To appear in *Quantitative Finance*, Vol. 00, No. 00, Month 20XX, 1–25

Acknowledgements

We thank Bloomberg for providing the data. We also thank Pierpaolo Vivo and Alessia Annibale for their input. Anshul Verma wishes to thank EPSRC for providing funding. We also declare no conflicts of interest.

To appear in *Quantitative Finance*, Vol. 00, No. 00, Month 20XX, 2–25

A cluster driven log-volatility factor model: a deepening on the source of the volatility clustering

Keywords: Econophysics, Empirical Finance, Volatility Clustering, Clustering, Multi Factor Models, Dimensionality Reduction

Anshul Verma † [a] , Riccardo Junior Buonocore ‡ [a] and Tiziana Di Matteo § [a,b,c]

[a] Department of Mathematics, King's College London, The Strand, London, WC2R 2LS, UK
[b] Department of Computer Science, University College London, Gower Street, London, WC1E 6BT, UK
[c] Complexity Science Hub Vienna, Josefstaedter Strasse 39, A 1080 Vienna

(*Received 00 Month 20XX; in final form 00 Month 20XX*)

We introduce a new factor model for log volatilities that considers contributions, and performs dimensionality reduction, at a global level through the market, and at a local level through clusters and their interactions. We do not assume a-priori the number of clusters in the data, instead using the Directed Bubble Hierarchical Tree (DBHT) algorithm to fix the number of factors. We use the factor model to study how the log volatility contributes to volatility clustering, quantifying the strength of the volatility clustering using a new non parametric integrated proxy. Indeed finding a link between volatility and volatility clustering, we find that a global analysis reveals that only the market contributes to the volatility clustering. A local analysis reveals that for some clusters, the cluster itself contributes statistically to the volatility clustering effect. This is significantly advantageous over other factor models, since it offers a way of selecting factors in a statistical way, whilst also keeping economically relevant factors. Finally, we show that the log volatility factor model explains a similar amount of memory to a Principal Components Analysis (PCA) factor model and an exploratory factor model.

1. Introduction

Volatilities are an important factor for the estimation of risk (Bouchaud and Potters 2009) and for models aiming at dynamically modelling price and what the rational, fair price should be under such models (Hull and White 1987, Hull 2006). However, the effect of volatility clustering, and particularly its unclear link with how volatilities are correlated with each other, complicates this process. This causes a problem due to the high dimensionality of the correlation matrix between the log volatilities that is also subject to noise (Bun *et al.* 2017), which makes it difficult to identify meaningful information about what drives the volatility and volatility clustering. This problem is also relevant in multivariate volatility modelling since most popular methods such as multivariate General Autoregressive Conditional Heteroskedasticity (GARCH) (Bauwens *et al.* 2006), stochastic covariance (Clark 1973) and realised covariance (Andersen *et al.* 2003) suffer from the curse of dimensionality and an increase in the number of parameters needed. One such way of tackling this

†Corresponding author. Email: anshul.verma@kcl.ac.uk. +447740779724
‡Email: riccardo_junior.buonocore@kcl.ac.uk. +447549919717
§ Email: tiziana.di_matteo@kcl.ac.uk. +4402078482223

problem is through dimensionality reduction, which is a general class of methods that aims to reduce high dimensional datasets to a reduced form which is a faithful representation of the original dataset (Van Der Maaten *et al.* 2009), and is also related to noise reduction of the dataset.

One such method of dimensionality reduction of correlation matrices is Principal Component Analysis (PCA) (Jolliffe 1986). It aims to transform the original correlation matrix into an orthogonal basis. For square correlation matrices, which are those that we consider in this paper, this essentially means calculating the eigenvalues and their respective eigenvectors. The first eigenvector (called the first principal component) has the highest variance and explains most of the variability in the data, the second eigenvector (called the second principal component) has the second highest variance and explains less variability than the first principal component, and so on. The method has been applied to finance mainly through portfolio optimisation to produce sets of orthogonal portfolios (Darbyshire 2017). A paper which uses PCA in the context of volatility modelling is Alexander (2002), where the author extracts the first few principal components and uses them to calibrate a multivariate GARCH model, with a further extension proposed in Zhang and Chan (2009). The main drawback of PCA is that it is not clear how many principal components i.e. factors to keep, as either too many principal components are kept or the methods used to select the components are heuristic and subjective in nature (Jolliffe 1986). In Plerou *et al.* (2002), the authors suggest to keep the number of principal components according to the Marchenko-Pastur distribution with a further refinement made in Majumdar and Vivo (2012) and previously in Jackson (1993), however in Livan *et al.* (2011) it is pointed out valuable information may still be lost.

A highly related class of methods in dimensionality reduction are called factor models (Sharpe 1964, Roll and Ross 1980, Fama and French 1993, Chicheportiche and Bouchaud 2015). Factor models are used to describe the dynamical evolution of time series, assuming that there exist common factors through the asset's sensitivity, often called responsiveness, to changes in the value of these factors. Dimensionality reduction is then achieved through the description of the time series as the number of factors is smaller than the number of stocks. Factor models have widespread use in finance due to their relative (or at least superficial) simplicity in comparison to other models of returns series (Sharpe 1964, Fama and French 1993, 1996, Engel *et al.* 2015, Chicheportiche and Bouchaud 2015). Factor models can be split into two varieties: exploratory, which assume no underlying structure to the data, and confirmatory, which tests relationships between known factors (Thompson 2004).

However, similarly to PCA a question of how we should choose the factors arises. One such answer can be categorised by assuming that we have some prior knowledge of the factors. The simplest and earliest factor model which falls under this category is the Capital Asset Pricing Model (CAPM)(Sharpe 1964, Merton 1973, Zabarankin *et al.* 2014, Barberis *et al.* 2015). It emerges from the extremely popular Markowitz scheme of portfolio optimisation (Markowitz 1952), which says it is better to spread an investment across a class of stocks in order to reduce the total risk of the portfolio. CAPM develops this further by saying that the non diversifiable risk, or systematic risk, comes from the stock's exposure to changes in the market and the corresponding sensitivity to this change.

A very well known factor model which has multiple factors, rather than just one like CAPM, is the 3-factor Fama-French factor model (Fama and French 1992, 1993, 1996, Connor *et al.* 2012, Faff *et al.* 2014). In this factor model, the first factor comes again from the exposure to the market risk with two extra factors: the small minus big (SMB) and the high minus low (HML)(Fama and French 1993, 1996). The SMB factor follows the observation by Fama and French that stocks with a smaller market cap, which is the market value of the stock used as a proxy of size, tend to give additional returns. Equivalently, the HML factor represents the book/market ratio i.e. the ratio of the total value of the assets owned by the company associated to a stock relative to the stock's market value, and is positively correlated with additional returns. The aim of the HML factor is to evaluate whether stocks have been undervalued by the market, where the book/market ratio exceeds 1, and thus have the potential for larger returns. Recently, the Fama-French model has been extended to include 5 factors (Fama and French 2015). The arbitrage pricing theory (APT) is also a more generalised multi factor model, except it states that returns are a linear function of macro economic factors (Roll and

Ross 1980, Chen *et al.* 1986). In APT however, there is no indication of exactly how many and what factors should be included, which then introduces an ad-hoc nature to the types and numbers of factors included in the model.

The above factor models share the fact that the number and nature of the factors are somewhat exogenous in the sense that they are determined by economic intuition on what should drive financial returns. Unfortunately, it has been pointed out that there is weak evidence for CAPM (Fama and French 1992), both Fama-French 3 and 5 factor models and to some manifestations of the APT (Reinganum 1981, Faff 2004, Grauer and Janmaat 2010, Racicot and Rentz 2016), underlying the issue that these factors cannot explain the cross dependence of assets. Instead, there is a strand of literature which invokes factors that are extracted from the financial data itself thus meaning that the factors are endogenous (Malevergne and Sornette 2004, Tumminello *et al.* 2007, Chicheportiche and Bouchaud 2015). In essence, it has been shown that the collective action of assets is what induces the factors, giving support to this type of determination of factors (Malevergne and Sornette 2004), an approach we shall adopt here. Another difference is that the above factor models are mainly applied to returns rather than volatilities.

In this paper, we instead build a new factor model of log volatilities that aims to reduce the dimensionality by considering contributions globally from the market and more locally to the clusters and their interactions. The number of factors is fixed by the Directed Bubble Hierarchical Tree (DBHT) clustering algorithm (Song *et al.* 2012, Musmeci *et al.* 2015), which therefore means we make no prior assumption on the number of clusters and thus the number of factors to be considered. Using this factor model between volatilities, we aim to study the link between the univariate volatility clustering and the multivariate correlation structure of volatilities. We will see that whilst over the entire market the only significant contributor that affects the memory is the market, individual clusters may have different properties where the cluster contributions and interactions are more significant. This offers a method to statistically select factors based on memory reduction. We also note that for the clusters which significantly reduce their own memory are mostly made up by stocks from particular industries, offering an economic interpretation for the makeup of the cluster modes. We can thus select the factors in a statistical manner like in PCA, but also retain the appealing economic interpretation like in CAPM and Fama-French.

