i, σ_i is its conditional volatility, α_i is the mean reversion parameter, and the error $\varepsilon^i_t \sim N(0, 1)$ is i.i.d. Thus, considering both equations (2) and (3), the proposed model for the asset returns is:

$$\begin{aligned} R^e_{it} =& \beta_{it} R^e_{mt} + \epsilon_{it}, \\ \beta_{it} =& \overline{\beta_i} + \alpha_i \left(\beta_{it-1} - \overline{\beta_i} \right) + \sigma_i \varepsilon^i_t. \end{aligned} \tag{4}$$

3 Data

In this Section we present and describe the main features of the financial series employed in this study. Our empirical exercise is mainly based on the 1926-2004 Monthly CRSP database. Portfolios formed on Size (SIZE), Earning Price (E-P), Dividend Price (D-P) and Industry (IND) are from Kenneth French's website.
SIZE portfolios are constructed at the end of each June using the June market equity and NYSE breakpoints. The Market Value is computed as price times shares outstanding. The available sample period is from July 1926 to December 2004.
D-P portfolios are formed are formed on dividend price ratios at the end of each June using NYSE breakpoints. The dividend yield used in June of year t

is the total dividends paid from July of t-1 to June of t per dollar of equity in June of t. The available sample period is from July 1927 to December 2004.
E-P Portfolios are constructed with the earning price ratio at the end of each June using NYSE breakpoints. The earnings used in June of year t are total earnings before extraordinary items for the last fiscal year end in t-1. The sample period covers from July 1951 to December 2004.
Finally the selected Industry portfolios are Manufacturing (SIC codes 2000-3999), Utilities (SIC codes 4900-4999), Shops (SIC codes 5000-5999, 7000-7999), Money and Finance (SIC codes 6000-6999) and Other[2].
In order to better understand the empirical exercise, it is worth looking briefly at the basic characteristics of the analyzed market. Table 1 presents, for each of the analyzed portfolios, the mean and standard deviation of the return time series. Panel A of Table 1 presents the descriptive statistics for the SIZE based portfolios. During the entire sample period the SIZE portfolio, based on the lowest quintile, outperforms by 46 basis points the portfolio based on the highest quintile, confirming the well documented size effect (see Stattman (1980), Roseberg et al. (1985) and Fama and French (1995) among others). Panel B and C of Table 1 show the descriptive statistics for the E-P and D-P based portfolios respectively. In these cases, the portfolios based on the highest quintile systematically outperform the portfolios based on the lowest quintile, confirming the well known value effect. (Cfr. for example Basu (1983)). Finally Panel D, Table 1, presents the descriptive statistics of the chosen industry portfolios. During the entire sample the portfolios seem to have a similar volatility-return profile, except the Money portfolios that slightly outperform the others.

[Table 1 about here.]

4 Empirical Results

4.1 Estimation Procedure

The estimation of the model presented in equation (4) is performed using a Kalman filter, where the observation equation and state equation are specified as follows:

$$\begin{aligned} Y_t &= \Phi_t S_t + R\epsilon_t, \\ S_t &= A + F S_{t-1} + Q v_t. \end{aligned} \tag{5}$$

In the above state-space form Y_t is a column vector that stores the asset returns observed at time t; Φ_t is a column vector of the observable risk factor

[2] A detailed description, along with the data, is available at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french

(in our case the market index) and S_t is a column vector of the unobservable risk factor sensitivities. In our model specification, the unobservable variables are supposed to follow a simple mean reverting autoregressive process. Thus, A and F are respectively column vectors of the unconditional means and a [assets x assets] diagonal matrix with the autoregressive parameters on the diagonal. Furthermore, Q and R are diagonal matrices of the volatilities of the unobservable and the observable variables respectively. Finally ϵ_t and v_t are column vectors of error terms with a $N(0, I)$ probability distribution. To guarantee and facilitate the correct estimation of the process parameters some restrictions are imposed. For all processes the domain of the diffusion terms is restricted to be positive. Once the restriction is imposed, the Kalman filter is performed.

For implementing the algorithm we follow closely the procedure in Hamilton (1994). First of all, we initialize the state-vector S_t with its expected value:

$$S_{1|0} = A + FS_0, \tag{6}$$

where S_0 contains the guessed starting values of the state variables. The associated mean squared error (MSE i.e. the variance covariance matrix of the initialized state vector) can be computed as:

$$P_{1|0} = FP_{1|0}F' + Q'Q. \tag{7}$$

By using the well known result from matrix algebra,
$vec(ABC) = [(C' \otimes A)vec(B)]$, we can easily compute the MSE as:

$$vec(P_{1|0}) = [I - F \otimes F]^{-1} vec(Q'Q). \tag{8}$$

The second step of the algorithm consists of forecasting the observable variables and updating the Kalman filter. With the updates it is then possible to calculate the new estimates for the state variable vector and its variance covariance matrix. The forecast of the Y_t vector is computed as:

$$Y_{t|t-1} = \Phi S_{t|t-1}, \tag{9}$$

with a an estimation forecast error equal to:

$$\xi_t = \Phi(S_t - S_{t|t-1}), \tag{10}$$

and a covariance matrix of estimation forecast error:

$$\mathbb{E}\left[\xi_t' \xi_t\right] = \Phi P_{t|t-1} \Phi' + R'R. \tag{11}$$

Once we have calculated the estimation forecasts and the relative estimation errors, we can update the Kalman filter via:

$$\begin{aligned} S_{t|t} =& S_{t|t-1} + P_{t|t-1}\Phi'(\Phi P_{t|t-1}\Phi' + R'R)^{-1}\xi_t \\ P_{t|t} =& P_{t|t-1} - P_{t|t-1}\Phi'(\Phi P_{t|t-1}\Phi' + R'R)^{-1}\Phi P_{t|t-1}. \end{aligned} \tag{12}$$

