University of Konstanz

# Predicting Economic Time-Series using Dynamic Factor Models under Structural Breaks

*Author:*

Johannes Degn

*Supervisor:*

Ralf Brüggemann

*A thesis submitted in fulfilment of the requirements*

*for the Masters degree at the*

Department of Economics of the University of Konstanz

July 29, 2014

# Contents

# List of Figures

# List of Tables

# Abstract

This thesis applies recently developed methods for finding structural breaks in factor models. Bootstrap versions of the tests are considered which are shown to have improved power and size properties in low dimensional data sets. It is found that removing variables which have breaks in the corresponding factor loadings has a considerable positive effect on the root mean squared forecasting error. This improvement is consistent but less pronounced if noisy predictors are removed beforehand using the method of targeted predictors introduced by Bai and Ng (2008). Structural break tests can thus be considered a promising preselection mechanism which can complement other variable selection methods such as targeted predictors.

# Introduction

Factor models are used in a wide range of situations spanning from economic indicators over prediction of high-dimensional datasets. The basic idea is that there are unobservable factors driving the change of all variables in the data set. Put differently, it is assumed that the change of variables is composed of a common component affecting all variables in the set and an idiosyncratic component which is specific all variables in the set.

If the common factors can be derrived consistently they prove to be a valuable and more practical set of predictors for the variables of interest than the whole set of available variables. Nowadays, typically the factors are estimated using principal components analysis. Resultingly factor models harbor similarity to principal component regression[1] including the feature of dimensionality reduction which allows factor models to be estimated even if the number of columns exceeds the number of rows. This is precisely what makes factor models appealing in a macroeconomic forecasting context as there exist many international macroeconomic variables some of which are not covered over a very long time horizon. With factor models all this information can be used while traditional models had to rely on variable selection in order to reduce the number of parameters.

---

[1] In essence factor models calculated using principal components are the same as principal component regressions applied to a time-series context.

In a time-series framework an interesting question to ask is whether these common components are stable over time or if the relationship of the variables of interest with the underlying factors is prone to changes. Specificity about what is meant when structural breaks are referenced is added in section 3.

While there are recent additions to the field, the behavior of factor models under structural breaks is still being researched. There are at least two reasons why structural breaks are interesting in an econometric model. If we assume that strcutural breaks imply a change in the coefficients of the model at a certain period, this can be interesting in and off itself especially if the coefficients and the factors are to be interpreted directly or if factor models are estimated in the context of some economic theory. Additionally, from a forecasting perspective it is likely that a model estimated over a certain time interval does not generalize well onto another time frame if parameters are unstable. This makes limitting the observations to those which do not include periods with parameter instability attractive. For forecasting into the future it can e.g. be worthwhile to consider only the data after the structural break(s) has/have happened[2]. The hope is that the parameters of the model which are estimated relying only on the data after the last break might be closer to the true values which have generated the data to be forecast. This assumes that there will not be another structural break at the period which is to be predicted. E.g. Bai and Ng (2008) employ a similar method in their empirical application by considering estimates based on several sub samples allowing them to reason about parameter stability. It should be notet that Pesaran and Timmermann (2007) have shown that it can be slightly better to include same observations before the last break and not to fully cut-off any unstable periods.

Instead of cutting the length of the data set by excluding observations before the structural break, the width of the data set can be decreased, as well. This treatment is perhaps more tailored towards factor models because they are usually employed in a high-dimensional environment where many predictors are replaced by relatively few common factors. It is thus possible that the overall information content does not suffer too much if some of the variables are excluded and the factors are instead calculated on a limited subset. This approach can be especially attractive if the number of observations

---

[2]Alternatively, instead of splitting the data into multiple periods for each break, a single forecasting equation can be constructed which includes the structural break by introducing a dummy variable.

is limitted relative to the number of variables.

Or argued more directly and as is made explicit below, factor models assume the variable of interest to depend on the factors which in turn depend on the complete data set. If there is a structural break in the relationship of one of the variables with the factors, it is conceivable that this break propagates to the relationship between the factors and the variable of interest. Resultingly removing the variable with the break can help to keep the relationship stable. In the empirical application presented in this paper I find that this approach has a more favorable effect on forecasting accuracy than limiting the number of observations.

Structural breaks can be thought of as adding noise to the predictors. There exist already several methods to on how to deal with reducing the number of uninformative predictors prior to the estimation of factor models. Bai and Ng (2008) have developed the method of targeted predictors which is in essence a preselection tool for the predictors based on relevance for forecasting a variable of interest based on a restricted regression. Resultingly the approach to remove variables with associated structural breaks can be thought of as an addition to these methods.

This thesis treats testing for structural breaks in factor models and using knowledge about these breaks for forecasting.

The contribution is twofold: Firstly, the influence of structural breaks on factor models is examined. To this end feasible testing procedures developed by Breitung and Eickmeier (2011) are presented. Bootstrap versions of these chow type tests are developed and their performance is evaluated. Secondly, the information about the structural breaks is used for forecating. The performance of different model specifications is tested in the form of horse races. The competing models are:

1. A static factor model which ignores structural breaks

2. A dynamic factor model which ignores structural breaks

3. A dynamic factor model which takes structural breaks into account

4. Targeted dynamic factor models which ignore structural breaks

    (a) Under hard thresholding

(b) Under soft thresholding

5. Targeted dynamic factor model which take structural breaks into account

    (a) Under hard thresholding

    (b) Under soft thresholding

The empirical section shows that removing variables with associated breaks in the relationship usually improves the forecasting accuracy considerably. The targeted models which take structural breaks into account are interesting as a tool to analyze the interaction between the two different variable preselection schemes. The improvement in terms of forecasting precision is reduced if the data set is targeted prior to estimation which shows that targeted predictors and the removable of variables under the influence of structural breaks target a similar situation in the data.

The remainder of this paper is structured into 4 parts. Firstly factor models are introduced. Both static and dynamic models are presented along with some theory and estimation methods. Secondly structural breaks are infixed into the framework, their effects on the factors and loadings are looked at and testing methods are considered. Thirdly an empirical application is presented. Aforementioned models are estimated using a set of German and US American variables and prediction accuracy is compared. Finally the fourth part concludes.

Some notes on notation are in order. Throughout this thesis I refer to $X$ as a $T \times N$ matrix of potentially very many predictors. To stay consistent with existing literature I index $X$ with $i$ or $t$ interchangeably where $i$ stands for the column index and $t$ for the row index. In other words $X_t$ is the $N \times 1$ data vector for observation $t$ i.e. $t$ indexes along the longitudinal dimension, $X_i$ is the $T \times 1$ data vector for variable $i$ or $i$ indexes along the cross-sectional dimension. Also as is customary, throughout this paper unindexed capital letters denote matrices while unindexed lower case letters denote vectors except for innovation terms which are always in lower case letters. All variables containing data are assumed to be demeaned and normalized to have unit variance[3].

---

[3]I.e. every variable $y_t$ is derrived as $y_t = \frac{y_t^* - \bar{y}_t}{\sigma_y}$ where $y_t^*$ is the vector containing the original data, $\bar{y}_t$ is the sample average and $\sigma_y$ is the corrected sample standard deviation of $y$

# 1    Factor Models

As mentioned in the introduction, the idea behind factor models or diffusion index models is that there are underlying latent factors which explain the evolution of the observable variables collected in $X$. In other words all variables $X_i$ are thought to consist of a common component driven by the factors and an idiosyncratic component which is particular to each variable.

If factor models are to be used for forecasting they are ususally written in the form of a factor equation (1) and a forecasting equation (2).

$$X_t = \lambda(L)f_t + e_t \tag{1}$$

$$y_{t+h}^h = \alpha(L)W_t + \beta(L)\hat{f}_t + \varepsilon_{t+h} \tag{2}$$

$X_t$ is a $N \times 1$ vector of potentially very many predictors. $\lambda(L)$ is a $N \times q$ matrix of lag polynomials, $f_t$ is a $q \times 1$ vector of latent factors. $e_t$ is a $N \times 1$ vector of idiosyncratic errors which may or may not be serially auto-correlated. $W_t$ consists of variables which the researcher knows to be sufficiently relevant to enter the forecasting equation directly. Usually $W_t$ is taken to consist of s lags of $y_t$ and a constant term. $\hat{f}_t$ is an estimate of the factors in the factor equation (1). Estimation of the factors is treated below.

It is sometimes assumed that the error term follows a white noise process (i.e. independent and of mean $0$). In the vocabulary of factor models these are models with uncorrelated idiosyncratic components $e_t$ (more specifically $E(e_t e_t') = \Sigma = \text{diag}(\sigma_{e_1}^2, ..., \sigma_{e_N}^2)$ and $E(e_t) = 0$). However, "weak" dependence is more commonly assumed. Concretely, Geweke (1977) and Sargent and Sims (1977) distinguish between exact or strict Factor Models and approximate Factor Models. In essence the difference is composed of different assumptions for the error terms. Approximate factor models allow for limited correlatedness of the idiosyncratic innovations $e_{it}$ between periods while strict factor models do not. More precise deifinitions follow now and are taken from Breitung and Eickmeier (2006)

**Strict factor models**

Breitung and Eickmeier (2006) list the assumptions for strict factor models as follows:

For the innovations it is assumed $E(e_t) = 0$ and $E(e_t e_t') = \sigma = \text{diag}(\sigma_1^2, ..., \sigma_N^2)$. Also $E(f_t) = 0$ and $E(f_t f_t') = \Omega$. If $E(f_t) = 0$ it follows that $E(X_t) = 0$ which is enforced by demeaning the variables. Additionally $E(f_t e_t') = 0$ i.e. the innovations and the error terms are uncorrelated. It follows directly that $E(X_t X_t') = E[(\Lambda f_t + e_t)(\Lambda f_t + e_t)'] = E(\Lambda f_t f_t' \Lambda' + \Lambda f_t e_t' + e_t f_t' \Lambda' + e_t e_t') = \Lambda \Omega \Lambda' + \Sigma$

**Approximate factor models**

Approximate factor models losen the assumptions of strict factor models at the cost of assuming $N \to \infty$. $E(e_{it} e_{js})$ is allowed to be different from 0 but only weakly so. Similarly the factors $f_t$ and and the idiosyncratic error terms are allowed to be correlated but again only weakly so, such that $E(f_t e_t')$ does not have to be 0. But $N^{-1}\Lambda'\Lambda$ must converge to a positive definite limiting matrix $\Sigma_\Lambda$ which ensures that the factors influence each variable with a similar order of magnitude. Otherwise it could be that the loadings for some variables are 0 (Breitung and Eickmeier, 2006). Similiary it is assumed about the factors that $T^{-1} \sum_{t=1} T F_t F_t' \xrightarrow{p} \Sigma_F$. This allows for stationary processes e.g. an ARMA process.

Some more assumptions are necessary for the inferential theory[4] of Bai (2003) to hold. Namely these assumptions are the following Assumptions A through D from Bai (2003):

**Assumption A:** $E||F_t||^4 \leq M < \infty$ and $T^{-1} \sum_{t=1}^{T} F_t F_t' \xrightarrow{p} \Sigma_F$ for some $r \times r$

positive matrix $\Sigma_F$

**Assumption B:** $||\lambda_i|| \leq \bar{\lambda} < \infty$, and $||\Lambda'\Lambda/N - \Sigma_\Lambda|| \to 0$ for some $r \times r$

positive definite matrix $\Sigma_\Lambda$

---

[4]Most notably the results of Bai (2003) are consistency and normality of the factor estimations. The rate of convergence is shown to be $\min(\sqrt{N}, \sqrt{T})$.

**Assumption C:** $\exists M : 0 < M < \infty$ such that $\forall N, T,$

1.) $E(e_{it}) = 0, E|e_{it}|^8 \leq M$

2.) $E(e_s' e_t / N) = E(N^{-1} \sum_{i=1}^{N} e_{is} e_{it}) = \gamma_N(s,t), |\gamma_N(s,s)| \leq M \; \forall s$ and

$$T^{-1} \sum_{s=1}^{T} \sum_{t=1}^{T} |\gamma_N(s,t) \leq M$$

3.) $E(e_{it}' e_{jt}) = \tau_{ij,t}$ with $|\tau_{ij,t}| \leq |\tau_{ij}|$ for some $\tau_{ij}$ and for all $t$ and

$$N^{-1} \sum_{i=1}^{N} \sum_{j=1}^{N} |\tau_{ij} \leq M$$

4.) $E(e_{it}' e_{js}) = \tau_{ij,ts}$ and $(NT)^{-1} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} |\tau_{ij,ts}| \leq M$

$$\forall (t,s) E \left| N^{-1/2} \sum_{i=1}^{N} [e_{is} e_{it} - E(e_{is} e_{it})] \right|^4 \leq M$$

**Assumption D:** $E(\frac{1}{N} \sum_{i=1}^{N} ||\frac{1}{\sqrt{T}} \sum_{t=1}^{T} F_t e_{it}||^2) \leq M$

Note that I have changed the notation somewhat. Specifically the true values are not indicated by a $^0$.

Assumption A allows the factors the follow a stationary AR(p) process, Assumption B forces each factor to the variables $X_i$ via the repsective loadings, Assumption C allows for some degree of dependence and heteroskedasticity among the innovation terms which will be used below to simplify the calculation of the dynamic factor model and finally Assumption D allows some weak dependence between the factors and the innovation terms. More details can be found in Bai (2003).

Using these assumptions Bai (2003) shows that the factors and loadings can be estimated consistently. By placing some additional constraints on serial dependence (Assumption C), the eigenvalues, and by invoking a central limit theorem Bai (2003) also proofs normality of the principal component estimator using a simultaneous limit theorem. In other words: as $T, N \to \infty$, Bai (2003) get the following limiting distributions for the principal components factor estimates:

$(i)$ If $\sqrt{N}/T \to 0$ then for each t

$$\sqrt{N}(\tilde{F}_t - H'F_t) \xrightarrow{d} N(0, V^{-1}Q\Gamma_t Q'V^{-1})$$

Where $Q, \Delta_t$ and $V$ are defined in Bai (2003) and H is some rotation matrix

and

$(ii)$ If $\liminf \sqrt{N}/T \geq \tau > 0,$ then

$$T(\tilde{F}_t - H'F_t) = O_p(1)$$

The first part establishes normality for the principal component estimator. Note that the true value of the factors can only be estimated up to a rotation $H$ which is due to the common component being seperated into factors and loadings. This is discussed below when principal component estimation is introduced. The normality result is important because it allows researchers to establish significance and construct confidence intervals which is especially important if factor models are used to construct indices (Bai, 2003). Bai and Ng (2006) uses similar assumptions to show $\sqrt{T}$ consistency of the parameter estimates from factor augmented regressions[5] and to construct confidence intervals in those models.

As mentioned, an important advantage of factor models compared to simpler econometric models is that they do not necessarily suffer if the fraction $\frac{T}{N}$ becomes small whereas most models can not cope with more variables than observations. VAR models

---

[5]Factor augmented regressions are essentially what we do in the forecasting equation (2) where a normal regression model is augmented by estimates of the factors to account for the common component of the whole set of predictors.

for example have the problem that the number of parameters to estimate becomes too high quickly if $N$ is high relative to $T$. The reason why factor models do not have this problem is that firstly the factor equation is estimated. This gives estimates $\hat{f}_t$ of the factors. Then only the first $r$ eigenvectors are used in the forecasting equation where $r < N$[6]. This feature allows researchers in principle to append $X$ by further variables derived from the initial set. Bai and Ng (2008) for example use what they call squared principal components and quadratic principal components. The first of which appends squares of the $X_i$ to the initial matrix $X$ such that $X_t^{spc} = \{X_{it}, X_{it}^2\}$ and the second of which adds the cross products of $X_i$ such that $X_t^{qpc} = \{X_{it}, X_{it}X_{jt}\} \, \forall i \neq j$. The idea is to change the link function between $X_t$ and $f_t$ from being linear to allow for a nonlinear relationship. Bai and Ng (2008) also consider adding a squared term of the factors $f_t$ to the forecasting equation (2) to allow the volatility of the factors to influence the forecast. Naturally, the value of these approaches depends heavily on what is being forecasted.

Apart from the distinction into strict or approximate factor models, a further distinction can be made between static and dynamic factor models. This distinction concerns the relationship between lags of $f_t$ and $X$. Static models set $\alpha(L)$ to be of lag order 0 while dynamic models alow for dependence with other lag orders.

## 1.1 Static Factor Models

For static factor models $X_t$ depends only on $f_t$ and not on lags of $f_t$ i.e. factors enter only contemporaneously. Resultingly a static factor model is the case in which the lag order in equation (1) is 0.

$$X_t = \Lambda F_t + \bar{e}_t = C_t + e_t \tag{3}$$

Where $X$ is a $T \times N$ Matrix of predictors. $F$ is a $T \times r$ factor matrix. $\Lambda$ is a $N \times r$ loadings matrix. $\bar{e}$ is a $T \times N$ matrix of idiosyncratic errors. Resultingly $X_t$ is a $N \times 1$ vector and $F_t$ is of dimension $r \times 1$.

---

[6]Typically $r$ is much smaller than $N$. E.g. the famous one factor model of intelligence by Spearman (1904) (in which factor models were developed the first time) has $r = 1$.

$C$ is called the common component. Note that depending on whether authors index by row, column or both, several alternative ways of writing down the factor model equation are used in the literature.

$$X_{it} = \Lambda_i'F_t + e_{it} = C_{it} + e_{it} \tag{4}$$

$$X = F\Lambda' + e = C + e \tag{5}$$

$$X_i = F_i\Lambda_i + e_i = C_i + e_i \tag{6}$$

For the sake of completeness $X_i$, $F_i$ and $e_i$ are $T \times 1$ vectors while $\Lambda_i$ is $1 \times 1$. In the literature $\Lambda_i$ is also sometimes written as $\lambda$.

Loosely and graphically speaking each $\Lambda_i$ maps the variables $X_i$ which are the columns of $X$ to the corresponding factor $F_i$.

## 1.2   Dynamic Factor Models

In applications there will usually be some kind of time dependence structure. Thus, dynamic factor models introduce a dynamic process into the factors.

$$f_t = \Psi(L)f_{t-1} + \eta_t \qquad \text{(VAR representation)} \tag{7}$$

In practice it is usually easier to ignore the dynamic process in the factors $f_t$ and follow e.g. Stock and Watson (2005) among others in assuming a dynamic proccess in the innovation term of the factor equation (1) instead. This captures the dynamic process while avoiding the technical difficulty in estimating the dynamic factors themselves. The underlying thought is that there will be some residual autocorrelation if the dynamic process of the factors is not modelled explicitely which can be caught by an AR(p) process for the innovation term (Breitung and Tenhofen, 2011). Additionally the lag polynomial $\lambda(L)$ in the factor equation (1) captures the noncontemporaneous interactions of the factors $f_t$ and $X_t$.

For the purpose of this paper I assume the following AR processes for the innovation

term of the dynamic factor models (Breitung and Eickmeier, 2011).

$$e_{it} = \varrho_{i,1} e_{i,t-1} + ... + \varrho_{i,p_i} e_{i,t-p_i} + \xi_{it} \text{ where } \xi_{it} \text{ is white noise} \tag{8}$$

### 1.2.1 Static interpretation of Dynamic factor models

Dynamic factor models can be written as static factor models in a linear state space environment. This gives the static factor equation (3) and $\Phi(L)F_t = G\eta_t$ where $\Phi(L)$ is defined such that equivalence to (7) holds and $F_t$ is defined below as the stacked lagged dynamic factors $f_t$ (Stock and Watson (2011)). The result are as follows.

