https://doi.org/10.15388/vu.thesis.388
https://orcid.org/0000-0002-6561-2679

VILNIUS UNIVERSITY

Saulius Jokubaitis

Sparse models: theory and applications

DOCTORAL DISSERTATION

Natural Sciences,
Mathematics (N 001)

VILNIUS 2022

The dissertation was prepared between 2017–2021 at Vilnius University.
Academic supervisor – Prof. Habil. Dr. Remigijus Leipus (Vilnius University, Natural Sciences, Mathematics – N 001).

Dissertation Defence Panel:

Chairman – Prof. Habil. Dr. Vydas Čekanavičius (Vilnius University, Natural Sciences, Mathematics – N 001).

Members:
Prof. Habil. Dr. Kęstutis Kubilius (Vilnius University, Natural Sciences, Mathematics – N 001),
Assoc. Prof. Dr. Jurgita Markevičiūtė (Vilnius University, Natural Sciences, Mathematics – N 001),
Prof. Dr. Anne Philippe (Nantes Université, Natural Sciences, Mathematics – N 001),
Prof. Dr. Jonas Šiaulys (Vilnius University, Natural Sciences, Mathematics – N 001).

The dissertation will be defended at a public meeting of the Dissertation Defence Panel at 3 pm. on 28th of October 2022 in room 102 of the Faculty of Mathematics and Informatics, Vilnius University.
Address: 24 Naugarduko st., LT-03225, Vilnius, Lithuania
Tel.: +370 5 219 3050; e-mail: mif@mif.vu.lt.

The text of this dissertation can be accessed at the library of Vilnius University as well as on the website of Vilnius University:
www.vu.lt/lt/naujienos/ivykiu-kalendorius

https://doi.org/10.15388/vu.thesis.388
https://orcid.org/0000-0002-6561-2679

VILNIAUS UNIVERSITETAS

Saulius Jokubaitis

Retų struktūrų modeliai: teorija ir taikymai

DAKTARO DISERTACIJA

Gamtos mokslai,
Matematika (N 001)

VILNIUS 2022

Disertacija rengta 2017–2021 metais Vilniaus universitete.
Mokslinis vadovas – prof. habil. dr. Remigijus Leipus (Vilniaus universitetas, gamtos mokslai, matematika – N 001).

Gynimo taryba:

Pirmininkas – prof. habil. dr. Vydas Čekanavičius (Vilniaus universitetas, gamtos mokslai, matematika – N 001).

Nariai:
prof. habil. dr. Kęstutis Kubilius (Vilniaus universitetas, gamtos mokslai, matematika – N 001),
doc. dr. Jurgita Markevičiūtė (Vilniaus universitetas, gamtos mokslai, matematika – N 001),
prof. dr. Anne Philippe (Nantes Université, gamtos mokslai, matematika – N 001),
prof. dr. Jonas Šiaulys (Vilniaus universitetas, gamtos mokslai, matematika – N 001).

Disertacija ginama viešame Gynimo tarybos posėdyje 2022 m. spalio mėn. 28 d. 15 val., Vilniaus universiteto Matematikos ir informatikos fakultete, 102 auditorijoje. Adresas: Naugarduko g. 24, LT-03225, Vilnius, Lietuva, tel. +370 5 219 3050; el. paštas: mif@mif.vu.lt.

Disertaciją galima peržiūrėti Vilniaus universiteto bibliotekoje ir Vilniaus universiteto interneto svetainėje adresu:
https://www.vu.lt/naujienos/ivykiu-kalendorius

Contents

Notation and abbreviations

Notation	Description
$\mathbb{R}$	The set of real numbers.
$\mathbb{N}$	The set of natural numbers, $\mathbb{N} = \{0, 1, 2, ...\}$.
$\mathbb{N}_+$	The set of positive natural numbers, $\mathbb{N} = \{1, 2, 3, ...\}$.
I_n	$n \times n$ identity matrix.
$\mathbb{1}_A$	Indicator function of a set A.
$\stackrel{d}{=}$	Equality in distribution.
$\stackrel{d}{\to}$	Convergence in distribution.
$\stackrel{\mathbb{P}}{\to}$	Convergence in probability.
$\|x\|_q$	ℓ_q-norm of a vector $x = (x_1, \ldots, x_d)'$, $d \geq 1$, $q \geq 0$.
$\lfloor x \rfloor$	The integer part of a real number x (floor).
$\mathcal{O}(a_n)$	For sequences a_n, b_n, we write $b_n = \mathcal{O}(a_n)$ if there exists constants N and C such that $\|b_n\| \leq C\|a_n\|$ for every $n \geq N$.
$\mathcal{O}_P(a_n)$	For a set of random variables X_n, we write $X_n = \mathcal{O}_P(a_n)$ if for every $\varepsilon > 0$ there exists constants C_ε and n_ε such that $\mathbb{P}(\|X_n\| \leq C_\varepsilon a_n) > 1 - \varepsilon$ for every $n \geq n_\varepsilon$.
$o(a_n)$	For sequences a_n, b_n, we write $b_n = o(a_n)$ if $b_n/a_n \to 0$, $n \to \infty$.
$o_P(a_n)$	For a set of random variables X_n, we write $X_n = o_P(a_n)$ if $X_n/a_n \stackrel{\mathbb{P}}{\to} 0$, $n \to \infty$.
$x \gg y$	x is much greater than y. I.e., there exists some implicitly large C such that $x > Cy$.
diag	$D = \text{diag}(a_1, \ldots, a_n)$ is a diagonal matrix with elements $a_1, \ldots, a_n$ on the main diagonal.

Notation	Description
$\equiv$	$f(x) \equiv g(x)$, f and g are equivalent, if $f(x) = g(x)$ for any x.
$\sim$	$a_n \sim b_n$ means that $a_n/b_n \to 1$, as $n \to \infty$.
$\mathcal{N}(\mu, \sigma^2)$	Normal distribution with mean $\mu \in \mathbb{R}$ and variance $\sigma^2 > 0$.
$\mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$	Multivariate normal distribution of a p-dimensional random vector with mean $\boldsymbol{\mu} \in \mathbb{R}^p$ and covariance matrix $\Sigma \in \mathbb{R}^{p\times p}$.
$\mathrm{VG}(r, \theta, \sigma, \mu)$	Variance-Gamma distribution with parameters $r > 0$, $\theta \in \mathbb{R}$, $\sigma > 0$, $\mu \in \mathbb{R}$.
$\Gamma(\alpha, \beta)$	Gamma distribution with shape $\alpha > 0$ and rate $\beta > 0$.
$\varphi_X(t)$	Characteristic function of a random variable X at point $t \in \mathbb{R}$.
$\mathbb{E}X$	Expectation of a random variable X.
$\mathrm{Var}(X)$	Variance of a random variable X.
$\mathrm{Corr}(X, Y)$	Correlation between random variables X and Y.
$\mathrm{Cov}(X, Y)$	Covariance between random variables X and Y.
$\mathrm{tr}(A)$	Trace of a matrix A.
$\mathrm{Li}_2(x)$	Real dilogarithm function, $x \leq 1,\ x \in \mathbb{R}$.

Abbreviation	Description
i.i.d	Independent identically distributed.
SNR	Signal-to-noise ratio.
KMS	Kac–Murdock–Szegő, used defining a specific covariance structure.
CDF	Cumulative distribution function.
PDF	Probability density function.
Q-Q plot	A quantile-quantile plot.
RMSE	Root mean square error.
GDP	Gross domestic product.
GFCF	Gross fixed capital formation.
PFCE	Private final consumption expenditure.
PCA	Principal component analysis.
LASSO	Least absolute shrinkage selection operator.
LP	A combination of LASSO and principal components methods (LASSO-PC).

Abbreviation	Description
LARS	Least Angle Regression
L0Lq	Fast Best Subset Selection algorithm, $q \in 1, 2$.
SqL	Square-Root LASSO.
SqP	Square-Root LASSO-PC.
RL	Relaxed LASSO.
AdL	Adaptive LASSO.
AdP	Adaptive LASSO-PC.
L0P, L0LqP	A Fast Best Subset Selection algorithm, $q \in 1, 2$, combined with the method of principal components.
LPX	LASSO-PC-X. A modification of LASSO-PC algorithm, where the set of factors is expanded by including a set of original variables.
DF	Dynamic factor model.
SW, DI	Stock and Watson (2002) diffusion index model.
BNDF	A dynamic factor model, modified as suggested by Bai and Ng (2008).

Introduction

The recent availability of large datasets, combined with advances in the fields of statistics, machine learning and econometrics, have generated interest in predictive models with many possible predictors. In these cases, standard techniques such as ordinary least squares, maximum likelihood, or Bayesian inference with uninformative priors perform poorly, since the proliferation of regressors magnifies estimation uncertainty and produces inaccurate out-of-sample predictions. As a consequence, various new approaches aimed at dealing with this curse of dimensionality have become increasingly popular (Giannone et al. (2021)). In the current high-dimensional modelling literature two main schools of thought can be discerned, differing by the assumption of either *dense* or *sparse* structure for the underlying signal of the data generating process (Ng (2013), Chernozhukov et al. (2017)).

Under the assumption of *dense* structures, the prevailing approach is to make use of all of the available information for the analyst. While the specifics of the used approach does vary based on the application field, in the literature, among the more widely used dense methods, one could often find applications of factor models, ridge regression, weighted-average least squares (WALS) and numerous other (see, e.g., Tikhonov (1963), Hoerl and Kennard (1970), Magnus and De Luca (2016), Andreini et al. (2020)). For instance, such approaches find wide applications for many different problems with large datasets in economics, see, e.g., Stock and Watson (2002), Diebold (2003), De Mol et al. (2006) and Stock and Watson (2009). These approaches are able to effectively use all of the explanatory information passed to the model, even though the individual effect of any separate variable might be reduced to a relatively small size.

On the other hand, under the assumption of *sparse* structures we assume that only a small subset of explanatory variables are crucial and focus on the selection and estimation of such variables. After identifying the key explanatory components, the remaining information is then typically discarded or shrunk to zero. Among the more popular sparse approaches is the *Least Absolute Shrinkage Selection Operator* (LASSO) method, pioneered by Tibshirani (1996). This method employs the supervised variable selection for modelling and shows great potential when used for both variable selection and forecasting. Due to its simplicity of use and attractiveness in different contexts, it is favoured by both practitioners and the academics, which in turn yielded many useful extensions and modifications of the original methodology (see, e.g., Adaptive LASSO (Zou (2006)), Relaxed LASSO (Meinshausen (2007)), Square-Root LASSO (Belloni et al. (2011))). In Chapter 1 of this dissertation we present a detailed review of popular sparse methods.

Sparsity can be motivated on economic grounds in situations where a researcher believes that the economic outcome could be well-predicted by a small (relative to sample size) number of factors but is unsure about the identity of the relevant factors. For instance, Bai and Ng (2008) find that large amounts of high-frequency data could often considerably worsen the predictive performance, while Bulligan et al. (2015) show that assuming a sparse structure of the underlying data can improve the quality of short-run forecasts, when the final set of available information is refined by supervised selection before modelling. However, it is unrealistic to presume exact sparsity – typically it suffices to assume that all regressors potentially have a non-zero contribution to the regression function, but no more than a small number of unknown regressors are needed for approximating the regression function with a sufficient degree of accuracy, while the remaining (not included) are hidden by the variability of the irreducible error term.

Following the surge of advancements in the sparse modelling literature, the rise of new approaches led to solving a wide array of problems in the high-dimensional regression context, introducing many promising results. See, e.g., Shibata (1980), Ing (2007) for model selection problems in autoregressive time series models; Belloni and Chernozhukov (2011), Javanmard and Montanari (2014), Zhang and Zhang (2014), Caner and Kock (2018), Belloni et al. (2018), Gold et al. (2020), Ning et al. (2020), Guo et al. (2021) for applications on inference of high-dimensional models and high-dimensional instrumental variable (IV) regression models.

In the aforementioned and related work, reliance on the exact or approximate sparsity of the underlying model parameters is crucial in order to achieve optimal convergence rates for the corresponding parameter estimates (see, e.g., Bradic et al. (2021)).

Lately, however, there has been an influx of results presenting contradictory findings and suggesting that no one specific type of modelling approach can be universally best. It seems that no method comes as a free lunch. Hence, the accuracy of any specific method often largely depends on the underlying data generating process (see, e.g., Giannone et al. (2021)). Unsurprisingly, such observations inspired further refinements in the modelling approaches, leading to a growth of literature on various modifications and combinations of both sparse and dense methods, generalizing and capturing the best of both worlds (see, among many others, Targeted Diffusion Index models (Bulligan et al. (2015)), Lava (Chernozhukov et al. (2017)), Fast Best Subset Selection (Hazimeh and Mazumder (2020)), Spike-and-Slab priors (Giannone et al. (2021))), which show promising results in varying application setups. In Chapter 3 we explore the aforementioned ideas in greater detail by performing pseudo-real-time experiments for nowcasting and short-term forecasting the US and EU GDP expenditure components. We compare the short-term forecasts of LASSO and its' popular modifications with a few variants of factor models as good *dense* structure alternatives based on their known performance and popularity in the nowcasting literature. We find that in many cases the sparse methods are able to outperform the dense alternatives, however, that was not the case for every country or indicator. In addition, we propose an adjustment that combines variants of LASSO with the principal components method (LASSO-PC), that deemed to improve the forecasting performance. We argue that LASSO-PC could represent a viable combination of both sparse and dense approaches, thus the obtained results are consistent with the observations from the aforementioned literature.

Likely, the underlying reason behind the contradicting results relates with the viability of the assumption on sparse structures of the data generating process. With such evidence in mind, Giannone et al. (2021) suggest that sparsity should not be simply assumed as that is uncertain, hence, should only be used in the presence of strong statistical evidence. The straightforward approach is allowing for both sparse and dense solutions in the model specification, leaving the flexibility for the statistical method to select which solution is the most appropriate for a given problem (see, e.g., Chernozhukov et al. (2017), Cevid et al. (2020), Giannone et al. (2021)).

The universal nature and the flexibility of such approaches can come with a certain price in computational complexity and forecasting accuracy, through, e.g., limiting the underlying dynamics of the individual forecasts. An interesting alternative approach would be a statistical test for evidence on approximate sparsity in a given high-dimensional data set. However, the literature on testing for sparsity in a regression model is relatively scarce. Promising results were achieved by Dicker (2014, 2016) who consider the signal-to-noise ratio (SNR) estimation in the linear regression context, which, in some sense, can be understood as a dual problem to testing for sparsity. Such observation is supported by the findings of Carpentier and Verzelen (2019, 2021), who build on the existing results from the SNR literature and consider the test for exact sparsity in high-dimensional linear regression.

Motivated by the aforementioned ideas, in Chapter 2, we contribute to the literature of SNR and sparsity testing by considering a specific Kac–Murdock–Szegő (KMS) covariance structure for the explanatory variables and deriving both the exact and the asymptotic distributions for the suitably centered and normalized squared norm of the product between the predictor matrix, $\mathbb{X}$, and the outcome variable, Y, i.e., the statistic $\|\mathbb{X}'Y\|_2^2$, under rather nonrestrictive assumptions for the model parameters β_j. We deem that the derived results can be crucial for further advancements in the sparsity testing and the related literature. We note that the statistic $\|\mathbb{X}'Y\|_2^2$ is central in both the results by Dicker (2014) and Carpentier and Verzelen (2019). While the authors derived the relevant results through random matrix theory and Wishart distributions, we approach the problem through the use of variance-gamma distribution due to its attractive properties. Firstly, in addition to the derivation of the asymptotic distribution, such approach allows us to specify the exact distribution of the statistic. Secondly, our approach can be easily extended towards other types of norms, instead of ℓ_2-norm. Finally, we assume approximate sparsity of the underlying data generating process and perform a Monte Carlo simulation study, which demonstrates that the statistic approaches the limiting distribution fairly quickly even under high variable multi-correlation and relatively small number of observations, suggesting possible further applications to the construction of statistical testing procedures.

Aims and problems

The aim of this dissertation is to analyse the sparsity assumption and sparse methods in the high-dimensional linear regression framework. Our objectives are the following:

- analyse the background assumptions and implications behind important sparse modelling techniques, with greater focus on the LASSO and its popular modifications; inspect and compare the performance of the sparse methods with dense alternatives by conducting a pseudo-real-time short-term nowcasting experiment simulating the performance that would be observed in practice; examine the insights from recovered sparse structures within the analysed datasets; introduce a modification of the LASSO by combining variants of LASSO with principal components methodology;
- introduce a test statistic $\|\mathbb{X}'Y\|_2^2$ and derive its exact and asymptotic distribution; specify conditions that allow simplifying certain necessary calculations; demonstrate the accuracy of the results and possible shortcomings under the assumption of approximate sparsity in a Monte Carlo simulation study.

Novelty

We propose a combination of LASSO and principal components methods, which combine the sparse and dense structure approaches and show promising results when forecasting US and EU GDP expenditure components. Compared with dense methods, the proposed combination in some cases is able to match or outperform the well known and widely used dense models, while doing so at a significantly smaller computational complexity and with greater short-term forecasting control. In certain cases the proposed combination was able to generate similar or more accurate forecasts when compared with the best of sparse methods.

We derive exact and asymptotic distributions of the statistic $\|\mathbb{X}'Y\|_2^2$. The asymptotic results do not impose strong restrictions on the model parameters β_j, we only assume that $\|\beta\|_2 < \infty$. On the other hand, the related literature relies on exact sparsity or certain approximations to hold.

For the variable covariance structure, instead of the general covariance matrix that deems to be estimable for practical purposes, we consider a flexible KMS structure, which, when carefully chosen, is able to well approximate a wide array of possible covariance structures.

We highlight the application of variance-gamma distribution used to derive the aforementioned distributions. First, we contribute to the literature of variance-gamma by extending certain known results. The variance-gamma distribution is demonstrated to have very attractive properties when working with products of Gaussian random variables. Second, along with the asymptotic distribution, this approach allows us to formulate the exact distribution of the statistic for any number of observations, n, and number of variables, p, as a certain combination of Gaussian and gamma random variables. We deem that such a result is much easier to work with than directly considering characteristic or density functions of the statistic $\|\mathbb{X}'Y\|_2^2$.

By performing Monte Carlo simulation studies we find that under the assumption of approximate sparsity of the model parameter vector β, the resulting statistics $\|\mathbb{X}'Y\|_2^2$ fairly quickly approach the limiting distributions for various sets of initial parameters. We argue that these results demonstrate the applicability of the statistics in approximately sparse setting, hence, can be important in certain applications when considering testing for sparsity or constructing signal-to-noise ratio estimates.

Dissertation structure

This dissertation consists of three chapters, as well as Conclusions, Bibliography and Appendix chapters.

In Chapter 1 we present a short overview of sparse high-dimensional linear regression modelling. We present the existing definitions of sparsity and briefly review several popular sparse models, that are often found in various applications. We discuss the known shortcomings and modifications of the approaches and end the chapter by introducing the latest advances in the construction of tests of sparsity and their relation with the SNR literature.

In Chapter 2 we build on the results introduced in Chapter 1 and derive the exact and asymptotic distribution of a central statistic, found in the SNR and sparsity testing literature. The chapter consists of proofs to the main formulated theorems, together with certain important auxiliary results. We end the chapter by presenting a Monte Carlo simulation study, where we assume approximate sparsity of the underlying model and examine the resulting sample distributions of the statistics.

In Chapter 3 we conduct pseudo-real-time experiments in order to examine the short-term forecasting performance of the sparse models, previously established in Chapter 1. We compare the performance of both sparse and dense methods by forecasting GDP expenditure components in US and selected EU countries. In addition, we present a detailed overview and in-depth analysis of the performance gains brought by the proposed combination of LASSO and principal components.

Finally, in Appendix A we present technical lemmas and proofs of certain results, that were omitted from Chapter 2. In Appendix B we present additional details, tables and figures with results from the pseudo-real-time experiments, that were omitted from Chapter 3.

Dissemination

The results of this dissertation were presented in the following conferences:

1. *59-th conference of Lithuanian Mathematical Society*, Vilnius, Lithuania, 2018 June 18–19.

2. *12th International Vilnius Conference on Probability Theory and Mathematical Statistics and 2018 IMS Annual Meeting on Probability and Statistics (IMS-2018 Vilnius)*, Vilnius, Lithuania, 2018 July 2–6.

3. *20th IWH-CIREQ-GW Macroeconometric Workshop: Micro Data and Macro Questions*, hosted by IWH, Haale, Germany, 2019 October 29–30.

4. *13th International Conference of the ERCIM WG on Computational and Methodological Statistics, 14th International Conference on Computational and Financial Econometrics, Virtual Conference (CFE-CMS2020)*, London, United Kingdom, 2020 December 19–21.

5. *61-th conference of Lithuanian Mathematical Society*, Vilnius, Lithuania, 2020 December 4.

6. *7th International Conference on Time Series and Forecasting*, Gran Canaria, Spain (ITISE 2021), 2021 July 19–21.

7. *22nd European Young Statisticians Meeting (EYSM 2021)*, Athens, Greece (virtual presentation), 2021 September 6–10.

Publications

1. S. Jokubaitis, D. Celov and R. Leipus, Sparse structures with LASSO through principal components: Forecasting GDP components in the short-run. *International Journal of Forecasting* (2021), 37(2), 759–776, https://doi.org/10.1016/j.ijforecast.2020.09.005

2. S. Jokubaitis and R. Leipus, Asymptotic Normality in Linear Regression with Approximately Sparse Structure. *Mathematics* (2022), 10, no. 10: 1657. https://doi.org/10.3390/math10101657

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my supervisor Prof. Remigijus Leipus for all the guidance, patience and feedback during the studies. I would also like to thank Assoc. Prof. Dmitrij Celov for his invaluable insights and support along the way. I am also grateful to Prof. Jonas Šiaulys and Assoc. Prof. Jurgita Markevičiūtė for careful reading of the manuscript and their valuable comments and suggestions. Finally, I would like to thank my family for their support and encouragement.

Chapter 1

Sparsity in high-dimensional linear regression

In this chapter, we present a short overview of sparse high-dimensional linear regression modelling. In Section 1.1 we introduce the necessary notations, used throughout the chapter. In Section 1.2 we discuss the definition of approximate sparsity, along with some possible extensions and less restrictive alternative specifications. In Section 1.3 we introduce and briefly review several popular sparse models, often found in applications. We end the chapter with Section 1.4 by establishing the latest advances in the construction of tests of sparsity and their relation with the SNR literature.

1.1 Notations

Throughout this chapter we assume that $Y := (y_1, \ldots, y_n)' \in \mathbb{R}^{n\times 1}$ are n observations of outcome and $\mathbb{X} = (X_1, \ldots, X_n)' \in \mathbb{R}^{n\times p}$ are p-dimensional predictors with $X_1, \ldots, X_n$ being $p\times 1$ random vectors, $X_i = (X_{1,i}, \ldots, X_{p,i})'$, $p > n$. We model

$$Y = \mathbb{X}\beta + \varepsilon, \tag{1.1}$$

where $\beta = (\beta_1, \ldots, \beta_p)'$ and $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)'$ is the error term. Recall, $||v||_q := (\sum_{j=1}^d |v_j|^q)^{1/q}$ denotes the ℓ_q-norm, $q \geq 1$, and $||v||_0 := \sum_{j=1}^d \mathbb{1}_{\{v_j \neq 0\}}$ for $q = 0$, where $v \in \mathbb{R}^d$ for some $d \in \mathbb{N}_+$; I_n denotes $n \times n$ identity matrix.

1.2 Approximate sparsity

Sparsity refers to the condition that only $s \ll n$ elements of β are nonzero, but allows the identities of these elements to be unknown. Sparsity can be motivated on economic grounds in situations where a researcher believes that the economic outcome could be well-predicted by a small (relative to sample size) number of factors but is unsure about the identity of the relevant factors. However, it is unrealistic to presume exact sparsity – typically it suffices to assume that all regressors potentially have a non-zero contribution to the regression function, but no more than s of unknown regressors are needed for approximating the regression function with a sufficient degree of accuracy.

In this dissertation we assume approximate sparsity. We follow the definition by Belloni and Chernozhukov (2011), namely, we assume that the true β parameters have a form:

$$|\beta_j| \quad = \quad Aj^{-\alpha}, \quad j = 1, \ldots, p, \quad \alpha \geq 1, \tag{1.2}$$

where the ordering and $A \in \mathbb{R}$ is arbitrary. We require that $\alpha \geq 1$ in order to achieve asymptotic stability, however, depending on the structure of Σ, this could be potentially relaxed. Belloni and Chernozhukov (2011) argue, that for approximately sparse models, the regression function can be well approximated by a linear combination of relatively few important regressors, which is the reason of wide popularity of variable selection approaches such as LASSO and its various modifications (see Section 1.3 for a detailed review). At the same time, approximate sparsity allows all coefficients to be nonzero, which is a more plausible assumption in many real world settings.

According to Bradic et al. (2021), the assumption of approximate sparsity can be motivated similarly to Newey (1997), who assumes that the first $s = s_n$ series terms can approximate the nonparametric regression function well. We do not impose that the most important $s = s_n$ terms are the first s terms, since the identity of the most important terms is typically unknown (Belloni et al. (2011)). Therefore, series approximations are usually based on relatively few regressors, often many fewer than the sample size. In contrast, approximate sparsity allows for very many potential regressors (possibly many more than sample size), when relatively few important regressors give a good approximation, but the identity of those few is not known. Approximately sparse and series approximations are similar in that they both depend on a few regressors giving a good approximation. However, they differ in that series regression requires that the identity of the important regressors is known, while with approximate sparsity their identity need not be known. This difference is useful in high dimensional settings, where potentially very many regressors are needed to approximate a function of many variables. Typically, economics and statistics provide little guidance about which regressors are important. With approximate sparsity, such information is not needed, since very many terms can be included among the potential regressors.

Alternatively, Ing (2020) and Cha et al. (2021) find that in certain cases the assumption of approximate sparsity can be relaxed. They show that with the use of *orthogonal greedy algorithm* in conjunction with *high-dimensional Akaike's Information criterion* (OGA+HDAIC) it is possible to achieve optimal convergence rates under weaker sparsity assumptions than (1.2). The authors discern approximate sparsity into polynomial and exponential decay, as follows:

$$L j^{-\gamma} \;\leq\; |\beta_{(j)}| \;\leq\; U j^{-\gamma}, \tag{1.3}$$

and

$$L_1 \exp(-\beta j) \;\leq\; |\beta_{(j)}| \;\leq\; U_1 \exp(-\beta j), \tag{1.4}$$

where $0 < L \leq U < \infty$ and $0 < L_1 \leq U_1 < \infty$ and $|\beta_{(1)}| \geq \ldots \geq |\beta_{(p)}|$ is a rearrangement of model parameters in the decreasing order. In addition, it seems that for certain algorithms these conditions can be relaxed further. For instance, the assumption of polynomial decay (1.3) is relaxed by requiring that there exists $0 < M < \infty$ such that $\sum_{j=1}^{p} \beta_j^2 < M$; and that there exists $\gamma \geq 1$ and $0 < C_\gamma < \infty$, such, that for every $J \subseteq \mathcal{P} \equiv \{1, \ldots, p\}$, the following inequality holds:

$$\sum_{j \in J} |\beta_j| \;\;\leq\;\; C_\gamma \Big(\sum_{j \in J} \beta_j^2 \Big)^{(\gamma-1)/(2\gamma-1)}, \tag{1.5}$$

where γ controls the sparsity of the model parameters.

In a similar manner, the exponential decay (1.4) can be relaxed further by assuming that there exists $0 < M_0 < \infty$ such that $\max|\beta_j| \leq M_0$ and that there exists $M_1 > 1$, such that

$$\sum_{j \in J} |\beta_j| \quad \leq \quad M_1 \max|\beta_j|. \tag{1.6}$$

The authors differentiate between the two types of sparsity and argue that it is difficult to attain optimal convergence rates without knowing the degree of sparseness, thus they consider a data-driven approach in order to determine the suitable number of iterations of the necessary algorithms in order to generalize. Ing (2020) and Cha et al. (2021) find that their proposed methods perform adequately under the assumptions (1.5) or (1.6). Furthermore, they show that under the case of exact sparsity, LASSO and OGA+HDAIC share the same error rates, but the latter is more favourable for various time-series applications since they do not require the independence between X_t and ε_t, $t = 1, \ldots, n$. Importantly, Wu and Wu (2016) note that LASSO can also be extended for the aforementioned time series applications, albeit with higher error rates.

However, throughout this dissertation we focus on the approximate sparsity assumption, as defined by (1.2), due to its widest applicability throughout the most popular sparse methods and the fact that it also plays an important role in time series modelling (see, e.g., Shibata (1980) and Ing (2007) for model selection problems in autoregressive time series models).

1.3 Short overview of the LASSO methods

1.3.1 LASSO

The LASSO method (Tibshirani (1996)) is one of the possible solutions dealing with $p \gg n$ problem, where the ordinary least squares (OLS) method is infeasible due to the 'curse-of-dimensionality', when the number of estimated parameters is much larger than the sample size. LASSO attracts a lot of attention in the literature due to its favourable strictly convex optimization problem, the solution path of which can be effectively estimated by using the Least angle regression (LARS) algorithm (Efron et al. (2004)).

LASSO is an ℓ_1-norm penalized least squares algorithm that solves the following optimization problem:

$$\hat{\beta}_{\text{LASSO}} = \arg\min_{\beta} \, (Y - \mathbb{X}\beta)'(Y - \mathbb{X}\beta) + \lambda||\beta||_1, \tag{1.7}$$

where the hyperparameter $\lambda \in (0, \infty)$ is fixed. With the value of λ tending to ∞, the estimated coefficients are shrunken towards zero, and with a sufficiently large value of λ some of them are estimated as 0 due to the properties of the ℓ_1-norm retaining the most significant features. Besides, there exists such λ, starting from which none of the variables are included in the model as significant. However, decreasing the value of λ the significant variables are sequentially included to the modelled regression. That is, in relation to the hyperparameter λ, LASSO procedure can be interpreted as a stepwise regression with an additional shrinkage of the estimated model's parameters towards zero. Also, due to the shrinkage of the estimated coefficients it is often possible to increase the accuracy of the forecasts, since the shrunken coefficients are able to reduce the variance of the forecasts by more than increasing the bias, implying the solution to bias-variance trade-off.

A lot of attention in the literature is given particularly to the variable selection aspect of the LASSO. Zhao and Yu (2006) and Zou (2006) proposed an almost necessary and sufficient condition – the Irrepresentable Condition – which ensures asymptotically consistent variable selection of the LASSO. The authors have shown that 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, the cited authors prove that the consistency of the variable selection by the LASSO requires that the value of λ should grow at a faster pace than $\sqrt{n}$. However, Knight and Fu (2000) show that the LASSO estimator $\hat{\beta}_{\text{LASSO}}$ is $\sqrt{n}$-consistent only under given $\lambda = \lambda_n = \mathcal{O}(\sqrt{n})$ and under certain regularity conditions. Therefore, we cannot fully expect both consistent variable selection and parameter estimation.

Let $\mathcal{A} := \{j : \beta_j \neq 0\}$ denote the true data generating process, where $|\mathcal{A}| = p_0 < n$ and $\hat{\beta}(\delta)$ is a parameter estimator of a certain procedure δ. Then, by Zou (2006), the procedure δ is said to have Oracle Properties if for the estimator $\hat{\beta}(\delta)$ the following two asymptotic conditions hold:

(O1) let $\mathcal{A}_\delta := \{j : \hat{\beta}_j(\delta) \neq 0\}$, then $\mathcal{A}_\delta = \mathcal{A}$; i.e., the true subset of variables is identified;

(O2) $\sqrt{n}\big(\hat{\beta}(\delta)_\mathcal{A} - \beta_\mathcal{A}\big) \overset{d}{\to} \mathcal{N}(0, \Sigma^*)$, where Σ^* is the covariance matrix knowing the true subset model.

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

1.3.2 Adaptive LASSO

In the literature many modifications of the LASSO can be found aiming to overcome various shortcomings of the original method. One of the most popular is the Adaptive LASSO, allowing to define weights for each individual explanatory variable used in the model:

$$\hat{\beta}_{\text{adaLASSO}} = \arg\min_{\beta} (Y - \mathbb{X}\beta)'(Y - \mathbb{X}\beta) + \lambda w'|\beta|, \tag{1.8}$$

where $|\beta| := (|\beta_1|, \ldots, |\beta_p|)'$ and $w = (w_1, \ldots, w_p)'$ is a vector of fixed weights. Zou (2006) proves that when the weight vector w is data-driven and appropriately chosen, the Adaptive LASSO is able to achieve the Oracle Properties. In the literature often weights are suggested to be chosen as $\hat{w}_j := 1/|\hat{\beta}^*_j|^\gamma$, $j = 1, \ldots, p$, $\gamma > 0$, where $\hat{\beta}^*$ is $\sqrt{n}$-consistent estimator of β. When $p \leq n$, a natural choice could be $\hat{\beta}^* := \hat{\beta}_{\text{OLS}}$. However, when the OLS estimates are infeasible ($p > n$) or the data is strongly multicollinear, it is suggested to replace the $\hat{\beta}_{\text{OLS}}$ with $\hat{\beta}_{\text{ridge}}$, where coefficients are estimated by the Ridge regression, defined by a problem similar to LASSO (1.7), with the ℓ_1-norm of the imposed penalty replaced by the ℓ_2-norm. It is argued that with the sample size increasing, the weights of the insignificant variables should tend to infinity, while the weights of the significant variables should converge to some finite non-zero constants. This way, the method allows for an asymptotically unbiased simultaneous estimation of large coefficient and small threshold estimates.

The importance of the appropriate choice of the weight vector is also stressed in Huang et al. (2008), since under high multicollinearity or due to a large amount of noise variables the usage of Ridge or univariate OLS may lead to overall weaker results than the original LASSO. Medeiros and Mendes (2016) claim that in the case of $p \gg n$ 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 non-zero finite values for the significant variables. These requirements, according to the authors, under certain regularity conditions, should be satisfied by the ordinary LASSO or the Elastic Net estimators. Therefore, these methods can be used for the selection of optimal weights. In fact, following these requirements, Liu (2014) successfully extended the weight estimation to adjust for auto-regressive processes, and Konzen and Ziegelmann (2016) adjust the method for lagged effects.

Most importantly, Medeiros and Mendes (2016) show 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 – the features that are often observed in financial and macroeconomic data. Also, the proposed method allows the number of variables, both in the search space and retained in the final set $\mathcal{A}$, to grow together with the sample size 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, when working with macroeconomic 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 $\mathbb{X}$, the resulting growth rate of the dimension of the full dataset is only linear with respect to the size of the sample, but not polynomial. This suggests some degree of freedom to additionally include non-linearities through, e.g., interactions between the variables or their power transformations. Also, promising results were obtained in the simulation studies when analysing the forecasting performance of models with heavy-tailed residuals having GARCH structures, by using strongly correlated regressors as the explanatory variables (Medeiros and Mendes (2016)).

Finally, Liu (2014) observes that the Adaptive LASSO can be effectively performed by employing the LARS algorithm: let $\hat{W} := \text{diag}(\hat{w}_1, \ldots, \hat{w}_p)$, then the optimization problem of the Adaptive LASSO (1.8) can be rewritten as:

$$\begin{aligned} \hat{\beta}_{\text{adaLASSO}} &= \arg\min_{\beta} (Y - \mathbb{X}\hat{W}^{-1}\hat{W}\beta)'(Y - \mathbb{X}\hat{W}^{-1}\hat{W}\beta) + \lambda||\hat{W}\beta||_1 \\ &= \arg\min_{\tilde{\beta}} (Y - \tilde{\mathbb{X}}\tilde{\beta})'(Y - \tilde{\mathbb{X}}\tilde{\beta}) + \lambda||\tilde{\beta}||_1, \qquad (1.9) \end{aligned}$$

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

1.3.3 Relaxed LASSO

Another popular modification of the LASSO, dealing with some of its shortcomings, is the Relaxed LASSO (Meinshausen (2007)). The main idea of this method is to separate the selection of the significant variables and the estimation of the model's coefficients by introducing an additional hyperparameter ϕ. Denote,

$$\mathcal{M}_\lambda := \{1 \leq j \leq p \ : \ \hat{\beta}_j^\lambda \neq 0\}, \qquad (1.10)$$

which corresponds with the active set of variables, preselected by the LASSO method under a certain fixed value of λ. Then the Relaxed LASSO is estimated as:

$$\hat{\beta}_{\text{RL}} = \arg\min_{\beta} n^{-1}\big(Y - \mathbb{X}\{\beta \cdot \mathbb{1}_{\mathcal{M}_\lambda}\}\big)'\big(Y - \mathbb{X}\{\beta \cdot \mathbb{1}_{\mathcal{M}_\lambda}\}\big) + \phi\lambda||\beta||_1, \qquad (1.11)$$

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 $\mathcal{M}_\lambda \subset \{1, \ldots, p\}$:

$$\{\beta \cdot \mathbb{1}_{\mathcal{M}_\lambda}\}_k = \begin{cases} 0, & k \notin \mathcal{M}_\lambda, \\ \beta_k, & k \in \mathcal{M}_\lambda, \end{cases} \qquad (1.12)$$

for every $k \in \{1, \ldots, p\}$. In this way, the selection of significant variables is performed by using the 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 LASSO, that is, no correction of the estimated coefficients is performed, while when ϕ approaches zero the estimator converges to Post-LASSO estimator (Belloni and Chernozhukov (2013)).

Meinshausen (2007) prove that due to such separation of the variable selection, the consistent estimates of the model's coefficients are obtained with the usual $\sqrt{n}$ rate of convergence, independently from the growth rate of the available information set. In addition, the authors demonstrate that the Relaxed LASSO is able to outperform the LASSO, the extent of which is highly dependent on the signal-to-noise ratio. Larger improvements are expected when the signal is strong and the shrinkage on the selected components is not necessary.

The authors expect that for most sparse high-dimensional problems the estimator should lead to sparser estimators than LASSO and generate more accurate predictions at almost no extra computational cost.

1.3.4 Square-Root LASSO

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 (1.7) by taking the square root of the residual sum of squares term, as defined by the equation (1.13):

$$\hat{\beta}_{\text{sqrtLASSO}} = \arg\min_{\beta} n^{-1/2} \left((Y - \mathbb{X}\beta)'(Y - \mathbb{X}\beta)\right)^{1/2} + \lambda ||\beta||_1. \quad (1.13)$$

The attractiveness of the method follows from the original idea for LASSO presented in Bickel et al. (2009), where rate-optimal value of $\lambda = 2\sigma\sqrt{2\log(pn)/n}$ depends on unknown value of σ. Belloni et al. (2011) show that for the Square-Root LASSO the rate-optimal penalty level reduces to $\lambda = \sqrt{2\log(pn)/n}$, which makes it having no user-specified parameters and therefore tuning free.

1.3.5 Fast Best Subset Selection

Recently, Bertsimas et al. (2016) presented a mixed integer optimization (MIO) formulation for the best-subset selection problem, achieving certifiable global optimality, which was further developed and extended by Dedieu et al. (2021), Mazumder et al. (2022). Recall, the main motivation behind the LASSO and similar methods, trying to solve the curse of dimensionality – the core argument was the fact that the best subset selection, while viewed as the "holy grail" of estimators for sparse modelling in regression, was essentially unfeasible in real world scenarios due to arising computational complexities (Natarajan (1995)). In fact, Natarajan (1995) state the problem as being NP-hard. In this sense, modelling approaches that followed are sometimes seen as certain approximations or heuristics of the best subset selection, used mainly out of the necessity when the best subset selection was not computable. However, research by Hastie et al. (2017) demonstrate that this is not the case. The authors argue that different modelling procedures have different operating characteristics, i.e., give rise to different bias-variance trade-offs as the tuning parameters vary. In fact, in certain settings the bias-variance trade-off provided by the best subset selection may be more or less useful than the trade-off provided by the LASSO.

These observations are followed further by Hazimeh and Mazumder (2020), who propose a Fast Best Subset Selection algorithm by making use of the MIO and related algorithms, making the necessary computations feasible under certain heuristics. The authors propose the following optimization problem:

$$\hat{\beta}_{\text{L0Lq}} \quad = \quad \arg\min_{\beta} \frac{1}{2}(Y - \mathbb{X}\beta)'(Y - \mathbb{X}\beta) + \lambda_0 ||\beta||_0 + \lambda_q ||\beta||_q^q, \quad (1.14)$$

where $q \in \{1, 2\}$ determines the type of additional regularization. Here the hyperparameter $\lambda_0 \in [0, \infty)$ controls the number of variables selected, while the $\lambda_q \in [0, \infty)$ controls the shrinkage imposed on the estimates. In this dissertation, by setting $q = 0$, denoted by L0, we assume that (1.14) is reduced to the problem of best-subset selection. The goal of the approach is to regularize the overfitting behaviour of the best-subsets and obtain sparse models with good predictive power.

Importantly, Mazumder et al. (2022) argue, that the success of the performance of the best subset selection largely depends on the signal-to-noise ratio (SNR). If the noise level is relatively small, the best-subsets estimator leads to models with excellent statistical properties in terms of prediction, estimation and variable selection (see, e.g., Bühlmann and Van De Geer (2011), Zhang and Zhang (2012)). However, by Breiman (1996b), the situation is different when the noise level is high, when high instability of the best subset selection can be observed. In other words, in high-noise regimes the predictive performance of best-subsets deteriorates very quickly and thus is a significant drawback of the approach. Noteworthy, SNR alone does not control the difficulty of the underlying statistical problem; model parameters p, n, $\mathbb{X}$ and the underlying sparsity of the true model also affect the performance of the estimator.

Furthermore, Mazumder et al. (2022) demonstrate that in high noise-level settings the best-subsets estimator is not the right approach, leading to overfitting and poor predictive performance, which quickly deteriorates as SNR decreases. In those cases the method is outperformed by LASSO or the Ridge regression. The authors argue that the overfitting behaviour of best-subsets can be attributed to its aggressive search for the best feature subset and not performing any shrinkage on the selected coefficients. They conclude that the classical best-subsets estimator is not designed to be used in high-noise regimes.

By proposing the L0Lq approach, Mazumder et al. (2022) focus on the predictive performance instead of the variable selection aspect. The aim is to obtain a sparse linear model with predictive performance better than best-subsets and comparable with Ridge regression or the LASSO. The goal is an estimator with fewer non-zero coefficients than the LASSO, but of similar or greater predictive accuracy.

Depending on the choice of q, the estimator by (1.14) can be contrasted with the LASSO or Ridge regression, since the estimator (1.14) separates out the effects of shrinkage (via $\lambda_q||\beta||_q^q$) and sparsity (via $\lambda_0||\beta||_0$). The authors argue, that key benefits of such approach can be observed when the SNR is low – in such case, the shrinkage imparted via ℓ_q-regularization becomes critical, since the estimator (1.14) then prefers to optimize for positive λ_q values to produce a good predictive model. Further sparsification may occur when ℓ_1 penalty is chosen, since for cases when λ_0 would choose a large number of parameters, penalization from the ℓ_1 would act as a counter.

By the results from simulations and experiments on real datasets, Mazumder et al. (2022) expect superior predictive accuracy of L0L1 over L0L2 for larger SNR values. In addition, the authors observe that when the noise level is high, L0 performs poorly in terms of prediction accuracy. To mitigate its overfitting behaviour, L0 attempts to regularize by selecting very sparse models – the best predictive model for L0 has fewer non-zeros than the Oracle solution.

Moreover, based on the simulation results, the cited authors expect the L0L1 and L0L2 methods to perform similarly or better than the LASSO or Ridge regression. Importantly, they observe that the proposed approaches lead to estimators that are significantly sparser, while showing similar or better predictive performance when compared with LASSO and similar alternatives. In addition, they find L0 models to generate the most sparse results when compared with L0L1 and L0L2, however, L0 suffers in terms of the prediction accuracy. Noteworthy, as SNR increases, L0L1 and L0L2 is expected to become more similar to L0 in terms of both sparsity and the prediction accuracy.

1.4 Signal-to-Noise Ratio and Tests for Sparsity

The importance of SNR levels was highlighted in several places in the preceding section. Besides the assumption of underlying sparsity, the amount of noise present in the model can be a crucial aspect, determining the optimal choice and performance of the sparse high-dimensional linear model. By Mazumder (2020), when dealing with real world data and problems, a researcher could expect high levels of signal in, e.g., marketing and retail applications, compressed sensing, image processing and classification problems, while, on the other hand, high levels of noise should be expected in financial applications. The key takeaway is, even if a wide range of SNR values may occur in practice, most of the time practitioners may not know *a priori* what the SNR is for the application at hand. Thus, it is often useful to have a suite of tools that are applicable for both high and low SNR regimes.

However, the question of SNR estimation is much broader and not limited to sparse linear regressions. It is often of great importance in applications where the aim is to quantify to what extent the covariates explain the variation of the response variable (see, for instance, the discussion by Verzelen and Gassiat (2018), who focus on heritability estimation in genome-wide association studies, where sparsity-based methods could be questionable).

Denote SNR by:

$$\eta := \frac{\beta'\Sigma\beta}{\sigma_\varepsilon^2}, \tag{1.15}$$

where β denotes the parameter vector of the linear model, Σ denotes the covariance matrix of the covariates $\mathbb{X}$, and σ_ε^2 denotes the error variance.

In the literature two main lines of research can be discerned for estimating σ_ε^2 or η in a high-dimensional setting. Estimating the signal strength is, up to a parametric loss, equivalent to estimating the proportion of explained variation η, the quadratic function $\beta'\Sigma\beta$ or the noise level σ_ε^2. For instance, for the estimation of σ_ε^2, see Reid et al. (2016) for a thorough review, as well as Janson et al. (2017) for recent advances, the results of which strongly rely on the assumption of sparsity. Notably, Verzelen and Gassiat (2018) build on the results by Dicker (2014) and Dicker (2016) and consider the following U-type statistic:

$$V = \frac{Y'(\mathbb{X}\mathbb{X}' - \operatorname{tr}(\mathbb{X}\mathbb{X}')I_n/n)Y}{\|Y\|_2^2(n+1)}, \tag{1.16}$$

which is demonstrated to concentrate exponentially fast around $\beta'\Sigma^2\beta/\operatorname{Var}(y_1)$. Since the accuracy of their results largely depends on the level of sparsity of the true model parameters β, the authors combine their proposed statistics with the estimates by Square-Root LASSO whenever the data allows it.

Furthermore, Verzelen and Gassiat (2018) note than whenever β is dense, many previously considered approaches fail. Thus, the aforementioned results are extended further by Carpentier and Verzelen (2019) and Carpentier and Verzelen (2021) who use this observation and adapt (1.16) for testing of exact sparsity of the parameter vector β. The latter result is a very interesting and valuable take, since currently the literature on testing of exact or approximate sparsity in a high-dimensional linear regression setting is very scarce (see the review by Carpentier and Verzelen (2021)). We argue that these results opens up an interesting avenue for further research in both the related SNR and testing for sparsity literature.