Thus, new estimates for the state variable vector and its variance covariance matrix can be calculated as:

$$\begin{aligned} S_{t+1|t} &= A + FS_{t|t} \\ P_{t+1|t} &= FP_{t|t}F' + Q'Q. \end{aligned} \tag{13}$$

Last step of the Kalman filter procedure is to compute and maximize the log-likelihood function. In our model the log-likelihood to be maximized is :

$$\begin{aligned} \mathcal{L}_T &= \sum_{t=1}^{T} \mathcal{L}_t \\ &= \sum_{t=1}^{T} \left((2\pi)^{-\frac{1}{2}} |\Phi P_{t|t-1}\Phi' + R'R|^{-\frac{1}{2}} e^{\left[-\frac{1}{2}\xi'(\Phi P_{t|t-1}\Phi' + R'R)^{-1}\xi\right]} \right). \end{aligned} \tag{14}$$

In order to maximize the expression in equation (14) we choose to implement a genetic algorithm (GA) procedure. Two main features make GA more suitable than other optimization algorithms: first, a GA is usually more robust than other algorithms and it can tolerate approximate or even noisy design evaluation; second, a Genetic Algorithm can be efficiently parallelized and therefore take full advantage of a GRID based application. In the next subsection we briefly describe the implemented algorithm.

4.2 Genetic Algorithm

Genetic algorithms are search algorithms based on the mechanics of natural selection (see Goldberg (1989) for a complete reference). Following Poloni and Pediroda (1997), a genetic algorithm can be described with a pseudo-code structure such as:

```
do ng generation
      do nind individuals
              translate bits into variables
              compute objective
      end do
      do some statistics on the population individuals
      do Create a new population:
           by cross over:
              select individuals
              and reproduce
           by mutation:
              select individuals
              and mutate
      end do
end do
```

The key points of a GA are the operators used for selection and reproduction that are crucial for the robustness and the efficiency of the algorithm.
In order to understand the mechanism of a GA, we illustrate in the next subsection, some of the operators and functions used in our implementation.

4.2.1 Coding

For starting the algorithm, it is necessary to define the initial population, that is any collection of solutions that could reasonably span the whole solution space. In order to perform this task, we generated a random sampling over that space, as explained in Montgomery (1996) and Del Vecchio (1997).[3] Each design variable is then coded in a finite-length string; traditionally, GAs use binary numbers to represent such strings: a string has a finite length and each bit of a string can be either 0 or 1. For real function optimization, however, it is more natural to use real numbers: the length of the real-number string corresponds to the number of design variables (cfr. Daisuke Sasaki and Himeno (Barcelona, 11-14 September 2000)). We adopted this coding technique.
After the initial population is generated the process of selection is implemented. The selection (reproduction) operator selects chromosomes, according to their fitness function values, to choose a new generation. In the selection procedure, the well-fitted individuals have more chances to be selected. It is worth noting that it is not a deterministic choice: even solutions with a comparatively low fitness may be chosen and they may reveal good choices in the evolution of the algorithm (see Periaux et al. (1997)).
The three selection techniques usually used are:

Roulette wheel is the first and most popular operator. A selection probability proportional to its fitness is assigned to each individual in the population. The operator is robust but computationally intensive, moreover it could cause premature convergence if no scaling of fitness is applied.

Tournament overcomes the problem of fitness scaling and it is considered more efficient and robust than roulette wheel. The characteristic of a tournament is to keep the best of a group of individuals randomly selected. In our implementation we used this operator.

Local Geographic Selection elsewhere named as step-stone island model, is a particular case of Turnament Selection. The n-size individuals participating to the tournament are not selected randomly in the population but through a local random walk in the neighbourhoods of a given individual being the population distributed in a N dimensional grid.

Next step in the genetic algorithm is to fill up the new generation. The main way to perform this task is through the cross-over operator. Amongst the

[3] It is worth noting that, for avoiding local optimum solutions, the size of the population has to be 2 to 4 times the size of variables, as noted by Rao (1996).

cross-over operators one with the highest search robustness is the *two points cross-over*; in this operator, two points are randomly chosen and the genetic materials (i.e the design variables) are exchanged between the parent variables vectors, as shown below:

A	**0 0**	**1 0 1**	**1 0 1**	$\longrightarrow$	1 1	**1 0 1**	1 1 0	A’
B	1 1	1 1 0	1 1 0		**0 0**	1 1 0	**1 0 1**	B’

Another powerfull cross-over operator has been implemented: the *directional cross-over*; it assume that a ”direction of improvement” can be detected comparing the fitness value of two reference individuals. The schema is shown below:

(1) for all individuals i
(2) select individual $i1$, select individual $i2$
(3) create the new individual as:

$$\bar{x} = \bar{x}_i + S \cdot sign(F_i - F_{i1}) \cdot (\bar{x}_i - \bar{x}_{i1}) + T \cdot sign(F_i - F_{i2}) \cdot (\bar{x}_i - \bar{x}_{i2})$$

where S and T are random numbers in the interval $[0, 1]$, F is the value of the fitness function for the corrisponding vector of variables $\bar{x}$.
Finally in order to enhance population diversity, a mutation operator is performed. A mutation is a random change in the genetic material of a single individual; it is applied to genes by changing them with a low probability, P_m. In our case, a mutation means switching a bit 0 to 1 and vice versa. This operator enables the optimization to get out of local minima.[4] A mutation algorithm can be described as follows:

A’ 1 1 1 0 1 1 1 0 $\longrightarrow$ 1 1 1 0 **0** 1 1 0 A”

4.3 Results

In this subsection we address the in-sample accuracy of the presented model. First it is interesting to assess the capability on the employed optimization algorithm.

[Fig. 1 about here.]

