Forecasting market states

PIER FRANCESCO PROCACCI [1] and TOMASO ASTE [1,2]

[1]Department of Computer Science, UCL, Gower Street, London, WC1E6BT, UK
[2]Systemic Risk Centre, London School of Economics and Political Sciences, London, WC2A
2AE, UK

We propose a novel methodology to define, analyse and forecast market states. In our approach market
states are identified by a reference sparse precision matrix and a vector of expectation values. In our
procedure each multivariate observation is associated to a given market state accordingly to a minimisa-
tion of a penalized Mahalanobis distance. The procedure is made computationally very efficient and can
be used with a large number of assets. We demonstrate that this procedure is successfull at clustering
different states of the markets in an unsupervised manner. In particular, we describe an experiment with
one hundred log-returns and two states in which the methodology automatically associates states preva-
lently to per-and post crisis periods with one state gathering periods with average positive returns and
the other state periods with average negative returns, therefore discovering spontaneously the common
classification of 'bull' and 'bear' markets. In another experiment, with again one hundred log-returns and
two states, we demonstrate that this procedure can be efficiently used to forecast off-sample future market
states with significant prediction accuracy. This methodology opens the way to a range of applications
in risk management and trading strategies in the context where the correlation structure plays a central
role.

Keywords: Financial market states; Temporal clustering; Information Filtering Networks; TMFG;
LoGo; Sparse inverse covariance; Correlation Structure.

JEL Classification: C38, G61, G15, G17

1. Introduction

Markets do not always behave in the same way. In common terminology, there are periods of 'bull'
market in which prices are more likely to rise and periods of 'bear' market in which prices are
more likely to fall. These different 'states' of markets are commonly attributed in literature to un-
observable, or latent, regimes representing a set of macroeconomic, market and sentiment variables.

Many time series models presented in literature tried to capture this phenomenon. Among the
most popular methods, it is worth mentioning the TAR models (Tong 1978), trying to estimate
'structural breaks' in the time series process, and the Markov Switching models (Hamilton 1989),
where the change in regimes are parametrized by means of an unobserved state variable typi-
cally modelled as Markov chain. However, the application of TAR models in finance is frequently
criticized since it cannot be established with certainty when a structural break has occurred in
economic time series and the prior knowledge of major economic events could lead to bias in in-
ference (Campbell *et al.* 1997). Markov switching models, on the other hand, are highly affected
by the curse of dimensionality. In particular, for slightly more complex dynamics than the original
proposal (Hamilton 1989), we need to rely on variational inference techniques or MCMC methods
(Tsay 2005, Kim and Nelson 1999). This implies that, in a multivariate context and particularly if

we aim to extract information on the switching from the correlation structure, estimation becomes difficult to perform.

Other approaches focus on clustering of observations into groups: 'similar' data objects are discovered on the basis of some criteria for comparisons. Most works related to clustering of time series are classified into two categories: subsequence time series clustering and point clustering. Subsequence clustering involves the clustering of sliding windows of data points and usually aim at discover repeated patterns. Example are Dynamic Time Warping (Liao 2005), Hierarchical methods (Nevill-Manning and Witten 1997) or pattern discovery (Ren *et al.* 2017). In point clustering methods, instead, each multivariate observation at each time instance t is assigned to a cluster. In most popular approaches, however, this is done based on a distance metric (Grabarnik and Särkkä 2001, Focardi and Fabozzi 2004, Zolhavarieh *et al.* 2014, Hendricks *et al.* 2016, Hallac *et al.* 2016).

In a multivariate context, different 'states' of markets are not only reflected in the gains and losses, but also in the relative dynamics of prices. Indeed, the correlation structure changes between bull and bear periods indicating that there are structural differences in these market states. Most common approaches in the industry assume -for convenience- a stationary correlation structure (Duffie and Pan 1997, Black and Litterman 1992). However, it is well established that correlations among stocks are not constant over time (Lin *et al.* 1994, Ang and Bekaert 2002, Musmeci *et al.* 2016) and increase substantially in periods of high market volatility, with, asymmetrically, larger increases for downward moves (see, for example, (Ang and Chen 2002, Cizeau *et al.* 2010, Schmitt *et al.* 2013)). Indeed, various approaches have been proposed in literature to model and predict time-varying correlations. Examples are, for instance, the generalized autoregressive conditional heteroskedasticity (GARCH) models by (Bollerslev 1990) or the Dynamic Conditional Correlation (DCC) model by (Engle 2002). However, most of these models are not able to cope with more than a few assets due to the curse of dimensionality having numbers of parameters that increases super-linearly with the number of variables (Danielsson 2011). Other approaches have been focusing on the study of changes in a time-varying correlation matrix computed from a rolling window. This is, for instance, the case of estimators like the RiskMetrics (Longerstaey and Spencer 1996) or (Lee and Stevenson 2003). However, since these approaches use only a small part of the data, these estimators have large variances and, in case of high dimensionality, may lead to inconclusive estimates (Laloux *et al.* 1999).

