VILNIAUS UNIVERSITETAS

MATEMATIKOS IR INFORMATIKOS FAKULTETAS

Magistrinis darbas

Retų struktūrų analizė vertinant trumpalaikes BVP komponenčių prognozes

Sparse Structure Analysis with Applications to Short-Term Forecasting of the GDP Components

Saulius Jokubaitis

VILNIUS 2017

MATEMATIKOS IR INFORMATIKOS FAKULTETAS
EKONOMETRINĖS ANALIZĖS KATEDRA

Darbo vadovas dr. Dmitrij Celov

Darbas apgintas ____________________

Registravimo NR. ____________________

Retų struktūrų analizė vertinant trumpalaikes BVP komponenčių prognozes

Santrauka

Šiame darbe vertinamos trumpalaikės JAV BVP komponenčių išlaidų metodu prognozės, siekiant rodiklius įvertinti anksčiau, negu juos paskelbia oficialios statistikos institucijos. Tuo tikslu formuojamos išankstinės bei 1 ir 2 ketvirčių prognozės, naudojant mėnesinius indikatorius, skelbiamus su mažu uždelsimu. Sprendžiant didelės naudojamų duomenų dimensijos problemą darbe vadovaujamasi retos struktūros prielaida, tariant, jog tik keletas indikatorių yra svarbūs prognozuojant BVP komponentes. Kintamųjų atrankai ir prognozių vertinimui taikomas LASSO metodas, kartu su keliomis populiariomis jo modifikacijomis. Papildomai, darbe pasiūloma LASSO modifikacija, apjungianti LASSO bei pagrindinių komponenčių metodus, siekiant suteikti papildomo prognozavimo tikslumo. Modelių prognozavimo tikslumas įvertinamas atliekant pseudo-realiuosius eksperimentus nuo 2005 iki 2015, prognozuojant keturias BVP komponentes, rezultatai lyginami su ARMA prognozėmis. Gauti rezultatai byloja, jog LASSO metodai geba reikšmingai pagerinti BVP komponenčių prognozes bei atpažinti svarbiausius aiškinančiuosius kintamuosius. Pasiūlyta LASSO modifikacija kai kuriais atvejais suteikia papildomo prognozių tikslumo.

Raktiniai žodžiai : LASSO, išankstinis prognozavimas, pagrindinės komponentės, kintamųjų atranka, BVP komponentės

Sparse Structure Analysis with Applications to Short-Term Forecasting of the GDP Components

Abstract

The aim of this thesis 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

Contents

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 various its 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. One such example is the method of Bridge Equations, based on a single equation or a small scale systems of equations, capable of bridging the available high-frequency information with quarterly national accounts data. These methods are extensively used by practitioners and researchers due to its statistical simplicity and adequate performance under a carefully selected set of explanatory variables.

In this thesis we further follow 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) or Stakėnas (2012)) 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 thesis 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 underlying 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 Scheufele (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 modelling 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 subcomponents. 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 thesis is organized as follows: in chapter 1 we briefly review the problematic of nowcasting and various solutions when forecasting economic data, also we present a detailed overview of the LASSO and some of its popular modifications found in the literature. In chapter 2 we briefly describe the design of the information set used in the pseudo-real-time exercise. In chapter 3 we describe our proposed modification of combining the LASSO with principal components, where in the chapter 4 we present the results of pseudo-real-time exercise, during which 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.

Chapter 1

Literature review

1.1 Review of nowcasting

The definition of the term *nowcasting*, used in this thesis, can be understood as suggested by Banbura et al. (2013), who define it as the prediction of the present, the very near future and the very recent past. Nowcasting is relevant in macroeconomics, especially if we are interested in timely estimates of key macroeconomic indicators, which are available only with a significant delay. This is particularly true for those collected on a quarterly basis, such as the GDP and its components. For example, the first official estimate of the GDP in the US is published approximately one month after the end of the reference quarter, while in the Euro area the corresponding publication lag is 2-3 weeks longer.

The basic idea of nowcasting is to exploit the information, published comparatively earlier and possibly at higher frequencies than the target variable of interest in order to obtain an early estimate before the official figure is released. Such *soft* information may come in the form of various surveys and financial variables, which are usually forward oriented and can therefore indicate expectations of the participants in various markets. Additionally, plenty of *hard* information is available at a monthly frequency, such as the personal consumption, industrial production or indicators for imports and exports. Both types of indicators may provide an early indication of current signals in economic activity. When the timeliness of the data is especially important, one can entail the information from various surveys, which are usually the first monthly release of the quarter, or financial variables, released at a very high frequency.

Additionally, nowcasting may be of great interest when recognizing the fact, that in many cases the first releases of the GDP (i.e. its flash estimates) are themselves only preliminary estimates, since at the time of release the full information is usually not available to the agencies of national statistics, so incomplete and uncertain data is used. Therefore those figures are subject for potential revisions, making first estimates not entirely reliable and accurate[1]. Additionally, the revisions, while not that substantial, may continue for quite some time after the first release: usually the first estimate of the GDP is measured on an output basis only (measuring the supply side by production approach), while the final data is adjusted according to its output-based and expenditure-based (measuring the demand side) estimates, in addition to *benchmarking* and *reconciliation* with annual data (Bloem et al. (2001)). Having this in mind, economic agents are facing a filtering problem in forming their own optimal estimate, therefore it is of great value having the tools for an alternative timely nowcast estimation by using all of the available information at a certain time. However, Castle and Hendry (2013) points out that not all of the information available at a timely matter can project a good early nowcast[2], especially during a structural

[1]For example, Faust et al. (2005) report that revisions to GDP announcements are quite large in all G7 countries, mainly because of reversion to the mean, which they interpret as due to removing measurement noise.

[2]The authors cite a case during the financial crisis, where the models of the UK Office for National Statistics had "broken down", resulting in even less reliable estimates than normal.

break in the economy. Therefore it is important to develop a robust methodology, able to both detect and adapt to such breaks.

Nowcasting essentially involves obtaining statistically efficient projections of a modelled variable of interest on the available information set Ω_v, where v denotes the time of a particular data release (a vintage). It is worth noting that v is not related to the time index t, when describing the models, since due to frequent and asynchronous update and release time schedules of various indicators may result in the index v being of very high frequency and of different intervals (for example, more than one update of the data can occur during one single day). The information set Ω_v can be composed of data at a wide range of frequencies, from daily to annual. Also, as most variables are released asynchronously and with different publication lags, the time of the last available observation differs from series to series. Since in nowcasting it is useful to exploit all of the available information, this results in a "ragged-edge" problem of the information set Ω_v.

Formally, Banbura et al. (2013) define the information set as

$$\Omega_v := \{y_{t,j}^{k_j} : t = k_j, 2k_j, \ldots, T_j(v), \quad j = 1, 2, \ldots, J\},$$

where $y_{t,j}^{k_j}$ is the j-th time series from the total of J at the time moment t, observed at the frequency of k_j. Here without the loss of generality we can assume that our modelled variables of interest $y_{t,i}^{k_i}$, for some values of i, are also included in the information set. Also, the value of k here depends on the context: if both modelled and explanatory variables are of a monthly frequency, $k = 1$; if the explanatory variable is of a quarterly frequency – $k = 3$; daily – $k = 1/22$ and so on. $T_j(v)$ is a multiple of k_j, depending on the time period of the last available observation at a certain vintage v. Due to the asynchronous nature of the vintages and different frequencies of the data, in general $T_m(v) \neq T_n(v)$ for some $m \neq n$, which is the formal definition of the "ragged-edge" problem of the dataset.

1.2 Modelling economic data

In the literature there is no one best way to accurately model the economic indicators, therefore it is common to distinguish them into three main modelling groups: (Castle et al. (2013))

- derived from various economic theories;
- based on information extracted from other economic indicators or their factors;
- based on information contained in the variable of interest (also known as the mechanistic approach).

From the first group of methods, currently the most widely used are the Dynamic Stochastic General Equilibrium (DSGE) models, especially favoured by various Central Banks (see, i.e., Smets and Wouters (2002)). The equations of the model are determined and the restrictions are set in such a way, that the behaviour of both the consumers and the producers would fit a certain macroeconomic theory (i.e., the Real Business Cycle (RBC) theory, based on the neoclassical Solow-Swan economic growth model; or various modifications of the RBC, such as the modification by the New Keynesians Rotemberg and Woodford, and others). More recently, a continuing growth in popularity can be seen of the Global-VAR models (Dees et al. (2005), Pesaran et al. (2009)), with the help of which it is possible to define a structural model of some economic environment (i.e. the Euro area) and estimate the endogenous relationships not only inside one country, but also between various different countries of the whole economic environment. Such supplementary information can allow for a more accurate estimation of various indicators of interest, leading to a more accurate forecasting.

The second approach is based not so much on the dominant macroeconomic theories, but rather on the available and attainable information. The main underlying idea is that the growth of a certain economic indicator is likely to be a result of some particular signal or an impulse, felt in a part of the economy (i.e. in one or more particular markets), which, perhaps, in some markets can even be sensed at an earlier time – therefore the goal is to estimate the signal as accurately as possible (as a latent variable) and use it in forecasting. Recently the most popular methods of this approach are the Factor Models (also known as the Diffusion Index models, giving the extracted factors the interpretation of the diffusion indices from the Real Business Cycle theory), which are able to estimate the latent signal by using the method of principal components (Stock and Watson (2002), Stock and Watson (2009)).

Another popular method – the Bridge Equations (BE) method, in which a single regression is formed for every modelled variable of a lower frequency (often of a quarterly measure) against additional explanatory variables of a higher frequency (typically of a monthly measure), aggregated to the frequency of the modelled variable (see Hahn and Skudelny (2008), Dreschel and Scheufele (2013)). While being relatively simple, these methods are still often used by various institutions due to their convenient features: first, such methods maintain a good balance between the complexity of the model and the accuracy of the resulting forecasts, since in most cases a small, appropriately chosen set of explanatory variables can ensure a good forecast accuracy. Second, since the regression models are usually formed using only one single equation, the resulting forecasts can be easily decomposed according to the explanatory variables used, thus allowing for an easy way of explaining a story behind the particular form of the resulting forecasts, which can be of great use when communicating the results to the decision makers. Third, the convenient linear form of the regression model allows forming a straightforward relation between the forecast errors and the underlying indicators, thus providing a way to distinguish which variables are under-performing and may need refinement when forecasting them in their original frequency, or which variable may potentially be included in the regression only due to a spurious relation. (Bulligan et al. (2015))

On the other hand, BE models present two important drawbacks: first, since the models are usually formed by using a parsimonious set of explanatory variables, in order to achieve accurate forecasts, the data generating process should be, in principle, sparse. However, in reality such a fact is rarely ever known for sure, therefore it is likely that BE models can potentially leave out parts of useful information. In addition to this, the problem can become even more serious when the modelled variable is of a low frequency (i.e., quarterly) and the size of the historical data is not large enough – in such case, even if we have an appropriate set of indicators required for accurate modelling, some of the variables just could not be included to the regression due to the large resulting modelling errors, caused by increased inefficiency of estimating the model's parameters under a low number of degrees of freedom. That is, over time, with the size of the historical time series growing, due to the estimation of the coefficients getting more and more efficient, the structure of the model can change drastically, which may as a consequence decrease the forecasting performance. Second, the estimation of the BE models (and the choice of a particular set of explanatory variables used) often depends from the subjective judgement and experience of the econometrician. (Bulligan et al. (2015))

However, the factor models are perfectly able to address both drawbacks of the BE method: principal components can be extracted from a significantly larger set of indicators, moreover, some blocks of information can be accurately approximated by a small number of (orthogonal) factors, which by itself increases the efficiency of estimating the coefficients of the model. Additionally, both the methods of extracting the common factors from the data and the ways of performing rotations of the extracted components are largely automated and based on specific algorithms (Giannone et al. (2008), Angelini et al. (2008)), thus the only place where the subjectivity of the econometrician can have an active role is the choice of a number of principal components and the type of rotation to be used (or the decision not to use the rotations at all). When the performance of these two methods is compared on real data, Angelini et al. (2008) find that on average the factor models are able to generate overall more accurate forecasts than the BE models when forecasting the GDP of the Euro area.

On the other hand, the aforementioned shortcomings can be solved without leaving the BE methodology. Indeed, Bulligan et al. (2015) claim that a large amount of information can be used by applying soft-thresholding methods, i.e., the *least absolute shrinkage selection operator* (LASSO), which is able to solve the high-dimensionality problem by shrinking the coefficients of the significant explanatory variables, with conveniently setting the coefficients of the insignificant variables to zero. The authors claim that recently, with the amount of available information getting larger every day, the factor models can miss the underlying important signal, since the factors are usually extracted "blindly". That is, if for some modelled variable the significant latent factor is dominant only in a small set of variables, it can easily become overshadowed by other, much stronger and more expressed signals from larger sets of indicators (the problem is also known as the "too much information" problem). For this reason, it is suggested in the literature to preselect only a subset of the whole information by using the LASSO procedure (Bai and Ng (2008)). Since LASSO takes into account the information of the modelled variable, the explanatory variables selected by LASSO as significant are more likely to include the true leading signals. On the other hand, De Mol et al. (2006) show in their research that in cases when the data is characterised by strong multicollinearity (which is common for macroeconomic data), when the dimension of the data is growing over time, the forecasts of the factor models are seen to be increasingly more correlated with the forecasts of the LASSO model. This conclusion suggests that in such cases the performance of the BE models might be on par with the performance of factor models, as was seen in the cited paper. Moreover, under a high-dimensional dataset, since the factors, formed by the factor models, are basically linear combinations of all of the variables from the dataset, it is likely that some parts of the extracted signal will be consisting of noise, while the LASSO method tends to shrink-down or even restrict to zero the coefficients of the noise variables (asymptotically, under specific conditions, see appendix C), which can possibly lead to a more clearly extracted signal.

The third approach of modelling macroeconomic data is by only using the underlying information of the modelled variable. Some of the most widely used methods of this type are the exponential moving weighted average, the *Holt-Winters* method, ARIMA time series methods and many others (Makridakis and Hibon (2000)), which, while appear not to use a lot of the available information, can in some cases generate the most accurate forecasts. For example, Stock and Watson (2010), when forecasting inflation, find that the use of ordinary time series methods (such as the random walk model) may result in more accurate forecasts than the forecasts, generated by models with additional exogenous information included. Similar results are achieved by Castle et al. (2013), where the performance of factor models and time series models is compared by their accuracy in forecasting the US GDP – in some cases the performance of time series methods is just as good as of the factor models, therefore they are a suitable benchmark to be considered when evaluating the forecast accuracy of various models.

Overall, the factor models have been the main workhorse for the economists for a long time when forecasting macroeconomic indicators, however, it is clear that this method is not perfect and there is room for refinement. In a way, we can argue that since this method is based on the extraction of principal components, which by itself is "blind" and unsupervised, by giving it some additional control and setting it to the right direction we could gain some additional accuracy. It seems that the LASSO method, recently growing in popularity among both the practitioners and the academics, can provide the desired additional accuracy, especially when dealing with high-dimensional problems.

1.3 Review of the LASSO methods

Assume that n is the number of observations of the modelled variable Y_i, here $i = 1, \ldots, n$, p is the total number of explanatory variables X_j used, $j = 1, \ldots, p$, and $X = (X_1, \ldots, X_p)$. Additionally, we use the following notation for an ℓ_q norm, $q \geq 1$, throughout the thesis: 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 ℓ_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. Together with it (as alternatives to the Forward Stepwise, the Backward Stepwise or even the Best Subset Selection methods) there are several other similar methods proposed in the literature, capable of dealing with high-dimensionality problem: the Nonnegative Garotte (Breiman (1995), Yuan and Lin (2007)), SCAD (Fan and Li (2001)), Elastic Net (Zou and Hastie (2005)), Dantzig Selector (Candes and Tao (2007)), all of which, by choosing appropriate hyperparameters, are able to restrict the insignificant parameters of the model to 0. However, out of all of these methods the LASSO 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 ℓ_1 norm, which solves the (1.1) problem:

$$\hat{\beta}_{LASSO} = \arg\min_{\beta} \sum_{i=1}^{n} (Y_i - X_i\beta)^2 + \lambda||\beta||_1, \tag{1.1}$$

where the hyperparameter $\lambda \in (0, \infty)$ is fixed. With the value of λ 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 ℓ_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. The whole solution path of the model can be solved by using the LARS algorithm. In a sense, we can see the LASSO as a stepwise regression, since with a large enough value of λ, 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 λ, 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 (see theorem C.2). 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 λ nor the sample size n. Additionally, they prove that the consistency of the variable selection by the LASSO requires that the value of λ 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 $\lambda = \lambda_n = \mathcal{O}(\sqrt{n})$ and under some additional conditions, formed 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 $|\mathcal{A}| = p_0 < p$, 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 δ. Then, by the definition, formed by Fan and Li (2001), the procedure δ is said to have Oracle Properties if for the estimator $\hat{\beta}(\delta)$ the following holds (asymptotically):

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

The authors claim that every adequate procedure, along with various other optimality conditions, should also have the Oracle Properties. In this case we can note that the LASSO does not have the Oracle Properties.