Figure 1 help us in analyzing the computational performance of the Genetic Algorithm. It shows, in term of absolute value reached by the optimized likelihood function, the gain obtained increasing the generations size. Clearly the Genetic Algorithm has an asymptote that is reached, in our test, at 1000 generations. The maximum value attained for the log-likelihood function is

[4] An intuitive characteristic of the mutation operator is that the higher the probability of mutation the more the search process functions like a pure random search.

6581.9. It is worth noting that, with 500 generations, the attained value is 6447.88, thus while diminishing the number of generations by a factor of two would certainly help in speeding up the algorithm, the loss of accuracy is only of about 2%

Table 2 presents parameters estimation on the selected stock portfolios. By analyzing these results, we can draw some preliminary insight on the goodness of fit of the proposed model. First, the model seems to be able to explain a consistent part of the analyzed stock returns, with an R^2 that range from 0.65 for the Money industry portfolio to 0.98 for the highest quintile SIZE portfolio. This result is consistent with an relevant strand of the literature, started by Jagannathan and Wang (1996). In their paper a conditional capital asset pricing model with time varying betas and market risk premiums is tested. Using returns on human capital and aggregate wealth they are able to explain 57% of cross sectional stock returns variability.

Analyzing in more details the presented panels some other features are worth noting. In Panel A, where the SIZE portfolios are analyzed, the explanatory power of the model is increasing in size, with an increment of 30 percentage points in the statistics from the smallest to the biggest portfolio. This result is well documented in literature (see for example Banz (1981), Fama and French (1993) and Fama and French (1992)). Non surprisingly a related pattern is followed by the estimated volatility parameters for the SIZE portfolios: where the R^2 is higher the volatility tends to be smaller, with an order of magnitude in the first quintile versus the last quintile. Similar results can be inferred from Panel B and Panel C, where the estimated parameters are presented for E-P and D-P portfolios respectively. In these cases, even if the R^2 range is narrower, the variance of the growth stock portfolios seems to be better explained by the model. Again the same pattern for the volatility of the unobservable process is founded. Finally, Panel D presents the results for industry based portfolios. While the model performs well in most of the analyzed portfolios, it is worth noting its relative lack of accuracy for the Money portfolio with respect to the other industries.

[Table 2 about here.]

[Fig. 2 about here.]

5 An application to Risk Management

In this section we apply the estimation method proposed in Subsection 4.1 to a simple Value a Risk (VaR) exercise.

We processed our data using a computational GRID technology implemented in a national facility as part of the research project EGRID.

5.1 EGRID Project

As explained in details by Leto et al. in Leto et al. (2005), the EGRID project is a research project funded by MIUR[5]. The aim of the project is to investigate the role of GRID technologies in the field of complex systems applied to economics and finance. The MIUR evaluation committee assigned to the EGRID project a further specific task: to implement a GRID infrastructure allowing geographically distributed scientific communities involved in these projects to share economic and financial data as well as applications. A preliminary version of this infrastructure was released on October 9, 2004: it is based on European Data Grid (EDG) middleware and is hosted as an independent Virtual Organization (VO) within INFN-GRID.[6]
The EGRID project manage to successfully implement the facility with the following characteristics

- the possibility to handle approximately 1GB of data coming from various stock exchanges;
- data privacy and security, i.e. the access to this resource had to be secure, authorized and authenticated;
- check availability of machines to distribute the computing load.

In the Risk Management exercise proposed in this Section, we fully take advantage of the GRID infrastructure treating our application as multithread. Loosely speaking, multithreading can be defined as a programming technique that enables an application to handle more than one operation at the same time. A main application has been created and launched in a "server machine": this program manages the Genetic Algorithm and constantly listens to a port for communication with other programs running in "client machines" inside the GRID (cfr. Fig. 3). Each client application elaborates a particular configuration (a genetic individual of the generation) as required by the server. In this setting, the most challenging task was to make sure that multiple threads do not interfere with each other in an undesired way.

[Fig. 3 about here.]

In a Risk Management setting, the VaR indicates, in percentage terms, the maximum probable loss on a given portfolio, referring to a specific confidence interval and time horizon. Historically the VaR literature has been evolved following two main approaches: parametric and non parametric models (see Jorion (2000) for a complete reference). In the latter class of models we can

[5] Ministero dell'Istruzione, Università e Ricerca: Italian Ministry of Education, University and Research.
[6] The national computing grid infrastructure of INFN (Istituto Nazionale di Fisica Nucleare: Italian National Institute for Nuclear Physics).

pinpoint full valuation models as Historical Simulation and Monte Carlo Simulation. The Historical Simulation uses past empirical distribution of returns in order to simulate the probability distribution of future returns. The VaR is then calculated as the chosen percentile of the simulated distribution. On the contrary, Monte Carlo Simulation models are based on a simulation of predetermined risk factors which allow the risk manager to calculate the return distribution. Again the VaR is determined as the relevant percentile of the obtained distribution. On the other hand the parametric approach is based on the estimation of a single parameter and has imbedded the simplifying assumptions of normal distribution of returns and linearity of portfolio returns with respect to the considered risk factors. These two hypothesis imply a normal distribution for portfolio returns. Consequently, it is possible to describe the returns' distribution simply with the first two moments and thus, the VaR can be calculated using the relevant percentile from a standard Z-distribution. In our empirical exercise we use a simple parametric approach, based on the beta estimation performed in Section 4, for evaluating several stock portfolios of the US market. Using the model proposed in equation 4, it is straightforward to define the variance of a portfolio as:

$$\sigma_p^2 = w'\beta\beta' w\sigma_m^2 + w'\Sigma w, \tag{15}$$

where w indicates a column vector of assets weights, β is a column vector of the estimated risk sensitivities, σ_m^2 is the variance of the market factor and Σ the diagonal variance-covariance matrix of idiosyncratic variances. It is a well known result that, as the number of assets in portfolio increases, the idiosyncratic risk becomes negligible. Thus, for a well diversified portfolio we can calculate the Value a Risk as:

$$VaR = \alpha_z \sqrt{w'\beta\beta' w\sigma_m^2}\sqrt{t}; \tag{16}$$

where α_z indicates the relevant percentile of the Z-distribution and t is the chosen time horizon.

