Indexing Financial Stress

Rik de Wilde
334391

June 27, 2013

Abstract

As we have experienced one of the greatest downturns in worldwide economy, there is a high demand for preventing future recessions. It is possible to translate disruptions in the financial markets into financial stress. We use several previously untried methods for constructing a financial stress index. With these indexes, we are able to capture historical events in terms of financial stress, such as the default of Lehman Brothers and the huge decrease in housing prices. By this, the evaluated financial stress index seems useful, and more importantly, we can forecast high levels of stress in order to prevent future downturns. We find that the clustering algorithm with dissimilarities based on Euclidean distances and clustered with the partitioning around medoids (PAM) algorithm performs the best. Overall, this paper shows how financial stress can be indexed in various ways, and recommends that future research should account for financial stress indexes based on clustering algorithms.

1 Introduction

The recent financial crisis and the associated decline in economic activity have raised some important questions about economic activity and its links to the financial sector. It is believed that the crisis started when the housing bubble bursted and the mortgage backed securities turned out to be near worthless. The resulting turmoil spread across a number of asset classes and markets, which ultimately led to the collapse of major financial institutions. Researchers are interested whether it is possible to signal events in the market, which we call financial stress. If we can signal financial stress within the market, it may be possible to prevent a recession. A recession is always preceded by an economic downturn, which is less severe than a recession. A downturn suggests the rate of economic growth is slowing down, whereas a recession is defined as a significant decline in economic activity spread across the economy, lasting more than a few months. By creating a financial stress index (FSI), researchers try to capture all economic activity within one single variable to prevent these recessions.

One of the first researchers to create an index to measure the financial stress of an economy were Illing and Liu (2006). To create this index, Illing and Liu tried many weighing techniques to assign weights to different financial variables. To evaluate their constructed index an internal survey was held at the national bank, to determine the financial crises that occurred in the past years. With this survey, they verified their constructed FSIs. The 'winning' FSI resulted in 13% type 1 errors, which means that a recession or downturn did occur, although it was not predicted by their FSI. A possible explanation is a downturn which is caused by unforeseen circumstances, such as 9/11. Another statistic error was more persistent in the FSI: 33% type 2 errors occurred, which means that a recession or downturn was predicted when in reality there was no downturn or recession. This makes sense, as it is not always the case that preceding episodes of stress lead to a downturn or recession.

Cardarelli et al. (2009) analyzed the experience of episodes of stress in banking, securities and foreign exchange markets in seventeen advanced economies. The paper finds that financial stress is often, but not always a precursor to an economic slowdown or recession. Banking stress in particular tends to lead to greater effects on downturns or recessions, despite the fact that financial innovation has increased the role of securities markets in many countries.

However, the main focus of this paper will be on the work of de Wilde et al. (2013), where FSIs were constructed based on multiple variables of the U.S. market. They chose to construct multiple FSIs, and pick the one which had the best out-of-sample performance. A total of eight indexes were constructed, of which three were chosen to be modeled with different autoregressive models: the multiple principal component FSI standardized recursively (PCRFSI), the market-based weighted sum FSI standardized recursively (WSRFSI) and standardized with a moving window (WSMFSI). Three models were then used to model and forecast these FSIs; an autoregressive moving average (ARMA) model, a vector autoregressive (VAR) model and a heterogeneous autoregressive (HAR) model. The PCRFSI turned out to be the best in-sample model, but the WSMFSI and WSRFSI outperformed the PCRFSI out-of-sample. The VAR model was also outperformed by the ARMA and HAR models. As the HAR model should outperform the ARMA model for h-step-ahead forecasts where h is large, and the WSMFSI was just slightly better performing in-sample than the WSRFSI, de Wilde et al. chose their 'best' FSI to be the WSMFSI modeled according to HAR theory.

We will extend the work previously done by de Wilde et al. by proposing alternative ways to construct FSIs, which will be compared and evaluated with their 'winning' FSI. We will consider the following construction methods: an index based on recursive principal component analysis (PCA) as in Erdogmus et al. (2004); an index based on supervised PCA as proposed in Bair et al. (2006); an index based on 'regular' PCA as a benchmark for aforementioned construction methods; indexes based on clustering methods in which we will follow Musetti; and an index based on a logistic regression which is derived of a footnote of Illing and Liu (2003). The constructed FSIs will be evaluated in-sample as done previously by Illing and Liu (2006), and then be modeled with autoregressive (AR) and HAR models to evaluate out-of-sample forecasts. We will also investigate if it is possible to average our variables on a monthly basis.

2 Methodology

As proposed in the introduction, we will construct multiple FSIs. The main goal of a FSI should be to capture a certain level of 'stress'. This stress should be visible in indicators of the financial market. Whenever a variable returns high values, given that this variable is stationary and standardized, we assume this indicates financial stress. The same should apply for a FSI: if there is stress on the financial market, high values should be returned, and if the market is stable the FSI should return values around zero (or even somewhat lower). Roughly, there are two main methods one can think of to create such an index. First, we can use the variables themselves and transform them in such a way they are usable as an index. This is done for the FSI based on a weighted sum of financial markets, the FSIs which are clustered and the FSI which is based on a logistic regression. Second, we can use PCA to convert the multiple series into a smaller subset of vectors which should capture the explained variance of the data set. We use PCA for our 'regular' PCA FSI, and for a recursive and a supervised FSI.

2.1 FSI based on weighted sum of financial markets

First, we will evaluate all our FSIs with the 'benchmark' FSI proposed by de Wilde et al., which is the WSMFSI. However, as we will explain later, we will standardize our data recursively and use this for all our FSIs. To make a good comparison, we will use the WSRFSI which differs slightly of the WSMFSI.

The WSRFSI is a weighted sum of financial markets, based on Bloomberg's U.S. Financial Conditions Index of Rosenberg (2009). The idea is simple: some variables capture movements of a specific market, and this should be visible within a FSI. Whenever some variables capture the same movements of a market, this is extrapolated within a FSI based on these variables. Hence, we weigh the movements of the variables with respect to the market they represent.

In terms of equations we define X_{tj} to denote variable $j \in [1, m]$ at observation $t \in [1, T]$. We add $p \in [1, k]$ representing the market and c_p representing its respective (proportional) loading,

$$\text{WSRFSI}_t = \sum_{j=1}^{m} \sum_{p=1}^{k} c_p \cdot X_{tjp}, \tag{1}$$

with, as we use proportional loadings

$$\sum_{p=1}^{k} c_p = 1. \tag{2}$$

Further explanation on this loading c_p is as follows. Consider p markets, where the loading of each market is $1/p$. Then, each market contains v_p variables of the dataset, of which its respective loading becomes

$$c_p = \frac{1/p}{v_p}. \tag{3}$$

Let us assume we define two different markets, the first represented by three variables, the second just by one. This means that the first three variables will have a loading equal to $c_1 = \frac{1/2}{3} = \frac{1}{6}$, and the last variable will have a loading equal to $c_2 = \frac{1/2}{1} = \frac{1}{2}$.

2.2 FSI based on PCA

PCA evaluates a number of eigenvectors, which represents independent linear combinations, equal to the number of variables in the data set, and a same number of eigenvalues evaluated from the covariance matrix of the data. It is a very popular method to minimize the rank of a matrix while a large amount of explained variance remains within the newly constructed matrix.

Let Σ denote the $m \times m$ covariance matrix of the data set, z_j the jth $m \times 1$ eigenvector and λ_j the jth 1×1 eigenvalue

$$\Sigma z_j = \lambda_j z_j. \tag{4}$$

These eigenvectors can then be multiplied by the data set itself, resulting in the principal components. Let X_{tj} denote the $T \times m$ dataset, Z a $m \times m$ matrix containing all eigenvectors and PC_{tj} a $T \times m$ matrix containing the sample principal components

$$X_{tj}Z = PC_{tj}. \tag{5}$$

The calculated principal components each explain a certain proportion of the data set, which we call the explained variance. The explained variance is evaluated by $\frac{\lambda_j}{m}$.

2.2.1 Regular PCA

We will construct a FSI based on regular PCA. This is partly done to evaluate our other FSIs based on PCA, which will be explained later. The regular PCA FSI will be constructed by using multiple principal components, which are included in the FSI weighed by their own eigenvalue. The weights will be scaled in such a way that these will sum up to one, to make sure this FSI can be compared with the other FSIs. We use multiple principal components as the use of only one principal components does not cover enough explained variance in our taste. The idea of using multiple principal components for the construction of a FSI was first employed by de Wilde et al. (2013).

The question rises how we should determine the amount of principal components we use for our FSI. Based on literature, three possible ways come to mind: first, we can use the famous SCREE-plot, where we look for a certain 'elbow' within the plotted eigenvalues per principal component. The elbow indicates that starting from this point, the explained variance will not decrease as quick as with previous components. This indicates that we should use all principal components before this point. We can also use a 'rule of thumb', which says to simply add every principal component with an eigenvalue greater than one. However, we use a certain 'benchmark' to evaluate how many principal components we should use for our FSI. The benchmark is set at a certain level of explained variance we want to capture within our FSI. We set the benchmark at 60% as we believe this suffices for enough explained variance.

2.2.2 Recursive PCA

An interesting approach of PCA for time series has been brought forward by Erdogmus et al. (2004). The proposed method of Erdogmus et al. updates the eigenvector and eigenvalue matrices simultaneously with every new sample such that the estimates approximately track their true values as would be calculated from the current sample estimate of the data covariance matrix. This method is called recursive PCA. This means that, for every $t \in [1, T]$ we perform PCA and store the values at t in another eigenvector and eigenvalue matrix. In theory, this should overcome a problem of PCA, which is that PCA performs well within-sample, but has poor results out-of-sample. By performing PCA recursively, we hope to overcome this last problem.

To create the PCRFSI, we again use multiple principal components. This is done in the same way as for regular PCA.

2.2.3 Supervised PCA

PCA tries to minimize the rank of a matrix by composing multiple principal components which covers the variance explained. However, the rank of the initial matrix may also be reduced by erasing variables which have little explanation power. This is basically the idea of Bair et al. (2006), which introduces a technique called supervised PCA. Supervised principal components is similar to conventional PCA except that it uses a subset of the variables, based on their association with the outcome. In a nutshell, the procedure comes down to the following:

1. Compute (univariate) standard regression coefficients for each variable.

2. Form a reduced data matrix consisting of only those variables whose univariate coefficient exceeds a certain threshold.

