Forecasting the US GDP Components in the short run

Saulius Jokubaitis,∗ Dmitrij Celov

1Faculty of Mathematics and Informatics, Vilnius University, Lithuania

∗To whom correspondence should be addressed; E-mail: saulius.jokubaitis@mif.vu.lt

The aim of this paper is to estimate short-term forecasts of the US GDP components by expenditure approach sooner than they are officially released by the national institutions of statistics. For this reason, nowcasts along with 1- and 2-quarter forecasts are estimated by using available monthly information, officially released with a considerably smaller delay. The high-dimensionality problem of the monthly dataset used is solved by assuming sparse structures for the choice of leading indicators, capable of adequately explaining the dynamics of the GDP components. Variable selection and the estimation of the forecasts is performed by using the LASSO method, together with some of its popular modifications. Additionally, a modification of the LASSO is proposed, combining the methods of LASSO and principal components, in order to further improve the forecasting performance. Forecast accuracy of the models is evaluated by conducting pseudo-real-time forecasting exercises for four components of the GDP over the sample of 2005-2015, and compared with the benchmark ARMA models. The main results suggest that LASSO is able to outperform ARMA models when forecasting the GDP components and to identify leading explanatory variables. The proposed modification of the LASSO in some cases show further improvement in forecast accuracy.

Key words: LASSO, nowcasting, principal components, variable selection, GDP components

JEL codes: C13, C38, C52, C53, C55

Introduction

Information on the current state of the economy is crucial for various economic agents and policy makers, since the choice of the appropriate policy stance relies on the knowledge of the macroeconomic situation in the country. Although there’s a number of indicators, covering many aspects of the economy available at a higher frequency, quarterly national accounts still play an important role guiding various economic decisions. Unfortunately, GDP and its components are officially released with a considerable delay after the reference period – for example, in the US the first estimates of GDP are released after 1 month after the reference period, where only the supply side of the economy is covered since only the GDP by production approach is estimated. However, if a certain economic institution is interested in the demand side of the economy, the publication of GDP by expenditure approach is released with an even longer delay, which can greatly complicate timely decision making.

On the other hand, various short-term indicators, such as business or consumer surveys, the industrial production, retail or external trade indexes are released at a monthly frequency and can be used to get an early picture of the evolution of the current economic activity in various sectors of the economy. One way of using the available information is by conducting a fundamental analysis, however, the information from the national accounts data still remains desirable for most decision makers. For this reason, a number of
econometric tools have been developed in order to extract the main underlying signals from the available
data and to estimate the national accounts data sooner than it is estimated by the agencies of national
statistics.

Many different methods for extracting the most important signals are studied in the literature, the
most popular of which are the Factor models and its various modifications, which are able to use all of
the available information in order to extract a reliable signal. However, lately, a lot of attention have
been focused on the idea that only a small subset of all of the available information might be enough for
adequate timely estimation of the GDP or its components – that is, the assumption of sparse structures
for the choice of explanatory covariates is made.

In this paper we further explore the sparse structure approach under a large amount of various monthly
economic indicators available, assuming that only a small subset of them are significant and should be
used in forecasting macroeconomic variables. Therefore, the main problem arising is of the optimal
selection of useful variables for the modelling. [Bai and Ng (2008)] find that in some cases large amounts
of high-frequency information can be too much, resulting in poor predictive performance, thus raising a
question of how much information is really needed for good predictions? Recent empirical results (for
example, [Bulligan et al. (2015)]) show that assuming a sparse structure of the underlying data provide
promising results, when the total set of available information is refined by supervised selection before the
application.

Recently a rapid growth in popularity among both the practitioners and the academics is seen of the
Least Absolute Shrinkage Selection Operator (LASSO) method, which employs the supervised variable
selection for modelling, and shows great potential in the literature when used for both variable selection
and the generation of forecasts of economic data. For this reason, in this paper we study the LASSO
method and some of its attractive modifications in greater detail, namely, the Square-Root LASSO, the
Adaptive LASSO and the Relaxed LASSO. Additionally, we propose a method, combining the Relaxed
LASSO approach with the method of Principal Components seeking to extract the significant underly-
ing information with greater accuracy, and we find evidence of further improvement of the forecasting
performance. The empirical performance of the models is evaluated by conducting a pseudo-real-time
short-run forecasting exercise of real, in chain-linked volumes sense, US GDP components by expenditure
approach, namely the Gross Fixed Capital Formation, Private Final Consumption Expenditure, Imports
and Exports of goods and services. During the exercise we estimate forecasts of 4 different forecast
horizons: the backcast of a previous quarter, the nowcast of the coinciding quarter and 1- and 2-quarter
forecasts over the sample of 2005-2015.

The motivation for looking at the demand breakdown, but not the GDP itself, is twofold. First, there
is evidence in the literature ([Dreschel and Schneufele (2013)]), that forecasting GDP by the bottom-up
approach can lead to a more accurate forecasts than when forecasting it directly. The main reason for
it is that by modeling each of the disaggregate separately, we are able to address the different underlying
structures of the subcomponents. For example, Gross Fixed Capital Formation and external trade
variables are much more volatile than aggregate GDP, while Private Consumption is typically smoother
than total activity (see [Artis et al. (2004)]). Therefore, it is interesting to study how do different models
compare in forecasting variables, that are behaving so differently over the business cycle. Second, it can
be seen that the business cycle behaviour of the aggregate GDP is very different from that of its subcom-
ponents. For example, investment tends to trough before GDP, while consumption only takes momentum
when an expansion is well under way, peaking only after the cycle. Therefore, forming models for the
subcomponents of GDP can not only improve the accuracy of the final aggregate GDP, but also act as a
complement to the final forecasts, providing a view on the main drivers of the economic activity, which by
itself may allow for a more accurate read on the cyclical phase for economic agents. Additionally, forming
a different model for each of the subcomponents allows for an inclusion of different sets of predictors
used, thus providing a richer story behind the specified equation.

The paper is organized as follows: in section 1 we present a detailed overview of the LASSO and some
of its popular modifications found in the literature. In section 2 we describe our proposed modification of combining the LASSO with principal components, and further expand into section 3 discussing the possible alterations of the approach. For an empirical study, in the section 4 we briefly describe the design of the information set and the settings for all models used in the pseudo-real-time exercise, while in the section 5 we present the results of the pseudo-real-time exercise, where the backcast, nowcast and 1- and 2-quarter forecasts are estimated. The forecasting performance of the models is then compared with the performance of ARMA models.

1 Review of the LASSO methods

1.1 LASSO

Throughout the paper we assume that \( n \) is the number of observations of the modelled variable \( Y \in \mathbb{R}^{1 \times n} \); \( p \) is the total number of explanatory variables \( X_j \in \mathbb{R}^{1 \times n} \) used, \( j = 1, \ldots, p \). We model \( Y = X' \beta + \varepsilon \), where \( \beta = (\beta_1, \ldots, \beta_p)' \), \( \varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)' \) is the error term, and \( X = (X'_1, \ldots, X'_p) \in \mathbb{R}^{p \times n} \) is the model matrix. Additionally, we use the following notation for an \( \ell_q \) norm, \( q \geq 1 \), throughout the paper: assume that \( v \in \mathbb{R}^d \) for some \( d \in \mathbb{N} \), then \( ||v||_q := (\sum_{j=1}^d |v_j|^q)^{1/q} \) denotes the \( \ell_q \) norm.

The LASSO method (Tibshirani (1996)) is one of the possible solutions when dealing with \( p > n \) problem, where the usual ordinary least squares (OLS) methods are infeasible due to the large amount of parameters needed to be estimated. It has attracted a lot of attention in the literature due to its convenient and widely studied by the academics strictly convex optimization problem, the solution path of which can be effectively estimated by using the LARS algorithm (Efron et al. (2004)).

LASSO is a penalized least squares algorithm with the penalty of an \( \ell_1 \) norm, which solves the following problem:

\[
\hat{\beta}_{\text{LASSO}} = \arg \min_{\beta} (Y - X' \beta)'(Y - X' \beta) + \lambda ||\beta||_1, \tag{1}
\]

where the hyperparameter \( \lambda \in (0, \infty) \) is fixed. With the value of \( \lambda \) growing, the estimated coefficients are shrunken towards zero, where with a sufficiently large value some of them are estimated as 0 due to the properties of the \( \ell_1 \) norm. This feature, allowing to restrict the insignificant parameters of the model to zero, is very convenient, since together with the estimation of the coefficients a selection of the significant features is performed. In a sense, we can see the LASSO as a stepwise regression, since with a large enough value of \( \lambda \), imposing a sufficiently strong penalty, no variable is included in the model as significant, however, by decreasing it by certain amounts we start to include the significant variables one by one to the modelled regression. That is, in relation to the hyperparameter \( \lambda \), in a way this procedure can be interpreted as a stepwise regression, with an additional shrinkage of the estimated values of the model’s parameters. Also, due to the aforementioned shrinkage of the estimated coefficients performed it is often possible to increase the accuracy of the forecasts, since the shrunken coefficients are able to reduce the variance of the forecasts, while increasing the bias (bias-variance trade-off).

A lot of attention in the literature is given particularly to the variable selection aspect of the LASSO. Zhao and Yu (2006) and Zou (2006) proposed an almost necessary and sufficient condition – the Irrepresentable Condition – which ensures asymptotically consistent variable selection of the LASSO. The authors have shown that under cases when a part of the insignificant variables are strongly correlated with the significant ones, the LASSO might not be able to consistently distinguish them apart, regardless of neither the chosen value of the hyperparameter \( \lambda \) nor the sample size \( n \). Additionally, they prove that the consistency of the variable selection by the LASSO requires that the value of \( \lambda \) should grow at a faster rate than of \( \sqrt{n} \). However, Knight and Fu (2000) proved that the LASSO estimator \( \hat{\beta}_n \) is \( \sqrt{n} \)-consistent only under given \( \lambda = \lambda_n = O(\sqrt{n}) \) and under certain conditions, defined in the cited paper. Therefore, we cannot fully expect both consistent variable selection and parameter estimation at the same time.
Let $\mathcal{A} := \{j : \beta_j \neq 0\}$ and assume that $|\mathcal{A}| = p_0 < n$, that is, the true data generating process is using a certain subset of our dataset, and assume that $\hat{\beta}(\delta)$ is a parameter estimator of a certain procedure $\delta$. Then, by the definition, formed by Fan and Li (2001), the procedure $\delta$ is said to have Oracle Properties if for the estimator $\hat{\beta}(\delta)$ the following holds (asymptotically):

1. Let $\mathcal{A}_0 := \{j : \hat{\beta}_j(\delta) \neq 0\}$, then $\mathcal{A}_0 = \mathcal{A}$; that is, the true subset of variables is selected as significant;
2. $\sqrt{n}(\hat{\beta}(\delta)_\mathcal{A} - \beta_\mathcal{A}) \xrightarrow{D} N(0, \Sigma)$, where $\Sigma$ is a covariance matrix of the true subset of significant variables, used by the data generating process.

It is often expressed by Fan and Li (2001) and other authors that every adequate procedure, along with certain other optimality conditions, should also have the Oracle Properties. In this case we can note that the LASSO does not have the Oracle Properties.

### 1.2 Adaptive LASSO

In the literature many variations and modifications of the LASSO can be found, all of which focused on overcoming various shortcomings of the method. One of the most popular is the Adaptive LASSO, allowing to define weights for each individual explanatory variable used in the model:

$$\hat{\beta}_{\text{adaLASSO}} = \arg \min_\beta (Y - X'\beta)'(Y - X'\beta) + \lambda w'|\beta|,$$

where we denote $|\beta| := (|\beta_1|, \ldots, |\beta_p|)'$ and $w = (w_1, \ldots, w_p)'$ is a vector of fixed weights. Zou (2006) proves that when the weight vector $w$ is data-driven and appropriately chosen, the Adaptive LASSO is able to achieve the Oracle Properties. In the literature often weights are suggested to be chosen as $\hat{w}_j := 1/|\hat{\beta}_j^\gamma|$, $j = 1, \ldots, p$, $\gamma > 0$, where $\hat{\beta}^\gamma$ is $\sqrt{n}$-consistent estimator of $\hat{\beta}$. When $p \leq n$, a natural choice can be $\hat{\beta} := \hat{\beta}_{\text{OLS}}$, however, in the case of $p > n$, when the OLS estimate is infeasible or the data is strongly multicollinear, it is suggested to replace the $\hat{\beta}_{\text{OLS}}$ with $\hat{\beta}_{\text{ridge}}$, where coefficients are estimated by the Ridge regression (defined by the $\ell_2$ problem, with additionally replacing the $\ell_1$ norm of the imposed penalty with the $\ell_2$ norm). The authors emphasize that with the sample size increasing, the weights for the insignificant variables should become inflated (to infinity), while the weights of the significant variables should converge to some finite non-zero constant. This way, the method allows for an (asymptotically) unbiased simultaneous estimation of large coefficient and small threshold estimates.