[Table 3 about here.]

[Fig. 4 about here.]

The proposed VaR measure is tested on a set of equally weighted portfolios based on the SIZE, E-P, D-P and Industry portfolios. The betas are estimated from the time-varying sensitivities as proposed above, while the volatility of the market is simply calculated as the historical standard deviation of the market index returns. The chosen confidence interval is 5% one side losses and the selected time horizon is one month. For assessing the accuracy of the calculated Value at Risk we perform a Proportion of Failure (POF) test based on Kupiec (1995), calculated as:

$$LR = -\, 2ln\left(\frac{p_0^x(1-p_0)^{(n-x)}}{p^x(1-p)^{(n-x)}}\right), \tag{17}$$

where p_0 is the probability of an exception implied by the chosen confidence interval, n is the sample size, x is the actual number of exception and p is the Maximum-Likelihood estimator x/n of p_0.
Basically this test performs a Likelihood-Ratio with 5% level, based on the number of exceedences in any given sample, where the null hypothesis is that the estimated value for the exceedences matches its exact value.
Given its definition, the test is asymptotically χ^2 distributed with one degree of freedom; thus if the value of the test statistic exceeds the critical value of 3.84, the Value at Risk model can be seen as not reliable with a 95% confidence level. Table 3 shows the performance of the Value at Risk measure via a backtesting. The obtained results are more than encouraging. In all the analyzed portfolios the POF statistic is well below its critical value. Thus, we do not reject the null hypothesis of a reliable VaR measure. In order to put our results in perspective, we estimate both the same VaR measure with an Exponential Moving Average (EWMA) estimation of the market volatility, and a full parametric Value at Risk following the procedure proposed by Riskmetrics.[7] In the whole sample of the analyzed portfolios, employing the EWMA volatility does not change the accuracy of the proposed VaR measure. More importantly, in two out of four cases (Panel C and Panel D Table 3) the VaR measure based on the model outperforms the full parametric VaR measure.
For further assessing the potential of a GRID structure in solving a Risk Management problem, we test our model on a portfolio composed by fifty stocks randomly selected from the CRSP database. Interestingly enough, with the use of the GRID infrastructure, we have obtained a reduction of computation time proportional, to a certain extent, with the number of available clients. In particular we measure the performance of a GRID infrastructure on a cluster of eight nodes. The speed, shown in Figure 5 Panel A, is increasing dramatically when 3 clients are employed, gaining 193 seconds with respect to a single node, with a decrease of execution time from 426 to 233 seconds, corresponding to a relative increase in performance of $45,3\%$. Employing 5 nodes is giving a further improvement in the performance with a relative speed-up of 12%. For more than 5 nodes the gain become negligible, with an average time of execution of 205 seconds. For further investigate the performance of the employed GRID cluster, we separate the computation time of our exercise in time employed by the Genetic Algorithm, time employed for communication amongst nodes and time for Kalman filter computation. Figure 5 Panel B, shows the employed time by the three pieces of the whole algorithm incrementally, displaying clearly where the bottlenecks arise. First of all, the GA

[7] for a complete reference see http://www.riskmetrics.com

is not parallelized in our implementation, thus it contribute with a constant amount of time to the entire time spent in executing the algorithm. Secondly, the communication time is also contributing nearly constantly to the total execution time, showing even a minor time increase when the number of clients increases. Third, the execution time employed by the Kalman filter is, as expected, gaining the most from the Grid architecture; this is mainly due to the parallel structure of its code, that is taking full advantage of a distributed computational capability. Finally it is worth noting that the performance of the VaR is comforting, with a POF statistics well above the 5% critical value for all the randomly selected fifty stocks portfolios.

[Fig. 5 about here.]

6 Conclusion

The estimation of systematic risk has been one of the most studied topics in empirical finance. Historically important research contributions were departing from the classical one factor constant beta model, exploring the two possibilities of multi factors models and time varying sensitivities respectively. This paper refers to the latter stream of literature by estimating time varying sensitivities where the betas are supposed to be unobservable. By Estimating the model via a Kalman filter augmented with a genetic optimization algorithm, we are able to explain a large part of the observed time series variance in several stock portfolios of the US market.

Furthermore we are able to calculate a Value at Risk measure, based on the proposed model, on a GRID computing architecture. In this context, the use of GRID computing offers an opportunity to enhance the solution of computational demanding problems with decentralized data retrieval.

Our results are more than promising in showing the accuracy of the proposed model coupled with the capability of the GRID architecture in dealing, in a reasonable amount of time, with CPU use intensive calculations and huge data retrieval queries.

References

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

Basu, S., 1983, The relationship between earnings yield, market value, and return for nyse common stocks: Further evidence, *Journal of Financial Economics*, 12, 129–156.

Black, A., P. Fraser, and D. Power, 1992, Uk unit trust performance 1980-1989: A passive time-varying approach, *Journal of Finance*, 16, 1015–1033.

Blume, M., March 1971, On the assessment of risk, *Journal of Finance*, 26, 1–10.

Blume, M., June 1975, Betas and their regression tendencies, *Journal of Finance*, 30, 785–795.

Bollerslev, T., Apr. 1986, Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics*, 31, 307–327.

Bollerslev, T., R. F. Engle, and J. M. Wooldridge, 1988, A capital asset pricing model with time-varying covariances, *Journal of Political Economy*, 96, 116–31.

Bos, T. and P. Newbold, Jan. 1984, An empirical investigation of the possibility of stochastic systematic risk in the market model, *Journal of Business*, 57, 35–41.