Hallac *et al.* (2017) introduced a clustering algorithm called TICC (Toeplitz Inverse Covariance Clustering), originally proposed for electric vehicles, where classification into states is constructed from a likelihood measure associate with a referential sparse precision matrix (inverse covariance matrix). Instead of considering each observation in isolation, however, in their approach they cluster short subsequences of observations so that the covariance matrix constructed on the subsequences provides a representation of the cross-time partial correlations. In this setting, then, by imposing a Toeplitz constraint to the precision matrix of each regime, the cross-time partial correlations are constrained to be constant and, hence, covariance-stationarity is enforced. This method has a number of appealing features from a financial perspective, although the structure of data considered by the authors is significantly different from noisy data in finance.

In this paper we build on (Hallac *et al.* 2017) and propose a similar Covariance based Clustering. However, we consider single observations and do not enforce Toeplitz structure on the precision matrix. We, therefore, call this methodology ICC - Inverse Covariance Clustering. Analogously to (Hallac *et al.* 2017), we also enforce temporal coherence by penalizing frequent switches between market states and favouring temporal consistency. Another difference is that we do not directly maximise likelihood but rather we assign states to clusters accordingly to their Mahalanobis distance (De Maesschalck *et al.* 2000). We experiment with this methodology in the context of financial time series and provide a detailed analysis of the role played by sparsity and temporal consistency, while assessing the significance of the clusters. Finally, we show that the

cluster classification can be used for one step ahead off-sample prediction.

Our approach simplify and clarify the definition of 'market state' by identifying each state with a sparse precision matrix and a vector of expectation values which are associated to a set of multivariate observations clustered together accordingly with a given procedure. In the following, the precision matrix of market state 'k' is denoted with $\boldsymbol{J}_k$ and it represents the structure of partial correlations between the system's variables. In the multivariate normal case, two nodes are conditionally independent if and only if the corresponding element of $\boldsymbol{J}_k$ is equal to zero. A sparse precision matrix provides an easily interpretable and intuitive structure of the market state with all the most relevant dependencies directly interconnected in a sparse network. Furthermore, sparsity reduces the number of parameters from order n^2 (with n the number of variables) to order n preventing overfitting (Lauritzen 1996) and filtering out noisy correlations (Barfuss *et al.* 2016, Musmeci *et al.* 2017).

The **segmentation procedure** uses a redesigned version of the Expectation Maximization (EM) algorithm (Dempster *et al.* 1977, McLachlan and Krishnan 1997). It starts by setting the number of clusters K (in the present paper we limit to $K = 2$) and assigns multivariate observations to clusters randomly. From these K sets of data we compute the sample means $\boldsymbol{\mu}_k$ and the precision matrices $\boldsymbol{J}_k$ and we then iteratively re-assign points to the cluster with smallest

$$\mathcal{M}_{t,k} = d_{t,k}^2 + \gamma \mathbb{1}\{\mathcal{K}_{t-1} \neq k\} \ . \tag{1}$$

where $\boldsymbol{X}_t = [x_{t,1}, x_{t,2}, ..., x_{t,n}]$ is the n-stocks multivariate observation at time t $(= 1, ..., T)$; $\boldsymbol{\mu}_k$ is the vector of the means for cluster k; $\boldsymbol{J}_k$ is the (sparse) precision matrix for cluster k; $d_{t,k}^2 = (\boldsymbol{X}_t - \boldsymbol{\mu}_k)^T \boldsymbol{J}_k \ (\boldsymbol{X}_t - \boldsymbol{\mu}_k)$ is the the square Mahalanobis distance of observation $\boldsymbol{X}_t$ in cluster k with respect to the cluster centroid $\boldsymbol{\mu}_k$; γ is a parameter penalizing state switching; $\mathcal{K}_{t-1}$ is the cluster assignment of the observation at time $t - 1$. We considered as well clustering with respect to maximum likelihood and minimum Euclidean distance, however we report only about the procedure with Mahalanobis distance which is the one that provides best results. Specifically, Euclidean distance is very efficient in distinguishing positive and negative returns but does not distinguish well between pre- and post-crysisis periods. The maximum likelihood instead identify very well the crisis period but then it is much less clean in classifying the 'bull' and 'bear' market states. Let us note that the used Mahalanobis distance clustering is producing high likelihood although not maximal.

The clustering assignment procedure is made computationally efficient by using the Viterbi algorithm (Viterbi 1967, Bishop 2006) that transforms an otherwise $O(K^T)$ procedure into $O(KT)$ (Appendix A). Further, the sparse precision matrix $\boldsymbol{J}_k$ is computed efficiently from the observations in each cluster by means of the TMFG-LoGo network filtering approach (Massara *et al.* 2015, Barfuss *et al.* 2016). TMFG-LoGo approach has proven to be more efficient and better performing, particularly when few data are available (Barfuss *et al.* 2016, Aste and Di Matteo 2017), with respect to other techniques such as GLASSO (Friedman *et al.* 2008). Implementation has been performed with and in-house built-for-purpose python package. This is the first time this methodology is introduced and applied to financial data and market states analytics.
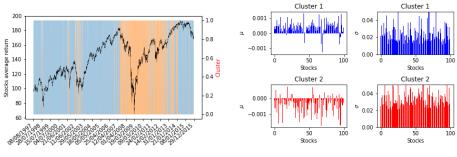
In this paper we report results for two experiments performed over a **dataset** of daily closing prices of $n = 2490$ US stocks entering among the constituents of the Russel 1000 index (*RIY index*) traded between 02/01/1995 and 31/12/2015. For each asset $i = 1, ..., n$, we calculated the corresponding daily log-returns $r_i(t) = \log(P_i(t)) - \log(P_i(t-1))$, where $P_i(t)$ is the closing price of stock i at time t.