The importance of the appropriate choice of the weight vector is also stressed in Huang et al. (2008), since under high multicorrelation or large amount of noise variables the usage of Ridge or univariate OLS may result in overall weaker results than the original LASSO. Medeiros and Mendes (2015) claim that in the case of $p > n$ it is sufficient that the weights are chosen by a zero-consistent estimator. That is, it is required that such estimator would generate sufficiently small coefficients for the insignificant variables, $n \to \infty$, and that they would converge to a non-zero finite constants for the significant variables. And these requirements, according to the authors, under certain additional conditions, should be satisfied by the ordinary LASSO or the Elastic Net estimators, therefore they can also be used for the selection of optimal weights. In fact, following these requirements, Liu (2014) were able to successfully extend the weight estimation to adjust for auto-regressive processes, and Konzen and Ziegelmann (2016) were able to adjust for lagged effects.

Most importantly, Medeiros and Mendes (2015) shows that the Adaptive LASSO can be widely applied when dealing with time series data. The authors allow for both the residuals and the regressors to be non-Gaussian and conditionally heteroscedastic, which is often observed when dealing with financial and macroeconomic data. They also allow the number of variables (both the candidates to the final model and the final selected ones by the procedure) to grow together with the size of the sample at a polynomial rate. Under these conditions it is shown that the variable selection by the Adaptive LASSO is consistent.
and that the Oracle Properties hold. The geometric growth rate of the number of variables is permitted under certain restrictions imposed on the residuals of the model, however, in reality, when working with economic variables, such a fast growth rate of the number of variables available is almost never observed. Even if we have a fixed set of variables, by additionally including lags of all of these variables into our design matrix $X$, the resulting growth rate of the dimension of the full dataset is only linear with respect to the size of the sample, but not polynomial. This suggests some degree of freedom to additionally include non-linearities through, i.e., interactions between the variables or their power transformations. Also, promising results were generated in the simulation studies by the cited authors when studying the forecasting performance of models with heavy-tailed residuals with GARCH structure, by using strongly correlated regressors as the explanatory variables.

Additionally, Liu (2014) observes that the procedure of the Adaptive LASSO can be effectively performed by employing the LARS algorithm: let $\hat{W} := \text{diag}(\hat{w}_1, \ldots, \hat{w}_p)$, then the optimization problem of the Adaptive LASSO (2) can be rewritten as:

$$\hat{\beta}_{\text{AdaLASSO}} = \arg\min_\beta (Y - X'\hat{W}^{-1}\hat{W}\beta)'(Y - X'\hat{W}^{-1}\hat{W}\beta) + \lambda||\hat{W}\beta||_1$$

$$= \arg\min_\beta (Y - \hat{X}'\hat{\beta}'(Y - \hat{X}'\hat{\beta}) + \lambda||\hat{\beta}||_1, \quad (3)$$

where $\hat{X}' = X'\hat{W}^{-1}$, and $\hat{\beta} = \hat{W}\beta$ (that is, $\hat{\beta}_j = \hat{w}_j\beta_j$, $\forall j$), therefore all of these parameters can be effectively estimated by using the LARS algorithm just as the ordinary LASSO method.

1.3 Relaxed LASSO

Another popular modification of the LASSO, dealing with some of its shortcomings, is the Relaxed LASSO (Meinshausen (2007)), the main idea of which is to separate the selection of the significant variables and the estimation of the model’s coefficients by introducing an additional hyperparameter $\phi$. Let $M_\lambda := \{1 \leq j \leq p : \hat{\beta}_j \neq 0\}$ denote the set of variables, preselected by the LASSO method under a certain fixed value of $\lambda$. Then the Relaxed LASSO is estimated as:

$$\hat{\beta}_{\text{relLASSO}} = \arg\min_\beta n^{-1}(Y - X'\{\beta \cdot 1_{M_\lambda}\})'(Y - X'\{\beta \cdot 1_{M_\lambda}\}) + \phi\lambda||\beta||_1, \quad (4)$$

where $\lambda \in [0, \infty)$ and $\phi \in (0, 1]$, with $1_{M_\lambda}$ being the indicator function, returning the value of 1 for those variables, that were selected by the LASSO as significant under a fixed $\lambda$. That is, for a fixed $\lambda$, the following holds for the set of significant variables $M_\lambda \subset \{1, \ldots, p\}$:

$$\{\beta \cdot 1_{M_\lambda}\}_k = \begin{cases} 0 & k \notin M_\lambda, \\ \beta_k & k \in M_\lambda, \end{cases} \quad (5)$$

for every $k \in \{1, \ldots, p\}$. In this way the selection of significant variables is performed by using the ordinary LASSO and estimating only the hyperparameter $\lambda$, while the appropriate estimation of the model’s parameters and the amount of shrinkage applied is refined by using a second hyperparameter $\phi$. When $\phi = 1$, the estimator coincides with the case of the ordinary LASSO, that is, no correction of the estimated coefficients is performed.

The authors prove that due to such separation of the variable selection, the consistent estimates of the model’s coefficients are estimated with the usual $\sqrt{n}$ rate of convergence, independently from the growth rate of the available information set.
1.4 Squared-Root LASSO

Another recent modification of the LASSO is the Squared-Root LASSO (Belloni et al. (2011)). The authors propose modifying the original formulation of the LASSO problem (1) by taking the square-root of the residual sum of squares term, as defined by the equation (6):

\[ \hat{\beta}_{\text{sqrtLASSO}} = \arg \min_{\beta} \frac{1}{n} \left( \frac{1}{2} (Y - X' \beta)'(Y - X' \beta) \right)^{1/2} + \lambda \| \beta \|_1. \] (6)

The attractiveness of the method follows Bickel et al. (2009), who shows that for the Square-Root LASSO the optimal penalty level reduces to \( \lambda = \sqrt{2 \log(p)/n} \), which makes it having no user-specified parameters and therefore tuning free.

2 Principal Components and LASSO

In this paper we propose a combination of the aforementioned LASSO modifications together with principal components in order to preserve specific strengths and to minimize the possible shortcomings for each of the methods combined.

First, we follow the arguments of Bai and Ng (2008), who show that the use of targeted predictors help achieving significantly better forecasts of macroeconomic data using factor models. Instead of the usual approach to factor model forecasting, where the principal components are extracted from the full data set, the authors suggest using only a subset of it, selected by a chosen hard/soft thresholding algorithm. In this way, an unsupervised algorithm becomes supervised one, because the choice of the targeted predictors now depends on the predicted variable. Therefore, following these arguments we propose employing the LASSO for subset selection (here we consider both the LASSO and the Adaptive LASSO, since the latter is expected to deliver more accurately selected variables due to its asymptotic properties under \( p > n \) even under many highly correlated variables). From here on, let’s assume that \( X^\lambda \in \mathbb{R}^{q \times n} \) is a pre-selected matrix of significant variables under a certain fixed \( \lambda \), where \( 0 < q \leq n \), and is scaled and centered.

Second, since we are interested in modelling specifically macroeconomic data, it is likely, that significant correlation will be observed, with some of the variables possibly even being nested. Therefore, our proposed modification stems from the initial variable selection step of the Relaxed LASSO: here, however, instead of a direct parameter re-estimation of the selected variables, as would be done on the second step of the Relaxed LASSO, we suggest rotating the data using the principal components methodology. In other words, we propose extracting the main latent factors \( F = L'X^\lambda \), where \( L \in \mathbb{R}^{q \times q} \) is a rotation matrix and \( F \in \mathbb{R}^{q \times n} \) is a principal component matrix.

The main idea here is to extract the main underlying information from the data as (orthogonal) latent factors and to model them instead. Due to probable inter-correlation and the (supervised) preselection done in the first step (we assume that the selected variables should be able to describe the macroeconomic process that we are interested in modelling) – it is likely that such data captures some common signals, driving the particular market or the economic sector in question. If we assume those signals being the main reason for macroeconomic growth, it is a good idea to model them instead of the data directly.

As for the parameter estimation, we expand on the idea, defined by equation (3), and estimate:

\[ \hat{\beta}_{\text{PCA-LASSO}} = \arg \min_{\beta} (Y - (X^\lambda)'LL'\beta)'(Y - (X^\lambda)'LL'\beta) + \lambda \| L' \beta \|_1 \]

\[ = \arg \min_{\tilde{\beta}} (Y - F'\tilde{\beta})'(Y - F'\tilde{\beta}) + \lambda \| L' \beta \|_1, \] (7)

since \( LL' = I \) holds by the definition of principal components, \( \tilde{\beta} = L' \beta \); all of which can be efficiently estimated using the LARS algorithm. It can be noted that \( LL' = I \) holds for any \( \tilde{q} \leq q \), so it is feasible.
to remove the redundant components (which explain very little of the total variance of the data and have very small loading coefficients) if there are any.

Our proposed approach differs from the one suggested by Bai and Ng (2008), firstly, because the number of significant factors are selected not by the usual selection, based on various information criterias (such as Akaike, Schwarz, t-statistics from OLS and similar), but by using the soft-thresholding LASSO approach. That is, both the selection of significant factors and the shrinkage of estimated parameters is done simultaneously in order to optimize the forecasting accuracy. The latter point is important, since we are interested in extracting predictive signals.

The obvious strength of such approach is when dealt with a large amount of data (i.e. from a particular market), driven by one or only a few leading factors. In that case the principal component transformation would allow us to extract those latent factors and estimate only the predictive ones using the LASSO.

The final coefficient vector $\hat{\beta}^* := L\hat{\beta}$ would be comprised of the same non-zero variables just as it would be in the classic (Relaxed or Adaptive) LASSO case, however, the estimated coefficient values would be set according to the significance and predictive performance in the latent space, rather than the direct one. In a way, such transformation would act as a filter, distinguishing the important underlying signals from the data and possibly allowing for a more accurate forecasting performance.

Secondly, in contrast to Bai and Ng (2008) and other similar factor forecasting related literature, we propose to base the final forecasts on the predictions of individual variables $X^\lambda_j$, $j = 1, \ldots, q$, rather than on the predicted significant factors $F_j$. That is, if we denote $X^\lambda_t = (X^\lambda_{1,t}, \ldots, X^\lambda_{q,t})'$ as the data at a time moment $t \in \{1, \ldots, n\}$, then for every $h > 0$, the forecasts $\hat{F}^*_{n+h} = (\hat{F}_{1,n+h}, \ldots, \hat{F}_{q,n+h})'$ can be calculated as

$$\hat{F}^*_{n+h} = \hat{X}^\lambda_{n+h}L = (\hat{X}^\lambda_{1,n+h}, \ldots, \hat{X}^\lambda_{q,n+h})' L,$$

where $L$ is known and $\hat{X}^\lambda_{j,n+h}$, for every $j = 1, \ldots, q$, are predicted using time series ARIMA methodology.

Such aggregation might induce a smaller loss in forecast accuracy of factors $F$ when we fail to accurately define a direct model. First, if the extracted factors are formed with weights of a similar size, such forecast aggregation is similar to bagging (bootstrap aggregating, (Breiman (1996))). Second, it is possible, that the data generating process of $F_j$ might be from some family of complex, long memory processes, therefore the aggregation of forecasts of simpler models of its components introduces some degree of freedom to make inaccurate estimations of components true models while still generating more accurate final predictions. Indeed, Granger (1980) has shown that the aggregation of a low-order AR/ARMA processes in certain cases may produce a process with more complex dynamics. Extending these ideas to aggregation of the forecasted series may recover such complex dynamics in the final forecasts of the original factors.

Additionally, it is worth noting that when the variables $X_j$ are orthogonal, due to the properties of Principal Components, the transformation would reduce to the base LASSO performance-wise, since the total amount of information would remain unchanged.

3 Tailoring the rotation of Principal Components

Having introduced the base idea of the transformation, it is possible to extend the method by adding a certain degree of supervision to the classic Principal Components approach. The classic transformation is convenient in such way, that the rotation matrix can be easily tailored against the specific modelled variable by modifying the scales or the angles of the components.

Scale modification, based on some criteria, might help the LASSO distinguishing the most important signals, since the components with larger scale will enter the solution path earlier than the other, smaller components. This idea is discussed by Stakėnas (2012), who observes significant improvement when using Weighted PCA or Generalized PCA for the extraction of factors when nowcasting Lithuanian GDP.
Alternatively, in this section we discuss the idea of angle modification by sparsifying the rotation matrix, i.e., through Sparse PCA (Zou et al. (2006)). The idea here is that we may achieve a more accurate factor representation by losing the orthogonality of the transformation. The main motivation is that, even though the data matrix $X$ is preselected by the LASSO as a matrix, containing mainly significant variables, it is not clear that by rotating the variables to the latent space, all of them will be significant there. In other words, if there are two strongly multicollinear variables, preselected by LASSO as significant, both having roughly the same estimated weights (possibly with different signs) in the loading matrix, it is possible that losing one of the two dimensions might not change the resulting factor estimate.