The structure of the paper is as follows: Section 2 desribes the dataset, Section 3 introduces a new factor model for log volatilities, Section 4 describes how we select factors based on their memory reduction using a new non parametric integrated proxy for the strength of the volatility clustering, Section 5 we explore how the empirical link between volatility clustering strength and volatility cross correlation can be explained. The penultimate Section 7 compares our factor model to a PCA inspired factor model and an exploratory factor analysis model in terms of their memory reduction performance. Finally, we draw some conclusions in Section 8.

2. Dataset

The dataset we shall use consists of the daily closing prices of 1270 stocks in the New York Stock Exchange (NYSE), National Association of Securities Dealers Automated Quotations (NASDAQ) and American Stock Exchange (AMEX) from 01/01/2000 to 12/05/2017, which makes 4635 points for each price time series. As anticipated in the introduction, we perform cross correlation analysis. We therefore make sure that the stocks are aligned through the data cleaning procedure described in 9.A, which leaves our dataset with $N = 1202$ stocks. We calculate the log-returns time series of a given stock i, $r_i(t)$, defined as:

$$r_i(t) = \ln p_i(t+1) - \ln p_i(t), \tag{1}$$

where $p_i(t)$ is the price time series of stock i, and $r_i(t)$ is a time series of length $T = 4364$. After standardising $r_i(t)$ so that it has zero mean and a variance of 1, we define the proxy we shall use for

the volatility as $\ln|r_i(t)|$ i.e. the log absolute value of returns (Taylor 1994).

3. Log-volatility factor model

In this section we describe a new factor model for log volatilities, which we shall use to uncover the relationship between the univariate volatility clustering effect and the cross correlations between volatilities. Let us recall that a general factor model is given by:

$$r_i(t) = \sum_{p=1}^{P} [\beta_{ip} f_p(t) + \alpha_{ip}] + \epsilon_i(t), \tag{2}$$

where $r_i(t)$ are the log returns for asset i, f_p are the $p = 1, 2, ..., P$ factors. β_{ip} is their respective sensitivities/responsiveness, which quantifies how $r_i(t)$ reacts to changes in f_p. α_{ip} is the intercept and $\epsilon_i(t)$ are residual terms with zero mean. Firstly, we define the log volatility term we want to study. Most stochastic volatility models (where the volatility is assumed to be random and not constant) assume that the returns for the stock i follow an evolution according to, which is (Gatheral 2011)

$$r_i(t) = \delta(t) e^{\omega_i(t)}, \tag{3}$$

where $\delta(t)$ is a white noise with finite variance and $\omega_i(t)$ are the log volatility terms. The exponential term encodes the structure of the volatility and how it contributes to the overall size of the return. Taking the absolute value of (3) and the log of both sides, Eq. (3) becomes

$$\ln|r_i(t)| = \ln|\delta(t)| + \omega_i(t), \tag{4}$$

from which we see that working with $\ln|r_i(t)|$ has the added benefit of making the proxy for volatility, $\omega_i(t)$ additive, which in turn makes the volatility more suitable for factor models. In its full shape our factor model reads as

$$\omega_i(t) = \beta_{i0} I_0(t) + \alpha_{i0} + \beta_{ik} I_k(t) + \sum_{k'=1}^{n-1} \beta_{ik'} I_{k'}(t) + \epsilon_i(t). \tag{5}$$

The factor model in (5) is therefore a special case of eq. (2), where here there are $n + 1 = P$ factors with $f_p(t) = I_{p-1}(t)$ for $p = 1, ..., n + 1$. β_{i0} is the responsiveness of stock i with respect to changes in $I_0(t)$, α_{i0} is the excess volatility compared to the market. The first two terms of eq. (5) represent the market factor, which is the widely observed effect of the market affecting all stocks i.e. the co-movement of all stocks at once Laloux *et al.* (1999)Plerou *et al.* (2002)Singh and Xu (2016). We also define the term $c_i(t)$

$$c_i(t) = \beta_{ik} I_k(t) + \sum_{k'=1}^{n-1} \beta_{ik'} I_{k'}(t) + \epsilon_i(t). \tag{6}$$

This represents the volatility that is not explained by the market, and where β_{ik} are the responsiveness for the k cluster mode $I_k(t)$ which i is a member of. In the sum in Eqns. (5) and (6), the $\beta_{ik'}$ are the responsiveness to changes in $I_{k'}(t)$ which are the cluster modes of the clusters $k' \neq k$ i.e. the clusters i is not a member of. In (6) the first term is for the cluster factor and represents the co-movement of the stock with its cluster. The sum in Eq. (6) represents the interactions the stock i has with other clusters, where the strength of the interactions are quantified and defined through the $\beta_{ik'}$. In the following subsections, we describe the fitting procedure of Eq. (5).

3.1. *Market Mode*

The log volatility term $\omega_i(t)$ in Eq. (5) can be rewritten as

$$\omega_i(t) = \beta_{i0} I_0(t) + \alpha_{i0} + c_i(t), \tag{7}$$

where $I_0(t)$ is defined as

$$I_0(t) = \sum_{i=1}^{N} \xi_i \ln |r_i(t)|. \tag{8}$$

ξ_i is the weight of stock i for the market mode, with the consequence that the weights define a pseudo-index. We see from Eq. (7) that $c_i(t)$ then also becomes the residue as a by product of performing the linear regression. In Table 1 of section 10, we show two examples of the regression coefficients for the market mode for two selected stocks Coca Cola Enterprises (KO) and Transoceanic (RIG). We report the values of β_{i0} and α_{i0} for the weighted scheme and for the equal weights scheme detailed in 9.B, along with their p values for the null hypothesis of each of the coefficients being 0. As we can see from Table 1, at the 5% level, the null hypothesis is rejected for all β_{i0} for both weighting schemes, which means that we can conclude that the β_{i0} are significant. For the α_{i0} the null hypothesis is rejected for both stocks in the equal weights case, and for the weighted case it is rejected only for RIG, and for these cases we can conlude that the α_{i0} are non-zero.

3.2. *DBHT output*

The next step of the calibration procedure concerns the identification of the clusters, which is relevant for the $c_i(t)$ term defined in eq. (6). Now, we need to find what the cluster structure is, which we do by first calculating $\mathbf{G}$, which is the cross correlation matrix between $c_i(t)$, defined as

$$G_{ij} = \frac{1}{T} \sum_{t=1}^{T} c_i(t) c_j(t). \tag{9}$$

We then apply the clustering algorithm to $\mathbf{G}$. We use the clustering algorithm after removing the market mode since this gives a more stable clustering (Borghesi *et al.* 2007). We shall use the Directed Bubble Hierarchical Tree, DBHT (Song *et al.* 2012, Musmeci *et al.* 2015, N. Musmeci 2016), to find the cluster membership of stocks. DBHT is used because as compared to other hierarchical clustering algorithms it provides the best performance in terms of information retrieval (Musmeci *et al.* 2015). Using the DBHT algorithm also means that we make no prior assumption on exactly how many factors for the clusters should be included, instead extracting them directly from the data. We can see from Table 2 that the DBHT algorithm identifies a total of $K = 29$ clusters, with the largest cluster comprising of 172 stocks and the smallest cluster comprising of 5 stocks. The average cluster size is 41.4.

3.3. *Cluster modes and interactions*

Once the number and composition of each cluster is identified, we can associate a factor to each cluster k. The interactions are then characterised through the responsiveness $\beta_{ik'}$ where $k \neq k'$ i.e. how $c_i(t)$ changes w.r.t to $I_{k'}(t)$. We define the cluster mode for cluster k, $I_k(t)$, again as a weighted average of volatilities for the assets in k

$$I_k(t) = \sum_{i \in cluster\ k} \xi_{ik} c_i(t). \tag{10}$$

ξ_{ik} is the weight for stock i which is in cluster k. From Eq. (6), we see that similarly to the market mode case, we can determine β_{ik}, $\beta_{ik'}$ and α_{ik}, $\alpha_{ik'}$ by linearly regressing $c_i(t)$ against $I_k(t)$ and $I_{k'}(t)$. We use elastic net regression Zou and Hastie (2005) to find β_{ik} and $\beta_{ik'}$ to take into account the possibility of $I_k(t)$ and $I_{k'}(t)$ being correlated, whilst also allowing for some of the $\beta_{ik'}$ to be 0 as i may not interact with cluster k'. More details about elastic net regression are provided in 9.C.

4. Empirical link between volatility clustering and volatility cross correlation

As anticipated in the introduction, we choose which factors are relevant for the decomposition in eq. (5), by measuring what the impact is of each cluster on the volatility clustering. Before turning our attention to this analysis, let us introduce the volatility clustering proxy we use in the rest of the paper.