Chapter 2

Asymptotic normality in linear regression with approximately sparse structure

In this chapter, we derive the exact and asymptotic distributions of the statistic $\|\mathbb{X}'Y\|_2^2$. First, we motivate the reasoning and the importance behind the derived results, and introduce the necessary assumptions and notations used in order to formulate the main results, which are presented in Section 2.1. Next, in Section 2.2 we present a brief overview of convenient properties of the variance-gamma distribution which are crucial to certain proofs of the main results. In Section 2.3 we establish auxiliary lemmas, used throughout the proofs of the main results. The proofs are presented in Section 2.4, in addition, we formulate convenient expressions for certain important terms, which are used in Section 2.5, where we assume approximate sparsity of the underlying model parameters and perform a Monte Carlo simulation study.

Throughout this chapter we consider a linear regression model by (1.1). In addition, we assume that $\mathbb{X} = (X_1, \ldots, X_n)' \in \mathbb{R}^{n \times p}$ are p-dimensional predictors with $X_1, \ldots, X_n$ being i.i.d. $p \times 1$ random vectors $X_i = (X_{1,i}, \ldots, X_{p,i})'$, which are normally distributed with zero mean and covariance matrix Σ, denoted $X_i \stackrel{d}{=} \mathcal{N}_p(0, \Sigma)$, $i = 1, \ldots, n$. Furthermore, we assume that the covariance matrix Σ has a form

$$\Sigma = (\varrho^{|i-j|})_{i,j=1}^{p} = \begin{bmatrix} 1 & \varrho & \dots & \varrho^{p-1} \\ \varrho & 1 & \dots & \varrho^{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ \varrho^{p-1} & \varrho^{p-2} & \dots & 1 \end{bmatrix}, \qquad (2.1)$$

if $0 < |\varrho| < 1$, and $\Sigma = I_p$ if $\varrho = 0$ (here and below I_p denotes the $p \times p$ identity matrix). This matrix is often called the Kac–Murdock–Szegő (KMS) matrix, originally introduced in Kac et al. (1953). As the auto-correlation matrix of corresponding causal AR(1) processes, KMS matrix is positive definite and is considered due to the wide array of applications in the literature and its' well known spectral properties (see, e.g., Fikioris (2018) for a thorough literature review). When carefully chosen, such a structure could well-approximate a wide array of possible covariance structures (see, e.g., Yang et al. (2021) for a more general approach with various Toeplitz covariance structures). Furthermore, $\varepsilon := (\varepsilon_1, \dots, \varepsilon_n)' \in \mathbb{R}^{n\times 1} \stackrel{d}{=} \mathcal{N}(0, \sigma_\varepsilon^2 I_n)$ are unobserved i.i.d. errors with $\mathbb{E}\varepsilon_i = 0$, $\mathrm{Var}(\varepsilon_i) = \sigma_\varepsilon^2 > 0$. In practice, the assumption that $\mathbb{E}X_i = 0$ can be untenable and it may be appropriate to add an intercept to the linear model (1.1), however, for simplicity, throughout this chapter we will assume that the intercept is known and the variables are centered. Similar settings are considered when dealing with certain geospatial data, longitudinal studies, microarray data, research on approximate message passing algorithms (see, e.g., Liang and Zeger (1986), Rangan (2011), Vila and Schniter (2013), Dicker (2014), Diggle and Giorgi (2019), Patil and Kim (2020)).

This chapter is concerned with the derivation of the exact and asymptotic distributions for the suitably centered and normalized squared norm $\|\mathbb{X}'Y\|_2^2$ under the assumption of the KMS type covariance structure in (2.1), where p and n are assumed large. Throughout the chapter we assume that $p, n \to \infty$ and $p/n \to c \in (0, \infty)$. We are particularly interested in cases where $p > n$. Statistics of such form arise in various applications in the context of high-dimensional linear regression. One notable example is presented by (1.16), general moments of which are derived through random matrix theory and Wishart distributions (see, e.g., Dicker (2014)). Dealing with such statistics typically require strong restrictions on the model parameters β, however, in this chapter we only require that $\|\beta\|_2^2 < \infty$ is satisfied. Moreover, our results could be extended by using β-generating functions (e.g., parameters of FARIMA models). In comparison to the related papers, Carpentier and Verzelen (2021) assume exact sparsity, while Dicker (2014) require certain approximations to hold.

We approach the problem following an observation by Gaunt (2013) that the distribution of product of Gaussian random variables admits a variance-gamma distribution, which results in a set of attractive properties. We contribute to the literature on variance-gamma distribution by extending the results of Gaunt (2014, 2018, 2019). We demonstrate that, along with the derivation of the asymptotic distribution of $\|\mathbb{X}'Y\|_2^2$, this approach allows us to define the exact distribution of the statistic given any fixed values p, n, which can be expressed through a combination of gamma and normal random variables. In the related literature we were not able to find results for the exact distribution and asymptotic analysis of the statistic $\|\mathbb{X}'Y\|_2^2$ based on the variance-gamma distribution. Furthermore, we deem that such a result is much easier to work with than when considering the characteristic or density functions of $\|\mathbb{X}'Y\|_2^2$ directly. Therefore, in addition to the ℓ_2-norm statistic, we argue that the obtained results can be extended towards alternative forms of the statistic, e.g., by using a different norm, which would reduce the problem to manipulating variance-gamma distribution, thus suggesting possible further research cases and useful extensions.

Additionally, we examine a specific case of parameter β by considering $\beta_j = j^{-1}$, $j \geq 1$. Similar structures of the vector β are often found in the literature when approximate sparsity of the coefficients in the linear regression model (1.1) is assumed. See the discussion in Chapter 1 for a thorough review of wide applications and the importance of such assumption. Furthermore, by performing Monte Carlo simulations, we find that the empirical distributions of the corresponding statistic approach the limiting distribution reasonably quickly even for large values of ϱ and c. Such observations suggest that the assumption of sparse structure can be included in the applications and statistical tests, thus, the results of this chapter could be further extended following the literature on testing for sparsity or construction of signal-to-noise ratio estimators (see, e.g., Dicker (2014), Dicker and Erdogdu (2016), Carpentier and Verzelen

(2019), Carpentier and Verzelen (2021)).

In this chapter, $\stackrel{d}{=}$, $\stackrel{d}{\to}$ and $\stackrel{\mathbb{P}}{\to}$ denote the equality of distributions, convergence of distributions and convergence in probability, respectively. C stands for a generic positive constant which may assume different values at various locations. $\mathbb{1}_A$ denotes the indicator function of a set A.

2.1 Main results

In this section, we formulate the main results on the normality of statistic $\|\mathbb{X}'Y\|_2^2$. Introduce the notations:

$$\kappa_{1,p} := \sum_{k=1}^{p}\sum_{l=1}^{p}\beta_k\beta_l\varrho^{|k-l|} = \beta'\Sigma\beta, \tag{2.2}$$

$$\kappa_{2,p} := \sum_{k=1}^{p}\Big(\sum_{l=1}^{p}\beta_l\varrho^{|k-l|}\Big)^2 = \beta'\Sigma^2\beta, \tag{2.3}$$

$$\kappa_{3,p} := \sum_{k,l,j,j'=1}^{p}\beta_j\beta_{j'}\varrho^{|k-j|}\varrho^{|l-j'|}\varrho^{|k-l|} = \beta'\Sigma^3\beta. \tag{2.4}$$

It is easy to see that, under $\sum_{j=1}^{\infty}\beta_j^2 < \infty$, there exist limits

$$\kappa_i = \lim_{p\to\infty}\kappa_{i,p}, \; i = 1, 2, 3.$$

Obviously, $\kappa_{2,p} \geq 0$. Moreover, since $(\varrho^{|i-j|})_{i,j=1}^{p}$ is positive semi-definite, $\kappa_{i,p} \geq 0$, $i = 1, 3$. Indeed, $\sum_{k,l=1}^{p}\varrho^{|k-l|}a_k a_l \geq 0$, thus it suffices to take $a_k = \beta_k$ for $i = 1$ and $a_k = \sum_{j=1}^{p}\beta_j\varrho^{|k-j|}$ for $i = 3$.

Our first main result is the following theorem.

Theorem 2.1. *Assume the model in* (1.1) *with covariance structure in* (2.1)*. Let* $n \to \infty$ *and let* $p = p_n$ *satisfy*

$$p \to \infty, \qquad \frac{p}{n} \to c \in (0, \infty). \tag{2.5}$$

Let also the β_j *satisfy*

$$\sum_{j=1}^{\infty}\beta_j^2 < \infty. \tag{2.6}$$

Then

$$\frac{\|\mathbb{X}'Y\|_2^2 - n^2\kappa_{2,p} - pn(\kappa_{1,p}+\sigma_\varepsilon^2)}{n^{3/2}} \ \stackrel{d}{\to}\ \mathcal{N}(0,s^2), \qquad (2.7)$$

where variance s^2 *has the structure*

$$s^2 \ = \ 4\kappa_2^2 + 4(\kappa_1+\sigma_\varepsilon^2)\,(2\kappa_2 c+\kappa_3) + 2c(\kappa_1+\sigma_\varepsilon^2)^2\Big(c+\frac{1+\varrho^2}{1-\varrho^2}\Big). \quad (2.8)$$

Our second main result deals with the case where the centering sequence in (2.7) is modified to include the limiting values of $\kappa_{i,p}$, $i=1,2$.

Theorem 2.2. *Let the assumptions of Theorem 2.1 hold. In addition, assume that* $\sum_{j=p+1}^{\infty}\beta_j^2 = o(p^{-1/2})$ *and* $\sup_{j\ge 1}|\beta_j|j^\alpha<\infty$ *with* $\alpha>1/2$. *Then,*

$$\frac{\|\mathbb{X}'Y\|_2^2 - n^2\big(\kappa_2 + c(\kappa_1+\sigma_\varepsilon^2)\big)}{n^{3/2}} \ \stackrel{d}{\to}\ \mathcal{N}(0,s^2), \qquad (2.9)$$

where variance s^2 *has the structure by* (2.8).

The proofs of these theorems are given in Section 2.4.

Remark 2.1. For alternative expressions of κ_1, κ_2 and κ_3, see Lemma 2.5 below.

Define

$$\beta(x) \ := \ \sum_{j=1}^{\infty}\beta_j^2 x^j, \quad |x|\le 1.$$

The following corollary deals with the case when $\varrho=0$, i.e., $\Sigma=I_p$. The result easily follows from Theorem 2.2, noting that in this case $\kappa_i=\beta(1)$, $i=1,2,3$.

Corollary 2.1. *Assume a model* (1.1) *with covariance structure* $\Sigma=I_p$. *Let assumptions* (2.5) *and* (2.6) *be satisfied. In addition, assume that* $\sum_{j=p+1}^{\infty}\beta_j^2 = o(p^{-1/2})$ *and* $\sup_{j\ge 1}|\beta_j|j^\alpha<\infty$ *with* $\alpha>1/2$. *Then,*

$$\frac{\|\mathbb{X}'Y\|_2^2 - n^2(\beta(1)(1+c)+c\sigma_\varepsilon^2)}{n^{3/2}} \ \stackrel{d}{\to}\ \mathcal{N}(0,s^2), \qquad (2.10)$$

where

$$s^2 \ = \ 2\beta(1)^2\big(4+5c+c^2\big) + 4\beta(1)\sigma_\varepsilon^2\big(1+3c+c^2\big) + 2\sigma_\varepsilon^4(c+c^2). \ (2.11)$$

2.2 Properties of variance gamma distribution

In this section, we define and provide important properties of the variance-gamma distribution, which will be crucial in proving the main results of this chapter.

Recall that the variance-gamma distribution with parameters $r > 0$, $\theta \in \mathbb{R}$, $\sigma > 0$ and $\mu \in \mathbb{R}$ has density

$$\begin{aligned} f^{\mathrm{VG}}(x) &= \frac{1}{\sigma\sqrt{\pi}\Gamma(r/2)} \mathrm{e}^{\theta(x-\mu)/\sigma^2} \left(\frac{|x-\mu|}{2\sqrt{\theta^2+\sigma^2}} \right)^{(r-1)/2} \\ &\times K_{(r-1)/2}\left(\frac{\sqrt{\theta^2+\sigma^2}}{\sigma^2} |x-\mu| \right), \end{aligned} \tag{2.12}$$

where $x \in \mathbb{R}$, $K_\nu(x)$ is the modified Bessel function of the second kind. For a random variable Q with density (2.12) we write $Q \overset{d}{=} \mathrm{VG}(r, \theta, \sigma, \mu)$. Let $\Gamma(a, b)$, $a > 0$, $b > 0$, denote the gamma distribution with density

$$f^{\mathrm{G}}(x) = \frac{b^a}{\Gamma(a)} x^{a-1} \mathrm{e}^{-bx}, \quad x > 0.$$

It holds that

$$Q \overset{d}{=} \mu + \theta W_r + \sigma\sqrt{W_r} U, \tag{2.13}$$

where $W_r \overset{d}{=} \Gamma(r/2, 1/2)$, $U \overset{d}{=} \mathcal{N}(0,1)$, W_r and U are independent. The characteristic function of $Q \overset{d}{=} \mathrm{VG}(r, \theta, \sigma, \mu)$ has a form (see, e.g., Madan et al. (1998), Kotz et al. (2001))

$$\varphi_Q(t) = \frac{\mathrm{e}^{\mathrm{i}\mu t}}{(1 + \sigma^2 t - 2\mathrm{i}\theta t)^{r/2}}, \quad t \in \mathbb{R}. \tag{2.14}$$

We note the following properties of the variance-gamma distribution.

1. If $Q_1 \overset{d}{=} \mathrm{VG}(r_1, \theta, \sigma, \mu_1)$ and $Q_2 \overset{d}{=} \mathrm{VG}(r_2, \theta, \sigma, \mu_2)$ are independent random variables, then

$$Q_1 + Q_2 \overset{d}{=} \mathrm{VG}(r_1 + r_2, \theta, \sigma, \mu_1 + \mu_2).$$

2. If $Q \overset{d}{=} \mathrm{VG}(r, \theta, \sigma, \mu)$, then for any $a > 0$

$$aQ \overset{d}{=} \mathrm{VG}(r, a\theta, a\sigma, a\mu).$$

The following proposition is crucial for our purposes.

Proposition 2.1. (i) *If* $(\xi_1, \xi_2)' \stackrel{d}{=} \mathcal{N}_2(0, \Sigma)$, *where* $\Sigma = \begin{pmatrix} \sigma_1^2 & \varrho\sigma_1\sigma_2 \\ \varrho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$, *then*

$$\xi_1\xi_2 \stackrel{d}{=} \mathrm{VG}(1, \varrho\sigma_1\sigma_2, \sqrt{1-\varrho^2}\sigma_1\sigma_2, 0).$$

(ii) *If* $(\xi_{1j}, \xi_{2j})'$, $j = 1, \ldots, n$, *are i.i.d. random vectors with common distribution* $\mathcal{N}_2(0, \Sigma)$, *then*

$$\sum_{j=1}^{n} \xi_{1j}\xi_{2j} \stackrel{d}{=} \mathrm{VG}(n, \varrho\sigma_1\sigma_2, \sqrt{1-\varrho^2}\sigma_1\sigma_2, 0)$$

and

$$\sum_{j=1}^{n} \xi_{1j}\xi_{2j} \stackrel{d}{=} \sigma_1\sigma_2(\varrho W_n + \sqrt{1-\varrho^2}\sqrt{W_n}U),$$

where $W_n \stackrel{d}{=} \Gamma(n/2, 1/2)$ *and* $U \stackrel{d}{=} \mathcal{N}(0,1)$ *are independent random variables.*

(iii) *Assume that* $(\xi_{1j}^{(1)}, \ldots, \xi_{1j}^{(p)}, \xi_{2j})'$, $j = 1, \ldots, n$, *are i.i.d. copies of* $(\xi_1^{(1)}, \ldots, \xi_1^{(p)}, \xi_2)' \stackrel{d}{=} \mathcal{N}_{p+1}(0, \Sigma^{(p)})$ *and let* $\varrho^{(kl)} := \mathrm{Corr}(\xi_1^{(k)}, \xi_1^{(l)})$, $\varrho^{(k)} := \mathrm{Corr}(\xi_1^{(k)}, \xi_2)$, $(\sigma_1^{(k)})^2 := \mathrm{Var}(\xi_1^{(k)})$, $\sigma_2^2 := \mathrm{Var}(\xi_2)$, $k, l = 1, \ldots, p$. *Then*

$$\begin{pmatrix} \sum_{j=1}^n \xi_{1j}^{(1)}\xi_{2j} \\ \vdots \\ \sum_{j=1}^n \xi_{1j}^{(p)}\xi_{2j} \end{pmatrix} \stackrel{d}{=} \begin{pmatrix} \sigma_1^{(1)}\sigma_2(\varrho^{(1)}W_n + \sqrt{1-(\varrho^{(1)})^2}\sqrt{W_n}U_1) \\ \vdots \\ \sigma_1^{(p)}\sigma_2(\varrho^{(p)}W_n + \sqrt{1-(\varrho^{(p)})^2}\sqrt{W_n}U_p) \end{pmatrix},$$

where $(U_1, \ldots, U_p)' \stackrel{d}{=} \mathcal{N}_p(0, \Sigma_U)$, $\Sigma_U = (\sigma_U^{(kl)})$ *with*

$$\sigma_U^{(k,l)} = \mathbb{E}U_kU_l = \frac{\varrho^{(kl)} - \varrho^{(k)}\varrho^{(l)}}{\sqrt{1-(\varrho^{(k)})^2}\sqrt{1-(\varrho^{(l)})^2}}, \quad k, l = 1, \ldots, p. \tag{2.15}$$

Proof. The statements in (i), (ii) are well-known, see e.g. Gaunt (2019). The proof of part (iii) follows from Lemma 2.1. □

Lemma 2.1. *Assume that* $(\xi_1^{(1)}, \ldots, \xi_1^{(p)}, \xi_2)'$ *has distribution* $\mathcal{N}_{p+1}(0, \Sigma^{(p)})$ *and let* $\varrho^{(kl)} := \mathrm{Corr}(\xi_1^{(k)}, \xi_1^{(l)})$, $\varrho^{(k)} := \mathrm{Corr}(\xi_1^{(k)}, \xi_2)$,

$(\sigma_1^{(k)})^2 := \mathrm{Var}(\xi_1^{(k)})$, $\sigma_2^2 := \mathrm{Var}(\xi_2)$, $k, l = 1, \ldots, p$. *Then*

$$\begin{pmatrix} \xi_1^{(1)}\xi_2 \\ \vdots \\ \xi_1^{(p)}\xi_2 \end{pmatrix} \stackrel{d}{=} \begin{pmatrix} \sigma_1^{(1)}\sigma_2\big(\varrho^{(1)}W_1 + \sqrt{1-(\varrho^{(1)})^2}\sqrt{W_1}U_1\big) \\ \vdots \\ \sigma_1^{(p)}\sigma_2\big(\varrho^{(p)}W_1 + \sqrt{1-(\varrho^{(p)})^2}\sqrt{W_1}U_p\big) \end{pmatrix},$$

where $W_1 \stackrel{d}{=} \Gamma(1/2, 1/2)$, $(U_1, \ldots, U_p)'$ *is, independent of* W_1, *zero mean normal vector with covariances in* (2.15).

Proof. It suffices to prove that for any $(t_1, \ldots, t_p) \in \mathbb{R}^p$ it holds

$$\Big(\sum_{k=1}^p t_k\xi_1^{(k)}\Big)\xi_2 \stackrel{d}{=} \sigma_2\sum_{k=1}^p t_k\sigma_1^{(k)}\big(\varrho^{(k)}W_1 + \sqrt{1-(\varrho^{(k)})^2}\sqrt{W_1}U_k\big). \tag{2.16}$$

Since

$$\sum_{k=1}^p t_k\xi_1^{(k)} \stackrel{d}{=} \mathcal{N}\Big(0, \sum_{k,l=1}^p t_kt_l\varrho^{(kl)}\sigma_1^{(k)}\sigma_1^{(l)}\Big), \quad \xi_2 \stackrel{d}{=} \mathcal{N}(0, \sigma_2^2),$$

by Proposition 2.1(i) we obtain that

$$\begin{aligned}&\Big(\sum_{k=1}^p t_k\xi_1^{(k)}\Big)\xi_2 \\ &\stackrel{d}{=} \mathrm{VG}\Bigg(1, \sigma_2\sum_{k=1}^p t_k\varrho^{(k)}\sigma_1^{(k)}, \sigma_2\sqrt{\sum_{k,l=1}^p t_kt_l\sigma_1^{(k)}\sigma_1^{(l)}(\varrho^{(kl)} - \varrho^{(k)}\varrho^{(l)})}, 0\Bigg).\end{aligned} \tag{2.17}$$

For the right-hand side of (2.16) write

$$\begin{aligned}&\sigma_2\sum_{k=1}^p t_k\sigma_1^{(k)}\big(\varrho^{(k)}W_1 + \sqrt{1-(\varrho^{(k)})^2}\sqrt{W_1}U_k\big) \\ &= \Big(\sigma_2\sum_{k=1}^p t_k\sigma_1^{(k)}\varrho^{(k)}\Big)W_1 + \Big(\sigma_2\sum_{k=1}^p t_k\sigma_1^{(k)}\sqrt{1-(\varrho^{(k)})^2}U_k\Big)\sqrt{W_1}.\end{aligned}$$

Here, by (2.15),

$$\begin{aligned}
\sigma_2 \sum_{k=1}^{p} t_k \sigma_1^{(k)} \sqrt{1-(\varrho^{(k)})^2} U_k
&\stackrel{d}{=} \sigma_2 \Big(\sum_{k,l=1}^{p} t_k t_l \sigma_1^{(k)} \sigma_1^{(l)} \sqrt{1-(\varrho^{(k)})^2} \sqrt{1-(\varrho^{(l)})^2} \mathbb{E}(U_k U_l) \Big)^{1/2} U_1 \\
&= \sigma_2 \Big(\sum_{k,l=1}^{p} t_k t_l \sigma_1^{(k)} \sigma_1^{(l)} (\varrho^{(kl)} - \varrho^{(k)} \varrho^{(l)}) \Big)^{1/2} U_1.
\end{aligned}$$

Note that $U_1 \stackrel{d}{=} \mathcal{N}(0,1)$. So that,

$$\begin{aligned}
\sigma_2 \sum_{k=1}^{p} t_k \sigma_1^{(k)} \big(\varrho^{(k)} W_1 + \sqrt{1-(\varrho^{(k)})^2} \sqrt{W_1} U_k \big) &\stackrel{d}{=} \Big(\sigma_2 \sum_{k=1}^{p} t_k \sigma_1^{(k)} \varrho^{(k)} \Big) W_1 \\
&+ \sigma_2 \Big(\sum_{k,l=1}^{p} t_k t_l \sigma_1^{(k)} \sigma_1^{(l)} (\varrho^{(kl)} - \varrho^{(k)} \varrho^{(l)}) \Big)^{1/2} \sqrt{W_1} U_1,
\end{aligned}$$

which, by representation (2.13), has the same VG distribution as that in (2.17). This proves (2.16). □

2.3 Some auxiliary lemmas

In this section, we establish some auxiliary results that will be used in the proofs of Theorems 2.1 and 2.2. Here and throughout this chapter we remove the upper indices when working with triangular schemes of random variables, e.g., $(V_1, \dots, V_p) \equiv (V_1^{(p)}, \dots, V_p^{(p)})$, whenever it is clear from the context.

Lemma 2.2. *Let* $V = (V_1, \dots, V_p)' \stackrel{d}{=} \mathcal{N}_p(0, \Sigma_V^{(p)})$, *where* $\Sigma_V^{(p)}$ *is positive definite covariance matrix and* $\operatorname{tr}((\Sigma_V^{(p)})^2) = o(p^2)$, $p \to \infty$. *Then*

$$\frac{1}{p} \sum_{k=1}^{p} \big(V_k^2 - \mathbb{E} V_k^2 \big) \stackrel{\mathbb{P}}{\to} 0 \quad as \quad p \to \infty. \tag{2.18}$$

If, in addition, $p^{-1} \operatorname{tr}(\Sigma_V^{(p)}) \to 1$, *then*

$$\frac{1}{p} \sum_{k=1}^{p} V_k^2 \stackrel{\mathbb{P}}{\to} 1 \quad as \quad p \to \infty. \tag{2.19}$$

Proof. Due to the Spectral Theorem, we have

$$V'V \;=\; \sum_{k=1}^{p} V_k^2 \;\stackrel{d}{=}\; \sum_{j=1}^{p} \lambda_j^{(p)} \tilde{Z}_j^2, \tag{2.20}$$

where $\tilde{Z}_j$ are i.i.d. standard normal variables and $\lambda_1^{(p)}, \ldots, \lambda_p^{(p)}$ are the eigenvalues of $\Sigma_V^{(p)}$. Observe from (2.20) that

$$\mathbb{E}V'V \;=\; \sum_{j=1}^{p} \lambda_j^{(p)} \;=\; \operatorname{tr}(\Sigma_V^{(p)}), \tag{2.21}$$

$$\operatorname{Var}(V'V) \;=\; \operatorname{Var}\Big(\sum_{j=1}^{p} \lambda_j^{(p)} \tilde{Z}_j^2\Big) \;=\; 2\sum_{j=1}^{p} (\lambda_j^{(p)})^2 = 2\operatorname{tr}((\Sigma_V^{(p)})^2). \tag{2.22}$$

Thus, by (2.21)–(2.22), for any $\epsilon > 0$

$$\mathbb{P}\Big(\Big|\frac{1}{p}(V'V - \mathbb{E}V'V)\Big| > \epsilon\Big) \;\le\; \frac{\operatorname{Var}(V'V)}{p^2\epsilon^2} \;\to\; 0, \quad p \to \infty,$$

and the relation in (2.18) follows due to assumption $\operatorname{tr}((\Sigma_V^{(p)})^2) = o(p^2)$. Finally, if $p^{-1}\operatorname{tr}(\Sigma_V^{(p)}) \to 1$, by (2.21), the result (2.18) leads to (2.19). □

Remark 2.2. The assumption on matrix $\Sigma_V = \Sigma_V^{(p)}$ in Lemma 2.2, requiring that $\operatorname{tr}(\Sigma_V^2) = o(p^2)$, is not overly restrictive: assume, for example, that $\Sigma_V = (\sigma^{(i,j)})$ is any KMS type covariance matrix, as in (2.1). Then it is straightforward to see that

$$\begin{aligned}\operatorname{tr}(\Sigma_V^2) \;&=\; \sum_{i,j=1}^{p} (\sigma^{(i,j)})^2 \;=\; \sum_{i,j=1}^{p} \varrho^{2|i-j|} \\ &=\; \sum_{|m|<p} (p - |m|)\varrho^{2|m|} \;\le\; p \sum_{|m|<p} |m|\varrho^{2|m|} \;=\; \mathcal{O}(p).\end{aligned}$$

Lemma 2.3. *Assume that $\tilde{Z}_1, \tilde{Z}_2, \ldots$ are i.i.d. $\mathcal{N}(0,1)$ random variables. For any $p \in \mathbb{N}$ define*

$$\zeta_j^{(p)} \;:=\; \nu_j^{(p)}(\tilde{Z}_j^2 - 1) + \gamma_j^{(p)}\sqrt{p}\tilde{Z}_j, \quad j = 1, \ldots, p, \tag{2.23}$$

where $\nu_j^{(p)}$, $j = 1, \ldots, p$, are positive scalars, and $\gamma_j^{(p)}$, $j = 1, \ldots, p$, are *real scalars, such that*

$$\sum_{j=1}^{p}(\nu_j^{(p)})^3 = o\Bigg(\Big(\sum_{j=1}^{p}\mathrm{Var}\big(\zeta_j^{(p)}\big)\Big)^{3/2}\Bigg), \tag{2.24}$$

$$p\sum_{j=1}^{p}(\gamma_j^{(p)})^2\nu_j^{(p)} = o\Bigg(\Big(\sum_{j=1}^{p}\mathrm{Var}\big(\zeta_j^{(p)}\big)\Big)^{3/2}\Bigg) \tag{2.25}$$

with $\mathrm{Var}(\zeta_j^{(p)}) = 2(\nu_j^{(p)})^2 + p(\gamma_j^{(p)})^2$. *Then, as* $p \to \infty$,

$$\frac{\sum_{j=1}^{p}\zeta_j^{(p)}}{\sqrt{\sum_{j=1}^{p}\mathrm{Var}(\zeta_j^{(p)})}} \overset{d}{\to} \mathcal{N}(0,1). \tag{2.26}$$

Proof. The proof uses the method of cumulants and is structured as follows:

1. we establish the moment-generating function of $\zeta_j^{(p)}$, $M_{\zeta_j^{(p)}}(t) := \mathbb{E}e^{t\zeta_j^{(p)}}$, and $\log\big(M_{\zeta_j^{(p)}}(t)\big)$;

2. we find $G(t;p)$ which corresponds to the cumulant generating function of the sum $\sum_{j=1}^{p}\zeta_j^{(p)}$;

3. we find $K(t;p) := G\Bigg(\frac{t}{\sqrt{\sum_{j=1}^{p}(2(\nu_j^{(p)})^2 + p(\gamma_j^{(p)})^2)}};p\Bigg)$, which corresponds to the cumulant generating function of the left hand side of (2.26);

4. finally, in order to prove (2.26), we show that the cumulants $\varkappa_j^{(p)}$, generated by $K(t;p)$, satisfy $\varkappa_1^{(p)} = 0$, $\varkappa_2^{(p)} = 1$ and $\varkappa_d^{(p)} \to 0$, $d = 3, 4, \dots$, as $p \to \infty$.

Step 1. First, rewrite

$$\zeta_j^{(p)} = \nu_j^{(p)}\Bigg(\tilde{Z}_j + \frac{\gamma_j^{(p)}\sqrt{p}}{2\nu_j^{(p)}}\Bigg)^2 - \nu_j^{(p)} - \frac{(\gamma_j^{(p)})^2 p}{4\nu_j^{(p)}}. \tag{2.27}$$

Here, $\psi_j^{(p)} := \Big(\tilde{Z}_j + \frac{\gamma_j^{(p)}\sqrt{p}}{2\nu_j^{(p)}}\Big)^2$ has a non-central chi-squared distribution with the following moment-generating function:

$$M_{\psi_j^{(p)}}(t) \;:=\; \mathbb{E}e^{t\psi_j^{(p)}} \;=\; (1-2t)^{-1/2}\exp\left\{\left(\frac{\gamma_j^{(p)}}{2\nu_j^{(p)}}\right)^2 tp(1-2t)^{-1}\right\}, \quad (2.28)$$

for $|t| < \frac{1}{2}$. Therefore, by (2.27) and (2.28),

$$\begin{aligned}
M_{\zeta_j^{(p)}}(t) &= M_{\psi_j^{(p)}}(\nu_j^{(p)}t)\exp\left\{-t\nu_j^{(p)} - tp\left(\frac{\gamma_j^{(p)}}{2\nu_j^{(p)}}\right)^2\right\} \\
&= \left(1-2\nu_j^{(p)}t\right)^{-\frac{1}{2}}\exp\left\{\frac{(\gamma_j^{(p)})^2}{4\nu_j^{(p)}}tp\left(1-2\nu_j^{(p)}t\right)^{-1} - t\left(\nu_j^{(p)} + \frac{(\gamma_j^{(p)})^2 p}{4\nu_j^{(p)}}\right)\right\},
\end{aligned}$$

for $|t| < (2\nu_j^{(p)})^{-1}$, and

$$\begin{aligned}
\log\left(M_{\zeta_j^{(p)}}(t)\right) &= \left(\frac{\gamma_j^{(p)}}{2\nu_j^{(p)}}\right)^2 pt\nu_j^{(p)}\left(1-2\nu_j^{(p)}t\right)^{-1} - \frac{1}{2}\log\left(1-2\nu_j^{(p)}t\right) \\
&\quad - t\left(\nu_j^{(p)} + \frac{(\gamma_j^{(p)})^2 p}{4\nu_j^{(p)}}\right) \\
&= \frac{1}{2}\left((\gamma_j^{(p)})^2 p + 2(\nu_j^{(p)})^2\right)t^2 + \frac{(\gamma_j^{(p)})^2 p}{2}\sum_{k=3}^{\infty} t^k 2^{k-2}(\nu_j^{(p)})^{k-2} \\
&\quad + \frac{1}{2}\sum_{k=3}^{\infty}\frac{2^k(\nu_j^{(p)})^k t^k}{k}.
\end{aligned}$$

Step 2. Since $\zeta_1^{(p)}, \ldots, \zeta_j^{(p)}$ are independent, we have that

$$\begin{aligned}
G(t;p) &= \sum_{j=1}^{p}\log M_{\zeta_j^{(p)}}(t) = \frac{t^2}{2}\sum_{j=1}^{p}\left((\gamma_j^{(p)})^2 p + 2(\nu_j^{(p)})^2\right) \\
&\quad + \frac{p}{2}\sum_{k=3}^{\infty}2^{k-2}t^k\sum_{j=1}^{p}(\gamma_j^{(p)})^2(\nu_j^{(p)})^{k-2} + \frac{1}{2}\sum_{k=3}^{\infty}\frac{2^k}{k}t^k\sum_{j=1}^{p}(\nu_j^{(p)})^k.
\end{aligned}$$

Step 3. It is straightforward to see that

$$K(t;p) = G\left(\frac{t}{\sqrt{\sum_{j=1}^{p}\left(2(\nu_j^{(p)})^2 + p(\gamma_j^{(p)})^2\right)}};p\right)$$

$$
\begin{aligned}
&= \frac{t^2}{2} + \frac{1}{2}\sum_{k=3}^{\infty} 2^{k-2} t^k \frac{p\sum_{j=1}^p (\gamma_j^{(p)})^2 (\nu_j^{(p)})^{k-2}}{\big(\sum_{j=1}^p (2(\nu_j^{(p)})^2 + (\gamma_j^{(p)})^2 p)\big)^{k/2}} \\
&+ \frac{1}{2}\sum_{k=3}^{\infty} \frac{2^k}{k} t^k \frac{\sum_{j=1}^p (\nu_j^{(p)})^k}{\big(\sum_{j=1}^p (2(\nu_j^{(p)})^2 + (\gamma_j^{(p)})^2 p)\big)^{k/2}} \quad = \quad \sum_{k=1}^{\infty} \varkappa_k^{(p)} \frac{t^k}{k!},
\end{aligned}
$$

where $\varkappa_1^{(p)} = 0$, $\varkappa_2^{(p)} = 1$, and for $k \geq 3$,

$$
\varkappa_k^{(p)} = \frac{k!2^{k-3} p \sum_{j=1}^p (\gamma_j^{(p)})^2 (\nu_j^{(p)})^{k-2} + (k-1)!2^{k-1} \sum_{j=1}^p (\nu_j^{(p)})^k}{\big(\sum_{j=1}^p (2(\nu_j^{(p)})^2 + (\gamma_j^{(p)})^2 p)\big)^{k/2}}. \tag{2.29}
$$

Step 4. In order to prove that (2.26) holds, it remains to show that, as $p \to \infty$, $\varkappa_d^{(p)} \to 0$ for all $d \geq 3$. By (2.29), it is equivalent to showing that for any fixed $k \geq 3$, as $p \to \infty$,

$$
\frac{\sum_{j=1}^p (\nu_j^{(p)})^k}{\big(\sum_{j=1}^p \big(2(\nu_j^{(p)})^2 + (\gamma_j^{(p)})^2 p\big)\big)^{k/2}} \to 0, \tag{2.30}
$$

$$
\frac{p\sum_{j=1}^p (\gamma_j^{(p)})^2 (\nu_j^{(p)})^{k-2}}{\big(\sum_{j=1}^p \big(2(\nu_j^{(p)})^2 + (\gamma_j^{(p)})^2 p\big)\big)^{k/2}} \to 0. \tag{2.31}
$$

In order to prove (2.30) we use induction. The case for $k = 3$ holds by assumption. Assuming that (2.30) holds for fixed $k \geq 3$, we have

$$
\begin{aligned}
&\frac{\sum_{j=1}^p (\nu_j^{(p)})^{k+1}}{\big(\sum_{j=1}^p \big(2(\nu_j^{(p)})^2 + (\gamma_j^{(p)})^2 p\big)\big)^{(k+1)/2}} \\
&\leq \frac{\big(\sum_{j'=1}^p (\nu_{j'}^{(p)})^2\big)^{1/2} \sum_{j=1}^p (\nu_j^{(p)})^k}{\big(\sum_{j=1}^p \big(2(\nu_j^{(p)})^2 + (\gamma_j^{(p)})^2 p\big)\big)^{(k+1)/2}} \\
&\leq \frac{\big(\sum_{j'=1}^p \big(2(\nu_{j'}^{(p)})^2 + (\gamma_{j'}^{(p)})^2 p\big)\big)^{1/2} \sum_{j=1}^p (\nu_j^{(p)})^k}{\big(\sum_{j=1}^p \big(2(\nu_j^{(p)})^2 + (\gamma_j^{(p)})^2 p\big)\big)^{(k+1)/2}} \\
&= \frac{\sum_{j=1}^p (\nu_j^{(p)})^k}{\big(\sum_{j=1}^p \big(2(\nu_j^{(p)})^2 + (\gamma_j^{(p)})^2 p\big)\big)^{k/2}} \quad \to \quad 0,
\end{aligned}
$$

concluding that (2.30) holds for all $k \geq 3$. The proof for (2.31) is analogous: the case for $k = 3$ holds by assumption, thus, we repeat the same arguments as with (2.30) and conclude that (2.31) holds for all $k \geq 3$. This concludes the proof of the lemma. $\square$

2.4 Proof of the main result

In this section, we give the proofs of Theorems 2.1 and 2.2. Throughout the proofs, we express corresponding constants in terms of $\kappa_{i,p}$ and κ_i, $i = 1, 2, 3$, introduced in (2.2)–(2.4). Recall that $\kappa_{i,p} \geq 0$, and, by Remark 2.3, $\kappa_i < \infty$, for $i = 1, 2, 3$.

Proof of Theorem 2.1. Write

$$\|\mathbb{X}'Y\|_2^2 \quad = \quad H_1^2 + \cdots + H_p^2 \quad =: \; H,$$

where

$$H_k \quad := \quad \sum_{j=1}^{n} X_{k,j}\Big(\sum_{l=1}^{p} \beta_l X_{l,j} + \varepsilon_j\Big), \quad k = 1, \ldots, p.$$

Denote $Z_j := \sum_{l=1}^{p} \beta_l X_{l,j} + \varepsilon_j$, $j = 1, \ldots, n$. By covariance structure (2.1) and $X_{k,j} \stackrel{d}{=} \mathcal{N}(0,1)$, $\varepsilon_j \stackrel{d}{=} \mathcal{N}(0, \sigma_\varepsilon^2)$, we have $Z_j \stackrel{d}{=} \mathcal{N}(0, \sigma_Z^2)$, where $\sigma_Z^2 = \sum_{l,l'=1}^{p} \beta_l \beta_{l'} \varrho^{|l-l'|} + \sigma_\varepsilon^2$ and $\mathrm{Cov}(X_{k,j}, Z_j) = \sum_{l=1}^{p} \beta_l \varrho^{|k-l|}$.

Applying Proposition 2.1(iii) with $\xi_{1j}^{(k)} = X_{k,j}$, $\xi_{2j} = Z_j$, and $\sigma_1^{(k)} = 1$, $\sigma_{2,p} = \sigma_Z$, $\theta_k^{(p)} := \varrho^{(k)} = \sigma_Z^{-1} \sum_{l=1}^{p} \beta_l \varrho^{|k-l|}$, where $\varrho^{(kl)} = \varrho^{|k-l|}$, we obtain that

$$\|\mathbb{X}'Y\|_2^2 \quad \stackrel{d}{=} \quad \sigma_{2,p}^2 \sum_{k=1}^{p} \Big(\theta_k^{(p)} W_n + \sqrt{1 - (\theta_k^{(p)})^2}\sqrt{W_n} U_k\Big)^2,$$

where $W_n \stackrel{d}{=} \Gamma(n/2, 1/2)$ and $(U_1, \ldots, U_p)' \stackrel{d}{=} \mathcal{N}(0, \Sigma_U^{(p)})$ with $\Sigma_U^{(p)} = (\sigma_U^{(k,l)})$ defined as (see (2.15)):

$$\sigma_U^{(k,l)} \quad = \quad \frac{\varrho^{|k-l|} - \theta_k^{(p)} \theta_l^{(p)}}{\sqrt{1 - (\theta_k^{(p)})^2}\sqrt{1 - (\theta_l^{(p)})^2}}, \quad k, l = 1, \ldots, p. \tag{2.32}$$

By expanding the square we can write

$$\|\mathbb{X}'Y\|_2^2 \quad \stackrel{d}{=} \quad \sigma_{2,p}^2 \Big((W_n - \mathbb{E}W_n + \mathbb{E}W_n)^2 \sum_{k=1}^{p} (\theta_k^{(p)})^2$$

$$
\begin{aligned}
&+\ 2W_n^{3/2}\sum_{k=1}^{p}\theta_k^{(p)}\sqrt{1-(\theta_k^{(p)})^2}U_k \\
&+ (W_n-\mathbb{E}W_n)\sum_{k=1}^{p}\big(1-(\theta_k^{(p)})^2\big)U_k^2 \\
&+\ \mathbb{E}W_n\sum_{k=1}^{p}\big(1-(\theta_k^{(p)})^2\big)U_k^2\Big).
\end{aligned} \tag{2.33}
$$

By further rearranging the right-hand side, we have

$$
\frac{\|\mathbb{X}'Y\|_2^2}{n^{3/2}} \overset{d}{=} I_1+I_2+I_3+I_4, \tag{2.34}
$$

where

$$
I_1 := \frac{\sigma_{2,p}^2}{n^{3/2}}(W_n-\mathbb{E}W_n)^2\sum_{k=1}^{p}(\theta_k^{(p)})^2, \tag{2.35}
$$

$$
I_2 := \frac{\sigma_{2,p}^2}{n^{3/2}}(W_n-\mathbb{E}W_n)\Big(2\mathbb{E}W_n\sum_{k=1}^{p}(\theta_k^{(p)})^2+\sum_{k=1}^{p}(1-(\theta_k^{(p)})^2)U_k^2\Big), \tag{2.36}
$$

$$
\begin{aligned}
I_3 &:= \frac{\sigma_{2,p}^2}{n^{3/2}}\,2W_n^{3/2}\sum_{k=1}^{p}\theta_k^{(p)}\sqrt{1-(\theta_k^{(p)})^2}U_k \\
&+\ \frac{\sigma_{2,p}^2}{n^{3/2}}\,\mathbb{E}W_n\sum_{k=1}^{p}\big((1-(\theta_k^{(p)})^2)U_k^2-1\big),
\end{aligned} \tag{2.37}
$$

$$
I_4 := \frac{\sigma_{2,p}^2}{n^{3/2}}\Big(p\mathbb{E}W_n+(\mathbb{E}W_n)^2\sum_{k=1}^{p}(\theta_k^{(p)})^2\Big). \tag{2.38}
$$

We will show that, as $p,n\to\infty$, $p/n\to c\in(0,\infty)$, the term $I_1=o_P(1)$, while the terms I_2 and I_3 are asymptotically normal. More precisely, we will show that $I_2\overset{d}{\to}\mathcal{N}(0,s_1^2)$ and $I_3\overset{d}{\to}\mathcal{N}(0,s_2^2)$, where s_1^2 and s_2^2 are given by (2.44) and (2.62) below. Here, since W_n and $(U_1,\ldots,U_p)'$ are mutually independent for each n, it follows that $I_2+I_3\overset{d}{\to}\mathcal{N}(0,s_1^2+s_2^2)$. Finally, the term I_4 defines the mean of the statistic, i.e.

$$
\frac{\|\mathbb{X}'Y\|_2^2}{n^{3/2}}-I_4 \overset{d}{\to} \mathcal{N}(0,s_1^2+s_2^2). \tag{2.39}
$$

Thus, we will conclude by establishing that $I_4 = \sqrt{n}(\kappa_{2,p} + pn^{-1}(\kappa_{1,p} + \sigma_\varepsilon^2))$, while $s_1^2 + s_2^2 = s^2$, as in the statement of the theorem.