Further, some of the variables (for convenience in the following text denoted as $Z$, where $Z \subset \{X_j : j = 1, \ldots, q\}$) used might be orthogonal to all other preselected variables, meaning that the principal component solution does not extract the correct factor of it from the latent space. That is, in the latent space, ideally, they would form a direction, where the coordinate vector would have zeros for all other variables. However, in usual principal components that is mostly not the case, since every extracted factor is a linear combination of all of the variables used, even if the weights are close to zero, it’s unlikely for them to be exactly zero. For discussion purposes let’s assume that the model matrix $X^\lambda$ is reordered such, that the block of variables $Z \in \mathbb{R}^{q_0 \times n}, q_0 < q$, is able to capture significant explanatory signals to our modelling problem, orthogonal to the remaining block $\bar{X}$, the LASSO will try to extract as much information as possible from those variables. Since the extracted factor matrix has the following structure:

$$F = L'X^\lambda = L' \cdot (Z', \bar{X}')'$$

the $j$-th factor component will have the following structure:

$$f_j = [L'_j] \cdot (Z', \bar{X}')' = [\Lambda_{q_0 \times q_0}, \Phi_{q \times (q_0 - q)}] \cdot (Z', \bar{X}')'$$

Therefore LASSO, while trying to reconstruct the signal from $Z$, will include too many factors $f_j$ to the final solution, since each one of them will include some information from $Z$. Some of those factors would not be included if part of the rotation matrix would have zeros (the $\Phi$ block in this example). Let’s assume that $G \subset \{1, \ldots, q\}$ is a set of indices denoting factors $f_j$, which have been selected as significant by the LASSO in the final solution only because non-zero loading weights in $\Phi_j$. Then, it is clear, that with every additional $f_r$, $r \in G$, included we will add some noise in the scale of $\Phi_r, \bar{X}'$ to the data. And the more such factors are selected, the closer the PCA-LASSO solution is to the usual Relaxed LASSO solution.

From this discussion we can see the benefit of adding an additional step to the PCA-LASSO procedure. One way is to modify the loading matrix $L$ to introduce some sparseness to it (for example, by using Sparse PCA (SPCA), Zou et al. (2006)). An alternative way could be including the preselected data matrix $X^\lambda$ together with the extracted factors $F$ and model them together as $\tilde{F} = (F', X')'$ using the LASSO. While it may seem that in this way no new information is added, it potentially gives the LASSO additional degrees of freedom to select the necessary combination of variables, in essence recovering the best fitting loading matrix. Thus, the Sparse PCA, base LASSO or the PCA-LASSO would potentially become specific cases, depending on the estimated values of $F_j$ coefficients.

4 Preliminaries: data preparation

In this paper the four main components of the US GDP by the expenditure approach are considered: Gross Fixed Capital Formation, Private Final Consumption Expenditure, Imports and Exports of goods and services, all of which are quarterly and seasonally adjusted by the source.

The monthly data used as explanatory variables are various openly available indicators from the databases of FRED (St. Louis Bank of Federal Reserves) and IMF (International Monetary Fund) from
1980 to 2015, with up to 2000 various macroeconomic time series used in total. Each time series used in the modelling were either seasonally adjusted by the source or by using the X13-ARIMA-SEATS procedure for seasonal adjustment. Additionally, in order to avoid the problem of spurious regression, every time series were stationarized and appropriately adjusted for normality and against outliers. Since we are dealing with a large number of variables and are relying on automatic algorithms for variable selection without defining any prior expectations, strict rules and conservative approach for data cleaning are extremely important to rule out as much as possible noise (see appendix A for specific heuristics and tests applied).

As we want to compare the forecasting performances of different models in a realistic setting, during the pseudo-real-time experiments, the results of which are presented in the chapter 5, we reconstructed the pseudo-real-time dataset for every iteration of the exercise by adjusting the amount of available data by the appropriate release lag for each monthly indicator (see Table 1 for detailed illustration). During a full quarter at least three major updates on the dataset are possible for every different month of the quarter, however, in this paper the results presented are of the last month of the full quarter. Since we are also interested in nowcasting performance, analyzing results of 3rd month helps separating it from forecasting due to available indicators with low publication lag. In the other case we would just be comparing the predictive performance of the ARIMA models, used for the individual predictions of the selected monthly indicators, which is already inspected using 1- and 2-quarter forecasts.

The monthly variables used in the main dataset were aggregated to quarterly by averaging, with additionally up to four quarterly lags included, and the ragged edges of the dataset have been filled by using the ARIMA time series methods.

## 5 Pseudo-real-time forecasting experiments

In this chapter the results of pseudo-real-time forecasting exercise over 2005Q1 – 2014Q4 are presented. As described in section 4, we produce 4 forecasted values for each quarter: one backcast, one nowcast and two forecasts of 1- and 2-quarters ahead, accordingly. We model 4 components of the GDP by expenditure approach: Gross Fixed Capital Formation (GFCF), Private Final Consumption Expenditure (PFCE), Imports and Exports of goods and services. Overall these selected components reflect the main drivers of the economy: the domestic demand largely consists of private consumption and the investments, while the foreign supply is indicated by the international trade, which reflects the economic relations with the foreign sector and the openness of the economy. International competitiveness is also important, since it drives the search for new innovative and advanced solutions.

Out of these 4 variables the hardest to accurately predict is the GFCF, since it is composed of investments in many different industries. The investment spending is necessarily forward looking and hopeful, therefore it expands rapidly during an economic boom, when investors expect that the future will require the great productive capacity, and falls rapidly when such expectations evaporate. Therefore it is
the most volatile of the four. Additionally, for private consumption, imports and exports, there are good leading monthly indicators available, allowing for an easier nowcast. On the other hand, no such variables are available for the GFCF, at least to the knowledge of the authors of this paper, making generation of good nowcasts a more challenging task. Therefore our main focus in this chapter is forecasting the GFCF. The results of forecasting the 3 remaining variables are then briefly reviewed.

5.1 Setting of the exercise

For every modelled component of the GDP we consider the following models and their respective notations: the Square-Root LASSO (in the tables denoted as: Sqrt), Relaxed LASSO (Relaxed), Adaptive LASSO (Adaptive), classic LASSO (LASSO) and a proposed combination of LASSO with the principal components of the data, preselected by the Adaptive LASSO (AdaPCA) or by the classic LASSO (PCA), as described in section 2.

Further, since it is interesting to inspect the gains brought solely by performing the rotation of the data to its principal components, in some particular cases we analyse the alternative cases for the AdaPCA models, where the preselected variables are fixed, but the rotation to the principal components is not performed. In a way it becomes a mix between Relaxed LASSO and Adaptive LASSO, hence denoted as AdaRL in the tables.

As discussed in the section, it is of interest to inspect the effects of added sparseness to the loading matrix of the principal components for the AdaPCA models. However, in none of the cases any significant gains were found when using specifically the SPCA method instead of the classic PCA, therefore the results are omitted from the tables. On the other hand, in some cases there were significant gains in forecasting accuracy when using both the rotated and original data. In a way we can understand it as a cross between the AdaPCA and AdaRL methods, hence the notation AdaPCAX in the tables.

Each of the model is estimated using the cross-validated hyperparameters unless specified differently, where each is chosen so as to maximize the out-of-sample accuracy, where the weights for Adaptive LASSO are chosen using Ridge estimates with $\gamma = 1$ (see equation (2)).

As a benchmark for these models we use ARMA($p,q$) models (ARMA), where the orders ($p,q$) are selected to minimize the Akaike information criterion during each quarter of the exercise.

Additionally, in some cases we present the results of models, where instead of the cross-validated out-of-sample hyperparameters we use such, that provide a more parsimonious result. In those cases the results are presented with an additional number next to their abbreviation (i.e. LASSO5, indicating that the model is consistently selecting 5 significant variables during the exercise). It is worth inspecting such results, since the optimally selected hyperparameter can result in a model of a too dense structure – in this case the LASSO wins a small amount of validation accuracy, but brings an additional forecast uncertainty with each of the additional variable included. On the other hand, the inclusion of more than the optimal number of variables as in the discussion may in some cases be a safer bet: if some of the individual ARMA predictions are generated very inaccurately, the larger number of other variables could act in a correcting way, while with a smaller amount of variables only the shrinkage effect of the LASSO would reduce the resulting negative effects. Therefore, it is of interest to inspect alternatives under different choices of $\lambda$.

It is important to note that during the forecasting exercise, for each of the model in question, both the number and type of variables used were reselected during every quarter when new data vintages has become available. Instead of tailoring the set of possible variables to match the predicted variable, we allow models themselves to select the significant parts – it is likely that some indicators that drove the growth of certain markets in the past are not that significant at a later time, hence they can be replaced in a timely matter with new indicator (i.e. a new market emerges). For this reason two different approaches to the forecasting exercise were taken: first, we employ an expanding forecasting window exercise with the sample data starting from 1982Q1 to 2004Q4, and the window is expanded by adding one additional quarter during every iteration. The size of the window was chosen to be not too large, so that there would
be an appropriate amount of available historical monthly indicators, but large enough for the models to be able to select a large amount of significant variables if needed. However, under such approach it is likely that some of the variables are selected in order to capture the historical dynamics of the modelled series rather than of the more present ones. Second, in order to avoid the possible downsides of an expanding window, a rolling window approach was also employed by using a 12-year sample window, starting from a window of 1993Q1 to 2004Q4, and rolling it by adding an additional quarter to both the start and the end of the window during every iteration. The size of the rolling window was chosen such that it would capture at least one full business cycle.

To indicate how well did the models forecast, we present the ratio of the RMSE of the LASSO models to the ARMA models, in addition to RMSE and pairwise Diebold-Mariano (DM, Diebold (2015)) tests on the forecast errors for the expanding window exercise, and pairwise Giacomini and White (2006) (GW) tests for the rolling window exercise.

5.2 Expanding window: Gross Fixed Capital Formation

5.2.1 Main findings
In this section we present the results of forecasting GFCF during the period of 2005Q1-2014Q4. The results of RMSE of the forecasted values are presented in the table 2 for all of the models and for all 4 forecast horizons. Additionally, the total results were divided into three periods (2005-2008, 2008-2011, 2011-2015) in order to compare the models during different stages of the economy: during the stable growth of 2005-2008, the crisis period of 2008-2011, and the stabilization period of 2011-2015. Dividing up the total results is also useful in inspecting whether there are models, dominating every other competitor in the "horse race". Also, in the table 3 are presented the results of Relative RMSE, when in comparison with the performance of the benchmark ARMA models.

Table 2: RMSE of models forecasts during pseudo-real-time experiments for Gross Fixed Capital Formation, here the bolded values are the smallest ones for every row, and for every block the last line denotes the total forecast accuracy for the full time period of 2005Q1-2014Q4.

|               | Sqrt | LASSO | Adaptive | Relaxed | PCA     | AdaPCA | ARMA |
|---------------|------|-------|----------|---------|---------|--------|------|
| **Back**      |      |       |          |         |         |        |      |
| 05-08         | 0.509| 0.708 | 0.185    | 0.242   | 0.216   | 0.089  | —    |
| 08-11         | 0.681| 1.339 | 0.585    | 0.194   | 0.463   | 0.713  | —    |
| 11-15         | 0.652| 0.771 | 0.338    | **0.223**| 0.313   | 0.468  | —    |
| 05-15         | 0.601| 0.946 | 0.377    | **0.221**| 0.330   | 0.428  | —    |
| **Now**       |      |       |          |         |         |        |      |
| 05-08         | 0.829| 0.930 | 0.817    | 0.958   | **0.761**| 0.857  | 1.246|
| 08-11         | 1.605| 1.867 | 1.516    | **1.383**| 1.478   | 1.509  | 3.115|
| 11-15         | 1.080| **1.075** | 1.130 | 1.114   | 1.094   | 1.116  | 1.228|
| 05-15         | 1.185| 1.315 | 1.166    | 1.155   | **1.116**| 1.161  | 2.038|
| **Fore1Q**    |      |       |          |         |         |        |      |
| 05-08         | 1.078| 1.176 | **0.898**| 1.025   | 1.131   | 0.992  | 1.363|
| 08-11         | 2.372| 2.616 | 2.106    | **2.037**| 2.127   | 2.056  | 3.669|
| 11-15         | 1.082| **1.039** | 1.113 | 1.174   | 1.170   | 1.078  | 1.208|
| 05-15         | 1.588| 1.720 | 1.443    | 1.450   | 1.503   | **1.431**| 2.332|
| **Fore2Q**    |      |       |          |         |         |        |      |
| 05-08         | 1.263| 1.328 | 1.263    | 1.331   | 1.327   | **1.214**| 1.396|
| 08-11         | 2.770| 3.012 | **2.455**| 2.478   | 2.547   | 2.455  | 4.046|
| 11-15         | 1.076| 1.052 | 1.095    | 1.146   | 1.114   | 1.080  | **0.997**|
| 05-15         | 1.833| 1.979 | 1.692    | 1.726   | 1.757   | **1.680**| 2.553|
Table 3: Relative (to ARMA models') RMSE of models forecasts during pseudo-real-time experiments for Gross Fixed Capital Formation, here the bolded values are the smallest ones for every row, and for every block the last line denotes the total forecast accuracy for the full time period of 2005Q1-2014Q4.