4.1. *Volatility Clustering*

Volatility clustering is one of the so called stylised facts of financial data, and expresses the idea that returns are not independent since volatilities are autocorrelated (Cont 2001, Chakraborti *et al.* 2011). The autocorrelation function (ACF) $\kappa(L)$ is defined as

$$\kappa(L) = corr(\ln|r(t+L)|, \ln|r(t)|) \tag{11}$$

$$= \frac{\boldsymbol{E}\left[\ln|r(t+L)|\ln|r(t)|\right]}{\sigma^2}, \tag{12}$$

where L is the lag and σ^2 is the variance of the process of $\ln|r(t)|$ and note that we use log absolute value returns as a proxy for volatility. The interpretation of this result is that large changes in returns are usually followed by other large changes in returns, or that the returns retain a memory of previous values (Mandelbrot 1997). For this reason, volatility clustering can also be called the memory effect. $\kappa(L)$ has also been assumed to follow a power law decay:

$$\kappa(L) \sim L^{-\beta^{\text{vol}}}, \tag{13}$$

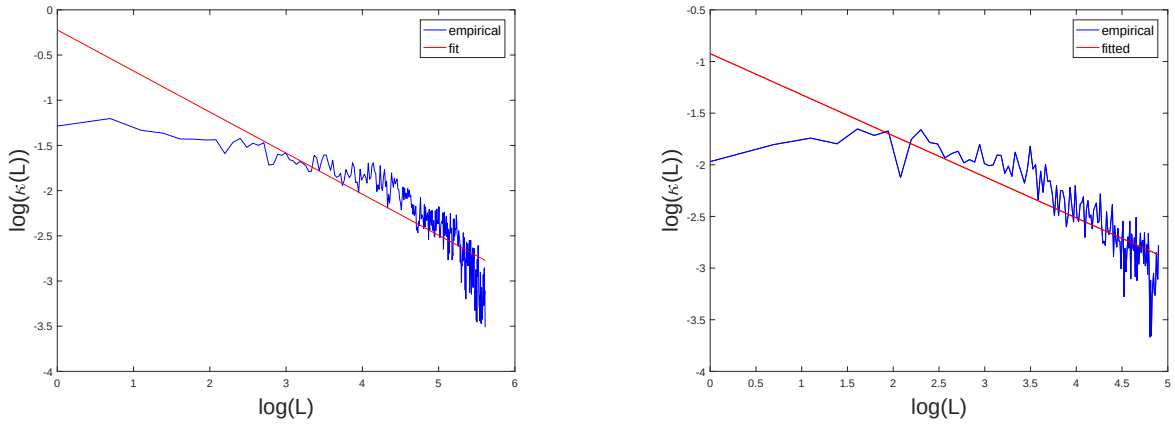
where β^{vol} describes the strength of the memory effect. A lower value of β^{vol} indicates that more memory of past values is kept. To compute β we transform eq. (13) into loglog scales and compute the slope of the linear best fit, which gives us the exponent β^{vol}. We shall compute β^{vol} using the Theil-Sen procedure rather than using standard least squares since it is more robust to outliers Theil (1992). We report in figure 1 the function $\kappa(L)$ for Coca Cola Enterprises Inc. in figure 1a and Transoceanic in figure 1b, both in loglog scale, with the linear best fit also plotted. We define the entries E_{ij} of the empirical volatility cross correlation $\mathbf{E}$ as

$$E_{ij} = \sum_{t=1}^{T} \ln|r_i(t)|\ln|r_j(t)|. \tag{14}$$

The proxy used for the volatility cross correlation is the average cross correlation for stock i, ρ_i^{vol}, is defined as

$$\rho_i^{\text{vol}} = \frac{1}{N-1}\sum_{i\neq j=1}^{N} E_{ij} \tag{15}$$

Using the proxies for volatility clustering and the volatility cross correlation, Micciche (2013) finds



(a) Coca Cola Enterprises Inc. $\beta^{\text{vol}} = 0.4544$ (b) Transoceanic $\beta^{\text{vol}} = 0.3975$

Figure 1.: Empirical ACF of the absolute returns (blue solid lines) for Coca Cola Co. (KO) in figure 1a and Transocean (RIG) in figure 1b in log-log scale. The linear best fit is also shown in red dashed lines.

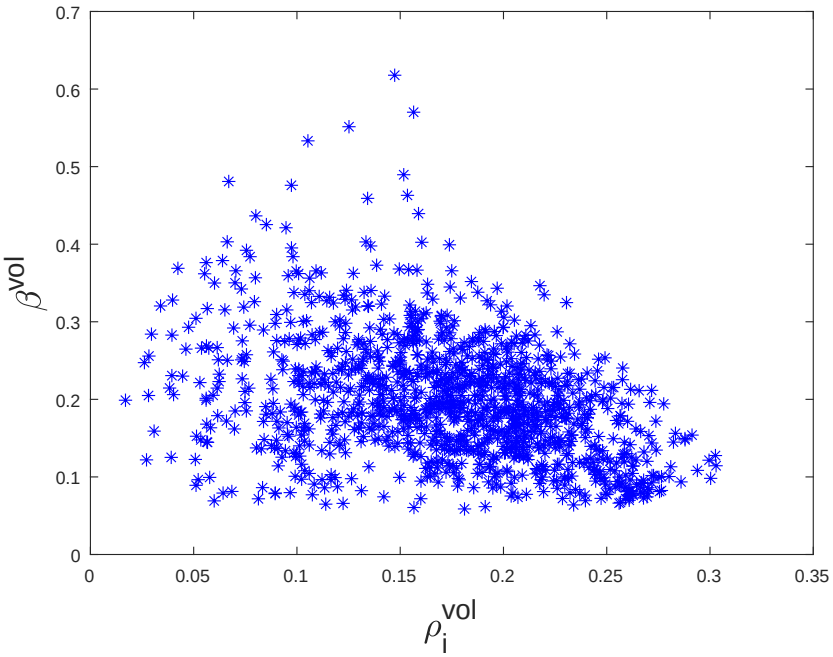
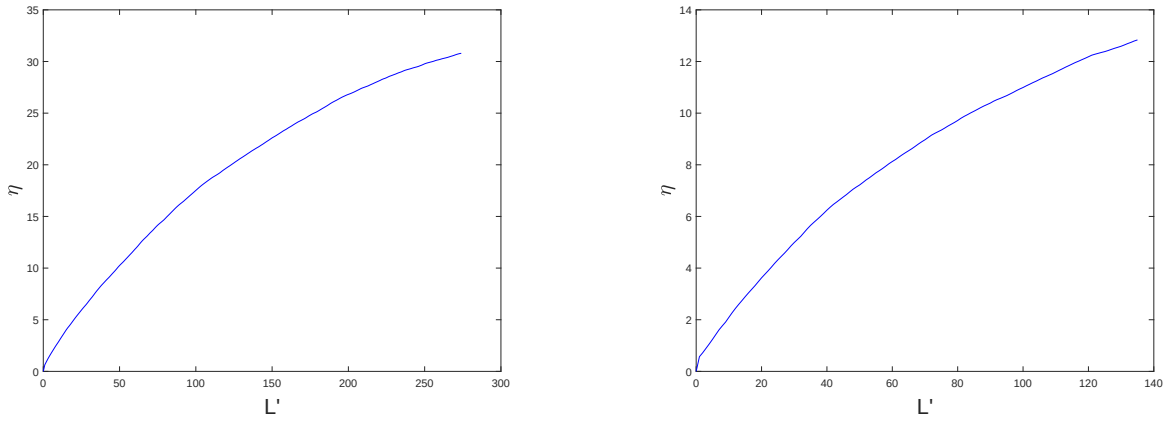


Figure 2.: Negative dependence between ρ_i^{vol} and β_i^{vol}. The negative relationship was tested using 1 sided Spearman's rank correlation at the 5% level with the null hypothesis of there being no correlation and was rejected, which confirms the result of Micciche (2013) on our data.

a negative relationship between ρ_i^{vol} and β_i^{vol}, which we confirm holds on our data set of daily data and using $\ln|r(t)|$, rather than the original high frequency data and $|r(t)|$ used in Micciche (2013), in figure 2. The main consequence of this result is that it implies that the more the volatility of a stock i is linked to other stocks, the stronger the memory effect and thus it retains more information about previous values of volatility, linking the strength of volatility clustering with the cross correlation matrix between volatilities.



(a) Coca Cola Enterprises Inc. (b) Transoceanic

Figure 3.: Integrated proxy η as a function of the lag L' where η is integrated over [1,L'] until $L' = L_{cut}$. Fig. 3a is for Coca Cola Co. and fig. 3b for Transocean

4.2. *Non parametric memory proxy*

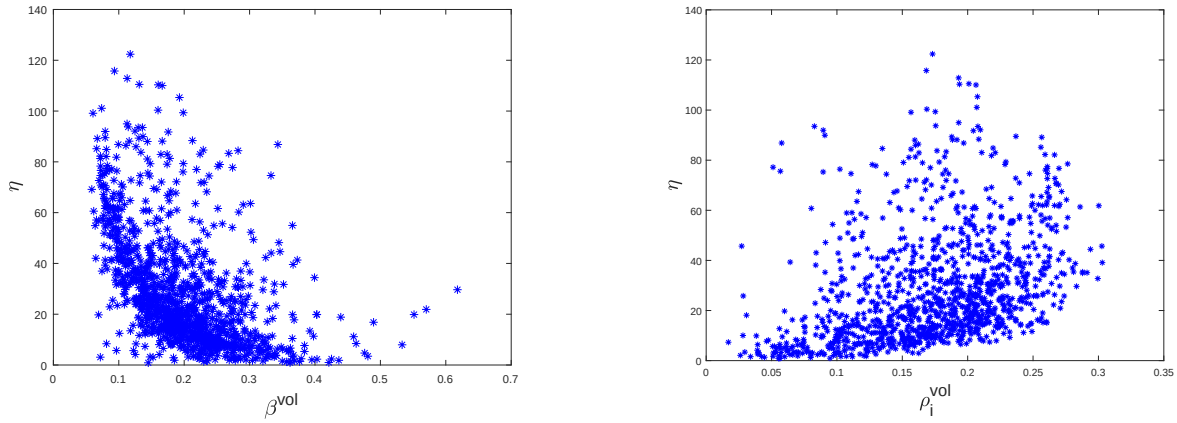
As already mentioned, the β^{vol} power law exponent that is fitted to the autocorrelation function of the absolute returns is a proxy for the strength of the memory effect: the lower the beta the stronger the memory effect. The use of the power law to quantify the memory effect is parametric as we *assume* the tail decays as a power law through the exponent β. In light of this, we instead introduce a new model free proxy, η, by integrating the autocorrelation function over time lags L until L_{cut}, which we define as the standard Bartlett Cut at the 5% level Box *et al.* (2015).