3. Compute the first (or first few) principal components of the reduced data matrix.

So, we should regress all our variables separate on some 'outcome' of stress, and then select those who live up to some threshold. We will regress the variables on a 0/1-vector, which is based on a list of 'stress-events' relevant for our data set. The idea of creating a 0/1-vector stems from Cardarelli et al. (2009), where it was used to evaluate already constructed FSIs. This 0/1-vector will be explained more extensively in the next section. We select variables which have a statistically significant coefficient at a 5% level.

2.3 FSI based on cluster analysis

Cluster analysis essentially has the same idea as weighing the variables to their respective market, with the exception that we now somehow try to quantify these clusters instead of defining clusters based on theory. The standard procedure for clustering variables comes down to the following:

1. We should define a certain measure of dissimilarity (how far two objects are apart from each other) between the variables.
2. With this dissimilarity, we should have a method which clusters variables together.
3. Find a number of clusters which divides the variables in an optimal way.

There are many ways to do this. We should somehow have a guideline for clustering variables of a financial time series. Luckily, Musetti (2012) compares build-in clustering algorithms of the statistical program R for a financial time series. Measuring dissimilarity with Hoeffding's D and finding clusters using an algorithm known as partitioning around medoids (PAM), also known as k-medoids, proved to fit the financial time series the best. We will compare these clustering methods with more common approaches, and find an optimal way to perform cluster analysis for our data set.

2.3.1 Dissimilarity

The most common approach of finding dissimilarity between variables i and j is the Euclidian distance, which is specified as follows for a data matrix X,

$$d(i,j) = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \ldots + (x_{iT} - x_{jT})^2}. \tag{6}$$

Derived from Kaufman and Rousseeuw (2009), the measure of Euclidian distance satisfies the following mathematical requirements for measuring distance:

1. $d(i,j) \geq 0$
2. $d(i,i) = 0$
3. $d(i,j) = d(j,i)$
4. $d(i,j) \leq d(i,h) + d(h,j)$

The first condition merely states that all distance are nonnegative, the second that the distance to itself is equal to zero, the third that distances are symmetrical, and the fourth condition states that an alternative distance which includes another point should not decrease the direct distance between objects i and j. Kaufman and Rousseeuw (2009) defined dissimilarities as nonnegative numbers $d(i,j)$ that are small (close to zero) when i and j are near to each other, and become large when i and j are very different. This satisfies the conditions of the Euclidian distances, and as it is an intuitive approach to define dissimilarity between variables, it is widely applied for estimating dissimilarities.

An alternative approach of measuring dissimilarities is Hoeffding's D. Hoeffding (1948) proposed a test for the independence of two random variables with continuous distribution function.

As it is a non-parametric test, it does not assume any distribution in the population. Essential for Hoeffding's D is the rank order of the two variables i and j, on which D depends. Each set of (i_t, j_t) values are cut points for classification. The formula for Hoeffding's D is

$$D = 30\frac{(n-2)(n-3)A + B - 2(n-2)C}{n(n-1)(n-2)(n-3)(n-4)}, \tag{7}$$

where

- $A = \Sigma_t(Q_t - 1)(Q_t - 2)$,
- $B = \Sigma_t(R_t - 1)(R_t - 2)(S_t - 1)(S_t - 2)$,
- $C = \Sigma_t(R_t - 2)(S_t - 2)(Q_t - 1)$.

R_t is the rank of variable i, S_t is the rank of variable j, and Q_t equals 1 plus the number of points with both i and j values less than the tth point. We should keep in mind that a point in i and/or j can be tied: that is, we observe the same value in the data set. A point that is tied on only the i_t or j_t value contributes 1/2 to Q_t, if the other value is less than the corresponding value for the tth point. A point that is tied both on i_t and j_t contributes 1/4 to Q_t. Hoeffding's D lies on the interval $[-0.5, 1]$ if there are no tied ranks, with larger values indicating a stronger relationship between the variables.

2.3.2 Searching for clusters

A widely-applied method for clustering is the so-called k-means clustering, introduced by MacQueen et al. (1967). It aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean. The main idea is to define k centroids, one for each cluster. Note that we have to define how many clusters we want to include a priori. These centroids should be placed in a 'smart' way, because different locations cause different results. Therefore, the centroids are placed as far away as possible from each other. The next step is to take each point belonging to the data set and associate it to the nearest centroid. When all points are associated to a centroid, the first step is completed and an early 'cluster' has been made. At this point we recalculate k new centroids of the clusters resulting from the previous step. After we have these k new centroids, we again associate the data set points with the nearest new centroid. This procedure will apply until the k centroids do not change their location anymore.

Another method we consider is the PAM algorithm, which is also known as k-medoids clustering. It is similar to k-means clustering, as both algorithms break up the data set into groups and attempt to minimize the distance between the points labeled to be in a cluster. However, the PAM algorithm chooses data points as centers and tries to minimize distances between data points in an arbitrary number of planes, whereas k-means only focuses on $\mathbb{R}^2$. It could be more robust to outliers and noise this way.

The PAM algorithm is as follows:

1. Initialization: randomly select k of the n data points as medoids
2. Allocate each data point to the closest medoid (depending on how we measure dissimilarity)
3. For each associated data point o within a medoid m we swap m and o and compute the average dissimilarity of o to all data points associated to m.
4. We select medoid o with the lowest average dissimilarity, and repeat this for every medoid m.
5. We repeat steps 2 - 4 until there is no change in medoids anymore.

As we now have defined k-means and k-medoids clustering, we note that we should define our k clusters a priori. The question remains how we can determine which amount of k clusters fits our data set in an optimal way. We use the silhouette technique for determining our clusters of data. The silhouette technique has been proposed by Rousseeuw (1987). For each data point o, we let $a(o)$ be the average dissimilarity of o with respect to all other data points within its cluster. We can interpret $a(o)$ as how well matched o is within the cluster. We then find $b(o)$, which represents the smallest dissimilarity data point o has to another data point within another cluster than its own. The cluster with the lowest dissimilarity is said to be the 'neighboring cluster' of o. We define

$$s(o) = \frac{b(o) - a(o)}{\max(a(o), b(o))}, \tag{8}$$

which can be written as

$$s(o) = \begin{cases} 1 - a(o)/b(o), & \text{if } a(o) < b(o) \\ 0, & \text{if } a(o) = b(o) \\ b(o)/a(o) - 1, & \text{if } a(o) > b(o). \end{cases} \tag{9}$$

It is now fairly easy to see that $-1 \leq s(o) \leq 1$, where a value close to one means that the data has been perfectly clustered, and a value close to minus one implies that the data point has been vary badly matched with its neighboring cluster. The average s(o) of all clusters is a measure of how tightly grouped all data points are. By applying this measure for all our choices of k clusters, we can determine which number of clusters is optimal.

We also employ the elbow method. We plot all clusters, and look for the "elbow" in the plot, which indicates the number of clusters you pick. For the criterion here, we will use the within-cluster sum-of-squares. This makes sure that the clusters are well-defined within their respective cluster. Another rule we keep in mind while determining our amount of clusters is the 'rule of thumb': $k = \sqrt{n/2}$. Hence, we make sure we do not choose too many clusters.

2.3.3 FSI based on a logistic regression

The widely acknowledged paper of Illing and Liu (2006) was based on their (working) paper of three years earlier (Illing and Liu (2003)). In this version, a footnote reads:

> "Another possible approach is to use implicit weights from simple non-linear probability models, such as probit and logit. Usually, these models are used to estimate probabilities, where the dependent variable is dichotomous. For the purposes of calculating implicit weights, however, both sides of the equation would be the same concept (i.e., financial stress), just measured in two different ways. (...). The coefficients on the stress variables could then be interpreted as the vector of weights, and the estimated value for the variable on the left-hand side would be the implicit-weight FSI. There are numerous technical questions associated with this methodology that remain unanswered, so we leave this experiment for future work."

This is an interesting concept which is quite easy to implement, as we have have constructed a 0/1-vector which represents stress in our defined period "in another way". The logit regression we will use is a type of regression used for predicting the outcome of a categorical vector, i.e. our 0/1-vector. The reason we prefer logit over probit is that the coefficients of the logit regression represent the change in the logit for each unit change in the predictor, which can be interpreted straightforward, whereas the coefficients of a probit model indicate some sort of predicted probability increase or decrease. We have several options to construct our FSI: we could consider only variables which have a significant coefficient, which can be determined by using a Wald test (similar to a t-test). We also could compute the so-called odds ratio of our coefficients, which is used to determine the strength of association between two variables, which could also generate some interesting results on which variables could be more strongly associated to stress. For practical purposes, we choose to keep it simple and use the coefficients of our logit regression as weights for our FSI. We add that we scale these coefficients in such a way that they sum up to one.

2.4 Modeling

As we now have defined all our FSIs, there are many ways to evaluate our FSIs. A first thing that springs to mind is to 'look' at our constructed FSIs. We will do this by plotting all our FSIs together. However, instead of just taking their values, we plot the Z-scores of their indexes: this makes it more visible to see where the FSIs might be different. Another idea is derived from Illing and Liu (2006), and was also used by de Wilde et al. (2013), which is to calculate the percentage false negative errors, where our FSI does not capture a downturn while there was a downturn, and false positive errors, where our FSI captures a downturn, while there actually was no downturn. The percentage false negative errors is calculated by dividing the number of stress events the FSI fails to capture by the total of stress events we have chosen. Again, we use our 0/1-vector to verify whether there was a financial stress event. We define the FSI to capture a downturn if the FSI signals a high level of stress within a timespan of 20 days before the downturn occurs.