Additionally, Leng et al. (2006) have shown that in cases where the hyperparameter λ is chosen by minimizing the forecast errors of the resulting model, such variable selection performed is not consistent. That is, the final selected variables by the procedure are not necessarily from the true subset of variables, used by the data generating process. In other words, the probability to select the true underlying model is strictly restricted by a constant $C < 1$ even in cases when the used data matrix X is orthonormal under any sample size n. The intuition behind this result is simple – since the hyperparameter λ is directly related to the amount of shrinkage applied to the estimated coefficients, in order to minimize the resulting forecast errors the value of λ will be chosen such, that the amount of shrinkage applied on the mostly significant variables will be as small as the out-of-sample forecast errors allow. Therefore, the need to apply the smaller shrinkage than the one that would otherwise be chosen (note that $\lambda = 0$ would result in a simple OLS solution without any penalties) suggests choosing a smaller than optimal value of λ, thus resulting in a possible inclusion of noise variables in the final set of variables selected by the procedure.

Secondly, Tibshirani (2012) notes that despite the fact that LASSO is highly favoured among the practitioners in the cases of $p > n$, actually the optimization problem of the LASSO in such cases is not strictly convex, which results in the fact that there is no one unique solution to the problem except for the cases when the explanatory variables used in the modelling are generated from a continuous distribution function (in that case there exists a unique solution with probability equal to 1). Therefore, in this thesis, since we are working with continuous macroeconomic data, we can expect unique solutions even under the $p > n$ case. Additionally, due to the uniqueness of the solution of the optimization problem we should also expect adequate performance when selecting the significant explanatory variables due to the uniqueness of the set $\text{supp}(\hat{\beta})$.

In the literature there are many variations and modifications of the LASSO, all of which are trying to overcome 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:

$$\hat{\beta}_{adaLASSO} = \arg\min_{\beta} \sum_{i=1}^{n} (Y_i - X_i\beta)^2 + \lambda \sum_{j=1}^{p} w_j |\beta_j|, \tag{1.2}$$

where $w = (w_1, \ldots, w_p)$ is a vector of fixed weights. In the paper of Zou (2006) it is proven 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 the weights are suggested to be chosen as $\hat{w} := 1/|\hat{\beta}^*|^{\gamma}$, $\gamma > 0$, where $\hat{\beta}^*$ is $\sqrt{n}$-consistent estimator of β[3]. It is usually suggested to choose $\hat{\beta} := \hat{\beta}_{OLS}$, when $p < n$, however, when $p > n$ and the OLS estimate is infeasible or the data is strongly multicollinear, it is suggested to replace the $\hat{\beta}_{OLS}$ with $\hat{\beta}_{ridge}$, that is, with the estimated coefficients by the Ridge regression, defined by

[3] Though the authors note that this restriction can be weakened: let $\{b_n\} : b_n \to \infty$ and $b_n(\hat{\beta} - \beta) = \mathcal{O}_p(1)$. Then the oracle properties still hold if $\lambda_n = o(\sqrt{n})$ and $b_n^{\gamma}\lambda_n/\sqrt{n} \to \infty$.

the (1.1) problem just with additionally replacing the ℓ_1 norm of the imposed penalty with the norm of ℓ_2. Additionally, the $\hat{\beta}_{uni}$ estimates are suggested, which is a vector of coefficients, obtained by forming univariate OLS regressions by modelling Y_i against each of the explanatory variable $X_{i,j}$, $j = 1, \ldots, p$, separately.

The authors emphasise that in order to ensure that the Oracle Properties hold the formed weight vector $\hat{w}$ should be data-driven. 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. Therefore, this method allows for an (asymptotically) unbiased simultaneous estimation of large coefficient and small threshold estimates.

This model solves a convex optimization problem, therefore it does not have more than one local minimum point, and the global minimizer can be efficiently found by applying the usual efficient algorithms, used for solving the LASSO problem.

Zou (2006) proves that the Adaptive LASSO has the Oracle Properties under a fixed p, however, originally this method has been defined and proved under the case of $p < n$ by using the OLS method for estimating the weights w_j. Even though the use of Ridge regression is also allowed, it should be noted that the estimated coefficients of the Ridge regression are also dependent on the choice of its penalty parameter, which should also be appropriately chosen. Therefore an additional uncertainty is imposed, bringing in some possible instabilities by requiring two different hyperparameters to be estimated in order to solve the Adaptive LASSO optimization problem. Moreover, Huang et al. (2008) shows that the Oracle Properties can also hold when $p_n \to \infty, n \to \infty$ and stresses the importance of appropriate choice of the weight vector. The authors claim that $\hat{\beta}_{uni}$ allows to retain the Oracle Properties only in cases when the insignificant variables are weakly correlated with other insignificant variables, which is a rather strong assumption about the dataset, especially in cases when the lagged values of both the modelled and the explanatory variables are included. For this reason it is meaningful to research other possible choices for the optimal weights, in order to highlight the advantages of the Adaptive LASSO under the case of $p > n$.

Additionally, Zou (2006) demonstrates that the forecasting accuracy of the Adaptive LASSO can be much worse than the accuracy of the ordinary LASSO in the case when the weights are estimated by the OLS inefficiently. Especially strongly felt is the effect of the multicollinearity, since in that case the estimates of the OLS are highly unstable.

Concerning the optimal weights, 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. Assume that $w = (w_1, \ldots, w_p)$ is a weight vector and $n \to \infty$, then in general the requirements for choosing the optimal weights are:

- there exists $\xi : \quad 0 < \xi < 1$ and a sufficiently large, positive constant c_w, for which
$$\min_{j=s+1,\ldots,p} n^{-\xi/2} w_j > c_w \sqrt{\frac{s}{\psi}},$$
with probability, converging to 1 as $n \to \infty$; where s is the number of significant, non-zero parameters, and ψ is the according eigenvalue of the covariance matrix. In other words, it is required that weights of the insignificant variables should diverge at a certain rate;
- there exists $w_{max} < n^{\xi/2}$, such that $\sum_{j=1}^{s} w_j^2 < s w_{max}$ with probability, converging to 1. In other words, the weights of the significant variables are restricted from above with a non-decreasing sequence $\{w_{max}\}$.

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.

Indeed, by using the aforementioned results, Liu (2014) proposed using weights in the form of $\hat{w}_j = |\hat{\beta}_{j,OLS}|^{-\gamma_1} \cdot A_j^{-\gamma_2}$, where $A_j = \sum_{k=j}^{h} |\hat{\rho}_{kk}|^{\gamma_0}$, here $\hat{\rho}_{kk} = \mathrm{Cor}(y_i, y_{i+k} \mid y_{i+1}, \dots, y_{i+k-1})$, and $\gamma_0, \gamma_1, \gamma_2 > 0$, accordingly, are hyperparameters, chosen by cross-validation. The main idea proposed here is that under the assumption that part of the macroeconomic variables can be well approximated by AR(h) processes, $h > 0$, the estimated autocorrelation of such variables should also have an effect to their estimated weights, where h is the order of the autocorrelation, estimated by the usual methods employed in the time series analysis (i.e., by minimizing the estimated value of the *Akaike*, *Schwarz* or similar information criterions).

Additionally, Konzen and Ziegelmann (2016) propose modifying the weights according to the number of time periods the particular variable is lagged: the authors claim that any variable used in the modelling should be of a higher significance (hence with a smaller estimated weight) if it is observed without lag or with only a small number of lags, since then the information brought to the regression would be more recent and more useful for forecasting. The authors suggest modifying the weights to $\hat{w}_j = 1/(|\hat{\beta}_j|\alpha(1-\alpha)^l)^\gamma$, where $\alpha \in [0, 1]$, $\gamma > 0$ – hyperparameters, chosen by cross-validation, and l corresponds to the number of lags used of the j-th variable.

Moreover, Medeiros and Mendes (2015) showed 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 a property, often found 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 a possibility of additionally including non-linearities through interactions between the variables of a certain order 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 $W = \mathrm{diag}(\hat{w}_1, \dots, \hat{w}_p)$, then the optimization problem of the Adaptive LASSO (1.2) can be rewritten as:

$$\begin{aligned}\hat{\beta}_{adaLASSO} &= \arg\min_{\beta}(Y - XW^{-1}W\beta)'(Y - XW^{-1}W\beta) + \lambda||W\beta||_1 \\ &= \arg\min_{\tilde{\beta}}(Y - \tilde{X}\tilde{\beta})'(Y - \tilde{X}\tilde{\beta}) + \lambda||\tilde{\beta}||_1, \qquad (1.3)\end{aligned}$$

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

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 ϕ. Let $\mathcal{M}_\lambda := \{1 \le k \le p \ : \ \hat{\beta}_k^\lambda \ne 0\}$ denote the set of variables, preselected by the LASSO method under a certain fixed value of λ.

	Algorithm 1: LARS algorithm for the Adaptive LASSO
1.	Define $\tilde{X} = XW^{-1}$, that is, $\tilde{X}_j = X_j/\hat{w}_j, j = 1, \ldots, p$.
2.	Use LARS algorithm to estimate the $\tilde{\beta}(\lambda) = \beta_{LASSO}(\lambda, Y, \tilde{X})$
3.	Obtain the final coefficient vector $\hat{\beta}_{adaLASSO} = W^{-1}\tilde{\beta}(\lambda)$.

Then the Relaxed LASSO is estimated as:

$$\hat{\beta}_{reLASSO} = \arg\min_{\beta} n^{-1} \sum_{i=1}^{n} \left(Y_i - X_i\{\beta \cdot \mathbb{1}_{\mathcal{M}_\lambda}\}\right)^2 + \phi\lambda||\beta||_1, \tag{1.4}$$

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

$$\{\beta \cdot \mathbb{1}_{\mathcal{M}_\lambda}\}_k = \begin{cases} 0 & k \notin \mathcal{M}_\lambda, \\ \beta_k & k \in \mathcal{M}_\lambda, \end{cases} \tag{1.5}$$

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 λ, while the appropriate estimation of the model's parameters and the amount of shrinkage applied is refined by using a second hyperparameter ϕ. When $\phi = 1$, the estimator coincides with the case of the ordinary LASSO, that is, no correction of the estimated coefficients is performed.

Such proposed methodology is especially useful when in the information set used in the modelling, among all of the available variables, there exist a few, which are extremely important and capable of bringing a lot of valuable explanatory information for the modelled variable. If the hyperparameter λ is chosen in order to minimize the out-of-sample forecast errors of the model, it will usually be chosen such, that the shrinkage applied to those strongly significant variables would be as small as the out-of-sample performance allows. Since a small amount of shrinkage applied requires a relatively small value of penalty, hence a small value of λ, this allows for an inclusion of additional, insignificant, noise variables to the model. However, this problem is solved by the Relaxed LASSO through the use of the hyperparameter ϕ: first, the value of λ is chosen such that there would be no insignificant noise variables included together with the significant ones in the model. However, in this case, due to the large resulting value of λ chosen, the amount of shrinkage applied will possibly be too large. Second, by the use of the hyperparameter $\phi \in (0, 1]$ the amount of shrinkage applied to the preselected variables is corrected. In the limiting case of $\phi \to 0$, the coefficients of the model $\mathcal{M}_\lambda$ are estimated using the OLS.

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.

Another recent modification of the LASSO is the Square-Root LASSO (Belloni et al. (2011)). The authors propose modifying the original formulation of the LASSO problem (1.1) by taking the square-root of the residual sum of squares term, as defined by the equation (1.6):

$$\hat{\beta}_{sqrtLASSO} = \arg\min_{\beta} \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - X_i\beta)^2} + \lambda||\beta||_1. \tag{1.6}$$

The main motivation for such modification is the observation by Bickel et al. (2009), who show that under certain conditions the optimal value of λ for the LASSO problem has the form of $\lambda = \sigma \cdot 2\sqrt{2\log(p)/n}$, where σ is the unknown standard deviation of the error term of the true model. Here the true model is assumed as $Y_i = X_i\beta_0 + \sigma\varepsilon_i$, where $\beta_0 \in \mathbb{R}^p$ is the true parameter value, $\sigma > 0$, regressors X_i are $p-$dimensional, allowing for the case of $p > n$, with i.i.d noise $\varepsilon_i, i = 1, \ldots, n$ from a certain distribution (mainly the case of $\mathcal{N}(0,1)$ is analysed, however, the authors argue that asymptotically their proofs will be valid without imposing normality due to the moderate deviation theory). Additionally, β_0 here is assumed to be sparse, that is, $\text{supp}(\beta_0)$ has $s < n$ elements.

In order to estimate the optimal value of λ without knowing the true value of σ, two ways are suggested in the literature: first, to estimate σ by iterating from a conservative starting value (usually the standard deviation around the sample mean), however, it is found that the accurate estimation of σ when $p > n$ may be as difficult as the original problem of variable selection. Second, by employing the cross-validation, which is often used in practice and produces good results. On the other hand, the authors show that for the Square-Root LASSO the optimal penalty level is independent of σ, that is, it reduces to $\lambda = \sqrt{2\log(p)/n}$, which makes it having no user-specified parameters and therefore tuning free. As in the case for the ordinary LASSO, the minimization problem (1.6) is globally convex, allowing for an easy polynomial-time computation.

Additionally, another modification of the LASSO is the Random LASSO (Wang et al. (2011)). The authors highlight the following serious shortcoming of the LASSO method: if among all of the possible explanatory variables there are some strongly intercorrelated, the LASSO will only choose one or a few as the most significant, while the other will be shrunken down to zero, even though that might not be the most efficient solution, since potentially useful explanatory information may get omitted (though, the authors note that such a fact is mostly important when working with biological data). Additionally, in the case of $p > n$, the final solution of the LASSO will not include more than n estimated non-zero parameters, while in reality we are never sure that the $(n + 1)$'st variable was actually insignificant. That is, asymptotically, when $n \to \infty$, the method will find the true subset of significant variables, while in reality, when working with samples of finite (and often not very large) sizes, in many cases the final resulting forecasts may be relatively worse only because of the small number of selected variables. Even though when working with macroeconomic data such a fact is not necessarily a big problem, the significance of this problem might increase a lot when, for example, working with economic data from countries, where long set of historical data is unavailable.

In order to solve the aforementioned problems the authors propose a Random LASSO algorithm, highlighting the parallel with the method of Random Forests: by using the bootstrap method, for every iteration $b_1 \in \{1, \ldots, B\}$ randomly select $q_1 \leq p$ variables from the full dataset and estimate them by using the (Adaptive) LASSO. The coefficients of insignificant variables (and of all those, that were not selected by the random draw) are set as zero. Then the importance of all of the coefficients are measured as I_j: $\forall j$, $I_j = |B^{-1}\sum_{b_1=1}^{B}\hat{\beta}_j^{(b_1)}|$. The idea here is that the truly insignificant variables, even if at some iterations selected by the LASSO as significant, will have a coefficient equal to or very close to zero, during some iterations the sign of the coefficient might differ, when compared with other iterations. Therefore, the mean value of all the coefficients for such variables should be very close to zero. On the other hand, for truly significant variables the estimated coefficients should on average be relatively large and consistent.

Additionally, a second bootstrap is performed, where with each new iteration $b_2 \in \{1, \ldots, B\}$ a random draw is made to select $q_2 \leq p$ variables, where the probability of drawing the j-th variable is defined by the value of I_j, and the LASSO coefficients are re-estimated a second time. The final result is $\hat{\beta}_j = B^{-1}\sum_{b_2=1}^{B}\hat{\beta}_j^{(b_2)}$.

It can be noted that this method is quite calculation-intensive, since the recommended number of iterations performed is $B \approx 1000$ or larger. Additionally, in order to obtain optimal results one should use cross-validation to estimate the hyperparameters q_1, q_2 and λ. During the simulation studies the authors found that the selection of highly correlated variables is performed much more effectively, also, the coefficients of such variables are estimated more efficiently than when using the Elastic Net method.