2. Clustering

As mentioned in the introduction, our primary goal is to efficiently cluster noisy, multivariate time series into meaningful regimes, while controlling for temporal consistency. In this first experiment, we considered the entire dataset between 02/01/1995 and 31/12/2015 and estimated two referential market states. In order to explore the role of each building block of our algorithm and to compare it to a traditional baseline method, we investigate five models:

- a) ICC Model - Sparse precision matrix and temporal consistency
- b) ICC Model - Full precision matrix and temporal consistency
- c) ICC Model - Sparse precision matrix
- d) ICC Model - Full precision matrix
- e) Gaussian Mixture Model - Full Covariance

Model (a) is the present proposed ICC methodology. Model (b) considers full precision matrices J_k instead of sparse ones. Model (c) relaxes temporal consistency allowing for $\gamma = 0$ in Eq. 1. Model (d) has $\gamma = 0$ full precision matrices. Finally, Model (e) is a conventional Gaussian Mixture Model (Bishop 2006) that has been chosen as a baseline method given the similarities with the ICC approach. We analysed and compared the resulting clusters both in terms of market properties to which the two clusters are associated and in terms of temporal consistency. First, we focused on a subset of 100 stocks chosen at random among those that have been continuously traded throughout the observed period. Random choice of the basket is to avoid selection bias. We then consider random resamplings to assess the robustness when different stocks are considered.

(a) Time series segmentation results

(b) Mean (*left*) and standard deviation (*right*) of each stock for each temporal cluster

Figure 1.: Clustering segmentation for experiment 1 over the whole dataset. Panel (a) reports the cumulative average return at each time t across the 100 stocks; in this picture, the blue background corresponds to time instances assigned to Cluster 1 and the orange background correspond instead to time instances assigned to Cluster 2. Panel (b) reports mean and standard deviation of each of the 100 stocks respectively computed using the returns assigned to each of the 2 clusters. We observe that Cluster 1 exhibits positive mean returns (*'bull'* state) and lower levels of volatility for all the considered stocks, while for cluster 2 all the stocks present negative mean returns (*'bear'* state) and higher levels of volatility.

We optimized the temporal consistency parameter by grid-searching as described in Appendix A and used $\gamma = 1.313$ for ICC Sparse (a) and $\gamma = 0.88$ for ICC Full (b) in both the experiments presented in this paper. The two referential precision matrices, J_1 and J_2, obtained with this experiment had 344 non-zero entries (dependency network edges) of which 181 were common to both states showing a good level of differentiation, but also significant overlaps between the two market states. The number of points assigned to each cluster were respectively 2895 for cluster 1 and 1904 for cluster 2. Figure 1 reports with colored background the points' assignment for the
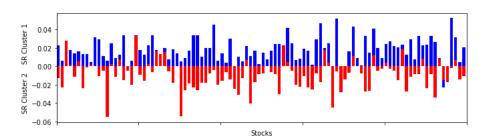
Figure 2.: Estimated Sharpe Ratio (SR) for each of the 100 stocks in the sample. The blue bars report the SR computed from log-returns in Cluster 1, whereas the red bars report the SR computed from log-returns in Cluster 2.

two clusters. We can observe there is a good spatial consistency. For instance, the average number of consecutive days in cluster 1 is 27.8 days. We also note that cluster 1 (blue background) tend to be associated with periods of rising market prices whereas cluster 2 (orange background) appears more present during crisis and market downturns. We indeed discovered that -automatically- the methodology assigns 'bull' market periods (positive mean returns) to cluster 1 and 'bear' market periods (negative mean returns) to cluster 2. We can for instance observe in Figure 1(a) that 82 consecutive observations during the 2001-2002 *.com* bubble crisis and 126 consecutive observations during the 2007-2008 global financial crisis have been assigned to the *bear* cluster 2. From Fig.1(b) we observe that the *bull* cluster 1 has, indeed, average positive returns for all stocks whereas the *bear* cluster 2 has average negative returns. Furthermore, also the standard deviations are different between the two cluster assignments.