Solving the lag operator, we can write equation (1) in the form

$$X_{it} = \lambda'_{i1} f_t + ... + \lambda'_{im} f_{t-m} + e_{it} \tag{9}$$

Following Bai and Ng (2002) this can be rewritten as

$$X_{it} = \Lambda_i^{*'} F_t + e_{it}$$

$$\text{where } \Lambda_i^* = \begin{bmatrix} \lambda_{i1} \\ \lambda_{i2} \\ \vdots \\ \lambda_{im} \end{bmatrix} \text{ and } F_t = \begin{bmatrix} f_t \\ f_{t-1} \\ \vdots \\ f_{t-m} \end{bmatrix}$$

To keep consistency between variable naming schemes denote the dimension of $F_t$ as $r \times 1$ and the dimension of $\Lambda_i^*$ as $N \times r$. Thus a dynamic factor model can always be rewritten in static form and resultingly the procedure introduced below can be used to estimate the $r$ "static factors"[7] $F_t$ as specified below. It can be seen that the dimension of the resulting stacked vector of "static" factors is $q(m+1)$ thus we can set $r = q(m+1)$. This shows that there is a relationship between $r$ and $q$ if the dynamic factor model is written as a static factor model. To identify $q$ it remains to estimate $m$.

There exist several ways to identify the dynamic factors from there[8]. Firstly Gi-

---

[7]Breitung et al. (2004) call the $f_t$ "structural factors" and the $F_t$ "reduced form" factors.
[8]Note that identifying the dynamic factors might not be of interest for forecasting but rather that it is a necessary condition in order to interpret the factors in a dynamic factor model. For forecasting

annone et al. (2002) suggest estimating the VAR $F_t = CF_{t-1} + \kappa_t$ and then to apply principal component analysis on an estimate of the residual covariance matrix (see Breitung et al. (2004) or Giannone et al. (2002) for details). Breitung et al. (2004) exploit the relationship between $q$ and $r$ in order to find information criteria which can determine $q$ and the lag order of the $\lambda$. This paper does not estimate the dynamic factors directly as in the application the factors are primarily used for the second stage regression to estimate the forecasting equation.

Imagine that the true data has been generated by a dynamic factor model. If instead of using the dynamic factor model specification and finding the dynamic factors from the static factors, a researcher simply estimates a static factor model it can be seen above that the number of factors $r$ will be larger than the number of dynamic factors $q$. Most authors use the approach briefly hinted at above to mimic the dynamic process in the factors by lagging the static factors. This does not reproduce the exact model if the true data generating process follows (7) but it accounts for some time dependence and is much simpler because it only entails estimation of the static factors from the data and lagging them in the first stage regression.

## 1.3  Estimation of the factors

### 1.3.1  Static factors

There exist now a number of ways to estimate the static factors $F$ in (5). Ultimately what is looked for is a solution to the minimization of the squared error.

$$\min_{F_1,\ldots,F_T,\Lambda} \frac{1}{NT} \sum_{t=1}^{T} (X_t - \Lambda F_t)'(X_t - \Lambda F_t) \tag{10}$$

Since both $\Lambda$ and $F_t$ are unspecified, there are infinitely many solutions to (10) and the problem is not uniquely identified. To achieve identification, assumptions can most conveniently be put on either $\Lambda$ or $F_t$[9]. Usually $\Lambda$ is normalized such that $N^{-1}\Lambda'\Lambda = I_r$.

---

purposes, in order to choose the number of factors and the number of the factor lags used in the forecasting equation (2) one can also follow Bai and Ng (2008) in using the BIC criterion to choose both simultaneously.

[9]Bai and Ng (2013) consider several alternative normalizations. They also show that direct interpretation of the factors and loading matrices are possible if one is willing to additionally assume

Alternatively $T^{-1}F'F = I_r$ is also frequently used. The choice of normalization can have a noticeable impact on the computational demand of calculating a solution for the minimization problem in (10) especially if the factors are to be calculated many times in some simulation study. The columns spaces spanned by the estimates of $F$ are equivalent with both normalizations (Stock and Watson, 2011).

Stock and Watson (2011) list four ways which are being applied in the literature to get a solution $\hat{F}$ and $\hat{\Lambda}$ to the minimization problem (10). Firstly a Gaussian process can be assumed which allows for maximum likelihood estimation using the Kalman filter. To do this a dynamic factor model can be rewritten as a static factor model in a linear state space form as introduced above.

Secondly principal components (and other nonparametric averaging methods) can be used to estimate both $\hat{F}$ and $\hat{\Lambda}$ at the same time. This requires relatively weak assumptions and is particularly easy and computationally efficient to implement which explains the popularity of this approach. This is also the approach taken in this thesis. Additionally the principal component estimation can be generalized by weighting the minimization problem with the variance matrix of the error term to take account of error variances which are not prooportional to the identity matrix (see Stock and Watson (2011) for details). Evidence of performance improvements by using the generalized principal components approach over the standard principal components estimator are mixed. Resultingly the simpler principal components estimator is used here. It is described in more detail below and in appendix A. As mentioned above, an unattractive feature of principal component estimation is that it can not identify the "true" factors $f_t$ but rather a transformation $Qf_t$ for some undefined rotation matrix $Q$ which results in some difficulty in interpreting the factors directly. This is due to the solution of the minimization problem not being unique. If the factors are used for forecasting this feature is not of much importance, however.

It should be noted that principal components estimation of the factors is asymptotically equivalent to maximum likelihood estimation if normality of the error terms is assumed (Bai, 2003).

Thirdly mixture approaches between the two methods from above can be constructed.

diagonality of $\Lambda'\Lambda$ or that $\Lambda$ is a block matrix of 2 submatrices with one submatrix being triangular.

Fourthly and finally Bayesian methods can be used to get estimations of the factors and loadings. These can have computational advantages over the maximum likelihood approach and can be useful if the assumption of Gaussian error terms is unappealing (Stock and Watson, 2011).

### 1.3.2   Principal component estimation of the static factors

The principal component method estimates $\hat{F}$ and $\hat{\Lambda}$ as follows. Normalizing $F$ such that $T^{-1}F'F = I_r$ gives $\hat{F} = \sqrt{T} * eigenvectors_r(XX')$ where $eigenvectors_i(A)$ returns the i eigenvectors corresponding to the i largest eigenvalues of a square matrix $A^{10}$. Since $\hat{X} = \hat{F}'\hat{\Lambda}$, it follows directly that we can estimate $\Lambda$ by $\hat{\Lambda}' = (\hat{F}'\hat{F})^{-1}\hat{F}'X$. due to the normalization $\hat{F}'\hat{F} = T$ and we can write $\hat{\Lambda} = X'\hat{F}/T$.

If, alternatively, we apply the normalization $N^{-1}\Lambda'\Lambda = I_r$ the solution to (10) becomes $\hat{\Lambda} = \sqrt{N} * eigenvectors_r(X'X)$. This solution to the optimization problem is shown in appendix B). Then we can follow from $\hat{X} = \hat{F}\hat{\Lambda}'$ that $\hat{F} = X\hat{\Lambda}(\hat{\Lambda}'\hat{\Lambda})^{-1}$. Again due to the normalization $\Lambda'\Lambda = N$ and we can write $\hat{F} = X\hat{\Lambda}/N$.

Note that the normalization $T^{-1}F'F = I_r$ requires us to calculate the eigenvectors of $XX'$ which has dimension $T \times T$ while the alternative $N^{-1}\Lambda'\Lambda = I_r$ requires us to calculate the eigenvectors of $X'X$ which is $N \times N$. Depending on the relative sizes of $T$ and $N$ it can make a considerable difference in terms of computational burden which method is applied.

Also note that what is being estimated here are in fact rotations $FH$ of the true factors $F$ and $\Lambda H^{-1}$ of the true loadings matrix. However, $\hat{F}\hat{\Lambda}' = (FH)(H^{-1}\Lambda')$ estimates the common component (Bai and Ng, 2013).

Stock and Watson (1998) and Stock and Watson (2002) show that the principal component estimator remains consistent if $\Lambda$ varies a bit over time as long as $\frac{T}{N} \to 0$. The case where $\Lambda$ varies strongly when we speak of structural breaks is discussed below.

---

[10]Eigenvalues and eigenvectors can be calculated with most standard statistical packages. The analytical solution can be seen in appendix A.

### 1.3.3 The number of factors

One important issue that remains to be addressed is the question of how many factors should be used in the factor equation. There exist many different ways that this choice can be made. Firstly there are simple rules of thumb. E.g. some researchers only consider factors with eigenvalues greater than 1 (see Larsen and Warne (2010)). The intuition behind this rule is that the variance of the variables has been normalized to 1. Since the variance of the factors is equal to the corresponding eigenvalues (see appendix A) this implies that only factors with more information about variation than any single predictor is included. Other methods also focus on the percentage of variation explained but simply require the cumulative variance to be higher than a certain threshold value e.g. 90% of variation. Then there are methods which rely on a similar idea but employ graphical methods. One approach is to inspect the resulting scree plots[11]. The intuition is to only take factors into account until the addition of another factor does more additional variance than a certain threshold. Basing on this, formal tests have been derrived which rely on the relative size of the eigenvalues (Stock and Watson, 2011). Additionally Larsen and Warne (2010) stress that since the eigenvalues are summary statistics they are prone to sampling errors which allows for the construction of confidence intervals.

This list is not comprehensive, there are many more such methods and they are not all listed here. However, an additional method needs to be mentioned. The information criteria developed by Bai and Ng (2002) are perhaps more appealing than the approaches considered so far as they formalize the costs and benefits of an additional factor explicitly (Stock and Watson, 2011). Cragg and Donald (1997) show in a Monte Carlo analysis that $T \to \infty$ and $N \to \infty$ makes that classical theory for predicting the number of factors performs badly. Resultingly information criteria as in Bai and Ng (2002) should be used which take into account that both $T$ and $N$ can grow. Bai and Ng (2002) show that their results hold under heteroskedasticity and weak serial correlation and are thus applicable to approximate factor models.

Bai and Ng (2002) consider 6 criteria, the first three of which they call $PCp$ criteria

---

[11]Scree plots plot the ordered eigenvalues against the number of factors (i.e. eigenvalues and corresponding eigenvectors) which results in a falling curve.

(Panel $C_p$ criteria). These criteria result from a generalization of Mallows' $C_p$ (Mallows, 1973). The second set of 3 criteria refine the usual criteria used in time-series analysis to depend on both $N$ and $T$. Bai and Ng (2002) highlight that the criteria $IC_{p1}$, $IC_{p2}$, $PC_{p1}$ and $PC_{p2}$ are specifically suited for principal component estimation. Bai and Ng (2002) present a montecarlo study to highlight the differences between the criteria. Notably the $PC_p$ criteria require the specification of a maximum number of factors, the choice of which is quite arbitrary[12]. Resultingly this paper uses the $IC_p$ criteria. Specifically the $IC_{p2}$ criterion because it seems to be more conservative (i.e. it usually underestimates $r$ whereas $IC_{p2}$ overestimates at times).

In order to understand the criteria better, tables 1 and 2 of Bai and Ng (2002) have been replicated using both different values for $r$ and $kmax$. If $r$ is chosen as in Bai and Ng (2002), the replicated results are almost identical to the original calculations and thus are not reported[13]. The results can be found in two tables in appendix C. Tables I through VIII in Bai and Ng (2002) in essence show 3 points. Firstly, traditional criteria perform poorly in estimating the number of factors if compared to the 6 presented criteria. Of the newly developed information criteria $PC_{p1}$, $PC_{p2}$, $IC_{p1}$ and $IC_{p2}$ perform well even under heteroskedasticity and autocorrelation. Secondly, if $\min\{N,T\}$ becomes too small, the performance of the criteria suffers which is highlighted by the bottom 5 results in the tables. Especially $PC_{p3}$ suffers in this regard. In the given specifications the cut-off value for $\min\{N,T\}$ appears to be 40 or 60 if the variance of the innovations is high. Thirdly, $PC_{p3}$ and $IC_{p3}$ are less robust to a small $\min\{N,T\}$. The replication results in appendix Bai and Ng (2002) harbor three additional insights. Firstly as $r$ increases, the requirements for $\min\{N,T\}$ are also higher. For $r = 7$ and $r = 9$ it seems $\min\{N,T\}$ should be at least 100. Secondly the choice of $kmax$ can matter a lot. Especially for small samples the specification of $kmax = 8$ used in Bai and Ng (2002) choose the maximum number of factors frequently when the $N$ and $T$ are too small. Almost all results in the last 5 rows set $\hat{k} = 8$ whereas the replicated results often report a lower number of factors if $kmax = \text{ceil}(\min\{T,N\})$. Noticeably, this

---

[12]In their montecarlo study Bai and Ng (2002) set the maximum number of factors to be 8 noting that they used the rule in Schwert (2002) and that one could consider $kmax = \text{int}[(\min\{N,T\}/100)^{1/4}]$. Tests which are not presented here show that choosing a different $kmax$ can result in the $PC_p$ criteria to become quite unstable. Especially high values of $kmax$ often result in the maximum number of factors chosen such that $r = kmax$

[13]I.e. differences are so small that they can be explained by sampling variation.

also happens in the replicated results for $r \in \{1, 3, 5\}$ and $kmax = \text{ceil}(\min\{T, N\})$. If $kmax = \text{ceil}(\min\{T, N\})$ $IC_{p3}$ predicts $\hat{k} = 50$ in the $T = 100, N = 100$ case for $r \in \{1, 3, 5, 7, 9\}$. Thus, the choice of $kmax$ can have a huge impact on the prediction of the number of factors. Luckily it is typically easy to see when the information criteria choose all the factors. In those cases the plausibily of the results should be checked by hand.

An interesting question if one is ultimately interested in forecasting is whether the "true" number of factors is of any interest at all or if the number of factors used in the forecasting equation (2) should perhaps be different from the predicted $r$ in the static model or $q$ in the dynamic model. The answer can not be easily given. Similar to Bai and Ng (2008) we try to dodge this question somewhat by using the BIC criterion and the out-of-sample error to choose a model among a list of models with differing number of factors (or number of lags of the variable of interest).

Breitung and Eickmeier (2011) among others, however, note that if one is primarily interested in forecasting, the number of dynamic factors can in principle be ignored and a static factor model with a higher number of factors can be estimated instead. The results presented in the empirical seciont, however, indicate that the forecasting accuracy of models which take dynamic effects into account is higher than static models.

### 1.3.4 Interpreting the factors

If the factors resulting from factor models are to be interpreted, it is necessary to identify the true factors rather than a transformation.

As has been noted, a difference has to be made for dynamic factor models which have been described above and static factor models. As Breitung et al. (2004) argue and as has been made explicit in section 1.2.1 in a dynamic setting the biggest challenge is identifying the dynamic factors from the static factor estimates if the principal components approach is taken. If one of the several approaches described above is taken, interesting economic insights can be gotten from the factor estimates. The relative importance of the factors for the variable of interest can be seen in the significance and size of the respective parameters in the second stage regression (2). Additionally the

composition of the factors in terms of the initial variables contained in $X$ can be seen in the factor loadings $\Lambda$. Most authors who employ factor models try to examine the composition of the factors in terms of the initial variables in this way[14]. Breitung et al. (2004) for example identify two static factors, one of which has a dynamic component and enters with a lag. They find that the first factor is given by short term interest rates while the second factor is composed mostly of PPI inflation of different countries.

Another way to evaluate and interpret the factors is to look at the percentage of variation explained by the factors for each variable individually and use this to draw conclusions about the factors. Stock and Watson (2005) forecast inflation in VAR and estimate impulse responses of a monetary policy shock applied by the central bank. To that end they use a factor model on 132 of macroeconomic variables. They find the first factor to explain a high percentage of variation in production and hours-worked related series such as industrial production and capacity utilization whereas the second factor explains variation in interest rates, consumption and stock prices. The second and the third factor in combination can explain variations in inflation (taken together they account for 65% of inflation in their data).

The fact that factors seem to "specialize" in a choice of predictors which they explain is possibly a convenient by-product of principal component estimation as a method to estimate the factors. Principal component estimation forces the factors to be orthogonal and thus might make it more likely for factors to undergo the observed "specialization".

While these approaches sometimes allow for insightful conclusions it is not always possible to relate a single or sufficiently few factors to the variation of the predictors $X$. Sometimes variables can only be sufficiently explained by a combination of several factors which does not allow for easy and clean economic interpretations.

## 1.4 Forecasting

The forecasting equation (2) can be estimated using a simple linear regression of $y_{t+h}$ on $\hat{F}_t$, $y_t$, lags of $y_t$ and depending on the model specification lags of $\hat{F}_t$.

The performance of factor models for forecasting purposes is most pronounced if the

---

[14]It is instructive to read the original paper by Spearman (1904) which finds that many different sources of intelligence depend on a single commong factor. While this claim has been refuted since and the result seems to depend strongly on the chosen data set, Spearman (1904) provides an interesting example of how the factor loadings can be interpreted and their relevance checked.

number of variables used in the factor equation is high (Stock and Watson, 2011). This makes sense because more variables improve the fit of the factor equation. However, there is a trade-off between quantity and quality of variables[15] that should not be ignored. This case will be adressed in the next subsection.

It has been noted above that the number of factors in forecasting equation(2) and factor equation (1) need not be the same. Similarly the number of factor lags which the forecasting equation (2) sets to be the same for all factors could be left flexible between factors. While this is likely to improve forecasting performance somewhat, this paper abstains from the additional complication.

Actual point forecasts for the variable of interest can be gotten by rescaling the demeaned and normalized forecast back to the original form i.e. by multiplying the original standard error of the series and then adding the population average.

### 1.4.1 Targeted predictors

Bai and Ng (2008) follow the idea that if factor models are ultimately going to be used for forecasting it might make sense to "target" the set of predictors $X_t$ to the task of forecasting the variable of interest $y_t$. The idea stems from Boivin and Ng (2006) who find that selecting only the "informative" variables can yield better results than taking a bigger set of variable which include the same "informative" variables but also additional ones. Bai and Ng (2008) thus provides a formal rule for which variables should be used for forecasting. To that end they use hard and soft thresholding rules to preselect among the predictors $X_t$ those that have a meassurable influence on $y_t$. Resultingly, a subset $\tilde{X}_t \subset X_t$ is selected for estimating the factor equation, the hope being that then the factors are better suited for predicting $y_t$.

For hard thresholding they regress $y_t^h$ on $W_{t-h}$ and $X_{it-h}$ for each $i = 1, ..., N$[16]. Then they compare the resulting t-statistics to a threshold and discard those predictors whose t-statistics do not exceed that threshold.

For soft thresholding Bai and Ng (2008) perform penalized regressions in the form of LASSO, elastic net and Least angle regression (see Tibshirani (1996), Zou and Hastie

---

[15]I.e. a smaller set of variables with a high predictive power can outperform a larger set of variables which includes the variables from the smaller set
[16]Note that the $N$ regressions include only the lags and single predictors $X_i$ as independent variables.

(2005), Efron et al. (2004) for details) using the fact that these regression specifications set the parameters of variables which are not "important" to 0. The idea is to perform a regularized regression i.e. a normal least squares regression with the addendum of a restriction on the size of the parameters $\beta$. LASSO sets the restriction to be that $||\beta||$ is not bigger than some constant. The resulting optimization problem reads

$$\min \text{RSS} + \lambda \sum_{i=1}^{N} |\beta_j|$$

Ridge regression takes a similar approach but uses the $L_2$ norm $\sum_{i=1}^{N} \beta_i^2 \leq c$ for some penalty term $c$. The elastic net weighs between the two regularization mechanisms keeping both the sparsity of the LASSO estimator and the property of the Ridge estimator to include groups of variables which are correlated together Zou and Hastie (2005). Least angle regression is a variable selection mechanism which updates coefficients stepwise in a way that keeps the angle between variables which are most correlated with the residuals of the currently chosen set equal until (in the next step) a new variable has to be added to the set of most correlated variables.