2.4.1 Heterogeneous Autoregressive Model

Besides these more qualitative approaches, we will model our FSIs according to the Heterogeneous Autoregressive (HAR) model. The HAR model was introduced by Corsi (2009) to model the realized volatility of foreign exchange indexes. He created a simple model with a long memory in order to use as much information from the past as possible to forecast the future foreign exchange indexes. To include information from the past month and week, he averaged the realized variance from the corresponding time period and included them as one variable. We can use a similar approach for our FSIs. Although this model is relatively simple and does not formally belong to the class of long memory models, the model is able to capture a lot of information from the past by only using three variables. For simplicity, we assume there are 20 trading days in one month. Instead of including 20 AR terms, we create a model that divides these terms and averages them to a daily, weekly or monthly variable.

$$\mathrm{FSI}_t^{(m)} = \frac{1}{15}(\mathrm{FSI}_{t-6} + \mathrm{FSI}_{t-2} + \ldots + \mathrm{FSI}_{t-20}). \tag{10}$$

In the same way we also create a variable that represents the FSI of the past trading week.

$$\mathrm{FSI}_t^{(w)} = \frac{1}{4}(\mathrm{FSI}_{t-2} + \mathrm{FSI}_{t-2} + \ldots + \mathrm{FSI}_{t-5}).^1 \tag{11}$$

Using these two variables and a simple AR(1) term we now have a HAR model with the lagged daily, weekly and monthly value of the FSI as explanatory variables.

$$\mathrm{FSI}_t = \alpha + \beta_1 \mathrm{FSI}_{t-1}^{(d)} + \beta_2 \mathrm{FSI}_{t-1}^{(w)} + \beta_3 \mathrm{FSI}_{t-1}^{(m)} + \epsilon_t. \tag{12}$$

The values of the parameters α, β_1, β_2 and β_3 can be estimated using OLS.

Promising as the HAR model sounds, it has already been included by de Wilde et al. (2013) in their research for an optimal model. It proved to outperform a vector autoregressive (VAR) model, and tied with the autoregressive moving average (ARMA) model for one-step-ahead forecasting. As the HAR model has a long-term memory and the ARMA model does not, the HAR model was preferred above the ARMA model. What has been forgotten, however, is to evaluate the HAR model to a simple AR(1) model: are the included weekly and monthly terms really improving the forecasts? This is why the HAR model will be evaluated now by comparing its results to a simple AR(1) model.

2.5 Forecasting

To compare the two different models, we will make two forecasts of different time periods. First, we will forecast the last year of our data set, 2012. This is to understand how our two models

[1] Note that the definition of this HAR model differs of Corsi (2009), where the daily and weekly realized volatilities were added to the weekly and monthly term, respectively. If we would do the same, we increase the correlation between the terms, putting more loading on the recent observations and negatively influencing β_2 and β_3.

are forecasting under relative stress-free periods: according to our own 0/1-vector, there were no relevant stress events within this time period. The second forecast will be made for the year 2008, when stress was high. It is interesting to see how these forecasts will perform.

The forecasts that will be made are one-step-ahead, that is,

$$\widehat{\text{FSI}}_{t+1}^{HAR} = \alpha + \beta_1 \text{FSI}_t^{(d)} + \beta_2 \text{FSI}_t^{(w)} + \beta_3 \text{FSI}_t^{(m)} \tag{13}$$

$$\widehat{\text{FSI}}_{t+1}^{AR} = \alpha + \beta \text{FSI}_t. \tag{14}$$

To evaluate these forecasts, we compute the root-mean-squared errors (RMSE) of the models

$$\text{RMSE(h)} = \sqrt{\frac{1}{P-h+1} \sum_{t=0}^{P-h} e^2_{N+h+i|N+i}}, \tag{15}$$

where P are the realizations of the standard errors e for $t = N+h, ..., N+h+P-1$ where N represents our sample period out of our T observations and h the number of steps-ahead.

We use the Diebold-Mariano statistic of Diebold and Mariano (2002) to test for a significant difference between the different models. The null hypothesis to be tested is that the RMSEs are equal, by comparing the difference of RMSE of two models

$$d_t = e^2_{A,t|t-h} - e^2_{B,t|t-h}, \tag{16}$$

where $e^2_{i,t|t-h}$ with $i = A, B$ are the RMSE of the different models. Therefore, the null hypothesis becomes $h_0 : d_t = 0$. The sample mean loss differential $d = \frac{1}{p} \sum_{t=0}^{P-1} d_{N+h+i}$ divided by its sample standard deviation follows an asymptotically standard normal distribution, given a sequence of P realizations d_t. The Diebold-Mariano test-statistic is

$$\text{DM} = \frac{\bar{d}}{\hat{\sigma}_{d_t}/N}, \tag{17}$$

where $\hat{\sigma}_{d_t}$ is the variance of d_t. We can therefore reject h_0 on a 5% confidence level if DM > 1.96.

The forecasts will be one-step-ahead as we include the AR(1) model: the forecasts will yield poor results if we include higher steps-ahead forecasts. We do realize that we are trying to forecast financial stress and therefore, it seems strange to forecast only one-step-ahead. This is why we will pick the 'winning' FSIs and will re-model these into monthly averaged FSIs; that is, we monthly-average our observations, and perform the same techniques on these new observations as we did for our daily observations. We will not do this initially, as we think it is more accurate to use the daily observations instead of monthly-averaged variables. However, we do want to know if our 'winning' FSI will yield good results as well for this specific data.

3 Data

The variables which were provided to us are listed in Table 1, and span a period of 1999 up to 2012. Based on Kliesen and Smith (2006), Illing and Liu (2006) and Brave and Butters (2012) we construct extra spreads which are also included in Table 1. Most of these spreads are self-explanatory, thus we will highlight only a few spreads. For instance, we estimate the volatility of the financial market by regressing with VIX on SPY, where VIX is an index tracking the implied volatility of S&P500 options and SPY is an index based on the 'confidence' in S&P500 options, and then taking the residuals as (unexplained) volatility of the financial market, denoted as 'financial σ'. We also construct a banking's β as follows: $\beta = \frac{\text{cov(XLF,SPY)}}{\text{var(SPY)}}$, where XLF indicates the confidence in the financial market, specifically the financial assets of S&P500. With these new variables, we choose to exclude variables in our analysis to prevent a leverage on variables which were already chosen in these spreads, such as stand-alone AAA. This leads to a total of fifteen indicators.

Indicators	Explanation in case of increase
High Yield ('risky bonds')	Compensation for risk increases, which indicates less secure return
LIBOR	Compensation for risk of interbank loans
DFF (Daily Federal Funds rate)	Compensation for risk for overnight loans for depository institutions
VIX: implied volatility (of S&P500 options)	Indicates the market's expectation of stock market. Increase shows more stress
EUDex (€/$ exchange rate)	The higher the rate, the less valuable the dollar, indicates a decrease of the U.S. economy
2Y, 10Y, 30Y (2, 10, 30 year government bonds)	Increase indicates compensation for increased risk on government bonds
Yield curve: 10y - 3M treasury bill	Indicates that the short term risk decreases
AAA-10Y	Indicates that the return on AAA (rated by Standard and Poor's) bonds increases
Baa-10Y	Indicates that the return on Baa (rated by Moody's) increases
HY-10Y	Indicates that the return on high yield bond increases
TED-spread (LIBOR - yield on treasury bills)	Indicates that the interbank rate increases, compensates for higher risk
Banking's β	Returns of the banking sector are more volatile, than the return on the overall market
Financial σ	Indicates the volatility of the financial market

Table 1: List of indicators. Variables were provided to us, and are originally from the website of the Federal Reserve. Note that the spreads starting from 'yield curve' were constructed ourselves.

3.1 Data modification

Our data has to be polished some more before we can use it for constructing our indexes. First, we have some observations missing due to operational failure. The longest 'missing' period was 4 trading days. We assume values for these days by simply interpolating the values around the observation(s). Note that we therefore assume these days to be 'eventless'. Second, we correct our data to become stationary. Stationarity is used in time series analysis, to confirm that within a period of time, the mean and/or variance does not change. We test for stationarity by employing the Augmented Dicky-Fuller (ADF) test. The ADF test tests the null hypothesis of $\phi = 1$, which is

$$y_t = \alpha + y_{t-1} + \epsilon_t, \tag{18}$$

against the alternative of $\phi < 1$, where ϕ is the coefficient of y_{t-1}, in the following regression

$$y_t = \alpha + \delta t + \phi y_{t-1} + \epsilon_t. \tag{19}$$

Here y_t denotes variable X_j of the dataset, and t the trend component. If the null hypothesis is rejected, we can assume no unit root is present. If the null hypothesis is not rejected, a unit root is present and we have to adjust the variable for its non-stationary characteristic. This is done by taking the first difference ΔX_j. If we again find a unit root, we take the second differences.

We find that taking the first differences is an appropriate measure for stationarity regarding our data set, that is, all indicators are stationary after taking the first differences.

Last, we standardize our data in order to properly compare it with each other. We keep in mind that, as we standardize our data for a time series, we want to get a clear overview of how the indicators were evaluated at the respective time. Therefore, we employ the same method as de Wilde et al. (2013) by standardizing our data recursively. The idea is to standardize observation t based on the mean and standard deviation of observations $[1, t]$. As this is done for every observation, we are standardizing in a recursive manner. The advantage of standardizing recursively is that we account for all historical data and leave out the, though available, future data.

3.2 Financial stress events

In order to create some of our FSIs, and to evaluate whether our FSIs capture financial stress, we make a list of all relevant financial stress events between the years 1999 and 2012. We created our list of events with the help of the website http://www.dof.ca.gov, which contain all financial, political and natural developments which have influenced California's economic indicators. We believe that most of the financial events within the state California is a good benchmark for the entire U.S. financial market. The list of these events has been provided in the appendix, with a graphical representation of the distribution of the events. Unsurprisingly, there are a lot of stress events in the last six years. With the help of these financial stress events, we construct the 0/1-vector as explained earlier. This vector is created as follows: we define there was stress in the market 3 days before the event, and 3 days after the event. Based on the total of our events (35), and the total of trading days between 1999 and 2012, we have stress events in about 5.73% of our observations. This is low, but we have many observations, and it would be worrying for the U.S. economy if we found a significant higher percentage.

3.3 Construction of indexes

As we have now taken every step needed, we move on to the construction of the various indexes.

3.3.1 FSI based on weighted sum of financial markets

For the FSI based on a weighted sum of financial markets, the WSRFSI, we allocate our variables to four specific markets: the money market, bond market, debt/equity market and the banking sector. Variables are allocated to the different markets based on the appendix of Brave and Butters (2012), which was also done by de Wilde et al. (2013). This results in Table 2.