|                | Sqrt | LASSO | Adaptive | Relaxed | PCA | AdaPCA | ARMA |
|----------------|------|-------|----------|---------|-----|--------|------|
| **Now**        |      |       |          |         |     |        |      |
| 05-08          | 0.67 | 0.75  | 0.66     | 0.77    | 0.61| 0.69   | 1    |
| 08-11          | 0.52 | 0.60  | 0.49     | 0.44    | 0.47| 0.48   | 1    |
| 11-15          | 0.89 | **0.88** | 0.92     | 0.91    | 0.89| 0.91   | 1    |
| 05-15          | 0.58 | 0.65  | 0.57     | 0.57    | 0.55| 0.57   | 1    |
| **Fore1Q**     |      |       |          |         |     |        |      |
| 05-08          | 0.79 | 0.86  | **0.66** | 0.75    | 0.83| 0.73   | 1    |
| 08-11          | 0.65 | 0.71  | 0.57     | **0.56** | 0.58| 0.56   | 1    |
| 11-15          | 0.90 | **0.86** | 0.92     | 0.97    | 0.97| 0.89   | 1    |
| 05-15          | 0.68 | 0.74  | 0.62     | 0.62    | 0.64| 0.61   | 1    |
| **Fore2Q**     |      |       |          |         |     |        |      |
| 05-08          | 0.90 | 0.95  | 0.90     | 0.95    | 0.95| **0.87** | 1    |
| 08-11          | 0.69 | 0.75  | **0.61** | 0.61    | 0.63| 0.61   | 1    |
| 11-15          | 1.07 | 1.05  | 1.09     | 1.14    | 1.11| 1.08   | 1    |
| 05-15          | 0.72 | 0.78  | **0.66** | 0.68    | 0.69| 0.66   | 1    |

Table 4: This table reports the p-value of the Diebold Mariano test for equal predictive ability with squared differences. The null hypothesis is that the column model has the same forecasting performance as of the row model against a two-sided alternative. Bolded values marks p-values smaller than 0.1.

|                | Sqrt | LASSO | Adaptive | Relaxed | PCA | AdaptivePCA | ARMA |
|----------------|------|-------|----------|---------|-----|-------------|------|
| **Nowcast**    |      |       |          |         |     |             |      |
| Sqrt           | -    | 0.05  | 0.79     | 0.72    | 0.42| 0.79        | 0.023 |
| LASSO          | -    | -     | 0.2      | 0.22    | 0.15| 0.24        | 0.064 |
| Adaptive       | -    | -     | 0.86     | 0.39    | 0.89| 0.023       |      |
| Relaxed        | -    | -     | -        | 0.36    | 0.93| 0.933       |      |
| PCA            | -    | -     | -        | -       | 0.45| 0.022       |      |
| AdaptivePCA    | -    | -     | -        | -       | -   | 0.027       |      |
| ARMA           | -    | -     | -        | -       | -   | -           |      |
| **1Q**         |      |       |          |         |     |             |      |
| Sqrt           | -    | 0.2   | 0.17     | 0.37    | 0.52| 0.18        | 0.056 |
| LASSO          | -    | -     | 0.19     | 0.29    | 0.34| 0.19        | 0.092 |
| Adaptive       | -    | -     | -        | 0.9     | 0.25| 0.05        | 0.05  |
| Relaxed        | -    | -     | -        | -       | 0.12| 0.74        | 0.081 |
| PCA            | -    | -     | -        | -       | -   | 0.13        | 0.093 |
| AdaptivePCA    | -    | -     | -        | -       | -   | -           | 0.053 |
| ARMA           | -    | -     | -        | -       | -   | -           |      |
| **2Q**         |      |       |          |         |     |             |      |
| Sqrt           | -    | 0.22  | 0.38     | 0.5     | 0.52| 0.35        | 0.11  |
| LASSO          | -    | -     | 0.29     | 0.35    | 0.34| 0.28        | 0.16  |
| Adaptive       | -    | -     | -        | 0.36    | 0.2 | 0.11        | 0.11  |
| Relaxed        | -    | -     | -        | -       | 0.5 | 0.1         | 0.14  |
| PCA            | -    | -     | -        | -       | -   | 0.16        | 0.13  |
| AdaptivePCA    | -    | -     | -        | -       | -   | -           | 0.1   |
| ARMA           | -    | -     | -        | -       | -   | -           |      |

The results reveal that when forecasting the GFCF most of the models provide a rather similar forecasting performance, with the ordinary LASSO and ARMA models having the worst accuracy overall. When comparing the results with the benchmark, in almost every case all of the models are able to predict with better accuracy than the benchmark ARMA model, with one exception of a 2-quarter forecast over the period of 2011-2015, where the ARMA models are able to outperform every other model by a small margin. Such a result is consistent with the literature: for example, D’Agostino and Giannone (2006) highlights the fact that during relatively steady growths (the authors analysed the Great Moderation period in particular, where a sizeable decline in volatility of output and price measures was observed) even sophisticated models can fail to outperform simple AR models. Therefore, analysis of the recession period of 2008-2011 is the most interesting one, since then we are comparing the performance of models during a unique event with no historical precedent.

Overall these results further emphasise the value of additional monthly data included in the modelling,
especially during the more volatile periods of 2005-2011.

Additionally, it is evident that both the Adaptive LASSO and the Relaxed LASSO are able to increase the predictive performance of the regular LASSO just as expected. Moreover, the results show that the usage of PCA in the estimation can additionally improve the predictive performance of the models: in all of the forecast horizons either the PCA or AdaPCA method generates the most accurate overall forecasts, while during the different (spliced) time periods the worst results are not worse than ones of the Adaptive LASSO.

In Table 4 the p-values of the DM test are reported, indicating the estimated significance of the models predictive abilities when compared with each other over the full testing period. Firstly, it can be noted that all of the LASSO models are able to outperform the ARMA benchmark with p-values lower than 0.1 when testing the nowcasts and 1-quarter forecasts. When comparing the pairwise results between the models, in most cases the significance is much weaker, however, with 15% significance the PCA model generates more accurate nowcasts than the LASSO. Since the PCA uses the same variables as the LASSO for every quarter of the exercise, these results suggest that the proposed transformation can increase the predictive accuracy.

Table 5: RMSE of selected models forecasts during pseudo-real-time experiments for Gross Fixed Capital Formation, here the bolded value is the smallest one for every row in the block, and for every block the last line denotes the total forecast accuracy for the full time period of 2005Q1-2014Q4.

| Back     | AdaPCA15 | AdaRL15 | AdaPCA20 | AdaRL20 | AdaPCA30 | AdaRL30 |
|----------|----------|---------|----------|---------|----------|---------|
| 05-08    | 0.554    | 0.705   | 0.528    | 0.598   | 0.447    | 0.549   |
| 08-11    | 0.899    | 1.048   | 0.798    | 1.027   | 0.673    | 0.938   |
| 11-15    | 0.775    | 0.983   | 0.609    | 0.909   | 0.609    | 0.934   |
| 05-15    | 0.730    | 0.902   | 0.637    | 0.846   | 0.542    | 0.794   |

| Now      |          |         |          |         |          |         |
|----------|----------|---------|----------|---------|----------|---------|
| 05-08    | 0.821    | 0.832   | 0.799    | 0.776   | 0.740    | 0.777   |
| 08-11    | 1.048    | 1.525   | 1.510    | 1.523   | 1.514    | 1.484   |
| 11-15    | 1.147    | 1.170   | 1.128    | 1.183   | 1.091    | 1.250   |
| 05-15    | 1.200    | 1.195   | 1.150    | 1.191   | 1.120    | 1.179   |

| Fore1Q   |          |         |          |         |          |         |
|----------|----------|---------|----------|---------|----------|---------|
| 05-08    | 0.979    | 1.052   | 0.987    | 0.991   | 0.957    | 0.990   |
| 08-11    | 2.210    | 2.289   | 2.159    | 2.221   | 2.187    | 2.260   |
| 11-15    | 1.159    | 1.048   | 1.125    | 1.106   | 1.124    | 1.167   |
| 05-15    | 1.530    | 1.547   | 1.492    | 1.519   | 1.484    | 1.532   |

| Fore2Q   |          |         |          |         |          |         |
|----------|----------|---------|----------|---------|----------|---------|
| 05-08    | 1.120    | 1.217   | 1.181    | 1.152   | 1.179    | 1.247   |
| 08-11    | 2.551    | 2.665   | 2.566    | 2.591   | 2.667    | 2.692   |
| 11-15    | 1.059    | 1.021   | 1.048    | 1.001   | 1.062    | 1.039   |
| 05-15    | 1.694    | 1.771   | 1.710    | 1.721   | 1.759    | 1.781   |

5.2.2 PCA-LASSO: gains from transformation

In order to directly inspect the gains of using the principal component transformation on the (relaxed) data, a few more comparisons were made. First, in the table 5 are presented the results of forecasting GFCF when the number of preselected variables were fixed to 15, 20, and 30. Note that the results in tables 2 and 3 are generated by models with cross-validated hyperparameters, therefore the estimated number of significant variables may differ greatly during different time periods and between different models. In this case, in order to inspect the performance of the models in greater detail, we found that restricting the hyperparameter selection problem to select only a fixed amount of (the same) variables is useful. Therefore, in the table 5 we examine the results of two models: AdaPCA, where the preselected variables are transformed into principal components, and AdaRL, where no transformation is made, only the coefficients are re-estimated in the style of Relaxed LASSO. Not only is the number of variables used the same, but also all of the variables selected are the same. The results provide evidence that AdaPCA in some cases can improve the forecasting accuracy when compared with ordinary methods. Additionally, by comparing the predictive accuracy of the models with the Diebold-Mariano test we found that with 5%
significance the AdaPCA30 model generated significantly better 1-quarter forecasts than the AdaRL30 model. Overall the improvement can be visible even on a relatively sparse number of variables selected, but the results suggest that the gains from using the PCA transformation are larger when more variables are included in the estimation. This result is natural, since with larger samples we’re likely to include more intercorrelated variables, thus allowing for a clearer extraction of the common factors.

Moreover, the results from Table 10 suggest that additional forecasting accuracy can be gained when using a cross between the two methods (see AdaPCAX), where both the principal components and the original preselected data are included in the model. Since the performance is better than of AdaPCA or AdaRL, this provides evidence that tailoring the loading matrix can further reduce the noise from the data, as discussed in section 3. The latter results are also consistent over different number of variables selected.

Secondly, in the previously presented results in total two hyperparameters were used: first one for the selection of variables used in modelling, and second one for the second step selection and for the amount of shrinkage applied. However, it may also be useful to examine the differences between the forecasting accuracy under a number of different hyperparameter values. For this we chose a set of indicators, preselected as optimal by the LASSO, and ran the pseudo-real-time forecasting exercise over the period of 2011Q1-2014Q4 for two cases: first, where the rotation to the principal components is used and second, where no transformation is applied, here the latter in a way corresponds to the Relaxed LASSO. Additionally, for completeness, the case of the Adaptive LASSO was included as well. The results are presented in the Figure 5. Note that the Adaptive LASSO uses a different set of variables for the prediction, therefore an additional number is added in the graph to enumerate the sets of variables used (also, note the different corresponding scales of log(\(\lambda\))). The results show a slight increase in both the average forecast accuracy (mean RMSE) and a smaller standard deviation for many different values of the hyperparameter \(\lambda\) used. These results provide further evidence that the use of principal components transformation in some cases can provide additional gains in forecast accuracy.

5.2.3 Sparse structure: uncovering leading indicators

For the macroeconomist it can be of great value to inspect the leading indicators for the GFCF, therefore in the Figure 1 we present the top indicators often selected by the Adaptive LASSO during the pseudo-real-time experiments. We can see that there are several variables selected consistently during every period, therefore they can be understood as the key variables for explaining the investment in the US. Additionally, it’s interesting to note that some of the variables seem to form certain clusters, where one part is included only before the crisis, while the other part becoming significant after the crisis, indicating a possible structural break in the data.