$$\eta = \int_{L=1}^{L_{cut}} \kappa(L) dL \ , \tag{16}$$

where $\kappa(L)$ is the empirical autocorrelation matrix of the log absolute returns as a function of the lag L. With this proxy the larger the value of η the greater the degree of the memory effect (in the β proxy this corresponds to larger values of the exponent). We plot η as a function of the upper limit in the integrand of eq. (16), where the upper limit is allowed to be in the interval $[1, L_{cut}]$. As we can see from both plots in figure 3, the line is much smoother showing that the η proxy is much more robust with respect to the noisy signal of the empirical ACF. The median value reported across all stocks is 20.7318 ± 8.6901, where the error is computed across all stocks using the median absolute deviation (MAD) for η_i defined as

$$MAD = median\left(|\eta_i - median(\eta_i)|\right). \tag{17}$$

We have also plotted the β as a memory effect proxy vs η in figure 4a, which as expected shows a decreasing relationship between η and the β memory proxy, which is the one used in the literature, since a larger memory effect means a higher η, but lower β. This provides justification for our use of η. This proves that η is coherent with β^{vol} and thus can be used a proxy for the strength of the memory effect. Figure 4b which is a plot of ρ_i^{vol} vs η confirms the main result of Micciche (2013) using η instead of β^{vol}, and was tested using Spearman's rank correlation at the 5% level for the null hypothesis, which was rejected, of there being no correlation between ρ_i^{vol} and η versus alternative hypothesis of there being significant positive relationship. Our proxy can therefore also confirm the result of Micciche (2013).



(a) $\beta^{\rm vol}$ vs $\beta^{\rm vol}$ (b) $\rho_i^{\rm vol}$ vs η

Figure 4.: In figure 4a we plot the $\beta^{\rm vol}$ power law exponent proxy for the strength of the memory effect vs c the integrated proxy. In figure 4b we plot the relationship between $\rho_i^{\rm vol}$ and η defined in the text. The decreasing relationship in figure 4a and the increasing relationship in figure 4b was tested using the Spearman's rank correlation at the 5% level and was rejected in both cases.

5. Memory filtration

In this section, by means of the factor model introduced in Eq. (5) and also by means of the η proxy introduced in the previous subsection, we want to understand the origin of the empirical link between the memory strength and the volatility cross-correlation. This analysis will in turn be also fundamental for the cluster mode selection in our model. The main intuition is that the market mode, the cluster mode and the interaction modes all bring relevant information about the memory of a certain stock's time-series.

5.1. *Assessing the memory contributions*

Let us here describe the method we use in order to understand the contribution to the memory of each term in the factor model in eq. (5). For every time-series, say for stock i, we follow a step-by-step procedure, by measuring the value of the proxy η_i for the following four times:

(i) on the plain time-series $\eta_{i,PL}$;
(ii) on the residual time-series once the market mode is removed $\eta_{i,MM}$;
(iii) on the residual time-series once the market mode and the cluster mode (of the the cluster the stock belongs to) are removed $\eta_{i,CM}$;
(iv) on the residual time-series once market, cluster and interaction mode are all removed. In order to make a quantitative comparison $\eta_{i,IM}$.

The next step consists in assessing the memory reduction after each removal. We do so by computing the ratio of two subsequently computed value of η_i. For stock i thus we have that

(i) $\frac{\eta_{i,MM}}{\eta_{i,PL}}$ defines the reduction in memory induced by the market mode;
(ii) $\frac{\eta_{i,CM}}{\eta_{i,MM}}$ defines the reduction in memory induced by the cluster mode once the market mode is removed;
(iii) $\frac{\eta_{i,IM}}{\eta_{i,CM}}$ defines the reduction in memory induced by the interaction mode once the market mode and the cluster mode are removed.

According to the definition, if a ratio is below one it means that a memory reduction has occurred via the corresponding removal. In order to understand what is the average behaviour of these ratios we take the median of each of them computed on all stocks. So, for example, the average reduction

of memory induced by the market mode on a given set of stocks is $median(\frac{\eta_{i,MM}}{\eta_{i,PL}})$ computed over the index i. As for an error to associate to this measure we used the Median Average Deviation Sachs (2012), defined as for $\frac{\eta_{i,MM}}{\eta_{i,PL}}$

$$MAD\left(\frac{\eta_{i,MM}}{\eta_{i,PL}}\right) \tag{18}$$

$$= median\left(\left|\frac{\eta_{i,MM}}{\eta_{i,PL}} - median\left(\frac{\eta_{i,MM}}{\eta_{i,PL}}\right)\right|\right) \ , \tag{19}$$

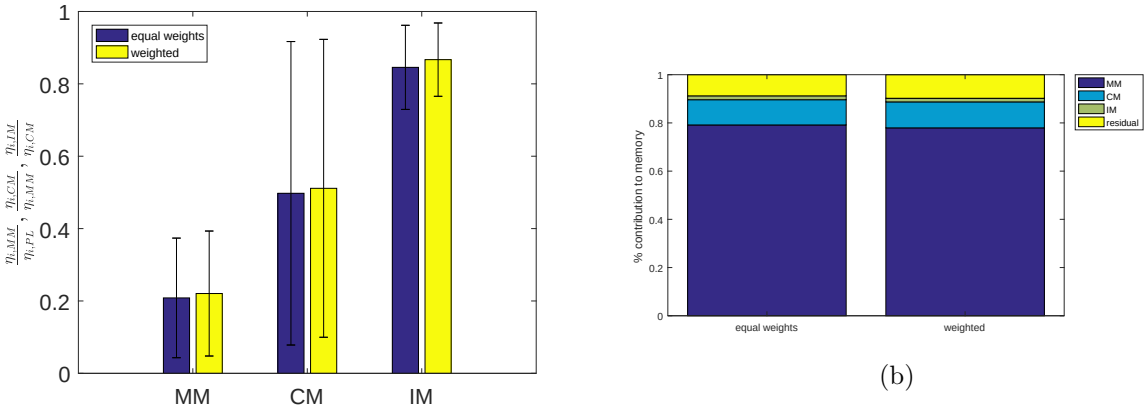
and similarly for $\frac{\eta_{i,CM}}{\eta_{i,MM}}$ and $\frac{\eta_{i,IM}}{\eta_{i,CM}}$. Both the median and the MAD were chosen because of their robustness against outliers. We regard as significant a reduction of memory on the given set of stocks for which the median plus the mad of the ratio are below one.

5.2. *Whole market analysis: finding the main source of memory*

We apply here the procedure described in the previous subsection to our dataset described in Section 2. For completeness, in Fig. 5 we report the result of our analysis for both the unweighted and the weighted schemes. Figure 5a reports the value of the ratios along with the errors (black vertical bars). We observe that in all cases the average value plus the error stays below one, which means that every term gives a meaningful contribution to the overall memory. However we also notice that, in particular for the reduction coming from the cluster mode, there is a large variablity among stocks. Figure 5b reports the same result but showing what is the contribution of each removal with respect to the overall memory. According to our analysis, the majority of the contribution comes from the market mode, which is than the main source of memory for the volatility. We also plot in figure 6 the cumulative of the fraction of stocks with at most the percentage of memory left reported on the x axis, after all contributions are removed. For example from figure 6 we find that 90% of all stocks have only 16.7% of their memory unexplained by all the contributions. We also note here that there is little difference in figure 6 between the weighted and unweighted versions so we shall herein use the unweighted scheme for most of the analysis. This analysis establishes that there is indeed a link between the log volatility and volatility clustering.

5.3. *Cluster-by-cluster analysis: selection criterion for factors*

In this subsection, instead of aggregating the result of the memory reduction over the whole market, we specialize and check what happens to the memory on a cluster-by-cluster basis. For brevity, we only discuss in detail the case of cluster 12 and cluster 22, as defined by the DBHT algorithm discussed in section 3.2, since they are quite informative about the different behaviour one can find at a cluster level. We repeat then the same analysis we performed in the previous subsection but report the behaviour of these two particular clusters. In figure 7 we report the result of our analysis for the unweighted scheme. Figure 7a reports the value of the ratios along with the errors (black vertical bars). Differently for the whole dataset, we see that from figure 7a, the cluster mode removes the vast majority of the memory for cluster 12, without any contribution coming from the market mode or from the interactions. Instead for cluster 22, we see from figure 7a that the market is the major contributor to the memory, whereas the cluster mode is reducing some the remaining memory to some extent and the interactions are again not giving much contribution. Figure 7b reports the same kind of result but relatively to the overall memory. These results suggest that a local analysis reveals a richer behaviour in how the terms in our log volatility factor model affect the memory effect, showing that there is also a link between the correlation structure of the log volatilities and the memory effect. Given these results, we argue that a good criteria for selecting statistically meaningful factors, among all cluster modes, to be included in the definition of our factor model, is



(a) (b)

Figure 5.: Results for the procedure described in section 5.1 across all stocks in the market. Figure 5a is the median of the ratio of the memory proxies for, starting from the left, $\frac{\eta_{i,MM}}{\eta_{i,PL}}$, $\frac{\eta_{i,CM}}{\eta_{i,MM}}$ and $\frac{\eta_{i,IM}}{\eta_{i,CM}}$, computed over the whole market. The blue bars are for the equal weights scheme and the yellow bars are for the weighted scheme. The black vertical bars represent the error among stocks memory reduction applied to the whole market. In figure 5b we plot the contribution to the memory effect of the market (MM), cluster (CM) and interactions (IM) as a percentage with respect to the overall memory. The residual is remaining percentage of memory that is unexplained by the contributors. The values are computed over the whole market. The left column is for the equal weights scheme and the right column is for the weighted scheme.