Braun, P. A., D. B. Nelson, and A. M. Sunier, Dec. 1995, Good news, bad news, volatility, and betas, *Journal of Finance*, 50, 1575–1603.

Brenner, M. and S. Smidt, Sep 1977, A simple model of non-stationarity of systematic risk, *Journal of Finance*, 32, 1081–1092.

Brooks, R. D., R. W. Faff, and M. D. McKenzie, 1998, Time-varying beta risk of australian industry portfolios: A comparison of modelling techniques, *Australian Journal of Management*, 23, Jun.

Chen, N.-F., R. Roll, and S. A. Ross, Jul. 1986, Economic forces and the stock market, *Journal of Business*59, 383–403.

Collins, D. W., J. Ledolter, and J. Rayburn, Jul. 1987, Some further evidence on the stochastic properties of systematic risk, *Journal of Business*, 60, 425–448.

Daisuke Sasaki, Shigeru Obayashi, K. S. and R. Himeno, Barcelona, 11-14 September 2000, Multiobjective aerodynamic optimization of supersonic wings using navier-stokes equations, *ECCOMAS 2000*.

Del Vecchio, R. J. *Design of experiments: a primer for technologists.* Hanser/Gardner, 1997.

Engle, R. F., Jul. 1982, Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation, *Econometrica*, 50, 987–1008.

Fabozzi, F. J. and J. C. Francis, 1978, Beta as a random coefficient, *Journal of Financial and Quantitative Analysis*, 13, 101–116.

Fama, E. and K. French, 1992, The cross-section of expected stock returns, *Journal of Finance*, 47, 427–465.

Fama, E. and K. French, 1993, Common risk factors in the returns of stocks and bonds, *Journal of Financial Economics*, 33, 3–56.

Fama, E. and K. French, 1995, Size and book-to-market factors in earnings and returns, *Journal of Finance*, 50, 131–155.

Francis, J. C., Dec. 1979, Statistical analysis of risk surrogates for nyse stocks, *Journal of Financial and Quantitative Analysis*, 14, 981–997.

Ghysels, E., Apr. 1998, On stable factor structures in the pricing of risk: Do time-varying betas help or hurt?, *Journal of Finance*, 53, 549–573.

Giannopoulos, K., Jun. 1995, Estimating the time varying components of international stock markets' risk, *European Journal of Finance*, 1, 129–164.

Goldberg, D. E., 1989, Genetic algorithms in search, optimization & machine learning, *Addison-Wesley Publishing Company.*

Groenewold, N. and P. Fraser, 1999, Time-varying estimates of capm betas, mathematics and computers in simulation, *Mathematics and Computers in Simulation*, pages 531–539.

Hamilton, J. D. *Time Series Analysis.* (Princeton University Press, NJ), 1994.

Jagannathan, R. and Z. Wang, Mar. 1996, The conditional capm and the cross-section of expected returns, *Journal of Finance*, 51, 3–53.

Jorion, P. *Value at Risk: the New Benchmark for Controlling Market Risk, 2nd ed.* (McGraw Hill, NY), 2000.

Kupiec, P. H., 1995, Techniques for verifying the accuracy of risk management models, *Journal of Derivatives*, 3, 73–84.

Leto, A., E. Corso, R. Murri, A. Terpin, C. Zoicas, and S. Cozzini. Egrid project: Experience report on the implementation of a grid infrastructure for the analysis of financial data. In Society, I. C., editor, *International Conference on Information Technology Coding and Computing (ITCC)*, 2005.

Montgomery, D. C. *Design and analysis of experiments.* John Wiley & Sons, fourth edition, 1996.

Ohlson, J. and B. Rosenberg, Jan. 1982, Systematic risk of the crsp equal-weighted common stock index: A history estimated by stochastic-parameter regression, *Journal of Business*, 55, 121–145.

Periaux, J., M. Sefrioui, and B. Mantel. *Genetic algorithm and evolution strategies in engineering and computer science*, chapter GA multi objective optimization strategies for electromagnetic backscattering. John Wiley & Sons, 1997.

Poloni, C. and V. Pediroda. *Genetic algorithm and evolution strategies in engineering and computer science*, chapter GA coupled with computationally expensive simulations: tools to improve efficiency. John Wiley & Sons, 1997.

Rao, S. S. *Engineering optimization - Theory and practice.* John Wiley & Sons, third edition, 1996.

Roll, R. and S. A. Ross, 1980, An empirical investigation of the arbitrage pricing theory, *Journal of Finance*, 35, 1073–1103.

Roseberg, B., K. Reid, and R. Lanstein, 1985, Persuasive evidence of market inefficiency, *Journal of Portfolio Management*, 11, 9–17.

Ross, S. A., Dec. 1976, The arbitrage theory of capital asset pricing, *Journal*

of Economic Theory, 13, 341–360.
Schwert, G. W. and P. J. Seguin, Sep. 1990, Heteroskedasticity in stock returns, *Journal of Finance*, 45, 1129–1155.
Stattman, D., 1980, Book values and stock returns, *The Chicago MBA: A Journal of Selected Papers*, 4, 25–45.
Sunder, S., Sep. 1980, Stationarity of market risk: Random coefficients tests for individual stocks, *Journal of Finance*, 35, 883–896.

Fig. 1. Genetic Algorithm Performance

This figure plots the performance, in term of absolute value of the obtained likelihood function, with respect to the number of simulations employed. The GA is employed on the optimization process of a fifty stocks portfolio, randomly selected, with a time span of 33 years. All the data are from the CRSP database.