First, consider I_1 defined in (2.35). We will show that $I_1 = o_P(1)$. Denote

$$c_2 \;:=\; \lim_{p\to\infty} \sum_{k=1}^{p} (\theta_k^{(p)})^2 = (\kappa_1 + \sigma_\varepsilon^2)^{-1}\kappa_2, \quad \sigma_2^2 \;:=\; \lim_{p\to\infty} \sigma_{2,p}^2 \;=\; \kappa_1 + \sigma_\varepsilon^2. \tag{2.40}$$

It is clear that $c_2 < \infty$ and $\sigma_2^2 < \infty$. Recall that, by CLT,

$$\frac{W_n - \mathbb{E}W_n}{n^{1/2}} \;\stackrel{d}{\to}\; \mathcal{N}(0, 2). \tag{2.41}$$

Therefore,

$$I_1 \;=\; \mathcal{O}(1) n^{-1/2} \Big(\frac{W_n - \mathbb{E}W_n}{n^{1/2}}\Big)^2 \;=\; o(1)\mathcal{O}_P(1) \;=\; o_P(1). \tag{2.42}$$

Second, consider I_2, defined in (2.36). We will show that

$$I_2 \;\stackrel{d}{\to}\; \mathcal{N}(0, s_1^2) \tag{2.43}$$

with s_1^2 given by

$$s_1^2 \;=\; 2\sigma_2^4(2c_2 + c)^2 \;=\; 8\kappa_2^2 + 8c(\kappa_1 + \sigma_\varepsilon^2)\kappa_2 + 2c^2(\kappa_1 + \sigma_\varepsilon^2)^2. \tag{2.44}$$

Rewrite

$$I_2 \;=\; \sigma_{2,p}^2 \frac{W_n - \mathbb{E}W_n}{n^{1/2}} \Big(\frac{2\mathbb{E}W_n}{n} \sum_{k=1}^{p} (\theta_k^{(p)})^2 + \frac{1}{n} \sum_{k=1}^{p} (1 - (\theta_k^{(p)})^2) U_k^2 \Big). \tag{2.45}$$

Applying (2.40) and (2.41) for the outer term of (2.45), we obtain

$$\sigma_{2,p}^2 \frac{W_n - \mathbb{E}W_n}{n^{1/2}} \;\stackrel{d}{\to}\; \mathcal{N}(0, 2\sigma_2^4).$$

We will show that the inner term of (2.45) approaches $2c_2 + c$. Since $\mathbb{E}W_n = n$, by (2.40) and assumption $p/n \to c$ it suffices to prove the convergence

$$\frac{1}{p}\sum_{k=1}^{p}(1-(\theta_k^{(p)})^2)U_k^2 \quad \stackrel{\mathbb{P}}{\to} \quad 1. \tag{2.46}$$

Denote matrix

$$A \quad := \quad \operatorname{diag}\left(1-(\theta_1^{(p)})^2,\ \dots\ ,1-(\theta_p^{(p)})^2\right). \tag{2.47}$$

To prove (2.46) we apply Lemma 2.2 with $V_j = \sqrt{1-(\theta_j^{(p)})^2}U_j$, $j = 1,\dots,p$, and $\Sigma_V^{(p)} = A^{1/2}\Sigma_U A^{1/2}$. Obviously, the conditions of Lemma 2.2 will hold if $\operatorname{tr}((A^{1/2}\Sigma_U A^{1/2})^2) = \mathcal{O}(p)$ and $p^{-1}\operatorname{tr}(A^{1/2}\Sigma_U A^{1/2}) \to 1$, as $p \to \infty$. Observe, that

$$\begin{aligned}
\operatorname{tr}((A^{1/2}\Sigma_U A^{1/2})^2) &= \operatorname{tr}((A\Sigma_U)^2) \\
&= \sum_{k=1}^{p}\sum_{k'=1}^{p}(1-(\theta_k^{(p)})^2)(1-(\theta_{k'}^{(p)})^2)(\sigma_U^{(k,k')})^2 \\
&= \sum_{k=1}^{p}\sum_{k'=1}^{p}\left(\varrho^{2|k-k'|} - 2\varrho^{|k-k'|}\theta_k^{(p)}\theta_{k'}^{(p)} + (\theta_k^{(p)})^2(\theta_{k'}^{(p)})^2\right) \\
&= \sum_{k=1}^{p}\sum_{k'=1}^{p}\varrho^{2|k-k'|} - 2(\kappa_{1,p}+\sigma_\varepsilon^2)^{-1}\kappa_{3,p} \\
&\quad + \ (\kappa_{1,p}+\sigma_\varepsilon^2)^{-2}\kappa_{2,p}^2 \\
&= \sum_{k=1}^{p}\sum_{k'=1}^{p}\varrho^{2|k-k'|} + o(p) \quad \sim \quad p\,\frac{1+\varrho^2}{1-\varrho^2},
\end{aligned} \tag{2.48}$$

since $\kappa_i < \infty$, $i = 1,2,3$ and $\kappa_{1,p} \geq 0$. Here we used (2.40) and the observation that

$$\sum_{k=1}^{p}\sum_{k'=1}^{p}\varrho^{|k-k'|}\theta_k^{(p)}\theta_{k'}^{(p)} \quad = \quad \frac{\kappa_{3,p}}{\kappa_{1,p}+\sigma_\varepsilon^2} \quad \to \quad \frac{\kappa_3}{\kappa_1+\sigma_\varepsilon^2}, \quad \text{as } p\to\infty. \tag{2.49}$$

Similarly, we have

$$\frac{1}{p}\operatorname{tr}(A^{1/2}\Sigma_U A^{1/2}) \quad = \quad \frac{1}{p}\sum_{k=1}^{p}(1-(\theta_k^{(p)})^2) \ = \ 1 \ - \ \frac{\kappa_{2,p}}{p(\kappa_{1,p}+\sigma_\varepsilon^2)} \quad \to \quad 1,$$

since, by Lemma A.4, $\kappa_{2,p} = o(p)$, while $\kappa_{1,p} \geq 0$, $\kappa_1 < \infty$. This concludes the proof of (2.46).

Next, consider I_3, defined by (2.37). We will show that

$$I_3 \ \stackrel{d}{\to} \ \mathcal{N}(0, s_2^2), \tag{2.50}$$

with s_2^2 defined in (2.62). Write

$$I_3 \ = \ \sigma_{2,p}^2 \Big(2\frac{W_n^{3/2}}{n^{3/2}} \mathbf{b}'U + n^{-1/2}(U'AU - p) \Big),$$

where $U \ = \ (U_1, \ldots, U_p)'$, A is defined by (2.47), and

$$\mathbf{b} \ = \ \Big(\theta_1^{(p)} \sqrt{1 - (\theta_1^{(p)})^2}, \ldots, \theta_p^{(p)} \sqrt{1 - (\theta_p^{(p)})^2} \Big)'.$$

Observe that $n^{-3/2} W_n^{3/2} \ \stackrel{\mathbb{P}}{\to} \ 1$ due to the Law of Large Numbers. Thus, since W_n and U are independent for any n and $p/n \to c$, it follows that

$$I_3 \ = \ \sigma_{2,p}^2 \Big(2\mathbf{b}'U + \sqrt{\frac{c}{p}} \big(U'AU - p \big) \Big) + o_P(1). \tag{2.51}$$

First, we consider the inner term of (2.51) and show, that, as $p \ \to \ \infty$,

$$2\mathbf{b}'U \ + \ \sqrt{\frac{c}{p}} \, (UAU' - p) \ \stackrel{d}{\to} \ V_2, \tag{2.52}$$

where $V_2 \stackrel{d}{=} \mathcal{N}(0, \sigma_2^{-4} s_2^2)$. Then, (2.50) readily follows from (2.51).

Recall, that $U \ \stackrel{d}{=} \ \mathcal{N}_p(0, \Sigma_U)$, $\Sigma_U > 0$. Further, let $\tilde{Z} \ \stackrel{d}{=} \ \mathcal{N}_p(0, I_p)$. Clearly, one has that $U \ \stackrel{d}{=} \ \Sigma_U^{1/2} \tilde{Z}$, where $\Sigma_U^{1/2}$ denotes the symmetric square root of Σ_U. By the Spectral Theorem, we construct $V := P'\tilde{Z}$, where $V \stackrel{d}{=} \mathcal{N}_p(0, I_p)$ and P is an orthogonal matrix that diagonalizes $\Sigma_U^{1/2} A \Sigma_U^{1/2}$, such, that $P' \Sigma_U^{1/2} A \Sigma_U^{1/2} P = \Lambda$, with $\Lambda = \mathrm{diag}(\lambda_1^{(p)}, \ldots, \lambda_p^{(p)})$ comprised of the eigenvalues of $\Sigma_U^{1/2} A \Sigma_U^{1/2}$. Then,

$$\begin{aligned}
\frac{\sqrt{c}}{\sqrt{p}} \big(U'AU - p \big) + 2\mathbf{b}'U \ &\stackrel{d}{=} \ \frac{\sqrt{c}}{\sqrt{p}} \big(V'\Lambda V - p \big) + 2\mathbf{b}'\Sigma_U^{1/2} PV \\
&= \ \frac{\sqrt{c}}{\sqrt{p}} \Big(\sum_{j=1}^{p} \Big(\lambda_j^{(p)} (V_j^2 - 1) + g_j^{(p)} \sqrt{p} V_j \Big) \Big) \\
&=: \ \frac{\sqrt{c}}{\sqrt{p}} \sum_{j=1}^{p} \widetilde{V}_j^{(p)}, \qquad (2.53)
\end{aligned}$$

where $(g_1^{(p)},\dots,g_p^{(p)}) = 2c^{-1/2}\mathbf{b}'\Sigma_U^{1/2}P$, and

$$\widetilde{V}_j^{(p)} \;:=\; \lambda_j^{(p)}(V_j^2-1) \;+\; g_j^{(p)}\sqrt{p}V_j, \quad j=1,\dots,p. \tag{2.54}$$

Clearly, $\mathbb{E}\widetilde{V}_j^{(p)}=0$ and $\mathbb{E}(\widetilde{V}_j^{(p)})^2 = 2(\lambda_j^{(p)})^2+(g_j^{(p)})^2p$. Therefore, proving the result (2.52) is equivalent to showing:

$$\frac{\sqrt{c}}{\sqrt{p}}\sum_{j=1}^{p}\widetilde{V}_j^{(p)} \;\xrightarrow{d}\; \mathcal{N}(0,\sigma_2^{-4}s_2^2), \tag{2.55}$$

where

$$\begin{aligned}\sigma_2^{-4}s_2^2 \;&=\; c\lim_{p\to\infty}p^{-1}\sum_{j=1}^{p}\mathbb{E}(\widetilde{V}_j^{(p)})^2\\ &=\; 2c\lim_{p\to\infty}p^{-1}\sum_{j=1}^{p}(\lambda_j^{(p)})^2 + c\lim_{p\to\infty}\sum_{j=1}^{p}(g_j^{(p)})^2. \end{aligned}\tag{2.56}$$

We prove (2.55) by applying Lemma 2.3 with $\nu_j^{(p)}=\lambda_j^{(p)}$ as the eigenvalues of $\Sigma_U^{1/2}A\Sigma_U^{1/2}$ and $\gamma_j^{(p)}=g_j^{(p)}$. By the conditions of Lemma 2.3, we need to show that the following holds:

$$\sum_{j=1}^{p}(\lambda_j^{(p)})^3 + p\sum_{j=1}^{p}(g_j^{(p)})^2\lambda_j^{(p)} \;=\; o\Bigg(\bigg(\sum_{j=1}^{p}\big(2(\lambda_j^{(p)})^2+(g_j^{(p)})^2p\big)\bigg)^{3/2}\Bigg). \tag{2.57}$$

First, observe that $p^{-1}\sum_{j=1}^{p}(2(\lambda_j^{(p)})^2+(g_j^{(p)})^2p)\to C\in(0,\infty)$. Indeed, we have that $\sum_{j=1}^{p}(g_j^{(p)})^2\to C_g\in(0,\infty)$, since

$$\begin{aligned}\sum_{j=1}^{p}(g_j^{(p)})^2 \;&=\; 4c^{-1}(\mathbf{b}'\Sigma_U^{1/2}P)(\mathbf{b}'\Sigma_U^{1/2}P)' \;=\; 4c^{-1}\mathbf{b}'\Sigma_U\mathbf{b}\\ &=\; 4c^{-1}\sum_{j=1}^{p}\sum_{j'=1}^{p}\theta_j^{(p)}\theta_{j'}^{(p)}\sqrt{1-(\theta_j^{(p)})^2}\sqrt{1-(\theta_{j'}^{(p)})^2}\sigma_U^{(j,j')}\\ &=\; 4c^{-1}\sum_{j=1}^{p}\sum_{j'=1}^{p}\theta_j^{(p)}\theta_{j'}^{(p)}\left(\varrho^{|j-j'|}-\theta_j^{(p)}\theta_{j'}^{(p)}\right)\\ &\to\; 4c^{-1}\left(\kappa_1+\sigma_\varepsilon^2\right)^{-1}\kappa_3 - 4c^{-1}\left(\kappa_1+\sigma_\varepsilon^2\right)^{-2}\kappa_2^2 \;=\; C_g\end{aligned}\tag{2.58}$$

by (2.40) and (2.49).

Next, by (2.48), we find that $p^{-1}\sum_{j=1}^{p}(\lambda_j^{(p)})^2 \to C_\lambda \in (0,\infty)$. Indeed, by (2.48), we have

$$\sum_{j=1}^{p}(\lambda_j^{(p)})^2 = \operatorname{tr}((\Sigma_U^{1/2}A\Sigma_U^{1/2})^2) = \operatorname{tr}((\Sigma_U A)^2)$$
$$= \sum_{j=1}^{p}\sum_{j'=1}^{p}\varrho^{2|j-j'|} + o(p) \sim p\,\frac{1+\varrho^2}{1-\varrho^2}. \tag{2.59}$$

Thus, by (2.58) and (2.59), it follows that $p^{-1}\sum_{j=1}^{p}(2c(\lambda_j^{(p)})^2 + (g_j^{(p)})^2 p) \to C \in (0,\infty)$ and condition (2.57) reduces to:

$$\sum_{j=1}^{p}(\lambda_j^{(p)})^3 + p\sum_{j=1}^{p}(g_j^{(p)})^2\lambda_j^{(p)} = o(p^{3/2}). \tag{2.60}$$

We show that (2.60) holds. For the first term of (2.60), we have

$$\sum_{j=1}^{p}(\lambda_j^{(p)})^3 = \operatorname{tr}((\Sigma_U^{1/2}A\Sigma_U^{1/2})^3) = \operatorname{tr}((\Sigma_U A)^3)$$
$$= \sum_{i,j,k=1}^{p}\big(1-(\theta_i^{(p)})^2\big)\big(1-(\theta_k^{(p)})^2\big)\big(1-(\theta_j^{(p)})^2\big)\sigma_U^{(i,j)}\sigma_U^{(i,k)}\sigma_U^{(k,j)}$$
$$= \sum_{i,j,k=1}^{p}\big(\varrho^{|i-j|}+\theta_i^{(p)}\theta_j^{(p)}\big)\big(\varrho^{|i-k|}+\theta_i^{(p)}\theta_k^{(p)}\big)\big(\varrho^{|k-j|}+\theta_k^{(p)}\theta_j^{(p)}\big)$$
$$= o(p^{3/2}), \tag{2.61}$$

where the last equality follows from Lemma A.5. For the second term of (2.60), observe, that by Hölder's inequality and (2.61),

$$p\sum_{j=1}^{p}(g_j^{(p)})^2\lambda_j^{(p)} \le p\Big(\sum_{j=1}^{p}|g_j^{(p)}|^3\Big)^{2/3}\Big(\sum_{j=1}^{p}(\lambda_j^{(p)})^3\Big)^{1/3}$$
$$= p^{3/2}\mathcal{O}(1)\Big(\frac{\sum_{j=1}^{p}(\lambda_j^{(p)})^3}{p^{3/2}}\Big)^{1/3} = o(p^{3/2}).$$

This concludes with (2.60), ensuring that the conditions of Lemma 2.3 hold.

Now we can establish the expression for s_2^2. By (2.40), (2.56), (2.58) and (2.59),

$$\begin{aligned}
s_2^2 &= \sigma_2^4 \lim_{p\to\infty} \sum_{j=1}^{p} \left(2p^{-1}c(\lambda_j^{(p)})^2 + c(g_j^{(p)})^2\right) \\
&= \sigma_2^4 \lim_{p\to\infty} \frac{2c}{p}\Bigg(\sum_{k=1}^{p}\sum_{k'=1}^{p} \varrho^{2|k-k'|} + o(p)\Bigg) + 4\sigma_2^4 \left(\kappa_1+\sigma_\varepsilon^2\right)^{-1}\kappa_3 \\
&- \ 4\sigma_2^4\left(\kappa_1+\sigma_\varepsilon^2\right)^{-2}\kappa_2^2 \\
&= 2c\,\frac{1+\varrho^2}{1-\varrho^2}\left(\kappa_1+\sigma_\varepsilon^2\right)^2 + 4(\kappa_1+\sigma_\varepsilon^2)\kappa_3 - 4\kappa_2^2. \qquad (2.62)
\end{aligned}$$

By (2.44) and (2.62), recalling that $s^2 = s_1^2 + s_2^2$, we have that

$$s^2 \;=\; 4\kappa_2^2 + 4(\kappa_1+\sigma_\varepsilon^2)\left(2\kappa_2 c + \kappa_3\right) + 2c(\kappa_1+\sigma_\varepsilon^2)^2\Big(c + \frac{1+\varrho^2}{1-\varrho^2}\Big). \quad (2.63)$$

Finally, consider I_4, defined by (2.38). Since $\mathbb{E}W_n = n$, we have that

$$I_4 \;=\; \frac{\kappa_{1,p}+\sigma_\varepsilon^2}{n^{3/2}}\Big(n^2\,\frac{\kappa_{2,p}}{\kappa_{1,p}+\sigma_\varepsilon^2} + pn\Big) \;=\; \sqrt{n}\Big(\kappa_{2,p} + \frac{p}{n}(\kappa_{1,p}+\sigma_\varepsilon^2)\Big). \quad (2.64)$$

By (2.34), having established the 4 parts by (2.35)–(2.38), we proved that (2.39) holds due to (2.42), (2.43), (2.50), (2.62), with terms (2.63) and (2.64), as in the statement of the theorem, thus concluding the proof. □

Before proceeding with the proof of Theorem 2.2, we establish the following lemma that ensures $\mathcal{O}(p^{-1/2})$ convergence rate for $\kappa_{1,p}$ and $\kappa_{2,p}$, appearing in Theorem 2.1, under additional restrictions for the parameters β_j.

Lemma 2.4. *Assume that* $\sum_{j=p+1}^{\infty}\beta_j^2 = o(p^{-1/2})$ *and* $\sup_{j\geq 1}|\beta_j|j^{\alpha} < \infty$, $\alpha > 1/2$, *and* $|\varrho| < 1$. *Then,*

1. $\kappa_1 = \kappa_{1,p} + o(p^{-1/2})$,
2. $\kappa_2 = \kappa_{2,p} + o(p^{-1/2})$.

Proof. For the proof see Appendix, Section A.2. □

Proof of Theorem 2.2. Rewrite the left-hand side of (2.9) as follows:

$$\begin{aligned}\frac{\|\mathbb{X}'Y\|_2^2 - n^2(\kappa_2 + c(\kappa_1 + \sigma_\varepsilon^2))}{n^{3/2}} &= \frac{\|\mathbb{X}'Y\|_2^2 - n^2\kappa_{2,p} - pn(\kappa_{1,p} + \sigma_\varepsilon^2)}{n^{3/2}} \\ &\quad + \sqrt{n}(\kappa_{2,p} - \kappa_2) + \sqrt{n}c(\kappa_{1,p} - \kappa_1) \\ &\quad + o(1).\end{aligned}$$

It remains to apply Lemma 2.4 and Theorem 2.1 in order to conclude the proof of the theorem. □

We end this section by deriving two supporting results that allows us to derive convenient alternative expressions for the terms κ_1, κ_2 and κ_3. For this, we introduce functions $\beta(\cdot)$ and $b(\cdot)$ by Definition 2.1 below, which, under the assumptions of Theorem 2.1 and a given structure of β_j's, requires only to evaluate the terms $\beta(1), \beta(\varrho), \beta(\varrho^2)$ and $b_1(\varrho), b_2(\varrho)$. Then, due to Lemma 2.5 below, the expressions for κ_1, κ_2 and κ_3 readily follow.

Definition 2.1. Assume that $\sum_{j=1}^{\infty} \beta_j^2 < \infty$ and $|\varrho| \leq 1$. Define,

$$\beta(\varrho) := \sum_{j=1}^{\infty} \beta_j^2 \varrho^j, \tag{2.65}$$

$$b_1(\varrho) := \sum_{j'=2}^{\infty} \sum_{j=1}^{j'-1} \beta_j \beta_{j'} \varrho^{j'-j}, \tag{2.66}$$

$$b_2(\varrho) := \sum_{j=2}^{\infty} \sum_{j'=1}^{j-1} \beta_j \beta_{j'} \varrho^{j+j'}, \tag{2.67}$$

and define the following quantities which involve derivatives of (2.65)–(2.67):

$$\beta^{(1)}(\varrho) := \varrho \frac{\mathrm{d}\beta(\varrho)}{\mathrm{d}\varrho} = \sum_{j=1}^{\infty} j \beta_j^2 \varrho^j, \tag{2.68}$$

$$b_1^{(1)}(\varrho) := \varrho \frac{\mathrm{d}b_1(\varrho)}{\mathrm{d}\varrho} = \sum_{j'=2}^{\infty} \sum_{j=1}^{j'-1} \beta_j \beta_{j'} \varrho^{j'-j}(j' - j), \tag{2.69}$$

$$b_2^{(1)}(\varrho) := \varrho \frac{\mathrm{d}b_2(\varrho)}{\mathrm{d}\varrho} = \sum_{j'=2}^{\infty} \sum_{j=1}^{j'-1} \beta_j \beta_{j'} \varrho^{j'+j}(j' + j), \tag{2.70}$$

$$b^{(2)}(\varrho) \quad := \quad \varrho^2 \frac{\mathrm{d}^2 b_1(\varrho)}{\mathrm{d}\varrho^2} + b_1^{(1)}(\varrho) \quad = \quad \sum_{j'=2}^{\infty} \sum_{j=1}^{j'-1} \beta_j \beta_{j'} \varrho^{j'-j} (j'-j)^2. \quad (2.71)$$

Note, that, by the rules of differentiation of power series, the functions (2.68)–(2.71) are well defined.

Lemma 2.5. *Let the assumptions of Theorem 2.1 hold. Let κ_1, κ_2 and κ_3 be given by (2.2), (2.3) and (2.4), respectively. Then, under notation in Definition 2.1, the following identities hold:*

1. $\kappa_1 \; = \; \beta(1) \; + \; 2b_1(\varrho)$,

2. $$\kappa_2 \; = \; \beta(1)\frac{1+\varrho^2}{1-\varrho^2} \; - \; \beta(\varrho^2)\frac{1}{1-\varrho^2} + \; 2\Big(b_1^{(1)}(\varrho) \; + \; b_1(\varrho)\frac{1+\varrho^2}{1-\varrho^2} \; - \; b_2(\varrho)\frac{1}{1-\varrho^2}\Big),$$

3. $$\kappa_3 = \frac{1}{1-\varrho^2}\big(3b_1^{(1)}(\varrho)(1+\varrho^2) - 2\big(b_2^{(1)}(\varrho) + \beta^{(1)}(\varrho^2)\big)\big) + b^{(2)}(\varrho) + \frac{1}{(1-\varrho^2)^2}(1+4\varrho^2+\varrho^4)(\beta(1)+2b_1(\varrho)) - \frac{1}{(1-\varrho^2)^2}(1+3\varrho^2)(\beta(\varrho^2)+2b_2(\varrho)).$$

Proof. See the proof in Appendix A.1. □

Remark 2.3. From the assumptions of Definition 2.1 it follows that $\beta(1), |\beta(\varrho)|, |b_1(\varrho)|, \; |b_2(\varrho)| \; < \; \infty$ for $|\varrho| \; < \; 1$. Thus, it follows from Lemma 2.5 that $\kappa_i < \infty$, $i = 1, 2, 3$.

Proof of Remark 2.3. Cases for $\beta(1)$ and $\beta(\varrho)$ follow straightforwardly from the assumptions. Consider $b_1(\varrho)$. Note, that for any p,

$$\begin{aligned}
|b_1(\varrho)| \;&\leq\; \sum_{l_1,l_2=1}^{\infty} |\beta_{l_1}||\beta_{l_2}||\varrho|^{|l_1-l_2|} \\
&=\; \sum_{l_1,l_2=1}^{\infty} \big(|\beta_{l_1}||\varrho|^{|l_1-l_2|/2}\big)\big(|\beta_{l_2}||\varrho|^{|l_1-l_2|/2}\big) \\
&\leq\; (1/2)\sum_{l_1,l_2=1}^{\infty} \big(\beta_{l_1}^2|\varrho|^{|l_1-l_2|} \;+\; \beta_{l_2}^2|\varrho|^{|l_1-l_2|}\big) \\
&=\; \sum_{l_1=1}^{\infty} \beta_{l_1}^2 \sum_{l_2=1}^{\infty} |\varrho|^{|l_1-l_2|} \;\leq\; \beta(1)\frac{1+|\varrho|}{1-|\varrho|} \;<\; \infty
\end{aligned}$$

by (A.24). In a similar manner, it is easy to see that $|b_2(\varrho)| \leq \beta(1)\frac{|\varrho|}{1-|\varrho|}$.
□

2.5 Approximate sparsity: an example

In this section, we study the case when coefficients β_j decay hyperbolically, i.e., $\beta_j = j^{-1}, j \geq 1$. This assumption is analogous to the assumption of approximate sparsity, as defined by Belloni and Chernozhukov (2011), reviewed in greater detail in Chapter 1.

In order to derive the quantities in Theorem 2.2, we apply the results of Lemma 2.5. For this, we establish the expressions for the quantities in Definition 2.1.

Define the real dilogarithm function (see, e.g., Morris (1979)):

$$\mathrm{Li}_2(x) \;=\; -\int_0^x \frac{\log(1-u)}{u}\,\mathrm{d}u, \quad x \leq 1, \quad x \in \mathbb{R}. \tag{2.72}$$

(Here and below, $\int_0^x = -\int_x^0$ if $x \leq 0$.) For $|x| \leq 1$ the real dilogarithm has a series representation,

$$\mathrm{Li}_2(x) \;=\; \sum_{k=1}^{\infty} \frac{x^k}{k^2}. \tag{2.73}$$

Then,

$$\beta(1) \;=\; \sum_{j=1}^{\infty} \frac{1}{j^2} \;=\; \frac{\pi^2}{6}, \qquad \beta(\varrho) \;=\; \sum_{j=1}^{\infty} \frac{\varrho^j}{j^2} \;=\; \mathrm{Li}_2(\varrho).$$

Additionally, we have

$$\frac{\mathrm{d}}{\mathrm{d}\varrho}\mathrm{Li}_2(\varrho) \;=\; -\frac{\log(1-\varrho)}{\varrho}. \tag{2.74}$$

Thus, by (2.68) and (2.74), we establish

$$\beta^{(1)}(\varrho) \;=\; \varrho\frac{\mathrm{d}}{\mathrm{d}\varrho}\beta(\varrho) \;=\; \varrho\frac{\mathrm{d}}{\mathrm{d}\varrho}\mathrm{Li}_2(\varrho) \;=\; -\log(1-\varrho).$$

Next, note that

$$\begin{aligned}
b_1(\varrho) &= \sum_{i=2}^{\infty}\sum_{j=1}^{i-1}\frac{\varrho^{i-j}}{ij} = \sum_{i=2}^{\infty}\sum_{k=1}^{i-1}\frac{\varrho^k}{i(i-k)} \\
&= \sum_{k=1}^{\infty}\varrho^k\sum_{i=k+1}^{\infty}\frac{1}{i(i-k)} = \sum_{k=1}^{\infty}\frac{\varrho^k}{k}\sum_{l=1}^{k}\frac{1}{l} \\
&= \sum_{l=1}^{\infty}\frac{1}{l}\sum_{k=l}^{\infty}\frac{\varrho^k}{k} = \sum_{l=1}^{\infty}\frac{1}{l}\int_0^{\varrho}\frac{x^{l-1}}{1-x}\,\mathrm{d}x \\
&= -\int_0^{\varrho}\frac{\log(1-x)}{x(1-x)}\,\mathrm{d}x = \frac{\log^2(1-\varrho)}{2} + \mathrm{Li}_2(\varrho), \qquad (2.75)
\end{aligned}$$

where we have used identities

$$\sum_{i=k+1}^{\infty}\frac{1}{i(i-k)} = \frac{1}{k}\sum_{l=1}^{k}\frac{1}{l},\ k \geq 1, \qquad \sum_{k=l}^{\infty}\frac{\varrho^k}{k} = \int_0^{\varrho}\frac{x^{l-1}}{1-x}\,\mathrm{d}x$$

and (2.72). Then, by (2.69), (2.74) and (2.75),

$$b_1^{(1)}(\varrho) = \varrho\frac{\mathrm{d}}{\mathrm{d}\varrho}b_1(\varrho) = -\frac{\log(1-\varrho)}{1-\varrho},$$

whereas by (2.71),

$$b^{(2)}(\varrho) = \varrho^2\frac{\mathrm{d}^2 b_1(\varrho)}{\mathrm{d}\varrho^2} + b_1^{(1)}(\varrho) = \frac{\varrho - \varrho\log(1-\varrho)}{(1-\varrho)^2}.$$

Further, note that

$$\begin{aligned}
b_2(\varrho) &= \sum_{i=2}^{\infty}\sum_{j=1}^{i-1}\frac{\varrho^{i+j}}{ij} = \sum_{i=2}^{\infty}\frac{\varrho^i}{i}\sum_{j=1}^{i-1}\frac{\varrho^j}{j} = \sum_{i=2}^{\infty}\frac{\varrho^i}{i}\int_0^{\varrho}\sum_{j=1}^{i-1}x^{j-1}\,\mathrm{d}x \\
&= \sum_{i=1}^{\infty}\frac{\varrho^{i+1}}{i+1}\int_0^{\varrho}\frac{1-x^i}{1-x}\,\mathrm{d}x \\
&= -\log(1-\varrho)\left(\sum_{i=1}^{\infty}\frac{\varrho^i}{i} - \varrho\right) - \int_0^{\varrho}\left(\sum_{i=1}^{\infty}\frac{\varrho^i}{i}\frac{x^{i-1}}{1-x} - \varrho\frac{1}{1-x}\right)\,\mathrm{d}x \\
&= -\log(1-\varrho)\sum_{i=1}^{\infty}\frac{\varrho^i}{i} - \int_0^{\varrho}\sum_{i=1}^{\infty}\frac{(\varrho x)^i}{i}\frac{1}{x(1-x)}\,\mathrm{d}x \\
&= \log^2(1-\varrho) + \int_0^{\varrho}\frac{\log(1-\varrho x)}{x(1-x)}\,\mathrm{d}x
\end{aligned}$$

$$= \frac{1}{2}\big(\log^2(1-\varrho) - \mathrm{Li}_2(\varrho^2)\big), \tag{2.76}$$

where the last equality follows from Lemma A.1. Next, by (2.69), (2.74) and (2.76) we have

$$b_2^{(1)}(\varrho) = \log\left(1-\varrho^2\right) - \frac{\varrho\log(1-\varrho)}{1-\varrho}.$$

Thus, we can apply Lemma 2.5(i) and arrive at the following expression for κ_1:

$$\kappa_1 = \frac{\pi^2}{6} + \log^2(1-\varrho) + 2\mathrm{Li}_2(\varrho). \tag{2.77}$$

Similarly, for κ_2, by collecting and simplifying the terms, by Lemma 2.5(ii) and Lemma A.1, we have

$$\begin{aligned} \kappa_2 &= \frac{1+\varrho^2}{1-\varrho^2}\Big(\frac{\pi^2}{6} + 2\mathrm{Li}_2(\varrho)\Big) - \frac{2\log(1-\varrho)}{1-\varrho} + \log^2(1-\varrho)\frac{\varrho^2}{1-\varrho^2} \\ &= \frac{1}{1-\varrho^2}\big((1+\varrho^2)\kappa_1 - \log^2(1-\varrho) - 2(1+\varrho)\log(1-\varrho)\big). \end{aligned} \tag{2.78}$$

Lastly, for κ_3, by Lemma 2.5(iii), through simplification of terms, we get

$$\begin{aligned} \kappa_3 &= \frac{1}{(1-\varrho^2)^2}\Big((1+4\varrho^2+\varrho^4)\Big(\frac{\pi^2}{6} + 2\mathrm{Li}_2(\varrho)\Big) + \log^2(1-\varrho)\varrho^2(1+\varrho^2) \\ &\quad - (3-\varrho+4\varrho^2)(1+\varrho)\log(1-\varrho) + \varrho(1+\varrho)^2\Big) \\ &= \frac{1}{(1-\varrho^2)^2}\Big((-1+\varrho+2\varrho^2)(1+\varrho)\log(1-\varrho) + \varrho(1+\varrho)^2 - 2\varrho^4\kappa_1\Big) \\ &\quad + \kappa_2\,\frac{1+3\varrho^2}{1-\varrho^2}. \end{aligned} \tag{2.79}$$

This allows us to apply Theorem 2.2 under the considered specification of the parameter β, and conclude with the following corollary.

Corollary 2.2. *Assume a model* (1.1) *with* (2.1) *covariance structure and consider* $\beta_j := j^{-1}$, $j = 1, \ldots, p$. *Let* $p = p_n$ *satisfies*

$$p \to \infty, \qquad \frac{p}{n} \to c \in (0, \infty).$$

Then

$$\frac{\|\mathbb{X}'Y\|_2^2 - n^2\big(\kappa_2 + c(\kappa_1+\sigma_\varepsilon^2)\big)}{n^{3/2}} \overset{d}{\to} \mathcal{N}(0,s^2), \tag{2.80}$$

where

$$s^2 = 4\kappa_2^2 + 4(\kappa_1+\sigma_\varepsilon^2)\,(2\kappa_2 c + \kappa_3) + 2c(\kappa_1+\sigma_\varepsilon^2)^2\Big(c + \frac{1+\varrho^2}{1-\varrho^2}\Big), \tag{2.81}$$

and κ_1*,* κ_2 *and* κ_3 *are defined by* (2.77)*,* (2.78) *and* (2.79)*, respectively.*

In order to illustrate the results of Corollary 2.2, we end this section with a Monte Carlo simulation study, where we generate 1000 independent replications of the statistic $\|\mathbb{X}'Y\|_2^2$. The data is generated following the assumptions of Corollary 2.2. We consider the following parameter values: $p = 100, 500, 1000, 1500, 2000, 3000$, $c = 1, 2, 5, 10$, $\sigma_\varepsilon^2 = 1, 2, 4, 10$. Due to the large number of resulting figures, we present only selected cases in Figures 2.1–2.9, which demonstrate certain disparities in greater detail. Figures show the empirical cumulative distribution function (CDF) and the empirical probability density function (PDF), together with the limiting CDF and PDF of $\mathcal{N}(0,s^2)$ by (2.80) for different parameter combinations. In addition, we present the corresponding Q-Q plots in order to inspect the tails of the resulting distributions in greater detail.

We find that for relatively small values of ϱ, the observed distribution of the statistic is fairly close to the limiting distribution even for small values of p, n and larger σ_ε^2, c (see, e.g., Figures 2.1–2.4). However, slower convergence is more evident with increasing values of ϱ. Furthermore, for moderate values of $\varrho, c, \sigma_\varepsilon^2$, only with larger values of p we observe adequate convergence towards the limiting distribution (see Figures 2.5–2.6). Similar behaviour is observed when the relation between the parameters $\varrho, c, \sigma_\varepsilon^2$ is appropriately controlled: e.g., in Figure 2.7, we see comparable results to those presented by Figure 2.6, where the effect of the increase of parameter value ϱ is countered by a smaller value of σ_ε^2. Alternatively, analogous effects can be achieved when reducing the values of c, instead.

Finally, slow convergence is observed for large values of $\varrho, c, \sigma_\varepsilon^2$, as expected (see Figures 2.8–2.9). In such cases, the simulation results suggest that even larger values of p, n would be needed for more accurate results.

2.6 Discussion

In this chapter we considered a specific KMS covariance structure due to its attractive properties and wide application possibilities for working with real world datasets. Moreover, our results could be extended further by considering a wider family of Toeplitz covariance structures. For instance, under specific constraints, one could employ the approaches proposed in Yang et al. (2021) in order to extend the application of our results towards more complex covariance structures of the data.

Furthermore, for future work, it would be interesting to expand and examine the results by removing the assumption of independence between the observations X_i, $i = 1, \ldots, n$.

Finally, we established both the exact and the asymptotic distributions of the statistic $\|\mathbb{X}'Y\|_2^2$ (see (2.34) and (2.7), (2.9)). Both distributions could be used for estimating β, σ_ε^2 or related measures (e.g., by applying the method of moments or maximum likelihood estimation) in future research. Such research direction could open up interesting avenues when compared with popular LASSO type methods in high-dimensional linear regression. Similar approach is taken by Dicker and Erdogdu (2016), who construct maximum likelihood estimators for the signal strength $\|\beta\|_2^2$ in a high-dimensional regression context. Note that the results by Dicker and Erdogdu (2016) are achieved under certain strong restrictions, which are consistent with the related literature (see, e.g., Bayati et al. (2013); Dicker (2014); Janson et al. (2017)). In our case, we impose weaker assumptions; therefore, both asymptotic and exact results could be used in order to extend the approaches in the aforementioned literature.

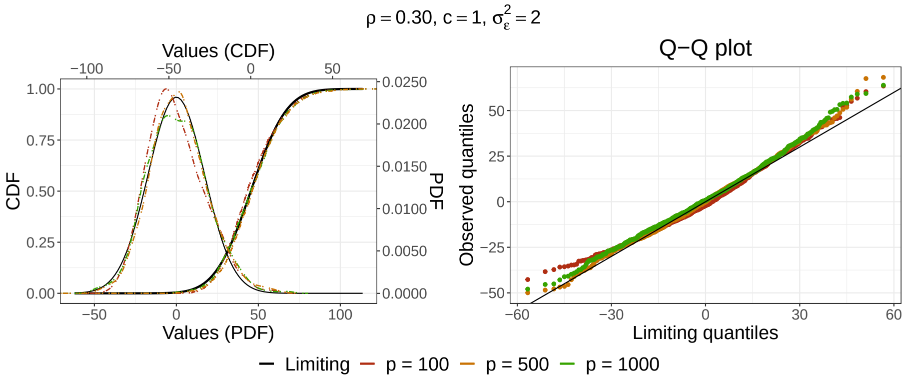

Figure 2.1: Comparison of the PDF and CDF (left) and the corresponding Q-Q plots (right) after 1000 replications from the Monte Carlo simulation of the statistic (2.80) with the limiting distribution $\mathcal{N}(0, s^2)$ by Corollary 2.2 (in black) for $\varrho = 0.3$, $c = 1$, $\sigma_\varepsilon^2 = 2$ and $p = 100, 500, 1000$.

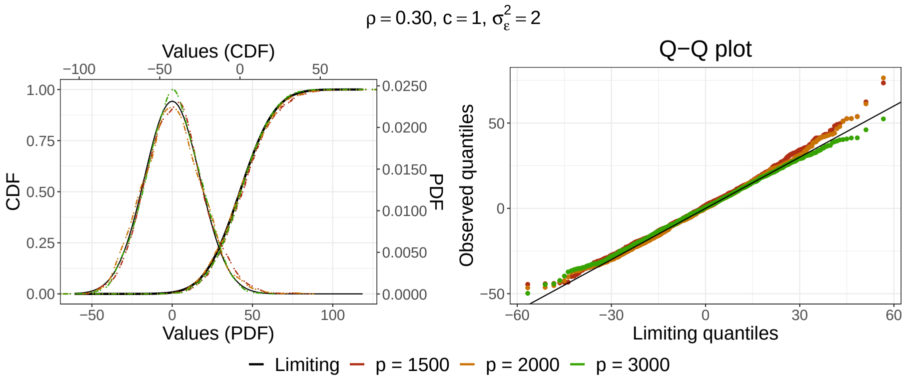

Figure 2.2: Comparison of the PDF and CDF (left) and the corresponding Q-Q plots (right) after 1000 replications from the Monte Carlo simulation of the statistic (2.80) with the limiting distribution $\mathcal{N}(0, s^2)$ by Corollary 2.2 (in black) for $\varrho = 0.3$, $c = 1$, $\sigma_\varepsilon^2 = 2$ and $p = 1500, 2000, 3000$.

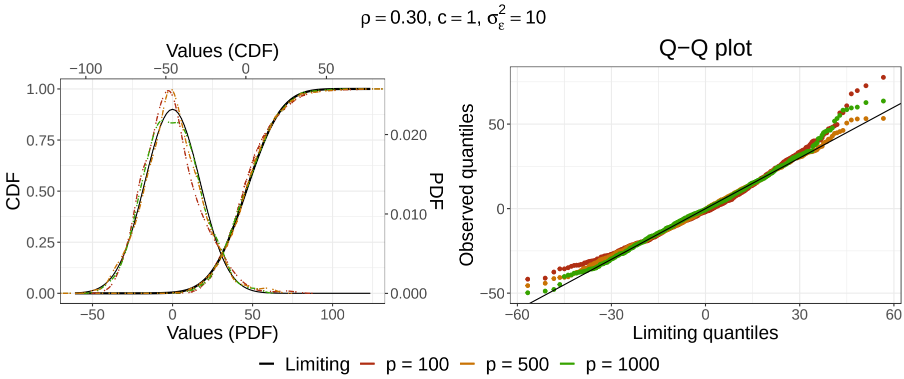

Figure 2.3: Comparison of the PDF and CDF (left) and the corresponding Q-Q plots (right) after 1000 replications from the Monte Carlo simulation of the statistic (2.80) with the limiting distribution $\mathcal{N}(0, s^2)$ by Corollary 2.2 (in black) for $\varrho = 0.3$, $c = 1$, $\sigma_\varepsilon^2 = 10$ and $p = 100, 500, 1000$.

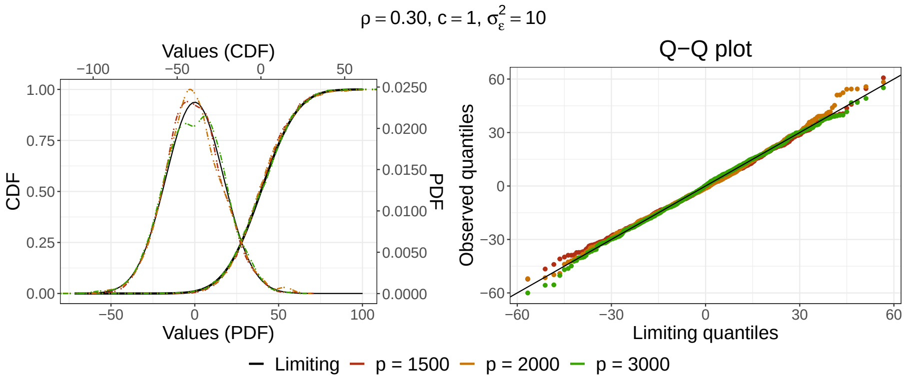

Figure 2.4: Comparison of the PDF and CDF (left) and the corresponding Q-Q plots (right) after 1000 replications from the Monte Carlo simulation of the statistic (2.80) with the limiting distribution $\mathcal{N}(0, s^2)$ by Corollary 2.2 (in black) for $\varrho = 0.3$, $c = 1$, $\sigma_\varepsilon^2 = 10$ and $p = 1500, 2000, 3000$.

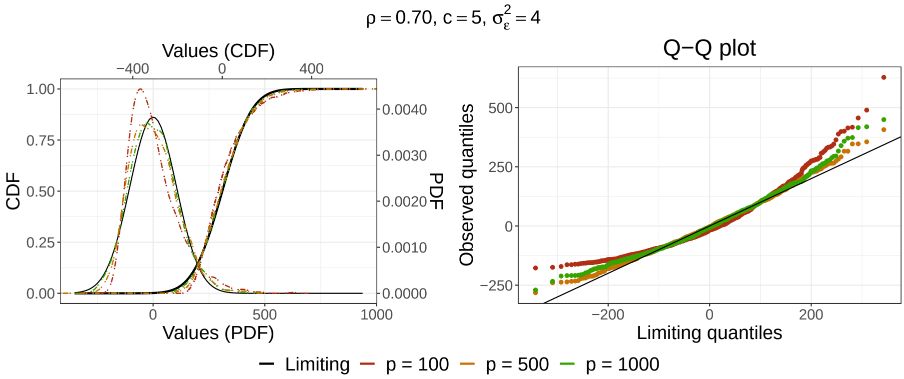

Figure 2.5: Comparison of the PDF and CDF (left) and the corresponding Q-Q plots (right) after 1000 replications from the Monte Carlo simulation of the statistic (2.80) with the limiting distribution $\mathcal{N}(0, s^2)$ by Corollary 2.2 (in black) for $\varrho = 0.7$, $c = 5$, $\sigma_\varepsilon^2 = 4$ and $p = 100, 500, 1000$.

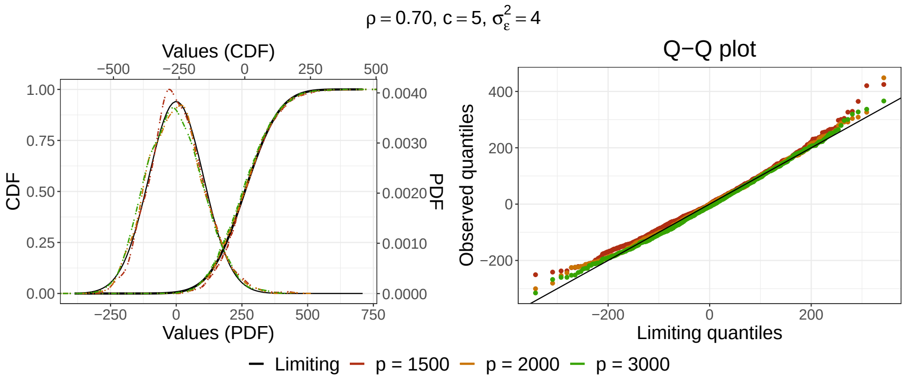

Figure 2.6: Comparison of the PDF and CDF (left) and the corresponding Q-Q plots (right) after 1000 replications from the Monte Carlo simulation of the statistic (2.80) with the limiting distribution $\mathcal{N}(0, s^2)$ by Corollary 2.2 (in black) for $\varrho = 0.7$, $c = 5$, $\sigma_\varepsilon^2 = 4$ and $p = 1500, 2000, 3000$.

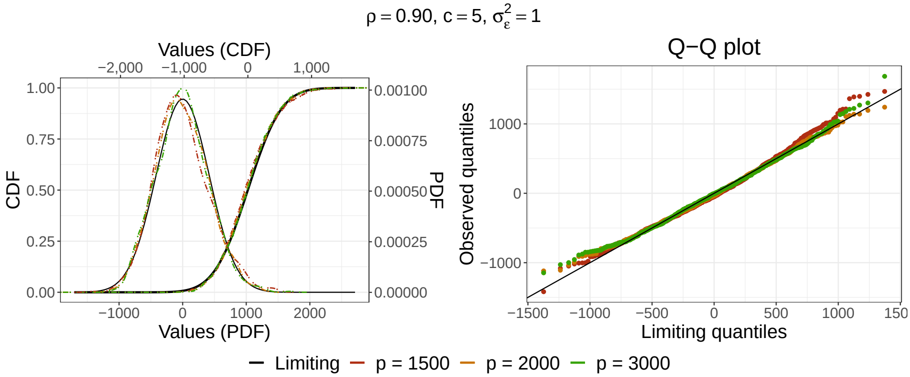

Figure 2.7: Comparison of the PDF and CDF (left) and the corresponding Q-Q plots (right) after 1000 replications from the Monte Carlo simulation of the statistic (2.80) with the limiting distribution $\mathcal{N}(0, s^2)$ by Corollary 2.2 (in black) for $\varrho = 0.9$, $c = 5$, $\sigma_\varepsilon^2 = 1$ and $p = 1500, 2000, 3000$.

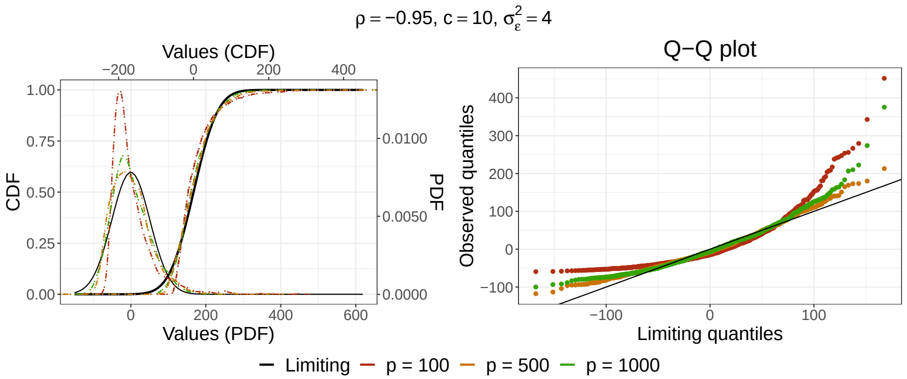

Figure 2.8: Comparison of the PDF and CDF (left) and the corresponding Q-Q plots (right) after 1000 replications from the Monte Carlo simulation of the statistic (2.80) with the limiting distribution $\mathcal{N}(0, s^2)$ by Corollary 2.2 (in black) for $\varrho = -0.95$, $c = 10$, $\sigma_\varepsilon^2 = 4$ and $p = 100, 500, 1000$.