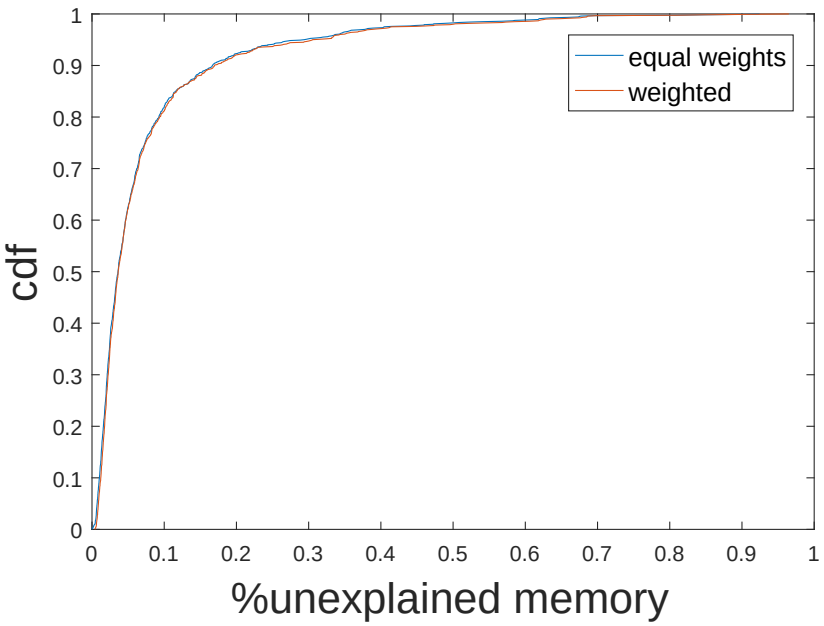
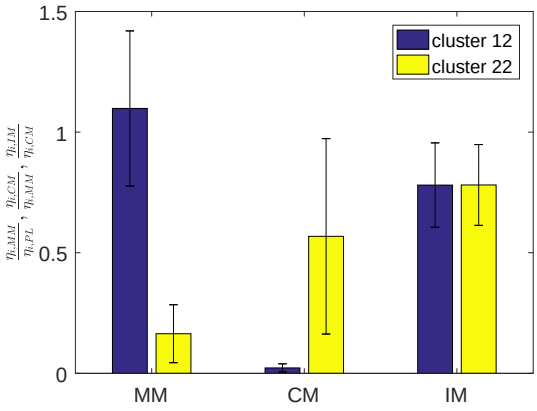
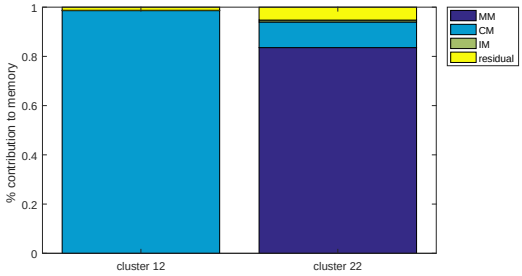


Figure 6.: Cumulative distribution of the fraction of stocks which have % residual memory left after all contributors of the model (market mode, cluster mode and interactions) are removed. The red line is for the weighted modes and the blue the equal weighted modes



(a)



(b)

Figure 7.: The same set of graphs as Fig. 5 except using the equal weights scheme and taking only stocks belonging to cluster 12 and 22. In figure 7a we plot the median ratio of, starting from the left, $\frac{\eta_{i,MM}}{\eta_{i,PL}}$, $\frac{\eta_{i,CM}}{\eta_{i,MM}}$ and $\frac{\eta_{i,IM}}{\eta_{i,CM}}$, computed over the stocks in cluster 12 for the blue bars and over stocks in cluster 22 for the yellow bars. The black vertical bars represent the error among stocks in cluster 12 for the blue bars and among stocks in cluster 22 for the yellow bars. Equal weighted modes are used. In figure 7b we plot the contribution to the memory effect of the market (MM), cluster (CM) and interactions (IM) as a percentage with respect to the overall memory. The residual is remaining percentage of memory that is unexplained by the contributors. The values are computed over all stocks in cluster 12 for the left column and over all stocks in cluster 22 for the right column. Equal weighted modes are used.

to choose those which achieve a significant reduction (in the sense of Section 5.1) to the memory of the stocks within their cluster. Table 2 summarizes the results of this procedure, reporting in the first column the cluster number (as given by the DBHT algorithm). The second column contains the number of stocks in each cluster and in the fourth column we show if the cluster mode reduces the memory of the stocks within that cluster significantly. As we can see we find that out of 29 clusters, 7 do not have a significant meaning to the memory, thus, according, to our criteria are discarded.

6. Economical interpretation of the clusters

Up till now, we have focused on determining the clusters via statistical tools. In this section we show that the clusters also have an economical interpretation. In figure 8, we show the cluster composition of each cluster identified through DBHT using the Industrial Classification Benchmark (ICB)supersector classification of common industries, with each colour representing a different supersector. In particular from figure 8, we observe that clusters are dominated by a particular supersector. For example, we see from figure 8 that clusters 12 and 22 show the presence of dominant supersectors: the real estate sector for cluster 12 and technology sector for cluster 22. In order to check that these identifications of dominant sectors are meaningful, we used the same hypothesis test as in Tumminello *et al.* (2011), Musmeci *et al.* (2015), which tests the null hypothesis that the cluster has merely randomly been assigned supersector classifications using the hypergeometric distribution versus the alternative hypothesis that the supersector is indeed dominating the cluster. Starting from a significance level of 5%, we additionally used a conservative Bonferroni correction for multiple hypothesis testing (Feller 2008) of $0.5N_{cl}N_{ICB}$ to reduce the level of significance, where N_{cl} is the number of clusters identified through DBHT and N_{ICB} is the number of ICB supersectors. This reduces the level of significance to 9.0×10^{-5}, reporting the p values to six decimal places. Table 2 of section 10 details the results of applying this process to all clusters, with the dominant

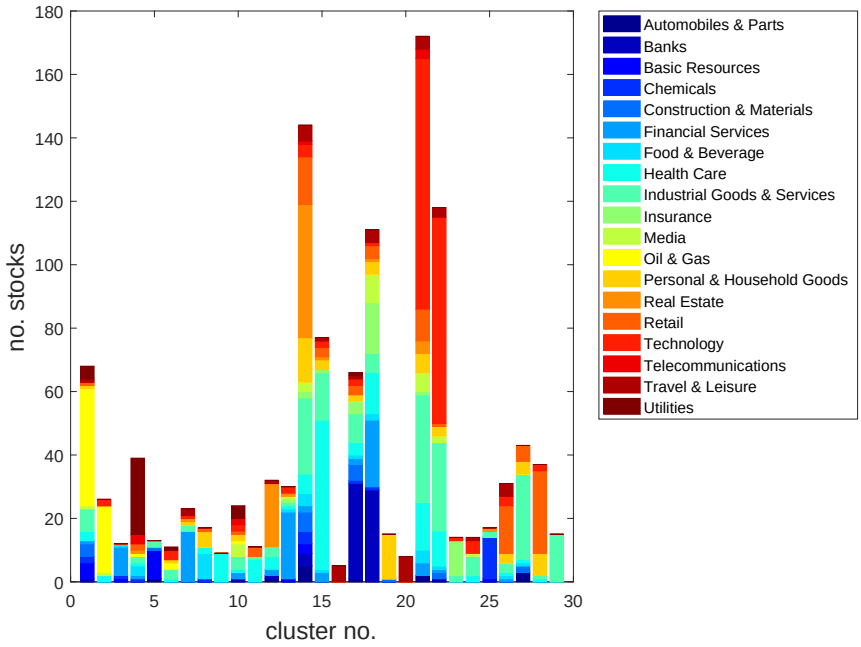


Figure 8.: Composition of DBHT clusters in terms of ICB supersectors. The x axis labels the clusters of DBHT and the y axis is the number of stocks in each cluster. The colours represent particular ICB supersector given in the key.

supersector denoted in the third column. We see from Table 2 that in 26 clusters, the cluster can indeed be matched to their dominating supersector, and of the clusters that significantly contribute to their own memory (see section 5.3), 19 correspond to their dominating supersector. This opens the possibility of choosing cluster modes for a further refinement of the factor model between log volatilties by choosing the cluster modes which reduce the memory statistically significantly after the market mode is removed, but also having an economic interpretation of being dominated by particular supersectors.