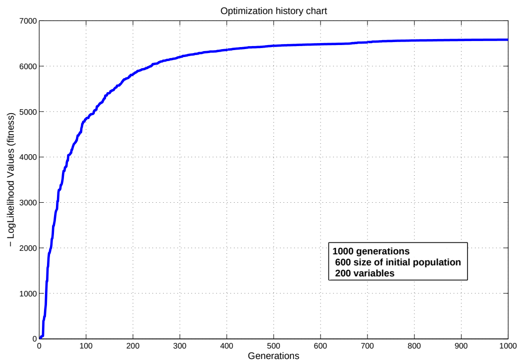
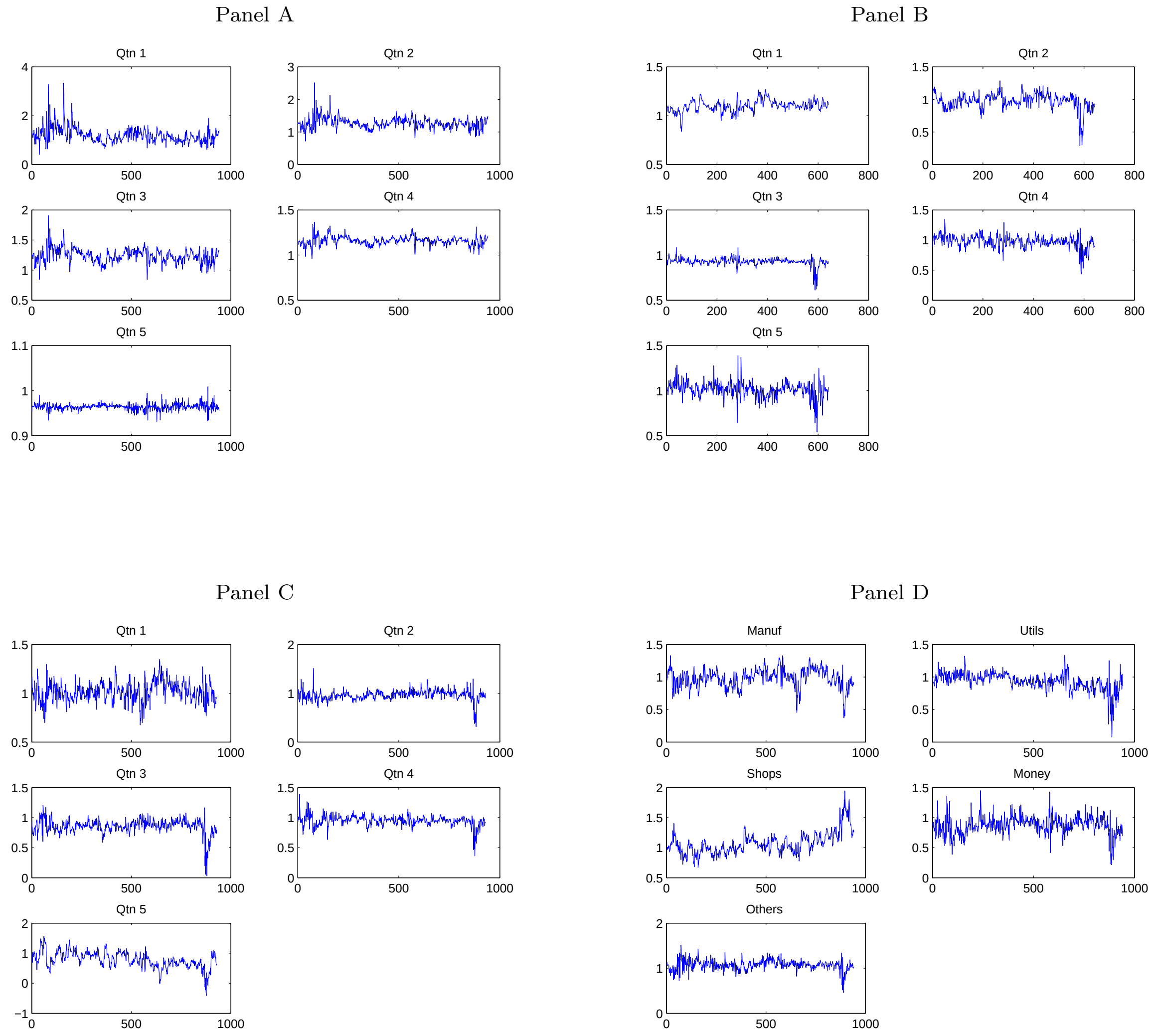


Fig. 2. Plot of Estimated Process

This figure plots the estimated path of the beta processes. Panel A through D show respectively the estimated processes from SIZE, E-P, D-P and Industries portfolios. For the E-P portfolios the sample size goes from July 1951 to December 2004, while for the D-P portfolios it goes from July 1927 to December 2004. The remaining data are from July 1926 to December 2004.



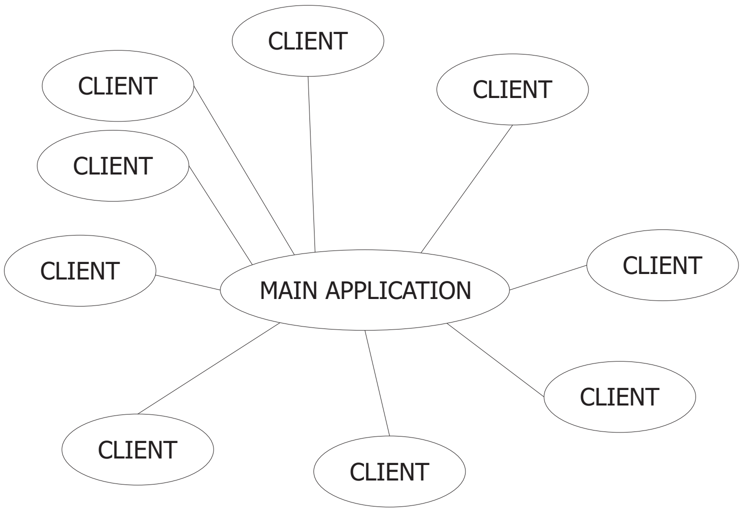


Fig. 3. A simple representation of a multithreading application.