Among the most frequently selected are the number of employees in the Construction services, which, together with the number of building permits (both not started and under construction) and building completions, in addition to the Consumer Price Index in the housing sector and industrial production for construction supplies, can form a rather detailed view of the situation in the market of the housing sector. As we know, the investment in construction takes up a large part of the total GFCF. Additionally, explaining the remaining investments in the country, a San Francisco Tech Pulse indicator is consistently selected, capturing the tendencies in the IT sector, which is understandable, since investing to efficient, state-of-the-art technology and R&D can significantly enhance the performance of various industries. Also, Coincident Economic Activity (CEA) Index is often selected. Noteworthy, that instead of the global index for the whole US some particular regions are consistently selected, i.e. Arizona, Virginia, Arkansas, Minnesota and other. Firstly, they are likely to be correlated when selected together, hence

\[1\] For other sample sizes and forecast horizons the differences were not that great, resulting in a larger estimated p-values of the DM test with squared errors and two-tailed alternative hypothesis.

\[2\] Note that there were some indicators, preselected as significant for a smaller amount of times, therefore for convenience they are omitted from the graph.
Figure 1: Most often selected variables during the expanding window pseudo-real-time experiments for Gross Fixed Capital Formation over 2005Q1-2014Q4. Number of times selected denotes only the number of the same variables selected (i.e. the variable and a one-quarter lag) but not the number of lag that was most oftenly selected.

the use of principal components to extract the underlying common factor, driving the economic activity in those regions, seem useful for a more efficient estimation. Secondly, it may be insightful to examine why are the particular regions selected instead of the total index for the US: i.e., according to OECD, Minnesota and Virginia seem to be among the top states when measured by the quality of housing (numbers of rooms per person, housing expenditures and etc.) and income per capita, while Arizona and Arkansas appear to be on the lower end of the scale, which suggest that the inclusion of these variables to the model in a way acted as a re-weighting of the total CEA for the US, where the “new” weights were re-estimated by the model and the selected regions acted as proxies for both the richer and poorer regions.

3Data published at https://www.oecdregionalwellbeing.org/
Table 6: Accuracies of models forecasts during pseudo-real experiments for Gross Fixed Capital Formation, here the bolded value is the smallest one for every row, and for every block the last line denotes the total error of the period 2005-2015.

|               | DirectPCA | AggregatedPCA |
|---------------|-----------|---------------|
| Back          |           |               |
| 05-08         | 0.528     | 0.528         |
| 08-11         | 0.564     | 0.564         |
| 11-15         | 0.694     | 0.694         |
| 05-15         | 0.585     | 0.585         |
| Now           |           |               |
| 05-08         | 0.708     | 0.843         |
| 08-11         | 1.477     | 1.330         |
| 11-15         | 1.148     | 1.158         |
| 05-15         | 1.135     | 1.096         |
| Fore1Q        |           |               |
| 05-08         | 1.299     | 1.034         |
| 08-11         | 2.892     | 2.073         |
| 11-15         | 1.104     | 1.131         |
| 05-15         | 1.917     | 1.459         |
| Fore2Q        |           |               |
| 05-08         | 1.254     | 1.297         |
| 08-11         | 3.496     | 2.475         |
| 11-15         | 1.723     | 1.077         |
| 05-15         | 2.262     | 1.698         |

5.2.4 Forecast aggregation

In order to evaluate the ideas of forecast aggregation, briefly discussed in section 2 under real data, the following additional experiment was made. In this exercise 20 significant variables were preselected by the Adaptive LASSO during every quarter of the pseudo-real-time exercise, however, instead of a second step estimation the following post-LASSO model was considered: using the first five principal components (when ordered by their variance explained) an OLS regression was made, treating the extracted factors as observable data. However, as discussed in section 2, two ways of forecasting those factors present themselves: first, by fitting an appropriate ARMA model for each of the component and forecasting them directly (DirectPCA); and second, by forecasting the preselected variables and aggregating their forecasts (AggregatedPCA). The resulting performance of both of these two methods are presented in Table 6 and a few conclusions arise.

First, we can see that the nowcasting performance is rather similar, with the aggregated method being able to explain the crisis period with greater accuracy than the direct method. However, such similar results can be expected, since some monthly information is already known during the nowcasted quarter, and since the factors compared are the same, such comparison essentially depends from the method used to fill the ragged edges.

Second, the forecasting performance for most of the periods is significantly (as suggested by the DM test for both 1-quarter and 2-quarter forecasts with 5% significance) improved when using the aggregated forecast method, with the biggest differences visible during the period of 2008-2011. It is likely that such a decrease in accuracy by forecasting directly can be caused by underestimating the complexity of the

\[4\] It is often found in the literature that a small number of principal components is usually enough when the initial data sample is not large. In our case with 20 variables selected this number seems optimal since it is not too large for efficient OLS estimation and not too small to be risking omission of significant data. Also, since the variables are preselected by the LASSO, it is likely that principal components, explaining the most variance, will be the most significant in the OLS estimation.

\[5\] In this exercise the ragged edges were filled using the Holt-Winters procedure. It is not the most commonly used method for such a problem, but we have found it producing adequate results. ARMA methods are also a good alternative, however we did not want to have coinciding nowcasts with the ones from the AggregatedPCA.
extracted factors – even if in this exercise the amount of sample data is relatively large, it may not be
large enough to efficiently estimate large numbers of ARMA parameters, and while the same holds for
forecasting the preselected variables, the aggregation of their forecasts appears to significantly improve
the results. The benefits of such aggregation, as discussed in section 2, seem to be twofold: firstly, by
forecast aggregation we create a more complex dynamics of the final forecasts than by forecasting directly,
and secondly, the applied aggregation can act in a self-correcting way, by smoothing out the possible cases
of highly “shooting” individual forecasts (if such cases occur while forecasting directly – the final forecast
can be highly inaccurate, while with aggregation the negative effect can be significantly diminished).

A few additional observations can be made from these results: first, because of the complexity of
modelling each individual component, the aggregation is feasible for only a small subset of variables used,
therefore it cannot be applied for large and dense problems. However, the complexity of the problem
with preselected (targeted) predictors essentially boils down to the complexity of the Relaxed LASSO
method. Second, it can be seen when comparing the results from Table 5 with Table 6 that since the 20
variables preselected are the same in both cases, the post-LASSO solution with using only the first few
extracted principal components can even lead to more accurate overall results than applying the LASSO
shrinkage.

5.3 Rolling window: Gross Fixed Capital Formation

In this section we repeat the exercise by switching to a 12-year rolling window instead of the expanding
window as in the section 5.2. The size of the window has been chosen in order to account for the likely
occurrence of structural breaks: since the business cycle tends to last around 5-7 years, we expect to
cover 1-2 cycles. The main motivation for such a comparison is to inspect whether there are “expired”
variables, i.e., those, that are consistently selected by the LASSO as significant only because they help
explaining the older historical data, but are less useful when forecasting during the later times, therefore
producing inaccuracies in the forecasts.

5.3.1 Main findings

The main results are presented in Tables 7 and 8. As in the previous case with the expanding forecast
window we can see that all of the models in comparison are able to outperform the ARMA benchmark,
with the LASSO and Square-Root LASSO overall producing the least accurate forecasts. Additionally,
the results show that in most cases the AdaPCA forecasts are not worse than the ones from the Adaptive
LASSO, with the largest improvement in the RMSE visible when inspecting the nowcasts over the crisis
period of 2008-2011. Moreover, the AdaPCAX method, combining both the original preselected data
and its rotation to the principal components, show some additional gains in forecasting accuracy when
compared to the AdaPCA: it produces more accurate nowcasts for every spliced period during the exercise,
and slightly better 1-quarter and 2-quarter forecasts, only with worse results for 1-quarter forecasts during
the 2008-2011. While the gains are not large, these results suggest that mixing the variables with their
principal components can further increase the forecasting performance, and overall the AdaPCAX showed
the highest forecast accuracy.

Additionally, all of the aforementioned results, using the dataset preselected by the Adaptive LASSO,
are more accurate than the ones, where the ordinary LASSO did the preselection (i.e. PCA, Relaxed),
providing evidence that in some cases the Adaptive LASSO is able to select better predictors than the
LASSO. This result is very important, since the selection of good predictors can be crucial in nowcasting
exercises.
5.3.2 Sparse structure: uncovering leading indicators

In the Figure 2 is presented the list of top variables, preselected by the Adaptive LASSO during the forecasting exercise. It can be seen that, similarly to the results from the expanding window exercise, most of the consistently selected indicators are explaining the construction and housing sectors in the US: the employment rate in the construction sector, together with numbers on building permissions and building completions, complemented by the Consumer Price Index in the housing sector provide a rather detailed view on the situation in the housing market. Additionally, just as in the case of the expanding window, the Coincident Economic Activity (CEA) for Virginia (and other states, such as Arizona, Arkansas and Minnesota, which were selected less often than for the Virginia, therefore not included in the figure among the top predictors) is also often found significant. Among other variables we find that the employment data from various states (Michigan, Arizona, Kentucky, Vermont, Florida and other) are often chosen when explaining the dynamics of GFCF. Also, it can be noted that the LIBOR interest rates are always included, reminding of the importance of the health of the global financial sector when explaining the investments: LIBOR is often served as a benchmark reference rate for various debt instruments (i.e. mortgages), often used by the investors.

However, we can note that the San Francisco Tech Pulse, indicating the health of the IT sector, is no longer included so often to the models, suggesting that it was likely more significant when explaining the historical data. With the rapid growth of the information technologies’ market during the 1995-2001 (note the dot-com bubble in the stock markets during that time), affecting the performance of various industries through rapid technological advancement, it is likely that much of the investment was aimed to the IT infrastructure. Additionally, the interest rates during that period were relatively low and many investors during that period were less risk averse than usual, likely causing a growth of investments in various sectors, correlating with the rapid growth of the IT sector.

Table 7: RMSE of models forecasts during rolling window pseudo-real-time experiments for Gross Fixed Capital Formation, here the bolded values are the smallest ones for every row, and for every block the last line denotes the total forecast accuracy for the full time period of 2005Q1-2014Q4.

|       | Sqrt | LASSO | Adaptive | PCA  | AdaPCA | AdaPCAX | Relaxed | ARMA |
|-------|------|-------|----------|------|--------|---------|---------|------|
| Back  |      |       |          |      |        |         |         |      |
| 05-08 | 0.619| 0.918 | 0.418    | 0.456| 0.363  | 0.198   | 0.430   |      |
| 08-11 | 0.555| 1.388 | 0.704    | 0.549| 0.587  | 0.187   | 0.136   |      |
| 11-15 | 0.549| 0.777 | 0.536    | 0.489| 0.452  | 0.245   | 0.367   |      |
| 05-15 | 0.555| 1.022 | 0.526    | 0.468| 0.438  | 0.219   | 0.346   |      |
|Now    |      |       |          |      |        |         |         |      |
| 05-08 | 0.921| 1.077 | 0.742    | 0.933| 0.809  | 0.799   | 0.860   | 1.316|
| 08-11 | 1.749| 1.985 | 1.470    | 1.875| 1.382  | 1.326   | 1.678   | 2.916|
| 11-15 | 1.188| 1.080 | 1.176    | 1.218| 1.103  | 1.100   | 1.226   | 1.342|
| 05-15 | 1.291| 1.405 | 1.182    | 1.360| 1.127  | 1.106   | 1.246   | 2.001|
|Fore1Q |      |       |          |      |        |         |         |      |
| 05-08 | 1.088| 1.183 | 0.952    | 1.124| 1.032  | 0.980   | 1.055   | 1.526|
| 08-11 | 2.377| 2.479 | 2.226    | 2.245| 2.168  | 2.213   | 2.325   | 3.467|
| 11-15 | 1.181| 1.182 | 1.209    | 1.183| 1.213  | 1.191   | 1.202   | 1.483|
| 05-15 | 1.623| 1.698 | 1.549    | 1.569| 1.52  | 1.534   | 1.586   | 2.327|
|Fore2Q |      |       |          |      |        |         |         |      |
| 05-08 | 1.276| 1.342 | 1.195    | 1.339| 1.230  | 1.187   | 1.205   | 1.660|
| 08-11 | 2.882| 3.024 | 2.906    | 2.813| 2.890  | 2.866   | 2.922   | 3.719|
| 11-15 | 1.204| 1.261 | 1.224    | 1.170| 1.232  | 1.217   | 1.207   | 1.337|
| 05-15 | 1.903| 2.005 | 1.901    | 1.874| 1.893  | 1.876   | 1.911   | 2.485|
Table 8: Relative (to ARMA models’) RMSE of models forecasts during rolling window pseudo-real-time experiments for Gross Fixed Capital Formation, here the bolded values are the smallest ones for every row, and for every block the last line denotes the total forecast accuracy for the full time period of 2005Q1-2014Q4.

| Date   | Sqrt LASSO | Adaptive PCA | AdaPCA | AdaPCAX | Relaxed | ARMA |
|--------|------------|--------------|--------|---------|---------|------|
| Now    |            |              |        |         |         |      |
| 05-08  | 0.70       | 0.82         | 0.56   | 0.71    | 0.62    | 0.60 | 0.65 | 1  |
| 08-11  | 0.60       | 0.68         | 0.5    | 0.64    | 0.47    | 0.45 | 0.57 | 1  |
| 11-15  | 0.89       | **0.80**     | 0.88   | 0.91    | 0.82    | 0.81 | 0.91 | 1  |
| 05-15  | 0.65       | 0.70         | 0.59   | 0.68    | 0.56    | 0.55 | 0.62 | 1  |
| Fore1Q |            |              |        |         |         |      |
| 05-08  | 0.71       | 0.78         | **0.62** | 0.74    | 0.68    | 0.64 | 0.69 | 1  |
| 08-11  | 0.69       | 0.72         | 0.64   | 0.65    | **0.63** | 0.63 | 0.67 | 1  |
| 11-15  | **0.80**   | 0.80         | 0.82   | 0.80    | 0.82    | 0.80 | 0.81 | 1  |
| 05-15  | 0.70       | 0.73         | 0.67   | 0.67    | 0.65    | **0.65** | 0.68 | 1  |
| Fore2Q |            |              |        |         |         |      |
| 05-08  | 0.77       | 0.81         | 0.72   | 0.81    | 0.74    | **0.71** | 0.72 | 1  |
| 08-11  | 0.77       | 0.81         | 0.78   | 0.76    | **0.76** | 0.78 | 0.78 | 1  |
| 11-15  | 0.90       | 0.94         | 0.92   | **0.88** | 0.92    | 0.91 | 0.90 | 1  |
| 05-15  | 0.77       | 0.81         | 0.76   | **0.75** | 0.76    | 0.75 | 0.76 | 1  |

Figure 2: Most often selected variables by the AdaptiveLASSO during the rolling window pseudo-real-time forecasting exercise for the Gross Fixed Capital Formation.