In this thesis we decided to use the LASSO, Square-Root LASSO, Relaxed LASSO and Adaptive LASSO methods due to their attractive properties, together with our proposed combination of LASSO and principal components, which is described in more detail in section 3.1. In all of the cases, except for the Square-Root LASSO, which has an optimal theoretical value of λ, the hyperparameters have been selected by cross-validation. For the Adaptive LASSO the weights used were formed using the estimates $\hat{\beta}_{ridge}$ with $\gamma = 1$, chosen as classic in the literature. Even though, as we discussed, the weights formed by using the coefficients from LASSO or Elastic Net are also a decent choice, we did not use them due to two main reasons: first, in the case of $p > n$ the LASSO will choose non-zero coefficients at most for n variables, thus the estimation of the Adaptive LASSO would somewhat reduce to the Relaxed LASSO, which would also select the same variables in the first step of the estimation. Second, if we are not certain that the LASSO is able to (asymptotically) consistently select the true significant variables, restricting the coefficients of possibly good explanatory variables to zero before using them in Adaptive LASSO seems too restrictive. As we know, with proper weights the Adaptive LASSO has the Oracle Property, therefore we don't want the weights to be too restrictive. The Elastic Net seems as a better choice for this, since it allows for the restriction of far less coefficients to zero, however the main shortcoming is that it requires the cross-validation of two separate hyperparameters, which might lead to relatively volatile results during the pseudo-real-time experiments conducted: if, for example, the cross-validated hyperparameters for one round of the experiment are selected such, that the solution coincides with the one of LASSO, choosing at most n weights for the Adaptive LASSO, and on the next round the selection coincides with the one of the Ridge Regression, choosing weights for all p variables, the final results for the Adaptive LASSO may be significantly different both in variable selection and in forecast accuracy. On the other hand, the estimates from Ridge regression seem to be adequate enough: first, only one hyperparameter is required for the estimation, second, the resulting weights should be reasonably similar throughout the whole forecasting exercise. Though for future work it would be interesting to focus on exploring all of the mentioned weights (and to look for other possible choices) for the Adaptive LASSO in greater detail.

Chapter 2

Preliminaries

2.1 Data preparation

In this thesis the four main components of the US GDP by the expenditure approach are modelled: Gross Fixed Capital Formation, Private Final Consumption Expenditure, Imports and Exports of goods and services, all of which are seasonally adjusted and measured in chain linked volumes on the quarterly basis.

The monthly data used as explanatory variables are various 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: 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, first by estimating and removing the deterministic part of the series (where the significance of it was estimated using the t-statistics, obtained from conducting OLS regression, with 5% significance level), 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 formed, 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 (where the significance was tested by using the *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 every case 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[4].

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:

[4]Though it can be noted that no series required more than 2 differences taken.

$$x(\lambda) = \frac{x^{\lambda} - 1}{\lambda},$$

where x is the time series tested, $\lambda \in \mathbb{R}$, and $x(\lambda) = log(x)$, if $\lambda = 0$, since such case is undefined in the previous definition. Here the optimal value of λ is estimated by maximizing the profile log-likelihood function $f(x, \lambda)$:

$$f(x, \lambda) = -\frac{n}{2} \log \left[\sum_{i=1}^{n} \frac{(x_i(\lambda) - \bar{x}(\lambda))^2}{n} \right] + (\lambda - 1) \sum_{i=1}^{n} \log(x_i),$$

where $\bar{x}(\lambda)$ is the mean value of the transformed series. (Box and Cox (1964))

When studying whether a logarithmic transformation is necessary for a particular series, it is useful optimizing the likelihood function when possible values of λ are under the domain of $\lambda \in [0, 1]$, thus narrowing the possible solution set. If the optimal estimated value of $\hat{\lambda}$ is reasonably close[5] to 1 – the transformation is not necessary, however, if it is close to 0 – the series are transformed by taking logarithms (transformations of x^q, $q \in (0, 1)$ were not used in this thesis since the main goal is not to normalize the data, but to extract and distinguish the multiplicative effects if such were present, instead of just shrinking them a bit, which would be performed in the case of transforming the data by the power of $q < 1$; additionally, log-transformation is useful since it does not heavily alter the interpretation of the data, because in certain cases the differences of logarithmized data are very close to the percentage growth of the original data).

Also, some of the available data has relatively large spikes 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[6]. 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.

Additionally, 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 4, 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. During a full quarter at least three updates on the dataset are possible for every different month of the quarter, however, in this thesis the results presented are of the last month of the full quarter, since in that case for some indicators with small enough publication lags there were some of the monthly information available for the coinciding quarter, resulting in the estimation of more accurate nowcasts. It is of interest to inspect the nowcasting performance of the models with some coinciding monthly information available, since 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. Also, since for every monthly indicator the day of a new release within a particular month differs, the release lags were calculated as of at the end of the month.

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]The transformation in this thesis was applied if the estimated value of λ was smaller than 0.8.

[6]It's even worse if the sudden shock is relatively recent, since it may strongly affect the individual forecasts of such a series.

2.2 Comparison of the forecast accuracy

In order to compare the forecasting performance of the models, we calculate the Root Mean Squared Error (RMSE) of their forecasts. Additionally, in order to study whether the additional uncertainty, brought with an increased amount of information included in the projection, has a positive effect, we compare the performance of LASSO models with the best ARMA model, chosen by minimizing the *Akaike* information criterion. The latter is able to minimize the mentioned uncertainty, since it employs the information within the modelled variable only. Therefore, for every model **m** we calculate the Relative RMSE in relation to the RMSE, obtained by the ARMA model (for convenience denoted as **AR**):

$$\text{Relative RMSE}(\mathbf{m}) = \frac{\sqrt{\sum_{t=T_1}^{T_2}\left(y_{j,t} - \hat{y}_{j,t}^{\mathbf{m}}\right)^2}}{\sqrt{\sum_{t=T_1}^{T_2}\left(y_{j,t} - \hat{y}_{j,t}^{\mathbf{AR}}\right)^2}} = \frac{\sqrt{\sum_{t=T_1}^{T_2}\left(\hat{\varepsilon}_{j,t}^{\mathbf{m}}\right)^2}}{\sqrt{\sum_{t=T_1}^{T_2}\left(\hat{\varepsilon}_{j,t}^{\mathbf{AR}}\right)^2}}, \tag{2.1}$$

where $\hat{y}_{j,t}^{\mathbf{m}}$ is the forecast by model **m** for j-th modelled variable at a time moment t, where the forecast error, accordingly, is $\hat{\varepsilon}_{j,t}^{\mathbf{m}} = y_{j,t} - \hat{y}_{j,t}^{\mathbf{m}}$. Alternatively, $\mathbf{m} = \mathbf{AR}$, when the model of interest is ARMA. T_1 denotes the first time moment of the forecasts during the pseudo-real-time forecasting exercise, while T_2 denotes the last one. In every case, when the forecasts of a certain model **m** are more accurate than the ones by ARMA model, the resulting Relative RMSE(**m**) is smaller than one.

Additionally, in order to estimate the significance of the possible differences in forecast accuracy a formal procedure is employed in this thesis. One commonly used method for such purposes is the *Diebold-Mariano* (DM) test, estimating whether the forecast accuracy of two models is statistically significantly equal. The null hypothesis of the test is:

$$H_0: \quad \mathbb{E}\left[L(y_{j,t} - \hat{y}_{j,t}^{\mathbf{m}_1}) - L(y_{j,t} - \hat{y}_{j,t}^{\mathbf{m}_2})\right] = 0, \tag{2.2}$$

where L is a chosen loss function (i.e., squared loss function: $L(x) := x^2, \forall x$) and $\mathbf{m}_1$ and $\mathbf{m}_2$ are some methods in comparison. In other words, it is tested whether during a certain time period, the forecasts, generated by two methods in comparison, are on average statistically significantly equal. If the null hypothesis is rejected, the test suggests that the model with smaller forecast error is significantly more accurate during that time period. It is convenient, that for this test a large class of various loss functions $L(\cdot)$ is suitable, also that the autocorrelations of the errors are tolerable (Diebold (2015)). In this thesis we used this test under a squared-loss function, thus giving larger penalties for bigger deviations. It is claimed in the cited literature, that instead of assumptions, made for the models of interest, this test forms assumptions on only the forecast errors, generated by the models. This means that we are allowed to compare possibly overfitted models without a great loss in the power of the test.

Assume that the difference between the forecast errors of some two models at a time period t under a certain loss function $L(\cdot)$ is denoted by $d_{1,2,t} = L(\hat{\varepsilon}_t^{\mathbf{m}_1}) - L(\hat{\varepsilon}_t^{\mathbf{m}_2})$. Then the assumptions[7] of the DM test are:

$$\begin{cases} \mathbb{E}(d_{1,2,t}) = \mu, \ \forall t, \\ \text{Cov}(d_{1,2,t}, d_{1,2,t-\tau}) = \gamma(\tau), \ \forall t, \\ 0 < \text{Var}(d_{1,2,t}) = \sigma^2 < \infty, \ \forall t, \end{cases}$$

with the test statistic DM_{12} defined as:

[7]These assumptions are sufficient, but may not be strictly necessary, as Diebold (2015) claims that less-restrictive types of mixing conditions may presumably be invoked.

$$DM_{12} = \frac{(T_1 - T_2)^{-1} \sum_{t=T_1}^{T_2} d_{1,2,t}}{\hat{\sigma}_d} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1),$$

where $\hat{\sigma}_d$ is the consistent estimate of the standard error (i.e. a HAC estimate) for the sum $\sum_{t=T_1}^{T_2} d_{1,2,t}$. (Diebold (2015))

Additionally, often used in the literature is the *Giacomini-White* (GW) (Giacomini and White (2006)) test. The authors of the test claim that when testing the unconditional null hypothesis of equal forecasting accuracy (2.2), basically the comparison of the specification of the models is made. That is, often a more correctly specified model will be suggested as the better one. They claim that by testing such hypothesis it is evaluated, whether the models of interest are better in describing the data generating process, but not necessarily whether they will continue to forecast with similar accuracy in the future. For this reason they introduce the conditional null hypothesis for equal forecasting accuracy, where the upcoming forecast is dependent on the currently available set of information $\mathcal{F}_t$. Therefore, the conditional null hypothesis is defined as:

$$H_0: \quad \mathbb{E}\big[L(y_{j,t+\tau} - \hat{y}^{\mathbf{m}_1}_{j,t+\tau}) - L(y_{j,t+\tau} - \hat{y}^{\mathbf{m}_2}_{j,t+\tau}) \mid \mathcal{F}_t\big] = 0 \quad \text{a. s. } \forall t, \tag{2.3}$$

where $\tau > 0$ is the forecast horizon.

The main difference between the latter and the unconditional null hypothesis is that the properties of the generated forecasts are evaluated according to the information, available at the time of the forecast, but not to the asymptotic properties of the models. The authors claim that the null hypothesis can be interpreted as a test, whether the two models in comparison will continue to forecast with the same forecasting accuracy, according to all of the currently available information. The main shortcoming of this method is that the test statistics of the proposed test converges to the defined limiting distributions only when a rolling window forecasting frame is used. For this reason, in this thesis we use this test only when evaluating the forecasts for the rolling window pseudo-real-time forecasting exercise.

In the general case the test statistic of the GW test is defined by the equation (2.4).

$$T_{n,m,\tau}(h) = n \left(n^{-1} \sum_{t=m}^{T-\tau} h_t d_{1,2,t+\tau} \right)' \hat{\Omega}_n^{-1} \left(n^{-1} \sum_{t=m}^{T-\tau} h_t d_{1,2,t+\tau} \right) \overset{\mathcal{D}}{\sim} \chi^2_q, \tag{2.4}$$

where h_t – $\mathcal{F}_t$-measurable test function, $\hat{\Omega}_n$ – the estimate of the covariance matrix for the forecast errors (i.e. using a *Newey-West* estimator), m – the estimation window length, n – the number of forecasts compared. A more detailed definition of the test is provided in Giacomini and White (2006).

Chapter 3

Estimation of the model

3.1 Principal Components and LASSO

In this thesis 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 using LASSO for subset selection (in this thesis both the LASSO and the Adaptive LASSO are used, since the latter is known to have the best asymptotic properties for a correct subset selection under $p > n$ with many highly correlated variables). From here on, let's assume that $X \in \mathbb{R}^{n \times q}$ is a preselected matrix of significant variables, where $0 < q \leq n$.

Second, since we are interested in modelling macroeconomic data, it is likely, that it will exhibit a significant correlation, some of the variables might be nested (i.e. should total unemployment in the country and unemployment of a particular age segment be included). Therefore, instead of a direct re-estimation, as would be done using Relaxed LASSO, we suggest rotating the data using the principal components methodology, thus extracting the main latent factors $F = XL$, where $L \in \mathbb{R}^{q \times q}$ is a rotation matrix and $F \in \mathbb{R}^{n \times q}$ 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. Since the data is likely to be correlated, and because of the (supervised) preselection done – 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, which is basically the idea of usual factor analysis in the literature.

However, from the literature reviewed it follows that LASSO can effectively and consistently estimate orthogonal variables. More so, the extension of the Adaptive LASSO or Relaxed LASSO is unnecessary for orthogonal data, since it does not violate any of the necessary conditions for the LASSO. Therefore, we expand on the idea of (1.3) and estimate:

$$\begin{aligned}\hat{\beta}_{fLASSO} &= \arg\min_{\beta} \ (Y - XLL'\beta)'(Y - XLL'\beta) + \lambda||L'\beta||_1 \\ &= \arg\min_{\tilde{\beta}} \ (Y - F\tilde{\beta})'(Y - F\tilde{\beta}) + \lambda||\tilde{\beta}||_1, \end{aligned} \tag{3.1}$$

since $LL' = I$ holds by the definition of principal components[8], and $F = XL$, $\tilde{\beta} = L'\beta$; 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. Also, since the transformation introduced by principal components does not remove any of the information from the data, it is possible, that after estimating coefficients $\tilde{\beta}$ using LARS, the rotation back to $\beta = L\tilde{\beta}$ can provide more accurate estimates than the straightforward LASSO without using the principal components.

Such 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 obvious strength of such approach is when dealt with a large amount of data (i.e. from a particular market), driven by one or a few leading factors, accompanied with a significant amount of noise, which in the latent space can be interpreted as various trends, behavioural patterns and signals from neighbouring markets, all of which can be insignificant to the main modelled variable. In such case the principal component transformation would allow us to extract some of those latent factors and estimate only the significant ones using the LASSO. The final coefficient vector $\hat{\beta}^* = L\hat{\tilde{\beta}}$ would be comprised of the same non-zero variables just as it would be in the ordinary (Relaxed or Adaptive) LASSO case, however, the estimated coefficient values would be set according to the significance in the latent space, rather than the direct one. So in a way, such transformation may act as a filter, distinguishing only the important underlying signals from the data, thus possibly allowing for a more accurate forecasting performance.

Secondly, in contrast to Bai and Ng (2008) and other factor forecasting related literature, we propose to base the final forecasts on the predictions of individual variables X rather than on the predicted significant factors F. That is, if we assume that $X = X_t = (X_{1t}, \ldots, X_{qt})$, for every $h > 0$, the forecasts $\hat{F}^*_{t+h}$ can be calculated as

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

where L is known and $\hat{X}_{j,t+h}$, for every $j = 1, \ldots, q$, are predicted using time series methodology (i.e. ARIMA models with appropriately selected parameters).

Typically in the literature the idea behind extracting factors is to reduce the complexity of the high-dimensionality problem by moving to the latent space, where only the significant signals, approximating the true factors, are of interest. Among many different approaches, one popular way is to forecast the approximated factor directly using *Kalman-Filter* or other time series methodologies (i.e. ARIMA also). That is, usually

$$\hat{F}_{t+h} = \hat{\alpha}(h)F_t \tag{3.2}$$

are constructed.

However, if we take into account the formulation of our problem in (3.1), the complexity of such indirect factor forecasting does not differ from the complexity of forecasting using just (Relaxed) LASSO, since the same number of variables are to be predicted individually[9]. In such case, it is possible that more information is used in generating the forecasts $\hat{X}_{t+h}$ directly than it is in generating $\hat{F}_{t+h}$ using the (3.2) approach.

[8] Here the data matrix X is scaled and centered.

[9] NB: in this and the following discussion ARIMA methods are regarded as the method for variable predictions, however it is worth noting that the forecast accuracy of individual variables $X_{jt}, j = 1, \ldots, q$, might be greatly improved by forming structural (i.e. VAR or similar) models for some of the variables (Jokubaitis (2015)).

Granger (1980) has shown that the aggregation of a low-order AR/ARMA processes may produce a process with complicated dynamics. I.e., they show that aggregation of $k >> 1$ independent AR(1) processes can generate an ARMA$(k, k-1)$ process, unless cancellation of roots occurs. Therefore, under the assumption that each of the modelled variable X_{jt} can be well approximated by a low-order ARMA process, it is clear that a factor, combined from such variables, might have a complex, even long memory structure[10]. Additionally, it is likely that because the aggregation of the predictor variables are made using factor loadings as weights, the parameters of aggregated long memory process will converge to zero at a fast rate[11], therefore a direct estimation of such parameters might be very hard or infeasible, considering the short historical time periods of many macroeconomic variables. However, each of the variables X_{jt} can be easily approximated by an appropriate (likely, also of a low order) ARMA process, therefore the complex structure of the forecasts from the true aggregated process can be retained indirectly, by aggregating the forecasts of each of the X_{jt} using the known factor loadings as weights.