To compare the two clusters on a risk-adjusted basis, we computed the Sharpe ratio (Sharpe 1966, 1994) for each stock in each cluster. We found for the *bull* cluster an average Sharpe ratio equal to 0.28, with 5th and 95th percentiles respectively equal to 0.15 and 0.41, while the *bear* cluster had average -0.38, with -0.6 and -0.15 as 5th and 95th percentiles. It is therefore clear that the two clusters have very different risk-return profiles. In order to further quantify the statistical significance of this difference, we considered the SR asymptotic distribution derived in (Opdyke 2007) to test for the risk-adjusted structural difference of the two clusters. We found that all of the 100 stocks in the sample had a Sharpe ratio significantly larger than 0 in the bull cluster and significantly lower then 0 in the bear cluster at a significance level $\alpha = 0.01$. Figure 2 reports the Sharpe ratios in the two clusters for the 100 stocks. In order to verify robustness and generality of the results we computed the same quantities for 100 other randomly chosen baskets of 100 stocks. For all resampled baskets of stocks we found a consistent clusterization in *bull* and *bear* regimes with had Sharpe ratios for at least 95% of stocks significantly larger than zero for the bull state and significantly smaller than zero for the bear state. Across the 100 resamplings, the two clusters had average number of elements respectively equal to 3087 and 1585.

Sparsity and Temporal Consistency. In order to assess the role of sparsity and temporal consistency, we performed the same analysis on the 'alternative' ICC Models (b)-(d) and the GMM (e).

Table 1 summarizes the number of stocks having significant Sharpe ratio in both clusters over 100 resamplings. In the table, each couple refers to the number of significant stocks in *bull* (left) and *bear* (right) states. We found that, in absence of temporal consistency constraints, both the ICC models (c, d) meaningfully classify clusters with and without sparsity. However, when temporal consistency is considered, ICC Full (b) is significantly affected by the constraint while ICC Sparse (a) provides robust results. GMM delivered the worst clusters in terms of risk/return significance.

	Median	5th **percentile**	95th **percentile**
GMM	(,)	(,)	(,)
ICC **Full,** $\gamma = 0$	(77,78)	(67,71)	(92,98)
ICC **Sparse,** $\gamma = 0$	(85,87)	(69,75)	(96,95)
ICC **Full,** $\gamma = 14.7$	(73,74)	(68,65)	(78,80)
ICC **Sparse,** $\gamma = 16$	(75,81)	(65,69)	(86,90)

Table 1.: Sharpe ratio significance for ('*bull*','*bear*') states. Median, 5th and 95th percentiles obtained from 100 random resamples of the stocks composing the dataset.

Focusing on temporal consistency, Table 2 reports the number of switches and the segment length resulting from the cluster assignments of the five models. When no temporal consistency is enforced (c,d), ICC provides the less temporal consistent results with small differences related to sparsity. This also explains the good results obtained by the models in terms of risk/return significance. When constrained to be temporal consistent, ICC Full (b) shows large variability in temporal consistency across samples with some having only a few switches over the whole period and others having several hundreds. ICC Sparse (a) is instead more consistent with a few hundred switches over the whole period which are about 1/3 of the switches in GMM (e).

Number of Switches			
	Median	5th **percentile**	95th **percentile**
GMM	785	540	874
ICC **Full,** $\gamma = 0$	1203	992	2176
ICC **Sparse,** $\gamma = 0$	1157	727	1421
ICC **Full,** $\gamma = 14.7$	204	120	306
ICC **Sparse,** $\gamma = 16$	208	54	298

Segment length			
	Median	5th **percentile**	95th **percentile**
GMM	5.07	2.4	11.8
ICC **Full,** $\gamma = 0$	3.3	1.68	4.38
ICC **Sparse,** $\gamma = 0$	3.5	2.8	6.65
ICC **Full,** $\gamma = 14.7$	22.64	14.6	38.26
ICC **Sparse,** $\gamma = 16$	23.6	18	55.27

Table 2.: Temporal consistency metrics. Number of switchings and Segment lengths over 100 resampligs.

3. Role of sparsity

In previous works (Barfuss *et al.* 2016), the TMFG-LoGo approach has proven to perform better than other filtering approaches including GLasso and Ridge providing the additional advantages of efficiency and fixed sparsity level with no need to calibrate hyperparameters (Massara *et al.* 2015). In this Section we motivate the choice of TMFG-LoGo filtering procedure in terms of statistical significance by comparing the performances of the TMFG-LoGo to the cross-validated Ridge l_2 penalized inverse covariance (Ridge) on our dataset. We considered the widely used Ridge penalization as robust estimate of the empirical inverse covariance matrix and compared it to TMFG-LoGo and show that, when applied to our dataset, TMFG-LoGo produces more stable likelihood results than Ridge. Coherently with our second experiment (Section 4), we used 40% of the data (from 31/12/2007 to 31/12/2015) as test set, and we considered as train sets the q observations preceding the test set (until 30/12/2007). The penalization parameter of Ridge was defined by cross validating within the train set. To compare TMFG-LoGo and the cross-validated Ridge we computed the log-likelihoods $\mathcal{L}_{s,k} = 1/2(\log |\mathbf{J}_k| - d_{s,k}^2 - p\log(2\pi))$ using the two covariance estimates and compared them. Figure 3 shows the likelihood observation-wise computed in train and in test using