Bond market	**25%**
HY	6.25%
Baa-10Y	6.25%
HY-10Y	6.25%
AAA-10Y	6.25%
Equity market	**25%**
VIX	25.00%
Money market	**25%**
LIBOR	3.13%
DFF	3.13%
EUDex	3.13%
2Y	3.13%
10Y	3.13%
30Y	3.13%
10Y-3M	3.13%
TED	3.13%
Banking sector	**25%**
Banking β	12.50%
Financial σ	12.50%

Table 2: These are the weights of the different indicators, respective to their market. Note that all weights sum up to 100%.

3.3.2 FSI based on PCA

For the FSI based on regular PCA, PCAFSI, we calculate the principal components and note that we exceed our threshold of 60% when we include 5 principal components. The FSI based on recursive PCA, PCRFSI, exceeds the threshold of 60% when we include 4 principal components. As for our FSI based on supervised PCA, PSCFSI, it is interesting to note which variables are statistically insignificant on a 5% level according to the created 0/1-vector. Five variables turn out to be insignificant: EUDex, 10Y, 30Y, 10Y-3M and AAA-10Y. We see that four of these variables are classified in our previous FSI as money market, and the other one as bond market. As both these markets are represented by many variables, we still have our hopes that we may find some more accurate results with this FSI. We exceed the threshold of 60% when we include 3 out of the 10 principal components.

3.3.3 FSI based on clustering

For our FSI based on cluster analysis, we construct a total of four potential FSIs: dissimilarity based on Euclidian distances or Hoeffding's D, and clustering according to k-means or the PAM algorithm. The resulting dissimilarity based on Hoeffding's D is presented in Table 3.

	HY	LIBOR	DFF	Vix	EUDex	2Y	10Y	30Y	10Y-3M	Baa-10Y	HY-10Y	TED	Banking β	AAA-10y	Financial σ
HY	1														
LIBOR	0.001195	1													
DFF	0.000562	0.002172	1												
Vix	0.008519	0.018033	0.001703	1											
EUDex	0.002684	0.000204	0.000144	0.000798	1										
2Y	0.001396	0.005325	0.002155	0.003695	0.005804	1									
10Y	0.004278	0.005603	0.009645	0.049164	0.000509	0.0029	1								
30Y	0.000605	0.004	0.0025	0.013846	0.0003	0.002322	0.166259	1							
10Y-3M	0.001347	0.000261	0.00043	0.002564	0.000231	0.132409	9.88E-05	0.000164	1						
Baa-10Y	0.004915	0.000846	0.000209	0.007807	0.001402	0.098789	0.001092	5.25E-05	0.037895	1					
HY-10Y	0.106348	0.000557	0.000456	0.008856	0.000333	0.148458	0.001247	5.28E-05	0.146491	0.078388	1				
TED	0.002608	0.017592	0.005127	0.009607	0.000347	0.002248	0.026531	0.017467	0.001636	0.000941	0.001247	1			
Banking β	0.000896	0.008215	0.00697	0.032423	2.80E-06	0.001633	0.045276	0.018784	0.000157	0.000803	0.000162	0.029255	1		
AAA-10y	0.00864	0.001296	0.000754	0.007249	0.000266	0.011385	0.001162	0.000154	0.004857	0.006459	0.011428	0.005556	0.001067	1	
Financial σ	0.000801	0.001456	0.000306	0.008011	0.003257	0.088905	0.001582	0.000223	0.030564	0.281557	0.043457	0.000671	0.001035	0.004238	1

Table 3: Values of Hoeffding's D, with $D = 30\frac{(n-2)(n-3)A+B-2(n-2)C}{n(n-1)(n-2)(n-3)(n-4)}$.

As we have already stated before, the interval of the dissimilarities lies between -0.5 and 1. It can be clearly seen however, that nearly all values are close to zero. This will likely result in poor clustering results, as there is hardly any distinction within the variables. We add, for the critical reader, that the problem does not appear to be a fault of programming; the algorithm of Hoeffding works fine for other data. Then, how can we explain these values? We presume that the problem lies within the ranking of the variables, and that we are ranking a time series here. This could mean that stress, which steadily grows as we have daily data, may not be captured as well as there are likely points somewhere in time that lie between the 'growth' of stress. It is likely that these points in time do not correlate with other variables. Ergo, the pattern of the stress may be very hard to define, and so the ranking of the variables may be somewhat arbitrary. This leads to smaller values for A,B and C, resulting in small values for D.

We list the outcomes of the silhouette plot in Table 4. The elbow plots are listed in the appendix, as these are similar to the silhouette results.

Nr. of clusters	Hoeff, PAM	Hoeff, kmeans	Euclid, PAM	Euclid, kmeans
2	0.016	-0.055	0.386	0.206
3	0.031	0.063	0.228	0.184
4	0.090	0.030	0.302	0.337
5	0.077	-0.004	0.400	0.268
6	0.115	0.030	0.352	0.332
7	0.196	0.126	0.417	0.465
8	0.297	0.291	0.445	0.536
9	0.387	0.362	0.570	0.518
10	0.582	0.466	0.641	0.689
11	0.623	0.536	0.668	0.649
12	0.694	0.530	0.719	0.764
13	0.917	0.917	0.824	0.892
14	0.945	0.972	0.945	0.972

Table 4: These are the values of the silhouette plot. Of course, the values of a high number of clusters are desirable, but do not make much sense if we want to truly cluster the data into groups.

It is clear that the dissimilarity based on Hoeffding is outperformed by Euclidian distances. The difference between PAM and k-means differs on how many clusters you choose to incorporate. We therefore decide to choose both methods as stress indexes: a PAMFSI and a KMEFSI. Note that the more clusters you choose, the better the data is clustered (with some small exceptions). By choosing the amount of clusters, we keep in mind the rule of thumb: the amount of clusters is more or less equal to the square root of the halved amount of variables. We also consider the respective increase of adding a new cluster. For instance, if we look at the third column (Euclid, PAM), we see the increase in the silhouette plot from the sixth cluster to the seventh cluster is equal to $0.417 - 0.352 = 0.065$, where the increase of the seventh to the eight cluster is equal to $0.445 - 0.417 = 0.028$. As this increase is lower, we think that seven clusters would be more appropriate than eight. Keeping in mind the rule of thumb and this method, we choose to include five clusters for the PAMFSI, and four clusters for the KMEFSI.

As we now have chosen our clusters, it may be interesting to see which variables tend to follow a similar pattern. The respective clusters of the PAMFSI and KMEFSI are listed in Table 5.

PAMFSI		**KMEFSI**	
#1	**20%**	**#1**	**25,00%**
HY	2.50%	LIBOR	3.57%
LIBOR	2.50%	DFF	3.57%
DFF	2.50%	EUDex	3.57%
EUDex	2.50%	2Y	3.57%
2Y	2.50%	10Y-3M	3.57%
10Y-3M	2.50%	Banking's β	3.57%
Banking's β	2.50%	AAA-10y	3.57%
AAA-10Y	2.50%		
#2	**20%**	**#2**	**25%**
VIX	20.00%	VIX	12.50%
		TED	12.50%
#3	**20%**	**#3**	**25%**
10Y	10.00%	10Y	12.50%
30Y	10.00%	30Y	12.50%
#4	**20%**	**#4**	**25%**
Baa - 10Y	6.66%	Baa-10Y	6.25%
HY-10Y	6.66%	HY-10Y	6.25%
Financial σ	6.66%	Financial σ	6.25%
		HY	6.25%
#5	**20%**		
TED	20.00%		

Table 5: Different weights of the KMEFSI and PAMFSI. Note that all weights sum up to one.

We note that the respective clusters do not seem to differ that much. There are, however, some small differences, where the prominent roles of VIX and TED in the PAMFSI strike out the most, as they account for 40% of the loading when added together on the PAMFSI. The KMEFSI seems to be divided more equally than the PAMFSI.

3.3.4 FSI based on logistic regression

The final FSI, LOGFSI, is constructed by using a logit regression. We present its respective loadings in Table 6, following the logistic regression.

	LOGFSI
HY	0.0775
LIBOR	0.1495
DFF	-0.0540
Vix	0.3903
EUDex	0.0134
2Y	0.0034
10Y	-0.0813
30Y	0.1881
10Y-3M	0.0337
Baa-10Y	-0.0871
HY-10Y	0.0957
TED spread	0.0987
Banking beta	-0.0134
AAA-10y	0.0445
Financial sigma	0.1410

Table 6: Coefficients of the logistic regression on the 0/1-vector created out of the list of financial events.

Two aspects are noteworthy, the first being the negative loadings. We emphasize that these are coefficients based on the 0/1-vector, where some variables may fail to capture the stress on the markets with respect to the other financial variables. There may also be some other cross-correlation effects which result in negative loadings. Second, we see that the VIX variable has an enormous loading compared to the other variables. Taking into account the respective loadings of the KMEFSI and PAMFSI, it seems VIX is an excellent indicator of financial stress, and follows its 'own' pattern.

4 Results

First, we evaluate our constructed FSIs qualitatively by plotting their Z-scores in Figure 1. We do this to make a good comparison possible, as now all FSIs are more or less within the same boundaries.

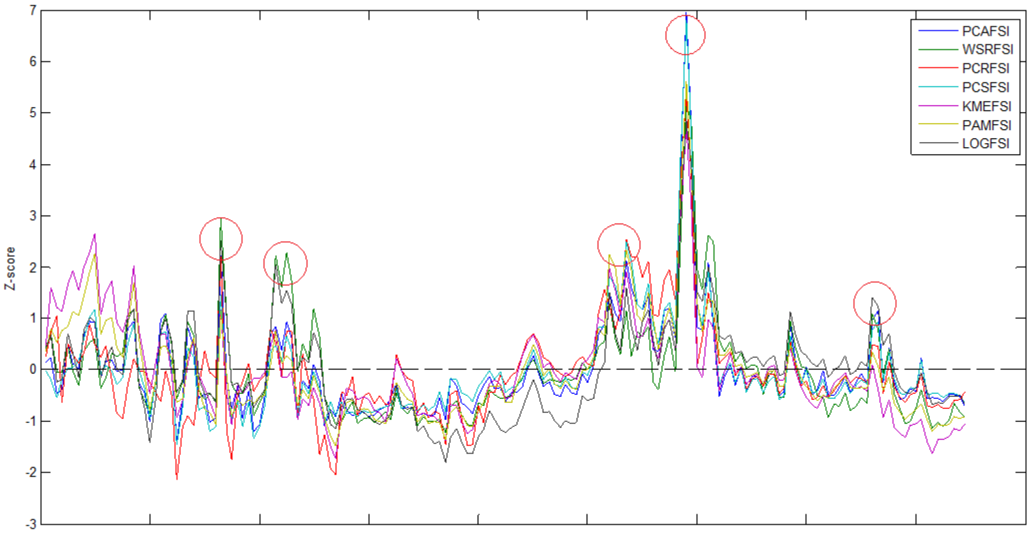

Figure 1: Graphic representation of FSI. We see that important financial stress events are represented in the FSI designs we chose.