As an illustration, let's assume that we can extract a strong factor from the data, which can explain our modelled macroeconomic variable with high accuracy (for example, $R^2 \approx 0.9$), and assume that such factor has an ARMA$(k, k-1)$ structure with $k >> 1$ as was discussed. Should the dataset be insufficient to correctly estimate most of the true model parameters (i.e., due to a small number of available data), it is likely that under various information criteria the best ARMA approximation would be of a low order. Additionally, because of the high in-sample accuracy the direct forecasts of such a factor would be very similar to a direct forecast of the modelled macroeconomic variable using only (benchmark) ARMA models. The point is, in such a case the forecasts would not differ much from the ones, where no additional (exogenous) information is used, except for the information of the main modelled variable, and that difference would shrink with increasing R^2. However, if we forecast such a factor indirectly, through forecasting each of the X_t, we would use all of the underlying information from the data, which could help attaining more accurate forecasts.

In the preceding discussion we assumed a simple case of AR(1) process for the explanatory variables. However, it is worth noting, that since in this thesis we are modelling macroeconomic variables, each of the predictor X_{jt} used can also be a process of a complex ARMA structure, since it can easily be a microeconomic variable, generated from several micro-variables (as discussed by Granger (1980), i.e., consumption, income, employment, production in various sectors and markets). Therefore, aggregation of such variables can lead to an even more complexly structured process than has been discussed.

Also, the aforementioned theory holds under the assumption of independence between the processes, however, Granger (1980) claims, that similar conclusions should also hold for correlated processes. In order to study the behaviour of the discussed processes under non-zero correlation in more detail, a Monte Carlo experiment has been conducted, the results of which are presented in section 3.3.

In addition to this, if the extracted factors are formed with weights of a similar size (that is, not only a few particular variables are dominant), such forecast aggregation in a way forms a parallel to other similar forecast aggregation methods in the literature, such as *bagging* (bootstrap aggregating, (Breiman (1996))), which are shown to improve the prediction accuracy in certain cases.

Also, the proposed aggregation might induce a smaller loss in cases where we fail to accurately define a suitable model for the process of interest: first, if we are unsuccessful in consistently selecting decent models for some of the predictors X_t, the aggregation

[10]Here long memory is understood in the sense of finitely nonsummable auto-covariance function. It is shown by Granger (1980) that such results may occur when the AR(1) models are generated with random coefficients and the aggregation scale is very large ($k \to \infty$).

[11]For example, if we assume that some of the variables have relatively small loading weights for a certain factor, their effect on the parameters of the aggregated ARMA process will be smaller than when applying aggregation with equal weights, i.e., by averaging.

of their resulting forecasts can stabilize the results to some extent (similarly to *bagging*); second, if we are able to identify appropriate models for the predictor variables, forecasting the main factor through aggregation of its variables might lead to an increase in forecast accuracy, as is demonstrated in section 3.3; third, it is possible, that the data generating process of X_t might be from some family of complex, long memory processes, therefore the aggregation of their forecasts introduces some degree of freedom to make inaccurate estimations of their true models while still generating more accurate final predictions.

Additionally, it is worth noting that in the case when the data set X is orthogonal, since the method of principal components is able to retain all of the information from the data, it would extract the same orthogonal data, just (possibly) resorted by the total variance explained (eigenvalues of the covariance matrix), therefore, the proposed method would boil down to the ordinary LASSO method.

3.2 Sparseness of the PCA loading matrix

Even though the data matrix X is preselected by the LASSO as a matrix, containing mainly significant variables, it is not clear that, first, 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[12], 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.

Second, some of the variables (for convenience in the following text denoted as $Z_t,\ t = 1, \ldots, n$, where $Z \subset X \in \mathbb{R}^{n\times q}$) 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. In the case when Z_t are very significant explanatory variables, the LASSO will try to extract as much information as possible from those variables. Since every principal component $f_{j,t}$ has the following structure:

$$f_{j,t} = \sum_{\ell=1}^{q} \lambda_{j,\ell} X_{\ell,t} = \Lambda_j Z_t + \Phi_j \bar{X}_t, \quad j = 1, \ldots, q,$$

where $\bar{X}_t$ are such that $\bar{X}_t \cup Z_t = X_t$, any such component is composed of Z_t together with the remaining variables. Therefore LASSO, while trying to reconstruct the most useful part of the Z_t, will include too many factors $f_{j,t}$ to the final solution. Some of those factors would not be included if the weights of Z_t would have been zero. Let's assume that $\mathcal{G} \subset \{1, \ldots, q\}$ is a set of indices denoting factors $f_{j,t}$, which have been selected as significant by the LASSO in the final solution only because of a significant loading weight for Z_t. Then, it is clear, that with every additional $f_{r,t},\ r \in \mathcal{G}$, included we will add some noise in the scale of $\Phi_r \bar{X}_t$ 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)). Another way, to prevent situations as described in the preceding paragraph, could be to include the preselected data matrix X_t together with the extracted factors F_t and model them together as $\bar{F}_t = (F_t, X_t)$ using the LASSO. While it may seem that in this way no new information is

[12]NB: even though the LASSO in theory should pick mainly uncorrelated variables due to the definition of the LARS algorithm, used in LASSO, in practice the author of this thesis found many cases, where the LASSO chooses several strongly multicollinear variables as the final ones, even when the shrinkage parameter λ is chosen by cross-validation as the most optimal one.

added, it can help LASSO distinguishing only the truly significant factors, while ignoring the factors $f_{r,t}, r \in \mathcal{G}$, since instead it would select the variables Z_t directly. Thus, the possible noise would be lowered in the scale of $\sum_{j\in\mathcal{G}} \Phi_j \bar{X}_t$. Also, such approach would add some robustness, preventing unnecessary transformations in cases, where the predictors X_t are orthogonal and the extraction of factors does not add any interesting projections, since they would not be more useful in explaining the modelled variable than the original data, hence would probably not be selected by the LASSO. However, it is also likely, that $\bar{F}_t$ would not be an orthogonal matrix anymore.

3.3 Monte Carlo experiment

In order to study the behaviour of aggregated AR processes' under more realistic conditions, we assume that $X \in \mathbb{R}^{n\times q}$, where $n = 100$ and $q = 20$ are generated by an AR(1) process with random parameters α from a known distribution. As we will work with stationary time series, the distribution for random coefficients have been restricted to $U[-0.85, 0.85]$. The following AR(1) processes were generated:

$$X_{j,t} = \alpha_j X_{j,t-1} + \varepsilon_{j,t}, \quad j = 1, \dots, q, \tag{3.3}$$

where $\varepsilon_t = (\varepsilon_{1,t}, \dots, \varepsilon_{q,t}) \sim \mathcal{N}(0, \Sigma)$ and Σ is not an identity matrix. That is, the Σ is chosen such that the average correlation between the variables would be around 0.5 with the main diagonal being normalized to one.

Additionally, we construct a factor f_t as a first principal component of the data X_t. Because the covariance matrix Σ is not an identity matrix, the first extracted factor has significantly non-zero loadings for most of the variables, similarly to what we would have in real situations with macroeconomic variables.

The goal of this experiment is to study the differences in forecasting accuracy between direct forecasting of the known factor f_t using usual time series methods and aggregating individually forecasted variables X_t, according to the factor loading matrix L.

The experiment is repeated 500 times, where in each iteration the variables are generated randomly and a first factor f_t is then extracted using principal component methodology ($F_t = X_t L,\ f_t \subset F_t$). In each iteration the factor is forecasted in two ways: first, by choosing an appropriate ARMA(p, q) model, where the orders p and q are chosen by minimizing the *Akaike* information criterion (which is known to be not too restrictive in choosing the order of the model); and second, by fitting an AR(1) model for each of the variable $X_{j,t},\ j = 1, \dots, q$, and aggregating its forecasts using the weights of the loading matrix L. The forecast horizon $h = 1, \dots, H$ is chosen to be $H = 8$ in order to study the longer dynamics of the forecasts.

In order to measure the accuracies of the forecasts several measures were used: the mean absolute error (MAE), root mean squared error (RMSE) and R-squared (R^2) :

- $MAE = \frac{1}{H}\sum_{h=1}^{H} |f_{t+h} - \hat{f}_{t+h}|$
- $RMSE = \sqrt{\frac{1}{H}\sum_{h=1}^{H}(f_{t+h} - \hat{f}_{t+h})^2}$
- $R^2 = 1 - \frac{\sum_{h=1}^{H}(f_{t+h}-\bar{f}_t)^2}{\sum_{h=1}^{H}(f_{t+h}-\hat{f}_{t+h})^2}$,

where $\bar{f}_t = \frac{1}{t}\sum_{j=1}^{t} f_j$ is the sample mean, and it can be conveniently set to 0 here as the variables are designed to be mean-centered. The idea behind using forecasted R^2 is to check not only the average accuracies of the forecasts, but also to inspect the amount of variance explained by the forecasts. Since we expect a randomly generated factor to have a complex structure, ideally the generated forecasts should also retain it. In this sense, a naive forecast (i.e. a sample mean) would be the least favourable forecast, since it does

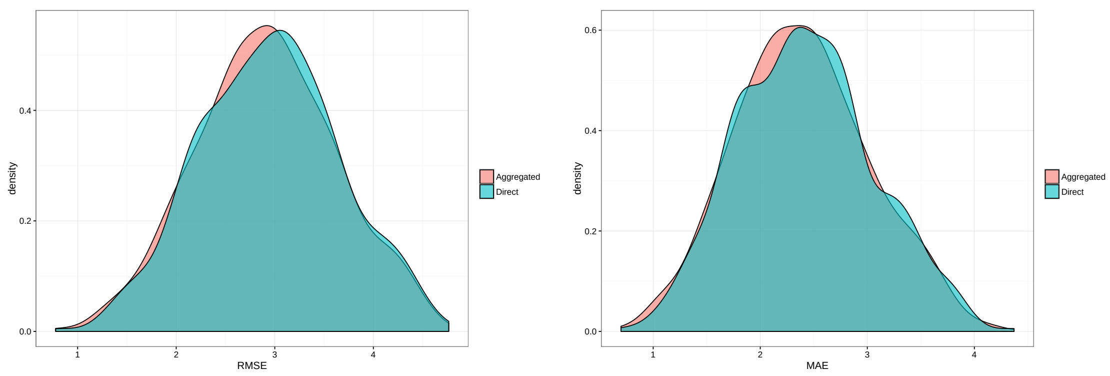

Figure 3.1: Density plot of root mean squared errors (RMSE) and mean absolute errors (MAE) of the forecasts, generated by both aggregated and direct predictions, $q = 20$.

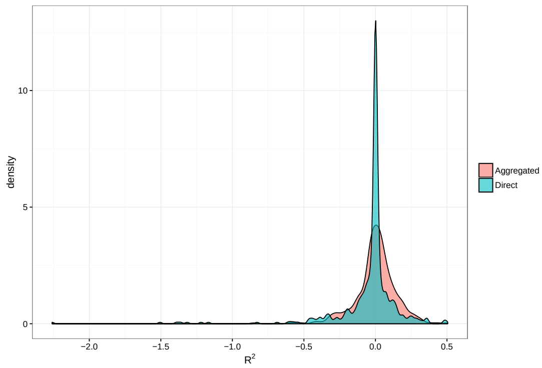

Figure 3.2: Density plot of R-squared (R^2) of the forecasts, generated by both aggregated and direct predictions, $q = 20$.

not predict any part of the underlying structure. Therefore, the point of R^2 here is to compare the differences between generated and naive forecasts.

The results are presented in figures 3.1 and 3.2 for $q = 20$ and in figures A.1 and A.2 for $q = 80$. It can be seen that while the difference is not big, overall the aggregated forecasts tend to generate more accurate forecasts: the sample distribution of the RMSE of the aggregated forecasts are shifted to left, when compared with the direct forecasts, similar results are seen when inspecting the distribution of the MAE. However, it is interesting to inspect the distribution of the R^2 of the forecasts: we can see from the figure 3.2 that direct forecasts are strongly concentrated around 0, which suggest that most of the times the directly modelled ARMA was not able to produce significantly better forecasts than a naive forecast. One of the reason for such a result is that in a relatively small sample the model was not able to recognize and estimate the true structure of the generated factor, therefore a low-order (including a zero-order) ARMA model was selected by the *Akaike* information criterion. On the other hand, the distribution of the aggregated forecasts, while also concentrated around 0, has a significantly smaller kurtosis, which suggest that it was able to explain a significantly larger amount of the true structure of the generated factor.

Also, in both of the cases many negative R^2 values can be seen, suggesting that in those cases a naive forecast would have been more accurate in explaining the factor. Similar effects, just to a greater scale, can be seen in pictures A.1 and A.2, where the number of indicators involved was increased from $q = 20$ to $q = 80$, resulting in a more complex resulting original factor.

An illustrative example of a few iterations from the experiment is presented in 3.3

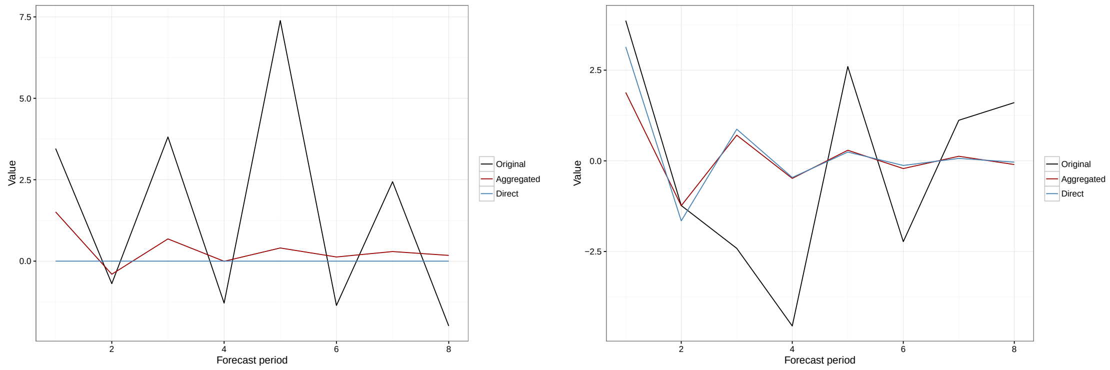

Figure 3.3: Example cases of aggregated and direct forecast realizations, when compared to the original data: (left) case, when aggregated forecasts succeed in capturing the true dynamics while the direct forecast fails; (right) case, when both of the forecasts coincide, with the direct forecast generating an overall more accurate predictions.

picture. In both of the cases the true original factor is plotted together with aggregated and direct forecasts. In the left picture we can see the example of aggregated forecasts being able to recover the structure of the factor, while the direct forecasts immediately flats out, failing to even select an adequate model – likely, that the structure of the factor is recognized as a white noise, rather than the factor itself. However, in the right picture we can see an example of a direct forecast being able to perfectly match its forecast with the one, generated by the aggregation of models, and even generating an overall more accurate fit.

Overall, the results of this experiment suggest that while the increase in forecast accuracy is not huge, the aggregation of the forecasts helps recovering some of the structure of the true factor, which is especially useful in moderately sized samples, where the usual direct time series modelling tend to choose a model of a low order.

Chapter 4

Pseudo-real-time forecasting experiments

In this chapter the results of pseudo-real-time forecasting exercise over 2005Q1 – 2014Q4 are presented. 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, 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.

For each of the modelled GDP component, the main models considered are: the Square-Root LASSO (in the tables denoted as: *Sqrt*), Relaxed LASSO (*Relaxed*), Adaptive LASSO (*Adaptive*), ordinary LASSO (*LASSO*) and a proposed combination of LASSO with the principal components of the data, preselected by the Adaptive LASSO (*AdaPCA*) or by the ordinary LASSO (*PCA*), as described in section 3.1. Additionally, it is interesting to inspect the gains brought solely by performing the rotation of the data to its principal components, therefore for some particular cases we analyse the alternative cases for the AdaPCA models, where the preselected variables are the same, but the rotation to the principal components is not performed (so in a way it's a mix between Relaxed LASSO and Adaptive LASSO, therefore in the tables denoted as *AdaRL*). Also, as discussed in the section 3.2, 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 the SPCA method instead of the ordinary 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 (so 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. 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 cross-validated hyperparameter can generate a model of a too dense structure, where an inclusion of a large number of variables can win only a small amount of accuracy, but bring an additional uncertainty with each of the additional variable included[13]. However, sparser versions of some models can produce a similar level of accuracy, but with a much smaller space of variables used.