7. Comparison with PCA and Exploratory Factor Analysis

In this section we compare the memory reduction performance of our model with a well established PCA inspired factor model (Jolliffe 1982) and exploratory factor analysis driven factor model. Firstly, we explain the importance of the PCA factor model. The PCA analysis gives a set of orthogonal eigenvectors that define mutually linearly uncorrelated portfolios that can be used to help define factor models by assigning each principal component a separate factor. However, as we have pointed out it is difficult to decide how many principal components we should keep. In our analysis, the number of principal components we keep in the PCA factor model shall be fixed to be the same as the number of factors in our factor model i.e. 20. PCA aims to explain the diagonal terms in the orthogonal basis of the correlation matrix $\mathbf{E}$, which is the correlation matrix between the $\ln|r_i(t)|$. Exploratory factor analysis (FA) on the other hand is more general, and aims to explain the off diagonal terms of $\mathbf{E}$, using the general linear model in (2). Again, there are problems selecting exactly how many factors should be included (Preacher *et al.* 2013), but we fix the number of factors in the FA model to be equal to the number of factors in our log volatility factor model i.e. 20. After extracting the factors, we apply a varimax rotation of the factors (Child 2006), which is commonly applied in factor analysis to improve understandability. In figure 9 we plot the cumulative distribution function of how much residual memory is left after removal of the factors for the log volatility factor

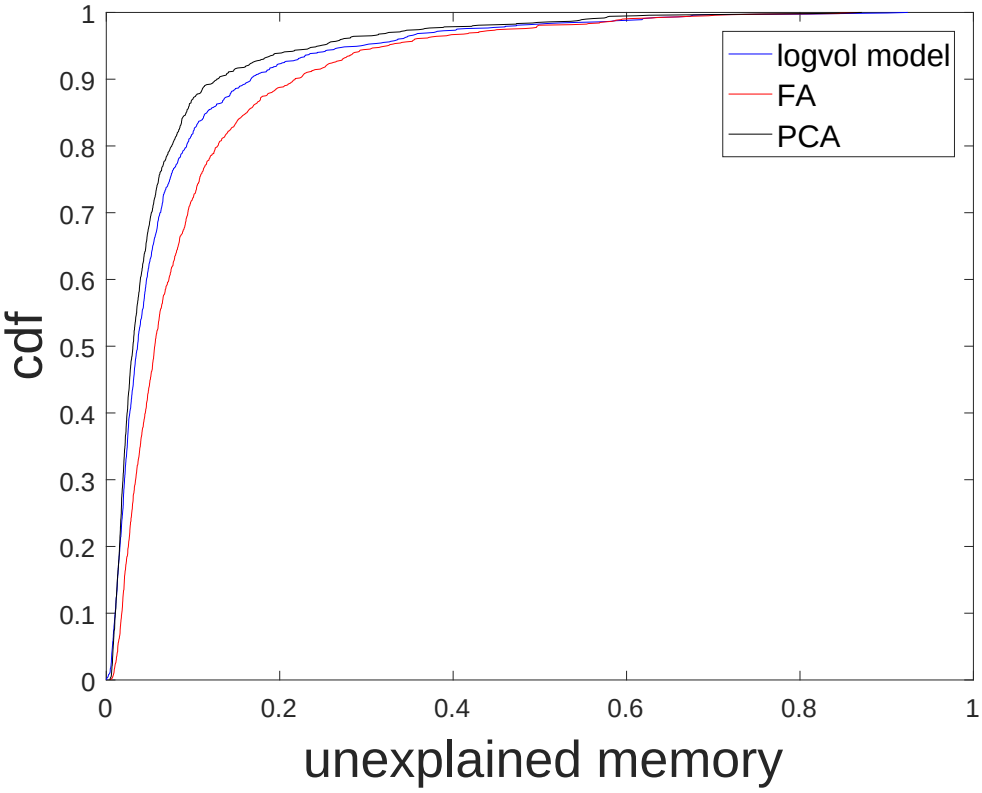


Figure 9.: Empirical cumulative distribution function of the unexplained residual memory for the factor model in blue line, the PCA in black, where we only take the first 23 principal components, and the exploratory factor analysis, where we use 23 factors and a varimax rotation.

model, FA model and PCA factor model as a percentage of the total memory before removal.

We see from figure 9 that 90% of all stocks only have a maximum of 16.7% residual memory left for the factor model of log volatility, whereas 90% of all stocks have a maximum of 12.7% of residual memory left, which means that the PCA factor model and the log volatility factor model both explain the memory to the same efficiency. For the exploratory factor model, we see that 90% of all stocks have 21.8% of their memory left, which is worse than the log volatility factor model and the PCA factor model, but still has a comparable performance. We can therefore conclude that the log volatility factor model explains the same amount of memory as the other two models, even after fixing the amount of factors to be the same in the PCA and exploratory factor model.

8. Conclusion

We proposed a new factor model for the log-volatility discussing how each term of the model affects the stylized fact of the volatility clustering. This reduces the information present in the linear correlation between the log volatilities to a global factor, which is the so-called market mode, and second and third local factors, which are the cluster mode and the interactions. Using a new non parametric, integrated proxy for the volatility clustering, we found that there is indeed a link between the volatility and volatility clustering. First, the dataset was examined globally, which revealed the market to account for the majority of the volatility clustering effect present in our dataset. However, a local cluster by cluster analysis instead reveals significant variability: in some clusters, the cluster mode itself may be contributing to the volatility clustering. This enabled us to select only statistically relevant cluster factors, reducing the information in the correlation between the log volatility

and the number of factors further. From these reduced set of factors, we can select factors that have an economic interpretation through the identification of their dominant ICB supersector, which decreased the number of relevant factors some more. This is significantly advantageous over other potential factor models that could be used for log volatility such as PCA and exploratory factor analysis since we do not subjectively select the number of factors, and also because the factors have a clearer economic interpretation through the identification of their dominant ICB supersector. A comparison of the log volatility factor model with PCA and an exploratory factor model reveals that they explain the same amount of memory in the dataset.

This work is particularly relevant for the field of volatility modelling, since most multivariate models such as multivariate extensions of GARCH, stochastic covariance and realised covariance models suffer from the curse of dimensionality and increase in the number of parameters. The log volatility factor model presented here could be used to help reduce the amount of parameters needed for these models through the identification of a reduced set of factors given by the procedure in this paper.

References

Alexander, C., Principal component models for generating large GARCH covariance matrices. *Economic Notes*, 2002, **31**, 337–359.

Andersen, T.G., Bollerslev, T., Diebold, F.X. and Labys, P., Modeling and forecasting realized volatility. *Econometrica*, 2003, **71**, 579–625.

Barberis, N., Greenwood, R., Jin, L. and Shleifer, A., X-CAPM: An extrapolative capital asset pricing model. *Journal of Financial Economics*, 2015, **115**, 1–24.

Bauwens, L., Laurent, S. and Rombouts, J.V., Multivariate GARCH models: a survey. *Journal of applied econometrics*, 2006, **21**, 79–109.

Borghesi, C., Marsili, M. and Miccichè, S., Emergence of time-horizon invariant correlation structure in financial returns by subtraction of the market mode. *Physical Review E*, 2007, **76**, 026104.

Bouchaud, J.P. and Potters, M., *Theory of financial risk and derivative pricing: from statistical physics to risk management*, 2009, Cambridge university press.

Box, G.E., Jenkins, G.M., Reinsel, G.C. and Ljung, G.M., *Time series analysis: forecasting and control*, 2015, John Wiley & Sons.

Bun, J., Bouchaud, J.P. and Potters, M., Cleaning large correlation matrices: tools from random matrix theory. *Physics Reports*, 2017, **666**, 1–109.

Chakraborti, A., Toke, I.M., Patriarca, M. and Abergel, F., Econophysics review: II. Agent-based models. *Quantitative Finance*, 2011, **11**, 1013–1041.

Chen, N.F., Roll, R. and Ross, S.A., Economic forces and the stock market. *Journal of business*, 1986, pp. 383–403.

Chicheportiche, R. and Bouchaud, J.P., A nested factor model for non-linear dependencies in stock returns. *Quantitative Finance*, 2015, **15**, 1789–1804.

Child, D., *The essentials of factor analysis*, 2006, A&C Black.

Clark, P.K., A subordinated stochastic process model with finite variance for speculative prices. *Econometrica: journal of the Econometric Society*, 1973, pp. 135–155.

Connor, G., Hagmann, M. and Linton, O., Efficient semiparametric estimation of the Fama–French model and extensions. *Econometrica*, 2012, **80**, 713–754.

Cont, R., Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 2001.

Darbyshire, J., *The volatility surface: a practitioner's guide*, Vol. 357, , 2017, Aitch & Dee Limited.

Engel, C., Mark, N.C. and West, K.D., Factor model forecasts of exchange rates. *Econometric Reviews*, 2015, **34**, 32–55.

Faff, R., A simple test of the Fama and French model using daily data: Australian evidence. *Applied Financial Economics*, 2004, **14**, 83–92.

Faff, R., Gharghori, P. and Nguyen, A., Non-nested tests of a GDP-augmented Fama–French model versus a conditional Fama–French model in the Australian stock market. *International Review of Economics & Finance*, 2014, **29**, 627–638.