Here we see the FSIs plotted over the time period 1999-2012. The highlighted events represent the stress of the terrorist attack of 9/11, the indictment of Arthur Anderson (15/07/2002), the drop in house prices (26/07/2007), the disturbance with Merrill Lynch and Lehman Brothers (19/09/2008) and the downgrade of the U.S. credit rating (04/08/2011). We see that all FSIs follow a similar pattern.

For all FSIs we calculate how many times they did not signal a financial stress event when there was a financial stress event (false negative error) and how many times they signal a financial stress event while there was no such thing (false positive error). The results are presented in Table 7.

	WSRFSI	**PCAFSI**	**PCRFSI**	**PCSFSI**	**KMEFSI**	**PAMFSI**	**LOGFSI**
False positive	65.93%	64.27%	62.50%	62.70%	67.79%	65.58%	66.67%
False negative	20.00%	8.57%	5.71%	8.57%	22.86%	20.00%	20.00%

Table 7: False positive and false negative errors of the FSIs. We add that all our FSI designs had trouble predicting all the financial events we listed, which can be found in our appendix.

Unsurprisingly, the FSIs based on PCA perform the best for this in-sample evaluation, PCRFSI generating the best results. The high amount of false positive errors is not worrying, as there can be stress in financial markets, but this does not always lead to a recession or downturn. In the next sections we will investigate the predictive power of the FSIs.

4.1 Forecast results

We present the results of the forecast errors in Table 8. The plotted figures of all FSIs, including their respective predictions in 2008 and 2012, can be found in the appendix.

2012	**WSRFSI**	**PCAFSI**	**PCRFSI**	**PCSFSI**	**KMEFSI**	**PAMFSI**	**LOGFSI**
RMSE AR	0.212	0.346	0.331	0.447	0.160	0.149	0.173
RMSE HAR	0.169	0.323	0.235	0.413	0.115	0.118	0.152
DM	5.601	2.984	8.000	3.073	6.978	5.419	4.278

Table 8: Forecasts are made with the formulas $\widehat{\text{FSI}}_{t+1}^{HAR} = \alpha + \beta_1 \text{FSI}_t^{(d)} + \beta_2 \text{FSI}_t^{(w)} + \beta_3 \text{FSI}_t^{(m)}$ and $\widehat{\text{FSI}}_{t+1}^{AR} = \alpha + \beta \text{FSI}_t$. We then calculated root-mean-square error values (RMSE(h) $= \sqrt{\frac{1}{P-h+1}\sum_{t=0}^{P-h} e^2_{N+h+i|N+i}}$, where $e = \widehat{\text{FSI}} - \text{FSI}$) and Diebold-Mariano statistic (DM $= \frac{\bar{d}}{\hat{\sigma}_{d_t}/N}$) of the year 2012.

First, these are the predictions of the forecast of the year 2012. We see that the RMSE is relatively small per FSI. This is due to the fact that there were not many shocks in this year, as we have no stress events defined in our list of events. Interestingly, we see that the PCAFSI results in smaller RMSEs than the PCSFSI. Maybe it was not such a good idea to exclude variables after all. The HAR model seems to have a smaller RMSE, and with the Diebold-Mariano statistic we conclude that all prediction errors are statistically significant from the AR model, as they all exceed the critical value of 1.96. Therefore, the HAR model would be the right choice. Note that the RMSEs of the KMEFSI and PAMFSI are substantially lower than WSRFSI and LOGFSI, whereas the RMSEs of PCAFSI, PCRFSI and PCSFSI are the highest. We can make some sort of 'distinction' between FSI based on clustering, FSI based on weighted variables and FSI based on PCA.

2008	**WSRFSI**	**PCAFSI**	**PCRFSI**	**PCSFSI**	**KMEFSI**	**PAMFSI**	**LOGFSI**
RMSE AR	0.617	1.464	1.549	1.695	0.508	0.490	0.670
RMSE HAR	0.564	1.364	1.453	1.578	0.474	0.455	0.658
DM	1.586	1.315	1.375	1.109	0.873	1.050	0.282

Table 9: Forecasts are made with the formulas $\widehat{\text{FSI}}_{t+1}^{HAR} = \alpha + \beta_1 \text{FSI}_t^{(d)} + \beta_2 \text{FSI}_t^{(w)} + \beta_3 \text{FSI}_t^{(m)}$ and $\widehat{\text{FSI}}_{t+1}^{AR} = \alpha + \beta \text{FSI}_t$. We then calculated root-mean-square error values (RMSE(h) $= \sqrt{\frac{1}{P-h+1}\sum_{t=0}^{P-h} e^2_{N+h+i|N+i}}$, where $e = \widehat{\text{FSI}} - \text{FSI}$) and Diebold-Mariano statistic (DM $= \frac{\bar{d}}{\hat{\sigma}_{d_t}/N}$) of the year 2008.

Second, we also construct forecasts for the year 2008. We see that all RMSEs have become higher, which is what we expected. Note that, in terms of RMSE, the PCAFSI now seems to outperform both the PCRFSI and PCSFSI. Again, the KMEFSI and PAMFSI have the lowest RMSE. However, we note that it is not possible to distinguish the AR model and the HAR model, based on the Diebold-Mariano statistic. This is interesting, as we would not expect a difference with the 2012 forecast. The difference is caused by the increased variance of d: the variance was very small in 2012, leading to statistically significant results, whereas the variance in 2008 is now higher.

Now, we should choose our own FSI. We stress that all FSIs are good candidates, as can be seen in our qualitative plot of the Z-scores of the FSIs. Based on the in-sample results, all FSIs based on PCA have the upper-hand. However, as we want to forecast stress, and particularly the

FSIs based on PCA perform worse, these FSIs are not what we are interested in. Based on RMSE, we choose the two FSIs based on clustering, the PAMFSI and KMEFSI, to be the 'winning' FSIs.

As we mentioned before, we try to forecast stress on the long term. Therefore, one-step-ahead forecasts based on daily observations may not make much sense if we truly want to predict financial crises. We monthly-average all our original data, and construct the KMEFSI and PAMFSI. Interestingly, the original amount of clusters (four) of KMEFSI stays the same: even the allocation of variables is the same. The PAMFSI now also has four clusters as this yields the best silhouette values. In Figure 2 we find the plotted Z-scores.

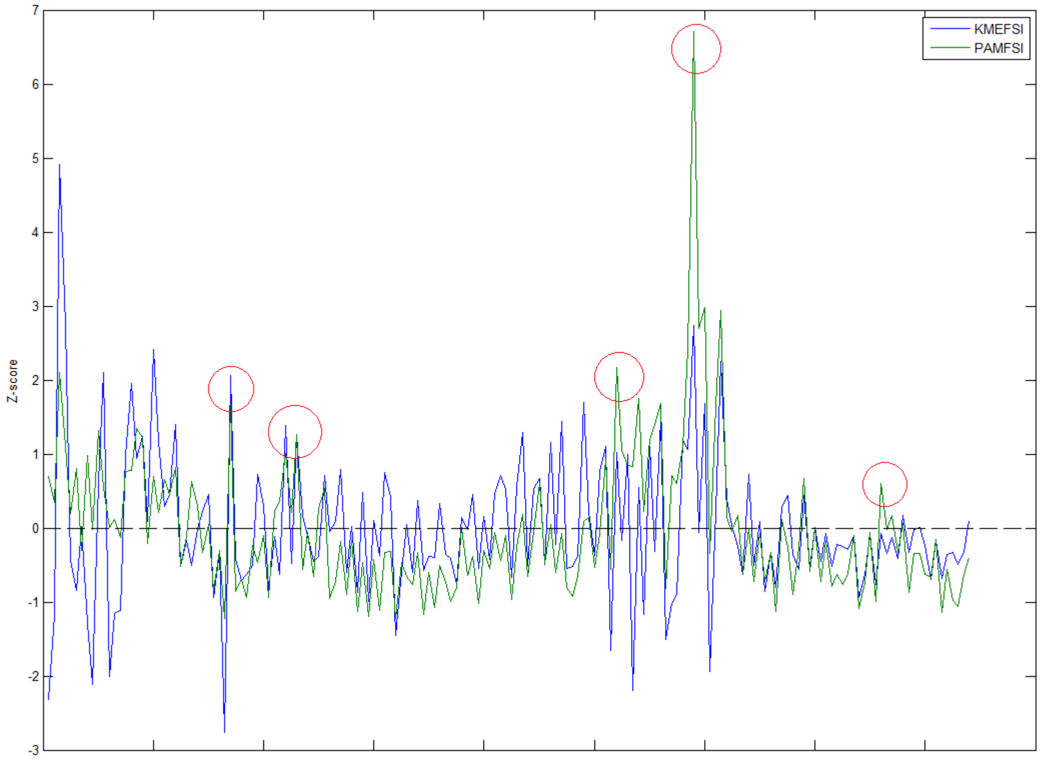

Figure 2: Graphic representation of monthly-averaged FSIs

For comparison reasons, we have highlighted the same events as earlier in this section. We see that the KMEFSI fails to capture the last three events, where the PAMFSI does capture these events. Overall, it seems that the pattern of the PAMFSI has not changed that much compared to the daily observations.

We can now choose our definite 'winning' FSI. The clustering algorithm of PAM seems to give us a good financial stress indicator, as both daily and monthly-averaged observations follow the same pattern, including important stress events of recent years. Forecasts have been made with the use of daily observed data, where the PAMFSI has a low RMSE compared to the other FSIs. Therefore, it seems appropriate to choose the PAMFSI as the best financial stress index.