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 the models themselves to select the significant parts, since it is likely, that some indicators, driving the growth of some of the markets in the past are not that significant at a later time, hence they can be replaced in a timely matter with new indicators, especially when 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 DM tests on the forecast errors for the expanding window exercise, and pairwise GW tests for the rolling window exercise.

4.1 Expanding window: Gross Fixed Capital Formation

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 4.1 for all of the models and for all 4 forecast horizons. Here the models were selected using the best parameters, chosen by the cross-validation (except for *Sqrt*, which uses the theoretically optimal value). 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

[13]Though it can be noted that since each variable included in the final model is individually forecasted using ARMA models, 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 relatively larger shrinkage would reduce the resulting negative effects. Also, the results from the Monte Carlo experiment holds here, with the cases of $q = 20$ and $q = 80$ proving the point.

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 4.2 are presented the results of Relative RMSE, when in comparison with the performance of the benchmark ARMA models.

Table 4.1: 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	1.089	**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	2.775	3.032	**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	1.75	**1.680**	2.553

Table 4.2: 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 4.3: 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	**0.033**
PCA	-	-	-	-	-	0.45	**0.022**
AdaptivePCA	-	-	-	-	-	-	**0.027**
ARMA	-	-	-	-	-	-	-
1Q							
Sqrt	-	0.2	0.17	0.37	0.52	0.18	**0.056**

Table 4.3: *(continued)*

Nowcast	Sqrt	LASSO	Adaptive	Relaxed	PCA	AdaptivePCA	ARMA
LASSO	-	-	0.19	0.29	0.34	0.19	**0.092**
Adaptive	-	-	-	0.9	0.25	0.65	**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	0.3	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 the table 4.3 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 4.4: 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.

	AdaPCA15	**AdaRL15**	**AdaPCA20**	**AdaRL20**	**AdaPCA30**	**AdaRL30**
Back						
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	**0.669**	0.909	**0.609**	0.934
05-15	**0.730**	0.902	**0.637**	0.846	**0.542**	0.794

Table 4.4: *(continued)*

	AdaPCA15	AdaRL15	AdaPCA20	AdaRL20	AdaPCA30	AdaRL30
Now						
05-08	**0.821**	0.832	0.799	**0.776**	**0.740**	0.777
08-11	1.608	**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	**2.531**	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

Additionally, 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 4.4 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 4.1 and 4.2 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 4.4 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[14]. 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 the table E.1 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. The latter results are also consistent over different number of variables used.

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 A.7. 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

[14] 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.

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 λ used. These results provide further evidence that the use of principal components transformation in some cases can provide additional gains in forecast accuracy.

For the macroeconomist it can be of great value to inspect the leading indicators for the GFCF, therefore in the figure 4.1 we present the top indicators[15], 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 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[16], 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.

Table 4.5: 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**

[15]Note that there were some indicators, preselected as significant for a smaller amount of times, therefore for convenience they are omitted from the graph.

[16]Data published at `https://www.oecdregionalwellbeing.org/`

Table 4.5: *(continued)*

	DirectPCA	**AggregatedPCA**
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**

Furthermore, in order to evaluate the ideas from section 3.3 under real data, the following additional experiment was made. Just like in the aforementioned discussions, 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)[17] an OLS regression was made, treating the extracted factors as observable data. However, as discussed in section 3.3, 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 the table 4.5, 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[18]. Second, the forecasting performance for most of the periods is significantly[19] 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 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 3.1, 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

[17]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 omittance 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.

[18]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*.

[19]The *Diebold-Mariano* test for both 1-quarter and 2-quarter forecasts with 5% significance rejects the null hypothesis of equal predictive performance between the models.

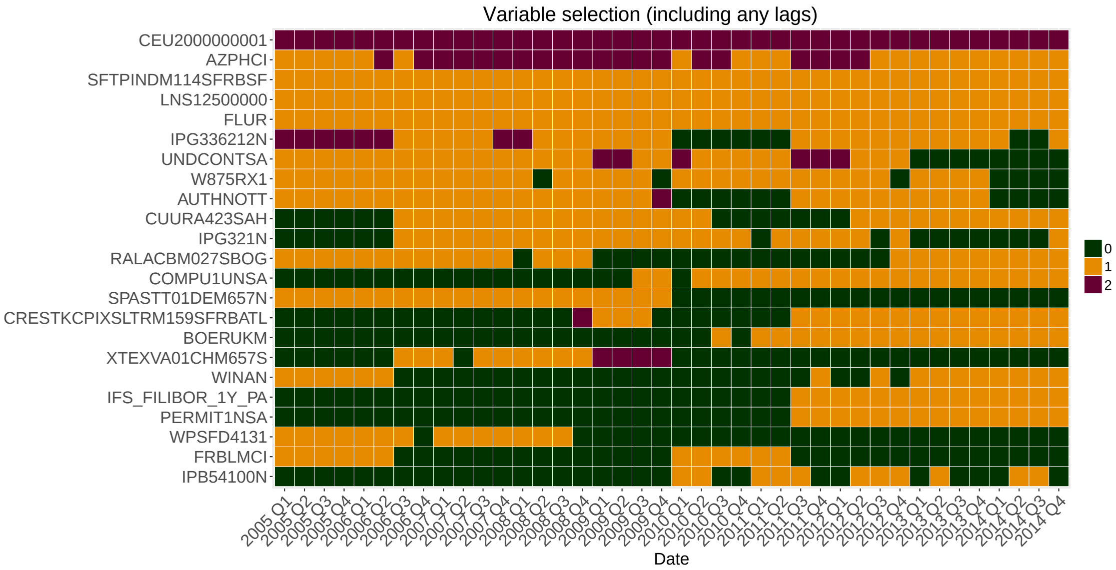

Figure 4.1: Most often selected variables during the expanding window pseudo-real-time experiments for Gross Fixed Capital Formation over 2005Q1-2014Q4. Number of times selected denotes only the number of the same variables selected (i.e. the variable and a one-quarter lag) but not the number of lag that was most oftenly selected.

when comparing the results from the table 4.4 with table 4.5, 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.

4.2 Rolling window: Gross Fixed Capital Formation

In this section we inspect the forecasting performance of the main models under a 12-year rolling window instead of the expanding window as in the section 4.1. 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 some variables, 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.

The main results are presented in the tables 4.6 and 4.7. 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.

In the figure 4.2 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 4.6: 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	**1.516**	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 4.7: 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.

	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**	0.78	0.77	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

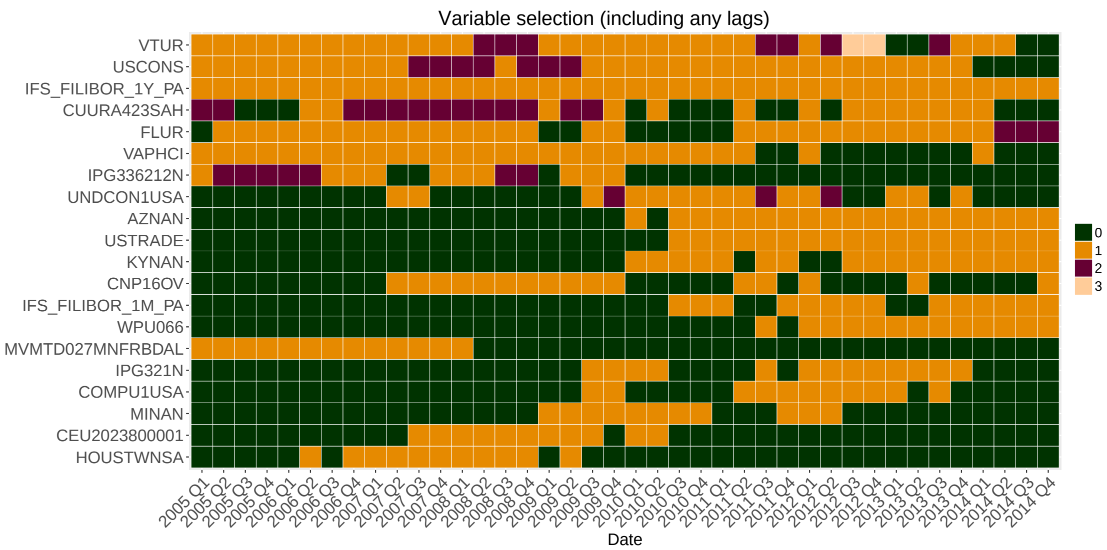

Figure 4.2: Most often selected variables by the AdaptiveLASSO during the rolling window pseudo-real-time forecasting exercise for the Gross Fixed Capital Formation.

4.3 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 fact 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 the tables E.2 and E.3 over a rolling 12-year window. First, it can be noted that in most cases the LASSO methods are able to forecast the consumption with a greater accuracy than the benchmark ARMA models, except for 1- and 2-quarter forecasts during the stable period of 2011-2015. The most accurate nowcasts overall are produced by the Relaxed LASSO models, though it can be seen that it is due to the most accurate performance during the period of 2008-2011. During the other remaining periods the PCA method is able to generate more accurate forecasts. 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 the table E.4, 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 4.3 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.

In addition to that, during the period of financial crisis often the indicators on unemployment rates in various regions were selected, indicating the obvious negative effects caused by growing unemployment in the country to the private consumption. Additionally, it is noteworthy that from the start of 2007, which is quite some time before the peak of the recession, several indicators for certain luxury goods have been consistently included in the models: namely, the employment in the retail trade sector of the automobile dealers and the consumer price index for sugar and sweets, included in the models from 2007 to 2012, covering the whole recession period. Unemployment is often used in representing the uncertainty on the labour market, regarding future income prospects. In the case of growing rates of unemployment, it serves as the indicator for the hikes in precautionary savings (for the rainy day) by households. On top of that, it indicates the diminished bargaining power of the trade unions, and vice versa in the case of a drop in unemployment. It is natural, that due to increasing savings some cutting-down on the unnecessary spending should occur, therefore, the indicators from the markets of various luxury goods should be the first ones to react to such a change in consumer behaviour. Identifying these indicators early before the crisis may be crucial in order to rapidly adapt to the change in behaviour and predict the oncoming fall in private consumption expenditure, which the LASSO models are able to accomplish.

Also, we can note the inclusion of industrial production index for durable goods. Similarly to indicators for luxury goods, spending on durable goods can also be readily postponed in times of economic weakness, therefore such indicators tend to be quite cyclical, capturing an important feature of the business cycle. During a drop in demand for such goods, when households or businesses delay purchases of durable goods, it is likely that the supply side will react accordingly. Therefore, this reaction may be reflected by such industrial production indicators.

4.4 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

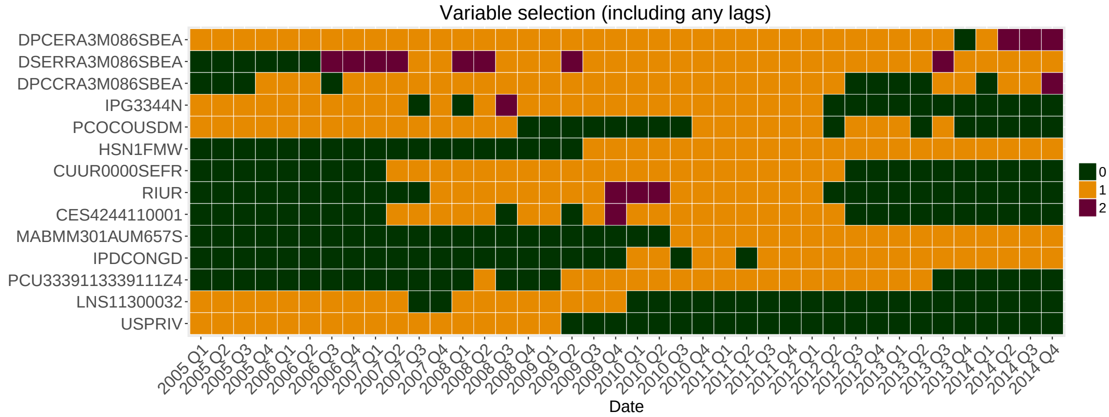

Figure 4.3: Most often selected variables by the AdaptiveLASSO during the rolling window pseudo-real-time forecasting exercise for the Private Final Consumption Expenditure.

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[20], 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 are presented in the tables E.5 and E.6 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 the table E.7, suggest that with 11% significance the Square-Root LASSO, LASSO and PCA are able to generate more accurate forecasts than the benchmark ARMA model.

By inspecting the monthly variables, selected when modelling Exports, where the most often selected are presented in the figure 4.4, as expected, we can see that the available main monthly indicators of exports of goods and services are included, both the total value, and often additionally the values of exports to the largest trading partners of the US: Canada, China, Korea, Mexico, Japan. Moreover, for some of these countries at times

[20]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.

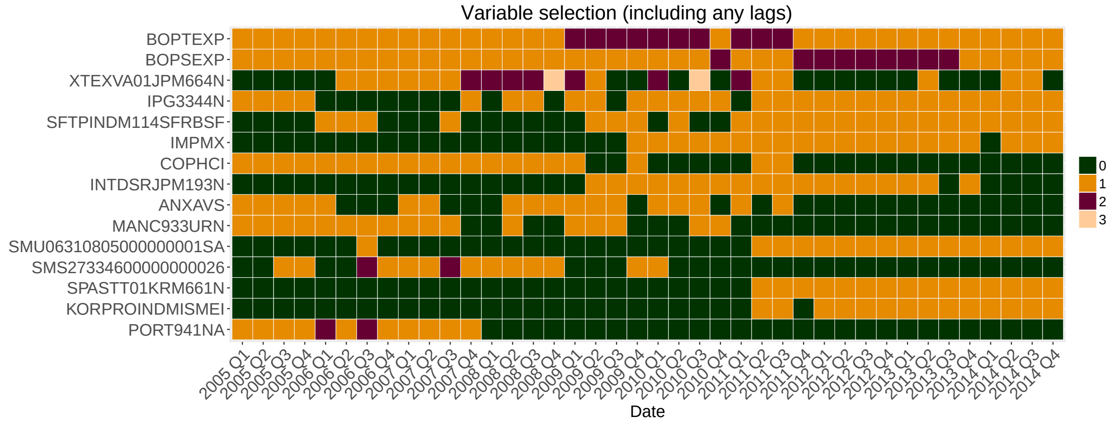

Figure 4.4: Most often selected variables by the AdaptiveLASSO during the rolling window pseudo-real-time forecasting exercise for the Exports.

additional indicators were included, able to somewhat reflect their economic situation, such as interest rates, total industry production or unemployment rate, indicating that the economic health of the largest trading partners of the US can also have an effect to the resulting trade balance numbers. Additionally, among the top variables included is the industrial production of electronic components, the observed growth of which can act as a leading indicator for growth in exports for some the electronic components; and the San Francisco Tech Pulse, indicating the health of the IT sector, which, in addition to explaining exports of electronic components, may also reflect the exports of computer software. Also, the value of shipments for nondefense capital goods is often included, which is a survey indicator, representing the shipments sent by a large part of the US manufacturers, therefore it is natural to expect that similar tendencies can appear in the national accounts data for exports.

Also, the results of forecasting Imports are presented in the tables E.8 and E.9 over a rolling 12-year window. 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 the table E.10, 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.

The monthly variables, most often selected for the modelling of Imports, are presented in the figure 4.5. As expected, the available monthly indicators of volume of imports of goods and services are constantly included. Again, as in the case of forecasting private consumption or exports, with *hard* monthly information already available, it is important to stress the value of accurate individual forecasting of these indicators. For example, one way of improving the individual forecasts of imports and exports might be by forming

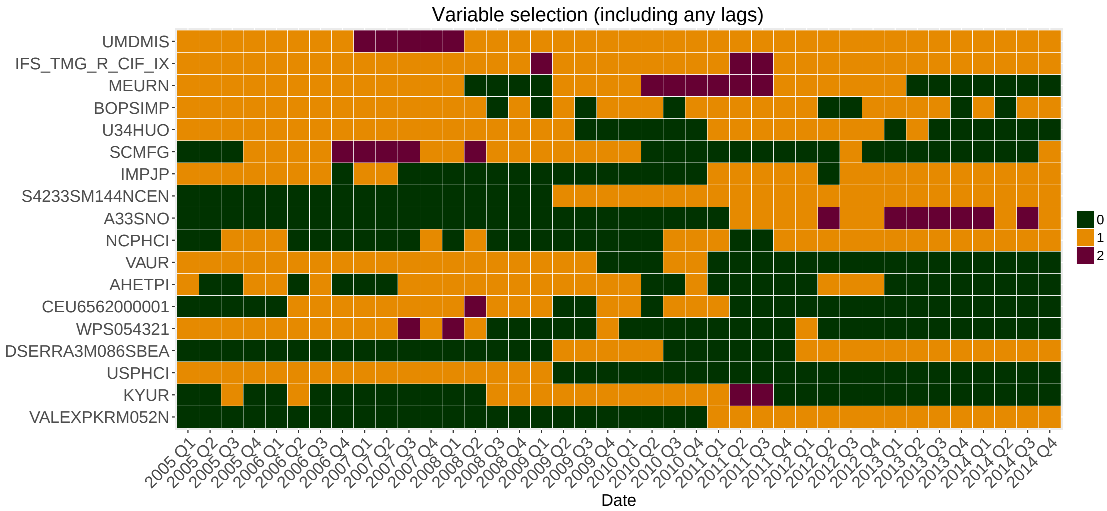

Figure 4.5: Most often selected variables by the AdaptiveLASSO during the rolling window pseudo-real-time forecasting exercise for the Imports.