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

Fama, E.F. and French, K.R., Common risk factors in the returns on stocks and bonds. *Journal of financial economics*, 1993, **33**, 3–56.

Fama, E.F. and French, K.R., Multifactor explanations of asset pricing anomalies. *The journal of finance*, 1996, **51**, 55–84.

Fama, E.F. and French, K.R., A five-factor asset pricing model. *Journal of Financial Economics*, 2015, **116**, 1–22.

Feller, W., *An introduction to probability theory and its applications*, Vol. 2, , 2008, John Wiley & Sons.

Gatheral, J., *The volatility surface: a practitioner's guide*, Vol. 357, , 2011, John Wiley & Sons.

Grauer, R.R. and Janmaat, J.A., Cross-sectional tests of the CAPM and Fama–French three-factor model. *Journal of banking & Finance*, 2010, **34**, 457–470.

Hull, J. and White, A., The pricing of options on assets with stochastic volatilities. *The journal of finance*, 1987, **42**, 281–300.

Hull, J.C., *Options, futures, and other derivatives*, 2006, Pearson Education India.

Jackson, D.A., Stopping rules in principal components analysis: a comparison of heuristical and statistical approaches. *Ecology*, 1993, **74**, 2204–2214.

Jolliffe, I.T., A note on the use of principal components in regression. *Applied Statistics*, 1982, pp. 300–303.

Jolliffe, I.T., Principal Component Analysis and Factor Analysis. In *Principal component analysis*, pp. 115–128, 1986, Springer.

Laloux, L., Cizeau, P., Bouchaud, J.P. and Potters, M., Noise dressing of financial correlation matrices. *Physical review letters*, 1999, **83**, 1467.

Livan, G., Alfarano, S. and Scalas, E., Fine structure of spectral properties for random correlation matrices: An application to financial markets. *Physical Review E*, 2011, **84**, 016113.

Lockhart, R., Taylor, J., Tibshirani, R.J. and Tibshirani, R., A significance test for the lasso. *Annals of statistics*, 2014, **42**, 413.

Majumdar, S.N. and Vivo, P., Number of relevant directions in principal component analysis and Wishart random matrices. *Physical review letters*, 2012, **108**, 200601.

Malevergne, Y. and Sornette, D., Collective origin of the coexistence of apparent random matrix theory noise and of factors in large sample correlation matrices. *Physica A: Statistical Mechanics and its Applications*, 2004, **331**, 660–668.

Mandelbrot, B.B., The variation of certain speculative prices. In *Fractals and Scaling in Finance*, pp. 371–418, 1997, Springer.

Markowitz, H., Portfolio selection. *The journal of finance*, 1952, **7**, 77–91.

Merton, R.C., An intertemporal capital asset pricing model. *Econometrica: Journal of the Econometric Society*, 1973, pp. 867–887.

Micciche, S., Empirical relationship between stocks cross-correlation and stocks volatility clustering. *Journal of Statistical Mechanics: Theory and Experiment*, 2013, **2013**, P05015.

Musmeci, N., Aste, T. and Di Matteo, T., Relation between financial market structure and the real economy: comparison between clustering methods. *PloS one*, 2015, **10**, e0116201.

N. Musmeci, T. Aste, T.D.M., Interplay between past market correlation structure changes and future volatility outbursts. *Scientific Reports 6*, 2016, **6**, 36320.

Plerou, V., Gopikrishnan, P., Rosenow, B., Amaral, L.A.N., Guhr, T. and Stanley, H.E., Random matrix approach to cross correlations in financial data. *Physical Review E*, 2002, **65**, 066126.

Preacher, K.J., Zhang, G., Kim, C. and Mels, G., Choosing the optimal number of factors in exploratory factor analysis: A model selection perspective. *Multivariate Behavioral Research*, 2013, **48**, 28–56.

Racicot, F.E. and Rentz, W.F., Testing Fama–FrenchâĂŹs new five-factor asset pricing model: evidence from robust instruments. *Applied Economics Letters*, 2016, **23**, 444–448.

Reinganum, M.R., The arbitrage pricing theory: some empirical results. *The Journal of Finance*, 1981, **36**, 313–321.

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

Sachs, L., *Applied statistics: a handbook of techniques*, 2012, Springer Science & Business Media.

Sharpe, W.F., Capital asset prices: A theory of market equilibrium under conditions of risk. *The journal of finance*, 1964, **19**, 425–442.

Singh, A. and Xu, D., Random matrix application to correlations amongst the volatility of assets. *Quanti-*

tative Finance, 2016, **16**, 69–83.
Song, W.M., Di Matteo, T. and Aste, T., Hierarchical information clustering by means of topologically embedded graphs. *PLoS One*, 2012, **7**, e31929.
Taylor, S.J., Modeling stochastic volatility: A review and comparative study. *Mathematical finance*, 1994, **4**, 183–204.
Theil, H., A rank-invariant method of linear and polynomial regression analysis. In *Henri TheilâĂŹs contributions to economics and econometrics*, pp. 345–381, 1992, Springer.
Thompson, B., *Exploratory and confirmatory factor analysis: Understanding concepts and applications.*, 2004, American Psychological Association.
Tumminello, M., Lillo, F. and Mantegna, R.N., Hierarchically nested factor model from multivariate data. *EPL (Europhysics Letters)*, 2007, **78**, 30006.
Tumminello, M., Micciche, S., Lillo, F., Piilo, J. and Mantegna, R.N., Statistically validated networks in bipartite complex systems. *PloS one*, 2011, **6**, e17994.
Van Der Maaten, L., Postma, E. and Van den Herik, J., Dimensionality reduction: a comparative. *J Mach Learn Res*, 2009, **10**, 66–71.
Zabarankin, M., Pavlikov, K. and Uryasev, S., Capital asset pricing model (CAPM) with drawdown measure. *European Journal of Operational Research*, 2014, **234**, 508–517.
Zhang, K. and Chan, L., Efficient factor garch models and factor-dcc models. *Quantitative Finance*, 2009, **9**, 71–91.
Zou, H. and Hastie, T., Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2005, **67**, 301–320.

9. Appendix

9.A. *Data cleaning process*

Our dataset cannot be used as it is since the price time-series are not aligned, which is due to the fact the some stocks have not been traded on certain days. In order to overcome this issue, we apply a data cleaning procedure which allows us to keep as many stock as possible. For example, we do not want to remove a stock just because it was not traded on few days in the given time-span. The main idea is to fill the gaps dragging the last available price and assuming that a gap in the price time-series corresponds to a zero log-return. At the same time we do not want to drag too many prices because a time-series filled with zeros would not be statistically significant. In light of this we remove from our dataset the time-series which are too short in a certain sense. The detailed procedure goes as follows:

(i) Remove from the dataset the price time-series with length less than p times the longest one;
(ii) Find the common earliest day among the remaining time-series;
(iii) Create a reference time-series of dates when at least one of the stocks has been traded starting from the earliest common date found in the previous step;
(iv) Compare the reference time-series of dates with the time-series of dates of each stock and fill the gaps dragging the last available price.

In this paper we chose $p = 0.90$ thus keeping as much as possible unmodified time-series. However, the results do not change if we pick a higher value of p.

9.B. *Weighting schemes*

Here we shall define the two types of weighting schemes used in this paper for the ξ_i and ξ_{ik} defined in (8) and (10) respectively. The first weighting scheme is based on the eigenspectrum of $\mathbf{E}$ and $\mathbf{G}$. It is useful now to explain the financial interpretation of the eigenvectors $\mathbf{v}$ with entries v_i and eigenvalue λ for $\mathbf{E}$. v_i can be seen as weights for a portfolio defined by $\mathbf{v}$. Measuring the risk from

the volatility of the portfolio via its variance, we see it is given by:

$$\frac{1}{T}\sum_t \left(\sum_i v_i \ln|r_i(t)|\right)^2 = \sum_{ij} v_i v_j E_{ij} = \lambda \tag{20}$$

Hence λ represents the risk from the volatility of the portfolio given by $\mathbf{v}$. We set $\xi_i = v_i$, where now v_i is the ith entry of the eigenvector corresponding to the largest eigenvalue of the empirical correlation matrix $\mathbf{E}$. This is called the market eigenvalue as it represents all stocks moving together Plerou *et al.* (2002), and is also portfolio of stocks that gives the risk of the market volatility mode through its corresponding eigenvalue. We could have also used a real index to determine the weights e.g. the Dow Jones, but Borghesi *et al.* (2007) showed that this does not effectively remove the influence of modes from returns compared to a pseudo-index.

The weights ξ_{ik} are established in a similar way to the market mode case, which we shall do by considering only the part of $\mathbf{G}$ which corresponds to members of the cluster. Defining a submatrix of $\mathbf{G}$

$$\mathbf{G}^{(k)} = \{\mathbf{G}\}_{(i,j)\in clusterk} \tag{21}$$

Where $\{...\}_{(i,j)\in cluster\ k}$ refers to only keeping the elements the matrix in which i and j are stocks in cluster k. Thus $\mathbf{G}^{(k)}$ is the square sub matrix of $\mathbf{G}$ corresponding to cluster k. This submatrix is the correlation matrix of a market which consists only of stocks which are part of cluster k. Hence, in exactly the same way as the market eigenvalue, the largest eigenvalue of $\mathbf{G}^{(k)}$ represents stocks of the cluster moving together, the value of the eigenvalue being the risk of the cluster market portfolio, and the related eigenvector giving the weights of such a portfolio. Therefore, the definition of the weights ξ_{ik} for cluster k are determined by setting $\omega_{ik} = v_i^{(k)}$, which is the ith entry of the eigenvector corresponding to the largest eigenvalue of $\mathbf{G}^{(k)}$. This is the weighting scheme used and is compared to the case of equal weights where $\xi_i = \frac{1}{N}$ and $\xi_{ik} = \frac{1}{m_k}$ in figures 5a, 5b and 6 thereafter the equal weights scheme is used.