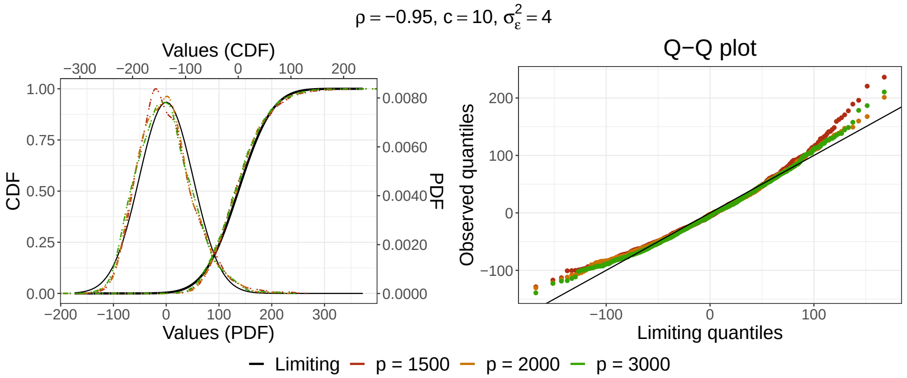

Figure 2.9: Comparison of the PDF and CDF (left) and the corresponding Q-Q plots (right) after 1000 replications from the Monte Carlo simulation of the statistic (2.80) with the limiting distribution $\mathcal{N}(0, s^2)$ by Corollary 2.2 (in black) for $\varrho = -0.95$, $c = 10$, $\sigma_\varepsilon^2 = 4$ and $p = 1500, 2000, 3000$.

Chapter 3

Sparse structures: nowcasting US and EU GDP components

In this chapter, we present a pseudo-real-time short-term forecasting experiment, where we focus on the main GDP expenditure components of US and selected EU countries. The goal is to investigate the assumption of sparse structures by comparing and examining the forecasting performance of the popular sparse methods, briefly presented in Chapter 1, with univariate ARMA models and well known dense alternatives, that are most commonly found in the nowcasting literature.

In Section 3.1 we describe the importance of accurate nowcasting of macroeconomic variables and introduce a variety of different approaches typically used by institutions. We highlight the possible benefits of sparse linear models in contrast to widely used dense approaches. In Section 3.2 we describe the algorithm and the motivation behind the proposed combination of LASSO and principal components methods. Furthermore, we argue that the method can be extended further by tailoring or adjusting the rotation matrix in some specific way, allowing further improvements in the forecasting performance. In Section 3.3 we describe the settings and methodology behind the pseudo-real-time experiments, the results of which are presented in Section 3.4.

3.1 Nowcasting and sparsity

Information on the current state of the economy is crucial for various economic agents and policymakers since the choice of the appropriate policy measures relies on the current knowledge of the macroeconomic situation in the country. Although there are some higher frequency indicators, covering many aspects of the economy, quarterly national accounts still play a pivotal role in guiding economic policy decisions. Unfortunately, the official release of GDP and its components occurs with a considerable delay after the reference period. For instance, the flash estimates of the US and EU GDP are released one month after the reference period, where only the supply side, i.e. production of goods and services, is covered. While the demand side reflected in GDP accounted by expenditure approach is released with an even longer two months delay. Bearing in mind economic policy implementation lags, the apt nowcasting and short-run forecasting tools are crucial for the regular monitoring and efficient countercyclical policy implementation. The latter is vitally important during severe downturn times like the global financial crisis of 2007–2008 followed by the sovereign debt crisis in the EU, and the COVID-19 led global lockdown.

At the same time, the agencies of national statistics announce various short-term monthly indicators, such as business or consumer surveys, the industrial production, retail or external trade indicators with a much shorter delay. The monthly data allow getting a preliminary picture of how the current economic activity evolves in various sectors of the economy. One way of using higher frequency information is by conducting the fundamental analysis. However, the quarterly and annual national accounts data remain the core in guiding most of the economic policy decisions, e.g. in the area of fiscal rules. The need to extract the underlying signals from the available higher frequency data and obtain the lower frequency national accounts data sooner than the official release by the agencies of national statistics led to the development of several econometric tools examined in this chapter.

The literature suggests many different methods for bridging the higher and lower frequency data, the mainstream of which are the factor models and their modifications (Bańbura et al. (2013)) altered to use all available information. Therefore, such methods possess a density feature. On the other hand, the sparsity feature of the data hints that only a small subset of all the available information might be sufficient for adequate and timely estimation of the GDP or its components (Bai and Ng (2008)).

Throughout this chapter, we further explore the sparse structure approach under a large amount of monthly economic indicators, assuming that only a small subset of them is significant and valuable in forecasting macroeconomic variables. Therefore, the underlying problem is the optimal selection of useful predictors. Bai and Ng (2008) find that large amounts of high-frequency data could often considerably worsen predictive performance, raising a question of how much of high-frequency data required for good predictions? On top of that, Bulligan et al. (2015) among others show that assuming a sparse structure of the underlying data improves the quality of short-run forecasts when the final set of available information is refined by supervised selection before the application of bridge equations modelling approach.

Recently a rapid growth in popularity among the practitioners and the academics is seen applying LASSO and similar sparse methods. In Section 1.3 we introduced and briefly described several well known approaches, which will be considered throughout this chapter. In addition, in Section 3.2 we propose a combination of the LASSO with the method of principal components seeking to extract the significant underlying information with greater accuracy. We evaluate the empirical performance of the models by conducting pseudo-real-time short-run forecasting experiments of real, in chain-linked volumes sense, GDP expenditure components of these 5 economies: US, France, Germany, Spain and Italy. We forecast the following GDP components: the Gross Fixed Capital Formation (GFCF), Private Final Consumption Expenditure (PFCE), Imports and Exports of goods and services. During the validation exercise, we estimate forecasts at four different forecasting horizons: the backcast of a previous quarter, the nowcast of the coinciding quarter and 1- and 2-quarter forecasts over the sample of 2005–2019. The results are presented in Section 3.4. We find evidence of further improvement of the forecasting performance while applying the proposed LASSO-PC method.

The motivation for looking at the demand breakdown and not the GDP itself is twofold. First, there is evidence in the literature (Heinisch and Scheufele (2018)), that forecasting GDP by the bottom-up approach can lead to more accurate forecasts than direct forecasts. This claim follows from the fact that we can examine the different underlying and sparse structures of the subcomponents by modelling them separately.

For example, investments and international trade are much more volatile than aggregate GDP, while private consumption is typically smoother than total activity (Artis et al. (2004)). Therefore, in this work, we aim to study how do different models compare in forecasting variables that are behaving so differently over the business cycle. Second, it is evident that the business cycle behaviour of the aggregate GDP is very different from that of its subcomponents. For example, investments tend to trough before GDP, while consumption only takes momentum when an expansion is well underway, peaking only after the cycle (Castro (2010)). 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. Recovering information on the main drivers of the economic activity by itself may allow for a more accurate read on the cyclical phase by the economic agents. Additionally, forming a different model for each of the subcomponents admits the inclusion of different sets of predictors providing a richer story behind the specified equation.

3.2 Principal components and LASSO

In this section, we propose a combination of variants of LASSO (for a detailed review see Section 1.3) together with principal components in order to preserve specific strengths and to minimize the possible shortcomings for each of the methods combined.

We follow the arguments of Bai and Ng (2008), who show that the use of targeted predictors with factor models helps achieve significantly better forecasts of macroeconomic data. Instead of the usual approach to factor model forecasting, where the principal components follow from the full data set, the authors suggest using only a subset selected by a chosen hard/soft thresholding algorithms. In this way, an unsupervised algorithm becomes supervised one, because the choice of the targeted predictors now depends on the predicted variable. Following these arguments, we propose employing the LASSO, Adaptive LASSO and Square-Root LASSO for variable selection. These modifications deemed to more accurately select variables due to their known asymptotic properties under $p \gg n$ allowing for many highly correlated variables. From here on, let us assume that $\mathbb{X}^{\lambda} \in \mathbb{R}^{n \times q}$ is a preselected matrix of significant variables under a fixed λ by some variant of LASSO, where $0 < q \leq n$, and the variables are scaled and centered.

Following these arguments, we present the proposed algorithm in Section 3.2.1, and follow with the explanation of the motivation behind each step in Section 3.2.2. In Section 3.2.3 we discuss further fine-tuning of the proposed algorithm.

3.2.1 LASSO-PC

We propose the following six-step algorithm:

1. Use LASSO or some of its variant[1] to select a subset of candidate indicators $\mathbb{X}^\lambda \in \mathbb{R}^{n\times q}$ from the full dataset $\mathbb{X} \in \mathbb{R}^{n\times p}$, under a certain fixed λ, were $0 < q \leq n$, and the data is assumed as scaled and centered.

2. Rotate $\mathbb{X}^\lambda$ by principal components to $F = \mathbb{X}^\lambda L$, where $L \in \mathbb{R}^{q\times q}$ is a rotation matrix and $F \in \mathbb{R}^{n\times q}$ is a principal components matrix.[2]

3. Forecast each series in $\mathbb{X}^\lambda$ by an appropriate ARIMA model. In our case, we forecast at the monthly frequency, and forecasts are aggregated back to quarterly frequency.

4. Aggregate time series forecasts to their factor representation using the rotation matrix L, i.e., $\hat{F}^*_{n+h} = \hat{X}^\lambda_{n+h} L$, for $h = 1, 2, \ldots$.

5. Use LASSO to estimate the coefficients in $Y_n = \hat{F}_n \beta + \varepsilon_n$.[3]

6. Produce the forecasts $\hat{Y}_{n+h} = \hat{F}_{n+h}\hat{\beta}$.

3.2.2 Motivation

Since we are interested in modelling macroeconomic data, it is likely that significant correlation will be observed, with some of the variables

[1]The interested reader may wish to try combining additional methods for the better choice of candidate variables, e.g., by employing the Rolling Window variable selection approach by Bulligan et al. (2015), or similar.

[2]Besides the classical PCA, one may consider alternatives, e.g., by fine-tuning the angles or scale of the rotation (briefly discussed in Section 3.2.3), or using another viable alternative, like Independent Component Analysis (see, for example, Kim and Swanson (2018)).

[3]Further, one may want to consider expanding the estimation by estimating the coefficients in $Y_n = (\hat{F}_n, X^\lambda_n)\beta + \varepsilon_n$. The motivation behind this step is discussed in Section 3.2.3.

possibly even being nested. Therefore, our proposed modification stems from the initial variable selection step of the Relaxed LASSO. However, instead of a direct parameter re-estimation of the selected variables, as would be done on the second step of the Relaxed LASSO, we propose rotating the data using the principal components methodology. In other words, we propose extracting the main latent factors $F = \mathbb{X}^{\lambda}L$, where $L \in \mathbb{R}^{q\times q}$ is a rotation matrix and $F \in \mathbb{R}^{n\times q}$ is a principal components matrix.

The main idea here is to extract the underlying information from the data as orthogonal latent factors and to model them instead. Due to probable inter-correlation and the supervised preselection done in the first step – it is likely that such data will capture prevalent 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 include them directly instead of the original data.

As for the parameter estimation, we expand on the idea, defined by equation (1.9), and estimate:

$$\begin{aligned}\hat{\beta}_{\text{LASSO-PC}} &= \arg\min_{\beta} \, (Y - \mathbb{X}^{\lambda}LL'\beta)'(Y - \mathbb{X}^{\lambda}LL'\beta) + \lambda||L'\beta||_1 \\ &= \arg\min_{\tilde{\beta}} \, (Y - F\tilde{\beta})'(Y - F\tilde{\beta}) + \lambda||\tilde{\beta}||_1, \qquad (3.1)\end{aligned}$$

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

Our proposed approach differs from the one suggested by Bai and Ng (2008), first of all, because the number of significant factors is selected not by the usual selection, based on various information criteria (such as *Akaike*, *Schwarz*, t-statistics from OLS and similar), but by using the LASSO approach. That is, both the selection of significant factors and the shrinkage of estimated parameters applied simultaneously deemed to improve the forecasting accuracy, which is the main objective of short-run forecasting.

The main strength of such an approach is when dealing with a large amount of data, driven by one or only a few leading factors. In that case, the principal component transformation would allow us to extract those latent factors and estimate only the predictive ones using the LASSO. The final coefficient vector $\hat{\beta}^* := \hat{\hat{\beta}}L$ would be comprised of the same non-zero variables just as it would be in the classic (Relaxed or Adaptive) LASSO case, but the estimated parameters would be set according to the significance and predictive performance in the latent space, rather than directly. Noteworthy, such transformation would act as a filter, distinguishing the essential underlying signals from the data and possibly allowing for more accurate forecasting performance.

Second, in contrast to Bai and Ng (2008) and other similar factor forecasting related literature, we propose to base the final forecasts on the predictions of individual variables X_j^λ, $j = 1, \ldots, q$, rather than on the predicted significant factors F_j. That is, if we denote $X_t^\lambda = (X_{t,1}^\lambda, \ldots, X_{t,q}^\lambda)$ as the data at a time moment $t \in \{1, \ldots, n\}$, then for every $h > 0$, the forecasts $\hat{F}_{n+h}^* = (\hat{F}_{n+h,1}, \ldots, \hat{F}_{n+h,q})$ can be calculated as

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

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

Such aggregation might induce a smaller loss in forecast accuracy of factors F when we fail to accurately define a direct model. First, forming the extracted factors with weights of a similar size, such forecast aggregation is equivalent to *bagging* (bootstrap aggregation as in Breiman (1996a)). Second, it is possible, that the data generating process of F_j might be from some family of complex, long memory processes. Therefore, the aggregation of forecasts of simpler models of individual components introduces some degree of freedom to make inaccurate estimations, while still generating more accurate final predictions. Indeed, Granger (1980) has shown that the aggregation of a low-order AR/ARMA processes in particular cases may produce a process with more complex dynamics. Extending these ideas to the aggregation of the forecasted series allows recovering such complex dynamics in the final forecasts of the original factors. We demonstrate the possible gains from such approach in this section by comparing the resulting forecasts both with direct factor forecasts (see, e.g., Table 3.5) and Kalman filter forecasts, used with the dynamic factors (see, e.g., Table 3.3 and similar).

Noteworthy, when the variables X_j are orthogonal, due to the properties of Principal Components, the transformation would reduce to the base LASSO performance, since the total amount of information would remain unchanged.

3.2.3 Tailoring the rotation of principal components

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

Scale modification might help the LASSO distinguishing the most important signals since the components with larger-scale will enter the solution path earlier than the other, smaller components. This idea is discussed by Stakėnas (2012), who observes significant improvement when using Weighted PCA or Generalized PCA for the extraction of factors when nowcasting Lithuanian GDP.

Alternatively, in this section, we discuss the idea of angle modification by sparsifying the rotation matrix, e.g., through Sparse PCA (Zou et al. (2006)). The idea here is that we may achieve a more accurate factor representation by losing the orthogonality of the transformation. The main motivation is that, even though the data matrix $\mathbb{X}^\lambda$ is preselected by the LASSO as a matrix, containing mainly significant variables, it is not clear that by rotating the variables to the latent space, all of them will be significant there. In other words, if there are two strongly correlated variables preselected by LASSO as significant, both having roughly the same estimated weights (possibly with different signs) in the rotation matrix, it is likely that losing one of the two dimensions might not change the predictive power of the resulting factor estimate.

Further, some of the variables denoted in what follows as Z, where $Z \subset \{X_j : j = 1, \ldots, q\}$, 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 standard principal components that is mostly not the case. Every extracted factor is a linear combination of all the variables used, even if the weights are close to zero, it's unlikely for them to be exactly zero. Let us assume that the model matrix $\mathbb{X}^\lambda$ is reordered such, that the block of variables $Z \in \mathbb{R}^{n \times q_0}, q_0 < q$, is able to capture significant explanatory signals to our modelling problem, orthogonal to the remaining block $\bar{\mathbb{X}}$, the LASSO will try to extract as much information as possible from those variables. Since the extracted factor matrix has the following structure:

$$F \;=\; \mathbb{X}^\lambda L \;=\; (Z,\ \bar{\mathbb{X}})L,$$

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

$$f_j = (Z,\ \bar{\mathbb{X}})L_j \;=\; (Z,\ \bar{\mathbb{X}})[\Lambda_{q \times q_0}, \Phi_{q \times (q-q_0)}]'_j.$$

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

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

3.3 Preliminaries: data preparation

In this chapter, we consider the four main components of the GDP by the expenditure approach: Gross Fixed Capital Formation (GFCF), Private Final Consumption Expenditure (PFCE), Imports and Exports of goods and services, all of which are quarterly and seasonally adjusted by the source.

The monthly data used as explanatory variables are various openly available indicators from the databases of St. Louis Bank of Federal Reserves (FRED) and Eurostat from 1990 to 2019, with up to 2400 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. Avoiding the problem of spurious regression, every time series were transformed to stationary and appropriately adjusted for normality and against additive outliers. Dealing with a large number of variables and relying on automatic algorithms for variable selection without any prior expectations, strict rules and conservative approach for data preparation allows ruling out as much as possible noise (see Appendix B.1 for specific heuristics and tests applied).

Comparing the forecasting performances of different models in a realistic setting, during the pseudo-real-time experiments, the results of which are presented in Section 3.4, 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 (see Table 3.1 for detailed illustration). During a full quarter, at least three updates on the dataset are possible for every different month of the quarter. However, in this study, the results presented are of the last month of the full quarter. Since we are also interested in nowcasting performance, analysing results of 3rd month helps to separate it from forecasting due to available indicators with low publication lag. Otherwise, we would just be comparing the predictive performance of the ARIMA models, used for the individual predictions of the selected monthly indicators, which is already inspected using 1- and 2-quarter forecasts.

The monthly variables used in the dataset were aggregated to quarterly by averaging, with up to four quarterly lags included, and

Table 3.1: The timing and data availability framework for the pseudo-real-time forecasting exercise

Target	GFCF	Backcast, Q1			Nowcast, Q2			Forecast, Q3			Forecast, Q4			Quarter
		1	2	3	4	5	6	7	8	9	10	11	12	Month
Group	Industrial Production													
	Surveys													
	Interest Rates													
	Stock Market Indexes													
	Trade													

Legend: Available data: Month of forecast: Forecasted data:

the ragged edges of the dataset were filled by using the ARIMA time series methods.

3.4 Pseudo-real-time forecasting experiments

This section presents the results of pseudo-real-time forecasting exercise over 2005Q1–2019Q1. The main goal is to gauge the real-world performance of the sparse methods in a high-dimensional environment. As discussed in Section 3.3, we consider as much macroeconomic data available as possible (potentially including some noise variables), and are interested in 4 specific forecast horizons: one backcast, one nowcast and two predictions of 1- and 2-quarters ahead. We focus on the main four components of the GDP by expenditure approach (GFCF, PFCE, Imports and Exports of goods and services) for five economies: the US, France, Italy, Spain and Germany. We present the main results and important discussion points using the US data, while the European countries are presented as a robustness check to inspect how well do the results translate between the different datasets.

In sum, the expenditure components reflect the main drivers of the economy: the domestic demand, mainly consisting of private consumption, investments and imports, and the foreign demand defined by exports. The international trade reflects the openness of the economy and drives international competitiveness that forces the search for new innovative and advanced solutions.

Out of these four variables, the most difficult to accurately predict is the GFCF composed of investments in many different industries. The investment spending deemed to be forward-looking and optimistic, expanding rapidly during an economic boom when investors expect that the future will require the extensive productive capacity, and falling fast when such expectations evaporate. Therefore, it is the most volatile of the four variables. For private consumption, imports and exports, there are "hard" coinciding monthly indicators available, allowing for an easier nowcast. However, such variables are absent for the GFCF making good nowcasts a more challenging task. Therefore, our primary focus in this section is forecasting the GFCF, presented in greater detail. The forecasting results of the three remaining variables are then briefly discussed.

The analysis aims to inspect the model robustness to the changing nature of the economy, testing if the models can reflect and respond to possible structural changes, as the economy evolves and matures. Therefore, we expect maximum flexibility by using all of the available data. Further, we seek to test the signal recovery and identification of the sparse structures. On the one hand, we allow the inclusion of possible noise variables in the datasets, for example, various indicators of foreign economies. On the other hand, this inclusion provides the possibility for the sparse models to recover significant signals from large trade partners and dependent parties without specific offsets imposed by the analyst. This exercise may present itself as a perfect tool for additional variable consideration to the proprietary datasets, likely enhancing any existing models with otherwise omitted stories, as we see later in this chapter.

3.4.1 Setting up the experiment

For every modelled component of the GDP, we consider the following models and their corresponding notations: the Square-Root LASSO (in the tables denoted as SqL), Relaxed LASSO (RL), Adaptive LASSO (AdL), classic LASSO (L), the L0Lq best subset selection models (L0, L0L1, L0L2) and a proposed combination LASSO-PC, where the data is preselected by a specific variant of the LASSO, i.e., Adaptive LASSO-PC (AdP), L0-PC (L0P), L0L1-PC (L0L1P), L0L2-PC (L0L2P), Square-Root LASSO-PC (SqP) or the LASSO-PC (LP), and similar, as described in Section 3.2. In what follows, the letter "P" is always appended to any model abbreviation where the PCA transformation is applied.

Seeking to inspect the gains brought solely by performing the rotation of the data to its principal components, we analyse the alternative cases for AdP models fixing the preselected variables without the rotation to the principal components. In a way, it becomes a mix between Relaxed LASSO and Adaptive LASSO, denoted as AdRL in the tables.

Additionally, in cases when there were significant gains in forecasting accuracy when using both the rotated and original data, we interpret such cases as a cross between the LASSO and LASSO-PC methods denoting them as LASSO-PC-X (LPX in the tables).

Each LASSO model is estimated using the cross-validated hyperparameters unless specified differently, where each hyperparameter is chosen to maximize the out-of-sample accuracy (Chetverikov et al. (2021)). Besides, the weights for Adaptive LASSO are chosen using Ridge estimates (as discussed in Section 1.3.2) with $\gamma \in \{0.001, 0.01, 0.1, 0.2, 0.3, 0.5, 1, 1.5, 2\}$ selected by cross-validation (Zou (2006)). The idea behind adding values of γ very close to 0 is to allow the Adaptive LASSO optimize under weights that can be close to the LASSO.

As benchmarks for the analysed models we use ARMA, where the orders are selected to minimize the corrected *Akaike* information criterion (AICc, Hyndman and Khandakar (2008)) during each quarter of the exercise. Additionally, we consider a few variants of factor models as good *dense* structure alternatives based on their known performance in the literature. Namely, we consider the static Diffusion-Index models (DI, Stock and Watson (2002)) and Dynamic Factor models (DF, Giannone et al. (2008), Banbura et al. (2011)), see Appendix, Section B.2.1 for a detailed model specification. For the dynamic factor estimation we use the two-step Kalman Filter realization with quarterly aggregation scheme (Angelini et al. (2011)). Additionally, for the DF models, we consider the targeted predictor approach of Bai and Ng (2008) (denoted in the tables as BNDF), i.e., instead of full model matrix, we preselect the candidate variables using the LASSO. Lastly, for more clarity and better comparison, during each iteration of the exercise, if some variant of LASSO was used for initial variable selection, e.g., for LASSO-PC, Relaxed LASSO and BNDF, the same variables will be used by these models. This was done to eliminate the uncertainty from randomness in variable selection. For all factor models, the number of significant factors was estimated based on the information criterion, defined by Bai and Ng (2002). Additionally, for the DF models, the number of quarterly factor lags is fixed to 2, while the number of shocks is estimated by information criterion, following the Bai and Ng (2007). See Appendix, Section B.2.3 for a detailed specification regarding the selection of aforementioned parameters.

Besides, we present the results of models, where, instead of the cross-validated hyperparameters, we use the fixed number of selected variables that provide a more parsimonious result. In those cases the results are presented with an additional number next to their abbreviation, e.g., RL15 in Table 3.4, indicating that the model is selecting top 15 significant variables during the exercise following Bulligan et al. (2015), while Bai and Ng (2008) claim that practical choice of top predictors should not exceed 30. It is worth inspecting such results since the optimally selected hyperparameter can result in a model of a too dense structure. In this case, the LASSO wins a small amount of validation accuracy but brings an additional forecast uncertainty with each of the extra variable included.

During the forecasting exercise, the subsets of variables are reselected for each analysed model and every quarter when the new pseudo-vintage of data is formed. Instead of tailoring the set of possible variables to match the predicted variable, we allow methods themselves to select the significant covariates. Likely, some indicators that drove the growth of particular markets in the past are not significant at a later stage. Hence, these past drivers can be replaced on time with new indicators reflecting new products and services. For this reason, we study two different approaches to forming the training sample: an expanding window approach and a rolling window approach.

For the expanding window experiments, we use the sample from 1992Q1 to 2004Q4, the window is expanded by adding one additional quarter during every iteration. The size of the window was chosen to be moderate so that there would be an appropriate amount of available historical monthly indicators, but sufficient for the methods to select a large number of significant variables if needed. Note, that under such an approach it is likely that some of the variables are capturing the historical dynamics of the modelled series early in the training sample rather than of the more recent events. Therefore, with a rolling window approach, we add some control over possible structural breaks (Bulligan et al. (2015)). We chose a 13-year rolling window to capture at least one full business cycle (see, e.g., Economic Cycle Research Institute (ECRI), de Bondt and Vermeulen (2018)): starting from 1992Q1 to 2004Q4, the window was rolled by adding a quarter to both the start and the end of the period during every iteration.

To indicate how well did the models forecast, we present the ratio of the RMSE of the LASSO models to the list of benchmark models, in addition to RMSE and pairwise Diebold-Mariano (DM, Diebold (2015)) tests on the forecast errors for the expanding window exercise, and pairwise Giacomini and White (2006) (GW) tests for the rolling window exercise (see Appendix, Sections B.2.4.1 and B.2.4.2 for detailed specification of the tests, respectively).

3.4.2 Expanding window: US Gross Fixed Capital Formation

3.4.2.1 Main findings

In this section, we present the results of forecasting GFCF during the period of 2005Q1–2019Q1. The results of RMSE of the forecasted values are presented in Table 3.2 for all models and for all 4 forecast horizons. The total period results were divided into three periods (2005Q1–2007Q4, 2008Q1–2014Q4, 2015Q1–2019Q1) aiming to compare the models during different stages of the economy: during the stable growth of 2005Q1–2007Q4, the crisis and the recovery period of 2008Q1–2014Q4, and the stable growth of 2015Q1–2019Q1. Dividing up the total results is also useful in inspecting whether there are models, dominating every other competitor in the "horse race". Also, in Table 3.3 the results of Relative RMSE are compared with the performance of the benchmark ARMA, BNDF, DF and SW models.

The results reveal that when forecasting the GFCF most of the sparse models provide a rather similar forecasting performance, including the BNDF with the LASSO based preselection of the set of variables. The dense benchmark models (SW and DF) seem to fall behind, mostly due to underpredicted shock period of 2008Q1–2014Q4. Finally, the ARMA benchmark models show the least overall accurate predictions, except for the period of 2005Q1–2007Q4 when forecasting 2-quarters ahead, where they are able to outperform every other model by a small margin. Such a result is consistent with the literature: e.g., D'Agostino and Giannone (2012) 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

Table 3.2: RMSE of models forecasts during pseudo-real experiments for GFCF. The labels B, N, Q1 and Q2 denotes the forecast horizons – Backcast, Nowcast, Forecast-1Q and Forecast-2Q, respectively. The bolded value is the smallest one for every column; the blocks correspond to different time periods, while the last block shows performance over the full 2005–2019 period.

Model	2005Q1–2007Q4				2008Q1–2014Q4				2015Q1–2019Q1				2005Q1–2019Q1			
	B	N	Q1	Q2	B	N	Q1	Q2	B	N	Q1	Q2	B	N	Q1	Q2
L	0.70	0.92	1.14	1.27	0.80	1.72	2.48	2.74	0.58	0.82	0.90	0.85	0.65	1.09	1.45	1.56
LP	0.53	**0.84**	1.10	1.23	0.53	1.49	**2.27**	2.75	0.49	0.86	0.91	0.84	0.50	1.01	**1.37**	1.56
LPX	0.45	0.85	1.06	1.22	0.50	1.47	2.31	2.73	0.49	0.88	0.93	0.84	0.47	1.01	1.38	1.56
RL	0.52	0.84	1.05	1.19	0.52	**1.44**	2.32	2.69	0.45	0.92	0.97	0.87	0.47	1.03	1.41	1.55
SqL	0.70	0.89	1.11	1.25	0.71	1.67	2.49	2.76	0.58	0.82	0.88	0.84	0.62	1.06	1.44	1.57
SqP	0.52	0.89	1.01	1.16	0.44	1.60	2.44	2.68	0.51	0.83	0.86	0.83	0.48	1.05	1.39	1.52
AdL	0.07	1.04	**0.98**	1.35	0.19	1.85	2.56	2.88	0.22	0.87	0.95	0.92	0.18	1.18	1.47	1.65
AdP	0.07	0.99	1.01	1.30	0.12	1.85	2.47	2.87	0.21	0.92	0.96	0.91	0.16	1.17	1.44	1.64
AdRL	**0.04**	1.01	1.00	1.35	**0.11**	1.86	2.45	2.82	**0.15**	0.92	0.96	0.91	**0.12**	1.20	1.42	1.62
L0	0.58	1.09	1.10	1.28	1.12	2.02	2.52	2.58	0.56	0.88	1.04	0.83	0.74	1.23	1.49	1.50
L0p	0.57	1.13	1.09	1.31	1.07	2.07	2.59	2.60	0.56	0.87	1.05	0.83	0.72	1.25	1.52	1.51
L0L1	0.82	0.92	1.21	1.37	0.95	1.67	2.34	2.51	0.61	0.84	0.90	**0.82**	0.74	1.10	1.42	**1.49**
L0L1p	0.80	0.94	1.23	1.38	0.92	1.70	2.42	2.54	0.59	0.86	0.93	0.84	0.73	1.13	1.47	1.51
L0L2	0.82	0.95	1.22	1.36	1.05	1.76	2.32	**2.50**	0.60	**0.74**	0.87	0.84	0.77	1.11	1.40	1.49
L0L2p	0.80	0.95	1.24	1.39	0.99	1.87	2.48	2.50	0.56	0.77	0.90	0.87	0.73	1.16	1.48	1.50
ARMA	0.86	0.84	1.04	**1.15**	1.98	2.56	3.16	3.44	0.92	0.99	1.02	1.07	1.24	1.48	1.77	1.91
BNDF	0.78	0.92	1.16	1.36	1.09	1.46	2.36	3.03	0.75	0.83	0.90	0.85	0.81	**1.00**	1.39	1.70
DF	1.12	1.18	1.42	1.57	1.16	1.49	2.30	2.82	0.87	1.14	1.14	1.04	0.98	1.18	1.50	1.71
SW	1.40	1.45	1.57	1.69	1.47	1.96	3.39	3.98	0.89	0.90	**0.83**	0.85	1.14	1.30	1.86	2.13

Table 3.3: US: GFCF. Relative RMSE of model forecasts during the expanding window pseudo-real-time experiments against four benchmarks: ARMA, BNDF, DF and SW models. The labels N, Q1 and Q2 denotes the forecast horizons – Nowcast, Forecast-1Q and Forecast-2Q, respectively. The bolded values correspond to the greatest accuracy within the forecast horizon, while the asterix denotes significant performance improvement as suggested by DM test with 5% significance.

Model	ARMA			BNDF			DF			SW		
	N	Q1	Q2	N	Q1	Q2	N	Q1	Q2	N	Q1	Q2
AdL	0.79	0.83	0.87	1.18	1.05	0.97	0.99	0.98	0.97	0.90	0.79	0.78
AdP	0.79	0.81	0.86	1.17	1.03	0.96	0.99	0.96*	0.96	0.90	0.77	0.77
AdRL	0.81	0.80	0.85	1.21	1.02	0.96	1.02	0.95*	0.95	0.92	0.76	0.76
L	0.73*	0.82	0.82	1.09	1.04	0.92	0.92	0.97*	0.92*	0.83	0.78	0.73
L0	0.83	0.84	0.79	1.23	1.07	0.89	1.04	0.99	0.88	0.94	0.80	0.70
L0L1	0.74*	0.80	0.78	1.11	1.02	0.88	0.93	0.95	0.87	0.84	0.76	0.70
L0L1p	0.76	0.83	0.79	1.13	1.05	0.89	0.95	0.98	0.88	0.86	0.79	0.71
L0L2	0.75*	0.79	**0.78**	1.11	1.01	**0.88**	0.93	0.94	**0.87**	0.85	0.75	**0.70**
L0L2p	0.79	0.83	0.79	1.17	1.06	0.89	0.98	0.98	0.88	0.89	0.79	0.70
L0p	0.85	0.86	0.79	1.26	1.09	0.89	1.06	1.01	0.89	0.96	0.81	0.71
LP	**0.68***	**0.77**	0.82	**1.02**	**0.98**	0.92	**0.86***	**0.91***	0.92*	**0.78***	**0.73**	0.73*
LPX	0.68*	0.78	0.82	1.02	0.99	0.92	0.86	0.92*	0.91*	0.78*	0.74	0.73*
RL	0.70	0.79	0.81	1.04	1.01	0.91	0.87	0.94	0.91*	0.79	0.76	0.73
SqL	0.72*	0.81	0.82	1.07	1.03	0.92	0.90*	0.96*	0.92*	0.81*	0.77	0.74
SqP	0.71	0.79	0.80	1.06	1.00	0.90	0.89	0.93*	0.89*	0.81	0.75	0.71*

models. Therefore, analysis of the recession period of 2008Q1–2014Q4 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 emphasize the value of additional monthly data included in the modelling, especially during the more volatile periods of 2005Q1–2014Q4.

Besides, the RL and SqL models are able to increase the predictive performance of the regular LASSO just as expected. However, the gains for Adaptive and L0Lq variants are not that evident: all of the L0Lq models were able to improve the 2-quarter forecast performance, with the L0L2 having the most accurate 2-quarter forecasts, without significant improvements for the nowcasts or 1-quarter forecasts. The comparison between these groups can be interesting from the variable selection perspective – in most cases throughout the experiment, the AdL models were selecting the median of 55 variables, while the RL and SqL ranged between 23-26 variables, and L0Lq models between 2-9 (L0 was the most sparse, with the median of 2 variables selected, which is consistent with the observations by Mazumder et al. (2022)). Evidently, these numbers translate well into the overall accuracy results: too large amount of variables selected likely introduced additional level of noise into the forecasts, while the too sparse models were likely able to only capture the long term dynamics, but lacking information for nowcasts and 1-quarter forecasts – the differences are most evident during the 2008Q1–2014Q4 period.

Moreover, the results show that the usage of PCA in the estimation can additionally improve the predictive performance of the models: in most of the cases proposed modifications are able to outperform their Relaxed counterparts (the LP vs RL, SqP vs SqL, AdP vs AdL), even though by a small margin. However, for the L0Lq models, likely due to the small amount of variables selected the transformation gains were not evident.

Additionally, in Table 3.3 we mark the significance in forecast gains based on the DM test with 5% significance. First, it should be noted that all of the LASSO models are able to outperform ARMA models when nowcasting. Further, the LP and SqL are able to significantly outperform DF and SW models in nowcasting, while in 1- and 2-quarter forecasting most variants of RL and SqL were able to significantly outperform DF

Table 3.4: RMSE of models forecasts during pseudo-real experiments for GFCF. The bolded value is the smallest one for every row within the same group by number of variables selected. For every block the last line denotes the total error for the period 2005–2019.

	LP10	**RL10**	**LP15**	**RL15**	**LP20**	**RL20**	**LP30**	**RL30**
Backcast								
2005Q1–2007Q4	**0.58**	0.64	0.50	**0.47**	0.39	**0.34**	0.29	**0.26**
2008Q1–2014Q4	**0.92**	0.94	**0.52**	0.55	0.58	**0.50**	**0.41**	0.42
2015Q1–2019Q1	**0.70**	0.70	**0.61**	0.62	0.62	**0.57**	0.52	**0.47**
2005Q1–2019Q1	**0.70**	0.71	**0.55**	0.56	0.54	**0.50**	0.44	**0.41**
Nowcast								
2005Q1–2007Q4	0.83	**0.78**	0.89	**0.81**	0.87	**0.84**	**0.90**	0.90
2008Q1–2014Q4	**1.52**	1.62	**1.62**	1.64	**1.58**	1.61	**1.67**	1.81
2015Q1–2019Q1	**0.80**	0.89	**0.85**	1.01	**0.82**	0.89	**0.85**	0.93
2005Q1–2019Q1	**1.00**	1.06	**1.05**	1.12	**1.02**	1.07	**1.08**	1.18
Forecast-1Q								
2005Q1–2007Q4	**1.05**	1.05	1.00	**0.93**	1.05	**1.03**	1.02	**0.97**
2008Q1–2014Q4	**2.46**	2.46	**2.40**	2.51	**2.51**	2.59	**2.44**	2.59
2015Q1–2019Q1	**0.85**	0.87	**0.87**	0.90	**0.88**	0.94	**0.89**	0.95
2005Q1–2019Q1	**1.42**	1.43	**1.39**	1.43	**1.44**	1.49	**1.42**	1.50
Forecast-2Q								
2005Q1–2007Q4	1.20	**1.16**	1.22	**1.14**	**1.26**	1.30	1.28	**1.22**
2008Q1–2014Q4	2.91	**2.77**	**2.66**	2.76	**2.83**	2.85	**2.72**	2.72
2015Q1–2019Q1	**0.84**	0.84	**0.83**	0.85	**0.83**	0.88	**0.81**	0.85
2005Q1–2019Q1	1.63	**1.57**	**1.52**	1.56	**1.60**	1.63	**1.54**	1.55

and SW models. Overall the most accurate nowcasts are generated by the BNDF model[4].

3.4.2.2 LASSO-PC: gains from transformation

Seeking to inspect directly the gains of using the principal component transformation on the (relaxed) data, a few more comparisons were made. First, in Table 3.4 the results of forecasting GFCF are presented, when the number of preselected variables were fixed to 10, 15, 20, and 30. Note that the results in Tables 3.2 and 3.3 are generated by models with cross-validated hyperparameters. Therefore, the estimated number of significant variables may differ greatly during different time periods

[4] It is worth highlighting the high computational complexity of the DF models in comparison to sparse methods, making the estimation very resource demanding when working with more than 2000 variables as in the case of this study, yet still showing similar (or weaker) forecasting performance to various sparse methods.

and between different models. In this case, we found that restricting the hyperparameter selection problem to select only a fixed amount of (the same) variables is useful.

In Table 3.4 we examine the results of two models: LP, where the preselected variables are transformed into principal components, and RL, where no transformation is made, only the coefficients are re-estimated following the Relaxed LASSO definition. In both cases the same variables are selected as significant. The results provide evidence that LP in some cases can improve the forecasting accuracy when compared with ordinary methods.

Indeed, while some accuracy gains are visible in most cases, with 5% significance the DM test suggests that LP15 and LP30 are able to generate significantly better 1-quarter forecasts than their respective RL versions. However, the results are less conclusive for nowcasts and 2-quarter forecasts due to larger estimated p-values of the DM test.

To sum up, 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 added. This result is expected, since with larger samples we are likely to include more tightly related variables, thus allowing for a clearer extraction of the common factors. Indeed, this becomes apparent in Section 3.4.2.3, where the 30 selected variables seem to form small correlated groups. While such groups may help extracting clearer signals through the PCA, the consequential collinearities can explain why in the RL case increasing the number of included variables did not improve the forecasting results.

In general, the RL associated results are based on two hyperparameters: λ for the selection of variables, and ϕ for the amount of shrinkage applied. Seeking better understanding how the values of hyperparameters affect the forecasting performance 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: the first, where the rotation to the principal components is used and the second, where no transformation is applied and which roughly follows the Relaxed LASSO ideas. Finally, the case of the AdL was included, which uses a different set of variables for the prediction. The results presented in Figure 3.1 show a slight improvement in both the average forecast accuracy – relative mean RMSE comparing with benchmark model – and a smaller

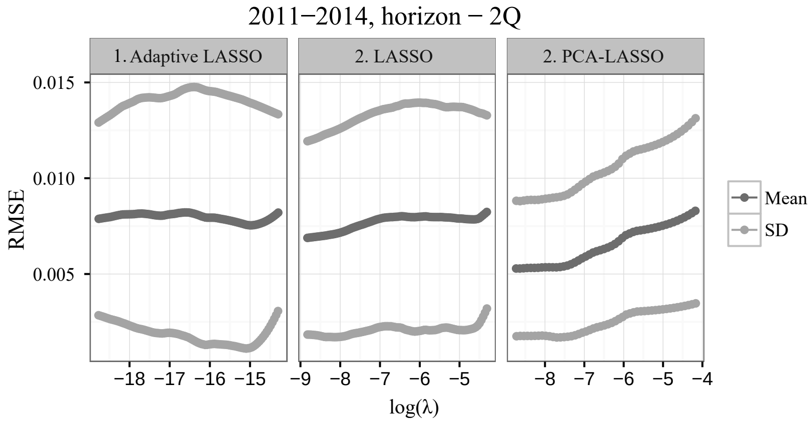

Figure 3.1: The results of forecasting accuracy during the pseudo-real-time experiments over 2011Q1–2014Q4 with the set of preselected variables being fixed for the whole period, and the numbers (1.) and (2.) enumerating the different sets of variables used.

standard deviation for many different values of the hyperparameter λ. These results provide further evidence that the use of principal components transformation might provide additional gains in forecast accuracy.

3.4.2.3 Sparse structures: implications from variable selection

The main advantage of sparse structures is higher interpretability of the selected subsets of indicators. Figure 3.2 depicts the top indicators frequently selected by the Relaxed LASSO during the pseudo-real-time experiments.

We can see that there are several variables selected consistently during every period, which we interpret as the main drivers for explaining the investments in the US. Noteworthy, some of the variables form particular clusters, where one part describes the pre-crisis period while the other part gets significant after the crisis, indicating a possible structural break in the underlying information.

Among the most frequently selected variables are the data reflecting the situation in the market of dwellings: the number of employees and employment rate in construction, the number of building permits and building completions, and industrial production of construction supplies. These variables likely capture the Residential Investments part of the

GFCF, following the evidence found in Lunsford (2015) and similar pre-screened variables detected for Italy in Bulligan et al. (2015).

Table 3.5: RMSE of models forecasts during pseudo-real experiments for GFCF. The bolded value is the smallest one for every row, and for every block the last line denotes the total error of the period 2005–2019.

	DirectPC	**AggregatedPC**
Backcast		
2005Q1–2007Q4	**0.53**	0.53
2008Q1–2014Q4	**0.47**	0.47
2015Q1–2019Q1	**0.57**	0.57
2005Q1–2019Q1	**0.54**	0.54
Nowcast		
2005Q1–2007Q4	0.95	**0.93**
2008Q1–2014Q4	1.45	**1.44**
2015Q1–2019Q1	0.89	**0.88**
2005Q1–2019Q1	1.03	**1.01**
Forecast-1Q		
2005Q1–2007Q4	1.21	**1.16**
2008Q1–2014Q4	2.55	**2.22**
2015Q1–2019Q1	0.95	**0.91**
2005Q1–2019Q1	1.51	**1.36**
Forecast-2Q		
2005Q1–2007Q4	1.39	**1.29**
2008Q1–2014Q4	**2.76**	2.76
2015Q1–2019Q1	0.89	**0.83**
2005Q1–2019Q1	1.61	**1.57**

Second, the productive investment decisions are often determined by the outlook in the related labour market, capturing the business cycle dynamics over different regions, industries and age groups. Such a broad scope may help refining the signals from the labour market. On top of that, Industrial Production Indices proxy the current state of the economic cycle and the expectations in aggregated demand.

In this regard, the use of principal components to extract the underlying common factor explaining the regional diversity in economic activity is deemed necessary for more efficient estimation of the model parameters. These results, however, contrast with Bok et al. (2018), who argue that disaggregated variables, such as employment by age or industry, show no substantial gains in prediction accuracy.

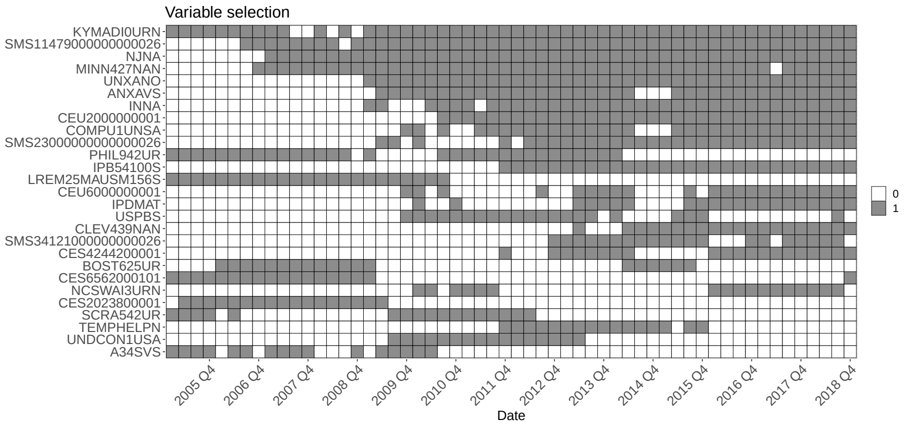

Figure 3.2: Most often selected variables during the expanding window pseudo-real-time experiments for GFCF over 2005Q1–2018Q4. The number of times selected denotes only the number of the same variables selected (e.g., the variable and a one-quarter lag) but not which lag was most often selected. Acronyms described in Appendix, Table B.1.

3.4.2.4 Forecast aggregation

To evaluate the ideas of forecast aggregation, briefly discussed in Section 3.2, under real data, we conduct the following additional experiment. In this experiment, first, we preselect a set of significant variables by the LASSO during every quarter of the pseudo-real-time exercise. Second, we consider the post-LASSO model, which applies an OLS regression, using the first five principal components ordered by their variance explained[5] and treating the assessed factors as explanatory variables. However, as discussed in Section 3.2, we examine two predicting approaches: the first – by fitting an appropriate ARMA model for each of the components and forecasting them directly (DirectPC); and the second – by forecasting the preselected variables and aggregating their forecasts (AggregatedPC). Table 3.5 presents the resulting performance of both of these approaches, and a few conclusions arise.