Fig. 4. Plot of Value at Risk Backtesting

This figure plots the results from a Value a Risk Backtesting. Portfolios are equally weighted and based on the Kenneth French portfolios. All returns are monthly value weighted. The decay factor chosen for the Exponential moving average is 0.97, while its rolling window is five years. The left column shows the actual returns with a VaR losses band calculated with the Full Model approach while the right column shows the losses band calculated with the Full EWMA approach.

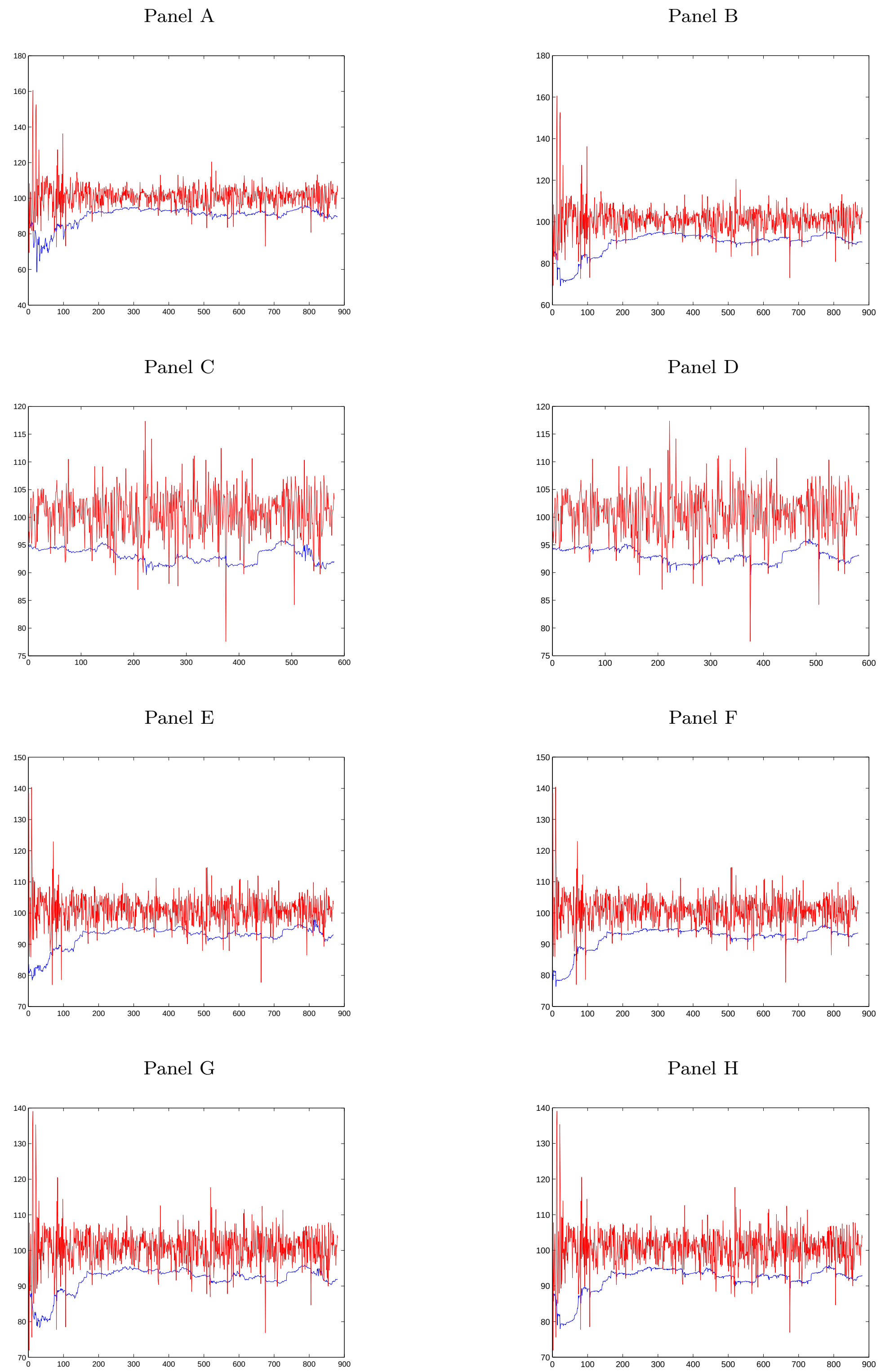


Fig. 5. Performance gain on a GRID architecture.

This figure plots the performance of a 8 nodes GRID cluster in performing a Risk Management application. The portfolio employed is generated randomly by picking fifty stocks from the CRSP database, with a time span of 33 years. Panel A shows the total computational time, while Panel B shows the time added, incrementally, to the total computational time by the Genetic Algorithm, the communication time and the Kalman filter algorithm respectively.

Panel A | Panel B

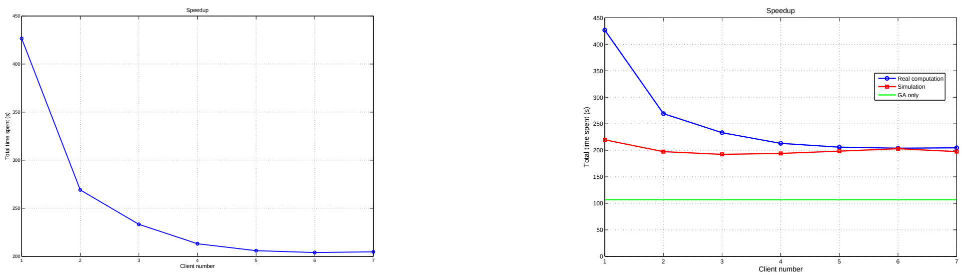


Table 1
Descriptive Statistics of Financial Series

This table reports the mean and standard deviation of the analyzed stock portfolios. The portfolios are from the Kenneth French website. All returns are monthly value weighted.