(Global) VAR models, that would be able to capture the deeper underlying structures by additionally including various domestic and international variables in the modelling.

Additionally, often selected are the indicators on the employment in the transportation and warehousing sector, together with various producer price indices for different commodities used in manufacturing. Also, the indicators for unfilled orders and inventories-to-shipments ratio for durable goods are constantly included, which are known as good leading indicators of the health of industrial production in the country, together with the average wage in the production sector and industrial production index – all possibly indicating, whether the industries are slowing down, bottlenecking or working healthily. For example, a growing amount of orders for the industry might indicate an increase in demand for the particular capital goods, where the manufacturing needs of fulfilling those orders may result in a future growth of imports of some of the materials used in the manufacturing, or it may just indicate a healthy state of the economy and act as a business cycle component.

Chapter 5

Conclusions

Short-term forecasting of quarterly components of the GDP rely on the availability of timely monthly information. In this thesis 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:

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 used 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, analysed in this thesis, 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 modifications of the original LASSO. Therefore, further gains can be expected with additional work on these methods. On the other hand, the studied 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 thesis 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 thesis, when combining the method of principal components with the LASSO, only the standard estimation procedure of the components was discussed, where the preselected variables are scaled before the analysis. However, Stakėnas (2012) has shown significant improvement when using Weighted PCA or Generalized PCA for the extraction of factors when nowcasting Lithuanian GDP, thus suggesting possible further improvement of the method.

References

Angelini, E., M. Banbura, and G. Runstler (2008). Estimating and forecasting the Euro Area monthly national accounts from a dynamic factor model. *ECB Working Paper, No. 953, October*.

Angelini, E., G. Camba-Mendez, D. Giannone, G. Runstler, and L. Reichlin (2008). Short-Term Forecasts of Euro Area GDP Growth. *ECB Working paper #949*.

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.

Banbura, M., D. Giannone, M. Modugno, and L. Reichlin (2013). Now-Casting and the Real-Time Data Flow. *ECARES working paper 2012-026*.

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.

Bloem, A. M., R. J. Dippelsman, and N. O. Maehle (2001). Quarterly national accounts manual – concepts, data sources, and compilation. *Washington, D.C.: International Monetary Fund*.

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. (1995). Better subset regression using the nonnegative garrote. *Technometrics 37*(4), 373–384.

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.

Candes, E. and T. Tao (2007). The dantzig selector: Statistical estimation when p is much larger than n. *Ann. Statist. 35*(6), 2313–2351.

Castle, J. and D. Hendry (2013). Forecasting and Nowcasting Macroeconomic Variables: A Methodological Overview. Economics Series Working Papers 674, University of Oxford, Department of Economics.

Castle, J. L., M. P. Clements, and D. F. Hendry (2013). Forecasting by factors, by variables, by both or neither? *Journal of Econometrics 177*(2), 305 – 319. Dynamic Econometric Modeling and Forecasting.

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

De Mol, C., D. Giannone, and L. Reichlin (2006). Forecasting using a large number of predictors: is Bayesian regression a valid alternative to Principal components? *ECB Working Papers* (700).

Dees, S., F. di Mauro, M. H. Pesaran, and L. V. Smith (2005). Exploring the international linkages of the the Euro Area: A Global VAR analysis. *ECB Working Papers* (568).

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).

Fan, J. and J. Lv (2008). Sure independence screening for ultrahigh dimensional feature space. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 70*(5), 849–911.

Faust, J., J. H. Rogers, and J. H. Wright (2005). News and noise in G-7 GDP announcements. *Journal of Money, Credit and Banking 37*(3), 403–419.

Ghaoui, L. E., V. Viallon, and T. Rabbani (2010). Safe feature elimination in sparse supervised learning. *CoRR abs/1009.4219.*

Giacomini, R. and H. White (2006). Tests of conditional predictive ability. *Econometrica 74*(6), 1545–1578.

Giannone, D., L. Reichlin, and D. Small (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics 55*(4), 665 – 676.

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

Hahn, E. and F. Skudelny (2008). Early estimates of Euro Area Real GDP Growh: A Bottom Up Approach from the production side. *ECB Working paper # 975*.

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

Jokubaitis, S. (2015). Short-term forecasting of domestic investments in Lithuania. Bachelor thesis (in Lith.).

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

Konzen, E. and F. A. Ziegelmann (2016). Lasso-type penalties for covariate selection and forecasting in time series. *Journal of Forecasting 35*(7), 592–612. for.2403.

Leng, C., Y. Lin, and G. Wahba (2006). A note on the Lasso and related procedures in model selection. *Statistica Sinica* (16).

Liu, Z. (2014). The Doubly Adaptive LASSO methods for time series analysis. *Electronic Thesis and Dissertation Repository, Paper 2321*.

Makridakis, S. and M. Hibon (2000). The M3-Competition: results, conclusions and implications. *International Journal of Forecasting 16*(4), 451 – 476. The M3- Competition.

Medeiros, M. C. and E. F. Mendes (2015). l1-Regularization of High-Dimensional Time-Series Models with Flexible Innovations. Textos para discussão 636, Department of Economics PUC-Rio (Brazil).

Meinshausen, N. (2007). Relaxed lasso. *Computational Statistics & Data Analysis 52*(1), 374 – 393.

Pesaran, M. H., T. Schuermann, and L. V. Smith (2009). Forecasting economic and financial variables with global VARs. *International Journal of Forecasting 25*(4), 642 – 675. Special section: Decision making and planning under low levels of predictability.

Smets, F. and R. Wouters (2002). An estimated stochastic dynamic general equilibrium model of the Euro Area. *ECB Working Papers* (171).

Stakėnas, J. (2012). Generating short-term forecasts of the Lithuanian GDP using factor models. Bank of Lithuania Working Paper Series 13, Bank of Lithuania.

Stock, J. and M. Watson (2002). Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economic Statistics 20*(2), 147–162.

Stock, J. and M. Watson (2009). Forecasting in dynamic factor models subject to structural instability. In N. Shephard and J. Castle (Eds.), *The Methodology and Practice of Econometrics: Festschrift in Honor of D.F. Hendry*, pp. 1–57. Oxford: Oxford University Press.

Stock, J. H. and M. W. Watson (2010). Modeling inflation after the crisis. Working Paper 16488, National Bureau of Economic Research.

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

Tibshirani, R., J. Bien, J. Friedman, T. Hastie, N. Simon, J. Taylor, and R. J. Tibshirani (2012). Strong rules for discarding predictors in lasso-type problems. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 74*(2), 245–266.

Tibshirani, R. J. (2012). The Lasso Problem and Uniqueness. *ArXiv e-prints*.

Wang, J., P. Wonka, and J. Ye (2012). Lasso screening rules via dual polytope projection. *CoRR abs/1211.3966*.

Wang, S., B. Nan, S. Rosset, and J. Zhu (2011). Random lasso. *Ann. Appl. Stat. 5*(1), 468–485.

Xiang, Z. J. and P. J. Ramadge (2012). Fast lasso screening tests based on correlations. In *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2137–2140.

Yuan, M. and Y. Lin (2007). On the non-negative garrotte estimator. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 69*(2), 143–161.

Zhao, P. and B. Yu (2006). On model selection consistency of Lasso. *Journal of Machine Learning, Research 7*.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association 101*(476), 1418–1429.

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

Zou, H., T. Hastie, and R. Tibshirani (2006). Sparse principal component analysis. *Journal of Computational and Graphical Statistics 15*(2), 265–286.

Appendix A

Graphs

A.1 Results from Monte Carlo Experiment

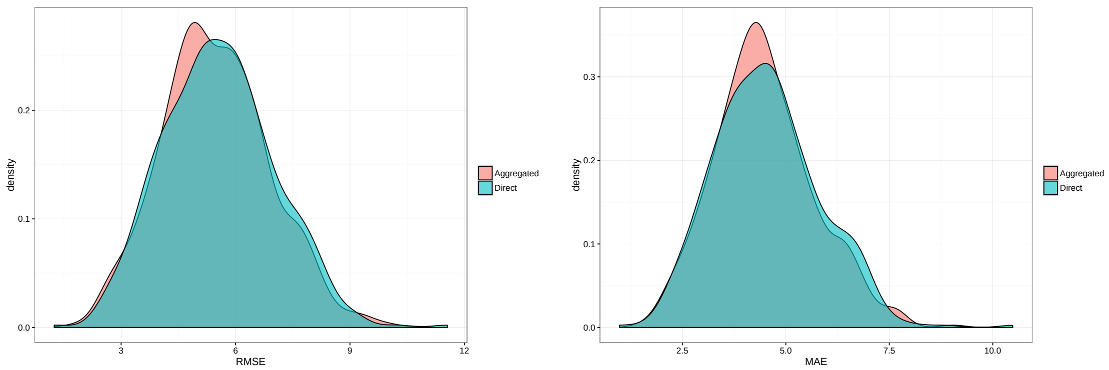

Figure A.1: Density plot of root mean squared errors (RMSE) and mean absolute errors (MAE) of the forecasts, generated by both aggregated and direct predictions, $q = 80$

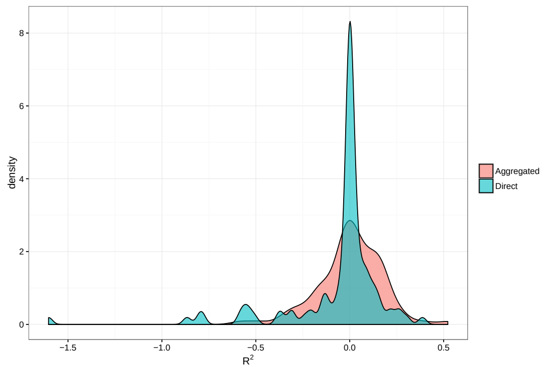

Figure A.2: Density plot of R-squared (R^2) of the forecasts, generated by both aggregated and direct predictions, $q = 80$

A.2 Results from the pseudo-real-time experiments

Figure A.3: Graphs illustrating all 4 of the generated forecasts during the pseudo-real-time experiments for the Gross Fixed Capital Formation over the expanding window for selected models, with the number of preselected variables is fixed to 30.

Figure A.4: Graphs illustrating all 4 of the generated forecasts during the pseudo-real-time experiments for Private Final Consumption Expenditure over the rolling window for best models.

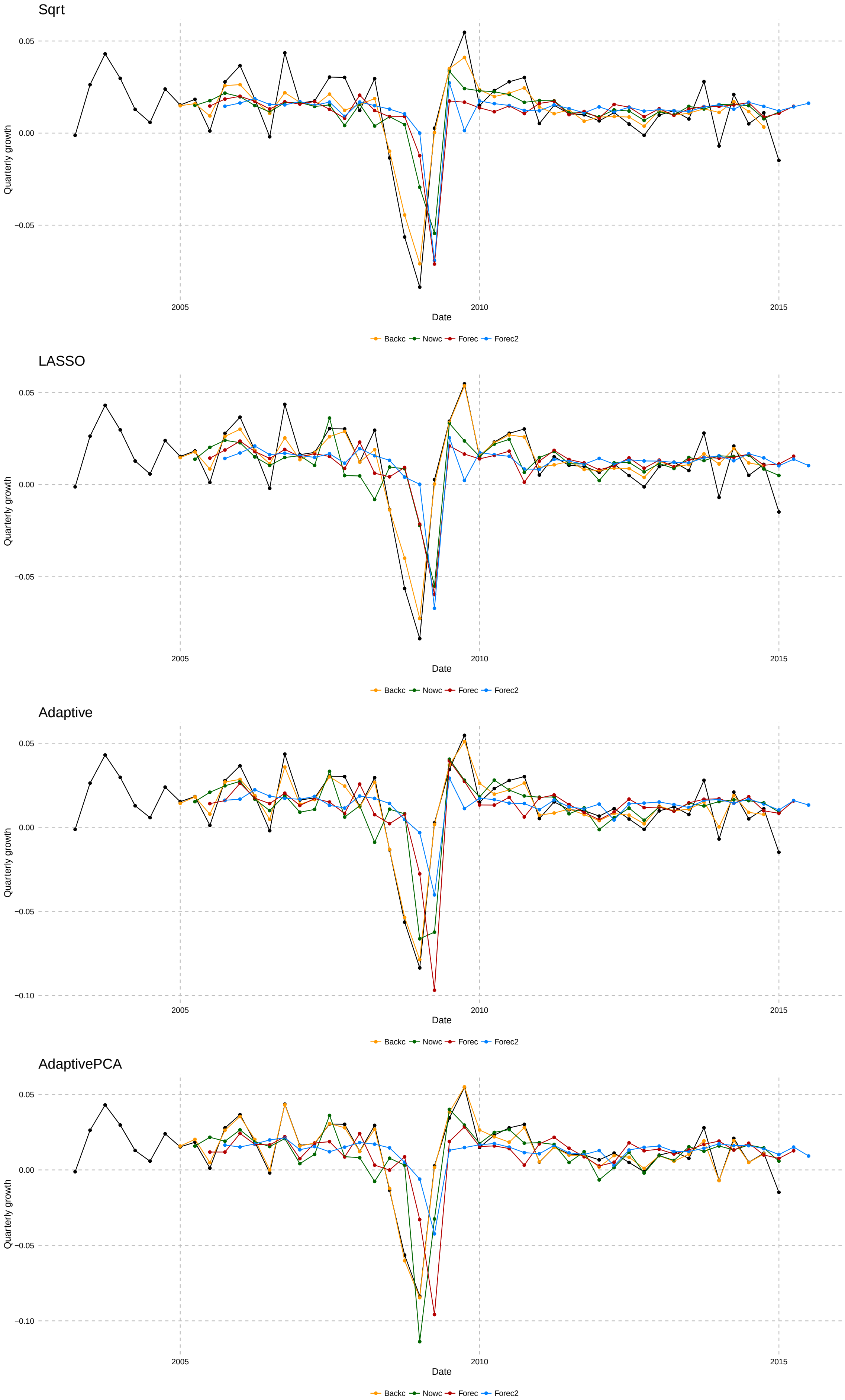

Figure A.5: Graphs illustrating all 4 of the generated forecasts during the pseudo-real-time experiments for Exports over the rolling window for best models.