[5]It is often found in the literature (e.g., Bai and Ng (2002)) that a small number of principal components is usually enough when the initial data sample is relatively sparse. In our case, we noticed that typically around 10 to 30 variables are selected during the exercise, so this number seems optimal since it is not too large for efficient OLS estimation and not too small to omit underlying 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.

First, we can see that the nowcasting performance is as expected very similar, with the aggregated method being able to explain the crisis period more accurately than the direct approach. Indeed, since some monthly information is already known during the nowcasted quarter and the factors compared are the same, such comparison essentially depends on the method used to fill the ragged edges[6].

Second, the forecasting accuracy for most of the periods visibly improved using the aggregated forecast method, with the highest differences during the crisis period of 2008Q1–2014Q4. An inferior forecasting performance by direct approach can likely be caused by underestimating the complexity of the extracted factors. Even if the amount of sample data is relatively large, it may be not enough for efficient estimation of all ARMA parameters. Although the same holds for forecasting the preselected variables, the aggregation of their forecasts appears to improve the results. In this case, the most significant gains were seen for the 1-quarter ahead forecasts. Besides, following the discussion in Section 3.2, we expect more persistent dynamics with a broader scope of information explaining each of the subcomponents.

The benefits of the forecast aggregation are twofold: the creation of a more complex dynamics of the final forecasts than by forecasting directly; and a self-correction by smoothing out inaccurate individual forecasts.

A few additional observations follow from these results. First, because of the complexity of modelling each component, the aggregation is feasible for only a small subset of variables and is not applicable for large and dense problems. However, the complexity of the problem with preselected (targeted) predictors essentially reduces to the complexity of the Relaxed LASSO method. Second, comparing the results from Table 3.2 with Table 3.5, it can be seen that since the variables preselected in both cases are the same, the Post-LASSO solution with using only the first few extracted principal components can even lead to comparable results with LASSO and LP.

[6] In this exercise the ragged edges were filled using the Holt-Winters approach. It is not the most commonly used method for such a problem, but we have found it producing adequate results. ARMA models are also a good alternative, but we did not want to have coinciding nowcasts with the ones from the AggregatedPC.

3.4.3 Rolling window: US Gross Fixed Capital Formation

In this section, we repeat the exercise by switching to a 13-year rolling window. The size of the window reflects the likely occurrence of structural breaks within both standard business cycle frequencies up to 8 years and supply-side driven medium run frequencies up to 13 years. The primary goal of the rolling window is to inspect the impact of old historical information, e.g., continuously chosen by the LASSO as significant only because they help to explain the historical data at the start of the sample but are less useful for the most recent predictions.

3.4.3.1 Main findings

The main results are presented in Tables 3.6 and 3.7. As in the expanding forecast window, we can see that most analysed models can outperform the dense benchmarks. The BNDF model generates the most accurate nowcasts, while L0L1 generates respectively the most accurate 1- and 2-quarter forecasts.

Most variants of the LASSO generates significantly more accurate nowcasts as suggested by the GW test (see Table 3.7) when compared with the ARMA and DF, SW benchmarks, however, the evidence against BNDF is lacking. On the other hand, the evidence is more visible for the 1- and 2-quarter ahead forecasts.

Overall results of the nowcasting performance tightly depend on the variable selection or the accuracy of signal extraction. Therefore, the differences are more visible when comparing dense methods with sparse, bearing in mind that BNDF uses the same variable selection to LP, LPX and RL methods. On the other hand, comparing LP with BNDF directly we are inspecting the gains from the additional shrinkage and factor selection, brought by LP. Finally, examining the forecasting performance, the different approaches for variable forecasting comes in play. For longer, 2-quarter ahead forecasts, we observe the higher gains of ARMA vs Kalman filter approaches, while the 1-quarter forecasts should also capture the forecast aggregation gains, following the discussion around Table 3.5 and the ideas from Section 3.2.

Noteworthy, when comparing the main results with the ones from the expanding window (see Tables 3.2 and 3.6), LP overall accuracy for the full data sample is very similar in both cases. However, by inspecting the nowcasting performance during different spliced periods, it appears that the rolling window setup leads to lower accuracy during the crisis period, but higher onward from 2015Q1. This finding provides evidence for the possible effect of structural changes in the composition of the selected data. In essence, the loss of accuracy during the crisis period suggests that the chosen 13-year window may be too small since the models miss information during the first periods of the estimation. To test the idea, we have repeated the estimation with a 16-year rolling window and were able to increase the nowcasting performance. Although beyond the scope of the exercise, this finding reveals that the size of the rolling window may be treated as a hyperparameter for structural change flexibility and can be further explored, in combination with the resulting sparse structures.

3.4.3.2 Sparse structures

In Figure 3.3, the list of top variables, preselected by the LASSO during the forecasting exercise, is presented. It can be seen that similar to the results from the expanding window estimation, most of the consistently selected indicators explain 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.

By comparing the differences between rolling and expanding windows, two main effects occur. The first effect is the inclusion of more informative variables that become accessible at later periods, e.g., additional indicators for the construction sector. The second is the flexible adjustment to structural changes, e.g., the shift of regions for the employment variables that can be related to different active economic activities, or the exclusion of information that was relevant only during the crisis period.

Table 3.6: RMSE of models forecasts during pseudo-real experiments with rolling window for US GFCF. The labels B, N, Q1 and Q2 denotes the forecast horizons – Backcast, Nowcast, Forecast-1Q and Forecast-2Q, respectively. The bolded value is the smallest one for every column, and every block corresponds to a splitted time period, while the last block shows the results for the full period over 2005–2019.

Model	2005Q1–2007Q4				2008Q1–2014Q4				2015Q1–2019Q1				2005Q1–2019Q1			
	B	**N**	**Q1**	**Q2**	**B**	**N**	**Q1**	**Q2**	**B**	**N**	**Q1**	**Q2**	**B**	**N**	**Q1**	**Q2**
L	0.49	0.86	1.02	1.12	0.85	1.84	2.62	2.95	0.44	0.77	0.90	0.93	0.57	1.10	1.50	1.67
LP	0.31	0.80	0.98	1.15	0.47	1.57	2.56	3.14	0.33	0.73	**0.89**	0.94	0.35	0.98	1.46	1.76
LPX	0.28	0.83	1.02	1.14	0.38	1.73	2.68	3.21	0.32	0.71	0.91	**0.92**	0.32	1.04	1.52	1.78
RL	0.19	0.77	0.97	1.14	0.38	1.81	2.66	3.16	0.29	0.79	0.92	0.92	0.28	1.09	1.52	1.76
SqL	0.55	0.84	1.04	1.16	0.69	1.78	2.67	2.96	0.44	0.77	0.90	0.92	0.52	1.08	1.51	1.68
SqP	0.38	0.84	1.00	1.16	0.43	1.74	2.62	2.99	0.36	0.75	0.91	0.94	0.37	1.06	1.49	1.69
AdP	0.07	0.71	0.82	1.04	0.15	1.87	2.54	2.94	0.09	0.74	0.94	0.96	0.10	1.09	1.45	1.66
AdL	0.10	0.72	0.84	1.03	0.19	1.91	2.58	2.91	0.12	0.75	0.94	0.93	0.14	1.12	1.47	1.64
AdRL	**0.04**	**0.63**	**0.81**	**1.02**	**0.07**	1.79	2.53	2.90	**0.06**	0.75	0.96	0.96	**0.06**	1.05	1.45	1.65
L0L1	0.77	1.17	1.25	1.23	1.01	1.76	**2.12**	**2.51**	0.58	0.86	0.91	0.94	0.74	1.19	**1.35**	**1.52**
L0L1p	0.75	1.15	1.27	1.23	0.98	1.83	2.15	2.57	0.51	0.91	0.93	0.93	0.70	1.22	1.37	1.54
L0	0.65	0.99	1.18	1.26	0.93	1.82	2.26	2.69	0.82	0.94	0.99	1.02	0.81	1.21	1.43	1.61
L0p	0.65	0.98	1.17	1.26	0.90	1.82	2.26	2.69	0.82	0.93	0.99	1.02	0.80	1.21	1.43	1.61
L0L2	0.75	0.99	1.26	1.23	1.01	1.94	2.50	2.61	0.56	0.81	0.94	0.94	0.73	1.20	1.51	1.55
L0L2p	0.75	1.00	1.26	1.24	0.94	1.99	2.61	2.70	0.54	0.84	0.93	0.92	0.70	1.23	1.54	1.58
ARMA	0.87	0.90	1.03	1.17	2.19	2.50	3.45	3.60	0.99	1.08	1.13	1.29	1.34	1.49	1.93	2.01
BNDF	0.54	0.83	1.06	1.20	0.75	**1.48**	2.72	3.65	0.55	**0.68**	0.89	0.97	0.56	**0.94**	1.55	1.99
DF	1.23	1.33	1.43	1.52	1.02	2.35	3.18	3.23	0.77	1.01	1.10	1.09	0.89	1.47	1.84	1.88
SW	1.31	1.42	1.53	1.59	1.74	2.32	3.40	4.15	0.86	0.95	1.01	1.14	1.18	1.46	1.92	2.28

Table 3.7: Relative RMSE to a set of benchmark models, during pseudo-real experiments for GFCF over the full period of 2005–2019. The bolded value corresponds to the best performance per forecast horizon. Asterix indicates significant differences in forecasting accuracy, when compared with the according benchmark, based on GW test with 5% significance.

Model	ARMA			BNDF			DF			SW		
	N	**Q1**	**Q2**	**N**	**Q1**	**Q2**	**N**	**Q1**	**Q2**	**N**	**Q1**	**Q2**
AdL	0.75*	0.76*	0.82*	1.20*	0.95	0.83	1.03	0.95*	0.84*	0.77*	0.77*	0.72*
AdP	0.73*	0.75*	0.83*	1.17	0.94	0.84	1.00	0.93*	0.85*	0.75*	0.76*	0.73*
AdRL	0.71*	0.75*	0.82*	1.13	0.94	0.83	0.97	0.94*	0.85*	0.72*	0.76*	0.72*
L	0.74*	0.78*	0.83*	1.18	0.97	0.84	1.01	0.97	0.86*	0.76*	0.78*	0.73*
L0	0.81	0.74*	0.80*	1.29*	0.93	0.81	1.11	0.92	0.83	0.83	0.75	0.71*
L0L1	0.80	**0.70***	**0.75***	1.27*	**0.88**	**0.76**	1.09	**0.87**	**0.78***	0.82	**0.71***	**0.67***
L0L1p	0.82	0.71*	0.77*	1.31*	0.89	0.77	1.12	0.88	0.79	0.84	0.71*	0.68*
L0L2	0.80	0.78*	0.77*	1.28*	0.98	0.78	1.10	0.97	0.80*	0.82*	0.79*	0.68*
L0L2p	0.82	0.80*	0.79*	1.31*	1.00	0.79	1.13	1.00	0.81	0.84	0.80*	0.69*
L0p	0.81	0.74*	0.80*	1.29*	0.92	0.81	1.11	0.92	0.83	0.83	0.74	0.71*
LP	**0.66***	0.76*	0.87*	**1.05**	0.95	0.88	**0.90**	0.94*	0.90*	**0.67***	0.76*	0.77*
LPX	0.70*	0.79*	0.88*	1.11	0.99	0.89	0.95	0.98	0.91*	0.71*	0.79*	0.78*
RL	0.73*	0.79*	0.88*	1.17*	0.98	0.89	1.00	0.98	0.90*	0.75*	0.79*	0.77*
SqL	0.72*	0.79*	0.83*	1.15*	0.98	0.84	0.99	0.98	0.86*	0.74*	0.79*	0.74*
SqP	0.71*	0.78*	0.84*	1.14*	0.97	0.85	0.97	0.96	0.87*	0.73*	0.78*	0.74*

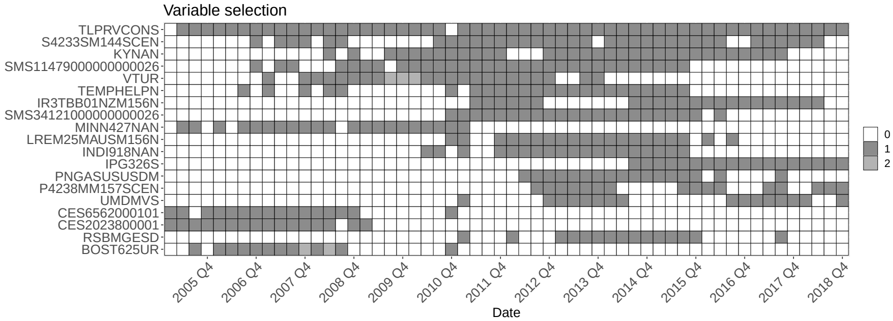

Figure 3.3: Most often selected variables by the LASSO during the rolling window pseudo-real-time forecasting exercise for the GFCF. The number of times selected denotes only the number of the same variables selected (e.g., the variable and a one-quarter lag) but not which lag was most often selected. Acronyms described in Appendix, Table B.2

3.4.4 US Private Final Consumption Expenditure

Compared with the investments, the behaviour of private consumption is quite different. First, it tends to show a much more stable growth than investments. Second, it does not immediately respond to the various stages of the business cycle – the private consumption tends to take the momentum only when the expansion of the current cycle is well underway, 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 particular markets some of the shocks could be perceived much earlier. Additionally, for nowcasting, it is especially convenient to use available "hard" monthly real personal consumption indicators released with a relatively small publication lag. Furthermore, this finding 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 "hard" monthly indicators would be generated by employing more sophisticated models, capable of including more explanatory information than benchmark models.

The results of forecasting the PFCE are presented in Tables B.6 and B.7 of the Appendix B.4 over a rolling 13-year window. The most accurate nowcasts are produced by the L0L1P models, with rather similar performance across most LASSO variants. Besides, the most accurate 1- and 2-quarter forecasts are generated by the BNDF and LP, respect-

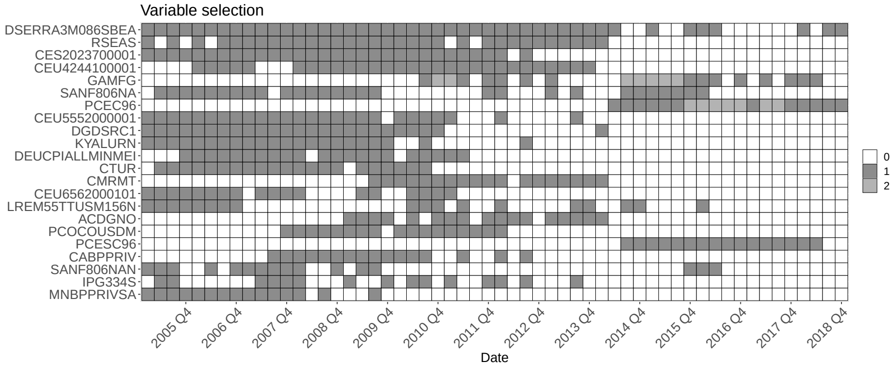

Figure 3.4: Most often selected variables by the LASSO during the rolling window pseudo-real-time forecasting exercise for the PFCE. The number of times selected denotes only the number of the same variables selected (e.g., the variable and a one-quarter lag) but not which lag was most often selected. Acronyms described in Appendix, Table B.3

ively. We note that while the nowcasting performance is rather similar to the benchmark models (as suggested by GW test, see Table B.8), most LASSO variants can significantly outperform DF and SW benchmarks for nowcasts and 1-quarter forecasts.

Figure 3.4 presents the top monthly variables most often preselected by the LASSO as significant. When inspecting the results, we find that the most frequently selected are the "hard" monthly indicators of real personal consumption expenditure (expenditures on goods and services) as was expected. Indeed, statistical agencies themself use "hard" data as the primary sources for their preliminary nowcasts, confirming that LASSO can identify the main leading indicators from a large set of available information.

3.4.5 US International Trade

Similarly to investments, the exports and imports of goods and services are more volatile than the aggregate GDP. The cyclical properties of international trade are quite appealing and follow from the balance of two forces. First, the 'loving-variety' economic agents that seek to smooth consumption using international trade. Second, the additional cyclical variability comes from the investments, that are permitted by the international capital flows. Although there usually exists a strong co-

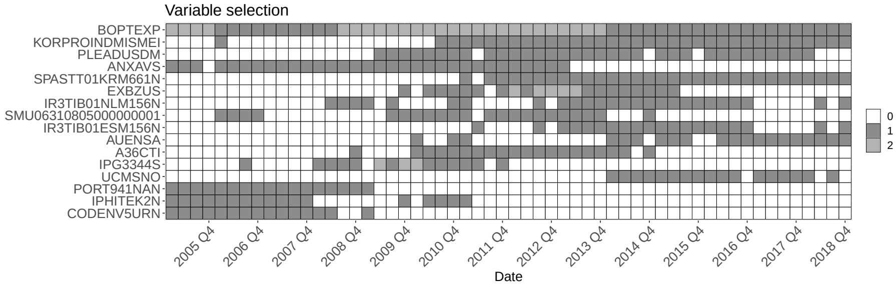

Figure 3.5: Most often selected variables by the LASSO during the rolling window pseudo-real-time forecasting exercise for the Exports. The number of times selected denotes only the number of the same variables selected (e.g., the variable and a one-quarter lag) but not which lag was most often selected. Acronyms described in Appendix, Table B.4

movement between exports and imports, the response to the shocks may have an opposite or supplementary impact on exports and imports. For instance, 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. Some strong demand shocks, on the contrary, might pass through the global economy leading to acceleration or are affected by global boom-bust cycles in economic activity. For example, during acceleration, an initial increase in imports due to the hike in the US domestic demand results in growing foreign exports and foreign income, which in turn increase domestic exports. As in the case with private consumption, the nowcasting of exports and imports is simplified by the existence of "hard" monthly indicators of external trade, published with a small delay.

The results of forecasting Exports and Imports are presented in Table B.6 of the Appendix B.4 over a rolling 13-year window.

When nowcasting the Exports, all variants of the LASSO can outperform the ARMA models, however, are very close to the nowcasts of BNDF, lacking evidence of significance by the GW test. Despite this, the lowest RMSE are generated by the LPX method. The latter provides further evidence that additional forecasting accuracy can follow from using a mixture between the two methods, where both the principal components and the original preselected data are included in the model. Finally, this finding shows that tailoring the rotation matrix can help to reduce the noise from the data, as discussed in Section 3.2.3.

When forecasting 1- and 2-quarters ahead, the results are even closer between all of the models, with BNDF having the most accuracy by a small margin. However, these results are expected due to the existence of "hard" monthly indicators – the nowcasting results highly depends on the accurate variable selection part, while the forecasts reflect the model structure and the accuracy of the monthly indicator forecasts. Since the forecasts are likely led by the monthly Exports of Goods and Services variable (see Figure 3.5, Table B.4), we do not expect significant forecast accuracy improvements due to forecast aggregation as per the discussion in Section 3.2.2, hence the very similar results between all of the model forecasts.

When nowcasting the Imports, AdRL shows the best result, significantly outperforming the DF model, as suggested by the GW test. Additionally, SqL generates the best 1-quarter forecasts, while SqP generates the best 2-quarter forecasts. The GW test results are weaker for the forecasts, suggesting that the large observed RMSE differences may be caused by a few distinct points, most likely during the crisis period.

3.4.6 European countries

As a robustness check, we repeat the estimation procedure for four largest Euro Area economies that are also the main contributors to the EU budget: Germany, France, Italy and Spain. Replicating the estimation, we cover the same four main components of the GDP accounted by the expenditure approach. Although, due to the varying availability of the historical data, we had to make certain adjustments. As with the US, it is of interest to cover the financial crisis period in the experiments. However, for some series, the available GDP data starts later than 1992Q1 – in such cases, the rolling window is set to expand until it reaches the previously defined 13-year window.

The results are presented in Appendix B.4. As can be expected, no single forecasting approach dominates in the "horse race". However, we note that unlike with the results for the US data, there were cases when the dense factor methods were able to outperform the sparse methods (see Appendix B.4, Tables B.12, B.13, B.19, B.20 for Spain GFCF and Exports, France PFCE and GFCF).

In fact, for Italy PFCE variable, both the sparse and dense methods were hardly outperforming the ARIMA models (see Appendix B.4,

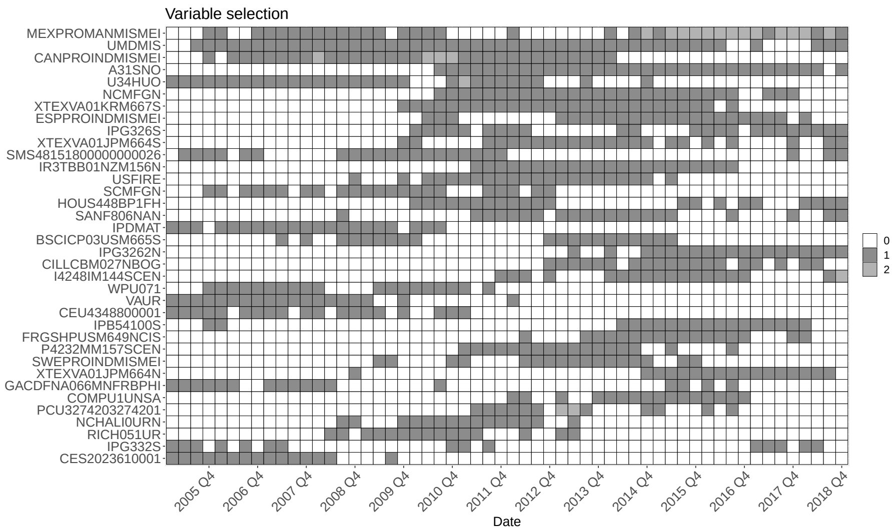

Figure 3.6: Most often selected variables by the LASSO during the rolling window pseudo-real-time forecasting exercise for the Imports. The number of times selected denotes only the number of the same variables selected (e.g., the variable and a one-quarter lag) but not which lag was most often selected. Acronyms described in Appendix, Table B.5

Table B.23) when nowcasting. This result naturally follows from the scarce availability of data and the short historical periods for the available explanatory variables.

Some variants of the proposed LASSO-PC modification were able to generate the most accurate nowcasts for France Exports, Italy Imports (see Appendix B.4, Tables B.21, B.26), and the most accurate 1- or 2-quarter forecasts for Spain (GFCF, Imports), Germany (PFCE, GFCF, Exports), France (GFCF, Exports, Imports), Italy (PFCE, GFCF, Imports) (see Appendix B.4, Tables B.12, B.14, B.15, B.16, B.17, B.20, B.21, B.22, B.23, B.24, B.26). While some of the gains are less evident according to the estimated statistical significance using the GW test, the results extend the findings from the US forecasting results. Noteworthy, in many cases the nowcasting performance for the DF and BNDF models were significantly improved from applying commonly used monthly aggregation strategies and the use of bridge equations for the nowcast estimation, as described in Bańbura et al. (2013), which were not used for the sparse methods. On the other hand, when comparing the 1- and 2-quarter forecasting performance, the effects of such nowcasting solutions are diminished and a clearer comparison can be made.

Comparing the performance by the forecasted GDP components, we conclude that for all four countries, the PCA modification consistently showed improvements for the GFCF. However, for the remaining variables, the performance gains were less evident and less consistent. These results confirm the ideas discussed in Sections 3.4 and 3.4.2 for the US data and provides further evidence that the structural complexity of the GFCF component and the lack of "hard" leading indicators translates well within Germany, Spain, France, and Italy. Hence, in such cases, we can expect tangible gains from applying LASSO-PC against traditional sparse methods.

Finally, in addition to inspecting only the best performing models, we examine the median overall performance rankings of every model in the "horse race". Table 3.8 presents the results, where the models are ordered by their median rankings throughout the "horse race". Median rankings can demonstrate the overall stability of the models, capturing the relative model performance under varying data quality and availability. In this case, it follows that the LP and LPX models are able to consistently generate adequate nowcasts and 1- and 2-quarter forecasts, while the remaining models show higher variability and higher probability of producing relatively poor forecasts.

Table 3.8: Median model rankings during the pseudo-real experiments for Euro Area GDP components over the full period of 2005–2019. The values are sorted by the 1-quarter forecast performance in the ascending order.

Model	Nowcast	Forecast-1Q	Forecast-2Q
LPX	7	5	5.5
LP	5.5	6	7
L0L1	9	7.5	12
L	8	8	7
SqL	8	8	8.5
SqP	8.5	8	8
AdP	8	8.5	7.5
RL	10	8.5	6.5
L0L2p	10	9	10.5
AdL	7.5	9.5	9.5
L0L1p	11.5	10	12
DF	11	11	10.5
BNDF	6.5	11.5	12.5
L0L2	13	11.5	8.5
AdRL	8.5	12	9.5
L0	16	15	16
L0p	16	16	15
SW	17.5	17	8.5
ARMA	16	18	18

Conclusions

In this dissertation important aspects of sparsity in high-dimensional linear regression models were studied. A particular attention was paid to approximate sparsity, which is the underlying assumption of many popular sparse modelling techniques. The empirical performance of sparse methods and its implications were investigated by a comprehensive pseudo-real-time short-term forecasting study. Below we summarize the main results of the dissertation.

In the first part of the dissertation, we studied the statistic $\|\mathbb{X}'Y\|_2^2$ and derived its exact and asymptotic distributions under certain conditions. We highlight the application of variance-gamma distribution used to derive the main results and deem that this approach can open up interesting avenues for future work, e.g., in order to derive alternative statistics of a similar nature.

By implementing a Monte Carlo simulation study under the assumption of approximate sparsity of the underlying model coefficients, we find that the empirical distribution of the statistic $\|\mathbb{X}'Y\|_2^2$ fairly quickly approaches the limiting distribution, suggesting applicability of the results in constructing tests for sparsity and the related SNR estimation literature.

For future research the results could be extended further by dropping the assumption of independence between the observations. Furthermore, following the ideas of Dicker and Erdogdu (2016), the derived exact distribution could be used in order to form likelihood functions, which could be applied for further improvements in estimating the SNR or $\|\beta\|_2^2$.

In the second part of the dissertation, we studied the forecasting accuracy of the LASSO and its widespread modifications, together with the proposed approach of combining LASSO with the method of principal components. We found evidence that these methods can distinguish and estimate the underlying explanatory variables for the nowcasting and short-run forecasting problem.

We find that all LASSO methods showed acceptable forecasting performance, outperforming the benchmark ARMA and factor models. The advantages of including additional explanatory monthly information were substantial during the crisis period of 2008Q1–2010Q4, where both the nowcasts and one- and two-quarter forecasts in most cases provided more accurate results than benchmark models. Furthermore, the number of variables selected by the methods was mostly relatively small, suggesting that the sparseness assumption for the data generating process holds.

Moreover, in most cases, the popular modifications of LASSO were 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.

Finally, while the LASSO is fit for generating adequate forecasts for different macroeconomic data, our suggested modification (LASSO-PC) showed additional gains in forecast accuracy. We found evidence that the proposed combination often generates more accurate forecasts than the Adaptive LASSO or the Relaxed LASSO, which already are substantial modifications of the original LASSO that have Oracle Properties. Therefore, we expect further gains in forecasting accuracy with additional work on these methods. On the other hand, the methods we discussed never find non-linearities if they are not in the initial search space. A more time consuming yet interesting extension, in this regard, would be to consider second- or third-order interaction terms between the variables or their power transformations, which might result in further improvement of the forecasting performance.

The short-term forecasting results might be improved further by working in two specific directions. First, the usage of default weights for the Adaptive LASSO often showed better pseudo-real-time forecasting performance than the standard LASSO and benchmark models. We suppose that the choice of the weights could be optimized or cross-validated to better deal with the high-dimensionality problem. Second, when combining the method of principal components with the LASSO, only the standard estimation procedure of the principal components was considered. However, the PCA part of the LASSO-PC might benefit further from tailoring the angles or scale of the rotation matrix for additional gains in forecasting accuracy.

Bibliography

Andreini, P., C. Izzo, and G. Ricco (2020). Deep dynamic factor models. *arXiv preprint arxiv:2007.11887*.

Angelini, E., G. Camba-Mendez, D. Giannone, L. Reichlin, and G. Rünstler (2011). Short-term forecasts of euro area gdp growth. *Econometrics Journal 14*(1), C25–C44.

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 (2002). Determining the Number of Factors in Approximate Factor Models. *Econometrica 70*(1), 191–221.

Bai, J. and S. Ng (2007). Determining the number of primitive shocks in factor models. *Journal of Business & Economic Statistics 25*(1), 52–60.

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, and L. Reichlin (2011). Nowcasting. In M. P. Clements and D. F. Hendry (Eds.), *Oxford Handbook on Economic Forecasting*. Oxford: Oxford University Press.

Bayati, M., M. A. Erdogdu, and A. Montanari (2013). Estimating lasso risk and noise level. *Advances in Neural Information Processing Systems 26*.

Bańbura, M., D. Giannone, M. Modugno, and L. Reichlin (2013). Now-Casting and the Real-Time Data Flow. In G. Elliott and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, Volume 2 of *Handbook of Economic Forecasting*, pp. 195–237. Elsevier.

Belloni, A. and V. Chernozhukov (2011). High dimensional sparse econometric models: An introduction. In *Inverse Problems and High-Dimensional Estimation*, pp. 121–156. Springer.

Belloni, A. and V. Chernozhukov (2013). Least squares after model selection in high-dimensional sparse models. *Bernoulli 19*(2), 521–547.

Belloni, A., V. Chernozhukov, D. Chetverikov, C. Hansen, and K. Kato (2018). High-dimensional econometrics and regularized GMM. *arXiv preprint arXiv:1806.01888*.

Belloni, A., V. Chernozhukov, and C. Hansen (2011). Inference for high-dimensional sparse econometric models. *arXiv preprint arXiv:1201.0220*.

Belloni, A., V. Chernozhukov, and L. Wang (2011). Square-root lasso: pivotal recovery of sparse signals via conic programming. *Biometrika 98*(4), 791–806.

Bertsimas, D., A. King, and R. Mazumder (2016). Best subset selection via a modern optimization lens. *The Annals of Statistics 44*(2), 813–852.

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

Bok, B., D. Caratelli, D. Giannone, A. M. Sbordone, and A. Tambalotti (2018). Macroeconomic nowcasting and forecasting with big data. *Annual Review in Economics 10*, 615–643.

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.

Bradic, J., V. Chernozhukov, W. K. Newey, and Y. Zhu (2021). Minimax semiparametric learning with approximate sparsity. *arXiv preprint arXiv:1912.12213*.

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

Breiman, L. (1996b). Heuristics of instability and stabilization in model selection. *The Annals of Statistics 24* (6), 2350–2383.

Bühlmann, P. and S. Van De Geer (2011). *Statistics for high-dimensional data: methods, theory and applications.* Springer Science & Business Media.

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

Caner, M. and A. B. Kock (2018). Asymptotically honest confidence regions for high dimensional parameters by the desparsified conservative lasso. *Journal of Econometrics 203* (1), 143–168.

Carpentier, A. and N. Verzelen (2019). Adaptive estimation of the sparsity in the gaussian vector model. *The Annals of Statistics 47* (1), 93–126.

Carpentier, A. and N. Verzelen (2021). Optimal sparsity testing in linear regression model. *Bernoulli 27* (2), 727–750.

Castro, V. (2010). The duration of economic expansions and recessions: More than duration dependence. *Journal of Macroeconomics 32* (1), 347–365.

Cevid, D., P. Bühlmann, and N. Meinshausen (2020). Spectral deconfounding via perturbed sparse linear models. *Journal of Machine Learning Research 21* (232), 1–41.

Cha, J., H. D. Chiang, and Y. Sasaki (2021). Inference in high-dimensional regression models without the exact or l^p sparsity. *ArXiv preprint arXiv:2108.09520*.

Chernozhukov, V., C. Hansen, and Y. Liao (2017). A lava attack on the recovery of sums of dense and sparse signals. *Annals of Statistics 45* (1), 39–76.

Chetverikov, D., Z. Liao, and V. Chernozhukov (2021). On cross-validated lasso in high dimensions. *The Annals of Statistics 49*(3), 1300–1317.

de Bondt, G. and P. Vermeulen (2018). Business cycle duration dependence and foreign recessions. *ECB Working Paper November 22, 2018*(No. 2205).

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).

Dedieu, A., H. Hazimeh, and R. Mazumder (2021). Learning sparse classifiers: Continuous and mixed integer optimization perspectives. *Journal of Machine Learning Research 22*(135), 1–47.

Dicker, L. H. (2014). Variance estimation in high-dimensional linear models. *Biometrika 101*(2), 269–284.

Dicker, L. H. (2016). Ridge regression and asymptotic minimax estimation over spheres of growing dimension. *Bernoulli 22*(1), 1–37.

Dicker, L. H. and M. A. Erdogdu (2016). Maximum likelihood for variance estimation in high-dimensional linear models. In *Artificial Intelligence and Statistics*, pp. 159–167. PMLR.

Diebold, F. and R. Mariano (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics 13*(3), 253–63.

Diebold, F. X. (2003). Big data dynamic factor models for macroeconomic measurement and forecasting. In L. H. M. Dewatripont and S. Turnovsky (Eds.), *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress of the Econometric Society*, pp. 115–122.

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).

Diggle, P. J. and E. Giorgi (2019). *Model-based geostatistics for global public health: methods and applications.* Chapman and Hall/CRC.

D'Agostino, A. and D. Giannone (2012). Comparing alternative predictors based on large-panel factor models. *Oxford bulletin of economics and statistics 74*(2), 306–326.

Efron, B., T. Hastie, I. Johnstone, and R. Tibshirani (2004). Least angle regression. *The Annals of Statistics 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).

Fikioris, G. (2018). Spectral properties of Kac–Murdock–Szegő matrices with a complex parameter. *Linear Algebra and its Applications 553*, 182–210.

Gaunt, R. E. (2013). *Rates of Convergence of Variance-Gamma Approximations via Stein's Method.* Ph. D. thesis, The Queen's College, University of Oxford.

Gaunt, R. E. (2014). Variance-gamma approximation via stein's method. *Electronic Journal of Probability 19*, 1–33.

Gaunt, R. E. (2018). Products of normal, beta and gamma random variables: Stein operators and distributional theory. *Brazilian Journal of Probability and Statistics 32*(2), 437–466.

Gaunt, R. E. (2019). A note on the distribution of the product of zero-mean correlated normal random variables. *Statistica Neerlandica 73*(2), 176–179.

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

Giannone, D., M. Lenza, and G. E. Primiceri (2021). Economic predictions with big data: The illusion of sparsity. *Econometrica 89*(5), 2409–2437.

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.

Gold, D., J. Lederer, and J. Tao (2020). Inference for high-dimensional instrumental variables regression. *Journal of Econometrics 217*(1), 79–111.

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

Guo, Z., D. Ćevid, and P. Bühlmann (2021). Doubly debiased lasso: High-dimensional inference under hidden confounding. *ArXiv preprint arXiv:2004.03758*.

Harvey, D., S. Leybourne, and P. Newbold (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting 13*(2), 281–291.

Hastie, T., R. Tibshirani, and R. J. Tibshirani (2017). Extended comparisons of best subset selection, forward stepwise selection, and the lasso. *arXiv preprint arXiv:1707.08692*.

Hazimeh, H. and R. Mazumder (2020). Fast best subset selection: Coordinate descent and local combinatorial optimization algorithms. *Operations Research 68*(5), 1517–1537.

Heinisch, K. and R. Scheufele (2018). Bottom-up or direct? Forecasting German GDP in a data-rich environment. *Empirical Economics 54*(2), 705–745.

Hoerl, A. E. and R. W. Kennard (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics 12*(1), 55–67.

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

Hyndman, R. and Y. Khandakar (2008). Automatic time series forecasting: The forecast package for r. *Journal of Statistical Software, Articles 27*(3), 1–22.

Ing, C.-K. (2007). Accumulated prediction errors, information criteria and optimal forecasting for autoregressive time series. *The Annals of Statistics 35*(3), 1238–1277.

Ing, C.-K. (2020). Model selection for high-dimensional linear regression with dependent observations. *The Annals of Statistics 48*(4), 1959–1980.

Janson, L., R. F. Barber, and E. Candes (2017). Eigenprism: inference for high dimensional signal-to-noise ratios. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 79*(4), 1037–1065.

Javanmard, A. and A. Montanari (2014). Confidence intervals and hypothesis testing for high-dimensional regression. *The Journal of Machine Learning Research 15*(1), 2869–2909.

Kac, M., W. Murdock, and G. Szegő (1953). On the eigen-values of certain Hermitian forms. *Journal of Linear Rational Mechanics and Analysis 2*, 767–800.

Kim, H. H. and N. R. Swanson (2018). Mining big data using parsimonious factor, machine learning, variable selection and shrinkage methods. *International Journal of Forecasting 34*(2), 339–354.

Knight, K. and W. Fu (2000). Asymptotics for lasso-type estimators. *The Annals of Statistics 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.

Kotz, S., T. Kozubowski, and K. Podgórski (2001). The laplace distribution and generalizations: A revisit with applications to communications, economics, engineering, and finance. Boston. Birkhäuser.

Kwiatkowski, D., P. C. Phillips, P. Schmidt, and Y. Shin (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics 54*(1-3), 159–178.

Liang, K.-Y. and S. L. Zeger (1986). Longitudinal data analysis using generalized linear models. *Biometrika 73*(1), 13–22.

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

Lunsford, K. G. (2015). Forecasting residential investment in the united states. *International Journal of Forecasting 31*(2), 276–285.

Madan, D. B., P. P. Carr, and E. C. Chang (1998). The variance gamma process and option pricing. *Review of Finance 2*(1), 79–105.

Magnus, J. R. and G. De Luca (2016). Weighted-average least squares (WALS): a survey. *Journal of Economic Surveys 30*(1), 117–148.

Maximon, L. C. (2003). The dilogarithm function for complex argument. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences 459*(2039), 2807–2819.

Mazumder, R. (2020). Discussion of "Best Subset, Forward Stepwise or Lasso? Analysis and Recommendations Based on Extensive Comparisons". *Statistical Science 35*(4), 602–608.

Mazumder, R., P. Radchenko, and A. Dedieu (2022). Subset selection with shrinkage: Sparse linear modeling when the snr is low. *Operations Research*.

Medeiros, M. C. and E. F. Mendes (2016). l1-regularization of high-dimensional time-series models with non-Gaussian and heteroskedastic errors. *Journal of Econometrics 191*(1), 255–271.

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

Morris, R. (1979). The dilogarithm function of a real argument. *Mathematics of Computation 33*(146), 778–787.

Natarajan, B. K. (1995). Sparse approximate solutions to linear systems. *SIAM journal on computing 24*(2), 227–234.

Newey, W. K. (1997). Convergence rates and asymptotic normality for series estimators. *Journal of Econometrics 79*(1), 147–168.

Newey, W. K. and K. D. West (1986). A simple, positive semi-definite, heteroskedasticity and autocorrelationconsistent covariance matrix.

Ng, S. (2013). Variable selection in predictive regressions. In *Handbook of Economic Forecasting*, Volume 2, pp. 752–789. Elsevier B.V.

Ning, Y., S. Peng, and J. Tao (2020). Doubly robust semiparametric difference-in-differences estimators with high-dimensional data. *ArXiv preprint arXiv:2009.03151*.

Patil, A. R. and S. Kim (2020). Combination of ensembles of regularized regression models with resampling-based lasso feature selection in high dimensional data. *Mathematics 8*(1), 110.

Rangan, S. (2011). Generalized approximate message passing for estimation with random linear mixing. In *2011 IEEE International Symposium on Information Theory Proceedings*, pp. 2168–2172. IEEE.

Reid, S., R. Tibshirani, and J. Friedman (2016). A study of error variance estimation in lasso regression. *Statistica Sinica*, 35–67.

Shibata, R. (1980). Asymptotically efficient selection of the order of the model for estimating parameters of a linear process. *The Annals of Statistics*, 147–164.

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 & Economic Statistics 20*(2), 147–62.

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.

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

Tikhonov, A. N. (1963). On the solution of ill-posed problems and the method of regularization. In *Doklady Akademii Nauk*, Volume 151, pp. 501–504. Russian Academy of Sciences.

Verzelen, N. and E. Gassiat (2018, 11). Adaptive estimation of high-dimensional signal-to-noise ratios. *Bernoulli 24*(4B), 3683–3710.

Vila, J. P. and P. Schniter (2013). Expectation-maximization gaussian-mixture approximate message passing. *IEEE Transactions on Signal Processing 61*(19), 4658–4672.

Wu, W.-B. and Y. N. Wu (2016). Performance bounds for parameter estimates of high-dimensional linear models with correlated errors. *Electronic Journal of Statistics 10*(1), 352–379.

Yang, Y., J. Zhou, and J. Pan (2021). Estimation and optimal structure selection of high-dimensional toeplitz covariance matrix. *Journal of Multivariate Analysis 184*, 104739.

Zhang, C.-H. and S. S. Zhang (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 76*(1), 217–242.

Zhang, C.-H. and T. Zhang (2012). A general theory of concave regularization for high-dimensional sparse estimation problems. *Statistical Science 27*(4), 576–593.

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., T. Hastie, and R. Tibshirani (2006). Sparse Principal Component Analysis. *Journal of Computational and Graphical Statistics 15*(2), 265–286.

Appendix A

Technical lemmas

Throughout the proofs we use the notation C to mark generic constants, the specific values of which can change from line to line.

Lemma A.1. *Assume that* $|\varrho| < 1$. *Then,*

$$\int_0^\varrho \frac{\log(1-\varrho x)}{x(1-x)}\,\mathrm{d}x \quad = \quad -\frac{1}{2}\big(\operatorname{Li}_2(\varrho^2) + \log^2(1-\varrho)\big),$$

where Li_2 *denotes the real dilogarithm function. (Recall, that for* $\varrho < 0$, *by* $\int_0^\varrho$ *we denote* $-\int_\varrho^0$.*)*

Proof. Write,

$$\int_0^\varrho \frac{\log(1-\varrho x)}{x(1-x)}\,\mathrm{d}x \quad = \quad \int_0^\varrho \frac{\log(1-\varrho x)}{x}\,\mathrm{d}x \;+\; \int_0^\varrho \frac{\log(1-\varrho x)}{1-x}\,\mathrm{d}x.$$

By (2.72), we have

$$\int_0^\varrho \frac{\log(1-\varrho x)}{x}\,\mathrm{d}x \quad = \quad -\operatorname{Li}_2(\varrho^2). \tag{A.1}$$

It remains to show that

$$\int_0^\varrho \frac{\log(1-\varrho x)}{1-x}\,\mathrm{d}x \quad = \quad \frac{1}{2}\big(\operatorname{Li}_2(\varrho^2) - \log^2(1-\varrho)\big). \tag{A.2}$$

Indeed, by substitution $v = \varrho - \varrho x$, we have

$$\int_0^\varrho \frac{\log(1-\varrho x)}{1-x}\,\mathrm{d}x \quad = \quad \int_{\varrho-\varrho^2}^\varrho \frac{\log(1-\varrho+v)}{v}\,\mathrm{d}v$$

$$= \int_{\varrho-\varrho^2}^{\varrho} \frac{\log(1+\frac{v}{1-\varrho})}{v}\,\mathrm{d}v \;-\; \log^2(1-\varrho) \tag{A.3}$$

Further, by substitution $w = -\frac{v}{1-\varrho}$, we have

$$\begin{aligned}
\int_{\varrho-\varrho^2}^{\varrho} \frac{\log(1+\frac{v}{1-\varrho})}{v}\,\mathrm{d}v &= -\int_{-\frac{\varrho}{1-\varrho}}^{-\varrho} \frac{\log(1-w)}{w}\,\mathrm{d}w \\
&= \mathrm{Li}_2(-\varrho) - \mathrm{Li}_2\Big(-\frac{\varrho}{1-\varrho}\Big) \\
&= \mathrm{Li}_2(-\varrho) + \mathrm{Li}_2(\varrho) + \frac{1}{2}\log^2(1-\varrho) \qquad \text{(A.4)} \\
&= \frac{1}{2}\big(\mathrm{Li}_2(\varrho^2) + \log^2(1-\varrho)\big), \qquad \text{(A.5)}
\end{aligned}$$

where for (A.4)–(A.5) we apply the easily verifiable identities (see, e.g., Maximon (2003)):

$$\begin{aligned}
\mathrm{Li}_2\Big(\frac{x}{x-1}\Big) &= -\mathrm{Li}_2(x) - \frac{1}{2}\log^2(1-x), \quad x<1, \\
\mathrm{Li}_2(x) + \mathrm{Li}_2(-x) &= \frac{1}{2}\mathrm{Li}_2(x^2), \quad |x|<1.
\end{aligned}$$

Thus, (A.3) and (A.5) imply (A.2), which concludes the proof.