* Sample starting July 1927.

** Postwar data available from July 1951.

Panel A: Size Portfolios					
	Qtn 1	Qtn 2	Qtn 3	Qtn 4	Qtn 5
Entire Sample					
Mean	1.39%	1.26%	1.20%	1.12%	0.93%
Std	9.33 %	7.74%	7.07%	6.34%	5.25%
Postwar Sample					
Mean	1.27%	1.24%	1.19%	1.16%	1.00%
Std	5.88%	5.47%	5.02%	4.72%	4.11%

Panel B: E-P Portfolios**					
	Qtn 1	Qtn 2	Qtn 3	Qtn 4	Qtn 5
Postwar Sample					
Mean	0.84%	1.01%	1.10%	1.33%	1.46%
Std	4.90%	4.19%	4.23%	4.16%	4.71%

Panel C: D-P Portfolios *					
	Qtn 1	Qtn 2	Qtn 3	Qtn 4	Qtn 5
Entire Sample					
Mean	0.96%	0.98%	0.94%	1.12%	1.10%
Std	5.98%	5.36%	5.49%	5.49%	6.11%
Postwar Sample					
Mean	1.04%	1.07%	1.02%	1.19%	1.17%
Std	5.07%	4.44%	4.18%	4.00%	3.88%

Panel D: Industry Portfolios					
	Manuf	Utils	Shops	Money	Other
Entire Sample					
Mean	1.03%	0.97%	0.96%	1.13%	0.97%
Std	5.47%	5.59%	5.75%	5.86%	6.49%
Postwar Sample					
Mean	1.08%	1.02%	1.02%	1.23%	1.08%
Std	4.45%	4.08%	5.27%	5.04%	4.83%

Table 2
Parameter Estimation

This table reports the estimated parameters of the analyzed stock portfolios. The portfolios are from the Kenneth French website. All returns are monthly value weighted.
* Sample starting July 1927.
**Data available from July 1951.

Panel A: Size Portfolios					
	Qtn 1	Qtn 2	Qtn 3	Qtn 4	Qtn 5
$\overline{\beta}$	1.099	1.311	1.292	1.181	0.965
	(0.054)	(0.041)	(0.029)	(0.015)	(0.006)
α	0.850	0.817	0.839	0.785	0.320
	(0.020)	(0.031)	(0.030)	(0.061)	(0.131)
σ	0.045	0.040	0.015	0.005	0.003
	(0.003)	(0.006)	(0.002)	(0.001)	(0.001)
R^2	0.667	0.818	0.902	0.949	0.985

Panel B: E-P Portfolios**					
	Qtn 1	Qtn 2	Qtn 3	Qtn 4	Qtn 5
$\overline{\beta}$	1.115	1.013	0.925	1.002	1.011
	(0.025)	(0.032)	(0.019)	(0.031)	(0.031)
α	0.880	0.830	0.636	0.697	0.589
	(0.045)	(0.029)	(0.148)	(0.053)	(0.108)
σ	0.003	0.015	0.011	0.037	0.051
	(0.001)	(0.004)	(0.002)	(0.007)	(0.009)
R^2	0.899	0.903	0.863	0.803	0.748

Panel C: D-P Portfolios *					
	Qtn 1	Qtn 2	Qtn 3	Qtn 4	Qtn 5
$\overline{\beta}$	0.920	0.953	0.812	0.965	0.412
	(0.030)	(0.021)	(0.026)	(0.024)	(0.142)
α	0.841	0.750	0.801	0.710	0.972
	(0.026)	(0.032)	(0.028)	(0.054)	(0.008)
σ	0.017	0.018	0.019	0.025	0.014
	(0.003)	(0.002)	(0.004)	(0.004)	(0.003)
R^2	0.916	0.928	0.895	0.869	0.830

Panel D: Industry Portfolios					
	Manuf	Utils	Shops	Money	Other
$\overline{\beta}$	1.013	0.891	1.185	0.870	1.070
	(0.050)	(0.041)	(0.047)	(0.032)	(0.025)
α	0.935	0.898	0.916	0.756	0.758
	(0.016)	(0.021)	(0.015)	(0.045)	(0.038)
σ	0.009	0.015	0.013	0.036	0.023
	(0.003)	(0.007)	(0.002)	(0.004)	(0.004)
R^2	0.883	0.934	0.848	0.653	0.884

Table 3
Value at Risk Backtesting

This table reports the results of a Value at Risk Backtesting on the analyzed stock portfolios. The portfolios are equally weighted based on the Kenneth French portfolios. All returns are monthly value weighted. The decay factor chosen for the Exponential moving average is 0.97, while its rolling window is five years.
* Sample starting July 1927.
**Data available from July 1951.

Panel A: Size Portfolio			
	Expected	Actual	LR test
VaR Full Model	44.000	40.000	0.404
VaR EWMA Model	44.000	40.000	0.404
VaR Full EWMA	44.000	40.000	0.404

Panel B: E-P Portfolio**			
	Expected	Actual	LR test
VaR Full Model	29.000	28.000	0.040
VaR EWMA Model	29.000	27.000	0.156
VaR Full EWMA	29.000	27.000	0.156

Panel C: D-P Portfolio*			
	Expected	Actual	LR test
VaR Full Model	43.000	43.000	0.005
VaR EWMA Model	43.000	42.000	0.051
VaR Full EWMA	43.000	34.000	2.331

Panel D: Industry Portfolios			
	Expected	Actual	LR test
VaR Full Model	44.000	41.000	0.227
VaR EWMA Model	44.000	41.000	0.227
VaR Full EWMA	44.000	36.000	1.647