the TMFG-LoGo and Ridge precision matrices estimated over $q = 500$ observations. The TMFG-LoGo likelihoods are much more stable over time suggesting that the procedure was successful in filtering out noise. Table 3 reports details on mean, 5th and 95th percentiles of the likelihoods computed in the train and test set. As previously mentioned, TMFG-LoGo likelihoods are much more stable with 5th and 95th varying a few percent only for TMFG-LoGo and instead varying of more than one order of magnitude in Ridge. We found similar results for TMFG-LoGo and Ridge when different values of q are considered. Note that Ridge log likelihoods have large differences between train and test. This is a typical indication of overfitting. Conversely, TMFG presents small differences indicating that the LoGo procedure acts as a topological penalize.
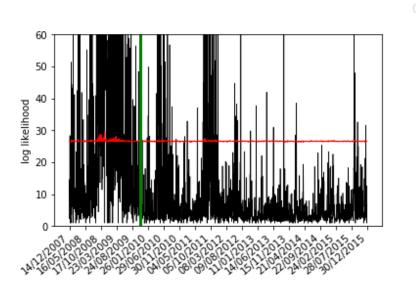

Figure 3.: Train and test log likelihood observation-wise using TMFG (red line) and Ridge (black line) precision matrices. The green vertical line divides train and test set. Ridge peaks reach values outside the range up to 320.

	Train set		
	Average	$5^{th} percentile$	$95^{th} percentile$
$\mathcal{L}_{Ridge}$	41.70	2.19	188.85
$\mathcal{L}_{TMFG}$	26.71	26.53	27.22

	Test Set		
	Average	$5^{th} percentile$	$95^{th} percentile$
$\mathcal{L}_{Ridge}$	8.08	1.39	27.64
$\mathcal{L}_{TMFG}$	26.55	26.44	26.73

Table 3.: TMFG and Ridge log likelihood metrics - means, 5th and 95th percentiles - computed in train (top panel) and test (bottom panel) set. TMFG and Ridge precision matrices are estimated using $q = 500$ observations.

4.   Forecasting

In the second experiment we used our methodology to forecast future states of the market form previous observations. To this end, we used the first 60% of the data (from 02/01/1995 to 31/12/2007)
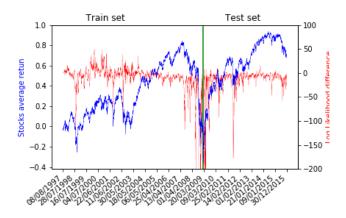
Figure 4.: Log likelihood ratio and mean returns across train and test sets. The log likelihood ratio of the two states $\mathcal{R}_t$ was computed using using the $\Delta = 28$ *days*. The green vertical bar indicates the end of the train set and the beginning of the test set. We estimated $\boldsymbol{J_1}$ and $\boldsymbol{J_2}$ in train and held it fixed for the computation of $\mathcal{R}_t$ also in the test set. The black horizontal line identifies $\mathcal{R}_t = 0$ level, *i.e.* the level above which the *bull* state is more likely. Coherently with previous findings, we can identify persistent market states with a more frequent bull market and regions of bear market.

as train set from which we extracted the two referential precision matrices and means ($\boldsymbol{J_1}$, $\boldsymbol{\mu_1}$) and ($\boldsymbol{J_2}$, $\boldsymbol{\mu_2}$) (note they are different form the ones of the first experiment in which we used the entire dataset instead). We then forecasted the probability that, given an observation at time t, the observation at a following time $t + h$ would belong to state k. This is achieved by performing a logistic regression using the log likelihood ratio of the two clusters (Neyman and Pearson 1933) from a rolling window of length Δ:

$$\mathcal{R}_t = \sum_{s=t-\Delta+1}^{t} \mathcal{L}_{s,1} - \mathcal{L}_{s,2} \ , \tag{2}$$

where $\mathcal{L}_{s,k} = 1/2(\log|\mathbf{J}_k| - d_{s,k}^2 - p\log(2\pi))$ is the log-likelihood of state . In our experiment, we considered $\Delta = 28$ days, since this is the average length of segments obtained from ICC (a) in the first experiment. Figure 4 provides a visual representation of the likelihood ratio computed for each cluster and of its evolution as compared to market movements. The green vertical line divides the train set form the test set. The logistic regression of market states $\mathcal{K}_t$ against the log likelihood ratio $\mathcal{R}_t$ can be written as

$$P(\mathcal{K}_{t+h} = 1, 2 \,|\, \mathcal{R}_t = x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}} \ , \tag{3}$$

where the parameters β_0 and β_1 are estimated through maximum likelihood (Bishop 2006). We estimated all parameters ($\boldsymbol{J_1}$, $\boldsymbol{J_2}$, $\boldsymbol{\mu_1}$, $\boldsymbol{\mu_2}$, γ, β_0 and β_1) in the train set and estimated a threshold or cut-off point of 0.57 by cross-validation in the train set. We then used these parameters to predict, in the test set, the next day state given the log-likelihood ratio $\mathcal{R}_t = x$. Specifically, we predict $\widehat{\mathcal{K}}_{t+1} = 1$ if $P(\mathcal{K}_{t+1} = 1 \,|\, \mathcal{R}_t = x) > 0.57$ and $\widehat{\mathcal{K}}_{t+1} = 2$ otherwise. For instance, for the day 30-Jan-2008 (test set) we predicted a *bear* state with probability $P(\mathcal{K}_{30-Jan} = 2 \,|\, \mathcal{R}_{29-Jan}) = 0.72$, where $\mathcal{R}_{29-Jan}$ was computed using the observations from 02-Jan to 29-Jan-2008 ($\Delta = 28$ days, all in the test set) and the parameters $\boldsymbol{\mu}_k$, $\boldsymbol{J}_k$, γ, β_0 and β_1 were the ones calibrated on the train set with data until 31/12/2007.