5.4 Private Final Consumption Expenditure

When compared with the investments, the behaviour of private consumption is quite different. First, it tends to show a much more stable and less volatile growth than the investments. Second, it does not immediately react to the various stages of the business cycle – it tends to take the momentum only when the expansion of the current cycle is well under way, with reaching the peak after the cycle. Therefore, it is easier to reflect various shocks in the economy when generating nowcasts for private consumption, since in certain markets some of the shocks could be felt at an earlier time. Additionally, for nowcasting,
it is especially convenient that there are hard monthly indicators available, which are released with a relatively small publication lag. Furthermore, the latter result highlights the importance of accurate individual forecasting of such monthly indicators. It is very likely, that 1- and 2-quarter forecasts of private consumption would be greatly improved if the forecasts of mentioned hard monthly indicators would be generated by employing more sophisticated models, capable of including more explanatory information than ARIMA models.

The results of forecasting the PFCE are presented in Table 11 over a rolling 12-year window. The most accurate nowcasts overall are produced by the Relaxed LASSO models, with very similar performance to the proposed PCA modification. Additionally, the most accurate 1- and 2-quarter forecasts are generated by the AdaPCA method.

By examining the results from the GW test, presented in Table 12, we can see that with 10% significance all of the LASSO modifications are able to generate significantly more accurate nowcasts than the ARMA models, with the greatest significance being suggested for the Adaptive PCA method.

In Figure 3 the top monthly variables are presented, most often preselected by the Adaptive LASSO as significant. When inspecting the results we find that the most often selected are the monthly indicators of real personal consumption expenditure (the index of total expenditure, together with the expenditures excluding food and energy; and expenditure on services) as was expected, since these are hard indicators and often used by statistical agencies as the primary sources for their own preliminary nowcasts. This result provides further evidence that LASSO is able to identify the main leading indicators from a large set of available information.

5.5 International Trade

Similarly to investments, the imports and exports of goods and services are much more volatile than the aggregate GDP. The cyclical properties of international trade are quite interesting, since they are determined by the balance of two forces: the desire of economic agents to smooth consumption using
international markets and the additional cyclical variability from the investments, that are permitted by the international capital flows. It can also be noted that there usually exists a strong co-movement between imports and exports. Even though one would expect that certain shocks may have an opposite effect on real exports and imports, it is likely that certain demand shocks might be transmitted across different countries and affected by global cycles, for example, an increase in imports due to the rise of domestic demand should result in a raise in foreign exports and foreign income, which in turn should raise the domestic exports. As in the case with private consumption, the nowcasting of these variables is simplified by the fact that there exists hard monthly indicators of external trade, published with a small delay.

The results of forecasting Exports and Imports are presented in Table 11 over a rolling 12-year window. First, it can be noted that all of the methods are able to outperform the ARMA models during nowcasting, with the AdaPCA method providing overall the most accurate nowcasts. Additionally, it may be interesting to see that both the PCA and AdaPCA are able to outperform every other model when nowcasting the crisis period of 2008-2011 by a large margin. Second, when comparing the forecast accuracy between the PCA and the Relaxed LASSO, in most cases we can see the PCA method providing both more accurate nowcasts and 1- and 2-quarter forecasts, further suggesting that the rotation to the principal components can in some cases provide additional forecasting accuracy, as was seen in the previous sections, since the variables preselected in the first step were the same during every period of the exercise for both methods.

Overall, the AdaPCA method generated the most accurate nowcasts and 2-quarter forecasts, while the 1-quarter forecasts were rather similar between most of the methods.

The results from the GW test, presented in Table 13, suggest that with 11% significance the Square-Root LASSO, LASSO and PCA are able to generate more accurate forecasts than the benchmark ARMA model.

Similar results can be seen for Imports data. First, it can be noted that the PCA method was able to generate both the most accurate overall nowcasts and the most accurate nowcasts during the financial crisis period of 2008-2011, with the AdaPCA being the second best. Additionally, the PCA method performed better than the Relaxed LASSO during the nowcasting. On the other hand, when comparing the 1-quarter and 2-quarter forecasts all of the methods performed similarly well, with the Square-Root LASSO being able to provide the most accurate 1-quarter forecasts.

The results from the GW test, presented in Table 14, suggest that with 10% significance all of the LASSO modifications are able to produce significantly better nowcasts than the benchmark ARMA models, with the highest significance found for the AdaPCA method, and with p-value of 0.1 the AdaPCA is suggested to be able to generate significantly better 1-quarter forecasts. Also, it can be noted that with 15% significance all of the models, excluding ordinary LASSO, are able to outperform the ARMA benchmark in 1-quarter forecasts.

6 Conclusions

Short-term forecasting of quarterly components of the GDP rely on the availability of timely monthly information. In this paper we studied the forecasting performance of the LASSO and its popular modifications, together with our proposed modification of combining LASSO with the method of principal components. This approach assumes a sparse structure of the available information set required for adequate modelling, therefore is able to distinguish and estimate the main important explanatory variables for the problem. The forecasting performance was studied by conducting a pseudo-real-time forecasting exercise, from which three main results emerge:

---

6 For example, appreciation of the real effective exchange rate can be expected to decrease exports due to the reduced price competitiveness, but increase imports by lowering the relative import prices.
First, all of the LASSO methods show good forecasting performance, outperforming the benchmark ARMA models. The advantages of including additional explanatory monthly information are substantial during the crisis period of 2008-2011, where both the nowcasts and 1- and 2-quarter forecasts in most cases provide more accurate results than the benchmark model. Furthermore, in most cases the number of variables selected by the methods was not large, suggesting that the sparseness assumption for the data generating process holds.

Second, in most cases the modifications of LASSO, discussed in this paper, are able to improve the forecasting accuracy of the LASSO, suggesting not only theoretical, but also practical usefulness of looking into the modifications of the classic LASSO method.

Third, while the LASSO is capable of generating adequate forecasts for different macroeconomic data, our suggested modification by combining the methodologies of LASSO and the principal components show additional gains in forecast accuracy, suggesting that there still is room for further improvement. Namely, we found evidence that in some cases the proposed combination was able to generate more accurate forecasts than the Adaptive LASSO or the Relaxed LASSO, which already are substantial modifications of the original LASSO. Therefore, further gains can be expected with additional work on these methods. On the other hand, the discussed methods never find non-linearities if they are not included into the initial information set. More time consuming, yet interesting extension would be to go for second/third order interaction terms between the variables or their power transforms, which might result in further improvement of forecasting performance.

As we have seen from the results of the forecasting exercise, the usage of weights by the Adaptive LASSO in some cases has successfully improved forecast accuracy when compared with LASSO, however, the weights chosen in this paper were rather conventional, most often suggested in the literature. While
the currently chosen weights overall generated good results, it is likely that they can be further improved by searching for other, more suitable weights (or an algorithm for their estimation) to optimally deal with the high-dimensionality problem.

Moreover, in this paper, when combining the method of principal components with the LASSO, only the standard estimation procedure of the components was discussed. However, the method can be further extended by tailoring the angles or scale of the rotation matrix for additional predictive gains.

References

Artis, M., M. Marcellino, and T. Proietti (2004). Dating Business Cycles: A Methodological Contribution with an Application to the Euro Area. *Oxford Bulletin of Economics and Statistics* 66(4), 537–565.

Bai, J. and S. Ng (2008). Forecasting economic time series using targeted predictors. *Journal of Econometrics* 146(2), 304 – 317. Honoring the research contributions of Charles R. Nelson.

Belloni, A., V. Chernozhukov, and L. Wang (2011). Square-root lasso: pivotal recovery of sparse signals via conic programming. *Biometrika* 98(4), 791.

Bickel, P. J., Y. Ritov, and A. B. Tsybakov (2009). Simultaneous analysis of Lasso and Dantzig selector. *Ann. Statist.* 37(4), 1705–1732.

Box, G. E. P. and D. R. Cox (1964). An Analysis of Transformations. *Journal of the Royal Statistical Society. Series B (Methodological)* 26(2), 211–252.

Breiman, L. (1996). Bagging Predictors. *Machine Learning* 24(2), 123–140.

Bulligan, G., M. Marcellino, and F. Venditti (2015). Forecasting economic activity with targeted predictors. *International Journal of Forecasting* 31(1), 188 – 206.

D’Agostino, A. and D. Giannone (2006). Comparing alternative predictors based on large-panel factor models. Working Paper Series 0680, European Central Bank.

Diebold, F. X. (2015). Comparing Predictive Accuracy, Twenty Years Later: A Personal Perspective on the Use and Abuse of Diebold–Mariano Tests. *Journal of Business & Economic Statistics* 33(1), 1–1.

Dreschel, K. and R. Scheufele (2013). Bottom-up or Direct? Forecasting german GDP in a Data-rich Environment. *Halle Institute for Economic Research*.

Efron, B., T. Hastie, I. Johnstone, and R. Tibshirani (2004). Least angle regression. *Ann. Statist.* 32(2), 407–499.

Fan, J. and R. Li (2001). Variable selection via Nonconcave Penalized Likelihood and its Oracle Properties. *Journal of the American Statistical Association* 96(458).

Giacomini, R. and H. White (2006). Tests of Conditional Predictive Ability. *Econometrica* 74(6), 1545–1578.

Granger, C. (1980). Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics* 14(2), 227 – 238.

Huang, J., S. Ma, and C.-H. Zhang (2008). Adaptive lasso for sparse high-dimensional regression models. *Statistica Sinica* 18(4), 1603–1618.

Knight, K. and W. Fu (2000). Asymptotics for lasso-type estimators. *Ann. Statist.* 28(5), 1356–1378.
Appendices

A Data preparation

Due to the nature of $p >> N$ problem, it is very easy for automatic algorithms such as LASSO to find significance in the noise of the data. With classic regression, such noise variables can be easily discarded through cross-validation, however, this may become a problem when shrinkage is applied and the spurious effects get smoothed out in out-of-sample forecasts. For this reason, strict set of rules were applied on variable prescreening.

First, by performing the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test, the stationarity of the time series was estimated (with 5% significance); second, the test for the unit roots was performed by using the Augmented Dickey-Fuller (ADF) test, starting from estimation and the removal of the deterministic part of the series, if such was observed; additionally, since the test statistic of the ADF test is based on the estimated value of the $t-$statistics from an arbitrary regression model, its resulting residuals were also inspected for the possible presence of heteroscedasticity. The main idea here is that if the resulting residuals are significantly heteroscedastic (according to Breusch-Pagan (BP) test with 5% significance level), the estimated value of the $t$-statistics might be biased, therefore in such cases additionally a nonparametric Philips-Perron (PP) unit-root test was performed, which is able to correct the possibly incorrect results of the ADF test by bootstrapping the critical values of the test statistic.

In all cases, the number of lags used in the arbitrary regressions were chosen by minimizing the Akaike information criterion. The time series was found as statistically significantly non-stationary if either KPSS or unit-root tests suggested non-stationarity with 5% significance. In such case the series
were transformed by differencing, after which the aforementioned procedure was repeated until the final series was found as significantly stationary.