9.C. *Elastic Net Regression*

Elastic net regression is used to find the values of β_{ik} and $\beta_{ik'}$ using Eq. (6). Further details of the use of this method is provided in this appendix. Elastic net regression Zou and Hastie (2005) is a hybrid version of ridge regularisation and lasso regression, thus providing a way of dealing with correlated explanatory variables (in our case $I_k(t)$ and $I_{k'}(t)$) and also performing feature selection, which takes into account non-interacting clusters $I_{k'}(t)$ that ridge regularisation would ignore. Elastic net regression solves the constrained minimisation problem

$$\min_{\boldsymbol{\beta}_i} \frac{1}{T}\sum_{t=1}^{T}\left(c_i(t) - \mathbf{I}(t)^\dagger \boldsymbol{\beta}_i\right)^2 + \lambda P_a(\boldsymbol{\beta}_i) \tag{22}$$

, where $\boldsymbol{\beta}_i$ is the vector of loadings given by $(\beta_{i1}, \beta_{i2}, \ldots, \beta_{iK})^\dagger$, $\mathbf{I}(t)$ is the matrix consisting of columns $(I_1(t), I_2(t), \ldots, I_{N_{cl}}(t))$ and λ and a are hyperparameters. $P_a(\boldsymbol{\beta}_i)$ is defined as

$$P_a(\boldsymbol{\beta}_i) = \sum_{j=1}^{M}\left((1-a)\frac{\beta_{ij}^2}{2} + a|\beta_{ij}|\right) \tag{23}$$

. The first term in the sum of Eq. (23) is the L_2 penalty for the ridge regularisation and the second term in the sum is the L_1 penalty for the lasso regression. Hence if $a = 0$ then elastic net reduces to ridge regression and if $a = 1$ then elastic net becomes lasso, with a value between the two controlling

the extent which one is preferred to the other. The determination of the a hyperparameter, controlling the extent of lasso vs ridge, and λ, for the ridge, is done using 10 cross validated fits Zou and Hastie (2005), picking the pair of (a, λ) that give the minimum prediction error. We show the values of β_{ik} and test the significance of the predictor $I_k(t)$ at the 5% level in Table 3 of section 10, where the p value is shown in brackets, using the significance test outlined in Lockhart *et al.* (2014).

10. Tables

	β_{i0}	α_{i0}
KO	0.0310 (0)	0.0015 (0.4764)
RIG	0.0248 (0)	0.1972 (0)

(a) weighted modes

	β_{i0}	α_{i0}
KO	1.1564 (0)	-0.0690 (0.0017)
RIG	0.9041 (0)	0.1426 (0)

(b) equal weights

Table 1.: This table shows the responsiveness to the market mode $I_0(t)$, β_{i0} and the corresponding excess volatility α_{i0} for stocks KO and RIG, calibrated as detailed in section 3.1. The p values shown in brackets are for the null hypothesis that both β_{i0} and α_{i0} are 0. Table 1a is for the weighted scheme and Table 1b for equal weights, which are detailed in 9.B.

cluster no.	no. stocks	dominant supersector	cluster sig
1	68	OG (0)	T
2	26	OG (0)	T
3	12	FS (0)	T
4	39	U (0)	T
5	13	BR (0)	T
6	11	IGS (0.089957)	T
7	23	FS (0)	T
8	17	FB (0)	F
9	9	HC (0)	T
10	24	IGS (0.355912)	T
11	11	HC (0)	F
12	32	RE (0)	T
13	30	FS (0)	T
14	144	RE (0)	T
15	77	HC (0)	T
16	5	TL (0)	T
17	66	B (0)	T
18	111	B (0)	T
19	15	PHG (0)	T
20	8	TL (0)	F
21	172	T (0)	T
22	118	T (0)	T
23	14	I (0)	F
24	12	IGS (0.003514)	T
25	17	C (0)	T
26	31	R (0)	T
27	43	IGS (0)	F
28	37	R (0)	F
29	15	IGS (0)	F

Table 2.: Table showing the cluster no. in the first column and the number of stocks in the second column. In the third column, we have the dominant ICB supersector (abbreviated to the first letters in each supersector, which are listed in figure 8). In brackets in the third column we have the p value of the hypothesis test which tests whether the most dominant supersector can be meaningfully identified from the cluster, which are given to 6 decimal places. The fourth column details whether the cluster mode significantly reduces the memory for that cluster.

	β_{ik}	
KO	0.9431(0)	0.8997(0)
RIG	0.9041 (0)	1.1265(0)

Table 3.: This table shows the responsiveness to the cluster mode $I_k(t)$, β_{ik} calibrated as detailed in section 3.3. P values shown in brackets test the significance of the predictor given by the cluster mode $I_k(t)$. The first column is for the weighted scheme and second is for equal weights, which are detailed in 9.B.

11. List of Figure Captions

(i) Figure 1: Empirical ACF of the absolute returns (blue solid lines) for Coca Cola Co. (KO) in figure 1a and Transocean (RIG) in figure 1b in log-log scale. The linear best fit is also shown in red dashed lines.
   a) Coca Cola Enterprises Inc. $\beta^{\text{vol}} = 0.4544$
   b) Transoceanic $\beta^{\text{vol}} = 0.3975$

(ii) Figure 2: Negative dependence between ρ_i^{vol} and β_i^{vol}. The negative relationship was tested using 1 sided Spearman's rank correlation at the 5% level with the null hypothesis of there being no correlation and was rejected, which confirms the result of Micciche (2013) on our data.

(iii) Figure 3: Integrated proxy η as a function of the lag L' where η is integrated over [1,L'] until $L' = L_{cut}$. Fig. 3a is for Coca Cola Co. and fig. 3b for Transocean.
   a) Coca Cola Enterprises Inc.
   b) Transoceanic

(iv) Figure 4: In figure 4a we plot the β^{vol} power law exponent proxy for the strength of the memory effect vs c the integrated proxy. In figure 4b we plot the relationship between ρ_i^{vol} and η defined in the text. The decreasing relationship in figure 4a and the increasing relationship in figure 4b was tested using the Spearman's rank correlation at the 5% level and was rejected in both cases.
   a) β^{vol} vs β^{vol}
   b) ρ_i^{vol} vs η

(v) Figure 5: Results for the procedure described in section 5.1 across all stocks in the market. Figure 5a is the median of the ratio of the memory proxies for, starting from the left, $\frac{\eta_{i,MM}}{\eta_{i,PL}}$, $\frac{\eta_{i,CM}}{\eta_{i,MM}}$ and $\frac{\eta_{i,IM}}{\eta_{i,CM}}$, computed over the whole market. The blue bars are for the equal weights scheme and the yellow bars are for the weighted scheme. The black vertical bars represent the error among stocks memory reduction applied to the whole market. In figure 5b we plot the contribution to the memory effect of the market (MM), cluster (CM) and interactions (IM) as a percentage with respect to the overall memory. The residual is remaining percentage of memory that is unexplained by the contributors. The values are computed over the whole market. The left column is for the equal weights scheme and the right column is for the weighted scheme.

(vi) Figure 6: Cumulative distribution of the fraction of stocks which have % residual memory left after all contributors of the model (market mode, cluster mode and interactions) are removed. The red line is for the weighted modes and the blue the equal weighted modes.

(vii) Figure 7: The same set of graphs as Fig. 5 except using the equal weights scheme and taking only stocks belonging to cluster 12 and 22. In figure 7a we plot the median ratio of, starting from the left, $\frac{\eta_{i,MM}}{\eta_{i,PL}}$, $\frac{\eta_{i,CM}}{\eta_{i,MM}}$ and $\frac{\eta_{i,IM}}{\eta_{i,CM}}$, computed over the stocks in cluster 12 for the blue bars and over stocks in cluster 22 for the yellow bars. The black vertical bars represent the error among stocks in cluster 12 for the blue bars and among stocks in cluster 22 for the yellow bars. Equal weighted modes are used. In figure 7b we plot the contribution to the memory effect of the market (MM), cluster (CM) and interactions (IM) as a percentage with respect to the overall memory. The residual is remaining percentage of memory that is unexplained by the contributors. The values are computed over all stocks in cluster 12 for the left column and over all stocks in cluster 22 for the right column. Equal weighted modes are used.

(viii) Figure 8: Composition of DBHT clusters in terms of ICB supersectors. The x axis labels the clusters of DBHT and the y axis is the number of stocks in each cluster. The colours represent particular ICB supersector given in the key.

(ix) Figure 9: Empirical cumulative distribution function of the unexplained residual memory for the factor model in blue line, the PCA in black, where we only take the first 23 principal components, and the exploratory factor analysis, where we use 23 factors and a varimax

rotation.