To assess the goodness of our approach we compared test set predictions with the classification performed over the whole period in the first experiment (see Fig.1). We used three metrics (Hastie *et al.* 2008) to assess the performance of our classification method: the True Positive Rate TPR (number of elements correctly assigned to cluster 1 divided by total number of elements in cluster 1), the True Negative Rate TNR (number of elements correctly assigned to cluster 2 divided by total number of elements in cluster 2) and Accuracy ACC (number of correct predictions in cluster 1 or 2 divided by total number of elements). In order to test for the robustness of our method, we randomly resampled the 100 stocks and performed the classification experiment considering the new dataset. We repeated this process 100 times and stored the three performance metrics TPR, TNR and ACC. Table 4 presents a summary of the results obtained. As we can see, ACC and TPR are higher than 50% at the 5th percentile. TNR presents a good median result, but a low 5th percentile showing that it can be difficult to correctly forecast. This indicates that there is a tendency to over-assign time-instances to cluster 1 (*bull* state) and conversely missing predictions for the less frequent *bear* state. Nonetheless, we verified (by using the hypergeometric distribution as reported in (Aste and Di Matteo 2017)) that these TNR are statistically significant at 0.01 level indicating that there is, indeed, significant prediction power also for the bear state. Let us stress that the present forecasting exercise is not optimized and there are several ways these performances can be improved. However, this is beyond the purpose of the present paper where we privileged simplicity over performances.

	Median	5th **percentile**	95th **percentile**
TPR	0.57	0.52	0.89
TNR	0.52	0.22	0.67
ACC	0.55	0.49	0.68

Table 4.: Out-of-sample performance metrics using the ICC log likelihood ratio as independent variable. Median, 5th and 95th percentiles obtained from 100 random resamples of the stocks composing the dataset.

To compare the previous results with a baseline method, we estimated the logistic regression in Eq. 3 using the fraction of stocks that at time $t-1$ were presenting positive returns as independent variable. Aim being to compare our ICC log-likelihood to a much simplified version of the the information about the correlation structure. Same estimation scheme is used and a treshold of 0.61 is obtained by cross validation. Results are reported in Table 5. While this simplified information still provides a median accuracy above 50%, the model has overall inferior performances with respect to the ICC log likelihood ratio case reported in Table 4.

	Median	5th **percentile**	95th **percentile**
TPR	0.39	0.41	0.63
TNR	0.55	0.35	0.77
ACC	0.51	0.42	0.59

Table 5.: Out-of-sample performance metrics using the fraction of positive strocks as independent variable. Median, 5th and 95th percentiles obtained from 100 random resamples of the stocks composing the dataset.

5. Conclusions

In this paper we presented a novel methodology to define, identify, classify and forecast market states. In addition to accuracy, intuitiveness and forecasting power, our procedure is numerically very efficient and able to process high dimensional datasets. We reported two experiments to illustrate that the method is efficient and reliable in identifying and predicting accurate and in-

terpretable structures in multivariate, non-stationary financial datasets. These two examples use two clusters and 100 variables, however we verified that analogous results hold for larger or smaller numbers of variables and similarly interesting classifications emerge also when three or more clusters are used. The choice of two clusters has been only motivated by simplicity. The fact that they turned out to be respectively populated mostly with average positive and negative returns associated with pre- and post-crisis periods was unexpected by us and opens potentials for completely novel ways to use multivariate analytics for the forecasting of stock market returns. This also greatly simplified the interpretation of these states as 'bull' and 'bear' markets. Of course, in reality, there are more than two market states and common definition of bull and bear markets are often blurry. In this work we did not attempt to optimize results favouring, instead, simplicity and interpretability and, therefore, there is a large open domain of exploration to refine the methodology. We also adopted several methodological choices that can be modified in future works. For instance, the segmentation with the Mahalanobis distance turned out o be a very powerful in the reported experiments, however there is a broad range of possible metrics for clustering and experiments with Euclidean distance or Likelihood also produce interesting results. Further, the choice of TMFG network over other possible information filtering networks or other sparsification methodologies can be investigated. Temporal consistency could had also being performed differently by using a hidden Markov model approach (see note in Appendix A). The choice of logistic regression to forecast market state is just one simple possibility among many regression options that might make better use of the information content of our regimes' structures. All these and other methodological choices has been motivated by simplicity and intuitiveness. Since one of the main achievement of our methodology is computational efficiency allowing to apply the methodology to high dimensional datasets, further work will include new sources of information (e.g. news, economic indicators, sentiment).