The sparsity property of these penalized regressions allows the choice of relevant variables to be those which have a coefficient different from 0. Note that ridge regression, although similar in construction, does not have this property because coefficients are shrunk in absolute size but not set to 0. In applications the three approaches named above typically give similar results. The practical problem with soft thresholding lies in the fact that the algorithms[17] provide solutions to restricted regression problem in the form of a solution path consisting of several steps where in each step the "most relevant" variable is added in the sense that the respective parameter is estimated and set to be different from 0. After the final step all variables are included in the model. Typically one choses the number of steps to be taken by the algorithm and thus the value of the restriction parameter using a cross-validation approach or alternatively sometimes Mallows $C_p$ is used. Bai and Ng (2008) propose to use the BIC or the AIC criteria to choose the size of the active set in a general application of the LARS-EN algorithm.

---

[17]In practice the LARS-EN (Least angle regression and Elastic net) algorithm can be used which computes Elastic Net solutions for all values of the penalization parameter using Least angle regressions. It has been shown by Efron et al. (2004) that the LARS-EN implementation includes the LASSO as special case (namely if the parameter $\lambda_2$ is set to 0. $\lambda_2$ weighs the penalty of the Elastic net between Ridge regression and the LASSO penalty).

However, in the framework of factor models it is of interest how the factors calculated from the chosen active set performs in the forecasting equation. This open question is answered here by considering the optimal size of the restriction parameter in a forward looking way. The RMSE of pseudo out-of-sample forecasts are calculated to choose the size of the targeted set of variables. This approach is presented in the empirical section.

The approach of targeting predictors, while attractive, requires additional thought in the presence of structural breaks. As has been stated, targeting the data prior to estimation can be interpreted as removing predictors which are not relevant or too "noisy" to be used in the prediction of the variable of interest. One of the approaches considered here in dealing with structural breaks once detected, is to remove the offending variables. The success of this method to deal with structural breaks is meassured using the RMSE. Resultingly it is unclear whether improved forecasting accuracy are the results of removing structural breaks or of removing a generally noisy predictor which happens to also have a structural break but would be removed also if it would not feature the structural break. Luckily this relationship can be somewhat untangled by targeting the data prior to testing the approach to deal with structural breaks. If the residual effect is positive it can be assumed that the method is succesfull especially since targeting the data in advance is likely to already account for some of the structural breaks. Put differently, if there is a residual positive effect of removing structural breaks it can be assumed that because some variable with structural breaks have already been removed by targeting the predictors, the approach of removing variables associated with breaks is successful.

## 1.5 Applications

Factor models have been applied in a wide range of problems both in economics and in other domains apart from economics, most notably perhaps psychology and biology. Specifically biology has an abundance of high-dimensional data sets on which factor models can be applied to great success. Within the realms of economics, prominently, there are applications in the calculation of indices. Analogeously to the factors in factor models, indices try to capture underlying tendencies applying to all variables.

Resultingly the underlying factors calculated from a data set can be seen as some kind of index. Examples for this type of application include Langetieg (1978) who calculates a performance index of stockholder gains from mergers. Kim and Nelson (1998) use Bayesian methods to estimate a regime switching factor model which indicates boom and bust business cycles. Altissimo et al. (2001) calculate a monthly inflation index using time-series data of differing frequencies.

Additionally factor augmented regressions can be used in which the factors are used as a way to account for a component common to all variables.

Schumacher (2010) employs international predictors to forecast German GDP which has influences the data choice for this thesis. The immediate justification is similar. Firstly it is intuitive that international data has an influence on an open economy directly. But secondly it is conceivable that there are underlying factors which influence both German GDP growth and international macroeconomic variables at the same time. However, to make sure that the information conveyed in the additional international predictors is of relevance for forecasting, Schumacher (2010) "targets" the predictors prior to estimating the factor equation. Schumacher (2010) does not deal with structural breaks.

Forni and Reichlin (1998) use the factors to estimate structural VAR equations to identify economy wide shocks.

Eickmeier and Breitung (2006) identify the sources of fluctuations in the euro area.

# 2   Structural breaks

In a macroeconomic setting it is easy to picture reasons fo why relationships between variables might change over time. Commonly cited reasons include the change of technology, different trade paterns, institutional and governmental changes. Given the relatively long time frame in macroeconomic data it is likely that most data sets contain breaks in at least a few of the variables. Several approaches to dealing with structural breaks have been developed in the economic literature. Firstly structural breaks could be included into the model making the model more sophisticated in a sense but also more complicated. To this end e.g. the approach of Bai and Perron (2003) can be

followed. Alternatively simpler model can be maintained while focussing on the part of the data which remains stable. Banerjee et al. (2007) among others consider the presence of structural breaks in a simulation study and find that using only the observations after the break is preferential.

The treatment of structural breaks in this paper follows Breitung and Eickmeier (2011) closely and is introduced below. The question of interest is how a researcher should react to structural breaks in the factor equation.[18]

The factor model under a structural breaks in period $T^*$ can be written as follows:

$$y_{it} = f'_t \lambda_i^{(1)} + \varepsilon_{it} \text{ for } t = 1, ..., T^* \tag{11}$$

$$y_{it} = f'_t \lambda_i^{(2)} + \varepsilon_{it} \text{ for } t = T^* + 1, ..., T \tag{12}$$

The intuition is the same as with simpler models: a structural break implies a different relationship between the variable of interest and the factors and hence different factor loadings for the periods before and after the breaks. Multiple structural breaks can be specified accordingly. In theory, structural breaks could be also thought of as affecting the forecasting equation (2). This is ignored here because structural breaks in equations of this type have been thoroughly researched.

Interestingly, estimating a static factor model without structural breaks if the true data generating process includes structural breaks, yields an analogy to ignoring dynamic factors if they were present in the data generating process: Breitung and Eickmeier (2011) note that the space estimated by the factors will be of higher dimension. This means that the number of factors $r$ predicted e.g. by the Bai and Ng (2002) information criteria will be higher than if no structural breaks were present. Breitung and Eickmeier (2011) note that similar to dynamic factor models it is enough to increase the number of factors if one is primarily interested in decomposing the common component from the idiosyncratic component (that is to say if one is interested foremost in forecasting).

---

[18]Structural breaks can refer to changes in the means and variances of variables or a change in the relationship of variables. I focus on the latter. From now on when I refer to structural breaks it is always assumed to mean breaks in the parameters of the factor equation (1). Breaks in the forecasting equation are ignored because the literature dealing with this kind of breaks is already large and there exist already a number of instructions for treatment.

If structural breaks are detected in one or more of the factors the question remains what should be done next. Since structural breaks consist of an instability in the factor loadings at a certain period it can be optimal in a forecasting sense to estimate the respective parameter after the structural break has occured in the hope that this new parameter value contains more accurate information about the future. There is of course a trade-off here because the data series is shorter afterwards which has a negative effect on prediction accuracy. An alternative approach consists of removing the variable containing the structural break from the data set. This approach allows all other series to enter the forecasting equation at the same length as before while the added noise through the structural breaks is removed. The idea is similar to targeted predictors which remove uninformative predictors from the data set in order to improve the ratio of information to noise. Because of the typically high dimension of the data matrix $X$, factor models appear to be well suited for the second approach. Which method should be chosen depends on the size of the structural break, the relative position of the break and the relative information content of the other variables. In practice it is necessary to test in order to weigh the benefits and costs. This is done in the empirical section.

## 2.1 Testing for structural breaks

Breitung and Eickmeier (2011) develop several Chow type tests for the static factor model under the assumptions of the strict factor model. They also present Quandt-Andrews type supremum tests to identify breaks at unknown dates. The test statistics are as follows.

$$\mathrm{lr}_i = T[\log(S_{0i}) - \log(S_{1i+S_{2i}})] \tag{13}$$

Where

$$S_{0i} = \sum_{t=1}^{T}(y_{it} - \hat{f}_t'\hat{\lambda}_i)^2, \ S_{1i} = \sum_{t=1}^{T_1^*}(y_{it} - \hat{f}_t'\hat{\lambda}_i^{(1)})^2 \text{ and } S_{2i} = \sum_{t=T_i^*+1}^{T_1^*}(y_{it} - \hat{f}_t'\hat{\lambda}_i^{(2)})^2$$

denote the residual sum of squares for the whole date range, the first subperiod up until the structural break and the period from the break to the end of the sample

respectively. $\hat{\lambda}^{(1)}$ and $\hat{\lambda}^{(2)}$ denote the esimated factor loadings calculated on the data before the break and after the break respectively.

$$\text{lm}_i = TR_i^2 \text{ where } R_i^2 \text{ is the r-squared from a regression } \hat{\varepsilon}_{it} = \theta_i' \hat{f}_t + \phi \hat{f}_t^* + \tilde{\varepsilon}_{it} \quad (14)$$

$$\text{wald}_i = \text{ Wald statistic for } H_0 : \Psi_i = 0 \text{ in the regression } y_{it} = \lambda_i' \hat{f}_t + \psi \hat{f}_t^* + \nu_{it} \quad (15)$$

Where

$$\hat{f}_t^* = \begin{cases} 0 \text{ for } t = 1, ..., T_i^* \\ \hat{f}_t \text{ for } t = T_i^* + 1, ..., T \end{cases}$$

$$\text{LM}^* = \frac{\left( \sum_{i=1}^N s_i \right) - rN}{\sqrt{2rN}} \quad (16)$$

The LM* statistic pertains to the joint test of no structural break in all the factor loadings. Note, however, that the assumptions required for the LM* statistic are those of a strict factor model, namely no interdependence in the error terms $e_t$ in (1).

The LR, LM and Wald statistics defined above can be used to define a test for a unknown break date as in Andrews (1993). Under mild assumptions Breitung and Eickmeier (2011) show that the supremum statistic defined over the LM statistics is distributed using the nonstandard distribution given in Andrews (1993) and retabulated due to calculation errors in Andrews (2003).

$$\mathscr{S}_{i,T}(\tau_0) = \sup_{\tau \in [\tau_0, 1-\tau_0]} (s_i^\tau) \text{ where } s_i^\tau \text{ refers to either the LR, LM or Wald statistic} \quad (17)$$

As mentioned in the introduction of the dynamic factor model, I follow Breitung and Eickmeier (2011) in assuming an AR(p) process for the innovation term as in equation (18):

$$e_{it} = \varrho_{i,1} e_{i,t-1} + ... + \varrho_{i,p_i} e_{i,t-p_i} + \xi_{it} \quad (18)$$

It is assumed that the process is stationary. These assumptions are then in line with the assumptions of the approximate factor model introduced above.

For dynamic factor models[19] Breitung and Eickmeier (2011) propose to use a GLS transformed model, i.e. to perform the regression:

$$\varrho_i(L)y_{it} = \lambda_i'[\varrho(L)\hat{f}_t] + \psi'[\varrho_i(L)\hat{f}_t^*] + \varepsilon_{it}^*$$

The lag polynomials are obtained by applying some information criterion to the $N$ regressions

$$\hat{\varepsilon}_{it} = \varrho_{i,t-1}\hat{u}_{i,t-1} + ... + \varrho_{i,p_i}\hat{u}_{i,t-p_i} + \tilde{\varepsilon}_{i,t}$$

Similar test statistics as above can then be applied to the results of the GLS regressions. Breitung and Eickmeier (2011) only present the LM-statistic arguing that it has the best size properties compared to the LR and Wald statistics.

$$\hat{\varrho}_i(L)y_{it} = \lambda_i'\left[\varrho_i(L)\hat{f}_t\right] + \psi_i'\left[\hat{\varrho}_i(L)\hat{f}_t^*\right] + \tilde{\varepsilon}_{it}^* \text{ for } t = p_i + 1, ..., T_i$$

Under heteroskedasticity the regression has to be weighted by the covariance matrix. If we also take account of the possibility of a break in the covariance matrix we get two distinct regressions

$$\frac{1}{\hat{\sigma}^{(1)}}\varrho_i(L)y_{it} = \lambda_i'\left[\frac{1}{\hat{\sigma}_i^{(1)}}\varrho_i(L)\hat{f}_t\right] + \left[\frac{1}{\hat{\sigma}_i^{(1)}}\varrho_i(L)\hat{f}_t^*\right] + \tilde{\varepsilon}_{it}^* \text{ for } t = p_i + 1, ..., T_i^*$$

$$\frac{1}{\hat{\sigma}^{(2)}}\varrho_i(L)y_{it} = \lambda_i'\left[\frac{1}{\hat{\sigma}_i^{(2)}}\varrho_i(L)\hat{f}_t\right] + \left[\frac{1}{\hat{\sigma}_i^{(2)}}\varrho_i(L)\hat{f}_t^*\right] + \tilde{\varepsilon}_{it}^* \text{ for } t = T_i * +1, ..., T$$

## 2.2 Bootstrapping Breitung and Eickmeier (2011)

In order to test whether the properties of the above test statistics can be improved upon I have repeated the montecarlo study presented in Breitung and Eickmeier (2011) tables 2 and 3 with the difference that the test statistics have been bootstrapped.

---

[19]Note that Breitung and Eickmeier (2011) call a model where the innovations are generated by an AR(p) model a dynamic model whereas the literature usually refers to a dynamic process in the factors as in equation (7). This is similar in effect but not quite the same.

The idea behind bootstrapping test statistics comes from the definition of the p-value. Denoting with $\tau$ the test statistic of interest, by $\hat{\tau}$ the realization of that test statistic given the observed data and by $F$ the cumulative distribution function $\tau$ under the null hypothesis we can borrow the notation from Davidson and MacKinnon (2004) and write the p-value as

$$p(\hat{\tau}) = 1 - F(\hat{\tau})$$

Resultingly the p-value can be estimated by estimating $F(\hat{\tau})$. To do this we will simulate new data which follows the DGP under $H_0$. The p-value is then the (estimated) probability of the test statistic exceeding the calculated value $\hat{\tau}$ if the hypothesis $H_0$ were true.

Again following Davidson and MacKinnon (2004) we denote with stars, values that are calculated from simulated data, $B$ denotes the number of simulated samples. We can then estimate the cumulative distribution function as

$$\hat{F}^*(x) = \frac{1}{B} \sum_{j=1}^{B} I(\tau_j^* \leq x) \text{ and } \hat{F}^*(\hat{\tau}) = \frac{1}{B} \sum_{j=1}^{B} I(\tau_j^* \leq \hat{\tau})$$

Resultingly

$$\hat{p}^*(\hat{\tau}) = 1 - \hat{\hat{F}}^*(\hat{\tau}) = 1 - \frac{1}{B} \sum_{j=1}^{B} I(\tau_j^* \leq \hat{\tau}) = \frac{1}{B} \sum_{j=1}^{B} I(\tau_j^* > \hat{\tau})$$

In the following two specifications of the bootstrap are considered. Firstly, the often used resampling of the residuals is applied. Secondly the wild bootstrap is used.

Resampling the residuals uses the definition of the factor equation (1) to generate new data. The innovation term $e_t$ is replaced with resampled residuals[20] gotten from the estimation of the model using the initial data.

$$X_t^* = \hat{f}_t \hat{\Lambda}' + \hat{e}_t^* \tag{19}$$

The wild bootstrap serves to deal with heteroskedasticity. The idea is to rescale the

---

[20]Resampling is done with replacement to model the empirical distribution. So the resampled residuals $e_t^*$ are given by setting $\hat{e}_t^* = \hat{e}_s$ where $s$ is drawn from $\{1, ..., T\}$ with replacement.

residuals using a white noise random variable $\nu_t$.

$$X_t^* = \hat{f}_t \hat{\Lambda}' + \hat{e}_t^+$$

Where $\hat{e}_t^+$ is set such that to $\hat{e}_{it}^+ = \hat{e}_{it} \nu_{it}$ and $\nu_{it}$ is distributed with mean 0 and variance 1. For the application I choose $\nu$ to be standard normal.

Tables with size and power properties can be found in the appendix as tables 11, 12, 13, 14. The tables show modest size improvements especially for small samples and the Wald and LR statistics which are slightly oversized in their regular versions. The power properties of the bootstrapped tests are quite similar to the regular versions of the tests although it could be argued that in small samples the bootstrapped versions outperformed the normal versions. Especially the small sample size results of the wild bootstrap seem to be improved. This could be an artifact of the relatively small number of repretitions. The results are not encouraging enough as to warrant an increase of the number of repetitions especially in light of the high computational burden although it would be interesting to see if the size properties of the wild bootstrap remain as favourable.

# 3   Empirical application

In this section the issue of forecasting economic time-series in the presence of structural breaks is approached. Specifically, focus is put on forecasting German GDP using German und US American macroeconomic time-series. The reason for including American time-series to predict German GDP is twofold. Firstly, Germany is active on international markets with the US being the second most important export market after France. Coupled with the big importance of the US American market in the world economy it is plausible that US data plays a role in predicting German economic development. Secondly, we are interested in forecasting time-series under structural breaks. There are at least two periods commonly associated with breaks in the world economy: the dot-com bubble of 2000 and the financial crisis of 2008, both of which originated in the US. The hope is that big ruptures in the world economy make a break of relationships between variables (and ultimately of the factor loadings) providing us with

meaningful data for the following analysis.

There are several ways to address the issue of structural breaks from a forecasting viewpoint. Firstly structural breaks can be ignored. As has been stated above, in this case factor models will tend to choose a higher number of factors than if no structural breaks are present which means that the space the factors span is of a higher dimension as the factors are orthogonal to each other and thus linearly independent.[21]

Secondly, structural breaks can be included in the model in the form of a dummy parameter interacting with the factor estimations $\hat{f}_t$. The forecasting equation of the model then results as

$$y_{t+h}^h = \alpha(L)W_t + \beta(L)\hat{f}_t + \gamma(L)\hat{f}_t^* + \varepsilon_{t+h} \tag{20}$$

where $\hat{f}_t^*$ is defined as above.

Thirdly, if predictors are to be targeted prior to estimating the factor equation it is conceivable to find that a subset of the predictors $\tilde{X}_t^{(1)} \subset X_t$ peforms well before and a different subset $\tilde{X}_t^{(2)} \subset X_t$ performs better after the structural break. This finding would imply that it could be beneficial to estimate the factor equation again using only the data after $T^*$ and discarding the rest.

## 3.1   Data choice and treatment

The data used for the horse race has been collected from the German Bundesbank (German central bank) and from the American Federal Reserve Bank of St. Louis.[22] Together the data set consists of $T = 82$ observations from $N = 169$ quarterly (or casted to quarterly) variables. The data spans from 1993-01 until 2013-04. The series are listed below in appendix F. The data set contains macroeconomic variables such as wages and salaries, employment levels, survey data as well as financial variables from Germany and the US.

Additionally an emphasis has been put on selecting several time-series which are well known to be subjected to big changes in the abovementioned periods such as "Housing

---

[21]How this affects the information criteria of Bai and Ng (2002) can be seen in table 1 of Breitung and Eickmeier (2011).

[22]More specifically the FRED2 public API has been used to extract selected series. The data from the Bundesbank have been scraped from the Bundesbank website.

starts".

The data series have been downloaded in seasonally adjusted form where appropriate. The data series have then been differenced (zero times, once or twice) until stationarity. To this end the number of differencing operations necessary to make the series stationary has been tested using the KPSS method developed in Kwiatkowski et al. (1992).[23]

Finally, the series have been demeaned and normalized such that $\frac{1}{T}\sum_{t=1}^{T} X_{it} = 0 \ \forall i$ and $\frac{1}{T}\sum_{t=1}^{T} X_{it}^2 = 1 \Leftrightarrow \sum_{t=1}^{T} X_{it}^2 = T \ \forall i$. This is achieved by subtracting from each series its sample average and dividing by its sample standard deviation.