Additionally, since most of the variables used were mainly economic indicators, which are usually described by multiplicative processes, it is useful to apply logarithmic transformation for some of them. For such series often a large deviation is observed in the top levels of the amplitude, which often results in severe non-normality of the data. However, by applying the logarithmic transformation the underlying multiplicative processes are transformed into additive, thus removing most of the explosive effects and to some level restricting its variance. Whether such a transformation is actually useful was decided by using the Box-Cox transformation (Box and Cox (1964)). A logarithmic transformation was applied if for a particular series if the Box-Cox procedure suggested the optimal value of $\lambda$ being reasonably close to 0. Transformations of $x^q$, $q \in (0, 1)$ were not used in this paper since the focus is on extracting and distinguishing the multiplicative effects if such were present, instead of just normalizing the data. Additionally, log-transformation is useful since it does not heavily alter the interpretation of the variable, because in certain cases the differences of logarithmized data are very close to the percentage growth of the original data.

Further, some of the available data has relatively large spikes (outliers) at certain time periods, with a comparably small volatility during the other remaining time periods, therefore such a variable may be included to the final model not as an explanatory variable, but rather as a dummy variable, helping the model fitting some of the sudden shocks in the data, but providing no additional information to the forecasts. Therefore, an additional heuristic rule have been applied to filter such variables from the final dataset: the variable was not included in the final dataset if the ratio of maximum to average value, when adjusted by standard deviation, was larger than 10. It was found that the inclusion of such variables to the final dataset resulted in much worse forecasting accuracy, especially during the crisis periods, when they were included in the models as dummy variables to explain the sudden shock.

### B Main variables used in the modelling

Table 9: Acronyms and full names of all of the variables, used in presenting the top preselected variables in the pseudo-real-time experiments, together with the transformation applied and the publication lag of the variable. Ordinary acronym corresponds to the one used in the FRED’s database, while the addition of "IFS" denotes the source of the data being the IMF IFS database.

| Acronym         | Transf. | Lag | Name                                                                 |
|-----------------|---------|-----|----------------------------------------------------------------------|
| AUTHNOTT        | $\Delta \ln$ | 1   | New Privately-Owned Housing Units Authorized, but Not Started: Total |
| AENAN           | $\Delta$ | 1   | All Employees: Total Nonfarm in Arizona                             |
| CES4244110001   | $\Delta \ln$ | 1   | All Employees: Retail Trade: Automobile Dealers                     |
| CEU2000000001   | $\Delta \ln$ | 1   | All Employees: Construction                                          |
| CEU2023800001   | $\Delta \ln$ | 1   | All Employees: Construction: Specialty Trade Contractors             |
| CEU6562000001   | $\Delta$ | 1   | All Employees: Education and Health Services: Health Care and Social Assistance |
| COMPU1USA       | $\Delta \ln$ | 1   | New Privately-Owned Housing Units Completed: 1-Unit Structures       |
| CRESTKPIXLTRM-159SFRBATL | $\Delta \ln$ | 1   | Sticky Price Consumer Price Index less Food, Energy, and Shelter     |
| CUUR08000SEFR   | $\Delta$ | 1   | Consumer Price Index for All Urban Consumers: Sugar and sweets       |
| CUU4243SAH      | $\Delta$ | 1   | Consumer Price Index for All Urban Consumers: Housing in Seattle-Tacoma-Bremerton, WA (CMSA) |
| DPCERA3M086SBEA | $\Delta \ln$ | 1   | Real personal consumption expenditures excluding food and energy      |
| DSERRA3M086SBEA | $\Delta \ln$ | 1   | Real personal consumption expenditures: Services (chain-type quantity index) |

---

7 It can be noted that no series required more than 2 differences taken.
8 The transformation in this paper was applied if the estimated value of $\lambda$ was smaller than 0.8.
9 It is even worse if the sudden shock is relatively recent, since it may strongly affect the individual forecasts of such a series.

25
| Acronym      | Transf. | Lag | Name                                           |
|--------------|---------|-----|-----------------------------------------------|
| FLUR         | ∆ln     | 1   | Unemployment Rate in Florida                  |
| FRBLMC1      | ∆       |     | Change in Labor Market Conditions Index       |
| HOUSTWNSA    | ∆       |     | Housing Starts in West Census Region          |
| HS1F1MW      | ∆ln     | 1   | New One Family Houses Sold in Midwest Census Region |
| IFP54106N    | ∆ln     | 1   | Industrial Production: Construction supplies  |
| IPDCONGD     | ∆ln     | 1   | Industrial Production: Durable Consumer Goods |
| IPG321N      | ∆ln     | 1   | Industrial Production: Durable manufacturing: Wood product |
| IPG331N      | ∆ln     | 1   | Industrial Production: Durable Goods: Agriculture, construction, and mining machinery |
| IPG344N      | ∆ln     | 1   | Industrial Production: Durable Goods: Semiconductor and other electronic component |
| IPG36212N    | ∆ln     | 1   | Industrial Production: Durable Goods: Truck trailer |
| KYAN         | ∆       |     | All Employees: Total Nonfarm in Kentucky      |
| KYUR         | ∆ln     | 1   | Unemployment Rate in Kentucky                 |
| LNS11300032  | ∆       |     | Labor Force Participation Rate: 20 years and over, Black or African American Women |
| LNS12500000  | ∆       |     | Employed, Usually Work Full Time              |
| MEURN        | ∆ln     | 1   | Unemployment Rate in Maine                    |
| MINAN        | ∆       |     | All Employees: Total Nonfarm in Michigan      |
| NEUR         | ∆ln     | 1   | Unemployment Rate in Nebraska                 |
| PCOCOUSDM    | ∆ln     | 1   | Global price of Cocoa                        |
| PCUS3339113339111Z4 | ∆ | 1 | Producer Price Index by Industry: Pump and Pumping Equipment Manufacturing: Industrial Pumps, Except Hydraulic Fluid Power Pumps |
| PERMIT1NSA   | ∆ln     | 1   | New Privately-Owned Housing Units Authorized by Building Permits: 1-Unit Structures |
| PORT941NA    | ∆       |     | All Employees: Total Nonfarm in Portland-Vancouver-Hillsboro, OR-WA (MSA) |
| RALACBM027SBOG | ∆ln | 1 | Residual (Assets Less Liabilities), All Commercial Banks |
| RUR          | ∆ln     | 1   | Unemployment Rate in Rhode Island             |
| SFTPPINDM114SFRBF | ∆ln | 1 | San Francisco Tech Pulse                      |
| SM27346000000000026 | ∆ | 1 | All Employees: Total Nonfarm in Minneapolis-St. Paul-Bloomington, MN-WI (MSA) |
| SMU06310805000000001SA | ∆ln | 1 | All Employees: Information in Los Angeles-Long Beach-Anaheim, CA (MSA) |
| UNDCON1USA   | ∆ln     | 1   | New Privately-Owned Housing Units Under Construction: 1-Unit Structures |
| UNDCONTSA    | ∆ln     | 1   | New Privately-Owned Housing Units Under Construction: Total |
| USCONS       | ∆       |     | All Employees: Construction                   |
| USTRADE      | ∆ln     | 1   | All Employees: Retail Trade                   |
| VAPHCI       | ∆       |     | Coincident Economic Activity Index for Virginia |
| VAUR         | ∆ln     | 1   | Unemployment Rate in Virginia                 |
| VUR          | ∆ln     | 1   | Unemployment Rate in Vermont                 |
| W875RX1      | ∆ln     | 1   | Real personal income excluding current transfer receipts |
| WPS054321    | ∆ln     | 1   | Producer Price Index by Commodity for Fuels and Related Products and Power: Industrial Electric Power |
| WPUP066      | ∆ln     | 1   | Producer Price Index by Commodity for Chemicals and Allied Products: Plastic Resins and Materials |
| AHETPI       | ∆²ln    | 1   | Average Hourly Earnings of Production and Nonsupervisory Employees: Total Private |
| AZPHCI       | ∆²ln    | 1   | Coincident Economic Activity Index for Arizona |
| COPHCI       | ∆²ln    | 1   | Coincident Economic Activity Index for Colorado |
| CNI160V      | ∆²ln    | 1   | Civilian Noninstitutional Population          |
| MVMT19027MNFRBDAL | ∆²ln | 1 | Market Value of Marketable Treasury Debt |
| NCPHICI      | ∆²      |     | Coincident Economic Activity Index for North Carolina |
| SCMFG        | ∆²      |     | All Employees: Manufacturing in South Carolina |
| USPHICI      | ∆²      |     | Coincident Economic Activity Index for the United States |
| WINAN        | ∆²      |     | All Employees: Total Private Industries       |
| WPSFD4131    | ∆²      |     | Producer Price Index by Commodity for Final Demand: Finished Goods Less Foods and Energy |
| A33SNO       | ∆ln     | 2   | Value of Manufacturers' New Orders for Durable Goods Industries: Machinery |
| ANXAVS       | ∆       | 2   | Value of Manufacturers' Shipments for Capital Goods: Nondefense Capital |
| BOPSEX       | ∆ln     | 2   | Exports of Services, Balance of Payments Basis |
| BOPSIMP      | ∆ln     | 2   | Imports of Services, Balance of Payments Basis |
| BOPTEXP      | ∆ln     | 2   | Exports of Goods and Services, Balance of Payments Basis |
| BOXTVL133S   | ∆ln     | 2   | U.S. Exports of Services - Travel             |
| IPS_FILIBOR_1M_PA | ∆ln | 2 | Interest Rates, London Interbank Offer Rate, 1-Month, Percent per Annum |
Table 9: (continued)

| Acronym                        | Transf. | Lag | Name                                                                 |
|--------------------------------|---------|-----|----------------------------------------------------------------------|
| IFS_FILIBOR_TV_PA             | ∆ ln    | 2   | Interest Rates, London Interbank Offer Rate, 1-Year, Percent per Annum |
| IMPJP                          | ∆       | 2   | U.S. Imports of Goods from Japan, Customs Basis                      |
| INTDSRIPM190N                  | ∆ ln    | 2   | Interest Rates, Discount Rate for Japan                              |
| MANCSQJURN                     | ∆ ln    | 2   | Unemployment Rate in Manchester, NH (NECTA)                          |
| S4233SM144NCEN                 | ∆ ln    | 2   | Merchant Wholesalers, Except Manufacturers’ Sales Branches and Offices |
| U34HUGU                        | ∆ ln    | 2   | Value of Manufacturers’ Unfilled Orders for Durable Goods Industries: Computers and Electronic Products: Electronic Components |
| UMDMIS                         | ∆ ln    | 2   | Ratio of Manufacturers’ Total Inventories to Shipments for Durable Goods Industries |
| XTEXVA01CHM657S                | ∆ ln    | 2   | Exports: Value Goods for Switzerland                                |
| XTEXVA01JPM664N                | ∆ ln    | 2   | Exports: Value Goods for the United States                          |
| BOERUKM                        | ∆       | 3   | Bank of England Policy Rate in the United Kingdom                    |
| KORPROINDMISMEI               | ∆ ln    | 3   | Production of Total Industry in Korea                                |
| MABMMJ301AUM657S              | ∆ ln    | 3   | M3 for Australia                                                    |
| IFS_TMG_R_CIF_IX               | ∆ ln    | 3   | Goods, Volume of Imports, Index                                     |
| IFS_TXG_R_PDB_IX               | ∆ ln    | 3   | Goods, Volume of Exports, US Dollars, Index                         |
| IR3TIB01PLM156N                | ∆ ln    | 3   | 3-Month or 90-day Rates and Yields: Interbank Rates for Poland      |
| HILILTO1CHM156N                | ∆       | 3   | Long-Term Government Bond Yields: 10-year: Main (Including Benchmark) for Switzerland |
| SPASTT01IDEM657N               | ∆ ln    | 3   | Total Share Prices for All Shares for Germany                       |
| SPASTT01KRM661N                | ∆ ln    | 3   | Total Share Prices for All Shares for the Republic of Korea         |
| VALEXPKRM052N                  | ∆ ln    | 3   | Goods, Value of Exports for Republic of Korea                       |

C Tables

Table 10: RMSE of models forecasts during pseudo-real-time experiments for Gross Fixed Capital Formation, here the bolded values are the smallest ones for every row, and for every block the last line denotes the total forecast accuracy for the full time period of 2005Q1-2014Q4.