References

Ang, A. and Bekaert, G., International Asset Allocation with Time-Varying Correlations. *Review of Financial Studies*, 2002, **15**, 1137–1187.

Ang, A. and Chen, J., Asymmetric correlations of equity portfolios. *Journal of Financial Economics*, 2002, **63**, 443–494.

Aste, T. and Di Matteo, T., Causality network retrieval from short time series. *arXiv preprint arXiv:1706.01954*, 2017.

Barfuss, W., Massara, G.P., di Matteo, T. and Aste, T., Parsimonious modeling with Information Filtering Networks. *Phys. Rev. E*, 2016, **94**, 062306.

Bishop, C.M., *Pattern Recognition and Machine Learning*, 2006, Springer-Verlag New York.

Black, F. and Litterman, R., Global Portfolio Optimization. *Financial Analysts Journal*, 1992, **48**, 28–43.

Bollerslev, T., Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model. *The review of economics and statistics*, 1990, **4**, 498–505.

Campbell, J.Y., , Lo, A.W. and MacKinlay, A.C., *The Econometrics of Financial Markets*, 1997 (Princeton University Press: Princeton, N.J.).

Cizeau, P., Potters, M. and Bouchaud, J.P., Correlation structure of extreme stock returns. *Quantitative Finance*, 2010, **1**, 217–222.

Danielsson, J., *Financial risk forecasting: the theory and practice of forecasting market risk with implementation in R and Matlab*, 2011 (Wiley-Blackwell: Hoboken).

De Maesschalck, R., Jouan-Rimbaud, D. and Massart, D.L., The mahalanobis distance. *Chemometrics and intelligent laboratory systems*, 2000, **50**, 1–18.

Dempster, A.P., Laird, N.M. and Rubin, D.B., Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 1977, **39**, 1–38.

Duffie, D. and Pan, J., An Overview of Value at Risk. *The Journal of Derivatives*, 1997, **4**, 7–49.

Engle, R., Dynamic Conditional Correlation. *Journal of Business & Economic Statistics*, 2002, **20**, 339–350.

Focardi, S.M. and Fabozzi, F.J., A methodology for index tracking based on time-series clustering. *Quantitative Finance*, 2004, **4**, 417–425.

Friedman, J., Hastie, T. and Tibshirani, R., Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 2008, **9**, 432–441.

Grabarnik, P. and Särkkä, A., Interacting neighbour point processes: Some models for clustering. *Journal of Statistical Computation and Simulation*, 2001, **68**, 103–125.

Hallac, D., Nystrup, P. and Boyd, S., Greedy Gaussian Segmentation of Multivariate Time Series. *ArXiv e-prints*, 2016.

Hallac, D., Vare, S., Boyd, S.P. and Leskovec, J., Toeplitz Inverse Covariance-Based Clustering of Multivariate Time Series Data. *CoRR*, 2017, **abs/1706.03161**.

Hamilton, J.D., A new approach to the economic analysis of nonstationary time series and the business cycle. *ECONOMETRICA*, 1989, **57**, 357–384.

Hastie, T., Tibshirani, R. and Friedman, J., *The elements of statistical learning*, 2008 (Springer: New York, USA).

Hendricks, D., Gebbie, T. and Wilcox, D., Detecting intraday financial market states using temporal clustering. *Quantitative Finance*, 2016, **16**, 1657–1678.

Kim, C.J. and Nelson, C., *State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications*, 1999 (The MIT Press: Cambridge, MA).

Laloux, L., Cizeau, P., Bouchaud, J.P. and Potters, M., Noise Dressing of Financial Correlation Matrices. *Phys. Rev. Lett.*, 1999, **83**, 1467–1470.

Lauritzen, S.L., *Graphical Models*, 1996 (Oxford University Press: Oxford).

Lee, S. and Stevenson, S., Time Weighted Portfolio Optimisation. *Journal of Property Investment & Finance*, 2003, **21**, 233–249.

Liao, W.T., Clustering of Time Series Data-a Survey. *Pattern Recogn.*, 2005, **38**, 1857–1874.

Lin, W.L., Engle, R. and Ito, T., Do Bulls and Bears Move across Borders? International Transmission of Stock Returns and Volatility. *Review of Financial Studies*, 1994, **7**, 507–38.

Longerstaey, J. and Spencer, M., *RiskMetrics Technical Document*, 1996, J.P. Morgan/Reuters.

Massara, G.P., di Matteo, T. and Aste, T., Network Filtering for Big Data: Triangulated Maximally Filtered Graph. *CoRR*, 2015, **abs/1505.02445**.

McLachlan, G. and Krishnan, T., *The EM algorithm and extensions*, 1997, Wiley.

Musmeci, N., Aste, T. and Di Matteo, T., What does past correlation structure tell us about the future? An answer from network filtering. *ArXiv e-prints*, 2016.

Musmeci, N., Nicosia, V., Aste, T., Di Matteo, T. and Latora, V., The multiplex dependency structure of financial markets. *Complexity*, 2017, **2017**, 13.