## 3.2 Model selection and evaluation

This section explores the different methods to selected and evaluate a model. The ultimate goal is always to forecast the true data as closely as possible. To this end I use the root mean squared forecasting error (RMSE) as an evaluation mechanism. $RMSE = \sqrt{\frac{1}{w}\sum_{s=t}^{t+h}[(y_s - \hat{y}_s)^2]}$ where $w$ denotes the length of the window to be forecasted and $t$ the period preceding the window being forecasted. The estimates $\hat{y}_t$ are gotten from estimating the forecasting equation and predicting $h$ steps into the future.

The IC$_{p2}$ criterion finds 4 static factors for the complete series. The scree plot depicted in figure 1 shows that the size of the eigenvalues drops of at approximately 4 factors which makes that choice seem appropriate in the sense that the factors which explain the biggest portion of the variance are contained in the first 4 factors..

The first factor is most strongly correlated with variables related to total US output and manufacturing. Namely the $X_i$ with a large absolute correlation with the first factor are "Industrial Production Index", "Industrial Production: Durable Manufacturing (NAICS)", "Industrial Production: Manufacturing (NAICS)", "ISM Manufacting: PMI Composite Index", "ISM Manufacturing Production Index" and "Capacity Utilization:

---

[23]The method consists of constructing an LM-test for the hypothesis that the variance $\sigma_\varepsilon^2$ is 0 (or equivalently that $\mu_t$ is constant) in

$$y_t = \beta' D_t + \mu_t + u_t$$
$$\mu_t = \mu_{t-1} + \varepsilon_t$$

where $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$ and $D_t$ captures deterministic components and time trends. This implies that the difference of $y_t$ denoted $\Delta y_t$ is stationary.
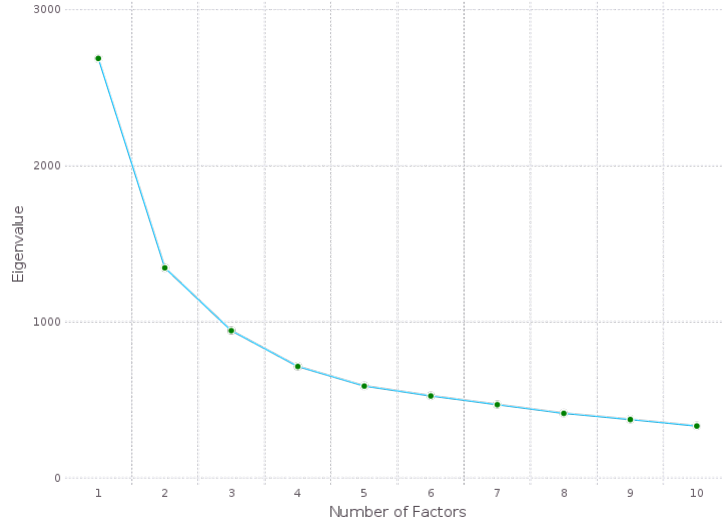
Figure 1: Scree plot of the eigenvalues calculated on the whole data set

Total Industry". The second factor mostly affects variables dealing with US construc-
tion and private housing: "Housing Starts in Midwest Census Region", "Housing Starts:
Total: New Privately Owned Housing Units Started", "New Privately-Owned Housing
Units Authorized by Building Permits: Total" and "All Employees: Construction". The
second factor is also correlated with some output related variables in Germany such as
"Output in the production sector", "Orders received" but also "Orders received" of the
German construction industry. For the third and the fourth factors the correlations are
less strong in absolute variables but one can perhaps say that they are mostly correlated
to hours worked and total employment variables in the US.

All in all, these results reflect the relative importance of US business cycles for the
chosen set of macroeconomic variables which can on the one hand be explained by the
fact that 120 of the 169 variables are composed of US macroeconomic variables but
on the other hand due to the relative sizes of the economies it is more likely that US
variables affect German variables and thus that they have a higher influence on the
underlying factors driving the evolution of the set of $X_i$.

If the number of static factors in the forecasting equation is predetermined using
criteria defined above, the number of lags of the variable of interest $y_t$ and the number
of lags of the static factors are still open. In the following I refer to the number of lags

31

of $y_t$ as $s$.

I follow Bai and Ng (2008) in selecting among model candidates using the BIC and the FPE criterion which are of course in-sample criteria.

Bai and Ng (2008) write the BIC and FPE criteria as follows

$$
\begin{aligned}
\text{BIC} &= \log(\hat{\sigma}_n^2) + n\frac{\log T}{T} \\
\text{FPE} &= \log(\hat{\sigma}_n^2) + n\frac{2}{T}
\end{aligned}
\tag{21}
$$

Where $\hat{\sigma}_n^2$ is the residual variance $\hat{e}'e$ from an estimation of the factor equation using $n$ predictors.

Since the point of the exercise is forecasting, additionally, the optimal model in terms of out-of-sample forecasting performance is found. This is done by calculating the root mean squared error of one step ahead pseudo out-of-sample forecasts.[24] This evaluation method follows Forni et al. (2005) and Bai and Ng (2008) among others. The foreacasting window $w$ over which the pseudo out-of-sample-forecasts are performed is set to $w = 15$ here.

This forecasting window makes sure that there are more observations than parameters for all possible models. Since the length of the data series is $T = 82$ this leaves 67 observations for the first forecast in the pseudo out-of-sample procedure. The maximum number of factors is set to 15 for static models and 10 for dynamic models. The maximum number of lags of $y_t$ is set to 10. This brings the maximal number of parameters to $n_{max} = 25$ for static models and $n_{max} = 50$ for dynamic models. The original number of predictors is $N = 169$ which highlights that factor models can be used as a dimensionality reduction method and that they are are an appropriate way to deal with the situation of $N >> T$ in this data set.

All models are compared to a benchmark model which consists of a simple moving average of the $w$ periods before period $t + h$ which is being forecasted. The window size for the moving average benchmark is the same as the forecasting window $w$.

---

[24]Pseudo out-of-sample forecasts are a method to simulated a forecasting situation in that the forecaster does as if he would not know the true realizations of the data after $t$. The forecaster tries to predict the value $y_{t+h}$ given all the information up to $t$. $t$ is then successively moved one period ahead. Here we set $h = 1$.

The first panel in table 1 shows the results of the in sample selection procedure for a static factor model. Three variants for choosing the number of factors are considered. Firstly the IC$_{p2}$ criterion from Bai and Ng (2002) is used to estimate the number of static factors for both the factor equation and the forecasting equation. The information criteria (BIC and FPE) are then used to select $s$, the number of lags of $y_t$. As a second choice, the decision over the number of factors *and* the number of lags of $y_t$ in the forecasting equation are left to the information criteria. I.e. the information criteria are minimized over different choices of $r$ and $s$. The same number of factors is used in the factor equation. A third possibility consists of leaving the choice of the number of factors for the factor equation to a Bai and Ng (2002) criterion and using the BIC or FPE criterion to choose the number of static factors in the forecasting equation along with the number of lags of $y_t$. I.e. the BIC or FPE are minimized over $r$ and $s$.

Both the BIC and FPE as well as the $IC_{p2}$ criteria select a large number of factors and are generally very erratic in their choice of a model. The fact that conventional information criteria should not be used for selecting the number of static factors has been extensively documented in the factor model literature and is no surprise. Another approach to model selection for forecasting consists of testing the model in an out-of-sample forecasting situation and selecting the best performing model in terms of the forecasting error. In a forecasting application this would imply that a certain fraction of the data has to be used for

The results of optimal forecasting model are reported in the second panel of table 1.

Since the variance of the data series has been normalized to 1 and the mean to 0 the RMSE of the simple average $y_t$ in the given window can be expected to be close to 1. The RMSE of the resulting out-of-sample forecasts is very large with 1.30 and 1.22 respectively. The benchmark model of a moving average reports an RMSE of 0.86 and an AR(6) model achieves an RMSE of 0.50 for the same time horizon.[25]

The best performing static factor model in terms of forecasting can be got if the BIC criterion for model selecting is replaced by the RMSE as can be seen in the second panel. I.e. the RMSE is minimized over choices of $r$ and $s$ rather than the BIC. This implies

---

[25]$p = 6$ is the optimal lag number in terms of RMSE.

using the out-of-sample performance as an evaluation criterion. In that case both $r$ and $s$ are shrunken to 1. While this outperforms the simple benchmark model in form of the moving average, the fact that the most parsimonious static factor model performs best in out-of-sample prediction hints at possible performance increases if dynamic effects are allowed. It is interesting to consider why the AR(6) model outperforms the static factor model with an RMSE of about 0.50 compared to 0.71. The static factor model does not allow for a dynamic interaction between the factors and the dependent variable. In other words the factors can only influence the $X_i$ contemporaneously. As will be seen below, the dynamic factor model can improve on this result by allowing non-contemporaneous interaction.

Table 1: Static factor model, model selection

|  | $r_{factor}$ | $r_{forecast}$ | $s$ | RMSE |
|---|---|---|---|---|
| | *In-Sample* | | | |
| BIC | 13 | 13 | 5 | 1.29 |
| FPE | 15 | 15 | 5 | 1.23 |
| $IC_{p2}$ & BIC | 4 | 13 | 5 | 1.30 |
| $IC_{p2}$ & FPE | 4 | 15 | 5 | 1.23 |
| $IC_{p2}$ | 4 | 4 | 10 | 1.21 |
| | *Out-of-Sample* | | | |
| RMSE | 1 | 1 | 1 | 0.71 |
| RMSE & $IC_{p2}$ | 4 | 1 | 1 | 0.72 |

**Notes:** The results are derrived by minimizing the criteria in the first column. If both an Bai and Ng (2002) criterion and a standard information criterion is given the former is applied to the factor equation and the latter to the forecasting equation. The maximum number of factors $r$ is set to 15. The maximum number number of the lags of the variable of interest $s$ is set to 10.

Repeating the exercise but allowing for lags of the factors to enter the forecasting equation, the results in table 2 are obtained.[26] Since the data used for the exercise consists of time-series it makes sense that a dynamic factor model is better suited for prediction here. In the case of the dynamic model the FPE criterion readily selects too many parameters which results in a higher RMSE. The $IC_{p2}$ criterion chooses $r = 4$

---

[26]Essentially, table 1 uses the same models as are used in table 2 with the added restriction $q = 0$ and the fact the maximal number of factors has been reduced to 10 because the factors enter the forecasting equation also in lagged form now which increases the number of parameters of the model quickly.

for the factor equation. For the forecasting equation, however, it seems to be better to increase the number of factors if the factors are also lagged in the forecasting equation. If the choice of the number of factors in both the factor and the forecasting equation is left to the $IC_{p2}$ which sets $r = 4$, the RMSE increases to about 0.70 as compared to 0.55 if only 2 factors, 3 lags of the dependent variable and 1 factor lag is chosen.

Also in this situation it does not necessarily make a difference in terms of RMSE if the number of factors differs in the factor equation and in the forecasting equation. This can be seen as in the optimal out-of-sample models they differ in both the static and the dynamic model while the RMSE stays the same. This no surprise as factor models relying on principal component estimation contain the most information in terms of variance in the first few factors.

In both the static and the dynamic factor model, choosing the number of static factors $r_{factor}$ in the factor equation using the $IC_{p2}$ criterion outperforms the other methods in terms of forecasting accuracy. Looking at table 2 it becomes clear why this is the case. Comparing panels a) and b) it is evident that the maximum number of factors, factor lags and lags of the dependent variable has to be chosen somewhat conservatively or else the model with the maximal number of factors and factor lags will be chosen. This is because the residuals are very small in these cases. If the number of factors is set by the $IC_{p2}$ criterion, the residuals are set "exogenously" to be big enough for the information criteria to perform well.

|  | $r_{factor}$ | $r_{forecast}$ | $s$ | $q$ | RMSE |
|---|---|---|---|---|---|
| | *In-Sample* | | | | |
| BIC | 2 | 2 | 3 | 1 | 0.55 |
| FPE | 10 | 10 | 10 | 3 | 2.91 |
| $IC_{p2}$ & BIC | 4 | 2 | 3 | 1 | 0.55 |
| $IC_{p2}$ & FPE | 4 | 10 | 10 | 3 | 2.91 |
| $IC_{p2}$ | 4 | 4 | 10 | 2 | 0.70 |
| | *Out-of-Sample* | | | | |
| RMSE | 7 | 7 | 1 | 3 | 0.47 |
| RMSE & $IC_{p2}$ | 4 | 7 | 1 | 3 | 0.47 |

(a) Maximum number of factor lags $q$ set to 3

|  | $r_{factor}$ | $r_{forecast}$ | $s$ | $q$ | RMSE |
|---|---|---|---|---|---|
| | *In-Sample* | | | | |
| BIC | 10 | 10 | 10 | 4 | 19.16 |
| FPE | 10 | 10 | 10 | 3 | 19.16 |
| $IC_{p2}$ & BIC | 4 | 10 | 10 | 4 | 19.16 |
| $IC_{p2}$ & FPE | 4 | 10 | 10 | 4 | 19.16 |
| $IC_{p2}$ | 4 | 4 | 10 | 4 | 0.78 |
| | *Out-of-Sample* | | | | |
| RMSE | 7 | 7 | 1 | 3 | 0.47 |
| RMSE & $IC_{p2}$ | 4 | 7 | 1 | 3 | 0.47 |

(b) Maximum number of factor lags $q$ set to 4

**Notes:** The results are derrived by minimizing the criteria in the first column. If both a Bai and Ng (2002) criterion and a standard information criterion is given the former is applied to the factor equation and the latter to the forecasting equation. The maximum number of factors $r$ is set to 10. The maximum number number of the lags of the variable of interest $s$ is set to 10. $q$ here refers to the number of lags of the factors in the forecasting equation, not to the number of dynamic factors as defined above. The maximum number of the factor lags $q$ is set to 3 and 4 respectively. The results using the traditional information criteria hinge on the maximum number of factor lags.

Table 2: Dynamic factor model, model selection

## 3.3 Introducing structural breaks

Structural breaks are estimated using Quandt-Andrews type supremum tests as defined above. The critival values for the test statistics are given in table I of Andrews (2003). The number of estimated breaks per period can be seen in table 3.

It is apparent that there are two periods with major changes in the loadings. Firstly, there are structural breaks reported in the loadings of 8 variables in the first quarter of 1999 which coincides approximately with the Russian default of 1998 and maybe the aftermath of the Asian financial crisis in 1997. In the following this period will be referred to as the first cluster of structural breaks. Secondly, a total of 22 changes in the loadings are reported for the period of the second quater of 2008 until the second quarter of 2009 which coincides with the recent financial crisis starting after an US housing bubble burst. This period will be referred to as the second cluster of structural breaks. It caused and was affected by severe macroeconomic changes which are quite obviously reflected in the given data set.

Variables with breaks in the periods around 2008 include several aggregate variables which deal with orders received by German companies, variables describing productivity and total output, variables describing the monetary base and loans given by banks, the consumer price index, the market rates for treasury bills of differing maturities and the "University of Michigan Consumer Sentiment" survey. These are all variables which can be expected to undergo structural breaks during the 2008 financial crisis. Eminently, "PERMITNSA" which counts the number of new privately-owned housing units authorized, undergoes a structural break in the third quarter of 2005. This is also the time when the housing bubble in the US was arguably at its peak in terms of housing price increases: Bernanke (2010) stresses that "[...] the most rapid price gains were in 2004 and 2005, when the annual rate of house price appreciation was between 15 and 17 percent".

Variables with breaks in the loadings at the first quarter of 1999 include German government debt, German exports and other trade related variables as well as variables describing bank term sheets and additionally, employment in the US durable goods

sector.

Table 3: Number of structural breaks per period

| 96:1 | 96:3 | 97:2 | 97:3 | 98:2 | 98:4 | 99:1 | 99:2 | 00:1 | 00:2 | 00:3 | 01:1 | 01:2 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | 1 | 2 | 1 | 1 | 1 | 8 | 1 | 1 | 2 | 1 | 3 | 1 |

| 03:2 | 04:3 | 04:3 | 05:1 | 05:2 | 05:3 | 07:1 | 07:2 | 07:3 | 07:4 | 08:1 | 08:2 | 08:3 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 4 | 2 |

| 08:4 | 09:1 | 09:2 | 09:3 | 09:4 | 10:1 |
|---|---|---|---|---|---|
| 7 | 7 | 2 | 1 | 1 | 3 |

***Notes:*** The entries show how many of the $N = 169$ variables are predicted to undergo a structural break in each period. Reported are the numbers of significant values of $\mathscr{S}_{i,T}$ at the 5% significance level. The total number of structural breaks found in the 169 variables is 64.

Figure 2: Structural breaks per quarter

(a) 5% significance level

(b) 1% significance level
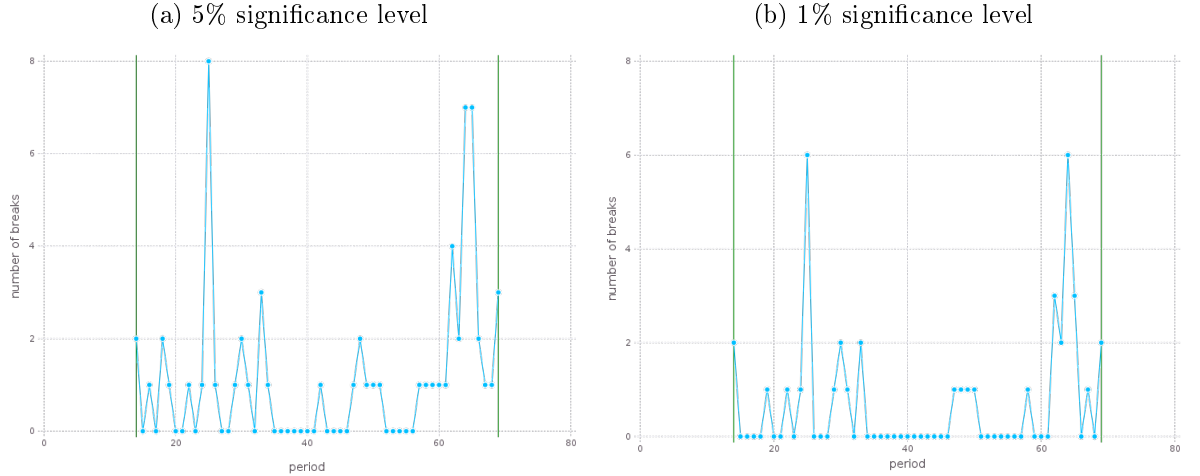


### 3.3.1   Removing variables which undergo structural breaks

In the following the question of how to react to the presence of structural breaks is approached as in section 2 above. Structural breaks in the factor loadings influence the performance of the forecasting equation through at least two channels. Firstly the presence of structural breaks decreases the signal to noise ratio in the data and thus the quality of the affected predictors. Secondly, Breitung and Eickmeier (2011) show that it is necessary to increase the number of factors $r$ under the presence of structural breaks in order to estimate the space spanned by the factors. If the number of static factors in the forecasting equation is set equal to the number of static factors in the

factor equation it becomes necessary to estimate a bigger number of parameters which can be an issue of practical relevance as has been seen above.

Either the length of the series is reduced by estimating the model only on data which is not subjected to structural breaks or the width of the data is reduced by removing the variables which are subjected to the structural break or a combination thereof. For the purpose of this paper we are interested in forecasting performance of factor models which ignore structural breaks versus factor models which take the structural breaks into account. Thus the results of models which restrict the amount of data to reduce the effect of structural breaks are compared with those that do not. It is worth noting that the comparison is not perfect because as has been argued in section 2 it can be better to reduce the number of predictors to a set which is "targeted" towards the prediction of the variable of interest. It is possible that removing the variables with structural breaks "accidentally" targets the set of variables towards predicting the variable of interest independent of the structural break. In other words an improved result is not necessarily only due to the structural breaks being removed but might be due to a noisy predictor being removed. To be precise a series with a structural break could be considered a noisy predictor as well. However, it might be noisy independent of the structural breaks. Resultingly, a robustness check consists of targeting the set of variables for both approaches prior to the estimation to remove that effect.