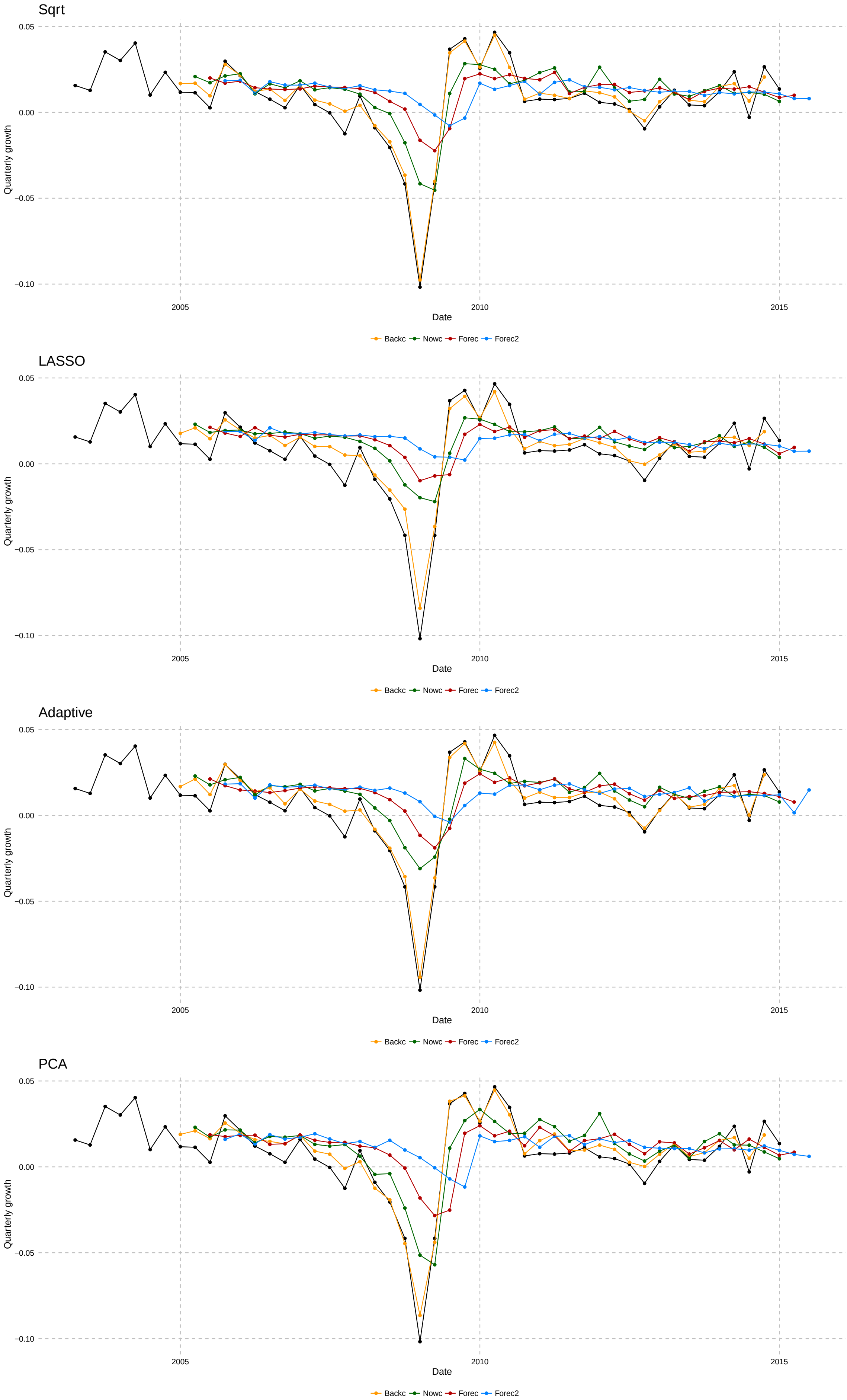

Figure A.6: Graphs illustrating all 4 of the generated forecasts during the pseudo-real-time experiments for Imports over the rolling window for best models.

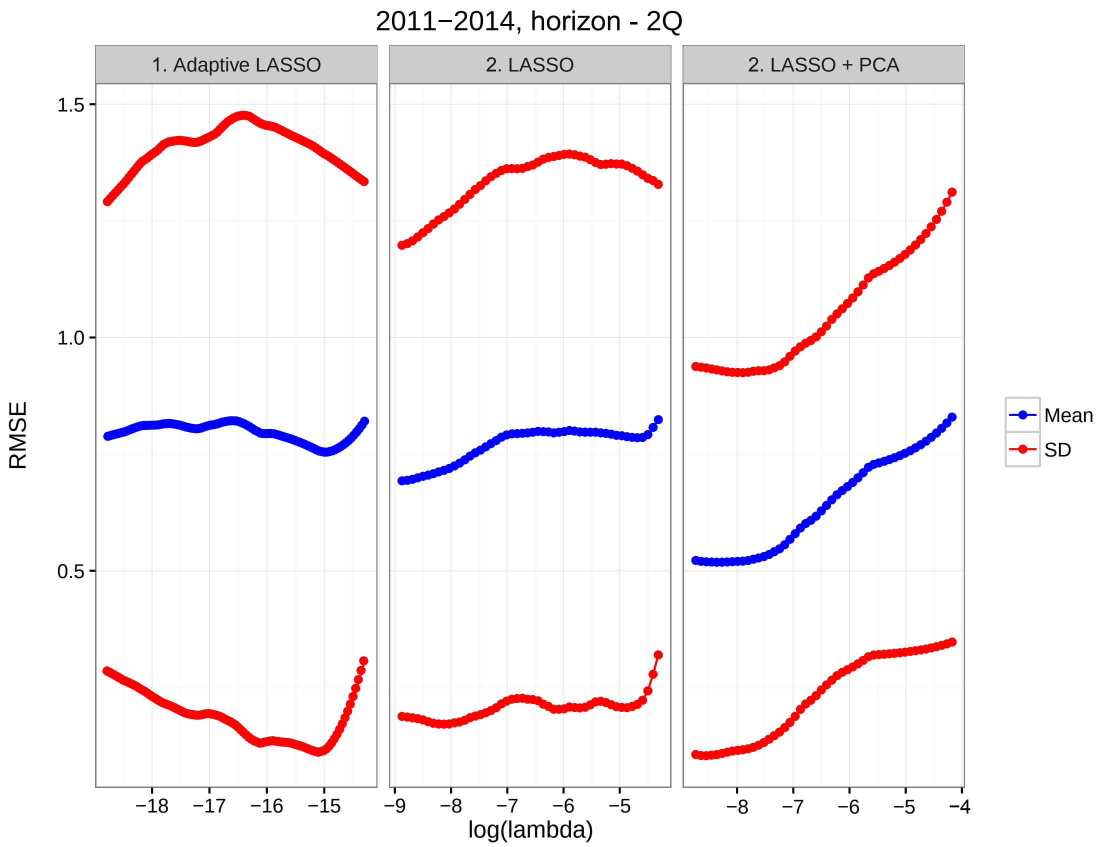

Figure A.7: The results of forecasting accuracy during the pseudo-real-time experiments over 2011-2014 with the set of preselected variables being fixed for the whole period, and the numbers (1.) and (2.) enumerating the different sets of variables used.

Appendix B

Main variables used in the modelling

Table B.1: 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
AZNAN	Δ	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	Δ	1	All Employees: Education and Health Services: Health Care and Social Assistance
COMPU1UNSA	$\Delta \ln$	1	New Privately-Owned Housing Units Completed: 1-Unit Structures
COMPU1USA	$\Delta \ln$	1	New Privately-Owned Housing Units Completed: 1-Unit Structures
CRESTKCPIXSLTRM-159SFRBATL	$\Delta \ln$	1	Sticky Price Consumer Price Index less Food, Energy, and Shelter
CUUR0000SEFR	Δ	1	Consumer Price Index for All Urban Consumers: Sugar and sweets
CUURA423SAH	Δ	1	Consumer Price Index for All Urban Consumers: Housing in Seattle-Tacoma-Bremerton, WA (CMSA)
DPCCRA3M086SBEA	$\Delta \ln$	1	Real personal consumption expenditures excluding food and energy (chain-type quantity index)
DPCERA3M086SBEA	$\Delta \ln$	1	Real personal consumption expenditures (chain-type quantity index)
DSERRA3M086SBEA	$\Delta \ln$	1	Real personal consumption expenditures: Services (chain-type quantity index)
FLUR	$\Delta \ln$	1	Unemployment Rate in Florida
FRBLMCI		1	Change in Labor Market Conditions Index
HOUSTWNSA	Δ	1	Housing Starts in West Census Region
HSN1FMW	$\Delta \ln$	1	New One Family Houses Sold in Midwest Census Region
IPB54100N	$\Delta \ln$	1	Industrial Production: Construction supplies
IPDCONGD	$\Delta \ln$	1	Industrial Production: Durable Consumer Goods
IPG321N	$\Delta \ln$	1	Industrial Production: Durable manufacturing: Wood product
IPG3331N	$\Delta \ln$	1	Industrial Production: Durable Goods: Agriculture, construction, and mining machinery
IPG3344N	$\Delta \ln$	1	Industrial Production: Durable Goods: Semiconductor and other electronic component
IPG336212N	$\Delta \ln$	1	Industrial Production: Durable Goods: Truck trailer
KYNAN	Δ	1	All Employees: Total Nonfarm in Kentucky
KYUR	$\Delta \ln$	1	Unemployment Rate in Kentucky
LNS11300032	Δ	1	Labor Force Participation Rate: 20 years and over, Black or African American Women
LNS12500000	Δ	1	Employed, Usually Work Full Time
MEURN	$\Delta \ln$	1	Unemployment Rate in Maine
MINAN	Δ	1	All Employees: Total Nonfarm in Michigan
NEUR	$\Delta \ln$	1	Unemployment Rate in Nebraska
PCOCOUSDM	$\Delta \ln$	1	Global price of Cocoa
PCU3339113339111Z4	Δ	1	Producer Price Index by Industry: Pump and Pumping Equipment Manufacturing: Industrial Pumps, Except Hydraulic Fluid Power Pumps
PERMIT1NSA	$\Delta \ln$	1	New Privately-Owned Housing Units Authorized by Building Permits: 1-Unit Structures

Table B.1: *(continued)*

Acronym	Transf.	Lag	Name
PORT941NA	Δ	1	All Employees: Total Nonfarm in Portland-Vancouver-Hillsboro, OR-WA (MSA)
RALACBM027SBOG	$\Delta \ln$	1	Residual (Assets Less Liabilities), All Commercial Banks
RIUR	$\Delta \ln$	1	Unemployment Rate in Rhode Island
SFTPINDM114SFRBSF	$\Delta \ln$	1	San Francisco Tech Pulse
SMS27334600000000026	Δ	1	All Employees: Total Nonfarm in Minneapolis-St. Paul-Bloomington, MN-WI (MSA)
SMU06310805000000001SA	$\Delta \ln$	1	All Employees: Information in Los Angeles-Long Beach-Anaheim, CA (MSA)
UNDCON1USA	$\Delta \ln$	1	New Privately-Owned Housing Units Under Construction: 1-Unit Structures
UNDCONTSA	$\Delta \ln$	1	New Privately-Owned Housing Units Under Construction: Total
USCONS	$\Delta \ln$	1	All Employees: Construction
USTRADE	$\Delta \ln$	1	All Employees: Retail Trade
VAPHCI	Δ	1	Coincident Economic Activity Index for Virginia
VAUR	$\Delta \ln$	1	Unemployment Rate in Virginia
VTUR	$\Delta \ln$	1	Unemployment Rate in Vermont
W875RX1	$\Delta \ln$	1	Real personal income excluding current transfer receipts
WPS054321	$\Delta \ln$	1	Producer Price Index by Commodity for Fuels and Related Products and Power: Industrial Electric Power
WPU066	$\Delta \ln$	1	Producer Price Index by Commodity for Chemicals and Allied Products: Plastic Resins and Materials
AHETPI	$\Delta^2 \ln$	1	Average Hourly Earnings of Production and Nonsupervisory Employees: Total Private
AZPHCI	$\Delta^2 \ln$	1	Coincident Economic Activity Index for Arizona
COPHCI	$\Delta^2 \ln$	1	Coincident Economic Activity Index for Colorado
CNP16OV	$\Delta^2 \ln$	1	Civilian Noninstitutional Population
MVMTD027MNFRBDAL	$\Delta^2 \ln$	1	Market Value of Marketable Treasury Debt
NCPHCI	Δ^2	1	Coincident Economic Activity Index for North Carolina
SCMFG	Δ^2	1	All Employees: Manufacturing in South Carolina
USPHCI	Δ^2	1	Coincident Economic Activity Index for the United States
USPRIV	$\Delta^2 \ln$	1	All Employees: Total Private Industries
WINAN	Δ^2	1	All Employees: Total Nonfarm in Wisconsin
WPSFD4131	Δ^2	1	Producer Price Index by Commodity for Final Demand: Finished Goods Less Foods and Energy
A33SNO	$\Delta \ln$	2	Value of Manufacturers' New Orders for Durable Goods Industries: Machinery
ANXAVS	Δ	2	Value of Manufacturers' Shipments for Capital Goods: Nondefense Capital Goods Excluding Aircraft Industries
BOPSEXP	$\Delta \ln$	2	Exports of Services, Balance of Payments Basis
BOPSIMP	$\Delta \ln$	2	Imports of Services, Balance of Payments Basis
BOPTEXP	$\Delta \ln$	2	Exports of Goods and Services, Balance of Payments Basis
BOXTVLM133S	$\Delta \ln$	2	U.S. Exports of Services - Travel
IFS_FILIBOR_1M_PA	$\Delta \ln$	2	Interest Rates, London Interbank Offer Rate, 1-Month, Percent per Annum
IFS_FILIBOR_1Y_PA	$\Delta \ln$	2	Interest Rates, London Interbank Offer Rate, 1-Year, Percent per Annum
IMPJP	Δ	2	U.S. Imports of Goods from Japan, Customs Basis
IMPMX	$\Delta \ln$	2	U.S. Imports of Goods from Mexico, Customs Basis
INTDSRJPM193N	$\Delta \ln$	2	Interest Rates, Discount Rate for Japan
MANC933URN	$\Delta \ln$	2	Unemployment Rate in Manchester, NH (NECTA)
S4233SM144NCEN	$\Delta \ln$	2	Merchant Wholesalers, Except Manufacturers' Sales Branches and Offices Sales: Durable Goods: Lumber and Other Construction Materials Sales
U34HUO	$\Delta \ln$	2	Value of Manufacturers' Unfilled Orders for Durable Goods Industries: Computers and Electronic Products: Electronic Components
UMDMIS	$\Delta \ln$	2	Ratio of Manufacturers' Total Inventories to Shipments for Durable Goods Industries
XTEXVA01CHM657S		2	Exports: Value Goods for Switzerland
XTEXVA01JPM664N	$\Delta \ln$	2	Exports: Value Goods for Japan
XTEXVA01USM664N	$\Delta \ln$	2	Exports: Value Goods for the United States
BOERUKM	$\Delta \ln$	3	Bank of England Policy Rate in the United Kingdom
KORPROINDMISMEI	$\Delta \ln$	3	Production of Total Industry in Korea
MABMM301AUM657S		3	M3 for Australia
IFS_TMG_R_CIF_IX	$\Delta \ln$	3	Goods, Volume of Imports, Index
IFS_TXG_R_FOB_IX	$\Delta \ln$	3	Goods, Volume of Exports, US Dollars, Index
IR3TIB01PLM156N	$\Delta \ln$	3	3-Month or 90-day Rates and Yields: Interbank Rates for Poland
IRLTLT01CHM156N	Δ	3	Long-Term Government Bond Yields: 10-year: Main (Including Benchmark) for Switzerland
SPASTT01DEM657N		3	Total Share Prices for All Shares for Germany
SPASTT01KRM661N	$\Delta \ln$	3	Total Share Prices for All Shares for the Republic of Korea
VALEXPKRM052N	$\Delta \ln$	3	Goods, Value of Exports for Republic of Korea

Appendix C

Selected theorems, referenced in the literature review

Assume that we are interested in modelling the data $y = (y_1, \dots, y_n)'$ and as the explanatory variables we are using $X = (X_1, \dots, X_p)$, where $X_j = (X_{1j}, \dots, X_{nj})'$. Additionally, let's assume that the data generating process has a linear form: $\mathbb{E}[y|X] = \beta_1 X_1 + \ldots + \beta_p X_p$, and that all of the data is centered, therefore we can ignore the constant term in the regression function.

Following Zou (2006) we require the following conditions for our data:

- $Y_i = X_i\beta + \varepsilon_i$, where $\varepsilon_i \sim i.i.d.(0, \sigma^2)$, $i = 1, \dots, n$.
- $n^{-1}X'X \to C$, where C is a positively defined matrix.

Assume that $\mathcal{A} = \{1, 2, \dots, p_0\}$, that is, the set of indices $\{1, \dots, p\}$ are ordered in such a way, that all of the significant variables are the first p_0 ones from the full dataset. Then

$$C = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix},$$

where C_{11} is a $p_0 \times p_0$ matrix.

Let's study the LASSO estimator $\hat{\beta}_{LASSO}$:

$$\hat{\beta}_{LASSO} = \arg\min_{\beta} ||Y - \sum_{j=1}^{p} X_j\beta_j||_2^2 + \lambda \sum_{j=1}^{p} |\beta_j|, \tag{C.1}$$

where $\lambda = \lambda_n$ depends from the sample size n. Assume that $\mathcal{A}_n = \{j : \hat{\beta}_{j,LASSO} \neq 0\}$. Then the variable selection by the LASSO is consistent if and only if $\lim_n \mathbb{P}(\mathcal{A}_n = \mathcal{A}) = 1$. Then Knight and Fu (2000) proves that the following lemmas hold:

Lemma C.1. *If $\lambda_n/n \to \lambda_0 \geq 0$, then $\hat{\beta}_{LASSO} \xrightarrow{p} \arg\min V_1$, where*

$$V_1(u) = (u - \beta)'C(u - \beta) + \lambda_0 \sum_{j=1}^{p} |u_j|$$

Lemma C.2. *If $\lambda_n/\sqrt{n} \to \lambda_0 \geq 0$, then $\sqrt{n}(\hat{\beta}_{LASSO} - \beta) \xrightarrow{\mathcal{D}} \arg\min V_2$, where*

$$V_2(u) = -2u'W + u'Cu + \lambda_0 \sum_{j=1}^{p} \left[u_j \text{sign}(\beta_j) \mathbb{1}_{\{\beta_j \neq 0\}} + |u_j| \mathbb{1}_{\{\beta_j = 0\}} \right],$$

where W is distributed by $\mathcal{N}(0, \sigma^2 C)$.