5 Conclusion

This paper has tried to index financial stress in numerous ways, continuing the work of de Wilde et al. (2013). We extend their research by introducing new techniques to index the stress, including new PCA techniques and clustering algorithms. The modeling itself is less prioritized, as we are mainly interested in indexing the financial stress.

All constructed FSIs follow a similar pattern, and capture relevant financial events. Based on in-sample evaluation, the FSIs based on PCA outperform the other FSIs; however, out-of-sample evaluation concludes that FSIs based on PCA are outperformed by the other FSIs. We find that the FSIs based on a clustering algorithm have a lower prediction error if we look at both the AR and HAR model. As we want to forecast stress on the long term, we construct the clustering FSIs based on monthly-averaged data. As it turns out, the FSI based on the PAM clustering algorithm follows a similar pattern as the daily observations, whereas the FSI based on k-means fails to capture important events.

Further research may focus on the different clustering algorithms. We chose to test only two alternatives against more traditional clustering algorithms, but it may be worthwhile to verify whether other clustering algorithms may even have better results. Of course, the same can be done for the modeling: there are many more models available besides the AR and HAR model. We also may want to exclude some variables we used in our analysis: a good example is the PCSFSI, where some variables were excluded as it turned out they were not influenced significantly by financial stress events.

We do think, besides all these further possible directions, that FSIs based on clustering algorithms should generate more attention in the future as there is some interesting work left. We have now evaluated Hoeffding's D against the Euclidian distances, and the PAM algorithm against the k-means algorithm. While the latter two are more common in practice and yield good results, there are many more algorithms for dissimilarities and cluster-optimization left to consider. The Hoeffding's D proved to be weak for our data, but other algorithms such as Pearson's correlation coefficient may be useful, especially since this algorithm depends on the correlation between variables, which may generate better cluster results for a time series than the ranking algorithm of Hoeffding's D. Alternatively, one could try to optimize the clusters with another algorithm than PAM or k-means, such as agglomerative nesting (AGNES) and divisive clustering (DIANA). As the common approaches already generate promising results, clustering algorithms may prove to be an interesting alternative for the future research on financial stress.

A Other tables and figures

Date	Financial Event
13/01/1999	Brazil devalues its currency sending U.S. stocks into a free fall.
21/01/1999	The 1998 trade deficit hit an all-time high of $175 billion, 58 percent more than the shortfall recorded in 1997.
27/07/1999	IMF approves stand-by credit for Russian Federation
01/10/1999	Fed establishes Century Date Change Special Liquidity Facility
03/01/2000	Y2K passes
11/01/2000	NASDAQ peaks above 4000, then begins to sharply decline
11/09/2001	Terrorists attack World Trade Center and the Pentagon. U.S. stock trading halts.
03/12/2001	Enron filed for bankruptcy.
15/07/2002	Arthur Anderson indicted
16/07/2002	The dollar sank against the euro for the first time in more than two years.
22/07/2002	WorldCom filed for bankruptcy protection.
30/07/2002	Sarbanes-Oxley Act passed
12/08/2002	U.S. Airways filed for bankruptcy.
21/02/2007	Rising default rates hitting sub prime mortgage industry
27/02/2007	Dow Jones industrial average down 416 points, biggest one-day point loss since 2001, after declining markets in China and Europe and a steep drop in durable goods orders triggered a massive sell-off on Wall Street.
26/07/2007	The Dow Jones industrial average dropped 311.50 points or 2.3 percent amid concerns about housing and credit markets.
31/07/2007	Bear Sterns liquidates two hedge funds investing in MBS
09/08/2007	The Dow Jones industrial average was down 387.18 points or 2.8 percent as worries about the global credit market sparked a broad sell-off in stocks.
21/01/2008	Global stock markets plunge. Federal funds rate target reduced from 4.25 percent to 3.5 percent, the biggest one-day interest rate reduction on record.
13/03/2008	Gold futures hit $1000 an ounce for the first time. Crude oil price tops $110 a barrel. Gas prices rise to another record high.
18/03/2008	JP Morgan agrees to buy Bear Stearns for a mere fraction of what it was once worth. Federal funds rate target reduced from 3 percent to 2.25 percent.
11/07/2008	Indy Mac Bank seized by federal regulators.
08/09/2008	The U.S. government takes over Fannie Mae and Freddie Mac.
19/09/2008	Merrill Lynch sold to Bank of America. Lehman Brothers files for bankruptcy protection. The Federal Reserve loans $85 billion to American International Group (AIG).
26/09/2008	Washington Mutual Bank failure, largest failure in terms of assets to date
03/10/2008	Emergency Economic Stabilization Act passed (TARP)
06/10/2008	Worst week for the stock market in 75 years. Fed provides $900 billion in short-term cash loans to banks. Fed makes emergency move to lend around $1.3 trillion directly to companies. Federal funds rate target reduced from 2 percent to 1.5 percent. The discount rate was cut to 1.75 percent. The Dow Jones industrial Average caps its worst week ever with its highest volatility day ever recorded in its 112 year history
24/11/2008	Citigroup requires government assistance
01/12/2008	Recession in the US began in December 2007, according to NBER.
16/01/2009	Bank of America requires government assistance
02/03/2009	Dow Jones Industrial Average drops below 7000 for the first time since 1997.
10/05/2010	EU, EGB and IMF announce $1 trillion aid package after Greek debt crisis
04/08/2011	Wall Street suffers worst sell-off in two years. S&P downgrades U.S. credit rating.
22/09/2011	Dow Jones industrials sees biggest two-day decline since December 2008.

Table 10: Total list of events we thought relevant for the U.S. financial market over the past fifteen years.

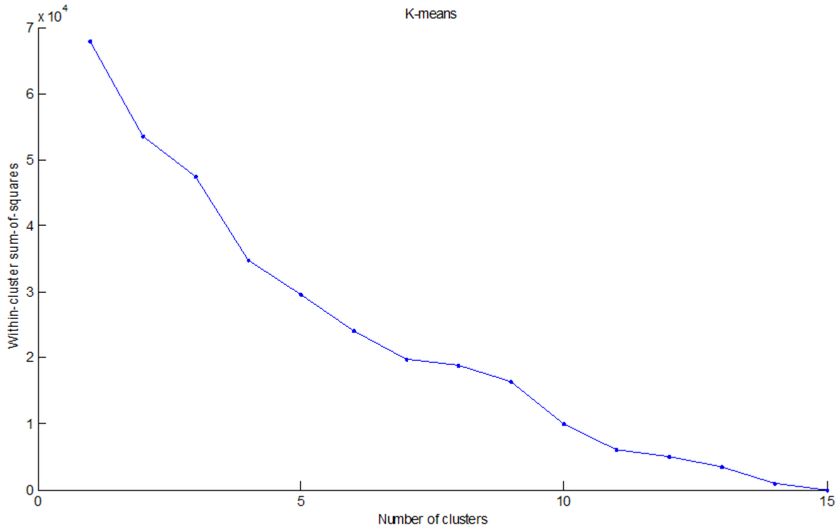

Figure 3: SCREE plot of the clusters based on the k-means clustering algorithm. It is hard to determine an elbow, as the line seems reasonably diagonal. However, keeping in mind the 'rule of thumb', we can see a small knick at the fourth cluster. We therefore choose to incorporate four clusters.

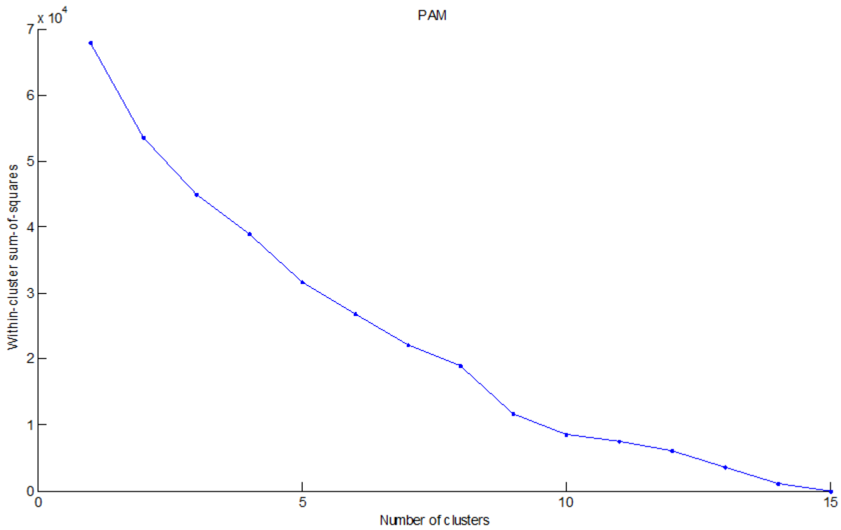

Figure 4: SCREE plot of the clusters based on the PAM clustering algorithm. It is hard to determine an elbow, as the line seems reasonably diagonal. However, keeping in mind the 'rule of thumb', we can see a small knick at the fifth cluster. We therefore choose to incorporate five clusters.

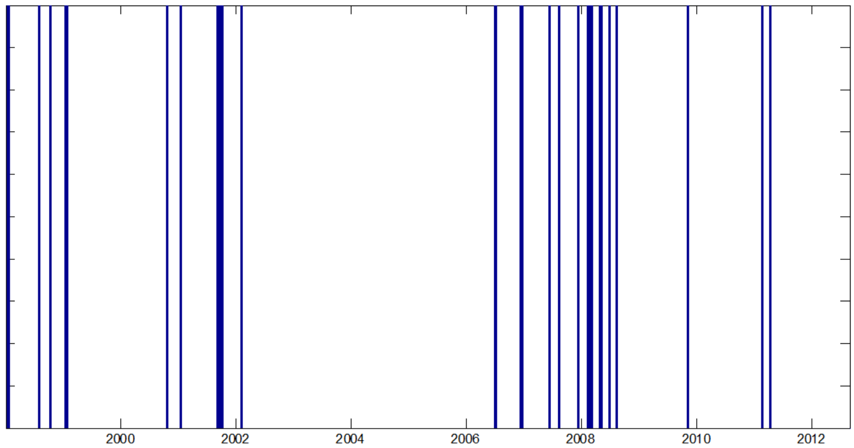

Figure 5: Barplot of the crisis events within our sample period. We see that 2008 was a troublesome year, whereas 2012 was calm.