A practical difficulty has to be addressed before the results are presented. Given the fact that there are two periods where a majority of structural breaks occur, each of which affects a different set of factor loadings, cutting off the values before the significant break periods would leave very few observations for estimating the new model.

If both periods with major structural breaks are treated in a way that only data after the structural breaks is used, this would leave at most 17 observations of which another 15 have to be removed for the first step in the pseudo out-of-sample forecasts.

In principle there is a simple way to overcome this issue. It can be remembered that some of the data columns are monthly series which have been merged into the quarterly data set by dropping two thirds of the observations. The data could be cast back into monthly form. Most of the series will have missing data which can be interpolated (e.g. GDP is only reported quarterly). While this method solves the problem in a technical

sense[27], it has several drawbacks: firstly no additional information is added by the interpolation although there are more observations available to the model. At the same time the imputation adds additional uncertainty. Traditional standard error estimates do not account for this and resultingly they may be downward biased (cf. Gelman and Hill (2006, chapter 25). It is thus likely that prediction based on this short data set performs badly in a forecasting sense. On the other hand, if predictions resulting from this shaved data set should perform better, it would constitute a strong indication that the structural breaks uncovered by the tests should not be ignored. In this case doing so clearly leads to a significant degradation of forecasting results. This has not been tested here.

Instead four approaches to removing the effects of structural breaks from the data are considered. First the variables in the second cluster of structural breaks between the second quarter of 2008 and the first quarter of 2009 are removed. Because this approaches ignores the first cluster of structural breaks in the first quarter of 1999, a second model is estimated to explore what happens if these series are excluded as well. A third model removes all series with structural breaks in the loadings. Finally, the fourth model leaves the *variables* with breaks in the first cluster in 1999 in the model but removes *observations* prior to the second quarter of 1999.

For all four approaches the best performing dynamic factor model in terms of RMSE is reported where the number of static factors in the factor equation is calculated using the $IC_{p2}$ criterion and the number of factors in the forecasting equation as well as the number of factor lags is chosen according to the minimum of the RMSE of out-of-sample forecasts. Structural breaks are detected at the 5% significance level.

The figures in appendix D show the number of significant structural breaks left if the above adjustements to the data have been made. The reason to check for structural breaks again after some variables have been removed or the observations prior to the second quarter of 1999 have been removed, is that the estimation of the factor equation changes with a reduced set of variables which has an impact on the statistics of the structural break tests.

---

[27]For the first step of the pseudo out-of-sample period there are only 8 data points for estimation. For every quarterly step of the pseudo out-of-sample procedure there are an additional 3 observations. Thus, the ratio of observations to predictors improves quickly with each step of the pseudo out-of-sample prediction.

Table 4 shows that the exclusion of variables improves the RMSE considerably compared to ignoring the structural breaks. For this data set, the more variables with structural breaks at the 5% significance level are removed, the lower the RMSE becomes.

| | $r_{factor}$ | $r_{forecast}$ | $s$ | $q$ | RMSE |
|---|---|---|---|---|---|
| | *Second break cluster removed* | | | | |
| | *$T = 82, N = 149$* | | | | |
| RMSE | 7 | 7 | 1 | 3 | 0.45 |
| RMSE & $IC_{p2}$ | 3 | 7 | 1 | 3 | 0.45 |
| | *First and second break cluster removed* | | | | |
| | *$T = 82, N = 141$* | | | | |
| RMSE | 6 | 6 | 4 | 2 | 0.44 |
| RMSE & $IC_{p2}$ | 3 | 6 | 4 | 2 | 0.44 |
| | *All variables with breaks removed* | | | | |
| | *$T = 82, N = 105$* | | | | |
| RMSE | 6 | 6 | 3 | 2 | 0.40 |
| RMSE & $IC_{p2}$ | 3 | 3 | 4 | 2 | 0.39 |
| | *Second break cluster removed, length trimmed* | | | | |
| | *$T = 57, N = 149$* | | | | |
| RMSE | 1 | 1 | 5 | 1 | 0.51 |
| RMSE & $IC_{p2}$ | 4 | 1 | 5 | 1 | 0.51 |

Table 4: Dynamic factor model, data adjusted for structural breaks

Figures for the number of structural breaks per period after variables have been dropped or the data length has been trimmed can be found in appendix D. Notably figure 6 shows a modest number of structural breaks at the 5% significance level *after* all structural breaks had been removed from the data set. Iterating the procedure by removing the variables with structural breaks again increases the RMSE to 0.72 which makes clear that blindly removing all variables which contain structural breaks need not be a good idea even though the previous results in table 4 seemed to indicate that removing as many variables with structural breaks as possible has a positive effect on RMSE.

### 3.3.2 Removing variables with breaks after targeting the data

It was found before that removing all variables with structural breaks can improve the forecasting performance. To test if targetting the predictors and removing the structural breaks has a similar effect, the variables with structural breaks are removed after the

data set has been targeted. To this end the following sub section considers the residual effect of removing the structural breaks after the effect of targeting the predictors has been accounted for. Both soft and hard thresholding rules are applied.

It should be noted as well that German GDP, -the variable of interest is included in the set of dependent variables in the factor equation (1) (as is customary) and breaks are sometimes detected in the factor loading of this variable if the data set is targeted. German GDP is still kept in the data set in all variable pre-selection schemes even if all other variables with breaks are removed from the set.

Starting with hard thresholding, targeting the predictors at the 10%, 5% and 1% significance level (see Bai and Ng (2008) for details) leaves 85, 78 and 43 variables in the data set respectively. Figure 3 shows that targetting the data removes a part of the two clusters of structural breaks found in the original data. It can be inspected in table 5 that with hard thresholding the RMSE is higher than it was for the best performing model of the original data set. For the complete data set the optimal model achieved an RMSE of 0.47. with hard thresholding the RMSE increases to about 0.50 for all three cut-off values. Regardless of the relatively bad performance of hard thresholding rules to target the predictors, removing the structural breaks from the targeted data set marginally improves the results.

| | **Hard thresholding** | | | | |
| | $r_{factor}$ | $r_{forecast}$ | $s$ | $q$ | RMSE |
|---|---|---|---|---|---|
| | $\alpha = 0.1$ | | | | |
| | $T = 82, N = 85$ | | | | |
| RMSE | 1 | 1 | 5 | 2 | 0.50 |
| RMSE & $IC_{p2}$ | 6 | 1 | 5 | 2 | 0.50 |
| | Hard Thresholding $\alpha = 0.05$ | | | | |
| | $T = 82, N = 78$ | | | | |
| RMSE | 1 | 1 | 5 | 2 | 0.50 |
| RMSE & $IC_{p2}$ | 6 | 1 | 5 | 2 | 0.50 |
| | $\alpha = 0.01$ | | | | |
| | $T = 82, N = 43$ | | | | |
| RMSE | 1 | 1 | 5 | 2 | 0.49 |
| RMSE & $IC_{p2}$ | 7 | 1 | 5 | 2 | 0.49 |

Table 5: Dynamic factor model, targeted data, hard thresholding

| | Hard thresholding, breaks removed | | | | | |
|---|---|---|---|---|---|---|
| | $r_{factor}$ | $r_{forecast}$ | $s$ | $q$ | RMSE | $\Delta$ |
| | $\alpha = 0.1$ | | | | | |
| | $T = 82, N = 51$ | | | | | |
| RMSE | 11 | 11 | 2 | 1 | 0.44 | 0.06 |
| RMSE & $IC_{p2}$ | 4 | 11 | 2 | 1 | 0.44 | 0.06 |
| | $\alpha = 0.05$ | | | | | |
| | $T = 82, N = 51$ | | | | | |
| RMSE | 1 | 1 | 5 | 2 | 0.50 | 0 |
| RMSE & $IC_{p2}$ | 4 | 1 | 5 | 2 | 0.50 | 0 |
| | $\alpha = 0.01$ | | | | | |
| | $T = 82, N = 28$ | | | | | |
| | Results omitted | | | | | |

**Notes:** $T$ and $N$ denote the dimension of the targeted data sets where all variables with evidence of structural breaks found at the 5% level using the sup LM test have been removed. The column denoted with $\Delta$ calculates the difference between the RMSE of the best factor model calculated on the targeted data with structural breaks removed and the results from table 5 where the structural breaks have not been removed. Note that the set of variables for hard thresholding with $\alpha = 0.05$ differs although the number of variables coincides. Also note that there are no results for $\alpha = 0.01$ because the estimated number of factors for this data set is 22 for wich there are no tabulated values to calculate the structural break tests with in Andrews (2003). The critical values have not been simulated here.

Table 6: Dynamic factor model, targeted data, structural breaks removed



(a) 5% significance level                    (b) 1% significance level
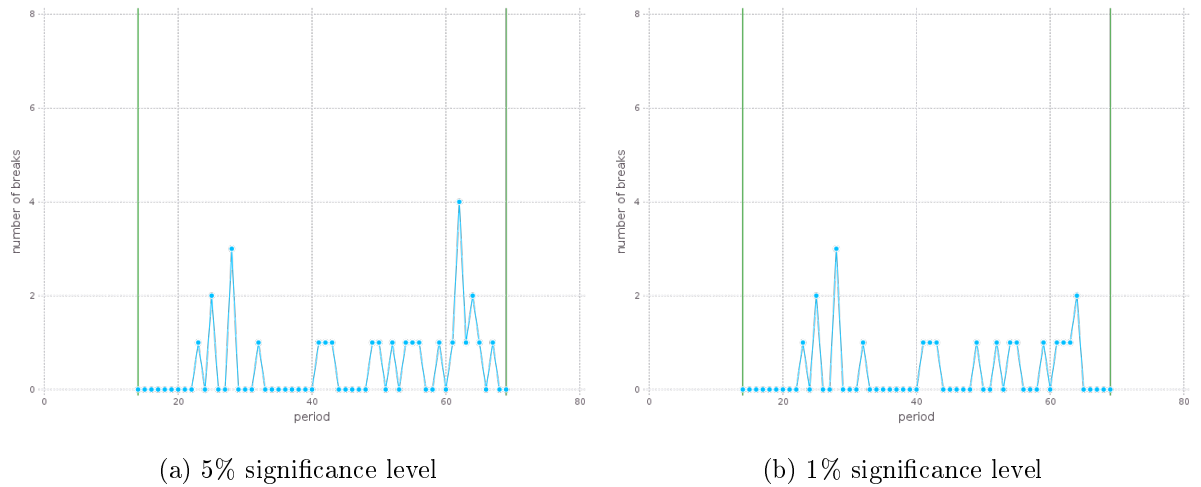
Figure 3: Structural breaks per quarter, targeted predictors

As noted in the introduction on targeted predictors above, the difficulty with soft thresholding lies in the choice of the number of steps $k$ in the LARS-EN algorithm. This coincides with the number of variables chosen as each step adds one variable to the active set. It is conceivable that if the thresholding is too strict, i.e. the regularization parameter too large, the LASSO will have kicked out most of the variables which contain structural breaks. A chosen set of results of the removal of structural breaks after applying soft thresholding using the LASSO can be seen in table 7.

The value of $k$ which yields the best performing targeted predictor model is $k = 39$ which gives an RMSE of just below 0.4 as can be inspected in table 7. The fact that the maximum number of factors is chosen is suspicious. However, if the maximum number of factors is increase the results remain constant. The same holds for $k = 30$ and for the number of factor lags.

| | Soft thresholding | | | | | |
| | $r_{factor}$ | $r_{forecast}$ | $s$ | $q$ | RMSE | $\Delta_{\text{RMSE}}$ |
|---|---|---|---|---|---|---|
| | $k = 39$ |
| | $T = 82, N = 39$ |
| RMSE | 15 | 15 | 3 | 1 | 0.4000 | $-0.0811$ |
| RMSE & $IC_{p2}$ | 2 | 15 | 3 | 1 | 0.4000 | $-0.0811$ |
| | $k = 32$ |
| | $T = 82, N = 32$ |
| RMSE | 13 | 13 | 1 | 1 | 0.4014 | $-0.0125$ |
| RMSE & $IC_{p2}$ | 2 | 13 | 1 | 1 | 0.4014 | $-0.0125$ |
| | $k = 40$ |
| | $T = 82, N = 40$ |
| RMSE | 15 | 15 | 3 | 1 | 0.4141 | $-0.0716$ |
| RMSE & $IC_{p2}$ | 2 | 15 | 3 | 1 | 0.4141 | $-0.0716$ |

**Notes:** The entries report detailed results of the three best performing optimal models for $k \in \{30, 32, 39\}$. Less detailed results for more values of $k$ can be found in table 15 in the appendix. $\Delta_{\text{RMSE}}$ denotes the change in RMSE if the optimal model is calculated from the targeted data set with variables containing structural breaks removed. The maximum number of factors is set to 15, the maximum number of factor lags is set to 3 and the maximum number of lags of $y_t$ is set to 5

Table 7: Dynamic factor model, targeted data, soft thresholding

The 39 variables contained in the best performing targeted predictor model include German output and employment variables, several variables describing German exports

and the trade balance. The US variables included contain output and employment variables, some production related financial variables such as "Commercial and Indstrial Loans", the consumer and producer price indices, construction and housing variables as well as monetary variables and the "6-Month Treasurey Bill: Secondary Market Rate". Interestingly the foreign exchange rate between Switzerland and the US is also added to the set of variables. It is likely that this variable captures the lack of other exchange rates in the data set.

The second best performing model removes some of the variables but because the second best model is nothing but an earlier point on the solution path, the other variables remain the same. The removed variables inlcude one of the German export variables, the US labor force participation rate, one of the US housing starts variables and markedly the producer price index.

Relative improvements of removing structural breaks for more values of $k$ and Diebold-Mariano test statistics[28] are shown in table 15. The table starts with $k = 26$ which is the first value in the LASSO solution path for which the resulting number of factors calculated by the $IC_{p2}$ criterion returns sensible values.[29] It can be seen that removing the variables with structural breaks can only improve the RMSE for some of the different sets of targeted data. This can be seen in the positive entries in the $\Delta_{\mathrm{RMSE}}$ column.

However, significant improvements in forecasting accuracy can only be confirmed for $k = 49$ at customary levels of significance and different forecasting performance can only be detected for 3 choices of $k$. It should also be noted that even this one rejection has to be taken with a grain of salt although significant at the 1% level. This is because if the 5% significance level is taken, at least one rejection can be expected due to chance in 25 tests and at the 1% level 1 false rejection is less likely but possible nonetheless.

---

[28]The models are estimated on different subsets of the data set. Strictly speaking the models could be seen as nested and another comparison procedure such as Clark and McCracken (2001) should be used depending on the specificity of the restriction. This corner case is ignored here.

[29]Sensible here means smaller than 20 which is the last value for which Andrews (2003) has tabulated critical values.

To further study the interaction between targeted predictors and structural breaks in a confined environment, in a final step, a monte carlo study is performed in which the simulated data contains a structural break in only one of the factor loadings. The size of the break and the information content for the prediction of $y$ is varied in one predictor. The idea is to meassure the difference in RMSE in this contained environment if the variable is removed versus keeping the variable in the model for varying sizes of the break and levels of information for predicting $y$. More precisely the Cholesky decomposition is used in order to generate data with varying degree of dependence $\delta$ between the dependent variable and the predictor which experiences a structural break of size $b$. All values of $\delta$ which are bigger than 0.1 imply that the last column is typically chosen first if the predictors are targeted. Table 8 shows that a higher $\delta$ decreases the likelihood of an improved forecasting accuracy which is intuitive because removing an informative predictor should decrease performance. It seems that an increase in the break size $b$ has the same effect but this effect seems to be less pronounced.

$$\text{DGP: } X_{it} = \sum_{j=1}^{r} \lambda_{ij} F_{tj} + \text{chol}(v(\delta))e_{it}, r = 1$$

$$\text{where } v(\delta) = \begin{bmatrix} 1 & 0.1 & \dots & \dots & 0.1 & \delta \\ 0.1 & 1 & 0.1 & \dots & 0.1 & 0.1 \\ \vdots & 0.1 & \ddots & 0.1 & \vdots & \vdots \\ \vdots & \vdots & 0.1 & \ddots & 0.1 & \vdots \\ 0.1 & \vdots & \vdots & 0.1 & \ddots & 0.1 \\ \delta & 0.1 & \dots & \dots & 0.1 & 1 \end{bmatrix}$$

|   |   | $\delta$ | | | | | | | | | |
|---|-----|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
|   |     | 0.1   | 0.2   | 0.3   | 0.4   | 0.5   | 0.6   | 0.7   | 0.8   | 0.9   | 0.99  |
| b | 0.0 | 0.040 | 0.036 | 0.018 | 0.020 | 0.024 | 0.024 | 0.020 | 0.018 | 0.014 | 0.028 |
|   | 0.1 | 0.032 | 0.024 | 0.026 | 0.026 | 0.026 | 0.008 | 0.026 | 0.010 | 0.020 | 0.004 |
|   | 0.2 | 0.026 | 0.020 | 0.020 | 0.024 | 0.018 | 0.018 | 0.018 | 0.014 | 0.014 | 0.008 |
|   | 0.3 | 0.042 | 0.040 | 0.020 | 0.008 | 0.014 | 0.020 | 0.010 | 0.024 | 0.022 | 0.002 |
|   | 0.4 | 0.040 | 0.036 | 0.020 | 0.026 | 0.020 | 0.008 | 0.010 | 0.026 | 0.010 | 0.018 |
|   | 0.5 | 0.030 | 0.028 | 0.022 | 0.016 | 0.016 | 0.018 | 0.020 | 0.018 | 0.014 | 0.008 |
|   | 0.6 | 0.040 | 0.040 | 0.014 | 0.012 | 0.004 | 0.016 | 0.030 | 0.004 | 0.008 | 0.012 |
|   | 0.7 | 0.042 | 0.036 | 0.020 | 0.028 | 0.014 | 0.020 | 0.018 | 0.016 | 0.010 | 0.012 |
|   | 0.8 | 0.036 | 0.034 | 0.024 | 0.016 | 0.024 | 0.016 | 0.018 | 0.016 | 0.006 | 0.010 |
|   | 0.9 | 0.030 | 0.034 | 0.018 | 0.014 | 0.026 | 0.014 | 0.014 | 0.014 | 0.010 | 0.012 |
|   | 1.0 | 0.028 | 0.024 | 0.016 | 0.012 | 0.020 | 0.026 | 0.012 | 0.020 | 0.020 | 0.018 |

**Notes:** The entries report the average rejection rate of $R = 500$ Diebold-Mariano tests of improved forecasting performance at the 5% significance level. The two competing models are static factor models, one of which includes the column with the structural break, the other does not. Optimal models are chosen using the RMSE as before. The structural break occurs in the mid observation $T^* = 25$ in the last column. The size of the break is given by $b$ and the correlation of the dependent variable in the first column with last column is given by $\delta$. $T = 50, N = 20$

Table 8: Montecarlo of DM-tests for varying break size and information

# 4 Conclusion

While factor models perform quite well in forecasting economic time-series, structural breaks can, if not accounted for, undermine the forecasting performance significantly. Not only does one have to be careful with the actual prediction results but also interpretting the factors directly can become unfeasible as the number of static factors changes under structural breaks.

Estimating different model specifications under structural breaks on a large macroeconomic data set including German and US macroeconomic variables shows that removing variables with structural breaks tends to improve forecasting accuracy considerably. In this data set the improvement of removing the variables which undergo structural breaks improves the performance more so than removing the pre-break observations. This result must not necessarily hold for different data sets and could be an artifact of its relative shortness.

The best performing models with breaks removed handily beat the best performing benchmark AR(p) model and the benchmark moving average model as well as dynamic factor models which do not account for breaks.

The best overall results are achieved using targeted perdictors with soft thresholding with $k = 39$ which amounts to keeping only roughly 23% of the original predictors.