□

Lemma A.2. *Assume that* $\sum_{j=1}^{\infty}\beta_j^2 < \infty$ *and* $|\varrho| < 1$. *Then, the following inequalities hold:*

1. $\Big|\sum_{l=p+1}^{\infty}\sum_{l'=l+1}^{\infty}\beta_l\beta_{l'}\varrho^{l'-l}\Big| \;\leq\; C\sum_{l=p+1}^{\infty}\beta_l^2.$

2. $\Big|\sum_{l=p+1}^{\infty}\sum_{l'=l+1}^{\infty}\beta_l\beta_{l'}\varrho^{l'-l}(l'-l)\Big| \;\leq\; C\sum_{l=p+1}^{\infty}\beta_l^2.$

3. $\Big|\sum_{l=1}^{p}\sum_{l'=p+1}^{\infty}\beta_l\beta_{l'}\varrho^{l'-l}\Big| \;\leq\; C\sum_{l=p+1}^{\infty}\beta_l^2.$

4. $\Big|\sum_{l=1}^{p}\sum_{l'=p+1}^{\infty}\beta_l\beta_{l'}\varrho^{l'+l}\Big| \;\leq\; C\sum_{l=p+1}^{\infty}\beta_l^2.$

Proof. See the proof in Appendix, Section A.3. □

Lemma A.3. *Assume that* $\sup_{j\geq 1}|\beta_j|j^{\alpha}<\infty$, $\alpha>1/2$ *and that* $|\varrho|<1$. *Then,*

$$\Big|\sum_{j=1}^{p}\beta_j\varrho^{p-j}\Big| = o(p^{-1/4}).$$

Proof. We have

$$\begin{aligned}
\Big|\sum_{j=1}^{p}\beta_j\varrho^{p-j}\Big| &\leq \sum_{j=1}^{\lfloor\sqrt{p}\rfloor}|\beta_j||\varrho|^{p-j}+\sum_{j=\lfloor\sqrt{p}\rfloor+1}^{p}|\beta_j||\varrho|^{p-j}\\
&\leq \sup_{j\geq 1}|\beta_j|\sum_{j=1}^{\lfloor\sqrt{p}\rfloor}|\varrho|^{p-j}+p^{-\alpha/2}\sum_{j=\lfloor\sqrt{p}\rfloor+1}^{p}|\beta_j|p^{\alpha/2}|\varrho|^{p-j}\\
&\leq \sup_{j\geq 1}|\beta_j|\sum_{j=1}^{\lfloor\sqrt{p}\rfloor}|\varrho|^{p-j}+p^{-\alpha/2}\sup_{j\geq 1}|\beta_j|j^{\alpha}\sum_{j=\lfloor\sqrt{p}\rfloor+1}^{p}|\varrho|^{p-j}\\
&\leq C\Big(\sum_{j=1}^{\lfloor\sqrt{p}\rfloor}|\varrho|^{p-j}+p^{-\alpha/2}\sum_{j=\lfloor\sqrt{p}\rfloor+1}^{p}|\varrho|^{p-j}\Big)\\
&\leq C\Big(|\varrho|^{p-\lfloor\sqrt{p}\rfloor}+p^{-\alpha/2}\Big).
\end{aligned}$$

Here we used the fact that $\sum_{j=\lfloor\sqrt{p}\rfloor+1}^{p}|\varrho|^{p-j}\to(1-|\varrho|)^{-1}<\infty$. Thus,

$$p^{1/4}\Big|\sum_{j=1}^{p}\beta_j\varrho^{p-j}\Big|\leq C\left(p^{1/4}|\varrho|^{p-\lfloor\sqrt{p}\rfloor}+p^{\frac{1}{4}-\frac{\alpha}{2}}\right)\ \to\ 0. \tag{A.6}$$

□

REMARK A.1. Obviously, the assumption $\sup_{j\geq 1}|\beta_j|j^{\alpha}<\infty$, for $\alpha>1/2$, implies that $\sum_{j=1}^{\infty}\beta_j^2<\infty$:

$$\sum_{j=1}^{\infty}\beta_j^2 \;=\; \sum_{j=1}^{\infty}\beta_j^2 j^{2\alpha}j^{-2\alpha} \;\leq\; \sup_{j\geq 1}\beta_j^2 j^{2\alpha}\sum_{k=1}^{\infty}k^{-2\alpha} \;<\; \infty.$$

Lemma A.4. *Assume that the assumptions of Theorem 2.1 hold. Then,*

$$\kappa_{2,p}=o(p).$$

Proof. Observe, that

$$\begin{aligned}
\kappa_{2,p} &= \sum_{k=1}^{p}\Big(\sum_{l=1}^{p}\beta_l\varrho^{|k-l|}\Big)^2 = \sum_{k=1}^{p}\sum_{l_1,l_2=1}^{p}\beta_{l_1}\beta_{l_2}\varrho^{|k-l_1|+|k-l_2|} \\
&\le \sum_{l_1,l_2=1}^{p}|\beta_{l_1}||\beta_{l_2}|\sum_{k=1}^{p}|\varrho|^{|k-l_1|+|k-l_2|} \\
&\le C\Big(\sum_{l=1}^{p}|\beta_{l_1}|\Big)^2 \qquad \text{(A.7)} \\
&= o(p)
\end{aligned}$$

where (A.7) follows from (A.24). Meanwhile, $\sum_{l=1}^{p}|\beta_{l_1}| = o(p^{1/2})$, since

$$\begin{aligned}
\sum_{l=1}^{p}|\beta_l| &= \sum_{l=1}^{\lfloor p^{1/2}\rfloor}|\beta_l| + \sum_{l=\lfloor p^{1/2}\rfloor+1}^{p}|\beta_l| \\
&\le p^{1/4}\Big(\sum_{l=1}^{\infty}\beta_l^2\Big)^{1/2} + p^{1/2}\Big(\sum_{l=\lfloor p^{1/2}\rfloor+1}^{\infty}\beta_l^2\Big)^{1/2} = o(p^{1/2}).
\end{aligned}$$

□

Lemma A.5. *Assume that* $\sum_{j=1}^{\infty}\beta_j^2 < \infty$ *and* $|\varrho| < 1$. *Define* $\theta_k^{(p)} = \sum_{j=1}^{p}\beta_j\varrho^{|k-j|}$. *Then,*

$$\Big|\sum_{i,j,k=1}^{p}\big(\varrho^{|i-j|}+\theta_i^{(p)}\theta_j^{(p)}\big)\big(\varrho^{|i-k|}+\theta_i^{(p)}\theta_k^{(p)}\big)\big(\varrho^{|k-j|}+\theta_k^{(p)}\theta_j^{(p)}\big)\Big| = o(p^{3/2}). \qquad \text{(A.8)}$$

Proof. See the proof in Appendix, Section A.4. □

A.1 Proof of Lemma 2.5

Here and throughout the proof we employ the notation as in Definition 2.1.
(i) Note that, by (2.65) and (2.67), we have

$$\kappa_{1,p} = \sum_{k=1}^{p}\beta_k^2 + 2\sum_{k=2}^{p}\sum_{l=1}^{k-1}\beta_k\beta_l\varrho^{k-l} \;\to\; \beta(1)+2b_1(\varrho) \quad \text{as } p\to\infty.$$

(ii) Write

$$\kappa_{2,p} = \sum_{l=1}^{p}\sum_{k=1}^{p}\beta_l^2\varrho^{2|k-l|} + 2\sum_{l'>l}\sum_{k=1}^{p}\beta_l\beta_{l'}\varrho^{|k-l|}\varrho^{|k-l'|}.$$

From here, it is straightforward to see that

$$\kappa_{2,p} \to \beta(1)\frac{1+\varrho^2}{1-\varrho^2} - \beta(\varrho^2)\frac{1}{1-\varrho^2} \tag{A.9}$$

$$+ \; 2\Big(b_1^{(1)}(\varrho) \; + \; b_1(\varrho)\frac{1+\varrho^2}{1-\varrho^2} \; - \; b_2(\varrho)\frac{1}{1-\varrho^2}\Big). \tag{A.10}$$

Technical details of the proof of (A.10) are presented in Appendix, Section A.5.

(iii) Consider

$$\kappa_{3,p} = \sum_{l=1}^{p}\beta_l^2 J_1(l) + 2\sum_{l<l'}\beta_l\beta_{l'}J_2(l,l'), \tag{A.11}$$

where

$$J_1(l) := \sum_{k,k'=1}^{p}\varrho^{|k-k'|}\varrho^{|k-l|}\varrho^{|k'-l|}\mathbb{1}_{\{l=l'\}}, \tag{A.12}$$

$$J_2(l,l') := \sum_{k,k'=1}^{p}\varrho^{|k-k'|}\varrho^{|k-l|}\varrho^{|k'-l'|}\mathbb{1}_{\{l<l'\}}. \tag{A.13}$$

Then, it is straightforward to see that, as $p\to\infty$, using the notation in Definition 2.1, we have that

$$\sum_{l=1}^{p}\beta_l^2 J_1(l) \to \beta(1)\frac{1+4\varrho^2+\varrho^4}{(1-\varrho^2)^2} - \beta(\varrho^2)\frac{1+3\varrho^2}{(1-\varrho^2)^2}$$

$$- \; \frac{2}{1-\varrho^2}\beta^{(1)}(\varrho^2), \tag{A.14}$$

and

$$\sum_{l'>l}\beta_l^2 J_2(l,l') \to \frac{1}{2(1-\varrho^2)^2}\big(b^{(2)}(\varrho)(1-\varrho^2)^2 + 3b_1^{(1)}(\varrho)(1-\varrho^4)$$

$$+ 2b_1(\varrho)(1+4\varrho^2+\varrho^4) - \; 2b_2^{(1)}(\varrho)(1-\varrho^2)$$

$$- \; 2b_2(\varrho)(1+3\varrho^2)\big). \tag{A.15}$$

Technical details of the proof of (A.14)–(A.15) are omitted here and presented in the Appendix, Section A.6. This concludes the proof.

A.2 Proof of Lemma 2.4

Part (i). Write

$$\kappa_1 - \kappa_{1,p} = 2\sum_{l=1}^{p}\sum_{l'=p+1}^{\infty} \beta_l\beta_{l'}\varrho^{l'-l} + \sum_{l=p+1}^{\infty}\sum_{l'=p+1}^{\infty} \beta_l\beta_{l'}\varrho^{|l-l'|}.$$

Here,

$$\left|\sum_{l=1}^{p}\sum_{l'=p+1}^{\infty} \beta_l\beta_{l'}\varrho^{l'-l}\right| = o(p^{-1/2})$$

by Lemma A.2(iii), while

$$\begin{aligned}\sum_{l=p+1}^{\infty}\sum_{l'=p+1}^{\infty} \beta_l\beta_{l'}\varrho^{|l-l'|} &= \sum_{l=p+1}^{\infty}\beta_l^2 + 2\sum_{l=p+1}^{\infty}\sum_{l'=l+1}^{\infty} \beta_l\beta_{l'}\varrho^{l'-l} \\ &\le C\sum_{l=p+1}^{\infty}\beta_l^2 = o(p^{-1/2})\end{aligned}$$

by Lemma A.2(i), which concludes the proof. □

Part (ii). Write

$$\begin{aligned}\kappa_2 - \kappa_{2,p} &= \Bigg(2\sum_{k=1}^{p}\sum_{l=1}^{p}\sum_{l'=p+1}^{\infty} + 2\sum_{k=p+1}^{\infty}\sum_{l=1}^{p}\sum_{l'=p+1}^{\infty} + \sum_{k=1}^{p}\sum_{l=p+1}^{\infty}\sum_{l'=p+1}^{\infty} \\ &\quad + \sum_{k=p+1}^{\infty}\sum_{l=1}^{p}\sum_{l'=1}^{p} + \sum_{k=p+1}^{\infty}\sum_{l=p+1}^{\infty}\sum_{l'=p+1}^{\infty}\Bigg)\beta_l\beta_{l'}\varrho^{|k-l|+|k-l'|} \\ &=: 2L_1 + 2L_2 + L_3 + L_4 + L_5. \end{aligned} \tag{A.16}$$

We have

$$\begin{aligned}L_1 &= \sum_{l=1}^{p}\sum_{l'=p+1}^{\infty} \beta_l\beta_{l'}\sum_{k=1}^{p}\varrho^{|k-l|+l'-k} \\ &= \sum_{l=1}^{p}\sum_{l'=p+1}^{\infty} \beta_l\beta_{l'}\left(\sum_{k=1}^{l}\varrho^{l'+l-2k} + \sum_{k=l+1}^{p}\varrho^{l'-l}\right)\end{aligned}$$

$$= \sum_{l=1}^{p} \sum_{l'=p+1}^{\infty} \beta_l \beta_{l'} \Big(\frac{\varrho^{l'-l}}{1-\varrho^2} - \frac{\varrho^{l+l'}}{1-\varrho^2} + (p-l)\varrho^{l'-l} \Big).$$

Thus,

$$\begin{aligned} |L_1| &\leq \sum_{l=1}^{p} \sum_{l'=p+1}^{\infty} |\beta_l||\beta_{l'}| \Big(\frac{|\varrho|^{l'-l}}{1-\varrho^2} + \frac{|\varrho|^{l+l'}}{1-\varrho^2} + (l'-l)|\varrho|^{l'-l} \Big) \\ &\leq C \sum_{j=p+1}^{\infty} \beta_j^2 \ = \ o(p^{-1/2}) \end{aligned}$$

by Lemma A.2(ii)–(iv). For the term L_2 we have

$$\begin{aligned} L_2 &= \sum_{l=1}^{p} \sum_{l'=p+1}^{\infty} \beta_l \beta_{l'} \sum_{k=p+1}^{\infty} \varrho^{k-l+|k-l'|} \\ &= \sum_{l=1}^{p} \sum_{l'=p+1}^{\infty} \beta_l \beta_{l'} \Bigg(\sum_{k=p+1}^{l'} \varrho^{l'-l} + \sum_{k=l'+1}^{\infty} \varrho^{2k-l-l'} \Bigg). \end{aligned}$$

Thus, by Lemma A.2(ii)-(iii),

$$|L_2| \ \leq \ \sum_{l=1}^{p} \sum_{l'=p+1}^{\infty} |\beta_l||\beta_{l'}| \Big((l'-l)|\varrho|^{l'-l} + \frac{\varrho^2}{1-\varrho^2}|\varrho|^{l'-l} \Big) \ = \ o(p^{-1/2}).$$

For the term L_3 we have

$$\begin{aligned} L_3 &= \sum_{l=p+1}^{\infty} \sum_{l'=p+1}^{\infty} \beta_l \beta_{l'} \sum_{k=1}^{p} \varrho^{l+l'-2k} \\ &= \sum_{l=p+1}^{\infty} \sum_{l'=p+1}^{\infty} \beta_l \beta_{l'} \Big(\frac{\varrho^{l'+l-2p}}{1-\varrho^2} - \frac{\varrho^{l+l'}}{1-\varrho^2} \Big). \end{aligned}$$

Thus,

$$\begin{aligned} |L_3| &\leq \sum_{l=p+1}^{\infty} \sum_{l'=p+1}^{\infty} |\beta_l||\beta_{l'}| \Big(\frac{|\varrho|^{l'+l-2p}}{1-\varrho^2} + \frac{|\varrho|^{l+l'}}{1-\varrho^2} \Big) \\ &\leq \frac{1}{1-\varrho^2} \Bigg(\Big(\sum_{l=p+1}^{\infty} |\beta_l||\varrho|^{l-p} \Big)^2 + \Big(\sum_{l=p+1}^{\infty} |\beta_l||\varrho|^{l} \Big)^2 \Bigg) \end{aligned}$$

$$\leq \; C \sum_{l=p+1}^{\infty} \beta_l^2 + o(p^{-1/2}) = o(p^{-1/2}),$$

since $\sum_{l=p+1}^{\infty} |\varrho|^{2(l-p)} < \infty$ and due to Hölder's inequality. Further, for the term L_4 we have

$$\begin{aligned} L_4 &= \sum_{l=1}^{p}\sum_{l'=1}^{p} \beta_l \beta_{l'} \sum_{k=p+1}^{\infty} \varrho^{2k-l-l'} = \frac{\varrho^2}{1-\varrho^2} \sum_{l=1}^{p}\sum_{l'=1}^{p} \beta_l \beta_{l'} \varrho^{(p-l)+(p-l)} \\ &= \frac{\varrho^2}{1-\varrho^2} \left(\sum_{l=1}^{p} \beta_l \varrho^{p-l} \right)^2 = o(p^{-1/2}), \end{aligned}$$

by Lemma A.3. Finally, for L_5 write

$$L_5 = \sum_{k=p+1}^{\infty} \sum_{l=p+1}^{\infty} \beta_l^2 \varrho^{2|k-l|} + 2 \sum_{k=p+1}^{\infty} \sum_{l,l'=p+1,\, l'>l}^{\infty} \beta_l \beta_{l'} \varrho^{|k-l|+|k-l'|}.$$

For the first summand, we have

$$\begin{aligned} \sum_{k,l=p+1}^{\infty} \beta_l^2 \varrho^{2|k-l|} &= \sum_{l=p+1}^{\infty} \beta_l^2 \sum_{k=p+1}^{\infty} \varrho^{2|k-l|} \mathbb{1}_{\{k \geq l\}} \\ &\quad + \sum_{l=p+1}^{\infty} \beta_l^2 \sum_{k=p+1}^{\infty} \varrho^{2|k-l|} \mathbb{1}_{\{k<l\}} \\ &= \sum_{l=p+1}^{\infty} \beta_l^2 \sum_{k=l}^{\infty} \varrho^{2k-2l} + \sum_{l=p+1}^{\infty} \beta_l^2 \sum_{k=p+1}^{l+1} \varrho^{2l-2k} \\ &= \sum_{l=p+1}^{\infty} \beta_l^2 \left(\left(\frac{1}{1-\varrho^2} \right) + \left(- \frac{\varrho^{2l-2p}}{1-\varrho^2} + \frac{1}{\varrho^2 (1-\varrho^2)} \right) \right) \\ &\leq C \sum_{l=p+1}^{\infty} \beta_l^2, \end{aligned} \tag{A.17}$$

where $C < \infty$. Similarly,

$$\begin{aligned} &\sum_{k=p+1}^{\infty} \sum_{l'>l}^{\infty} \beta_l \beta_{l'} \varrho^{|k-l|+|k-l'|} \\ &\qquad = \sum_{l=p+1}^{\infty} \sum_{l'=l+1}^{\infty} \sum_{k=p+1}^{\infty} \beta_l \beta_{l'} \varrho^{2k-l-l'} \mathbb{1}_{\{k \geq l' > l\}} \end{aligned} \tag{A.18}$$

$$+ \sum_{l=p+1}^{\infty} \sum_{l'=l+1}^{\infty} \sum_{k=p+1}^{\infty} \beta_l \beta_{l'} \varrho^{l'-l} \mathbb{1}_{\{l'>k\geq l\}} \quad \text{(A.19)}$$

$$+ \sum_{l=p+1}^{\infty} \sum_{l'=l+1}^{\infty} \sum_{k=p+1}^{\infty} \beta_l \beta_{l'} \varrho^{l'+l-2k} \mathbb{1}_{\{l'>l>k\}}. \quad \text{(A.20)}$$

For (A.18), write

$$\sum_{l=p+1}^{\infty} \sum_{l'=l+1}^{\infty} \sum_{k=p+1}^{\infty} \beta_l \beta_{l'} \varrho^{2k-l-l'} \mathbb{1}_{\{k\geq l'>l\}} = \sum_{l=p+1}^{\infty} \sum_{l'=l+1}^{\infty} \beta_l \beta_{l'} \frac{\varrho^{l'-l}}{1-\varrho^2}.$$

Thus, by Lemma A.2(i), we have

$$\left| \sum_{l=p+1}^{\infty} \sum_{l'=l+1}^{\infty} \sum_{k=p+1}^{\infty} \beta_l \beta_{l'} \varrho^{2k-l-l'} \mathbb{1}_{\{k\geq l'>l\}} \right| \leq C \sum_{l=p+1}^{\infty} \beta_l^2,$$

for $C < \infty$. Next, for (A.19), write

$$\sum_{l=p+1}^{\infty} \sum_{l'=l+1}^{\infty} \sum_{k=p+1}^{\infty} \beta_l \beta_{l'} \varrho^{l'-l} \mathbb{1}_{\{l'>k\geq l\}} = \sum_{l=p+1}^{\infty} \sum_{l'=l+1}^{\infty} \beta_l \beta_{l'} \varrho^{l'-l} (l'-l).$$

It follows from Lemma A.2(ii) that

$$\left| \sum_{l=p+1}^{\infty} \sum_{l'=l+1}^{\infty} \sum_{k=p+1}^{\infty} \beta_l \beta_{l'} \varrho^{l'-l} \mathbb{1}_{\{l'>k\geq l\}} \right| \leq C \sum_{l=p+1}^{\infty} \beta_l^2,$$

for $C < \infty$. Finally, for (A.20), we have

$$\sum_{k,l,l'=p+1}^{\infty} \beta_l \beta_{l'} \varrho^{l'+l-2k} \mathbb{1}_{\{l'>l>k\}}$$
$$= \sum_{l=p+2}^{\infty} \sum_{l'=l+1}^{\infty} \beta_l \beta_{l'} \left(\frac{\varrho^{(l-p)+(l'-p)}}{\varrho^2 - 1} + \varrho^2 \frac{\varrho^{l'-l}}{1-\varrho^2} \right).$$

Thus, by Lemma A.2(i),

$$\left| \sum_{k,l,l'=p+1}^{\infty} \beta_l \beta_{l'} \varrho^{l'+l-2k} \mathbb{1}_{\{l'>l>k\}} \right|$$
$$\leq \frac{1}{2} \sum_{l=p+2}^{\infty} \sum_{l'=l+1}^{\infty} \left(\beta_l^2 \varrho^{2(l'-p)} + \beta_{l'}^2 \varrho^{2(l-p)} \right) + C \sum_{l=p+1}^{\infty} \beta_l^2$$

$$\leq \quad C \sum_{l=p+1}^{\infty} \beta_l^2.$$

Hence, (A.17) and the estimates for (A.18)–(A.20) yield

$$|L_5| \quad \leq \quad C \sum_{l=p+1}^{\infty} \beta_l^2 \;=\; o(p^{-1/2}).$$

Equality (A.16) and estimates $|L_i| = o(p^{-1/2})$, $i = 1, \ldots, 5$, complete the proof of part (ii). $\square$

A.3 Proof of Lemma A.2

Proof. For inequality (i), we have that

$$\begin{aligned}
\Big| \sum_{l=p+1}^{\infty} \sum_{l'=l+1}^{\infty} \beta_l \beta_{l'} \varrho^{l'-l} \Big| \quad &\leq \quad \frac{1}{2} \sum_{l=p+1}^{\infty} \sum_{l'=l+1}^{\infty} \Big(\beta_l^2 |\varrho|^{l'-l} + \beta_{l'}^2 |\varrho|^{l'-l} \Big) \\
&= \quad \frac{1}{2} \Big(\frac{|\varrho|}{1-|\varrho|} \sum_{l=p+1}^{\infty} \beta_l^2 + \sum_{l'=p+1}^{\infty} \beta_{l'}^2 \Big(- \frac{|\varrho|^{l'-p}}{1-|\varrho|} + \frac{1}{1-|\varrho|} \Big) \Big) \\
&\leq \quad C \sum_{l=p+1}^{\infty} \beta_l^2.
\end{aligned}$$

For inequality (ii), note that

$$\begin{aligned}
&\Big| \sum_{l=p+1}^{\infty} \sum_{l'=l+1}^{\infty} \beta_l \beta_{l'} \varrho^{l'-l} (l'-l) \Big| \\
&\qquad \leq \quad \frac{1}{2} \sum_{l=p+1}^{\infty} \sum_{l'=l+1}^{\infty} \Big(\beta_l^2 |\varrho|^{l'-l}(l'-l) + \beta_{l'}^2 |\varrho|^{l'-l}(l'-l) \Big) \\
&\qquad = \quad \frac{|\varrho|}{(1-|\varrho|)^2} \sum_{l=p+1}^{\infty} \beta_l^2 \\
&\qquad \quad + \sum_{l'=p+2}^{\infty} \beta_{l'}^2 \frac{|\varrho|^{l'-p}}{(1-|\varrho|)^2} \big(-(l'-p)(1-|\varrho|) - |\varrho| \big) \\
&\qquad \quad + \sum_{l'=p+2}^{\infty} \beta_{l'}^2 \frac{|\varrho|}{(1-|\varrho|)^2}
\end{aligned}$$

$$
\begin{aligned}
&\leq \quad \frac{2|\varrho|}{(1-|\varrho|)^2}\sum_{l=p+1}^{\infty}\beta_l^2 - \sum_{l'=p+1}^{\infty}\beta_{l'}^2\frac{|\varrho|^{l'-p}(l'-p)}{1-|\varrho|} \\
&\quad - \sum_{l'=p+1}^{\infty}\beta_{l'}^2\frac{|\varrho|^{l'-p}}{1-|\varrho|} \\
&\leq \quad C\sum_{l=p+1}^{\infty}\beta_l^2.
\end{aligned}
$$

For (iii), see that

$$
\begin{aligned}
\left|\sum_{l=1}^{p}\sum_{l'=p+1}^{\infty}\beta_l\beta_{l'}\varrho^{l'-l}\right| &\leq \sum_{l=1}^{p}\sum_{l'=p+1}^{\infty}\left(\beta_l^2|\varrho|^{l'-l}+\beta_{l'}^2|\varrho|^{l'-l}\right) \\
&= \sum_{l=1}^{p}\beta_l^2\sum_{l'=p+1}^{\infty}|\varrho|^{l'-l} + \sum_{l'=p+1}^{\infty}\sum_{l=1}^{p}\beta_{l'}^2|\varrho|^{l'-l} \\
&= \sum_{l=1}^{p}\beta_l^2\frac{|\varrho|^{p-l+1}}{1-|\varrho|} + \sum_{l'=p+1}^{\infty}\beta_{l'}^2\left(\frac{|\varrho|^{l'-p}}{1-|\varrho|}-\frac{|\varrho|^{l'}}{1-|\varrho|}\right) \\
&\leq o(1) + C\sum_{l=p+1}^{\infty}\beta_l^2,
\end{aligned}
$$

where the last inequality follows from (A.25).

The proof of (iv) follow the same steps as that of (iii). □

A.4 Proof of Lemma A.5

Proof. First, note, that by (2.40), Lemma 2.5 and Remark 2.3, it holds that

$$
\lim_{p\to\infty}\sum_{k=1}^{p}(\theta_k^{(p)})^2 \quad = \quad c_2 \ < \ \infty. \tag{A.21}
$$

Then, we have that

$$
\begin{aligned}
&\sum_{i,j,k=1}^{p}\left(\varrho^{|i-j|}+\theta_i^{(p)}\theta_j^{(p)}\right)\left(\varrho^{|i-k|}+\theta_i^{(p)}\theta_k^{(p)}\right)\left(\varrho^{|k-j|}+\theta_k^{(p)}\theta_j^{(p)}\right) \\
&= \sum_{i,j,k=1}^{p}\left(\varrho^{|i-j|+|i-k|}+\varrho^{|i-j|}\theta_i^{(p)}\theta_k^{(p)} \ + \ \varrho^{|i-k|}\theta_i^{(p)}\theta_j^{(p)}+(\theta_i^{(p)})^2\theta_j^{(p)}\theta_k^{(p)}\right)
\end{aligned}
$$

$$\times \ \left(\varrho^{|k-j|} + \theta_k^{(p)}\theta_j^{(p)}\right)$$

$$= \sum_{i,j,k=1}^{p} \left(\varrho^{|i-j|+|i-k|+|k-j|} \ + \ \varrho^{|i-j|+|k-j|}\theta_i^{(p)}\theta_k^{(p)} \ + \ \varrho^{|i-k|+|k-j|}\theta_i^{(p)}\theta_j^{(p)}\right.$$

$$+ \ \varrho^{|k-j|}(\theta_i^{(p)})^2\theta_j^{(p)}\theta_k^{(p)} \ + \ \varrho^{|i-j|+|i-k|}\theta_k^{(p)}\theta_j^{(p)} \ + \ \varrho^{|i-j|}\theta_i^{(p)}(\theta_k^{(p)})^2\theta_j^{(p)}$$

$$\left. + \ \varrho^{|i-k|}\theta_i^{(p)}(\theta_j^{(p)})^2\theta_k^{(p)} \ + \ (\theta_i^{(p)})^2(\theta_j^{(p)})^2(\theta_k^{(p)})^2\right).$$

In order to show (A.8), it suffices to show that the following results hold, since the remaining cases will be symmetric:

1. $\left|\sum_{i,j,k=1}^{p} \varrho^{|i-j|+|i-k|+|k-j|}\right| = o(p^{3/2});$

2. $\left|\sum_{i,j,k=1}^{p} \varrho^{|i-j|+|k-j|}\theta_i^{(p)}\theta_k^{(p)}\right| = o(p^{3/2});$

3. $\sum_{i,j,k=1}^{p} \varrho^{|k-j|}(\theta_i^{(p)})^2\theta_j^{(p)}\theta_k^{(p)} = o(p^{3/2});$

4. $\sum_{i,j,k=1}^{p} (\theta_i^{(p)})^2(\theta_j^{(p)})^2(\theta_k^{(p)})^2 = o(p^{3/2}).$

Case (i). We have

$$\sum_{i,j,k=1}^{p} \varrho^{|i-j|+|i-k|+|k-j|} = \sum_{i=1}^{p}\sum_{k=1}^{p} \varrho^{2|i-k|} + 2\sum_{i>j}\sum_{k=1}^{p} \varrho^{|i-j|+|i-k|+|k-j|},$$

where

$$\sum_{i=1}^{p}\sum_{k=1}^{p} \varrho^{2|i-k|} = \sum_{i=1}^{p}\left(\sum_{k=1}^{i} \varrho^{2(i-k)} + \sum_{k=i+1}^{p} \varrho^{2(k-i)}\right) = \mathcal{O}(p), \quad \text{(A.22)}$$

and

$$\sum_{j=1}^{p}\sum_{i=j+1}^{p}\sum_{k=1}^{p} \varrho^{|i-j|+|i-k|+|k-j|}$$

$$= \sum_{j=1}^{p}\sum_{i=j+1}^{p}\sum_{k=1}^{j} \varrho^{2i-2k} + \sum_{j=1}^{p}\sum_{i=j+1}^{p}\sum_{k=j+1}^{i} \varrho^{2i-2j} + \sum_{j=1}^{p}\sum_{i=j+1}^{p}\sum_{k=i+1}^{p} \varrho^{2k-2j}$$

$$= \frac{\varrho^2}{(1-\varrho^2)^3}\left(p(1-\varrho^2)(3\varrho^{2p}+\varrho^2+2) - 2(1+2\varrho^2)(1-\varrho^{2p})\right)$$

so that

$$\left|\sum_{j=1}^{p}\sum_{i=j+1}^{p}\sum_{k=1}^{p}\varrho^{|i-j|+|i-k|+|k-j|}\right| \;=\; \mathcal{O}(p) \;=\; o(p^{3/2}).$$

Case (ii). We have,

$$\begin{aligned}\sum_{i,j,k=1}^{p}\varrho^{|i-j|+|k-j|}\theta_i^{(p)}\theta_k^{(p)} &\leq \sum_{i,j,k=1}^{p}\Big(\varrho^{2|i-j|+2|k-j|} \;+\; (\theta_i^{(p)})^2(\theta_k^{(p)})^2\Big)\\ &= \sum_{i,j,k=1}^{p}\varrho^{2|i-j|+2|k-j|} \;+\; p\sum_{i,k=1}^{p}(\theta_i^{(p)})^2(\theta_k^{(p)})^2.\end{aligned}$$

Observe, that by (A.21), we have $p\sum_{i,k=1}^{p}(\theta_i^{(p)})^2(\theta_k^{(p)})^2 = \mathcal{O}(p)$. Additionally,

$$\sum_{i,j,k=1}^{p}\varrho^{2|i-j|+2|k-j|} \;=\; \sum_{i,k=1}^{p}\varrho^{2|k-i|} \;+\; 2\sum_{j=1}^{p}\sum_{i=j+1}^{p}\sum_{k=1}^{p}\varrho^{2(i-j)+2|k-j|}.$$

We use (A.22) and note that

$$\begin{aligned}\sum_{j=1}^{p}\sum_{i=j+1}^{p}\sum_{k=1}^{p}\varrho^{2(i-j)+2|k-j|} &= \sum_{j=1}^{p}\sum_{i=j+1}^{p}\sum_{k=1}^{j}\varrho^{2(i-j)+2(j-k)}\\ &\quad+\sum_{j=1}^{p}\sum_{i=j+1}^{p}\sum_{k=j+1}^{p}\varrho^{2(i-j)+2(k-j)}\\ &= \frac{\varrho^2}{(1-\varrho^2)^3}\Big(p(1-\varrho^2)(\varrho^{2p}+\varrho^2+1)\\ &\quad-(1-\varrho^{2p})(3\varrho^2+\frac{1-2\varrho^{2p+2}}{1+\varrho^2})\Big)\\ &= \mathcal{O}(p).\end{aligned}$$

Thus, it follows that

$$\sum_{i,j,k=1}^{p}\varrho^{|i-j|+|k-j|}\theta_i^{(p)}\theta_k^{(p)} \;=\; \mathcal{O}(p) \;=\; o(p^{3/2}).$$

Case (iii). By (A.22) and (A.21),

$$\begin{aligned}\sum_{i,j,k=1}^{p}(\theta_i^{(p)})^2\varrho^{|k-j|}\theta_j^{(p)}\theta_k^{(p)} &= \sum_{i=1}^{p}(\theta_i^{(p)})^2\sum_{k,j=1}^{p}\varrho^{|k-j|}\theta_j^{(p)}\theta_k^{(p)} \\ &\leq \sum_{i=1}^{p}(\theta_i^{(p)})^2\sum_{k,j=1}^{p}\left(\varrho^{2|k-j|}+(\theta_j^{(p)})^2(\theta_k^{(p)})^2\right) \\ &= o(p^{3/2}).\end{aligned}$$

Case (iv). By (A.21),

$$\sum_{i,j,k=1}^{p}(\theta_i^{(p)})^2(\theta_j^{(p)})^2(\theta_k^{(p)})^2 = \left(\sum_{i=1}^{p}(\theta_i^{(p)})^2\right)^3 = o(p^{3/2}).$$

Thus, this concludes the proof of (A.8). □

A.5 Proof of result (A.10) of Lemma 2.5(ii)

Proof. Denote

$$K_1(l) := \sum_{k=1}^{p}\varrho^{2|k-l|}, \quad K_2(l,l') := \sum_{k=1}^{p}\varrho^{|k-l|+|k-l'|}.$$

Then, we can write

$$\sum_{k=1}^{p}\left(\sum_{l=1}^{p}\beta_l\varrho^{|k-l|}\right)^2 = \sum_{l=1}^{p}\beta_l^2K_1(l)+2\sum_{l'>l}\beta_l\beta_{l'}K_2(l,l'). \quad (A.23)$$

First, note that

$$K_1(l) = \frac{-\varrho^{-2l+2p+2}-\varrho^{2l}+\varrho^2+1}{1-\varrho^2} \to \frac{-\varrho^{2l}+\varrho^2+1}{1-\varrho^2} \quad \text{as } p\to\infty.$$

Then, using the notation by Definition 2.1, it follows by the Dominated Convergence Theorem (DCT) that

$$\sum_{l=1}^{p}\beta_l^2K_1(l) \to \sum_{l=1}^{\infty}\beta_l^2\frac{1+\varrho^2-\varrho^{2l}}{1-\varrho^2} = \beta(1)\frac{1+\varrho^2}{1-\varrho^2} - \beta(\varrho^2)\frac{1}{1-\varrho^2}.$$

Next, consider $K_2(l,l')$, $l' > l$. It's straightforward to see that,

$$\begin{aligned}
\sum_{k=1}^{l} \varrho^{|k-l|+|k-l'|} &= \sum_{k=1}^{l} \varrho^{l-k+l'-k} = \frac{(1-\varrho^{2l})\varrho^{l'-l}}{1-\varrho^2}, \\
\sum_{k=l+1}^{l'} \varrho^{|k-l|+|k-l'|} &= \sum_{k=l+1}^{l'} \varrho^{k-l+l'-k} = (l'-l)\varrho^{l'-l}, \\
\sum_{k=l'+1}^{p} \varrho^{|k-l|+|k-l'|} &= \sum_{k=l'+1}^{p} \varrho^{k-l+k-l'} = \frac{\varrho^{-l-l'+2}(\varrho^{2l'}-\varrho^{2p})}{1-\varrho^2}.
\end{aligned}$$

By simplifying, it follows that

$$\begin{aligned}
K_2(l,l') &= \frac{(l-l'-1)\varrho^{l'-l} + \varrho^{l+l'} + (-l+l'-1)\varrho^{l'+2-l} + \varrho^{2p+2-l-l'}}{\varrho^2-1} \\
&\to \frac{\varrho^{l'-l}((l'-l)(\varrho^2-1) - 1 - \varrho^2) + \varrho^{l+l'}}{\varrho^2-1} \quad \text{as } p\to\infty,
\end{aligned}$$

therefore, using the notation of Definition 2.1, we rewrite

$$\begin{aligned}
&\sum_{l'=2}^{\infty}\sum_{l=1}^{l'-1} K_2(l,l') \\
&\quad = \sum_{l'=2}^{\infty}\sum_{l=1}^{l'-1} \beta_l\beta_{l'} \left(\varrho^{l'-l}(l'-l) + \varrho^{l'-l}\frac{1+\varrho^2}{1-\varrho^2} - \varrho^{l'+l}\frac{1}{1-\varrho^2} \right) \\
&\quad = b_1^{(1)}(\varrho) + b_1(\varrho)\frac{1+\varrho^2}{1-\varrho^2} - b_2(\varrho)\frac{1}{1-\varrho^2},
\end{aligned}$$

which concludes the proof of (A.10). □

A.6 Proof of results (A.14)–(A.15) of Lemma 2.5(iii)

Proof. First, we establish the following observation:

$$\begin{aligned}
\sum_{l_2=1}^{p} |\varrho|^{|l_1-l_2|} &= |\varrho|^{l_1} \sum_{l_2=1}^{l_1} |\varrho|^{-l_2} + |\varrho|^{-l_1} \sum_{l_2=l_1+1}^{p} |\varrho|^{l_2} \\
&= |\varrho|^{l_1} \frac{|\varrho|^{-1}(|\varrho|^{-l_1}-1)}{|\varrho|^{-1}-1} + |\varrho|^{-l_1} \frac{|\varrho|^{l_1+1}(|\varrho|^{p-l_1}-1)}{|\varrho|-1}
\end{aligned}$$

$$= \frac{1+|\varrho|-|\varrho|^{l_1}-|\varrho|^{p-l_1+1}}{1-|\varrho|} \leq \frac{1+|\varrho|}{1-|\varrho|}. \tag{A.24}$$

Consider $J_1(l)$ in (A.12). By (A.24), write,

$$\begin{aligned} J_1(l) &= \sum_{k,k'=1}^{p} \varrho^{|k-k'|}\varrho^{|k-l|}\varrho^{|k'-l|}\mathbb{1}_{\{l=l'\}} \\ &= \sum_{k=1}^{p} \varrho^{2|k-l|} + 2\sum_{k<k'} \varrho^{|k'-k|+|k-l|+|k'-l|} \\ &= \frac{1}{1-\varrho^2}\left(1+\varrho^2-\varrho^{2l}-\varrho^{2(p-l+1)}\right) + 2\sum_{k<k'} \varrho^{|k'-k|+|k-l|+|k'-l|}. \end{aligned}$$

Observe, that

$$\begin{aligned} \sum_{l=1}^{p} \beta_l^2 \varrho^{2(p-l+1)} &= \sum_{l=1}^{\lfloor\sqrt{p}\rfloor} \beta_l^2 \varrho^{2(p-l+1)} + \sum_{l=\lfloor\sqrt{p}\rfloor+1}^{p} \beta_l^2 \varrho^{2(p-l+1)} \\ &\leq \varrho^{2(p-\lfloor\sqrt{p}\rfloor+1)} \sum_{l=1}^{\lfloor\sqrt{p}\rfloor} \beta_l^2 + \varrho^2 \sum_{l=\lfloor\sqrt{p}\rfloor+1}^{p} \beta_l^2 \to 0. \end{aligned} \tag{A.25}$$

Hence,

$$\begin{aligned} \sum_{l=1}^{p} \beta_l^2 \sum_{k=1}^{p} \varrho^{2|k-l|} &\to \frac{1+\varrho^2}{1-\varrho^2}\sum_{k=1}^{\infty} \beta_l^2 - \frac{1}{1-\varrho^2}\sum_{k=1}^{\infty} \beta_l^2 \varrho^{2l} \\ &= \beta(1)\frac{1+\varrho^2}{1-\varrho^2} - \beta(\varrho^2)\frac{1}{1-\varrho^2}. \end{aligned} \tag{A.26}$$

Similarly,

$$\begin{aligned} \sum_{k<k'} \varrho^{|k'-k|+|k-l|+|k'-l|} &= \sum_{k<k'} \varrho^{2l-2k}\mathbb{1}_{\{l\geq k\}} + \sum_{k<k'} \varrho^{2k'-2k}\mathbb{1}_{\{k<l\leq k'\}} \\ &\quad + \sum_{k<k'} \varrho^{2k'-2l}\mathbb{1}_{\{k'>l\}}. \end{aligned}$$

The first term can be rewritten,

$$\sum_{k<k'} \varrho^{2l-2k}\mathbb{1}_{\{l\geq k\}} = \sum_{k=1}^{l}\sum_{k'=k+1}^{l} \varrho^{2l-2k} = \frac{\varrho^{2l}\left(-l+l\varrho^2-\varrho^2\right)+\varrho^2}{\left(1-\varrho^2\right)^2}.$$

We get

$$\sum_{l=1}^{p}\beta_l^2\sum_{k<k'}\varrho^{2l-2k} = \sum_{l=1}^{p}\beta_l^2\frac{\varrho^{2l}\left(-l+l\varrho^2-\varrho^2\right)+\varrho^2}{\left(1-\varrho^2\right)^2}$$
$$\to \frac{\varrho^2}{(1-\varrho^2)^2}\beta(1) - \frac{1}{1-\varrho^2}\beta^{(1)}(\varrho^2) - \frac{\varrho^2}{(1-\varrho^2)^2}\beta(\varrho^2). \quad \text{(A.27)}$$

Similarly,

$$\sum_{k<k'}\varrho^{2k'-2k}\mathbb{1}_{\{k<l\le k'\}} = \sum_{k=1}^{l}\sum_{k'=l+1}^{p}\varrho^{2k'-2k}$$
$$= -\frac{\varrho^{2(p-l+1)}}{\left(1-\varrho^2\right)^2} - \frac{\varrho^{2(l+1)}}{\left(1-\varrho^2\right)^2} + \frac{\varrho^{2(p+1)}}{\left(1-\varrho^2\right)^2} + \frac{\varrho^2}{\left(1-\varrho^2\right)^2},$$

where due to (A.25),

$$\sum_{l=1}^{p}\beta_l^2\sum_{k<k'}\varrho^{2k'-2k}\mathbb{1}_{\{k<l\le k'\}} \to \sum_{l=1}^{\infty}\beta_l^2\left(\frac{\varrho^2}{\left(1-\varrho^2\right)^2} - \frac{\varrho^{2(l+1)}}{\left(1-\varrho^2\right)^2}\right)$$
$$= \frac{\varrho^2}{(1-\varrho^2)^2}\beta(1) - \frac{\varrho^2}{(1-\varrho^2)^2}\beta(\varrho^2). \quad \text{(A.28)}$$

Finally, observe that

$$\sum_{k<k'}\varrho^{2k'-2l}\mathbb{1}_{\{k'<l\}} = \sum_{k=l+1}^{p-1}\sum_{k'=k+1}^{p}\varrho^{2k'-2l}$$
$$= \frac{(p-l)\varrho^{2(p-l+1)}}{\left(1-\varrho^2\right)^2} - \frac{(p-l+1)\varrho^{2(p-l+2)}}{\left(1-\varrho^2\right)^2}$$
$$+ \frac{\varrho^4}{\left(1-\varrho^2\right)^2},$$

where due to (A.25), it remains to see that, as $p\to\infty$,

$$\sum_{l=1}^{p}\beta_l^2\sum_{k<k'}\varrho^{2k'-2l}\mathbb{1}_{\{k'<l\}} \to \sum_{l=1}^{\infty}\beta_l^2\frac{\varrho^4}{(1-\varrho^2)^2} = \frac{\varrho^4}{(1-\varrho^2)^2}\beta(1). \quad \text{(A.29)}$$

Finally, by collecting the terms of (A.27), (A.28) and (A.29) and simplifying, we get

$$\begin{aligned}\lim_{p\to\infty}\sum_{l=1}^{p}\beta_l^2\sum_{k<k'}^{p}\varrho^{|k'-k|+|k-l|+|k'-l|} &= \frac{\varrho^4}{(1-\varrho^2)^2}\beta(1)-\frac{\varrho^2}{(1-\varrho^2)^2}\beta(\varrho^2)\\ &+\ \frac{2\varrho^2}{(1-\varrho^2)^2}\beta(1)\ -\ \frac{1}{1-\varrho^2}\beta^{(1)}(\varrho^2)\ -\ \frac{\varrho^2}{(1-\varrho^2)^2}\beta(\varrho^2)\\ &=\ \beta(1)\Big(\frac{\varrho^2(\varrho^2+2)}{(1-\varrho^2)^2}\Big)-\beta(\varrho^2)\Big(\frac{2\varrho^2}{(1-\varrho^2)^2}\Big)\ -\ \frac{\beta^{(1)}(\varrho^2)}{1-\varrho^2}.\end{aligned}$$

Therefore, from (A.26), (A.27), (A.28) and (A.29),

$$\begin{aligned}\sum_{l=1}^{p}\beta_l^2 J_1(l)\ &\to\ \beta(1)\frac{1+\varrho^2}{1-\varrho^2}\ -\ \frac{1}{1-\varrho^2}\beta(\varrho^2)\\ &+\ 2\Big(\beta(1)\frac{\varrho^2(\varrho^2+2)}{(1-\varrho^2)^2}\ -\ \beta(\varrho^2)\,\frac{2\varrho^2}{(1-\varrho^2)^2}\ -\ \frac{1}{1-\varrho^2}\beta^{(1)}(\varrho^2)\Big)\\ &=\ \beta(1)\,\frac{\varrho^4+4\varrho^2+1}{(1-\varrho^2)^2}\ -\ \beta(\varrho^2)\,\frac{3\varrho^2+1}{(1-\varrho^2)^2}\ -\ \frac{2}{1-\varrho^2}\beta^{(1)}(\varrho^2),\end{aligned}$$

which concludes the proof of (A.14).

Next, consider $J_2(l,l')$ in (A.13). According to the arrangement of indices k,k,l,l', we have 9 cases:

1. $k,k'\in\{1,\dots,l\}$,
2. $k\le l,\ l<k'\le l'$,
3. $k\le l,\ l'<k'\le p$,
4. $l<k\le l',\ 1\le k'\le l$,
5. $k,k'\in\{l+1,\dots,l'\}$,
6. $l<k\le l',\ l'<k'\le p$,
7. $l'<k\le p,\ 1\le k'\le l$,
8. $l'<k\le p,\ l<k'\le l'$,
9. $k,k'\in\{l'+1,\dots,p\}$.