Lemma C.2 shows that the LASSO estimator is $\sqrt{n}$-consistent. However, the lemma C.1 guarantees the consistency only under the case of $\lambda_0 = 0$ due to the imposed penalty. On the other hand, if we study the asymptotics of the variable selection, lemma C.2 claims, that when $\lambda_n = \mathcal{O}(\sqrt{n})$, $\mathcal{A}_n$ cannot be equal to $\mathcal{A}$ with a positive probability. In other words, the authors prove that the following proposition holds:

Proposition C.1. *If* $\lambda_n/\sqrt{n} \to \lambda_0 \geq 0$*, then* $\limsup_n \mathbb{P}(\mathcal{A}_n = \mathcal{A}) \leq \kappa < 1$*, where* κ *is some constant, dependent from the true data generating model.*

Following this proposition, Zou (2006) proposes another interesting idea: let $\lambda_0 = \infty$, how would then the $\hat{\beta}_{LASSO}$ behave? Apparently, this coincides with the case when $\lambda_n/n \to 0$ and $\lambda_n/\sqrt{n} \to \infty$, where following this observation a lemma C.3 is proved, the proof of which can be found in the cited paper.

Lemma C.3. *Let* $\lambda_n/n \to 0$ *and* $\lambda_n/\sqrt{n} \to \infty$*, then* $\frac{n}{\lambda_n}(\hat{\beta}_{LASSO} - \beta) \xrightarrow{p} \arg\min V_3$*, where*

$$V_3(u) = u'Cu + \sum_{j=1}^{p} \left[u_j \text{sign}(\beta_j) \mathbb{1}_{\{\beta_j \neq 0\}} + |u_j| \mathbb{1}_{\{\beta_j = 0\}} \right].$$

It can be observed from the lemma C.3 that $\hat{\beta}_{LASSO}$ converges at a slower rate than of $\sqrt{n}$. Moreover, the limiting value is not random. Additionally, we can see that the optimal rate of convergence can be achieved only when $\lambda_n = \mathcal{O}(\sqrt{n})$, which in turn results in asymptotically inconsistent variable selection. Additionally, the authors stress that even by sacrificing some of the convergence rate of the estimates, the consistency of the variable selection is still not guaranteed.

Theorem C.1. *(Irrepresentable Condition). Let* $\lim_n \mathbb{P}(\mathcal{A}_n = \mathcal{A}) = 1$*. Then there exists a sign vector* $s = (s_1, \ldots, s_{p_0})'$*, where* $s_j = 1$ *or* -1*, such that the following holds:*

$$|C_{21}C_{11}^{-1}s| \leq 1, \tag{C.2}$$

where the inequality is understood component-wise.

If the C.2 inequality does not hold, the resulting variable selection of the LASSO is inconsistent. Authors claim that there are many situations in reality where the mentioned requirement is broken, however on the average case the LASSO is able to adequatelly perform the variable selection. Additionally, it is noted that when the LASSO is performed on orthogonal data, the consistency of the variable selection is guaranteed since the inequality C.2 will always hold.

Appendix D

Discussion on LASSO screening

In the context of this thesis it is worth highlighting the importance of the *screening* procedure. One of the main reasons for making LASSO method attractive for practitioners is the ability to identify the significant explanatory variables from the whole information set. However, when this set becomes very large ($p >> n$), the (analytical) solution of the LASSO problem (1.1) is not trivial, therefore the solution is usually obtained through various mathematical algorithms (i.e. LARS, coordinate descent and other), requiring serious computing capabilities. For this reason, the initial problem is simplified by performing *screening* procedure on the main information set. That is, this set is simplified by removing the *inactive* set of variables, consisting of truly insignificant variables. In this way, the mathematical algorithms are able to work with a reduced feature matrix, making the calculations faster and more efficient.

Currently existing screening methods for LASSO may be divided into two categories: the *heuristic* and the *safe* screening methods. Noteworthy that the heuristic screening methods are not able to fully guarantee that their inactive set is consisting of only truly insignificant variables. Most popular of such methods – SIS (Fan and Lv (2008)) and *strong rules* (Tibshirani et al. (2012)), the former of which is based on associations between explanatory and modelled variables, while the latter relies on the assumption that the absolute values of the inner products between features and the residue are *non-expansive* with respect to the parameter values. Wang et al. (2012) claim that this assumption does not always hold in reality, therefore when applying heuristic screening methods in practice, during every step of the LASSO estimation, the *Karush-Kuhn-Tucker* (KKT) conditions are also checked to ensure the correctness of the solution. In case of a violation, the screened set is weakened and the process is repeated. On the other hand, *safe* algorithms can guarantee that the discarded variables are absent from the true sparse model. The most popular *safe* methods are SAFE (Ghaoui et al. (2010)) and DOME (Xiang and Ramadge (2012)), based on an estimation of the dual optimal solution of the LASSO.

The main reason why highlighting these facts is important is the fact that the current software, developed for the efficient and fast estimation of the LASSO under large-scale problems is using the heuristic methods due to their relative simplicity and rapid calculations. For example, according to the authors of `glmnet` package for `R`, the use of heuristic *strong* method should not in reality produce inaccurate screening, even though the authors provide an example when the method fails. In addition to this, Wang et al. (2012) propose an even more efficient (calculation speed-wise) screening method, which is not heuristic and guarantees accurate screening, however, it is not yet implemented in `R`.

For this reason, while working with large-scale data in practice it may be useful in some cases to avoid the heuristic screening by either reducing to slower, but more accurate *safe* screening procedures for the LASSO, or, for example, skipping the screening procedure by iterating through smaller subsets, if the scale of the problem allows it.

Appendix E

Tables

Table E.1: 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.

	AdaPCA15	AdaRL15	AdaPCA15X	AdaPCA20	AdaRL20	AdaPCA20X	AdaPCA25	AdaRL25	AdaPCA25X
Back									
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	**0.501**
Now									
05-08	**0.821**	0.832	0.918	0.799	**0.776**	0.827	**0.746**	0.791	0.765
08-11	1.608	1.525	**1.522**	1.510	1.523	**1.386**	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									
05-08	**0.979**	1.052	1.017	**0.987**	0.991	1.026	**0.993**	1.005	1.009
08-11	2.210	2.280	**2.116**	2.159	2.221	**2.120**	2.217	2.316	**2.170**
11-15	1.159	**1.048**	1.159	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									
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	2.699	**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**

E.1 Private Final Consumption Expenditure

Table E.2: RMSE of models forecasts during rolling window pseudo-real-time experiments for Private Final Consumption Expenditure, 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	Relaxed	ARMA
Back							
05-08	0.166	0.183	0.0817	0.079	0.059	**0.056**	—
08-11	0.137	0.283	0.108	0.183	**0.098**	0.180	—
11-15	0.120	0.157	**0.116**	0.155	0.123	0.144	—
05-15	0.140	0.210	0.102	0.147	**0.097**	0.137	—
Now							
05-08	0.361	0.391	0.348	0.354	**0.313**	0.365	0.453
08-11	0.350	0.438	0.314	0.325	0.387	**0.285**	0.700
11-15	0.233	0.219	0.239	**0.212**	0.237	0.217	0.284
05-15	0.301	0.341	0.289	0.286	0.307	**0.279**	0.485
Fore1Q							
05-08	0.451	0.470	0.426	0.441	**0.406**	0.443	0.493
08-11	0.640	0.670	0.592	0.646	**0.589**	0.619	0.838
11-15	0.340	**0.309**	0.329	0.322	0.321	0.326	0.317
05-15	0.472	0.480	0.445	0.470	**0.437**	0.461	0.568
Fore2Q							
05-08	0.503	0.519	0.474	0.478	**0.429**	0.481	0.532
08-11	0.783	0.804	0.761	0.776	**0.753**	0.768	0.970
11-15	0.365	0.343	0.366	0.352	0.362	0.358	**0.326**
05-15	0.556	0.563	0.540	0.546	**0.526**	0.544	0.652

Table E.3: Relative (to ARMA models') RMSE of models forecasts during rolling window pseudo-real-time experiments for Private Final Consumption Expenditure, 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	Relaxed	ARMA
Now							
05-08	0.80	0.86	0.77	0.78	**0.69**	0.81	1
08-11	0.50	0.63	0.45	0.46	0.55	**0.41**	1
11-15	0.82	0.77	0.84	**0.75**	0.83	0.76	1
05-15	0.62	0.70	0.60	0.59	0.63	**0.58**	1
Fore1Q							
05-08	0.91	0.95	0.86	0.89	**0.82**	0.90	1
08-11	0.76	0.80	0.71	0.77	**0.70**	0.74	1
11-15	1.07	**0.98**	1.03	1.01	1.01	1.02	1
05-15	0.83	0.85	0.78	0.83	**0.77**	0.81	1
Fore2Q							
05-08	0.95	0.98	0.89	0.90	**0.81**	0.90	1
08-11	0.80	0.82	0.78	0.79	**0.77**	0.78	1
11-15	1.11	1.05	1.12	1.07	1.11	1.09	**1**
05-15	0.85	0.86	0.83	0.84	**0.81**	0.83	1

Table E.4: 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 marks 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	-	-	-	-	-	-	**0.054**
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	0.15	0.23	0.67
PCA	-	-	-	-	0.21	0.21	0.68
AdaptivePCA	-	-	-	-	-	0.28	0.38
Relaxed	-	-	-	-	-	-	0.65
ARMA	-	-	-	-	-	-	-

E.2 Exports

Table E.5: RMSE of models forecasts during rolling window pseudo-real-time experiments for Exports, 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	Relaxed	ARMA
Back							
05-08	0.975	0.700	0.444	1.078	**0.161**	0.994	—
08-11	0.819	0.649	**0.439**	0.888	0.454	0.551	—
11-15	0.732	0.622	0.473	0.774	**0.325**	0.712	—
05-15	0.847	0.668	0.464	0.915	**0.343**	0.768	—
Now							
05-08	1.431	1.405	**1.267**	1.590	1.335	1.721	1.632
08-11	3.099	3.392	3.020	2.587	**2.562**	3.018	4.259
11-15	1.057	**0.966**	1.044	1.264	1.087	1.149	1.073
05-15	2.041	2.166	1.967	1.887	**1.768**	2.089	2.698

Table E.5: *(continued)*

	Sqrt	LASSO	Adaptive	PCA	AdaPCA	Relaxed	ARMA
Fore1Q							
05-08	1.462	**1.380**	1.392	1.490	1.400	1.471	1.584
08-11	3.743	**3.478**	3.901	3.843	3.880	5.153	4.337
11-15	1.043	1.030	1.088	1.054	1.136	1.013	**0.950**
05-15	2.387	**2.233**	2.460	2.452	2.462	3.137	2.735
Fore2Q							
05-08	1.533	1.483	1.534	1.587	**1.476**	1.509	1.606
08-11	4.023	3.925	3.445	3.835	**3.444**	4.761	3.936
11-15	1.091	1.058	1.091	1.109	1.107	1.111	**0.989**
05-15	2.583	2.517	2.279	2.498	**2.272**	2.983	2.552

Table E.6: Relative (to ARMA models') RMSE of models forecasts during rolling window pseudo-real-time experiments for Exports, 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	Relaxed	ARMA
Now							
05-08	0.88	0.86	**0.78**	0.97	0.82	1.05	1
08-11	0.73	0.80	0.71	0.61	**0.60**	0.71	1
11-15	0.99	**0.90**	0.97	1.17	1.01	1.07	1
05-15	0.76	0.80	0.73	0.70	**0.66**	0.77	1
Fore1Q							
05-08	0.92	**0.87**	0.88	0.94	0.88	0.93	1
08-11	0.86	**0.80**	0.90	0.89	0.89	1.18	1
11-15	1.09	1.08	1.14	1.10	1.19	1.06	**1**
05-15	0.87	**0.82**	0.90	0.90	0.90	1.10	1
Fore2Q							
05-08	0.95	0.92	0.96	0.99	**0.92**	0.94	1
08-11	1.02	0.99	0.88	0.97	**0.88**	1.20	1
11-15	1.10	1.06	1.10	1.12	1.12	1.12	**1**
05-15	1.01	0.99	0.89	0.98	**0.89**	1.16	1

Table E.7: 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 marks 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	-	-	-	0.51	0.53	0.59	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	0.33	0.23
Adaptive	-	-	-	0.97	0.86	0.3	0.89
PCA	-	-	-	-	1	0.34	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
Adaptive	-	-	-	0.28	0.54	0.43	**0.14**
PCA	-	-	-	-	0.25	0.6	0.33
AdaptivePCA	-	-	-	-	-	0.51	0.15
Relaxed	-	-	-	-	-	-	0.53
ARMA	-	-	-	-	-	-	-

E.3 Imports

Table E.8: RMSE of models forecasts during rolling window pseudo-real-time experiments for Imports, 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	Relaxed	ARMA
Back							
05-08	**0.552**	0.838	0.692	0.782	0.800	0.647	—
08-11	**0.374**	0.833	0.582	0.547	0.451	0.428	—
11-15	0.430	0.592	**0.369**	0.607	0.410	0.520	—
05-15	**0.456**	0.763	0.549	0.647	0.563	0.521	—
Now							
05-08	1.140	1.305	1.226	1.163	**1.138**	1.213	1.507
08-11	2.286	2.922	2.629	**2.071**	2.234	2.322	5.385
11-15	1.204	1.125	1.098	1.243	1.126	1.221	**1.012**
05-15	1.626	1.946	1.782	**1.535**	1.585	1.659	3.277
Fore1Q							
05-08	1.250	1.400	1.331	**1.229**	1.268	1.298	1.464
08-11	**3.352**	3.609	3.475	3.469	3.513	3.383	5.636
11-15	1.108	1.140	1.073	1.120	**1.001**	1.127	1.006
05-15	**2.164**	2.327	2.232	2.219	2.226	2.188	3.443
Fore2Q							
05-08	**1.235**	1.378	1.288	1.304	1.255	1.243	1.503
08-11	4.215	4.280	4.247	4.297	4.214	**4.165**	5.358
11-15	1.036	1.048	1.056	1.007	**0.990**	1.051	1.052
05-15	2.636	2.690	2.661	2.684	2.628	**2.611**	3.335

Table E.9: Relative (to ARMA models') RMSE of models forecasts during rolling window pseudo-real-time experiments for Imports, 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	Relaxed	ARMA
Now							
05-08	0.76	0.87	0.81	0.77	**0.76**	0.80	1
08-11	0.42	0.54	0.49	**0.38**	0.41	0.43	1
11-15	1.18	1.11	1.08	1.22	1.11	1.20	**1**
05-15	0.50	0.59	0.54	**0.47**	0.48	0.51	1
Fore1Q							
05-08	0.85	0.96	0.91	**0.84**	0.87	0.89	1
08-11	**0.59**	0.64	0.62	0.62	0.62	0.6	1
11-15	1.10	1.13	1.06	1.11	**0.99**	1.12	1
05-15	**0.63**	0.68	0.65	0.64	0.65	0.64	1
Fore2Q							
05-08	**0.82**	0.92	0.86	0.87	0.83	0.83	1
08-11	0.79	0.80	0.79	0.80	0.79	**0.78**	1
11-15	0.98	0.99	1.00	0.96	**0.94**	0.99	1
05-15	0.79	0.81	0.80	0.80	0.79	**0.78**	1

Table E.10: 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
Nowcast							
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	0.33	**0.088**
AdaptivePCA	-	-	-	-	-	0.31	**0.06**
Relaxed	-	-	-	-	-	-	**0.099**
ARMA	-	-	-	-	-	-	-
1Q							
Sqrt	-	0.11	0.22	0.31	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	-	-	-	-	0.4	0.22	0.14
AdaptivePCA	-	-	-	-	-	0.53	0.1
Relaxed	-	-	-	-	-	-	0.14
ARMA	-	-	-	-	-	-	-
2Q							

Table E.10: *(continued)*

Nowcast	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
AdaptivePCA	-	-	-	-	-	0.62	0.25
Relaxed	-	-	-	-	-	-	0.22
ARMA	-	-	-	-	-	-	-