The removal of variables which feature structural breaks works similarly to the method of targeted predictors. While each method individually improves forecasting precision, there is little evidence that the two methods used in conjunction further improve the forecasting results. This hints at a trade-off. The method of removing structural breaks ignores the information contained in the predictor while soft-thresholding focusses on correlatedness over the whole time-frame without taking into account that the linear relationship might break down over time. Put differently: as long as two predictors are sufficiently correlated *overall* it does not matter if the most recent data follow a different data generating process if targeted predictors are used for variable pre-selection. Further research into variable pre-selection methods for factors models could also yield a method which generalizes targeted predictors and the approach to remove variables featuring structural breaks.

It should be noted that the approach to remove all variables which have significant

structural breaks is very crude and comparatively unfair with respect to targeted data under soft thresholding because the latter consists of a whole solution path of which the best solution is picked. Alternatively and perhaps more fairly one could consider removing variables step by step starting e.g. with variables which have a high associated sup LM statistic. This is likely to yield additional gains in forecasting performance.

Additionally and independently from the treatment of structural breaks, the model selection framework can be further refined. It is likely for example that allowing factors to enter the forecasting equation with different lag levels can help to boost the forecasting performance.

# Apendix

# A    Derrivation of Principal Components

The principal components of a data matrix $X$ are defined as the transformation of $X$ into the space spanned by the loadings [30] $u_i$ where each $u_i$ is a unit vector. The transformation is defined such that the transformed columns are uncorrelated (as are the loadings $u_i$) and such that the first transformed column (i.e. the first principal component) contains the largest amount of the variance of the original data, the second column the second largest amount (i.e. the second principal component), etc.

Let $X$ have mean 0 (i.e. subtract the mean from each observation if it does not). For the first principal component, we want to find a loading vector vector $u_1$ such that the variance along the projection of $x$ onto the first principal component direction $u_1$ is maximized where $u_1$ is a unit norm vector.

$$\max_{u_1:||u_1||=1} \frac{1}{T} \sum_{i=1}^{T} (x_i' u_1)^2 = \max_{u_1:u_1' u_1=1} u_1'(\frac{1}{T} \sum_{i=1}^{T} x_i' x_i) u_1 = \max_{u_1:u_1' u_1=1} u_1'(\Sigma) u_1$$

Setting up the Lagrangian yields

$$L(u_1, \lambda) = u_1' \Sigma u_1 - \lambda(u_1' u_1 - 1)$$

After taking the derivative with respect to $u_1$ and dividing by 2 we are left with

$$\frac{\partial L}{\partial u_1} = \Sigma u_1 - \lambda u_1 \stackrel{!}{=} 0$$

in other words $u_1$ is an eigenvector of $\Sigma$, the covariance matrix of x. The corresponding eigenvalue is $\lambda$ and we must have $\Sigma = \lambda$ for the variance of the principal component to be maximal under the constraint.

The maximal variance of $x' u_1$ is thus $V(x' u_1) = u_1' \Sigma u_1 = u_1' \lambda u_1$ and it is achieved for the highest eigenvalue $\lambda$ of the covariance matrix of x. The derrivation of the following principal components is accordingly subject to the additional constraint that the

---

[30]The $u_i$ are also called scores or principal component directions.

principal components are uncorrelated. In the case of the second principal component this implies $\mathrm{cov}(u_1'x, u_2'x) = u_1'\Sigma u_2 = 0$. The solution for the third and all following principal component is more complex but goes accordingly. The derrivations can be seen e.g. in Jolliffe (2005) or any other textbook on the subject. The solution for the $i$-th principal component is to set the loadings $u_i$ to be eigenvector which belongs to the $i$-th highest eigenvalue. Also, for each principal component $i$ we get that $\mathrm{var}(u_i'x) = \lambda_i$.

# B    Estimation of the Factor Equation

Estimators of the factor equation solve (10).

$$\min_{F_1,\ldots,F_T,\Lambda} V_r(\Lambda, F) = \min_{F_1,\ldots,F_T,\Lambda} \frac{1}{NT} \sum_{t=1}^{T} (X_t - \Lambda F_t)'(X_t - \Lambda F_t)$$

$$\text{s.t. } N^{-1}\Lambda'\Lambda = I_r$$

The solution can be gotten by first optimizing over $F_t$. The first order condition is:

$$\frac{\partial V_r}{\partial F_t} = -2X_t'\Lambda + 2\hat{F}_t'\Lambda'\Lambda \overset{!}{=} 0$$

Resultingly $\hat{F}_t = (\Lambda'\Lambda)^{-1}\Lambda' X_t$. Inserting that back into the optimization problem gives $\min_\Lambda \frac{1}{NT}X_t'X_t - 2X_t'\Lambda(\Lambda'\Lambda)^{-1}\Lambda'X_t + X_t'\Lambda(\Lambda'\Lambda)^{-1}\Lambda'\Lambda(\Lambda'\Lambda)^{-1}\Lambda'X_t = \min_\Lambda \frac{1}{NT}X_t'[I - \Lambda(\Lambda'\Lambda)^{-1}\Lambda']X_t$.

This is equivalent to the problem $\max_\Lambda \text{tr}\left\{(\Lambda'\Lambda)'^{-\frac{1}{2}}\lambda'(T^{-1}\Sigma_{t=1}^{T}X_tX_t')\Lambda(\Lambda'\Lambda)^{-\frac{1}{2}}\right\}$ which is the same as $\max_\Lambda \Lambda'\hat{\Sigma}_{XX}\Lambda$ s.t. $N^{-1}\Lambda'\Lambda = I_r$ where $\hat{\Sigma}_{XX}$ is the usual sample equivalent of the variance of $X$. This can be solved by setting $\hat{\Lambda}$ to eigenvectors of the $r$ largest eigenvalues of $\hat{\Sigma}_{XX}$.

# C  Replication of Bai and Ng (2002) for $r = 7$, $r = 9$

Table 9: Replication of Table I of Bai and Ng (2002) for $r = 7$

DGP: $X_{it} = \sum_{j=1}^{r} \lambda_{ij} F_{tj} + \sqrt{\theta} e_{it}, r = 7$

| N | T | $PC_{p1}$ | $PC_{p2}$ | $PC_{p3}$ | $IC_{p1}$ | $IC_{p2}$ | $IC_{p3}$ |
|---|---|---|---|---|---|---|---|
| 100 | 40 | 6.4 | 5.9 | 6.97 | 4.93 | 3.46 | 6.73 |
| 100 | 60 | 6.72 | 6.27 | 7.02 | 6.34 | 4.74 | 6.99 |
| 200 | 60 | 6.93 | 6.87 | 7.0 | 6.78 | 6.46 | 6.96 |
| 500 | 60 | 6.99 | 6.96 | 7.0 | 6.94 | 6.92 | 6.99 |
| 1000 | 60 | 7.0 | 6.99 | 7.0 | 6.99 | 6.97 | 7.0 |
| 2000 | 60 | 7.0 | 6.99 | 7.0 | 6.98 | 6.99 | 7.0 |
| 100 | 100 | 7.0 | 6.77 | 7.35 | 6.89 | 6.32 | 50.0 |
| 200 | 100 | 7.0 | 7.0 | 7.0 | 7.0 | 6.99 | 7.0 |
| 500 | 100 | 7.0 | 7.0 | 7.0 | 7.0 | 7.0 | 7.0 |
| 1000 | 100 | 7.0 | 7.0 | 7.0 | 7.0 | 7.0 | 7.0 |
| 2000 | 100 | 7.0 | 7.0 | 7.0 | 7.0 | 7.0 | 7.0 |
| 40 | 100 | 6.39 | 6.12 | 7.02 | 4.86 | 3.49 | 6.78 |
| 60 | 100 | 6.79 | 6.37 | 7.01 | 6.01 | 5.0 | 7.0 |
| 60 | 200 | 6.93 | 6.85 | 7.0 | 6.82 | 6.58 | 6.98 |
| 60 | 500 | 6.96 | 6.97 | 7.0 | 6.97 | 6.9 | 6.99 |
| 60 | 1000 | 7.0 | 6.99 | 7.0 | 6.99 | 6.96 | 7.0 |
| 60 | 2000 | 7.0 | 7.0 | 7.0 | 6.97 | 6.98 | 6.99 |
| 4000 | 60 | 7.0 | 7.0 | 6.99 | 6.98 | 6.99 | 6.99 |
| 4000 | 100 | 7.0 | 7.0 | 7.0 | 7.0 | 7.0 | 7.0 |
| 8000 | 60 | 6.99 | 7.0 | 7.0 | 7.0 | 6.98 | 7.0 |
| 8000 | 100 | 7.0 | 7.0 | 7.0 | 7.0 | 7.0 | 7.0 |
| 60 | 4000 | 7.0 | 7.0 | 7.0 | 7.0 | 7.0 | 6.99 |
| 100 | 4000 | 7.0 | 7.0 | 7.0 | 7.0 | 7.0 | 7.0 |
| 60 | 8000 | 6.99 | 7.0 | 7.0 | 6.99 | 7.0 | 6.99 |
| 100 | 8000 | 7.0 | 7.0 | 7.0 | 7.0 | 7.0 | 7.0 |
| 10 | 50 | 5.0 | 5.0 | 5.0 | 4.76 | 3.95 | 5.0 |
| 10 | 100 | 5.0 | 5.0 | 5.0 | 4.58 | 4.18 | 4.96 |
| 20 | 100 | 7.04 | 6.5 | 7.98 | 3.14 | 2.07 | 6.94 |
| 100 | 10 | 5.0 | 5.0 | 5.0 | 5.0 | 4.96 | 5.0 |
| 100 | 20 | 7.1 | 6.64 | 8.05 | 3.3 | 2.08 | 8.2 |

**Notes:** Estimated number of factors averaged over 1000 simulations. The true number of factors is $r$ and the maximum number of factors is ceil(min$\{T, N\}$)/2).

Table 10: Replication of Table I of Bai and Ng (2002) for $r = 9$

DGP: $X_{it} = \sum_{j=1}^{r} \lambda_{ij} F_{tj} + \sqrt{\theta} e_{it}, r = 9$

see Bai and Ng (2002) for specification of the DGP

| N | T | $PC_{p1}$ | $PC_{p2}$ | $PC_{p3}$ | $IC_{p1}$ | $IC_{p2}$ | $IC_{p3}$ |
|---|---|---|---|---|---|---|---|
| 100 | 40 | 6.51 | 5.93 | 8.06 | 3.55 | 1.49 | 7.89 |
| 100 | 60 | 7.02 | 6.25 | 8.77 | 5.4 | 2.74 | 8.9 |
| 200 | 60 | 7.74 | 7.31 | 8.65 | 7.33 | 6.26 | 8.74 |
| 500 | 60 | 8.09 | 8.02 | 8.52 | 8.18 | 7.97 | 8.58 |
| 1000 | 60 | 8.42 | 8.38 | 8.59 | 8.45 | 8.25 | 8.62 |
| 2000 | 60 | 8.47 | 8.32 | 8.59 | 8.67 | 8.56 | 8.64 |
| 100 | 100 | 8.03 | 6.93 | 9.01 | 7.68 | 4.85 | 50.0 |
| 200 | 100 | 8.85 | 8.48 | 8.99 | 8.92 | 8.58 | 9.0 |
| 500 | 100 | 9.0 | 8.99 | 9.0 | 9.0 | 8.99 | 9.0 |
| 1000 | 100 | 9.0 | 8.99 | 9.0 | 9.0 | 9.0 | 9.0 |
| 2000 | 100 | 9.0 | 9.0 | 9.0 | 9.0 | 9.0 | 9.0 |
| 40 | 100 | 6.63 | 6.06 | 8.08 | 3.38 | 1.66 | 7.79 |
| 60 | 100 | 7.03 | 6.33 | 8.85 | 5.51 | 2.68 | 8.94 |
| 60 | 200 | 7.79 | 7.43 | 8.6 | 7.1 | 6.05 | 8.73 |
| 60 | 500 | 8.16 | 7.98 | 8.61 | 8.28 | 7.96 | 8.66 |
| 60 | 1000 | 8.41 | 8.28 | 8.51 | 8.51 | 8.34 | 8.69 |
| 60 | 2000 | 8.4 | 8.4 | 8.58 | 8.57 | 8.56 | 8.74 |
| 4000 | 60 | 8.52 | 8.44 | 8.57 | 8.61 | 8.51 | 8.73 |
| 4000 | 100 | 9.0 | 9.0 | 9.0 | 9.0 | 9.0 | 9.0 |
| 8000 | 60 | 8.44 | 8.52 | 8.55 | 8.66 | 8.64 | 8.68 |
| 8000 | 100 | 9.0 | 9.0 | 9.0 | 9.0 | 9.0 | 9.0 |
| 60 | 4000 | 8.53 | 8.54 | 8.55 | 8.62 | 8.6 | 8.57 |
| 100 | 4000 | 9.0 | 9.0 | 9.0 | 9.0 | 9.0 | 9.0 |
| 60 | 8000 | 8.61 | 8.47 | 8.51 | 8.66 | 8.58 | 8.76 |
| 100 | 8000 | 9.0 | 9.0 | 9.0 | 9.0 | 9.0 | 9.0 |
| 10 | 50 | 5.0 | 5.0 | 5.0 | 4.66 | 3.29 | 5.0 |
| 10 | 100 | 5.0 | 5.0 | 5.0 | 4.32 | 3.59 | 4.77 |
| 20 | 100 | 7.25 | 6.84 | 8.1 | 1.76 | 1.28 | 7.21 |
| 100 | 10 | 5.0 | 5.0 | 5.0 | 4.96 | 4.74 | 5.0 |
| 100 | 20 | 7.37 | 7.02 | 8.28 | 1.92 | 1.25 | 8.89 |

***Notes:*** Estimated number of factors averaged over 1000 simulations. The true number of factors is $r$ and the maximum number of factors is ceil(min$\{T, N\}$)/2).

DGP: $X_{it} = \sum_{j=1}^{r} \lambda_{ij} F_{tj} + \sigma e_{it}, r = 1, \sigma_i \sim U(0.5, 1.5)$

| N | LR | LM | W | LR | LM | W |
|---|------|------|------|------|------|------|
|   | T=50 |    |    | T=100 |    |    |
| 20 | 0.0445 | 0.0445 | 0.0483 | 0.0405 | 0.0435 | 0.0443 |
| 50 | 0.0462 | 0.0456 | 0.0488 | 0.0524 | 0.0512 | 0.0508 |
| 100 | 0.0513 | 0.0516 | 0.0549 | 0.0458 | 0.0458 | 0.0475 |
| 150 | 0.0481 | 0.0487 | 0.0519 | 0.0483 | 0.0493 | 0.0487 |
| 200 | 0.0497 | 0.0498 | 0.0509 | 0.0481 | 0.0483 | 0.0472 |
|   | T=150 |    |    | T=200 |    |    |
| 20 | 0.0510 | 0.0540 | 0.0540 | 0.0515 | 0.0540 | 0.0525 |
| 50 | 0.0458 | 0.0474 | 0.0464 | 0.0504 | 0.0514 | 0.0460 |
| 100 | 0.0473 | 0.0463 | 0.0478 | 0.0472 | 0.0464 | 0.0478 |
| 150 | 0.0491 | 0.0493 | 0.0491 | 0.0499 | 0.0499 | 0.0511 |
| 200 | 0.0505 | 0.0514 | 0.0511 | 0.0489 | 0.0499 | 0.0497 |

**Notes:** Average rejection frequencies of $R = 100$ bootstrapped Wald, Lagrange multiplier and Likelihood ratio tests with $B = 100$ bootstrap repetitions. Results are averaged over the $N$ tests for each variable. The DGP features no structural break ($b = 0$). The tests are calculated at the 5% significance level.

Table 11: Size of Chow-tests, Table II of Bai and Ng (2002) Residual resampling

$$\text{DGP: } X_{it} = \sum_{j=1}^{r} \lambda_{ij} F_{tj} + \sigma e_{it}, r = 1, \sigma_i \sim U(0.5, 1.5)$$

| N | LR | LM | W | LR | LM | W |
|---|------|------|------|------|------|------|
| | T=50 | | | T=100 | | |
| 20 | 0.0251 | 0.0250 | 0.0552 | 0.0252 | 0.0351 | 0.0353 |
| 50 | 0.0522 | 0.04602 | 0.0501 | 0.0461 | 0.0483 | 0.0462 |
| 100 | 0.0421 | 0.0370 | 0.0502 | 0.0440 | 0.0481 | 0.0591 |
| 150 | 0.0420 | 0.0401 | 0.0553 | 0.0547 | 0.0547 | 0.0601 |
| 200 | 0.0435 | 0.0421 | 0.0495 | 0.0505 | 0.0502 | 0.0565 |
| | T=150 | | | T=200 | | |
| 20 | 0.0551 | 0.0496 | 0.0581 | 0.0352 | 0.0403 | 0.0593 |
| 50 | 0.0498 | 0.0574 | 0.0553 | 0.0346 | 0.0436 | 0.0374 |
| 100 | 0.0569 | 0.0466 | 0.0559 | 0.0457 | 0.0455 | 0.0487 |
| 150 | 0.0472 | 0.04464 | 0.0518 | 0.0447 | 0.0428 | 0.0453 |
| 200 | 0.0485 | 0.0432 | 0.0515 | 0.0465 | 0.0492 | 0.0555 |

***Notes:*** Average rejection frequencies of $R = 100$ bootstrapped Wald, Lagrange multiplier and Likelihood ratio tests with $B = 100$ bootstrap repetitions. Results are averaged over the $N$ tests for each variable. The DGP features no structural break ($b = 0$). The tests are calculated at the 5% significance level.

Table 12: Size of Chow-tests, Table II of Bai and Ng (2002) Wild bootstrap

$$\text{DGP: } X_{it} = \sum_{j=1}^{r} \lambda_{ij} F_{tj} + \sigma e_{it}, r = 1, \sigma_i \sim U(0.5, 1.5)$$

| b | LR | LM | W | LR | LM | W |
|---|---|---|---|---|---|---|
| | T=50 | | | T=100 | | |
| 0.1 | 0.0532 | 0.0542 | 0.0564 | 0.0642 | 0.0682 | 0.0693 |
| 0.2 | 0.0788 | 0.0784 | 0.0746 | 0.1134 | 0.1096 | 0.1112 |
| 0.3 | 0.1076 | 0.1064 | 0.1006 | 0.1598 | 0.1616 | 0.1541 |
| 0.5 | 0.1836 | 0.1806 | 0.1698 | 0.2934 | 0.2894 | 0.2828 |
| | T=150 | | | T=200 | | |
| 0.1 | 0.0721 | 0.0743 | 0.0712 | 0.0814 | 0.0804 | 0.0838 |
| 0.2 | 0.1466 | 0.1474 | 0.1442 | 0.1744 | 0.1716 | 0.1742 |
| 0.3 | 0.2298 | 0.2284 | 0.2262 | 0.2644 | 0.2644 | 0.2564 |
| 0.5 | 0.3654 | 0.3618 | 0.3662 | 0.4232 | 0.4264 | 0.4194 |

***Notes:*** Average rejection frequencies of $R = 100$ bootstrapped Wald, Lagrange multiplier and Likelihood ratio tests with $B = 100$ bootstrap repetitions. Results are averaged over the $N$ tests for each variable. The DGP features structural breaks of size $b$. The tests are calculated at the 5% significance level.