| Backcast | AdaPCA15 | AdaRL15 | AdaPCA15X | AdaPCA20 | AdaRL20 | AdaPCA20X | AdaPCA25 | AdaRL25 | AdaPCA25X |
|----------|----------|---------|-----------|----------|---------|-----------|----------|---------|-----------|
| 05-08    | 0.554    | 0.705   | 0.470     | 0.528    | 0.598   | 0.492     | 0.514    | 0.569   | 0.421     |
| 08-11    | 0.899    | 1.048   | 0.669     | 0.798    | 1.027   | 0.614     | 0.749    | 1.038   | 0.561     |
| 11-15    | 0.775    | 0.983   | 0.645     | 0.669    | 0.909   | 0.584     | 0.717    | 0.883   | 0.607     |
| 05-15    | 0.730    | 0.902   | 0.589     | 0.637    | 0.846   | 0.543     | 0.620    | 0.832   | 0.591     |

| Nowcast  | AdaPCA15 | AdaRL15 | AdaPCA15X | AdaPCA20 | AdaRL20 | AdaPCA20X | AdaPCA25 | AdaRL25 | AdaPCA25X |
|----------|----------|---------|-----------|----------|---------|-----------|----------|---------|-----------|
| 05-08    | 0.821    | 0.832   | 0.918     | 0.799    | 0.776   | 0.827     | 0.746    | 0.791   | 0.765     |
| 08-11    | 1.008    | 1.524   | 1.539     | 1.510    | 1.553   | 1.346     | 1.520    | 1.550   | 1.477     |
| 11-15    | 1.147    | 1.170   | 1.148     | 1.128    | 1.183   | 1.136     | 1.125    | 1.175   | 1.159     |
| 05-15    | 1.200    | 1.195   | 1.191     | 1.150    | 1.191   | 1.112     | 1.133    | 1.199   | 1.137     |

| Fore1Q   | AdaPCA15 | AdaRL15 | AdaPCA15X | AdaPCA20 | AdaRL20 | AdaPCA20X | AdaPCA25 | AdaRL25 | AdaPCA25X |
|----------|----------|---------|-----------|----------|---------|-----------|----------|---------|-----------|
| 05-08    | 0.979    | 1.052   | 1.017     | 0.987    | 0.991   | 1.026     | 0.993    | 1.065   | 1.009     |
| 08-11    | 2.210    | 2.280   | 2.116     | 2.159    | 2.221   | 2.120     | 2.217    | 2.316   | 2.170     |
| 11-15    | 1.150    | 1.048   | 1.150     | 1.125    | 1.106   | 1.129     | 1.141    | 1.099   | 1.156     |
| 05-15    | 1.530    | 1.547   | 1.485     | 1.492    | 1.519   | 1.483     | 1.511    | 1.555   | 1.499     |

| Fore2Q   | AdaPCA15 | AdaRL15 | AdaPCA15X | AdaPCA20 | AdaRL20 | AdaPCA20X | AdaPCA25 | AdaRL25 | AdaPCA25X |
|----------|----------|---------|-----------|----------|---------|-----------|----------|---------|-----------|
| 05-08    | 1.120    | 1.217   | 1.134     | 1.181    | 1.152   | 1.223     | 1.236    | 1.206   | 1.261     |
| 08-11    | 2.531    | 2.665   | 2.494     | 2.566    | 2.591   | 2.540     | 2.631    | 2.689   | 2.585     |
| 11-15    | 1.059    | 1.021   | 1.059     | 1.048    | 1.001   | 1.061     | 1.064    | 1.012   | 1.082     |
| 05-15    | 1.694    | 1.771   | 1.679     | 1.710    | 1.721   | 1.710     | 1.747    | 1.782   | 1.738     |

Table 11: Relative (to ARMA models’) RMSE of models forecasts during rolling window pseudo-real-time experiments for PFCE, Exports and Imports; here the bolded values are the smallest ones for every row during the full time period of 2005Q1-2014Q4.

|                      | Sqrt | LASSO | Adaptive | PCA | AdaPCA | Relaxed | ARMA |
|----------------------|------|-------|----------|-----|--------|---------|------|
| PFCE                 |      |       |          |     |        |         | 1    |
| Fore1Q               | 0.83 | 0.83  | 0.78     | 0.83| 0.77   | 0.81    | 1    |
| Fore2Q               | 0.85 | 0.86  | 0.83     | 0.84| 0.81   | 0.83    | 1    |

27
Table 11: (continued)

|          | Sqrt | LASSO | Adaptive | PCA | AdaPCA | Relaxed | ARMA |
|----------|------|-------|----------|-----|--------|---------|------|
| **Exports** |      |       |          |     |        |         |      |
| Nowcast  | 0.76 | 0.80  | 0.73     | 0.70| 0.66   | 0.77    | 1    |
| Fore1Q   | 0.87 | **0.82** | 0.90     | 0.90| 0.90   | 1.10    | 1    |
| Fore2Q   | 1.01 | 0.99  | 0.89     | 0.98| **0.89** | 1.16    | 1    |
| **Imports** |      |       |          |     |        |         |      |
| Nowcast  | 0.50 | 0.59  | 0.54     | **0.47** | 0.48 | 0.51    | 1    |
| Fore1Q   | 0.63 | 0.68  | 0.65     | 0.64| 0.65   | 0.64    | 1    |
| Fore2Q   | 0.79 | 0.81  | 0.80     | 0.80| 0.79   | **0.78** | 1    |

Table 12: This table reports the p-value of the *Giacomini-White* test for equal predictive ability with squared differences for Private Final Consumption Expenditure. The null hypothesis is that the column model has the same forecasting performance as of the row model against a two-sided alternative. Bolded values mark p-values smaller than 0.1.

|          | Sqrt | LASSO | Adaptive | PCA | AdaptivePCA | Relaxed | ARMA |
|----------|------|-------|----------|-----|-------------|---------|------|
| **Nowcast** |      |       |          |     |             |         |      |
| Sqrt    | -    | 0.43  | 0.68     | 0.53| 0.16        | 0.13    | **0.068** |
| LASSO   | -    | -     | 0.26     | **0.076** | 0.41 | 0.29    | 0.22 |
| Adaptive| -    | -     | -        | 0.9  | 0.36        | 0.61    | **0.082** |
| PCA     | -    | -     | -        | 0.32 | -           | 0.77    | **0.05**  |
| AdaptivePCA | -   | -     | -        | -   | -           | 0.17    | **0.044** |
| Relaxed | -    | -     | -        | -   | -           | -       | -      |
| ARMA    | -    | -     | -        | -   | -           | -       | -      |
| **1Q** |      |       |          |     |             |         |      |
| Sqrt    | -    | 0.0014 | 0.055 | **0.071** | 0.12 | 0.0048 | 0.54 |
| LASSO   | -    | -     | 0.16     | 0.32| 0.15        | 0.3     | 0.48 |
| Adaptive| -    | -     | -        | 0.33| 0.34        | 0.34    | 0.33 |
| PCA     | -    | -     | -        | 0.11| 0.63        | 0.43    | -     |
| AdaptivePCA | -     | -    | -        | -   | -           | 0.24    | 0.19 |
| Relaxed | -    | -     | -        | -   | -           | -       | 0.37 |
| ARMA    | -    | -     | -        | -   | -           | -       | -      |
| **2Q** |      |       |          |     |             |         |      |
| Sqrt    | -    | 0.33  | 0.35     | 0.19| **0.09**    | 0.036   | 0.7  |
| LASSO   | -    | -     | 0.36     | 0.31| 0.26        | 0.34    | 0.47 |
| Adaptive| -    | -     | -        | 0.87| 0.53        | 0.23    | 0.67 |
| PCA     | -    | -     | -        | 0.64| 0.21        | 0.21    | 0.68 |
| AdaptivePCA | -    | -     | -        | -   | -           | 0.28    | 0.38 |
| Relaxed | -    | -     | -        | -   | -           | -       | 0.65 |
| ARMA    | -    | -     | -        | -   | -           | -       | -      |

Table 13: This table reports the p-value of the *Giacomini-White* test for equal predictive ability with squared differences for Exports. The null hypothesis is that the column model has the same forecasting performance as of the row model against a two-sided alternative. Bolded values mark p-values smaller than 0.15.

|          | Sqrt | LASSO | Adaptive | PCA | AdaptivePCA | Relaxed | ARMA |
|----------|------|-------|----------|-----|-------------|---------|------|
| **Nowcast** |      |       |          |     |             |         |      |
| Sqrt    | -    | 0.3   | 0.43     | 0.34| 0.39        | 0.88    | **0.04** |
| LASSO   | -    | -     | 0.29     | 0.34| 0.23        | 0.88    | **0.11** |
| Adaptive| -    | -     | -        | 0.87| 0.53        | 0.23    | 0.15 |
| PCA     | -    | -     | -        | **0.64** | 0.36 | 0.11    | -     |
| AdaptivePCA | -     | -    | -        | -   | -           | 0.23    | 0.19 |
| Relaxed | -    | -     | -        | -   | -           | -       | 0.16 |
| ARMA    | -    | -     | -        | -   | -           | -       | -     |
| **1Q** |      |       |          |     |             |         |      |
| Sqrt    | -    | 0.5   | 0.54     | 0.6  | 0.66        | 0.32    | 0.48 |
| LASSO   | -    | -     | 0.36     | 0.38| 0.37        | 0.34    | 0.23 |
| Adaptive| -    | -     | -        | 0.97| 0.86        | 0.3     | 0.89 |
| PCA     | -    | -     | -        | 1   | 0.34        | 0.44    | **0.73** |
| AdaptivePCA | -     | -    | -        | -   | -           | 0.31    | 0.85 |
| Relaxed | -    | -     | -        | -   | -           | -       | 0.45 |
| ARMA    | -    | -     | -        | -   | -           | -       | -     |
| **2Q** |      |       |          |     |             |         |      |
| Sqrt    | -    | 0.3   | 0.35     | 0.6  | 0.33        | 0.5     | 0.48 |
| LASSO   | -    | -     | 0.42     | **0.12** | 0.39 | 0.45    | 0.38 |
Table 13: (continued)

|          | Sqrt | LASSO | Adaptive | PCA  | AdaptivePCA | Relaxed | ARMA |
|----------|------|-------|----------|------|-------------|---------|------|
| Adaptive | -    | -     | 0.28     | 0.54 | 0.43        | 0.14    |
| PCA      | -    | -     | 0.65     | 0.15 | 0.53        | -       |
| Relaxed  | -    | -     | -        | -    | -           | -       |
| ARMA     | -    | -     | -        | -    | -           | -       |

Table 14: This table reports the p-value of the Giacomini-White test for equal predictive ability with squared differences for Imports. The null hypothesis is that the column model has the same forecasting performance as of the row model against a two-sided alternative. Bolded values marks p-values smaller than 0.1.

|          | Sqrt | LASSO | Adaptive | PCA  | AdaptivePCA | Relaxed | ARMA |
|----------|------|-------|----------|------|-------------|---------|------|
| Sqrt     | 0.28 | 0.37  | 0.45     | 0.66 | 0.79        | 0.093   |
| LASSO    | -    | 0.43  | 0.4      | 0.25 | 0.27        | 0.42    |
| Adaptive | -    | -     | 0.46     | 0.12 | 0.42        | 0.078   |
| PCA      | -    | -     | 0.49     | 0.35 | 0.49        | 0.088   |
| AdaptivePCA | -  | -     | -        | -    | 0.31        | 0.06    |
| Relaxed  | -    | -     | -        | -    | -           | 0.099   |
| ARMA     | -    | -     | -        | -    | -           | -       |

1Q

|          | Sqrt | LASSO | Adaptive | PCA  | AdaptivePCA | Relaxed | ARMA |
|----------|------|-------|----------|------|-------------|---------|------|
| Sqrt     | -    | 0.11  | 0.22     | 0.52 | 0.62        | 0.14    |
| LASSO    | -    | 0.11  | 0.14     | 0.097| 0.096       | 0.18    |
| Adaptive | -    | -     | 0.24     | 0.43 | 0.29        | 0.14    |
| PCA      | -    | -     | 0.45     | 0.4  | 0.22        | 0.14    |
| AdaptivePCA | - | -     | -        | -    | 0.53        | 0.1     |
| Relaxed  | -    | -     | -        | -    | -           | 0.14    |
| ARMA     | -    | -     | -        | -    | -           | -       |

2Q

|          | Sqrt | LASSO | Adaptive | PCA  | AdaptivePCA | Relaxed | ARMA |
|----------|------|-------|----------|------|-------------|---------|------|
| Sqrt     | -    | 0.26  | 0.14     | 0.51 | 0.37        | 0.6     | 0.21 |
| LASSO    | -    | 0.77  | 0.17     | 0.67 | 0.13        | 0.35    |
| Adaptive | -    | -     | 0.18     | 0.15 | 0.15        | 0.32    |
| PCA      | -    | -     | 0.65     | 0.48 | 0.2         | 0.2     |
| AdaptivePCA | - | -     | -        | -    | 0.62        | 0.25    |
| Relaxed  | -    | -     | -        | -    | -           | 0.22    |
| ARMA     | -    | -     | -        | -    | -           | -       |
Figure 5: The results of forecasting accuracy during the pseudo-real-time experiments over 2011-2014 with the set of preselected variables being fixed for the whole period, and the numbers (1.) and (2.) enumerating the different sets of variables used.