Nevill-Manning, C.G. and Witten, I.H., Identifying Hierarchical Structure in Sequences: A Linear-time Algorithm. *Journal of Artificial Intelligence Research*, 1997, **7**, 67–82.

Neyman, J. and Pearson, E.S., IX. On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 1933, **231**, 289–337.

Opdyke, J., Comparing Sharpe Ratios: So Where are the P-Values?. *Journal of Asset Management*, 2007, **8**, 308–336.

Ren, L., Wei, Y., Cui, J. and Du, Y., A sliding window-based multi-stage clustering and probabilistic forecasting approach for large multivariate time series data. *Journal of Statistical Computation and Simulation*, 2017, **87**, 2494–2508.

Schmitt, T.A., Chetalova, D., Schäfer, R. and Guhr, T., Non-stationarity in financial time series: Generic features and tail behavior. *EPL (Europhysics Letters)*, 2013, **103**, 58003.

Sharpe, W.F., Mutual Fund Performance. *The Journal of Business*, 1966, **39**, 119–138.

Sharpe, W.F., The Sharpe Ratio. *The Journal of Portfolio Management*, 1994, **21**, 49–58.

Tong, H., On a Threshold Model. In *Pattern Recognition and Signal Processing*, edited by in C. H. Chen, 1978 (Sijthoff and Noordhoff: Amsterdam).

Tsay, R., *Analysis of financial time series*, 2. ed. , 2005 (Wiley: Hoboken, NJ).

Viterbi, A., Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory*, 1967, **13**, 260–269.

Zolhavarieh, S., Aghabozorgi, S. and Wah Teh, Y., A review of subsequent time series clustering. *The Scientific World Journal*, 2014, **2014**.

Appendix A: The Viterbi algorithm

Figure A1 provides a visualization of the problem of assigning points to clusters. Based on the parameters estimates ($\boldsymbol{\mu}_k$ and $\boldsymbol{J}_k$ via TMFG-LoGo) from the E step of the Expectation Maximization procedure, we compute the Mahalanobis distance of every multivariate observation obtaining, for each cluster k and for each observation t, a value $d_{t,k}^2 = (\boldsymbol{X}_t - \boldsymbol{\mu}_k)^T \boldsymbol{J}_k (\boldsymbol{X}_t - \boldsymbol{\mu}_k)$.
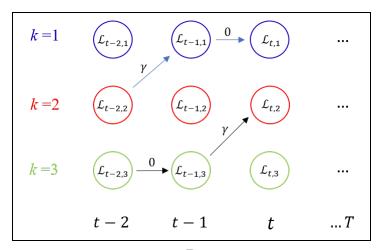

Figure A1.: Example of two among the K^T possible paths considering $K = 3$ clusters and T observations. $\mathcal{L}_{t,j}$ represents the log likelihood of the multivariate observation at time t if assigned to cluster j. If an observation is assigned to same cluster as the previous one, no penalty is applied, otherwise a *cost* weighted by the parameter γ is added.

We need to consider the best *sequence* of latent states which is not the set of best individual states. In particular, if we introduce a cost parameter γ that penalizes cluster switching, the problem complexity becomes combinatorial, since we need to account for the whole sequence or *path* of assignations. In particular, given K potential cluster assignment of T points (multivariate observations), the number of potential paths grows exponentially with the length of the chain to K^T possible assignments of points to clusters. Based on a dynamic programming approach, the Viterbi algorithm (Viterbi 1967) provides an efficient solution with complexity $O(KT)$ (*i.e., linear*) to this problem, searching the space of the paths and finding the most efficient path. The Viterbi algorithm in the convenient formulation by (Hallac *et al.* 2017) is sketched in 1.

A more general formulation can be implemented by describing the paths as Markov chains and introducing a transition probability between the states. However, under the Markov chain formalism the expression in Eq.2 for the likelihood ratio is no longer consistent because it implies implicitly iid multivariate observations.

Algorithm 1 Viterbi algorithm

Input

$d_{t,k}^2 = $ square Mahalanobis distance of observation t if assigned to state k

$\gamma = $ time consistency parameter

Initialize

      PreviousCost = array of K zeros

      CurrentCost = array of K zeros

      PreviousPath = array of K elements

      CurrentPath = array of K elements

for each observation $t = 1, ..., T$ **do**

    for each state $k = 1, ..., K$ **do**

        MinVal = index of minimum value of PreviousCost

        if PreviousCost$[MinVal] + \gamma > $ PreviousCost$[k]$ **then**

            CurrentCost$[k] = $PreviousCost$[k] - d_{t,k}^2$

            CurrentPath$[k] = $PreviousPath$[k]$.append$[k]$

        else

            CurrentCost$[k] = $PreviousCost$[MinVal] + \gamma - d_{t,k}^2$

            CurrentPath$[k] = $PreviousPath$[MinVal]$.append$[k]$

        PreviousCost=CurrentCost

        PreviousPath=CurrentPath

FinalMinVal=index of minimum value of CurrCost

FinalPath=CurrPath[FinalMinVal]