Table 13: Power of Chow-tests, Table III of Bai and Ng (2002) Residual resampling

DGP: $X_{it} = \sum_{j=1}^{r} \lambda_{ij} F_{tj} + \sigma e_{it}, r = 1, \sigma_i \sim U(0.5, 1.5)$

| b | LR | LM | W | LR | LM | W |
|---|---|---|---|---|---|---|
| | T=50 | | | T=100 | | |
| 0.1 | 0.0625 | 0.0684 | 0.0622 | 0.0921 | 0.0942 | 0.0961 |
| 0.2 | 0.0644 | 0.0562 | 0.0741 | 0.1042 | 0.1310 | 0.1142 |
| 0.3 | 0.0946 | 0.1041 | 0.1141 | 0.1583 | 0.1543 | 0.1665 |
| 0.5 | 0.1644 | 0.1622 | 0.1785 | 0.3122 | 0.3312 | 0.3264 |
| | T=150 | | | T=200 | | |
| 0.1 | 0.0581 | 0.0664 | 0.0686 | 0.1167 | 0.1041 | 0.1048 |
| 0.2 | 0.1528 | 0.1581 | 0.1684 | 0.1824 | 0.1823 | 0.1921 |
| 0.3 | 0.1981 | 0.2089 | 0.2147 | 0.2448 | 0.2389 | 0.2389 |
| 0.4 | 0.4583 | 0.3965 | 0.4117 | 0.4246 | 0.4243 | 0.4347 |

***Notes:*** Average rejection frequencies of $R = 100$ bootstrapped Wald, Lagrange multiplier and Likelihood ratio tests with $B = 100$ bootstrap repetitions. Results are averaged over the $N$ tests for each variable. The DGP features structural breaks of size $b$. The tests are calculated at the 5% significance level.

Table 14: Power of Chow-tests, Table III of Bai and Ng (2002) Wild bootstrap

# D Figures - Empirical Section



(a) 5% significance level

(b) 1% significance level
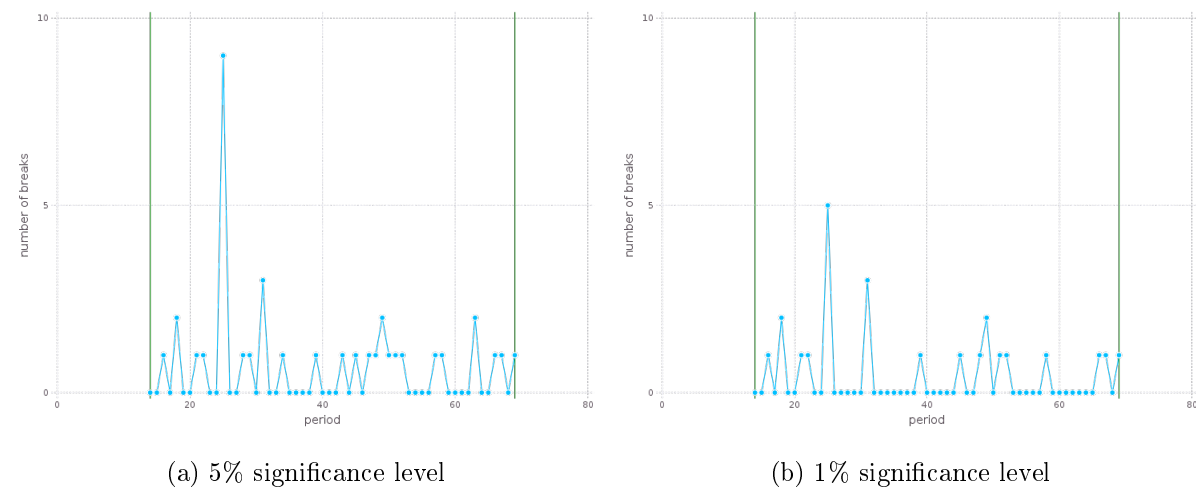
Figure 4: Second cluster removed



(a) 5% significance level

(b) 1% significance level
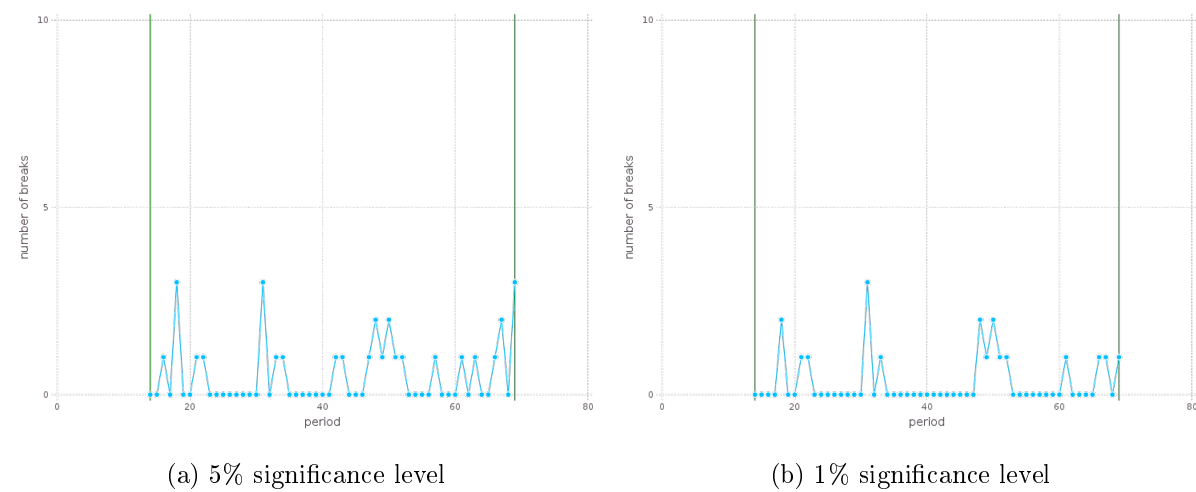
Figure 5: First and second cluster removed

(a) 5% significance level

(b) 1% significance level

Figure 6: All variables with breaks removed



(a) 5% significance level

(b) 1% significance level

Figure 7: Second cluster removed, data set trimmed before 1991 quarter 1

# E   Interaction: Targeted predictors, removal of breaks

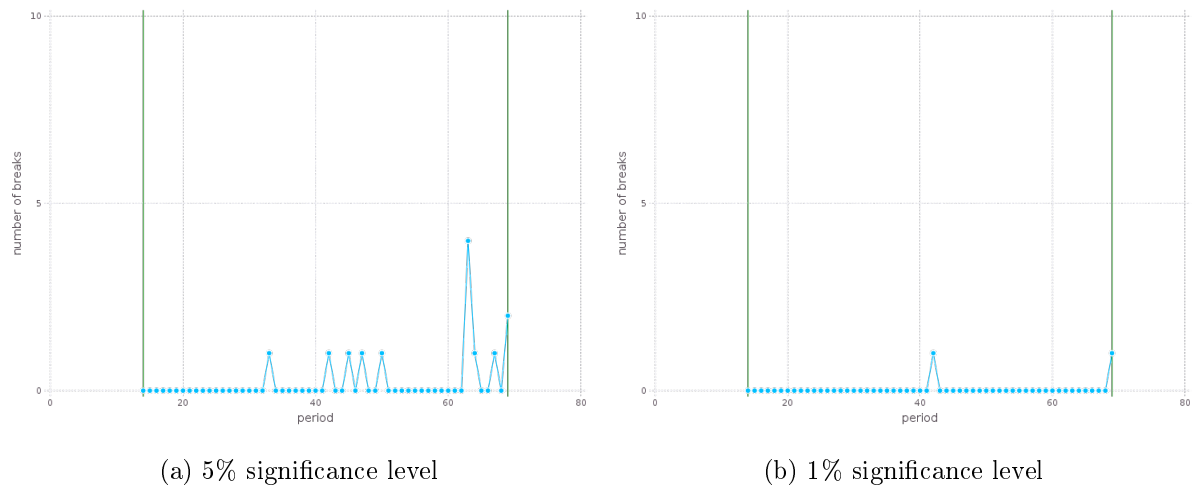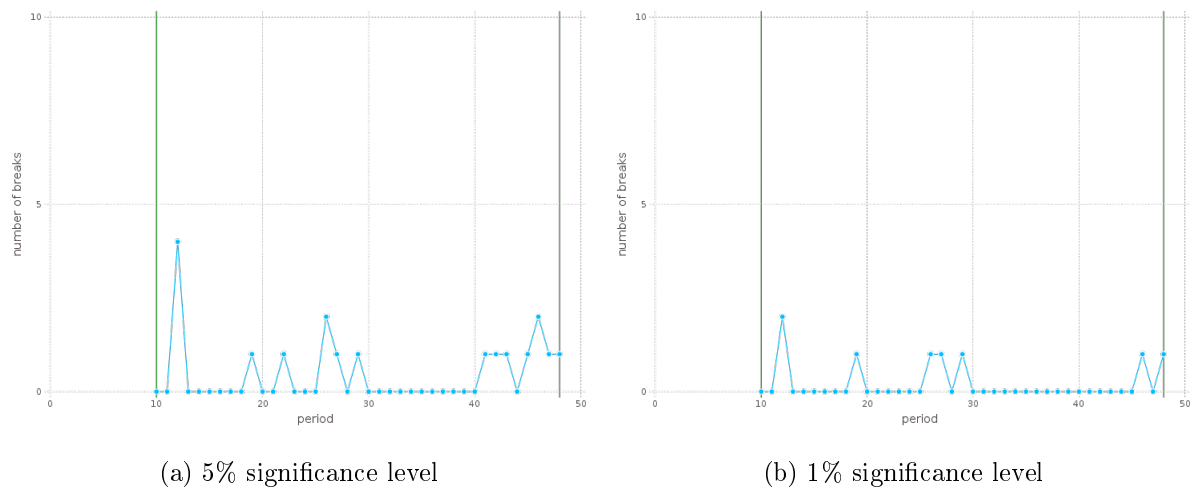| k | $N_{\text{breaks removed}}$ | $RMSE_{\text{tar}}$ | $RMSE_{\text{br}}$ | $\Delta_{\text{RMSE}}$ | DM | $p_>$ | $p_{\neq}$ |
|---|---|---|---|---|---|---|---|
| 26 | 18 | 0.4809 | 0.4564 | 0.0245 | 0.4965 | 0.3136 | 0.6272 |
| 27 | 19 | 0.4823 | 0.4602 | 0.0221 | -0.4547 | 0.6718 | 0.6563 |
| 28 | 18 | 0.4957 | 0.4528 | 0.0429 | -0.2979 | 0.6149 | 0.7701 |
| 29 | 18 | 0.4927 | 0.4806 | 0.0121 | 0.0883 | 0.4654 | 0.9309 |
| 30 | 17 | 0.4925 | 0.4959 | -0.0034 | 0.5863 | 0.2835 | 0.5670 |
| 31 | 18 | 0.4943 | 0.4999 | -0.0055 | -0.1871 | 0.5729 | 0.8543 |
| 32 | 21 | 0.4014 | 0.4139 | -0.0125 | -0.5408 | 0.7014 | 0.5971 |
| 33 | 21 | 0.4747 | 0.4295 | 0.0452 | -1.9228 | 0.9625 | 0.0751 |
| 34 | 22 | 0.4849 | 0.4501 | 0.0349 | -1.9577 | 0.9647 | 0.0705 |
| 35 | 24 | 0.4217 | 0.5230 | -0.1013 | -0.7701 | 0.7730 | 0.4541 |
| 36 | 24 | 0.4826 | 0.4620 | 0.0206 | -1.1949 | 0.8740 | 0.2520 |
| 37 | 25 | 0.4442 | 0.4530 | -0.0088 | -1.2802 | 0.8894 | 0.2213 |
| 38 | 26 | 0.4312 | 0.4785 | -0.0473 | -1.0385 | 0.8417 | 0.3167 |
| 39 | 27 | 0.4000 | 0.4811 | -0.0811 | -1.3551 | 0.9016 | 0.1968 |
| 40 | 28 | 0.4141 | 0.4856 | -0.0716 | -1.2543 | 0.8849 | 0.2303 |
| 41 | 27 | 0.4269 | 0.5076 | -0.0807 | -1.9239 | 0.9625 | 0.0749 |
| 42 | 28 | 0.4490 | 0.5048 | -0.0558 | -1.9349 | 0.9633 | 0.0735 |
| 43 | 28 | 0.4485 | 0.5084 | -0.0600 | -1.6533 | 0.9397 | 0.1205 |
| 44 | 29 | 0.4360 | 0.5192 | -0.0832 | -2.0236 | 0.9687 | 0.0625 |
| 45 | 30 | 0.4309 | 0.5237 | -0.0928 | -1.0432 | 0.8427 | 0.3145 |
| 46 | 31 | 0.4786 | 0.4948 | -0.0163 | -1.5163 | 0.9242 | 0.1517 |
| 47 | 32 | 0.4851 | 0.4943 | -0.0092 | -0.9166 | 0.8126 | 0.3748 |
| 48 | 33 | 0.4716 | 0.5000 | -0.0284 | -3.0493 | 0.9957 | 0.0087 |
| 49 | 33 | 0.5100 | 0.5000 | 0.0100 | 4.7581 | 0.0002 | 0.0003 |
| 50 | 35 | 0.5204 | 0.4579 | 0.0625 | -2.3135 | 0.9818 | 0.0364 |

Critical values for the Diebold-Mariano test of improved and differing accuracy:

$DM_{1\%} = 2.62$, $DM_{5\%} = 1.76$, $DM_{10\%} = 1.35$

$DM_{1\%} = 2.98$, $DM_{5\%} = 2.14$, $DM_{10\%} = 1.76$ (the relevant Statistic is abs(DM)

***Notes:*** Entries report the performance of the optimal model achieved through targeting the predictors only ($RMSE_{\text{tar}}$) and removing the structural breaks from the targeted data sets afterwards ($RMSE_{\text{br}}$). The data sets have been targeted using soft thresholding with $k$ steps. $N_{\text{breaks removed}}$ is the number of variables left in the model after structural breaks are removed. DM denotes the small sample corrected Diebold Mariano statistic from Harvey et al. (1997). $p_>$ and $p_{\neq}$ denote the p-values of a test for improved and differing forecasting accuracy respectively.

Table 15: RMSE after removing structural breaks from targeted data

# F   Data series

Table 16: Federal Reserve Saint Louis data series

| | Series ID | Series Description |
|---|---|---|
| 1 | A229RX0 | Real Disposable Personal Income: Per capita |
| 2 | AAA | Moody's Seasoned Aaa Corporate Bond Yield© |
| 3 | ACNGNO | Value of Manufacturers' New Orders for Consumer Goods: Consumer Nondurable Goods Industries |
| 4 | AHECONS | Average Hourly Earnings Of Production And Nonsupervisory Employees: Construction |
| 5 | AHEMAN | Average Hourly Earnings Of Production And Nonsupervisory Employees: Manufacturing |
| 6 | AHETPI | Average Hourly Earnings of Production and Nonsupervisory Employees: Total Private |
| 7 | AWHMAN | Average Weekly Hours of Production and Nonsupervisory Employees: Manufacturing |
| 8 | AWOTMAN | Average Weekly Overtime Hours of Production and Nonsupervisory Employees: Manufacturing |
| 9 | BAA | Moody's Seasoned Baa Corporate Bond Yield© |
| 10 | BOGMBASE | Monetary Base; Total |
| 11 | BOGNONBR | Non-Borrowed Reserves of Depository Institutions (DISCONTINUED SERIES) |
| 12 | BUSLOANS | Commercial and Industrial Loans, All Commercial Banks |
| 13 | BUSLOANSNSA | Commercial and Industrial Loans, All Commercial Banks |
| 14 | CCRETT01USM661N | Real Effective Exchange Rates Based on Manufacturing Consumer Price Index for the United States© |
| 15 | CES0600000007 | Average Weekly Hours of Production and Nonsupervisory Employees: Goods-Producing |

| 16 | CES0600000008 | Average Hourly Earnings of Production and Nonsupervisory Employees: Goods-Producing |
| --- | --- | --- |
| 17 | CES2000000008 | Average Hourly Earnings of Production and Nonsupervisory Employees: Construction |
| 18 | CES3000000008 | Average Hourly Earnings of Production and Nonsupervisory Employees: Manufacturing |
| 19 | CEU0500000008 | Average Hourly Earnings of Production and Nonsupervisory Employees: Total Private |
| 20 | CEU0600000007 | Average Weekly Hours of Production and Nonsupervisory Employees: Goods-Producing |
| 21 | CEU0600000008 | Average Hourly Earnings of Production and Nonsupervisory Employees: Goods-Producing |
| 22 | CEU1000000001 | All Employees: Mining and Logging |
| 23 | CEU2000000001 | All Employees: Construction |
| 24 | CEU3000000001 | All Employees: Manufacturing |
| 25 | CEU3000000007 | Average Weekly Hours of Production and Nonsupervisory Employees: Manufacturing |
| 26 | CEU3000000009 | Average Weekly Overtime Hours of Production and Nonsupervisory Employees: Manufacturing |
| 27 | CEU3100000001 | All Employees: Durable Goods |
| 28 | CEU3200000001 | All Employees: Nondurable Goods |
| 29 | CEU9000000001 | All Employees: Government |
| 30 | CIVPART | Civilian Labor Force Participation Rate |
| 31 | CLF16OV | Civilian Labor Force |
| 32 | CPALTT01USM661S | Consumer Price Index: Total All Items for the United States© |
| 33 | CPIAPPNS | Consumer Price Index for All Urban Consumers: Apparel |
| 34 | CPIAPPSL | Consumer Price Index for All Urban Consumers: Apparel |
| 35 | CPIAUCNS | Consumer Price Index for All Urban Consumers: All Items |
| 36 | CPIAUCSL | Consumer Price Index for All Urban Consumers: All Items |
| 37 | CPILFENS | Consumer Price Index for All Urban Consumers: All Items Less Food & Energy |
| 38 | CPILFESL | Consumer Price Index for All Urban Consumers: All Items Less Food & Energy |
| 39 | CPIMEDNS | Consumer Price Index for All Urban Consumers: Medical Care |
| 40 | CPIMEDSL | Consumer Price Index for All Urban Consumers: Medical Care |

| 41 | CPITRNNS | Consumer Price Index for All Urban Consumers: Transportation |
|----|----------|----------------------------------------------------------------|
| 42 | CPITRNSL | Consumer Price Index for All Urban Consumers: Transportation |
| 43 | CURRNS | Currency Component of M1 |
| 44 | CURRSL | Currency Component of M1 |
| 45 | DGORDER | Manufacturers' New Orders: Durable Goods |
| 46 | DMANEMP | All Employees: Durable goods |
| 47 | EXCAUS | Canada / U.S. Foreign Exchange Rate |
| 48 | EXJPUS | Japan / U.S. Foreign Exchange Rate |
| 49 | EXSZUS | Switzerland / U.S. Foreign Exchange Rate |
| 50 | EXUSUK | U.S. / U.K. Foreign Exchange Rate |
| 51 | FEDFUNDS | Effective Federal Funds Rate |
| 52 | GS1 | 1-Year Treasury Constant Maturity Rate |
| 53 | GS5 | 5-Year Treasury Constant Maturity Rate |
| 54 | HOUST | Housing Starts: Total: New Privately Owned Housing Units Started |
| 55 | HOUSTMW | Housing Starts in Midwest Census Region |
| 56 | HOUSTMWNSA | Housing Starts in Midwest Census Region |
| 57 | HOUSTNE | Housing Starts in Northeast Census Region |
| 58 | HOUSTNSA | Housing Starts: Total: New Privately Owned Housing Units Started |
| 59 | HOUSTS | Housing Starts in South Census Region |
| 60 | HOUSTSNSA | Housing Starts in South Census Region |
| 61 | HOUSTW | Housing Starts in West Census Region |
| 62 | HOUSTWNSA | Housing Starts in West Census Region |
| 63 | INDPRO | Industrial Production Index |
| 64 | IPBUSEQ | Industrial Production: Business Equipment |
| 65 | IPCONGD | Industrial Production: Consumer Goods |