Case 1:

$$\sum_{k,k'=1}^{p} \varrho^{|k-k'|}\varrho^{|k-l|}\varrho^{|k'-l'|}\mathbb{1}_{\{k,k'\leq l\}} = \varrho^{l+l'}\sum_{k,k'=1}^{l}\varrho^{|k-k'|-k-k'}$$
$$= \frac{(-(2l+1)\varrho^{2l}+(2l-1)\varrho^{2l+2}+\varrho^2+1)\varrho^{l'-l}}{(1-\varrho^2)^2}.$$

Case 2:

$$\sum_{k,k'=1}^{p} \varrho^{|k-k'|}\varrho^{|k-l|}\varrho^{|k'-l'|}\mathbb{1}_{\{k\leq l, l<k'\leq l'\}} = \varrho^{l+l'}\sum_{k=1}^{l}\sum_{k'=l+1}^{l'}\varrho^{-2k}$$
$$= \frac{(1-\varrho^{2l})(l'-l)\varrho^{l'-l}}{1-\varrho^2}.$$

Case 3:

$$\sum_{k,k'=1}^{p} \varrho^{|k-k'|}\varrho^{|k-l|}\varrho^{|k'-l'|}\mathbb{1}_{\{k\leq l, l'<k'\leq p\}} = \varrho^{l-l'}\sum_{k=1}^{l}\sum_{k'=l'+1}^{p}\varrho^{-2k+2k'}$$
$$= \frac{(1-\varrho^{2l})\varrho^{-l'-l+2}(\varrho^{2l'}-\varrho^{2p})}{(1-\varrho^2)^2}.$$

Case 4:

$$\sum_{k,k'=1}^{p} \varrho^{|k-k'|}\varrho^{|k-l|}\varrho^{|k'-l'|}\mathbb{1}_{\{l<k\leq l', 1\leq k'\leq l\}} = \varrho^{-l+l'}\sum_{k=l+1}^{l'}\sum_{k'=1}^{l}\varrho^{2k-2k'}$$
$$= \frac{(1-\varrho^{2l})\varrho^{l'-3l+2}(\varrho^{2l}-\varrho^{2l'})}{(1-\varrho^2)^2}.$$

Case 5:

$$\sum_{k,k'=1}^{p} \varrho^{|k-k'|}\varrho^{|k-l|}\varrho^{|k'-l'|}\mathbb{1}_{\{l<k,k'\leq l'\}} = \varrho^{-l+l'}\sum_{k,k'=l+1}^{l'}\varrho^{|k-k'|+k-k'}$$
$$= \varrho^{-l+l'}\left(\sum_{k=l+1}^{l'}\sum_{k':k'\geq k}^{l'}1+\sum_{k=l+1}^{l'}\sum_{k':k'\geq l+1}^{k-1}\varrho^{2k-2k'}\right)$$
$$= \frac{1}{2}\varrho^{l'-3l}\Bigg(\varrho^{2l}(-l'+l-1)(l-l')$$

$$+ \quad \frac{2\varrho^2\big((1-\varrho^2)\varrho^{2l}l' + \varrho^{2l'} - (l+1)\varrho^{2l} + l\varrho^{2l+2}\big)}{(1-\varrho^2)^2}\Bigg).$$

Case 6:

$$\begin{aligned}\sum_{k,k'=1}^{p} \varrho^{|k-k'|}\varrho^{|k-l|}\varrho^{|k'-l'|}\mathbb{1}_{\{l<k\leq l',l'<k'\leq p\}} &= \varrho^{-l-l'}\sum_{k=l+1}^{l'}\sum_{k'=l'+1}^{p}\varrho^{2k'} \\ &= \frac{(l'-l)\,\varrho^{-l'-l+2}(\varrho^{2l'}-\varrho^{2p})}{1-\varrho^2}.\end{aligned}$$

Case 7:

$$\begin{aligned}\sum_{k,k'=1}^{p} \varrho^{|k-k'|}\varrho^{|k-l|}\varrho^{|k'-l'|}\mathbb{1}_{\{l'<k\leq p,1\leq k'\leq l\}} &= \varrho^{-l+l'}\sum_{k=l'+1}^{p}\sum_{k'=1}^{l}\varrho^{2k-2k'} \\ &= \frac{(1-\varrho^{2l})\varrho^{l'-3l+2}(\varrho^{2l'}-\varrho^{2p})}{(1-\varrho^2)^2}.\end{aligned}$$

Case 8: We have

$$\begin{aligned}\sum_{k,k'=1}^{p} \varrho^{|k-k'|}\varrho^{|k-l|}\varrho^{|k'-l'|}\mathbb{1}_{\{l'<k\leq p,l<k'\leq l'\}} &= \varrho^{-l+l'}\sum_{k=l'+1}^{p}\sum_{k'=l+1}^{l'}\varrho^{2k-2k'} \\ &= \frac{\varrho^{-3l-l'+2}(\varrho^{2l}-\varrho^{2l'})(\varrho^{2l'}-\varrho^{2p})}{(1-\varrho^2)^2}.\end{aligned}$$

Case 9: We have

$$\begin{aligned}\sum_{k,k'=1}^{p} \varrho^{|k-k'|}\varrho^{|k-l|}\varrho^{|k'-l'|}\mathbb{1}_{\{l'<k,k'\leq p\}} &= \varrho^{-l-l'}\sum_{k,k'=l'+1}^{p}\varrho^{|k-k'|+k+k'} \\ &= \varrho^{-l-l'}\Bigg(\sum_{k=l'+1}^{p}\sum_{k'=l'+1}^{k}\varrho^{2k} + \sum_{k=l'+1}^{p}\sum_{k'=k+1}^{p}\varrho^{2k'}\Bigg) \\ &= \frac{\varrho^{-l'-l+2}}{(1-\varrho^2)^2}\big(2(1-\varrho^2)l'\varrho^{2p} + \varrho^{2l'} + \varrho^{2l'+2} \\ &\quad - (2p+1)\varrho^{2p} + (2p-1)\varrho^{2p+2}\big).\end{aligned}$$

Summing up the cases 1–9, we get

$$J_2(l,l') \;=\; \frac{1}{2(1-\varrho^2)^2}\Big(\varrho^{l'-l}\big((l'-l)^2(1-\varrho^2)^2 + 3(1-\varrho^4)(l'-l)$$

$$+\ 2(1+4\varrho^2+\varrho^4)\big)\ +\ \varrho^{l+l'}\big(-2(1-\varrho^2)(l'+l)-2-6\varrho^2\big)$$
$$+\ \varrho^{2p-l-l'}\big(2\varrho^2(1-\varrho^2)(l+l'-2p)-6\varrho^2-2\varrho^4\big)$$
$$+\ 2\varrho^2\big(\varrho^{2p+l'-l}+\varrho^{2p+l-l'}\big)\Big).$$

Observe, that

$$\Big|\sum_{l'=2}^{p}\sum_{l=1}^{l'-1}\beta_l\beta_{l'}\varrho^{2p-l-l'}(l+l'-2p)\Big| \ \le\ \Big|\sum_{l'=2}^{p}\sum_{l=1}^{l'-1}\beta_l\beta_{l'}\varrho^{p-l}\varrho^{p-l'}(p-l)\Big| + \Big|\sum_{l'=2}^{p}\sum_{l=1}^{l'-1}\beta_l\beta_{l'}\varrho^{p-l}\varrho^{p-l'}(p-l')\Big|. \tag{A.30}$$

The two summands of (A.30) are symmetric, therefore due to brevity we consider only the first term. The proof for the second term will be analogous. Note,

$$\Big|\sum_{l'=2}^{p}\sum_{l=1}^{l'-1}\beta_l\beta_{l'}\varrho^{p-l}\varrho^{p-l'}(p-l)\Big| \ \le\ \sum_{l'=2}^{p}|\beta_{l'}||\varrho|^{p-l'}\sum_{l=1}^{p-1}|\beta_l||\varrho|^{p-l}(p-l), \tag{A.31}$$

where

$$\sum_{l=1}^{p-1}|\beta_l||\varrho|^{p-l}(p-l)\ \le\ \Big(\sum_{l=1}^{p-1}\beta_l^2\Big)^{1/2}\Big(\sum_{l=1}^{p-1}\varrho^{2(p-l)}(p-l)^2\Big)^{1/2}\ <\ \infty$$

holds due to $\sum_{j=1}^{\infty}\beta_j^2<\infty$. It remains to note that

$$\begin{aligned}\sum_{l'=2}^{p}|\beta_{l'}||\varrho|^{p-l'} &= \sum_{l'=2}^{\lfloor\sqrt{p}\rfloor}|\beta_{l'}||\varrho|^{p-l'}+\sum_{l'=\lfloor\sqrt{p}\rfloor+1}^{p}|\beta_{l'}||\varrho|^{p-l'}\\ &\le |\varrho|^{p-\lfloor\sqrt{p}\rfloor}p^{1/4}\Big(\sum_{l'=2}^{\lfloor\sqrt{p}\rfloor}\beta_{l'}^2\Big)^{1/2}\\ &\quad+\Big(\sum_{l'=\lfloor\sqrt{p}\rfloor+1}^{p}\beta_{l'}^2\Big)^{1/2}\Big(\sum_{l'=\lfloor\sqrt{p}\rfloor+1}^{p}\varrho^{2(p-l')}\Big)^{1/2}\\ &\to 0,\end{aligned}$$

since $\sum_{l'=2}^{\lfloor\sqrt{p}\rfloor} \beta_{l'}^2 < \infty$ and $\sum_{l'=\lfloor\sqrt{p}\rfloor+1}^{p} |\varrho|^{2(p-l')} < \infty$. Thus, by (A.30) and (A.31), it follows that

$$\Big|\sum_{l'=2}^{p}\sum_{l=1}^{l'-1} \beta_l\beta_{l'}\varrho^{2p-l-l'}(l+l'-2p)\Big| \;\to\; 0, \text{ as } p\to\infty. \qquad \text{(A.32)}$$

Therefore, by DCT, as $p\to\infty$, we have

$$\begin{aligned}
\sum_{\substack{l,l'=1\\ l'>l}}^{p} \beta_l\beta_{l'}J_2(l,l') \;&\to\; \frac{1}{2(1-\varrho^2)^2}\Bigg(\sum_{l'>l}\beta_l\beta_{l'}\varrho^{l'-l}\big((l'-l)^2(1-\varrho^2)^2 \\
&\quad + 3(1-\varrho^4)(l'-l) + 2(1+4\varrho^2+\varrho^4)\big) \\
&\quad + \sum_{l'>l}\beta_l\beta_{l'}\varrho^{l'+l}\big(-2(1-\varrho^2)(l'+l) - 2 - 6\varrho^2\big) \\
&= \frac{1}{2(1-\varrho^2)^2}\big(b^{(2)}(\varrho)(1-\varrho^2)^2 + 3b_1^{(1)}(\varrho)(1-\varrho^4) \\
&\quad + 2b_1(\varrho)(1+4\varrho^2+\varrho^4) - 2b_2^{(1)}(\varrho)(1-\varrho^2) \\
&\quad - 2b_2(\varrho)(1+3\varrho^2)\big),
\end{aligned}$$

which concludes the proof of (A.15). □

Appendix B

Details of pseudo-real-time experiments

B.1 Data preparation

Due to the nature of $p \gg n$ problem, it is relatively easy for automatic selection algorithms such as LASSO to find the spurious signal in the noise. Applying linear regression, we could discard such noise variables through cross-validation. However, the cross-validation becomes problematic when shrinkage is applied, since the spurious effects get smoothed out in out-of-sample forecasts. For this reason, we use a strict set of rules for variable pre-screening.

First, by performing the *Kwiatkowski–Phillips–Schmidt–Shin* (KPSS, Kwiatkowski et al. (1992)) test, the stationarity of the time series was tested (with 5% significance). Second, the test for the unit-roots was performed by using the *Augmented Dickey-Fuller* (ADF) test, accounting for the probable deterministic part of the data. Additionally, since the test statistic of the ADF test is the estimated value of the $t-$statistics from an auxiliary regression model, its resulting residuals were also inspected for the possible presence of heteroscedasticity. If we could not reject the heteroscedasticity hypothesis, according to *Breusch-Pagan* (BP) test with 5% significance level, the estimated value of the t-statistics might be biased. Therefore, in such cases, we additionally perform nonparametric *Philips-Perron* (PP) unit-root test, which can correct the possibly incorrect results of the ADF test by bootstrapping the critical values of the test statistic.

In all cases, we determined the number of lags used in the auxiliary regressions 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 whole stationarity testing procedure was iterated until the final series was found as statistically significantly stationary[1].

Since most of the macroeconomic indicators used in the study follow multiplicative processes, it is instructive to apply logarithmic transformation where relevant. By using the logarithmic transformation, we transform the underlying multiplicative processes into additive, removing most of the explosive effects and normalizing the variability of the data. Finally, log-transformation does not alter the interpretation of the variables a lot, because the differences of log-transformed data approximate the percentage growths of the original data.

We test the relevance of logarithmic transformation by estimating and interpreting the parameter of Box-Cox transformation (Box and Cox (1964)). A logarithmic transformation was applied if the Box-Cox parameter estimate of the analysed indicator is smaller than 0.8. The higher threshold value was chosen to safeguard against possible structural breaks in the trend of the data. Transformations of $\frac{x^q-1}{q}$, $q \in (0, 1)$, were not used in this dissertation since the focus is on extracting and distinguishing the multiplicative effects if such were present, instead of just normalizing the data.

Finally, some of the available data has relatively large spikes (outliers) at particular periods, with comparably small volatility during the other remaining periods. Therefore, such a variable may be included in 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[2]. Therefore, we applied an additional heuristic rule to filter such variables from the final dataset: the variable was not included in the final dataset if the ratio of maximum to the average value, when adjusted by standard deviation, was higher than 10. We found that the inclusion of such variables to the

[1] It can be noted that no series required more than 2 differences to achieve stationarity.

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

final dataset resulted in worsened forecasting accuracy, especially during the crisis periods, when they were included in the models as dummy variables to explain the sudden shock.

B.2 Methodological details

B.2.1 Dynamic factor models

We follow the approach described by Giannone et al. (2008), Angelini et al. (2011) and Banbura et al. (2011). Assume the notation used in Chapter 3, i.e., that $X_t = (X_{t,1}, \ldots, X_{t,p})$, $t = 1, \ldots, n$, is a $1 \times p$ vector representing p stationary monthly time series, which have been standardized to mean zero and variance one. The general specification of the dynamic factor model is given by:

$$X_t' = \Lambda F_t' + \xi_t, \quad \xi_t \sim \mathcal{N}_p(0, \Sigma_\xi), \tag{B.1}$$

$$F_t' = \sum_{j=1}^{m} A_j F_{t-j}' + G\eta_t, \quad \eta_i \sim \mathcal{N}_q(0, I_q), \tag{B.2}$$

where F_t' is a $r \times 1$ vector of unobserved common factors and ξ_t is a $p \times 1$ vector of idiosyncratic components, which is assumed to be a multivariate white noise with diagonal covariance matrix Σ_ξ. Λ denotes the $p \times r$ factor loading matrix of the variables. Furthermore, $A_1, \ldots, A_m$ are $r \times r$ matrices of autoregressive coefficients, and $G = [I_q, \ 0_{q\times(r-q)}]'$ is a $r \times q$ matrix of full rank q, with η_t describing the white noise process of the shocks to the common factors. In such a model, the number of common factors (r) that is large relative to the number of common shocks (q) aims at capturing the lead and lag relations among variables along the business cycle.

In this dissertation we assume that $m = 2$ in order to make the results comparable between other considered modelling approaches. Furthermore, the number of significant factors r is estimated based on the information criterion by Bai and Ng (2002), while the number of shocks q is selected following Bai and Ng (2007). See Section B.2.3 for detailed specifications behind the choice of the aforementioned parameters.

Finally, the nowcasting model is estimated through (B.1)–(B.2) by employing a two-step Kalman Filter approach and a quarterly aggregation scheme, following Angelini et al. (2011).

B.2.2 Diffusion index models

We follow the Diffusion Index approach proposed by Stock and Watson (2002). The method first uses principal component analysis to estimate r factors F_t' from the predictors X_t', $t = 1, \dots, n$. Then, for a h-step-ahead forecast we consider the model

$$Y_{t+h} = \alpha_h + \beta_h(L)F_t' + \gamma_h(L)Y_t + \varepsilon_{t+h},$$

where the orders for $\beta_h(L)$ and $\gamma_h(L)$ are selected by BIC. Note, that here $\beta_h(L)$ and $\gamma_h(L)$ are lag polynomials, modelled as having finite orders, with L denoting the lag operator, i.e., $LF_t = F_{t-1}$. Furthermore, here Y_{t+h} is the $h-$step ahead variable to be forecasted.

The forecasts of Y_{n+h} are constructed using a two-step procedure. First, the sample data $\{X_t\}_{t=1}^n$ are used to estimate a time series of factors (the diffusion indexes), $\{F_t'\}_{t=1}^n$. Second, the estimators $\hat{\alpha}_h$, $\hat{\beta}_h(L)$, $\hat{\gamma}_h(L)$ are obtained by regressing Y_{t+h} onto a constant, F_t' and Y_t. The forecast of Y_{n+h} is then formed as $\hat{\alpha}_h + \hat{\beta}_h(L)F_n' + \hat{\gamma}_h(L)Y_n$.

B.2.3 Dynamic factor model parameter specification

Consider a factor model and the notation used in Section B.2.1. Assume that $\hat{F}^{(r)}$ denotes the $n \times r$ matrix of r estimated factors, e.g., by using principal components. Also, assume that $\hat{\Lambda}$ denotes the corresponding $r \times r$ loading matrix. Denote,

$$\hat{\sigma}_r^2 := \frac{1}{np}\sum_{i=1}^{n}\sum_{j=1}^{p}(X_{ij} - \hat{F}_i^{(r)}\hat{\lambda}_j^{(r)})^2, \tag{B.3}$$

where $\hat{F}_i^{(r)}$ is an $1 \times r$ vector, $r \leq p$, $i = 1, \dots, n$, and $\hat{\lambda}_j^{(r)}$ is an $r \times 1$ vector, corresponding to the $j-$th row of the loading matrix $\hat{\Lambda}$.

A model with $r+1$ factors can fit the data no worse than a model with r factors, but efficiency is lost as more factor loadings are being estimated. Thus, Bai and Ng (2002) define the following information criteria in order to choose the optimal number of factors r:

$$IC_1(r) := \ln(\hat{\sigma}_r^2) + r\Big(\frac{n+p}{np}\Big)\ln\Big(\frac{np}{n+p}\Big), \tag{B.4}$$

$$IC_2(r) := \ln(\hat{\sigma}_r^2) + r\Big(\frac{n+p}{np}\Big)\ln\min(n,p), \tag{B.5}$$

$$IC_3(r) := \ln(\hat{\sigma}_r^2) + r\Big(\frac{\ln \min(n,p)}{\min(n,p)}\Big), \quad \text{(B.6)}$$

where $\hat{\sigma}_r^2$ is given by (B.3). The chosen number of factors r will then correspond to $\arg\min_r IC_i(r)$ for $i = 1, 2, 3$. Bai and Ng (2002) suggests that the information criterions (B.4)–(B.6) are asymptotically equivalent, however may nevertheless give different results for finite samples.

Next, Bai and Ng (2007) define the selection of number of shocks q. Assume that the number of factors r is chosen by some information criteria (B.4)–(B.6), as described above. Given the estimated factors $\hat{F}_t^{(r)}$, $t = 1, \ldots, n$, the authors propose forming an r-dimensional VAR model by using $\hat{F}_t^{(r)}$. Using the specified VAR model, one should obtain the model residuals $\hat{u}_t$. Selecting too few lags might be problematic, thus the authors consider a VAR(2) model, suggesting that the results using higher order models should be similar. Given $\hat{u}_t$, a $r \times r$ covariance matrix $\hat{\Sigma}_u$ is constructed. Define the eigenvalues of $\hat{\Sigma}_u$ as $c_1 > c_2 \geq ... \geq c_r \geq 0$. Let

$$\hat{D}_k := \Big(\frac{c_{(k+1)}^2}{\sum_{j=1}^r c_j^2}\Big)^{1/2}, \quad k = 1, \ldots, r.$$

Then, for some $0 < m < \infty$ and $0 < \delta < 1/2$, define

$$K := \{k : \hat{D}_k < m / \min(n^{1/2-\delta}, p^{1/2-\delta})\}.$$

The authors propose choosing the number of shocks as $q = \min\{k \in K\}$. Furthermore, it is observed that choosing values $m = 1$ and $\delta = 0.1$ works well in practice.

B.2.4 Forecast accuracy comparison

Given observed performance records of any models in consideration, it is inevitably the case that one set of forecasts will appear more accurate than the other, even if only by a small margin. Therefore, a statistical test is needed in order to answer whether these observed differences are statistically significant. In this dissertation we employ two such tests, namely the Diebold-Mariano and Giacomini-White tests, presented in Sections B.2.4.1 and B.2.4.2, respectively.

B.2.4.1 Diebold-Mariano test

The Diebold-Mariano (DM, Diebold and Mariano (1995)) test is used for comparing forecast accuracy of any two methods of consideration. In this dissertation we employ the modified version of the test, as proposed by Harvey et al. (1997).

The original Diebold and Mariano (1995) framework of the test is described as follows. Suppose that a pair of h-steps-ahead forecasts have produced errors $(e_{1,t}, e_{2,t})$, $t = 1, \ldots, n$. The quality of a forecast is judged based on some specified function $g(e)$ of the forecast error e. Typically $g(e) = e^2$. The authors specify the following null hypothesis of equality of expected forecast performance:

$$H_0: \qquad \mathbb{E}(g(e_{1,t}) - g(e_{2,t})) \;=\; 0.$$

Thus, by defining $d_t := g(e_{1,t}) - g(e_{2,t})$, $t = 1, \ldots, n$, the test is constructed based on the sample mean:

$$\bar{d} = n^{-1} \sum_{t=1}^{n} d_t.$$

A known difficulty is that the series d_t is likely to be autocorrelated for any reasonably well-conceived set of forecasts, as is demonstrated by Harvey et al. (1997). Define,

$$\mathrm{V}(\bar{d}) \;:=\; n^{-1}\Big(\gamma_0 + 2\sum_{k=1}^{h-1} \gamma_k\Big), \tag{B.7}$$

where γ_k is the k-th autocovariance of d_t. These autocovariances can be estimated by:

$$\hat{\gamma}_k \;=\; n^{-1} \sum_{t=k+1}^{n} (d_t - \bar{d})(d_{t-k} - \bar{d}). \tag{B.8}$$

Thus, the DM test statistic is then defined as:

$$S_1 \;=\; \frac{\bar{d}}{\sqrt{\hat{\mathrm{V}}(\bar{d})}}, \tag{B.9}$$

where $\hat{\mathrm{V}}(\bar{d})$ is obtained by substituting (B.8) into (B.7). Under the null hypothesis, the statistic S_1 has an asymptotic standard normal distribution. Diebold and Mariano (1995) found that for moderately large samples the performance of the test was satisfactory when using $g(e) = e^2$. However, the test appeared to be over-sized for moderate numbers of sample observations.

In order to address certain shortcomings, Harvey et al. (1997) propose a set of modifications to the original test, resulting in the following statistic:

$$S_1^* = \left(\frac{n + 1 - 2h + n^{-1}h(h-1)}{n}\right)^{1/2} S_1,$$

which follows from employing an approximately unbiased estimator of the variance of $\bar{d}$, here S_1 is defined by (B.9). The authors propose that the critical values for the test should be chosen from the Student's t distribution with $(n-1)$ degrees of freedom.

B.2.4.2 Giacomini-White test

Giacomini and White (2006) expand the framework behind the DM test by considering a conditional approach for out-of-sample predictive ability testing. They focus on limited memory estimators, whose finite sample properties are preserved asymptotically. Such a setup arises, for instance, when the forecasts are generated by a "rolling window" forecasting approach, which is a common strategy in many applications.

Assuming a specified forecast horizon $h > 0$, the authors propose testing the following conditional hypothesis:

$$H_0 : \quad \mathbb{E}(g(e_{1,t+h}) - g(e_{2,t+h}) \mid \mathcal{G}_t) = 0,$$

where $\mathcal{G}_t$ is some set of information.

The "rolling window" estimation scheme is defined as follows. Let n be the total sample size, m being the size of the first estimation window. The first $h-$step-ahead forecast is formulated at time m using data indexed $1, \ldots, m$, and these forecasts are compared against the observation Y_{m+h}. At time $m+1$ the second set of forecasts is formulated using the previous m observations (although in a general case the window size can vary), which are then compared with the realization Y_{m+h+1}. Iterating this procedure yields $q := n - h - m + 1$ out-of-sample forecasts and relative forecast errors.

The sequence of out-of-sample forecasts is evaluated by some loss function g. In this dissertation we use $g(e) = e^2$. Define $d_t := g(e_{1,t}) - g(e_{2,t})$, $t = 1, \ldots, n$.

For $h = 1$, the conditional test statistic is defined as:

$$T_1 \; := \; q\bar{Z}_q' \hat{\Omega}_q^{-1} \bar{Z}_q,$$

where

$$\bar{Z}_q \; := \; q^{-1} \sum_{t=m}^{n-1} Z_{t+1}, \qquad Z_{t+1} \; := \; h_t d_{t+1},$$

and h_t is some $r \times 1$ test function, controlling the information included in the conditional set $\mathcal{G}_t$. The authors consider $h_t = (1, d_t)'$ for the test function, meanwhile, setting $h_t = 1$ would reduce the test to the unconditional variant. Finally,

$$\hat{\Omega}_q \; = \; q^{-1} \sum_{t=m}^{n-1} Z_{t+1} Z_{t+1}'.$$

Giacomini and White (2006) demonstrate that $T_1 \xrightarrow{d} \chi^2_r$, as $q \to \infty$.

Similarly, for $h > 1$, the conditional test statistic is defined as:

$$T_2 \; := \; q\bar{Z}_q' \hat{\Omega}_q^{-1} \bar{Z}_q,$$

where

$$\bar{Z}_q \; := \; q^{-1} \sum_{t=m}^{n-h} Z_{t+h}, \qquad Z_{t+h} \; := \; h_t d_{t+h},$$

and $\hat{\Omega}_q$ is a $r \times r$ covariance matrix of Z_{t+h}, estimated by Newey and West (1986) procedure. I.e., Giacomini and White (2006) specify it as follows:

$$\begin{aligned} \hat{\Omega}_q \; := \; & q^{-1} \sum_{t=m}^{n-h} Z_{t+h} Z_{t+h}' \\ & + \; q^{-1} \sum_{j=1}^{h-1} \sum_{t=m+j}^{n-h} w_j [Z_{t+h} Z_{t+h-j}' + Z_{t+h-j} Z_{t+h}'], \end{aligned}$$

where w_j is an appropriately set weight function, e.g., $w_j = 1 - j/h$. The authors demonstrate that $T_2 \xrightarrow{d} \chi^2_r$, as $q \to \infty$.

B.3 Main variables used in the modelling

Table B.4: US Exports (rolling window): acronyms and full names of all of the variables, used in presenting the modelling results. The acronyms correspond to those originally used in the FRED database.

Acronym	Transf.	Lag	Description
A36CTI	$\Delta^1 \ln$	2	Value of Manufacturers' Total Inventories for Durable Goods Industries: Transportation Equipment: Heavy Duty Trucks
ANXAVS	Δ^1	1	Value of Manufacturers' Shipments for Capital Goods: Nondefense Capital Goods Excluding Aircraft Industries
AUENSA	$\Delta^1 \ln$	2	Auto Exports
BOPTEXP	$\Delta^1 \ln$	2	Exports of Goods and Services, Balance of Payments Basis
CODENV5URN	$\Delta^1 \ln$	2	Unemployment Rate in Denver County/city, CO
EXBZUS	$\Delta^1 \ln$	1	Brazil / U.S. Foreign Exchange Rate
IPG3344S	$\Delta^2 \ln$	1	Industrial Production: Durable Goods: Semiconductor and other electronic component
IPHITEK2N	$\Delta^2 \ln$	1	Industrial Production: Computers, communications equipment, and semiconductors
IR3TIB01ESM156N	Δ^1	1	3-Month or 90-day Rates and Yields: Interbank Rates for Spain©
IR3TIB01NLM156N	Δ^1	1	3-Month or 90-day Rates and Yields: Interbank Rates for the Netherlands©
KORPROINDMISMEI	$\Delta^1 \ln$	2	Production of Total Industry in Korea©
PLEADUSDM	$\Delta^1 \ln$	1	Global price of Lead©
PORT941NAN	Δ^1	1	All Employees: Total Nonfarm in Portland-Vancouver-Hillsboro, OR-WA (MSA)
SMU06310805000000001	$\Delta^1 \ln$	1	All Employees: Information in Los Angeles-Long Beach-Anaheim, CA (MSA)
SPASTT01KRM661N	$\Delta^1 \ln$	1	Total Share Prices for All Shares for the Republic of Korea©
UCMSNO	Δ^1	2	Value of Manufacturers' New Orders for Construction Materials and Supplies Industries

Table B.5: US Imports (rolling window): acronyms and full names of all of the variables, used in presenting the modelling results. The acronyms correspond to those originally used in the FRED database.

Acronym	Transf.	Lag	Description
A31SNO	$\Delta^1 \ln$	1	Value of Manufacturers' New Orders for Durable Goods Industries: Primary Metals

Table B.5: *(continued)*

Acronym	Transf.	Lag	Description
BSCICP03USM665S		1	Business Tendency Surveys for Manufacturing: Confidence Indicators: Composite Indicators: OECD Indicator for the United States©
CANPROINDMISMEI	Δ^1	3	Production of Total Industry in Canada©
CES2023610001	$\Delta^1 \ln$	1	All Employees: Construction: Residential Building
CEU4348800001	Δ^1	1	All Employees: Transportation and Warehousing: Support Activities for Transportation
CILLCBM027NBOG	$\Delta^1 \ln$	1	Commercial and Industrial Loans, Large Domestically Chartered Commercial Banks
COMPU1UNSA	$\Delta^1 \ln$	1	New Privately-Owned Housing Units Completed: 1-Unit Structures
ESPPROINDMISMEI	$\Delta^1 \ln$	2	Production of Total Industry in Spain©
FRGSHPUSM649NCIS	$\Delta^1 \ln$	1	Cass Freight Index: Shipments©
GACDFNA066MNFRBPHI		0	Current General Activity; Diffusion Index for FRB - Philadelphia District
HOUS448BP1FH	$\Delta^1 \ln$	1	New Private Housing Units Authorized by Building Permits: 1-Unit Structures for Houston-Sugar Land-Baytown, TX (MSA)
I4248IM144SCEN	$\Delta^1 \ln$	2	Merchant Wholesalers, Except Manufacturers' Sales Branches and Offices Sales: Nondurable Goods: Beer, Wine, and Distilled Alcoholic Beverages Inventories
IPB54100S	$\Delta^1 \ln$	1	Industrial Production: Construction supplies
IPDMAT	$\Delta^1 \ln$	1	Industrial Production: Durable Materials
IPG3262N	$\Delta^1 \ln$	1	Industrial Production: Nondurable Goods: Rubber product
IPG326S	Δ^1	1	Industrial Production: Nondurable manufacturing: Plastics and rubber products
IPG332S	$\Delta^1 \ln$	1	Industrial Production: Durable manufacturing: Fabricated metal product
IR3TBB01NZM156N	$\Delta^1 \ln$	1	3-Month or 90-day Rates and Yields: Bank Bills for New Zealand©
MEXPROMANMISMEI	Δ^1	3	Production in Total Manufacturing for Mexico©
NCHALI0URN	$\Delta^1 \ln$	2	Unemployment Rate in Halifax County, NC
NCMFGN	$\Delta^2 \ln$	1	All Employees: Manufacturing in North Carolina
P4232MM157SCEN		2	Merchant Wholesalers, Except Manufacturers' Sales Branches and Offices Sales: Durable Goods: Furniture and Home Furnishings Sales
RICH051UR	$\Delta^1 \ln$	2	Unemployment Rate in Richmond, VA (MSA)
SANF806NAN	$\Delta^2 \ln$	1	All Employees: Total Nonfarm in San Francisco-Oakland-Hayward, CA (MSA)
SCMFGN	Δ^2	1	All Employees: Manufacturing in South Carolina
SMS48151800000000026		1	All Employees: Total Nonfarm in Brownsville-Harlingen, TX (MSA)
SWEPROINDMISMEI	$\Delta^1 \ln$	2	Production of Total Industry in Sweden©
U34HUO	$\Delta^1 \ln$	2	Value of Manufacturers' Unfilled Orders for Durable Goods Industries: Computers and Electronic Products: Electronic Components

Table B.5: *(continued)*

Acronym	Transf.	Lag	Description
UMDMIS	$\Delta^1 \ln$	2	Ratio of Manufacturers' Total Inventories to Shipments for Durable Goods Industries
USFIRE	Δ^2	1	All Employees: Financial Activities
VAUR	$\Delta^1 \ln$	1	Unemployment Rate in Virginia
WPU071	$\Delta^1 \ln$	1	Producer Price Index by Commodity for Rubber and Plastic Products: Rubber and Rubber Products
XTEXVA01JPM664N	$\Delta^1 \ln$	2	Exports: Value Goods for Japan©
XTEXVA01JPM664S	$\Delta^1 \ln$	2	Exports: Value Goods for Japan©
XTEXVA01KRM667S	$\Delta^1 \ln$	2	Exports: Value Goods for the Republic of Korea©

B.4 Tables

Table B.6: RMSE of models forecasts during rolling window pseudo-real-time experiments for US PFCE, Exports and Imports; here the bolded values are the smallest ones for every row during the full time period of 2005Q1–2018Q4

	LP	**LPX**	**RL**	**SqL**	**SqP**	**AdP**	**AdL**	**AdRL**	**SW**	**BNDF**	**DF**	**ARMA**
PFCE:												
Backcast	0.08	0.07	0.07	0.14	0.10	**0.03**	0.04	0.03	0.44	0.27	0.37	0.36
Nowcast	0.29	0.29	0.29	**0.28**	0.29	0.29	0.29	0.29	0.53	0.30	0.48	0.38
Forecast-1Q	0.38	0.38	0.38	0.40	**0.37**	0.39	0.39	0.39	0.57	0.37	0.54	0.42
Forecast-2Q	**0.43**	0.44	0.44	0.46	0.44	0.44	0.44	0.44	0.69	0.46	0.55	0.47
Exports:												
Backcast	0.46	0.49	0.48	0.70	0.51	0.20	0.25	**0.15**	1.28	1.00	1.35	1.66
Nowcast	1.40	**1.35**	1.43	1.38	1.51	1.48	1.46	1.48	1.69	1.36	1.78	2.14
Forecast-1Q	2.19	2.20	2.21	2.11	2.18	2.12	2.22	2.19	2.25	**2.00**	2.13	2.32
Forecast-2Q	2.32	2.31	2.33	2.27	2.34	2.31	2.34	2.35	2.50	2.38	**2.15**	2.32
Imports:												
Backcast	0.35	0.32	0.29	0.47	0.30	0.13	0.15	**0.10**	1.81	1.04	1.36	1.88
Nowcast	1.29	1.24	1.20	1.24	1.26	1.21	1.21	**1.19**	1.42	1.37	1.95	2.23
Forecast-1Q	1.75	1.69	1.66	**1.60**	1.63	1.70	1.63	1.62	2.25	1.95	2.35	2.84
Forecast-2Q	2.23	2.25	2.26	2.25	**2.16**	2.27	2.28	2.25	2.66	2.74	2.38	3.02

Table B.1: US GFCF (expanding window): acronyms and full names of all of the variables, used in presenting the modelling results. The acronyms correspond to those originally used in the FRED database.

Acronym	Transf.	Lag	Description
ANXAVS	Δ^1	1	Value of Manufacturers' Shipments for Capital Goods: Nondefense Capital Goods Excluding Aircraft Industries
CACONS	$\Delta^1 \ln$	1	All Employees: Construction in California
CES6562000101	Δ^1	1	All Employees: Education and Health Services: Health Care
CUUR0000SETC	$\Delta^1 \ln$	1	Consumer Price Index for All Urban Consumers: Motor vehicle parts and equipment
IAWARR1URN	$\Delta^1 \ln$	2	Unemployment Rate in Warren County, IA
IMPMX	$\Delta^1 \ln$	2	U.S. Imports of Goods from Mexico, Customs Basis
IPDMAN	$\Delta^1 \ln$	1	Industrial Production: Durable Manufacturing (NAICS)
KYMADI0URN	$\Delta^1 \ln$	2	Unemployment Rate in Madison County, KY
LNS12027659	$\Delta^1 \ln$	1	Employment Level: Less than a High School Diploma, 25 years and over
LNU07400000	$\Delta^1 \ln$	1	Labor Force Flows Employed to Unemployed: 16 Years and Over
LREM25MAUSM156S	Δ^1	1	Employment Rate: Aged 25-54: Males for the United States
MAURN	$\Delta^1 \ln$	1	Unemployment Rate in Massachusetts
MINAN	Δ^1	1	All Employees: Total Nonfarm in Michigan
MNBP1FH	$\Delta^1 \ln$	1	New Private Housing Units Authorized by Building Permits: 1-Unit Structures for Minnesota
MOBPPRIVSA	$\Delta^1 \ln$	1	New Private Housing Units Authorized by Building Permits for Missouri
MVMTD027MNFRBDAL	$\Delta^2 \ln$	1	Market Value of Marketable Treasury Debt
PITAGRIDX	Δ	3	Economic Conditions Index for Pittsburgh, PA (MSA)
S4233SM144SCEN	$\Delta^1 \ln$	2	Merchant Wholesalers, Except Manufacturers' Sales Branches and Offices Sales: Durable Goods: Lumber and Other Construction Materials Sales
SEAT653NA	$\Delta^1 \ln$	1	All Employees: Total Nonfarm in Seattle-Tacoma-Bellevue, WA (MSA)
SMS11479000000000026	Δ	1	All Employees: Total Nonfarm in Washington-Arlington-Alexandria, DC-VA-MD-WV (MSA)
SMS34121000000000026	Δ	1	All Employees: Total Nonfarm in Atlantic City-Hammonton, NJ (MSA)
UNXANO	Δ^1	1	Value of Manufacturers' New Orders for Capital Goods: Nondefense Capital Goods Excluding Aircraft Industries
WPU014102	$\Delta^1 \ln$	1	Producer Price Index by Commodity for Farm Products: Slaughter Chickens
WPU054321	$\Delta^1 \ln$	1	Producer Price Index by Commodity for Fuels and Related Products and Power: Industrial Electric Power

Table B.2: US GFCF (rolling window): acronyms and full names of all of the variables, used in presenting the modelling results. The acronyms correspond to those originally used in the FRED database.

Acronym	Transf.	Lag	Description
CES2023800001	$\Delta^1 \ln$	1	All Employees: Construction: Specialty Trade Contractors
CES6562000101	Δ^1	1	All Employees: Education and Health Services: Health Care
INDI918NAN	Δ^1	1	All Employees: Total Nonfarm in Indianapolis-Carmel-Anderson, IN (MSA)
IPG326S	Δ^1	1	Industrial Production: Nondurable manufacturing: Plastics and rubber products
IR3TBB01NZM156N	$\Delta^1 \ln$	1	3-Month or 90-day Rates and Yields: Bank Bills for New Zealand©
KYNAN	Δ^1	1	All Employees: Total Nonfarm in Kentucky
LREM25MAUSM156N	Δ^1	1	Employment Rate: Aged 25-54: Males for the United States©
MINN427NAN	Δ^1	1	All Employees: Total Nonfarm in Minneapolis-St. Paul-Bloomington, MN-WI (MSA)
P4238MM157SCEN		2	Merchant Wholesalers, Except Manufacturers' Sales Branches and Offices Sales: Durable Goods: Machinery, Equipment, and Supplies Sales
PNGASUSUSDM	$\Delta^1 \ln$	1	Global price of Natural Gas, US Henry Hub Gas©
RSBMGESD	Δ^1	1	Retail Trade: Building Materials, Garden Equipment and Supplies Dealers
S4233SM144SCEN	$\Delta^1 \ln$	2	Merchant Wholesalers, Except Manufacturers' Sales Branches and Offices Sales: Durable Goods: Lumber and Other Construction Materials Sales
SMS11479000000000026		1	All Employees: Total Nonfarm in Washington-Arlington-Alexandria, DC-VA-MD-WV (MSA)
SMS34121000000000026		1	All Employees: Total Nonfarm in Atlantic City-Hammonton, NJ (MSA)
TEMPHELPN	Δ^1	1	Professional and Business Services: Temporary Help Services
TLPRVCONS	$\Delta^1 \ln$	2	Total Private Construction Spending
UMDMVS	Δ^1	1	Value of Manufacturers' Shipments for Durable Goods Industries
VTUR	$\Delta^1 \ln$	1	Unemployment Rate in Vermont

Table B.3: US PFCE (rolling window): acronyms and full names of all of the variables, used in presenting the modelling results. The acronyms correspond to those originally used in the FRED database.

Acronym	Transf.	Lag	Description
ACDGNO	Δ^1	2	Value of Manufacturers' New Orders for Consumer Goods: Consumer Durable Goods Industries
CABPPRIV	$\Delta^1 \ln$	1	New Private Housing Units Authorized by Building Permits for California
CES2023700001	$\Delta^1 \ln$	1	All Employees: Construction: Heavy and Civil Engineering Construction
CEU4244100001	Δ^1	1	All Employees: Retail Trade: Motor Vehicle and Parts Dealers
CEU5552000001	Δ^1	1	All Employees: Financial Activities: Finance and Insurance
CEU6562000101	Δ^1	1	All Employees: Education and Health Services: Health Care
CMRMT	$\Delta^1 \ln$	2	Real Manufacturing and Trade Industries Sales
CTUR	$\Delta^1 \ln$	1	Unemployment Rate in Connecticut
DEUCPIALLMINMEI	Δ^1	2	Consumer Price Index of All Items in Germany©
DGDSRC1	$\Delta^1 \ln$	1	Personal Consumption Expenditures: Goods
DSERRA3M086SBEA	Δ^1	1	Real personal consumption expenditures: Services (chain-type quantity index)
GAMFG	Δ^2	1	All Employees: Manufacturing in Georgia
IPG334S	$\Delta^2 \ln$	1	Industrial Production: Durable manufacturing: Computer and electronic product
KYALURN	$\Delta^1 \ln$	2	Unemployment Rate in Allen County, KY
LREM55TTUSM156N	Δ^2	1	Employment Rate: Aged 55-64: All Persons for the United States©
MNBPPRIVSA	Δ^1	1	New Private Housing Units Authorized by Building Permits for Minnesota
PCEC96	$\Delta^1 \ln$	1	Real Personal Consumption Expenditures
PCESC96	$\Delta^1 \ln$	1	Real Personal Consumption Expenditures: Services
PCOCOUSDM	$\Delta^1 \ln$	1	Global price of Cocoa©
RSEAS	Δ^1	1	Retail Trade: Electronics and Appliance Stores
SANF806NA	$\Delta^2 \ln$	1	All Employees: Total Nonfarm in San Francisco-Oakland-Hayward, CA (MSA)
SANF806NAN	$\Delta^2 \ln$	1	All Employees: Total Nonfarm in San Francisco-Oakland-Hayward, CA (MSA)

Table B.7: RMSE of models forecasts during rolling window pseudo-real-time experiments for US PFCE, Exports and Imports; here the bolded values are the smallest ones for every row during the full time period of 2005Q1–2018Q4

	L	**L0L1**	**L0**	**L0L2**	**L0L1p**	**L0p**	**L0L2p**	**SW**	**BNDF**	**DF**	**ARMA**
PFCE:											
Backcast	0.11	0.18	0.25	0.19	**0.17**	0.25	0.18	0.44	0.27	0.37	0.36
Nowcast	0.28	0.29	0.43	0.32	**0.27**	0.42	0.30	0.53	0.30	0.48	0.38
Forecast-1Q	0.39	0.44	0.50	0.43	0.43	0.49	0.42	0.57	**0.37**	0.54	0.42
Forecast-2Q	**0.43**	0.46	0.54	0.47	0.46	0.53	0.48	0.69	0.46	0.55	0.47
Exports:											
Backcast	0.63	0.67	0.99	0.97	**0.55**	0.98	0.86	1.28	1.00	1.35	1.66
Nowcast	1.41	1.41	1.57	1.57	1.45	1.59	1.58	1.69	**1.36**	1.78	2.14
Forecast-1Q	2.12	2.34	2.33	2.41	2.38	2.32	2.45	2.25	**2.00**	2.13	2.32
Forecast-2Q	2.26	2.30	2.29	2.32	2.30	2.29	2.35	2.50	2.38	**2.15**	2.32
Imports:											
Backcast	**0.38**	0.77	1.07	1.02	0.75	1.06	0.83	1.81	1.04	1.36	1.88
Nowcast	**1.21**	1.42	1.68	1.54	1.31	1.72	1.62	1.42	1.37	1.95	2.23
Forecast-1Q	1.66	1.74	2.07	1.74	**1.65**	2.06	1.80	2.25	1.95	2.35	2.84
Forecast-2Q	2.27	**2.21**	2.32	2.35	2.26	2.28	2.35	2.66	2.74	2.38	3.02

Table B.8: Relative RMSE of model forecasts during the rolling window pseudo-real-time experiments for US PFCE against four benchmarks: ARMA, BNDF, DF and SW models. The labels N, Q1 and Q2 denotes the forecast horizons – Nowcast, Forecast-1Q and Forecast-2Q, respectively. The bolded values correspond to the greatest accuracy within the forecast horizon, while the asterix denotes significant performance improvement as suggested by GW test with 5% significance.

Model	**ARMA**			**BNDF**			**DF**			**SW**		
	N	**Q1**	**Q2**	**N**	**Q1**	**Q2**	**N**	**Q1**	**Q2**	**N**	**Q1**	**Q2**
AdL	0.77	0.92	0.94	0.97	1.05	0.96	0.61*	0.72*	0.81	0.54*	0.68*	0.64*
AdP	0.76	0.92	0.92	0.96	1.05	0.94	0.60*	0.72*	0.80	0.54*	0.68*	0.63*
AdRL	0.77	0.92	0.93	0.98	1.05	0.95	0.61*	0.72*	0.80	0.55*	0.68*	0.63*
L	0.75	0.93	0.92	0.95	1.05	0.94	0.60*	0.72*	0.79*	0.53*	0.68*	0.63*
L0	1.13	1.19	1.14	1.43	1.36	1.16	0.89	0.93	0.98	0.80	0.88	0.78*
L0L1	0.76	1.04	0.98	0.96	1.19	1.00	0.60*	0.81	0.84*	0.54*	0.77	0.67*
L0L1p	**0.72***	1.03	0.97	**0.91**	1.17	0.99	**0.57***	0.80	0.83	**0.51***	0.76	0.66*
L0L2	0.85	1.02	0.99	1.07	1.16	1.02	0.67*	0.80	0.86	0.60*	0.75	0.68*
L0L2p	0.80	1.01	1.01	1.01	1.15	1.03	0.63*	0.79	0.87	0.57*	0.75	0.69*
L0p	1.11	1.18	1.13	1.41	1.34	1.15	0.88	0.91	0.97	0.79	0.87	0.77*
LP	0.78	0.90	**0.92**	0.99	1.03	**0.94**	0.62*	0.70*	**0.79***	0.55*	0.67*	**0.63***
LPX	0.77*	0.92	0.94	0.98	1.04	0.96	0.61*	0.71*	0.81*	0.55*	0.68*	0.64*
RL	0.77	0.91	0.92	0.98	1.03	0.94	0.61*	0.71*	0.80	0.55*	0.67*	0.63*
SqL	0.75	0.96	0.96	0.95	1.09	0.98	0.59*	0.74*	0.83	0.53*	0.71*	0.66*
SqP	0.77	**0.89**	0.93	0.97	**1.02**	0.95	0.60*	**0.69***	0.80*	0.54*	**0.66***	0.63*

Table B.9: Relative RMSE of model forecasts during the rolling window pseudo-real-time experiments for US Exports against four benchmarks: ARMA, BNDF, DF and SW models. The labels N, Q1 and Q2 denotes the forecast horizons – Nowcast, Forecast-1Q and Forecast-2Q, respectively. The bolded values correspond to the greatest accuracy within the forecast horizon, while the asterix denotes significant performance improvement as suggested by GW test with 5% significance.