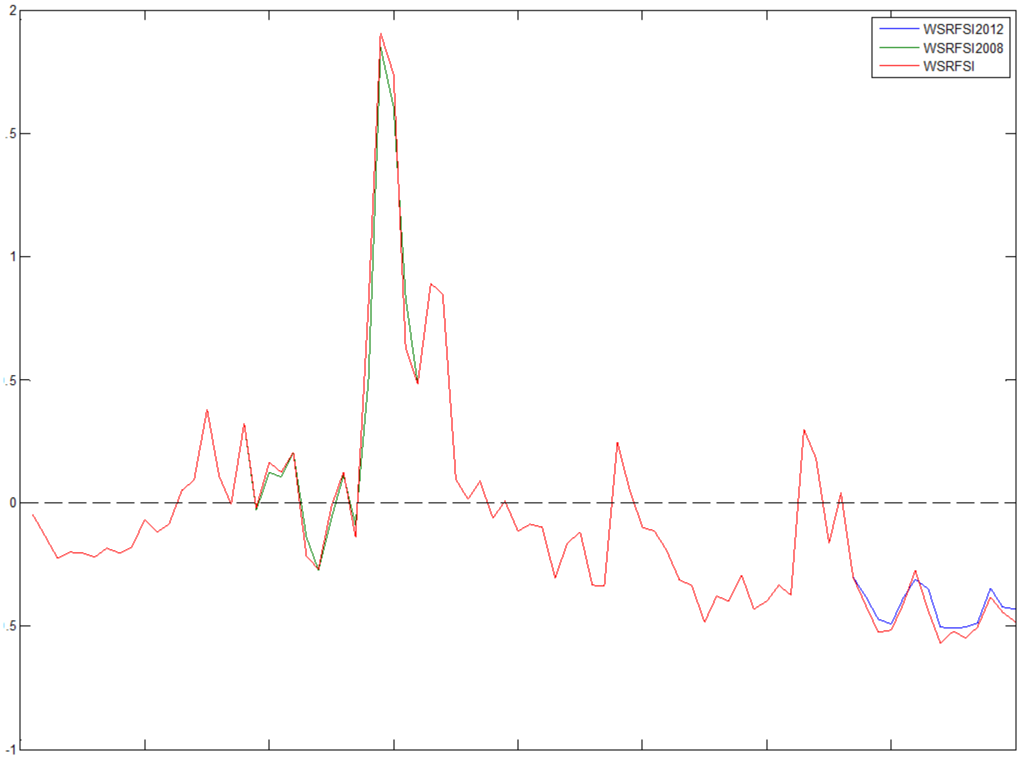

Figure 6: Plot of WSRFSI including forecasts of 2008 and 2012.

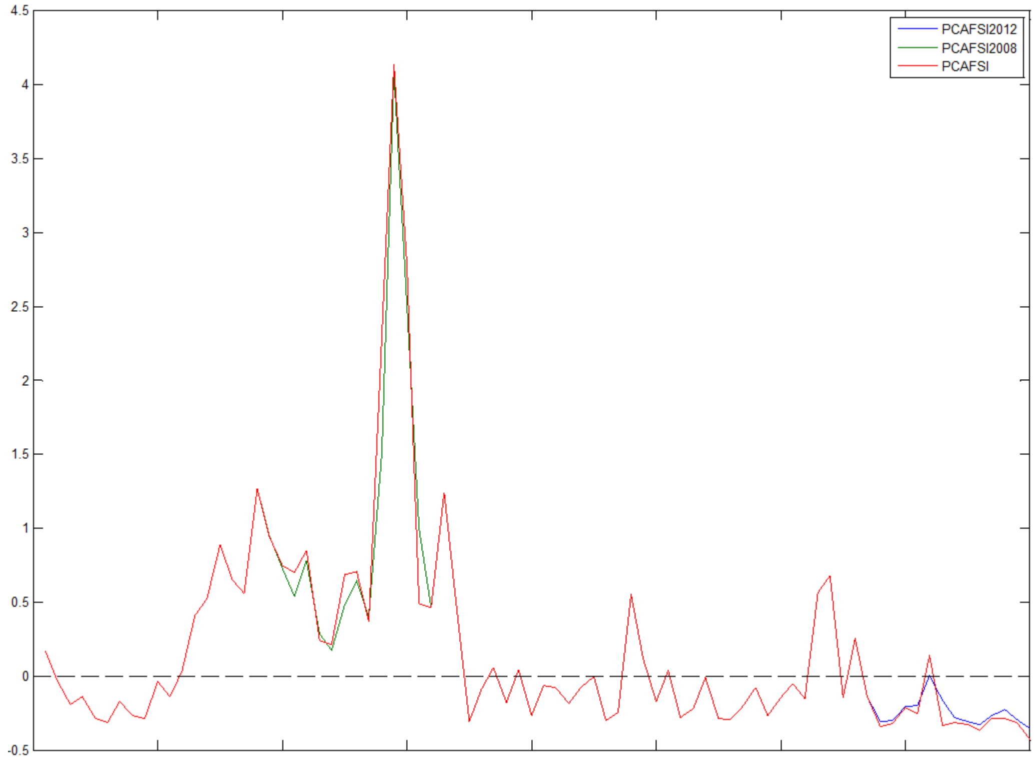

Figure 7: Plot of PCAFSI including forecasts of 2008 and 2012.

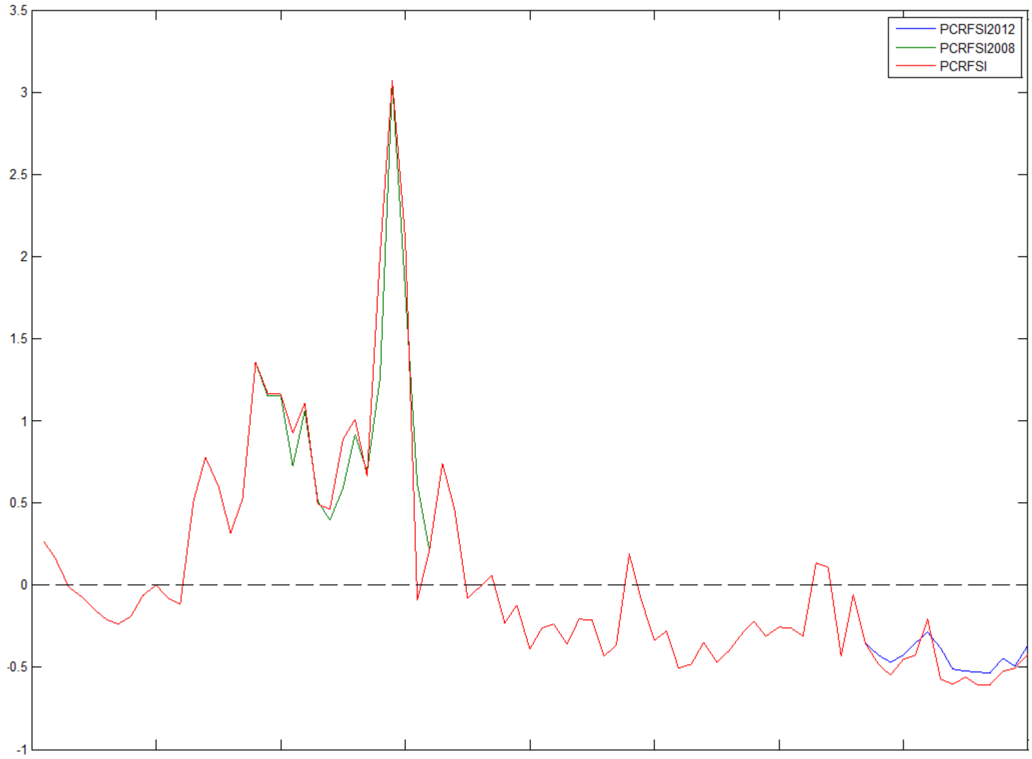

Figure 8: Plot of PCRFSI including forecasts of 2008 and 2012.

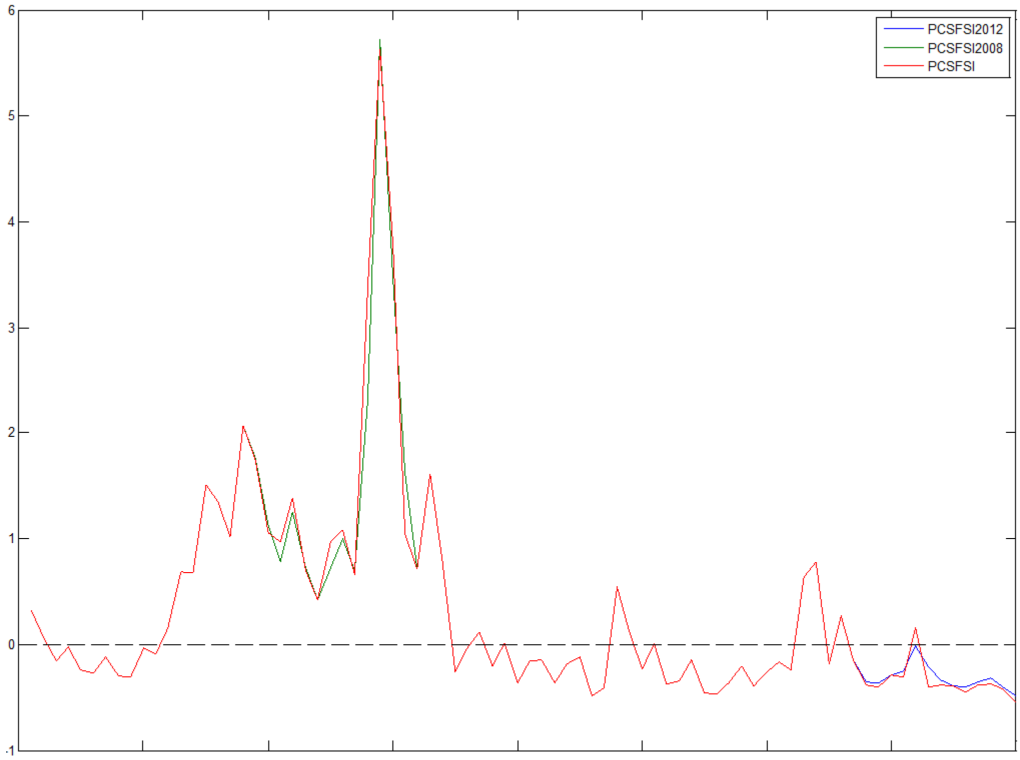

Figure 9: Plot of PCSFSI including forecasts of 2008 and 2012.