| 66 | IPDCONGD | Industrial Production: Durable Consumer Goods |
|----|----------|-----------------------------------------------|
| 67 | IPDMAN | Industrial Production: Durable Manufacturing (NAICS) |
| 68 | IPFINAL | Industrial Production: Final Products (Market Group) |
| 69 | IPFUELN | Industrial Production: Fuels |
| 70 | IPFUELS | Industrial Production: Fuels |
| 71 | IPMAN | Industrial Production: Manufacturing (NAICS) |
| 72 | IPMAT | Industrial Production: Materials |
| 73 | IPMINE | Industrial Production: Mining |
| 74 | IPNCONGD | Industrial Production: Nondurable Consumer Goods |
| 75 | ISRATIO | Total Business: Inventories to Sales Ratio |
| 76 | LNS12032197 | Employment Level - Part-Time for Economic Reasons, Nonagricultural Industries |
| 77 | LNU01000000 | Civilian Labor Force |
| 78 | LNU01300000 | Civilian Labor Force Participation Rate |
| 79 | M1NS | M1 Money Stock |
| 80 | M1SL | M1 Money Stock |
| 81 | M2NS | M2 Money Stock |
| 82 | M2SL | M2 Money Stock |
| 83 | MABMM301USM189S | M3 for the United States© |
| 84 | MANEMP | All Employees: Manufacturing |
| 85 | MCUMFN | Capacity Utilization: Manufacturing (NAICS) |
| 86 | MICH | University of Michigan Inflation Expectation© |
| 87 | NAPM | ISM Manufacturing: PMI Composite Index© |
| 88 | NAPMEI | ISM Manufacturing: Employment Index© |
| 89 | NAPMII | ISM Manufacturing: Inventories Index© |
| 90 | NAPMNOI | ISM Manufacturing: New Orders Index© |

| 91 | NAPMPI | ISM Manufacturing: Production Index© |
|---|---|---|
| 92 | NAPMSDI | ISM Manufacturing: Supplier Deliveries Index© |
| 93 | NDMANEMP | All Employees: Nondurable goods |
| 94 | NEWORDER | Manufacturers' New Orders: Nondefense Capital Goods Excluding Aircraft |
| 95 | PAYEMS | All Employees: Total nonfarm |
| 96 | PAYNSA | All Employees: Total nonfarm |
| 97 | PERMITNSA | New Privately-Owned Housing Units Authorized by Building Permits: Total |
| 98 | PPIACO | Producer Price Index: All Commodities |
| 99 | PPICRM | Producer Price Index: Crude Materials for Further Processing |
| 100 | PPIFCF | Producer Price Index: Finished Consumer Foods |
| 101 | PPIFGS | Producer Price Index: Finished Goods |
| 102 | PPIITM | Producer Price Index: Intermediate Materials: Supplies & Components |
| 103 | RRSFS | Real Retail and Food Services Sales |
| 104 | SRVPRD | All Employees: Service-Providing Industries |
| 105 | TB3MS | 3-Month Treasury Bill: Secondary Market Rate |
| 106 | TB6MS | 6-Month Treasury Bill: Secondary Market Rate |
| 107 | TCU | Capacity Utilization: Total Industry |
| 108 | TOTBUSIMNSA | Total Business Inventories |
| 109 | TOTBUSIRNSA | Total Business: Inventories to Sales Ratio |
| 110 | UCNGNO | Value of Manufacturers' New Orders for Consumer Goods: Consumer Nondurable Goods Industries |
| 111 | UEMPMEAN | Average (Mean) Duration of Unemployment |
| 112 | UMCSENT | University of Michigan: Consumer Sentiment© |
| 113 | UNEMPLOY | Unemployed |
| 114 | UNRATE | Civilian Unemployment Rate |
| 115 | UNRATENSA | Civilian Unemployment Rate |
| 116 | USCONS | All Employees: Construction |
| 117 | USGOOD | All Employees: Goods-Producing Industries |
| 118 | USGOVT | All Employees: Government |
| 119 | USMINE | All Employees: Mining and logging |
| 120 | W875RX1 | Real personal income excluding current transfer receipts |

Table 17: Bundesbank data series

|  | Series ID | Series Description |
|---|---|---|
| 121 | BBK01. JQA000 | National accounts/Overall economic view/At current prices/ GDP |
| 122 | BBDE1. M.DE.N. AEA1. A2P300000. F.C.I10.L | Orders received / At constant prices / Germany / Industry / Unadjusted figure |
| 123 | BBDE1. M.DE.N. BAA1. P2XF00000. G.C.I10.L | Output in the production sector / At constant prices / Germany / Main construction industry / Unadjusted figure |
| 124 | BBDE1. M.DE.W. AEA1. A2P310000. F.V.I10.A | Orders received / At current prices / Germany / Industry / Calendar adjusted only |
| 125 | BBDE1. M.DE.W. AEA1. P2XF00000. B2.V.I10.A | Orders received / At current prices, flows / Germany / Main construction industry / Calendar adjusted only |
| 126 | BBDE1. M.DE.Y. AEA1. A2P300000. F.V.I10.A | Orders received / At current prices / Germany / Industry / Calendar and seasonally adjusted |
| 127 | BBDE1. M.DE.Y. AEA1. A2R200050. F.V.I10.A | Orders received / At current prices, flows / Germany / 20+21 Manufacture of chemicals, chemical products; basic pharmaceutical products and pharmaceutical preparations / Calendar and seasonally adjusted |
| 128 | BBDE1. M.DE.Y. AEA1. P2XF00000. B2.C.I10.A | Orders received / At constant prices / Germany / Main construction industry / Calendar and seasonally adjusted |
| 129 | BBDE1. M.DE.Y. AEA1. P2XF02000. B2.V.I10.A | Orders received / At current prices, flows / Germany / Main construction industry / Calendar and seasonally adjusted |
| 130 | BBDE1. M.DE.Y. AEB1. A2P310000. F.V.I10.A | Orders received from the domestic market / At current prices / Germany / Industry / Calendar and seasonally adjusted |

| 131 | BBDE1. M.DE.Y. AEB1. N2C290000. F.C.I10.A | Orders received from the domestic market / At constant prices / Germany / 20+21 Manufacture of chemicals, chemical products; basic pharmaceutical products and pharmaceutical preparations / Calendar and seasonally adjusted |
|---|---|---|
| 132 | BBDE1. M.DE.Y. AEB5. A2P300000. F.V.I10.A | Orders received from abroad / At current prices / Germany / Industry / Calendar and seasonally adjusted |
| 133 | BBDE1. M.DE.Y. BA01. A2P200000. F.C.I10.A | Produktionsergebnis je Beschäftigten / Deutschland / Bergbau und Gewinnung von Steinen und Erden sowie Verarbeitendes Gewerbe (B + C) / Angaben für fachliche Betriebsteile / kalender- und saisonbereinigt |
| 134 | BBDE1. M.DE.Y. BAA1. A2P000000. G.C.I10.A | Output in the production sector / Germany / Production sector including construction (B -F) / Calendar and seasonally adjusted |
| 135 | BBDE1. M.DE.Y. BEA1. P2XF00000. BA.P.ABA.A | Employed persons (including proprietors and family workers) / Persons / Germany / Main construction industry / Calendar and seasonally adjusted |
| 136 | BBDP1. M.DE.N. APA1.G. GP09SA000000. I10.A | Construction price index / Germany / Unadjusted figure / Total |
| 137 | BBDP1. M.DE.N. VPI.C. A00000. I10.L | Consumer price index / Germany / Unadjusted figure / Total |
| 138 | BBDP1. M.DE.Y. APA1.G. GP09SA000000. I10.A | Indices of foreign trade prices - exports, total / Germany / Calendar and seasonally adjusted / Total |
| 139 | BBDP1. M.DE.Y. EPL.L. L1A00000000. I10.A | Index of producer prices of industrial products sold on the domestic market / Germany / Calendar and seasonally adjusted / Total |
| 140 | BBEE1. M.DE.AAA. XY0B02. R.AACPB. M00 | Indicator of the German economy's price competitiveness against 24 selected industrial countries, based on the deflators of total sales |
| 141 | BBEX3. M.XAU. USD.EA. AC.C06 | Price of gold in London / morning fixing * / 1 ounce of fine gold = USD ... |
| 142 | BBK01. BU1132 | Central government debt - total debt |
| 143 | BBK01. DU7802 | Basic pay rates, overall economy, excluding ancillary benefits, excluding one-off payments, on a monthly basis, Germany |
| 144 | BBK01. ES4710 | Balance on current account / Seasonally adjusted |
| 145 | BBK01. EU2001 | Foreign trade / Exports (fob) |

| 146 | BBK01. EU4170 | Income / Total / Receipts |
|---|---|---|
| 147 | BBK01. EU4670 | Financial transactions / Other investment / Balance |
| 148 | BBK01. EU4680 | Foreign investment in Germany / Other investment / Total / Balance |
| 149 | BBK01. OU0092 | Lending to domestic banks (MFIs) / Credit balances and loans / All categories of banks |
| 150 | BBK01. OU0144 | Medium and long-term lending to general government / Loans / Medium-term / All categories of banks |
| 151 | BBK01. OU0149 | Medium and long-term lending to general government / Loans / Long-term / All categories of banks |
| 152 | BBK01. OU0337 | Medium and long-term lending / to general government / Securities / All categories of banks |
| 153 | BBK01. OU0778 | Lending to banks (MFIs) / Total / All categories of banks |
| 154 | BBK01. OU1429 | Lending to banks (MFIs) / Balances and loans / All categories of banks |
| 155 | BBK01. OU1825 | Capital / Total / All categories of banks |
| 156 | BBK01. TUD317 | MONEY STOCK M1 (FROM 2002, EXCLUDING CURRENCY IN CIRCULATION) / GERMAN CONTRIBUTION |
| 157 | BBK01. TVS303J | Geldmenge M1 / Veränderung saisonbereinigt / Jahresrate / EWU |
| 158 | BBK01. TXI355 | TOTAL ASSETS OR LIABILITIES / GERMAN CONTRIBUTION |
| 159 | BBK01. US02CC | Employees subject to social security contributions / Total / Western Germany / Seasonally adjusted / 1 / 2 |
| 160 | BBK01. US0CC1 | Employees subject to social security contributions / Total / Eastern Germany / Seasonally adjusted / 1 / 2 |
| 161 | BBK01. USBA14 | Employment in Germany / Employed persons in Germany / Germany / Seasonally adjusted / 1 / 2 |
| 162 | BBK01. UUCY01 | Unemployment rate (unemployment as a percentage of the civilian labour force) / Germany / 1 / 2 / 3 |
| 163 | BBK01. WU0028 | Gross sales of domestic debt securities at nominal value / Total |
| 164 | BBK01. WU3141 | DAX price index / End 1987 = 1000 / End of month |
| 165 | BBK01. XS7085 | Exports / European countries / Seasonally adjusted |
| 166 | BBK01. XU4204 | German exports / special trade / values / 1 / 2 / 3 / 4 |
| 167 | BBK01. XU4205 | German imports / special trade / values / 1 / 2 / 3 / 4 |
| 168 | BBK01. XU4206 | German foreign trade balance / special trade / values / 1 / 2 / 3 / 4 |
| 169 | BBK01. XX3739 | Exports of the Federal Republic of Germany / Special Trade / Total / 1 / 2 / |

# G   Replication and data files

All replication files can be found at `https://github.com/joidegn/Masterarbeit` or can be gotten from the author by request.

A result from writing this thesis is the FactorModels Julia package which has been released in the public domain as free software licensed under the MIT License a copy of which can be found at `http://opensource.org/licenses/MIT`. Julia, is a powerful, open-source technical computing language which can be found at `http://julialang.org/`. The FactorModels package does the heavy lifting for the replication files and has to be installed in order to run them. Instructions on how to replicate the results are provided both with the package and the replication files.

# References

Filippo Altissimo, Antonio Bassanetti, Riccardo Cristadoro, Mario Forni, Marc Hallin, Marco Lippi, Lucrezia Reichlin, and Giovanni Veronese. Eurocoin: a real time coincident indicator of the euro area business cycle. Technical report, CEPR discussion paper, 2001.

Donald WK Andrews. Tests for parameter instability and structural change with unknown change point. *Econometrica: Journal of the Econometric Society*, pages 821–856, 1993.

Donald WK Andrews. Tests for parameter instability and structural change with unknown change point: A corrigendum. *Econometrica*, pages 395–397, 2003.

Jushan Bai. Inferential theory for factor models of large dimensions. *Econometrica*, 71 (1):135–171, 2003.

Jushan Bai and Serena Ng. Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221, 2002.

Jushan Bai and Serena Ng. Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions. *Econometrica*, 74(4):1133–1150, 2006.

Jushan Bai and Serena Ng. Forecasting economic time series using targeted predictors. *Journal of Econometrics*, 146(2):304–317, 2008.

Jushan Bai and Serena Ng. Principal components estimation and identification of static factors. *Journal of Econometrics*, 176(1):18–29, 2013.

Jushan Bai and Pierre Perron. Computation and analysis of multiple structural change models. *Journal of applied econometrics*, 18(1):1–22, 2003.

A Banerjee, M Marcellino, and I Masten. Forecasting macroeconomic variables using diffusion indexes in short samples with structural change. 2007.

Ben S Bernanke. Monetary policy and the housing bubble. In *speech at the annual meeting of the American Economic Association, Atlanta, Georgia*, volume 3, 2010.

Jean Boivin and Serena Ng. Are more data always better for factor analysis? *Journal of Econometrics*, 132(1):169–194, 2006.

Jörg Breitung and Sandra Eickmeier. Dynamic factor models. *Allgemeines Statistisches Archiv*, 90(1):27–42, 2006.

Jörg Breitung and Sandra Eickmeier. Testing for structural breaks in dynamic factor models. *Journal of Econometrics*, 163(1):71–84, 2011.

Jörg Breitung and Jörn Tenhofen. Gls estimation of dynamic factor models. *Journal of the American Statistical Association*, 106(495):1150–1166, 2011.

Jörg Breitung, Uta Kretschmer, and Deutsche Bundesbank. Identification and estimation of dynamic factors from large macroeconomic panels. Technical report, Discussion Paper, http://www. ect. unibonn. de/discussion. htm, 2004.

Todd E Clark and Michael W McCracken. Tests of equal forecast accuracy and encompassing for nested models. *Journal of econometrics*, 105(1):85–110, 2001.

John G Cragg and Stephen G Donald. Inferring the rank of a matrix. *Journal of econometrics*, 76(1):223–250, 1997.

Russell Davidson and James G MacKinnon. *Econometric theory and methods*, volume 5. Oxford University Press New York, 2004.

Bradley Efron, Trevor Hastie, Iain Johnstone, and Robert Tibshirani. Least angle regression. *The Annals of statistics*, 32(2):407–499, 2004.

Sandra Eickmeier and Jörg Breitung. How synchronized are new eu member states with the euro area? evidence from a structural factor model. *Journal of Comparative Economics*, 34(3):538–563, 2006.

Mario Forni and Lucrezia Reichlin. Let's get real: a factor analytical approach to disaggregated business cycle dynamics. *The Review of Economic Studies*, 65(3):453–473, 1998.

Mario Forni, Marc Hallin, Marco Lippi, and Lucrezia Reichlin. The generalized dynamic factor model. *Journal of the American Statistical Association*, 100(471):830–840, 2005. doi: 10.1198/016214504000002050. URL `http://dx.doi.org/10.1198/016214504000002050`.

Andrew Gelman and Jennifer Hill. Missing-data imputation. In *Data Analysis Using Regression and Multilevel/Hierarchical Models*, pages 529–544. Cambridge University Press, 2006. ISBN 9780511790942. URL `http://dx.doi.org/10.1017/CBO9780511790942.031`. Cambridge Books Online.

J Geweke. The dynamic factor analysis of economic time series', in (dj aigner and as golberger, eds) latent variables in socio-economic models, 1977.

Domenico Giannone, Lucrezia Reichlin, and Luca Sala. *Tracking Greenspan: systematic and unsystematic monetary policy revisited*. Centre for Economic Policy Research, 2002.

David Harvey, Stephen Leybourne, and Paul Newbold. Testing the equality of prediction mean squared errors. *International Journal of forecasting*, 13(2):281–291, 1997.

Ian Jolliffe. *Principal component analysis*. Wiley Online Library, 2005.

Chang-Jin Kim and Charles R Nelson. Business cycle turning points, a new coincident index, and tests of duration dependence based on a dynamic factor model with regime switching. *Review of Economics and Statistics*, 80(2):188–201, 1998.

Denis Kwiatkowski, Peter CB Phillips, Peter Schmidt, and Yongcheol Shin. Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of econometrics*, 54(1):159–178, 1992.

Terence C Langetieg. An application of a three-factor performance index to measure stockholder gains from merger. *Journal of Financial Economics*, 6(4):365–383, 1978.

Ross Larsen and Russell T Warne. Estimating confidence intervals for eigenvalues in exploratory factor analysis. *Behavior research methods*, 42(3):871–876, 2010.

Colin L Mallows. Some comments on c p. *Technometrics*, 15(4):661–675, 1973.

M Hashem Pesaran and Allan Timmermann. Selection of estimation window in the presence of breaks. *Journal of Econometrics*, 137(1):134–161, 2007.

Jean-Pierre Roth. The swiss franc within an euro environment.

Thomas J Sargent and Christopher A Sims. Business cycle modeling without pretending to have too much a priori economic theory. *New methods in business cycle research*, 1:145–168, 1977.

Christian Schumacher. Factor forecasting using international targeted predictors: the case of german gdp. *Economics Letters*, 107(2):95–98, 2010.

G William Schwert. Tests for unit roots: A monte carlo investigation. *Journal of Business & Economic Statistics*, 20(1):5–17, 2002.

Charles Spearman. " general intelligence," objectively determined and measured. *The American Journal of Psychology*, 15(2):201–292, 1904.

James H Stock and Mark W Watson. Diffusion indexes. Technical report, National bureau of economic research, 1998.

James H Stock and Mark W Watson. Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, 20(2):147–162, 2002.

James H Stock and Mark W Watson. Implications of dynamic factor models for var analysis. Technical report, National Bureau of Economic Research, 2005.

James H Stock and Mark W Watson. Dynamic factor models. *Oxford Handbook of Economic Forecasting. Oxford University Press, USA*, pages 35–59, 2011.

Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1):267–288, 1996.

Hui Zou and Trevor Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2): 301–320, 2005.

# Declaration

1. I hereby declare that this thesis, entitled:

### Predicting Economic Time-Series using Dynamic Factor Models under Structural Breaks

is a result of my own work and that no other than the indicated aids have been used for its completion. Material borrowed directly or indirectly from the works of others is indicated in each individual case by acknowledgement of the source and also the secondary literature used. This work has not previously been submitted to any other examining authority and has not yet been published.

2. After completion of the examining process, this work will be given to the library of the University of Konstanz, where it will be accessible to the public for viewing and borrowing. As author of this work, I agree to this procedure.

Konstanz, _____        _____

# Erklärung

1. Ich versichere hiermit, dass ich die vorliegende Arbeit mit dem Thema:

### Predicting Economic Time-Series using Dynamic Factor Models under Structural Breaks

selbständig verfasst und keine anderen Hilfsmittel als die angegebenen benutzt habe. Die Stellen, die anderen Werken dem Wortlaut oder dem Sinne nach entnommen sind, habe ich in jedem einzelnen Falle durch Angaben der Quelle, auch der benutzten Sekundärliteratur, als Entlehnung kenntlich gemacht. Die Arbeit wurde bisher keiner anderen Prüfungsbehörde vorgelegt und auch noch nicht veröffentlicht.

2. Diese Arbeit wird nach Abschluss des Prüfungsverfahrens der Universitätsbibliothek Konstanz übergeben und ist durch Einsicht und Ausleihe somit der Öffentlichkeit zugänglich. Als Urheber der anliegenden Arbeit stimme ich diesem Verfahren zu.

Konstanz, den _____        _____