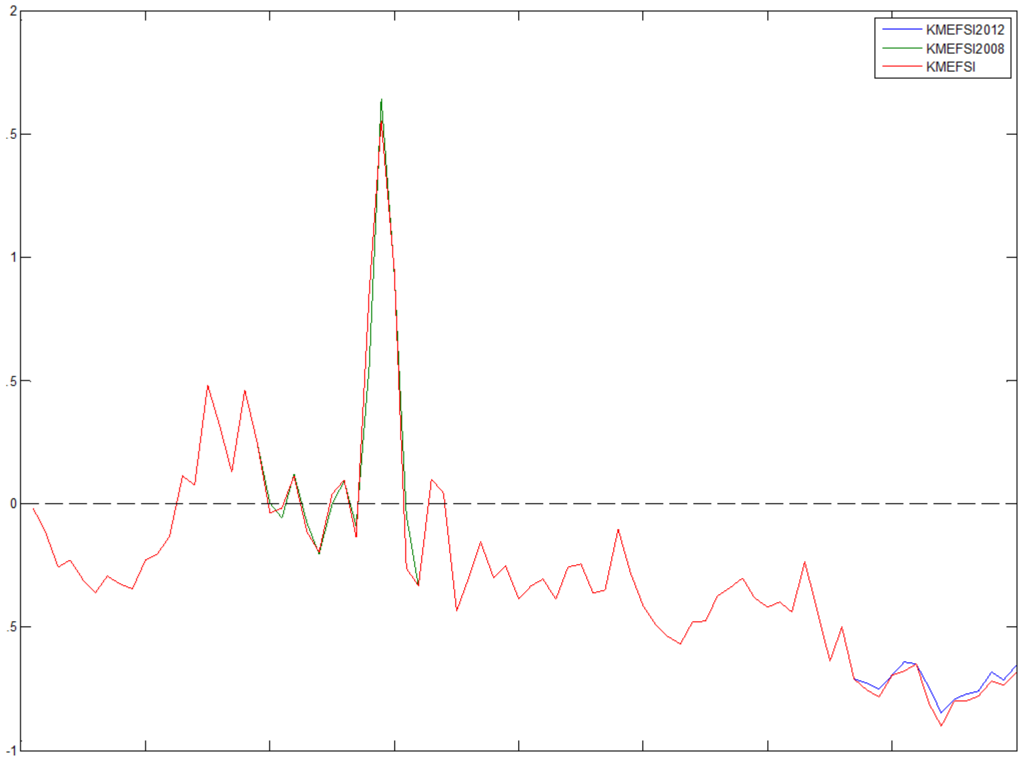

Figure 10: Plot of KMEFSI including forecasts of 2008 and 2012.

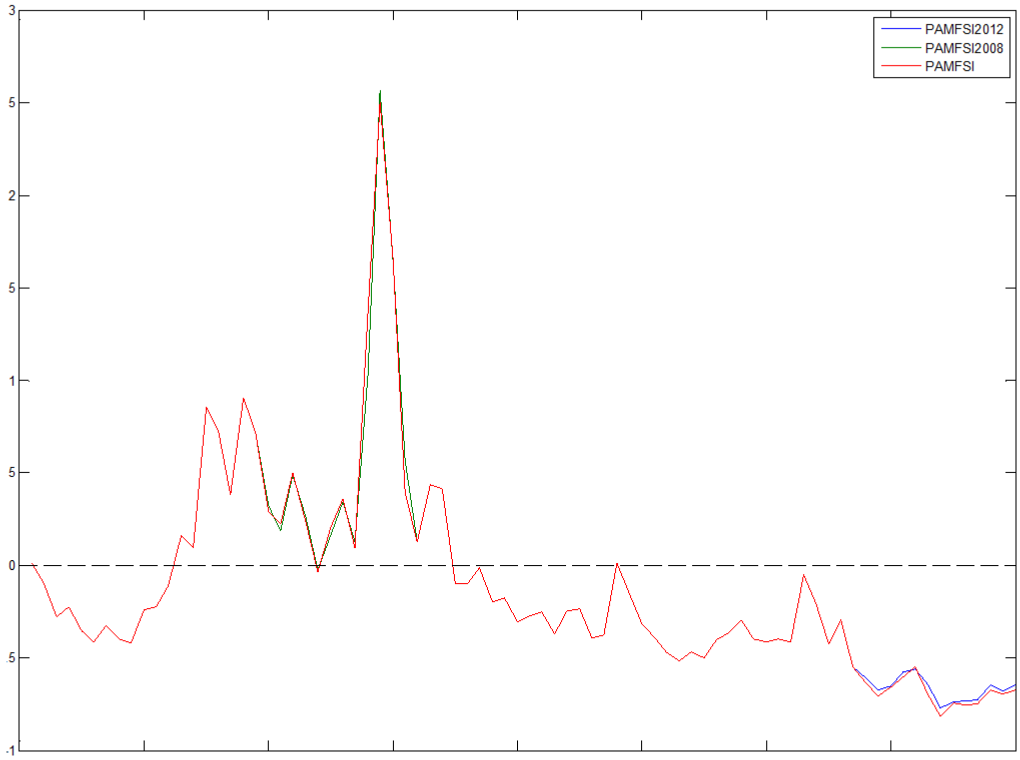

Figure 11: Plot of PAMFSI including forecasts of 2008 and 2012.

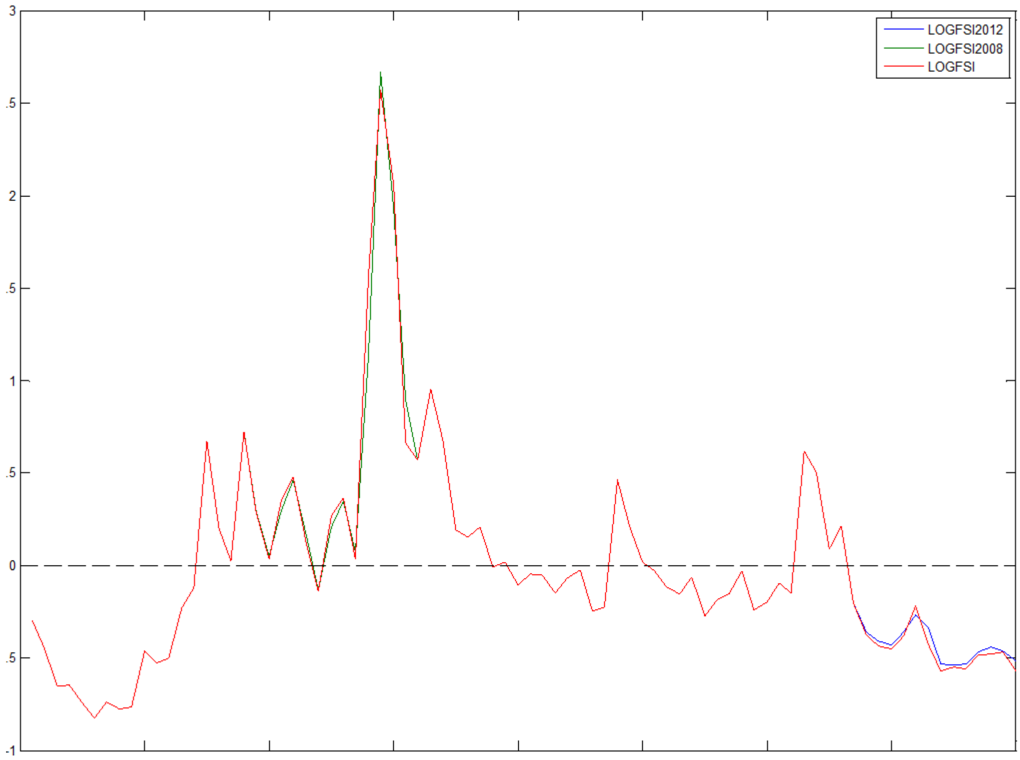

Figure 12: Plot of LOGFSI including forecasts of 2008 and 2012.

References

Eric Bair, Trevor Hastie, Debashis Paul, and Robert Tibshirani. Prediction by supervised principal components. *Journal of the American Statistical Association*, 101(473), 2006.

Scott Brave and A Butters. Diagnosing the financial system: Financial conditions and financial stress. *International Journal of Central Banking*, 8(2):191–239, 2012.

Roberto Cardarelli, Selim Elekdag, and Subir Lall. Financial stress, downturns, and recoveries. *IMF Working Papers*, pages 1–58, 2009.

Fulvio Corsi. A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7(2):174–196, 2009.

Rik de Wilde, Josephine Boerman, Philo Meerman, and Yan Kit Lee. Modeling the financial stress index. Erasmus University Rotterdam, Bachelor project, 2013.

Francis X Diebold and Robert S Mariano. Comparing predictive accuracy. *Journal of Business & economic statistics*, 20(1), 2002.

Deniz Erdogmus, Yadunandana N Rao, Hemanth Peddaneni, Anant Hegde, and Jose C Principe. Recursive principal components analysis using eigenvector matrix perturbation. *EURASIP Journal on Applied Signal Processing*, 2004:2034–2041, 2004.

Wassily Hoeffding. A non-parametric test of independence. *The Annals of Mathematical Statistics*, 19(4):546–557, 1948.

Mark Illing and Ying Liu. *An index of financial stress for Canada*. Bank of Canada, 2003.

Mark Illing and Ying Liu. Measuring financial stress in a developed country: An application to canada. *Journal of Financial Stability*, 2(3):243–265, 2006.

Leonard Kaufman and Peter J Rousseeuw. *Finding groups in data: an introduction to cluster analysis*, volume 344. Wiley-Interscience, 2009.

Kevin L Kliesen and Douglas C Smith. Measuring financial market stress. *Economic Synopses*, 2006.

James MacQueen et al. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1, page 14. California, USA, 1967.

Ana Teresa Yanes Musetti. Clustering methods for financial time series. Master's thesis.

Michael Rosenberg. Financial conditions watch. *New York: Bloomberg*, 2009.

Peter J Rousseeuw. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20:53–65, 1987.