Model	**ARMA**			**BNDF**			**DF**			**SW**		
	N	**Q1**	**Q2**	**N**	**Q1**	**Q2**	**N**	**Q1**	**Q2**	**N**	**Q1**	**Q2**
AdL	0.68*	0.96	1.01	1.07	1.11	0.98	0.82	1.04	1.09	0.86	0.99	0.94
AdP	0.69*	0.91	1.00	1.09	1.06	0.97	0.83	0.99	1.07	0.88	0.94	0.93
AdRL	0.69*	0.94	1.01	1.09	1.10	0.99	0.83	1.03	1.09	0.88	0.97	0.94
L	0.66*	0.91	**0.97**	1.03	1.06	**0.95**	0.79*	0.99	**1.05**	0.83	0.94	**0.90**
L0	0.73	1.00	0.98	1.15	1.16	0.96	0.88	1.09*	1.06	0.93	1.03	0.92
L0L1	0.66*	1.01	0.99	1.04	1.17	0.97	0.79	1.10*	1.07*	0.84	1.04*	0.92
L0L1p	0.68*	1.02	0.99	1.06	1.19	0.97	0.81	1.11*	1.07*	0.86	1.05*	0.92
L0L2	0.73*	1.04	1.00	1.15	1.21	0.98	0.88	1.13*	1.08	0.93	1.07*	0.93
L0L2p	0.74*	1.05	1.01	1.16	1.23	0.99	0.89	1.15*	1.09	0.94	1.09*	0.94
L0p	0.74	1.00	0.98	1.17	1.16	0.96	0.89	1.09	1.06	0.94	1.03	0.91
LP	0.65*	0.94	1.00	1.02	1.09	0.98	0.78	1.02	1.08	0.83	0.97	0.93
LPX	**0.63***	0.95	0.99	**0.99**	1.10	0.97	**0.76**	1.03	1.07	**0.80**	0.98	0.92
RL	0.67*	0.95	1.00	1.05	1.11	0.98	0.80	1.04	1.08	0.85	0.98	0.93
SqL	0.65*	**0.91**	0.98	1.02	**1.06**	0.95	0.78*	**0.99**	1.05	0.82	**0.94**	0.91
SqP	0.70*	0.94	1.01	1.11	1.09	0.98	0.85	1.02	1.09	0.89	0.97	0.94

Table B.10: Relative RMSE of model forecasts during the rolling window pseudo-real-time experiments for US Imports against four benchmarks: ARMA, BNDF, DF and SW models. The labels N, Q1 and Q2 denotes the forecast horizons – Nowcast, Forecast-1Q and Forecast-2Q, respectively. The bolded values correspond to the greatest accuracy within the forecast horizon, while the asterix denotes significant performance improvement as suggested by GW test with 5% significance.

Model	**ARMA**			**BNDF**			**DF**			**SW**		
	N	**Q1**	**Q2**	**N**	**Q1**	**Q2**	**N**	**Q1**	**Q2**	**N**	**Q1**	**Q2**
AdL	0.54	0.57	0.75	0.88	0.84	0.83	0.62*	0.69	0.95	0.85	0.72	0.86
AdP	0.54	0.60	0.75	0.89	0.87	0.83	0.62*	0.72	0.95	0.85	0.76	0.85
AdRL	**0.53**	0.57	0.74	**0.87**	0.83	0.82	**0.61***	0.69	0.94	**0.84**	0.72	0.84
L	0.54	0.58	0.75	0.88	0.85	0.83	0.62*	0.70	0.95	0.85	0.74	0.85
L0	0.76	0.73	0.77	1.23*	1.07	0.85	0.86	0.88	0.97	1.18	0.92	0.87
L0L1	0.64	0.61	0.73	1.04	0.90	0.81	0.73	0.74	0.93	1.00	0.77	0.83
L0L1p	0.59	0.58	0.75	0.95	0.85	0.82	0.67	0.70	0.95	0.92	0.73	0.85
L0L2	0.69	0.61	0.78	1.13	0.89	0.86	0.79	0.74	0.98	1.08	0.77	0.88
L0L2p	0.73	0.63	0.78	1.19	0.92	0.86	0.83	0.76	0.98	1.14	0.80	0.88
L0p	0.77	0.73	0.76	1.25*	1.06	0.83	0.88	0.88	0.96	1.20	0.92	0.86
LP	0.58	0.62	0.74	0.94	0.90	0.81	0.66*	0.75	0.93	0.91	0.78	0.84
LPX	0.56*	0.60	0.75	0.90	0.87	0.82	0.64*	0.72	0.94	0.87	0.75	0.85
RL	0.54	0.58	0.75	0.88	0.85	0.83	0.62	0.71	0.95	0.84	0.74	0.85
SqL	0.55	**0.56**	0.75	0.90	**0.82**	0.82	0.64*	**0.68**	0.94	0.87	**0.71**	0.85
SqP	0.57	0.58	**0.71**	0.92	0.84	**0.79**	0.65*	0.69	**0.90**	0.89	0.73	**0.81**

Table B.11: Spain: PFCE. Relative RMSE of model forecasts during the rolling window pseudo-real-time experiments against four benchmarks: ARMA, BNDF, DF and SW models. The labels N, Q1 and Q2 denotes the forecast horizons – Nowcast, Forecast-1Q and Forecast-2Q, respectively. The bolded values correspond to the greatest accuracy within the forecast horizon, while the asterix denotes significant performance improvement as suggested by GW test with 5% significance.

Model	**ARMA**			**BNDF**			**DF**			**SW**		
	N	**Q1**	**Q2**	**N**	**Q1**	**Q2**	**N**	**Q1**	**Q2**	**N**	**Q1**	**Q2**
AdL	1.07	1.33	1.19	1.04	1.19	1.14	1.03	1.10	1.13	0.97	0.91	0.92
AdP	1.03	1.29	1.22	1.01	1.15	1.17	1.00	1.06	1.16	0.94	0.88	0.94
AdRL	1.04	1.34	1.19	1.01	1.20	1.15	1.00	1.11	1.13	0.95	0.92	0.92
L	1.02	1.26*	1.19	0.99	1.12	1.14	0.98	1.04	1.13	0.93	0.86	0.92
L0	1.16	1.52	1.16	1.13	1.36	1.12	1.12	1.26	1.10	1.06	1.04	0.89
L0L1	**0.97**	1.24	1.03	**0.95**	1.11	0.99	**0.94**	1.03	0.98	**0.89**	0.85	0.79*
L0L1p	1.03	1.29	1.10	1.01	1.15	1.06	1.00	1.07	1.04	0.94	0.88	0.85
L0L2	1.01	**1.16**	**0.96**	0.98	**1.04**	**0.92**	0.97	**0.96**	**0.91**	0.92	**0.79**	**0.74***
L0L2p	1.05	1.24	1.02	1.02	1.11	0.98	1.01	1.03	0.96	0.96	0.85	0.78*
L0p	1.12	1.50	1.16	1.09	1.35	1.12	1.08	1.24	1.10	1.02	1.03	0.90
LP	1.04	1.21	1.17	1.01	1.08	1.12	1.00	1.00	1.11	0.95	0.83	0.90
LPX	1.03	1.20*	1.17	1.00	1.08	1.12	0.99	1.00	1.11	0.94	0.82	0.90
RL	1.06	1.24*	1.18	1.04	1.11	1.13	1.02	1.02	1.12	0.97	0.84	0.91
SqL	1.00	1.23*	1.13	0.97	1.10	1.08	0.96	1.02	1.07	0.91	0.84	0.87
SqP	1.06	1.31*	1.24	1.03	1.17	1.19*	1.02	1.08	1.18	0.96	0.89	0.96

Table B.12: Spain: GFCF. Relative RMSE of model forecasts during the rolling window pseudo-real-time experiments against four benchmarks: ARMA, BNDF, DF and SW models. The labels N, Q1 and Q2 denotes the forecast horizons – Nowcast, Forecast-1Q and Forecast-2Q, respectively. The bolded values correspond to the greatest accuracy within the forecast horizon, while the asterix denotes significant performance improvement as suggested by GW test with 5% significance.

Model	**ARMA**			**BNDF**			**DF**			**SW**		
	N	**Q1**	**Q2**	**N**	**Q1**	**Q2**	**N**	**Q1**	**Q2**	**N**	**Q1**	**Q2**
AdL	0.97	0.91	0.93	1.25	0.96	0.95	1.21	1.13	1.05	1.02	0.80	0.81
AdP	1.02	0.93	0.92	1.32	0.98	0.94	1.28	1.15	1.04	1.07	0.82	0.80
AdRL	1.06	0.94	0.91	1.36	0.99	0.92	1.32	1.16	1.03	1.11	0.83	0.79
L	0.91	0.89	0.92	1.17	0.93	0.93	1.13	1.10	1.04	0.95	0.78	0.80
L0	1.05	0.91	0.94	1.36*	0.96	0.96	1.31	1.12	1.07	1.11	0.80	0.82
L0L1	0.95	0.93	0.94	1.23	0.98	0.95	1.19	1.15	1.06	1.00	0.82	0.81
L0L1p	0.91	0.91	0.92	1.17	0.95	0.93	1.13	1.12	1.03	0.95	0.79	0.79
L0L2	0.96	0.94	0.97	1.23	0.99	0.98	1.20	1.16	1.09	1.01	0.83	0.84
L0L2p	0.93	0.88	0.92	1.20	0.93	0.93	1.16	1.09	1.04	0.98	0.77	0.80
L0p	1.07	0.91	0.95	1.37*	0.96	0.96	1.33	1.13	1.07	1.12	0.80	0.82
LP	**0.90**	**0.86**	**0.89**	**1.16**	**0.91**	**0.90**	**1.13**	**1.06**	**1.00**	**0.95**	**0.76**	**0.77**
LPX	0.91	0.87	0.90	1.16	0.91	0.91	1.13	1.07	1.01	0.95	0.76	0.78
RL	0.95	0.87	0.90	1.22	0.92	0.91	1.19	1.08	1.02	1.00	0.77	0.78
SqL	0.94	0.93	0.95	1.21	0.98	0.96	1.17	1.15	1.08	0.99	0.82	0.83
SqP	0.92	0.91	0.92	1.19	0.95	0.93	1.15	1.12	1.04	0.97	0.80	0.79

Table B.13: Spain: Exports. Relative RMSE of model forecasts during the rolling window pseudo-real-time experiments against four benchmarks: ARMA, BNDF, DF and SW models. The labels N, Q1 and Q2 denotes the forecast horizons – Nowcast, Forecast-1Q and Forecast-2Q, respectively. The bolded values correspond to the greatest accuracy within the forecast horizon, while the asterix denotes significant performance improvement as suggested by GW test with 5% significance.

Model	**ARMA**			**BNDF**			**DF**			**SW**		
	N	**Q1**	**Q2**	**N**	**Q1**	**Q2**	**N**	**Q1**	**Q2**	**N**	**Q1**	**Q2**
AdL	0.73	0.78	0.90	0.84	0.93	0.98	0.93	0.89	0.91*	0.79	0.76	0.82
AdP	0.72	0.79	0.88	0.83	0.95	0.97	0.93	0.91	0.90*	0.78	0.78	0.81
AdRL	0.73	0.78	**0.88**	0.84	0.93	**0.96**	0.94	0.89	**0.89***	0.79	0.76	**0.80**
L	0.75	0.80	0.89	0.86	0.96	0.98	0.96	0.91*	0.91*	0.81	0.78	0.81
L0	0.88	0.91	1.03	1.01	1.08	1.13	1.13	1.03	1.05	0.95	0.89	0.94
L0L1	0.85	0.88	0.97	0.98	1.05	1.07	1.09	1.01	0.99	0.92	0.86	0.89
L0L1p	0.87	0.86	0.97	1.00	1.03	1.07	1.11	0.98	0.99	0.94	0.84	0.89
L0L2	0.92	0.93	1.00	1.07	1.11	1.10	1.19	1.06	1.01	1.00	0.91	0.91
L0L2p	0.92	0.86	0.99	1.06	1.03	1.08	1.18	0.98	1.00	1.00	0.84	0.90
L0p	0.88	0.91	1.04	1.01	1.09	1.14	1.12	1.04	1.05	0.95	0.89	0.95
LP	0.75	0.77	0.91	0.87	0.92	1.00	0.96	0.88	0.93	0.81	0.75	0.83
LPX	0.71	0.76	0.89	0.82	0.91	0.98	0.91	0.87*	0.91*	0.77	0.74	0.81
RL	**0.70**	**0.75**	0.89	**0.81**	**0.90**	0.98	**0.90**	**0.86***	0.90*	**0.76**	**0.74**	0.81
SqL	0.77*	0.83	0.95	0.89	1.00	1.04	0.99	0.95*	0.96	0.83	0.81	0.86
SqP	0.72	0.85	0.94	0.83	1.01	1.03*	0.92	0.97	0.95	0.78	0.83	0.86

Table B.14: Spain: Imports. Relative RMSE of model forecasts during the rolling window pseudo-real-time experiments against four benchmarks: ARMA, BNDF, DF and SW models. The labels N, Q1 and Q2 denotes the forecast horizons – Nowcast, Forecast-1Q and Forecast-2Q, respectively. The bolded values correspond to the greatest accuracy within the forecast horizon, while the asterix denotes significant performance improvement as suggested by GW test with 5% significance.

Model	**ARMA**			**BNDF**			**DF**			**SW**		
	N	**Q1**	**Q2**	**N**	**Q1**	**Q2**	**N**	**Q1**	**Q2**	**N**	**Q1**	**Q2**
AdL	0.81	0.79	0.84	1.22	0.83	0.95	1.20	0.93	0.91	1.01	0.73	0.81
AdP	**0.75**	0.81	0.84	**1.12**	0.85	0.94	**1.10**	0.95	0.91	**0.94**	0.74	0.80
AdRL	0.80	0.79	0.83	1.20	0.84	0.94	1.18	0.93	0.90*	1.00	0.73	0.80*
L	0.80	0.79	0.81	1.20	0.83	0.92	1.18	0.92	0.88*	1.00	0.72	0.78
L0	0.80	0.88	0.98	1.21	0.93	1.10	1.19	1.04	1.06	1.01	0.81	0.94
L0L1	0.76	0.85	0.88	1.15	0.90	0.99	1.13	1.00	0.96	0.96	0.78	0.85
L0L1p	0.80	0.86	0.89	1.21	0.91	1.00	1.19	1.01	0.97	1.01	0.79	0.85
L0L2	0.78	0.85	0.86	1.17	0.90	0.97	1.15	1.00	0.94	0.97	0.78	0.83
L0L2p	0.78	0.85	0.86	1.17	0.89	0.97	1.15	0.99	0.94	0.98	0.78	0.83
L0p	0.79	0.88	0.98	1.19	0.93	1.10	1.17	1.04	1.06	0.99	0.81	0.94
LP	0.81	**0.79**	**0.81**	1.21	**0.83**	**0.91***	1.19	**0.92**	**0.88***	1.01	**0.72**	**0.78***
LPX	0.79	0.79	0.82	1.19	0.84	0.92	1.17	0.93	0.89*	0.99	0.73	0.78
RL	0.80	0.79	0.81	1.21	0.83	0.91	1.19	0.92	0.88*	1.01	0.72	0.78
SqL	0.82	0.88	0.89	1.23	0.93	1.00	1.21	1.03	0.97	1.02	0.81	0.86
SqP	0.85	0.83	0.87	1.28	0.88	0.97	1.26	0.98	0.94	1.06	0.77	0.83

Table B.15: Germany: PFCE. Relative RMSE of model forecasts during the rolling window pseudo-real-time experiments against four benchmarks: ARMA, BNDF, DF and SW models. The labels N, Q1 and Q2 denotes the forecast horizons – Nowcast, Forecast-1Q and Forecast-2Q, respectively. The bolded values correspond to the greatest accuracy within the forecast horizon, while the asterix denotes significant performance improvement as suggested by GW test with 5% significance.

Model	**ARMA**			**BNDF**			**DF**			**SW**		
	N	**Q1**	**Q2**	**N**	**Q1**	**Q2**	**N**	**Q1**	**Q2**	**N**	**Q1**	**Q2**
AdL	0.86	1.03	1.04	0.89	1.09	1.06	0.86	1.06	1.04	0.75	0.99	1.01
AdP	0.92	1.03	1.05	0.95	1.09	1.06	0.91	1.07	1.04	0.80	1.00	1.02
AdRL	0.88	1.05	1.05	0.91	1.11	1.06	0.88	1.08	1.04	0.77	1.01	1.01
L	0.91	0.97	0.99	0.94	1.03	1.00	0.91	1.00	0.99	0.79*	0.94	0.96
L0	0.97	1.14	1.02	1.00	1.20	1.03	0.96	1.17	1.01	0.84	1.09	0.98
L0L1	0.84	0.95	1.08	0.87	1.00	1.09	0.84	0.98	1.07	0.74	0.91	1.04
L0L1p	0.94	1.03	1.18	0.97	1.09	1.19	0.94	1.07	1.17	0.82	0.99	1.14
L0L2	0.93	0.94	0.96	0.96	1.00	0.97*	0.92	0.97	0.96	0.81	0.91	0.93
L0L2p	0.91	**0.92**	0.95	0.94	**0.98**	0.96	0.91	**0.95**	0.95	0.79	**0.89**	0.92
L0p	0.97	1.12	1.00	1.00	1.18	1.01	0.97	1.16	0.99	0.85	1.08	0.97
LP	0.85	0.93	0.97	0.88	0.98	0.99	0.85	0.96	0.97	0.74	0.90	0.94
LPX	0.82	0.95	0.98	0.85	1.00	0.99	0.82	0.98	0.97	0.72	0.91	0.95
RL	**0.79**	0.98	1.00	**0.82**	1.03	1.02	**0.79**	1.01	1.00	**0.69**	0.94	0.97
SqL	0.90	0.95	0.96	0.94	1.00	0.97*	0.90	0.98	0.96	0.79*	0.91	0.93
SqP	0.89	0.97	**0.95**	0.92	1.02	**0.96**	0.89	1.00	**0.95**	0.78	0.93	**0.92**

Table B.16: Germany: GFCF. Relative RMSE of model forecasts during the rolling window pseudo-real-time experiments against four benchmarks: ARMA, BNDF, DF and SW models. The labels N, Q1 and Q2 denotes the forecast horizons – Nowcast, Forecast-1Q and Forecast-2Q, respectively. The bolded values correspond to the greatest accuracy within the forecast horizon, while the asterix denotes significant performance improvement as suggested by GW test with 5% significance.

Model	**ARMA**			**BNDF**			**DF**			**SW**		
	N	**Q1**	**Q2**	**N**	**Q1**	**Q2**	**N**	**Q1**	**Q2**	**N**	**Q1**	**Q2**
AdL	0.77	0.95	0.96	0.95	1.01	1.00	0.89	1.07	1.02	0.79	0.93	1.01
AdP	0.78	0.94	0.96	0.97	1.01	1.00	0.90	1.07	1.02	0.81	0.93	1.01
AdRL	0.82	0.96	0.95	1.01	1.02	0.98	0.95	1.08	1.00	0.84	0.94	0.99
L	0.89	0.91	0.95	1.10	0.97	0.99	1.03	1.03	1.01	0.92	0.90	1.00
L0	**0.68**	0.94	0.97	**0.85**	1.00	1.00	**0.79**	1.06	1.03	**0.70**	0.93*	1.02
L0L1	0.84	0.93	0.97	1.04	0.99	1.00	0.97	1.05	1.03	0.87	0.92*	1.02
L0L1p	0.81	0.94	0.97	1.00	1.00	1.00*	0.93	1.06	1.02	0.83	0.92*	1.01
L0L2	0.87	0.91	0.96	1.08	0.97	0.99	1.01	1.03	1.02	0.90	0.90*	1.00
L0L2p	0.85	**0.91**	0.95	1.06	**0.97**	0.98	0.99	**1.02**	1.00	0.88	**0.89***	0.99
L0p	0.69	0.94	0.97	0.86	1.00	1.00	0.80	1.06	1.03	0.71	0.92*	1.02
LP	0.89	0.94	0.95	1.10	1.00	0.98	1.03	1.06	1.01	0.92	0.93	1.00
LPX	0.90	0.95	**0.94**	1.12	1.01	**0.97**	1.04	1.07	**0.99**	0.93	0.93	**0.98**
RL	0.87	0.95	0.95	1.07	1.01	0.98	1.00	1.07	1.00	0.89	0.94	0.99
SqL	0.83	0.93	0.97	1.03	0.99	1.00	0.96	1.04	1.02	0.86	0.91*	1.01
SqP	0.85	0.93	0.97	1.05	1.00	1.00	0.98	1.05	1.03	0.87	0.92	1.02

Table B.17: Germany: Exports. Relative RMSE of model forecasts during the rolling window pseudo-real-time experiments against four benchmarks: ARMA, BNDF, DF and SW models. The labels N, Q1 and Q2 denotes the forecast horizons – Nowcast, Forecast-1Q and Forecast-2Q, respectively. The bolded values correspond to the greatest accuracy within the forecast horizon, while the asterix denotes significant performance improvement as suggested by GW test with 5% significance.

Model	**ARMA**			**BNDF**			**DF**			**SW**		
	N	**Q1**	**Q2**	**N**	**Q1**	**Q2**	**N**	**Q1**	**Q2**	**N**	**Q1**	**Q2**
AdL	0.63	0.74	0.97	1.33	0.97	1.08	0.91	0.90	1.04	0.66	0.79	1.07
AdP	0.64*	0.75	0.94	1.35	0.99	1.05	0.92	0.92	1.01	0.66*	0.80	1.04
AdRL	0.63	0.72	0.97	1.34	0.96	1.08	0.92	0.89	1.04	0.66	0.78	1.07
L	0.65	0.82	0.98	1.38	1.08	1.09	0.95	1.00	1.05	0.68	0.88	1.08
L0	0.80	0.75	0.95	1.69*	0.99	1.06	1.16	0.92	1.02	0.83	0.80	1.05
L0L1	0.69	**0.72**	0.89	1.46	**0.95**	0.99	1.00	**0.88**	0.95	0.72	**0.77**	0.98
L0L1p	0.69	0.74	**0.88**	1.46	0.97	**0.98**	1.00	0.90	**0.95**	0.72	0.79	**0.97**
L0L2	0.72	0.79	0.91	1.52	1.05	1.02	1.04	0.97	0.98	0.75	0.85	1.01
L0L2p	0.70	0.81	0.92	1.49	1.07	1.02	1.02	0.99	0.99	0.73	0.86	1.02
L0p	0.79	0.74	0.94	1.68	0.98	1.05	1.15	0.91	1.01	0.83	0.79	1.04
LP	0.61	0.77	0.98	1.28	1.02	1.10	0.88	0.95	1.06	0.63	0.83	1.09
LPX	0.58	0.79	1.00	1.23	1.04	1.11	0.84	0.96	1.07	0.60	0.84	1.10
RL	0.58	0.80	1.00	1.24	1.06	1.12	0.85	0.98	1.08	0.61	0.86	1.11
SqL	0.64*	0.81	0.92	1.37	1.07	1.03	0.94	0.99	0.99	0.67*	0.87	1.02
SqP	**0.56***	0.77	0.90*	**1.18**	1.02	1.01	**0.81***	0.94	0.97	**0.58***	0.82	1.00

Table B.18: Germany: Imports. Relative RMSE of model forecasts during the rolling window pseudo-real-time experiments against four benchmarks: ARMA, BNDF, DF and SW models. The labels N, Q1 and Q2 denotes the forecast horizons – Nowcast, Forecast-1Q and Forecast-2Q, respectively. The bolded values correspond to the greatest accuracy within the forecast horizon, while the asterix denotes significant performance improvement as suggested by GW test with 5% significance.

Model	**ARMA**			**BNDF**			**DF**			**SW**		
	N	**Q1**	**Q2**	**N**	**Q1**	**Q2**	**N**	**Q1**	**Q2**	**N**	**Q1**	**Q2**
AdL	0.81	0.92	1.00	1.11	0.96	1.02	1.03	1.29	1.07	0.99	0.95	1.17
AdP	0.79	0.92	0.99*	1.07	0.96	1.01	0.99	1.29	1.06	0.95	0.95	1.16
AdRL	0.83	0.93	1.00*	1.13	0.97	1.01	1.04	1.31	1.07	1.00	0.96	1.17
L	0.83	0.98	**0.99***	1.13	1.02	**1.00**	1.05	1.37	**1.05**	1.00	1.01	**1.16**
L0	0.73	**0.87**	1.09	0.99	**0.90**	1.11	0.92	**1.22**	1.17	0.88	**0.90**	1.28
L0L1	0.72	0.88	1.14	0.98	0.92	1.16	0.91	1.24	1.22	0.87	0.91	1.34
L0L1p	0.78	0.92	1.17	1.05	0.96	1.19	0.98	1.29	1.25	0.94	0.95	1.37
L0L2	**0.71**	0.90	1.11	**0.97**	0.94	1.12	**0.90**	1.26	1.18	**0.86**	0.93	1.29
L0L2p	0.75	0.94	1.17	1.03	0.98	1.19	0.95	1.32	1.25	0.91	0.97	1.37
L0p	0.73	0.87	1.08	0.99	0.91	1.10	0.92	1.22*	1.16	0.88	0.90	1.27*
LP	0.84	0.92	1.03	1.14	0.96	1.04	1.05	1.30	1.09	1.01	0.96	1.20
LPX	0.84	0.92	1.03	1.14	0.96	1.05	1.06	1.29	1.10	1.02	0.95	1.21
RL	0.84	0.95	1.03	1.14	0.99	1.04	1.05	1.33	1.10	1.01	0.98	1.20
SqL	0.78	0.97	0.99*	1.06	1.01	1.01	0.99	1.36	1.06	0.95	1.00	1.16
SqP	0.78	0.97	1.04	1.06	1.01	1.05	0.98	1.36	1.11	0.94	1.00	1.21

Table B.19: France: PFCE. Relative RMSE of model forecasts during the rolling window pseudo-real-time experiments against four benchmarks: ARMA, BNDF, DF and SW models. The labels N, Q1 and Q2 denotes the forecast horizons – Nowcast, Forecast-1Q and Forecast-2Q, respectively. The bolded values correspond to the greatest accuracy within the forecast horizon, while the asterix denotes significant performance improvement as suggested by GW test with 5% significance.

Model	**ARMA**			**BNDF**			**DF**			**SW**		
	N	**Q1**	**Q2**	**N**	**Q1**	**Q2**	**N**	**Q1**	**Q2**	**N**	**Q1**	**Q2**
AdL	0.94	1.05	1.06	0.85	0.64	0.69	1.02	1.07	1.02	0.93	1.00	1.03
AdP	0.96	1.06	1.06	0.87	0.64	0.69	1.04	1.08	1.02	0.95	1.01	1.03
AdRL	0.94	1.08	1.08	0.84	0.66	0.70	1.02	1.10	1.03	0.92	1.03	1.05
L	0.98	**0.98**	1.03	0.88	**0.59**	0.67	1.05	**1.00**	0.98	0.96	**0.93**	1.00
L0	1.13	1.11	1.12*	1.01	0.67	0.73	1.22	1.13	1.07*	1.11	1.06	1.08*
L0L1	0.93	1.05	1.10	0.84*	0.64	0.71	1.01	1.08	1.05	0.92	1.00	1.06
L0L1p	0.95	1.11	1.14*	0.85	0.67	0.74	1.03	1.13	1.09*	0.93	1.06	1.10
L0L2	**0.92**	1.06	1.07	**0.83***	0.64	0.70	**1.00**	1.08	1.03	**0.91**	1.01	1.04
L0L2p	0.96	1.06	1.14*	0.86	0.64	0.74	1.04	1.08	1.09*	0.94	1.01	1.11*
L0p	1.12	1.11	1.11*	1.01	0.67	0.72	1.21	1.13	1.07*	1.10	1.06	1.08*
LP	0.94	1.10	1.08	0.85	0.67	0.70	1.02	1.12	1.03*	0.93	1.04	1.04
LPX	0.98	1.08	1.09	0.88	0.65	0.71	1.06	1.10	1.04*	0.96	1.03	1.06
RL	1.01	1.07	1.10	0.90	0.65	0.71	1.09	1.09	1.05*	0.99	1.01	1.06
SqL	0.96	0.99	**1.01**	0.86*	0.60	**0.66**	1.04	1.01	**0.97**	0.94	0.94	**0.98**
SqP	0.97	1.06	1.09	0.87	0.64	0.71	1.05	1.08	1.04	0.95	1.01	1.05

Table B.20: France: GFCF. Relative RMSE of model forecasts during the rolling window pseudo-real-time experiments against four benchmarks: ARMA, BNDF, DF and SW models. The labels N, Q1 and Q2 denotes the forecast horizons – Nowcast, Forecast-1Q and Forecast-2Q, respectively. The bolded values correspond to the greatest accuracy within the forecast horizon, while the asterix denotes significant performance improvement as suggested by GW test with 5% significance.

Model	**ARMA**			**BNDF**			**DF**			**SW**		
	N	**Q1**	**Q2**	**N**	**Q1**	**Q2**	**N**	**Q1**	**Q2**	**N**	**Q1**	**Q2**
AdL	1.04	0.76	0.62	0.14	0.11	0.04	1.35	0.96	0.97	0.88	0.86	1.00
AdP	1.04	0.76	0.61	0.14	0.11	0.04	1.36*	0.98	0.96	0.88	0.88	0.99
AdRL	1.05	0.79	0.63	0.14	0.11	0.04	1.36	1.01	0.98	0.88	0.90	1.01
L	0.95	0.77	0.62	0.13	0.11	0.04	1.24	0.98	0.97	0.81	0.88	1.00
L0	1.01	0.81	0.71	0.13	0.11	0.04	1.32	1.04	1.11	0.86	0.93	1.14
L0L1	0.92	0.79	0.68	0.12	0.11	0.04	1.19	1.01	1.06	0.78	0.91	1.10
L0L1p	0.96	0.81	0.69	0.13	0.11	0.04	1.25*	1.03	1.08	0.81	0.93	1.11
L0L2	0.87	0.78	0.65	0.12	0.11	0.04	1.14	0.99	1.01	0.74*	0.89	1.04
L0L2p	**0.86**	0.74	0.62	**0.11**	0.10	0.04	**1.12**	0.94	0.97	**0.73**	0.84	1.00
L0p	1.01	0.81	0.71	0.13	0.12	0.04	1.32	1.04	1.11	0.86	0.93	1.14
LP	0.90	0.72	0.57	0.12	0.10	0.03	1.18	0.92	0.89	0.77	0.82	0.92
LPX	0.93	0.73	0.58	0.12	0.10	0.03	1.21	0.93	0.91	0.79	0.83	0.94
RL	0.95	0.73	**0.57**	0.13	0.10	**0.03**	1.24	0.93	**0.89**	0.81	0.83	**0.92**
SqL	0.95	0.77	0.62	0.13	0.11	0.04	1.24	0.98	0.96	0.81	0.88	0.99
SqP	0.92	**0.71**	0.59	0.12	**0.10**	0.04	1.20	**0.90***	0.91	0.78	**0.81**	0.94

Table B.21: France: Exports. Relative RMSE of model forecasts during the rolling window pseudo-real-time experiments against four benchmarks: ARMA, BNDF, DF and SW models. The labels N, Q1 and Q2 denotes the forecast horizons – Nowcast, Forecast-1Q and Forecast-2Q, respectively. The bolded values correspond to the greatest accuracy within the forecast horizon, while the asterix denotes significant performance improvement as suggested by GW test with 5% significance.

Model	**ARMA**			**BNDF**			**DF**			**SW**		
	N	**Q1**	**Q2**	**N**	**Q1**	**Q2**	**N**	**Q1**	**Q2**	**N**	**Q1**	**Q2**
AdL	0.80	0.93	0.89	0.75	1.00	0.93	0.84	0.94	0.96	0.74	0.95	1.04
AdP	0.85	0.94	0.90	0.80	1.02	0.94	0.89	0.95	0.97	0.78	0.96	1.04
AdRL	0.83	0.94	0.91	0.78	1.02	0.95	0.87	0.95	0.98	0.76	0.96	1.06*
L	0.80*	0.91	0.89	0.75	0.98	0.93	0.84	0.92	0.95	0.73	0.93	1.03
L0	1.10	1.13	1.01	1.04	1.22	1.05	1.16	1.14	1.08	1.01	1.16	1.17
L0L1	0.91	0.91	0.89	0.85	0.99	0.93	0.95	0.93	0.96	0.83	0.94	1.04
L0L1p	0.85	0.87	0.90	0.80	0.94	0.93	0.90	0.88	0.96	0.78	0.89	1.04
L0L2	1.04	1.01	0.89	0.97	1.09	0.93	1.09	1.02	0.95	0.95	1.03	1.03
L0L2p	1.02	1.02	0.91	0.96	1.10	0.95	1.08	1.03	0.98	0.94	1.04	1.06
L0p	1.11	1.11	0.99	1.04	1.20	1.03	1.16	1.12	1.06	1.01	1.14	1.15
LP	**0.73***	**0.85**	0.88	**0.69**	**0.92**	0.92	**0.77**	**0.86**	0.94	**0.67**	**0.88**	1.02
LPX	0.78	0.89	0.89	0.73	0.97	0.92	0.82	0.90	0.95	0.71	0.92	1.03
RL	0.77	0.89	0.89	0.73	0.97	0.93	0.82	0.90	0.95	0.71	0.91	1.03
SqL	0.83*	0.90	0.89	0.78	0.97	0.93	0.88	0.91	0.96	0.76	0.92	1.04
SqP	0.86	0.86	**0.88**	0.80	0.94	**0.91**	0.90	0.87	**0.94**	0.78	0.89	**1.02**

Table B.22: France: Imports. Relative RMSE of model forecasts during the rolling window pseudo-real-time experiments against four benchmarks: ARMA, BNDF, DF and SW models. The labels N, Q1 and Q2 denotes the forecast horizons – Nowcast, Forecast-1Q and Forecast-2Q, respectively. The bolded values correspond to the greatest accuracy within the forecast horizon, while the asterix denotes significant performance improvement as suggested by GW test with 5% significance.

Model	**ARMA**			**BNDF**			**DF**			**SW**		
	N	**Q1**	**Q2**	**N**	**Q1**	**Q2**	**N**	**Q1**	**Q2**	**N**	**Q1**	**Q2**
AdL	0.81*	0.74	0.80*	1.04	0.94	1.03	1.08	0.86	0.95	0.84	0.76	1.04
AdP	0.79*	0.74	0.80*	1.00	0.94	1.02	1.04	0.87	0.94	0.81	0.76	1.03
AdRL	0.85	0.76	0.81*	1.08	0.96	1.03	1.13	0.88	0.95	0.88	0.78	1.04
L	0.78*	0.79	0.80*	0.99	1.00	1.02	1.03	0.92	0.94	0.80	0.81	1.03
L0	0.86	0.99	0.97	1.10	1.25	1.25	1.14	1.15	1.15	0.89	1.02	1.26
L0L1	0.81*	0.98	0.96*	1.03	1.23	1.23	1.07	1.14	1.13	0.83	1.00	1.24
L0L1p	0.81*	0.96	0.94*	1.04	1.21	1.21	1.08	1.12	1.11	0.84	0.98	1.21
L0L2	0.78*	1.04	0.99	1.00	1.31	1.27	1.04	1.21	1.17	0.81	1.06	1.28
L0L2p	0.77*	0.98	0.97	0.99	1.23	1.24	1.03	1.14	1.14	0.80	1.00	1.25
L0p	0.87	0.98	0.97	1.11	1.24	1.24	1.15	1.15	1.14	0.89	1.01	1.25
LP	0.80	0.75	0.78*	1.02	0.95	1.00	1.06	0.87	0.92	0.82	0.77	1.01
LPX	0.76*	0.74	**0.77***	0.98	0.93	**0.99**	1.01	0.86	**0.91**	0.79	0.76	**1.00**
RL	0.81	0.75	0.78*	1.03	0.94	1.00	1.07	0.87	0.91	0.83	0.76	1.00
SqL	**0.74***	0.77	0.82*	**0.94**	0.98	1.05	**0.98**	0.90	0.96	**0.76**	0.79	1.06
SqP	0.83	**0.70**	0.78*	1.06	**0.89**	1.00	1.10	**0.82**	0.92	0.85	**0.72**	1.01

Table B.23: Italy: PFCE. Relative RMSE of model forecasts during the rolling window pseudo-real-time experiments against four benchmarks: ARMA, BNDF, DF and SW models. The labels N, Q1 and Q2 denotes the forecast horizons – Nowcast, Forecast-1Q and Forecast-2Q, respectively. The bolded values correspond to the greatest accuracy within the forecast horizon, while the asterix denotes significant performance improvement as suggested by GW test with 5% significance.

Model	**ARMA**			**BNDF**			**DF**			**SW**		
	N	**Q1**	**Q2**	**N**	**Q1**	**Q2**	**N**	**Q1**	**Q2**	**N**	**Q1**	**Q2**
AdL	1.08	1.04	0.84	0.96	0.93	0.85	0.85	0.92	0.79	0.75*	0.80*	0.71*
AdP	1.12	1.06	0.86	1.00	0.94	0.87	0.89	0.93	0.81	0.78*	0.81*	0.73*
AdRL	1.15	1.09	0.86	1.02	0.97	0.86	0.91	0.96	0.81	0.80*	0.83*	0.72*
L	1.02	0.89	0.81	0.90	0.79	0.82	0.81*	0.78	0.76*	0.71*	0.68*	0.68*
L0	1.15	0.99	0.84	1.02	0.88	0.85	0.91	0.87	0.79	0.81	0.75*	0.71*
L0L1	1.08	0.90	0.83	0.96	0.81	0.83	0.86	0.80	0.78	0.76*	0.69*	0.70*
L0L1p	1.13	0.92	**0.81**	1.00	0.82	**0.81**	0.89	0.81	**0.76***	0.79	0.71*	**0.68***
L0L2	1.40*	1.18	1.09	1.25*	1.05	1.09	1.11	1.04	1.02	0.98	0.90	0.91
L0L2p	1.07	0.89	0.85	0.95	0.79	0.86	0.85	0.78	0.80*	0.75*	0.68*	0.72*
L0p	1.11	0.96	0.82	0.99	0.86	0.82	0.88	0.85	0.77*	0.78*	0.74*	0.69*
LP	1.01	0.93	0.90	0.90	0.83	0.90	0.80	0.82	0.84	0.71*	0.71*	0.75*
LPX	1.02	0.94	0.90	0.91	0.84	0.91	0.81	0.83	0.85	0.71*	0.72*	0.76*
RL	1.01	0.95	0.87	0.90	0.85	0.88	0.80	0.84	0.82	0.71*	0.73*	0.73*
SqL	**0.99**	**0.87**	0.84	**0.88**	**0.77**	0.84	**0.78***	**0.76**	0.79*	**0.69***	**0.66***	0.70*
SqP	1.00	0.97	0.86	0.89	0.86	0.87	0.79*	0.85	0.81*	0.70*	0.74*	0.73*

Table B.24: Italy: GFCF. Relative RMSE of model forecasts during the rolling window pseudo-real-time experiments against four benchmarks: ARMA, BNDF, DF and SW models. The labels N, Q1 and Q2 denotes the forecast horizons – Nowcast, Forecast-1Q and Forecast-2Q, respectively. The bolded values correspond to the greatest accuracy within the forecast horizon, while the asterix denotes significant performance improvement as suggested by GW test with 5% significance.

Model	**ARMA**			**BNDF**			**DF**			**SW**		
	N	**Q1**	**Q2**	**N**	**Q1**	**Q2**	**N**	**Q1**	**Q2**	**N**	**Q1**	**Q2**
AdL	0.92	0.99	0.86	1.11	1.06	0.88	0.85	0.88	0.87	0.81	0.88	0.85*
AdP	**0.91**	0.99	0.85*	**1.10**	1.07	0.87*	**0.84**	0.89	0.86	**0.80**	0.88	0.84*
AdRL	0.93	1.01	0.85*	1.11	1.09	0.87	0.86	0.91	0.86	0.81	0.90	0.84*
L	0.97	0.95	0.88	1.17	1.02	0.90*	0.90	0.85	0.89	0.85	0.85*	0.87*
L0	0.93	0.92	0.94	1.12	0.99	0.96	0.86	0.82	0.95	0.82*	0.82*	0.93*
L0L1	0.95	0.91	0.87	1.14	0.98	0.89*	0.88	0.82*	0.88*	0.83	0.81*	0.86*
L0L1p	0.93	**0.91**	0.86	1.12	**0.97**	0.88	0.86	**0.81**	0.87	0.82	**0.81***	0.85*
L0L2	1.07	0.98	0.92	1.29*	1.05	0.94	0.99	0.88	0.94	0.94	0.87*	0.91
L0L2p	1.14*	1.02	0.90	1.37*	1.10	0.92	1.06	0.92	0.92	1.01	0.91	0.89*
L0p	0.92	0.92	0.94	1.11	0.98	0.97	0.86	0.82	0.96	0.81*	0.82*	0.93
LP	0.97	0.94	0.85*	1.17	1.01	0.87*	0.90	0.84	0.86*	0.85	0.83*	0.84*
LPX	0.95	0.92	0.84*	1.15	0.98	0.86*	0.88	0.82	0.85*	0.84	0.81*	0.83*
RL	0.95	0.94	**0.83***	1.14	1.01	**0.85***	0.88	0.84	**0.84***	0.84	0.83*	**0.82***
SqL	0.95	0.95	0.89	1.14	1.02	0.91*	0.88	0.85	0.90	0.83	0.84*	0.88*
SqP	0.96	0.93	0.84*	1.15	1.00	0.86*	0.88	0.84	0.85*	0.84	0.83*	0.83*

Table B.25: Italy: Exports. Relative RMSE of model forecasts during the rolling window pseudo-real-time experiments against four benchmarks: ARMA, BNDF, DF and SW models. The labels N, Q1 and Q2 denotes the forecast horizons – Nowcast, Forecast-1Q and Forecast-2Q, respectively. The bolded values correspond to the greatest accuracy within the forecast horizon, while the asterix denotes significant performance improvement as suggested by GW test with 5% significance.

Model	**ARMA**			**BNDF**			**DF**			**SW**		
	N	**Q1**	**Q2**	**N**	**Q1**	**Q2**	**N**	**Q1**	**Q2**	**N**	**Q1**	**Q2**
AdL	0.91	0.94	1.05	1.13	0.99	1.02	0.80*	0.94	1.07	0.80	1.01	1.14*
AdP	0.91	0.93	**1.02**	1.14	0.97	**1.00**	0.80*	0.92	**1.05**	0.81	1.00	**1.12**
AdRL	**0.88**	0.92	1.05	**1.09**	0.96	1.02	**0.77***	0.91	1.07	**0.78**	0.99	1.15*
L	0.88	0.95	1.05	1.10	0.99	1.02	0.77*	0.94	1.08	0.78	1.02	1.15
L0	1.09	1.02	1.13	1.36*	1.07	1.10	0.95	1.02	1.15	0.96	1.10	1.23
L0L1	0.92	0.92	1.03	1.15	0.96	1.00	0.81	0.91	1.06	0.82	0.99	1.13
L0L1p	0.97	0.91	1.04	1.21	0.95	1.01	0.85	0.90	1.06	0.86	0.98	1.13
L0L2	0.95	**0.88**	1.04	1.18	**0.93**	1.01	0.83*	**0.88**	1.06	0.84	**0.95**	1.13
L0L2p	0.94	0.90	1.07	1.17	0.94	1.04	0.82*	0.90	1.09	0.83	0.97	1.17
L0p	1.09	1.02	1.12	1.36*	1.07	1.09	0.95	1.02	1.15	0.96	1.10	1.23
LP	0.90	0.95	1.05	1.12	0.99	1.03	0.79*	0.94	1.08*	0.80	1.02	1.15*
LPX	0.93	0.96	1.06	1.16	1.01	1.04	0.81*	0.96	1.09	0.82	1.04	1.16*
RL	0.93	0.98	1.07	1.16	1.03	1.04	0.82*	0.97	1.09	0.82	1.06	1.17*
SqL	0.90*	0.94*	1.02	1.13	0.98	1.00	0.79*	0.93	1.05	0.80	1.01	1.12
SqP	0.93	0.94	1.04	1.15	0.99	1.01	0.81*	0.93	1.06*	0.82	1.01	1.13*

Table B.26: Italy: Imports. Relative RMSE of model forecasts during the rolling window pseudo-real-time experiments against four benchmarks: ARMA, BNDF, DF and SW models. The labels N, Q1 and Q2 denotes the forecast horizons – Nowcast, Forecast-1Q and Forecast-2Q, respectively. The bolded values correspond to the greatest accuracy within the forecast horizon, while the asterix denotes significant performance improvement as suggested by GW test with 5% significance.

Model	**ARMA**			**BNDF**			**DF**			**SW**		
	N	**Q1**	**Q2**	**N**	**Q1**	**Q2**	**N**	**Q1**	**Q2**	**N**	**Q1**	**Q2**
AdL	0.85	0.74	0.80	1.04	0.83	0.81	0.91	0.92	1.02	0.86	1.01	1.06
AdP	0.81	0.74	0.79	0.99	0.83	0.80	0.87	0.92	1.02	0.82	1.01	1.05
AdRL	0.88	0.76*	0.78	1.06	0.85	0.79	0.94	0.94	1.01	0.88	1.04	1.04
L	0.81	0.74	0.82	0.99	0.83	0.82	0.87	0.92	1.05	0.82	1.01	1.08*
L0	0.99	0.91	0.97	1.20*	1.02*	0.98	1.06	1.13*	1.25*	1.00	1.24*	1.29*
L0L1	0.86	0.75*	0.89	1.04	0.84	0.90	0.92	0.93	1.14*	0.87	1.03	1.18*
L0L1p	0.91	0.82	0.99	1.10*	0.92	1.00	0.97	1.02	1.27	0.92	1.13	1.31
L0L2	0.89	0.73	0.80	1.08	0.82	0.81	0.96	0.91	1.02	0.90	1.00	1.06
L0L2p	0.88	0.76	0.84	1.07	0.85	0.85	0.94	0.94	1.09	0.89	1.04	1.12
L0p	0.97	0.91	0.97	1.18*	1.01*	0.98	1.04	1.12*	1.25*	0.98	1.24*	1.28*
LP	0.79	0.80	0.85	0.96	0.90	0.86	0.84	0.99	1.09	0.80	1.10	1.12
LPX	0.82	0.79	0.84	0.99	0.89	0.84	0.87	0.98	1.07	0.82	1.08	1.11
RL	0.81	0.77	0.81	0.99	0.86	0.82	0.87	0.95	1.04	0.82	1.05	1.07
SqL	0.81	0.74*	0.81	0.98	0.83	0.82	0.86	0.92	1.04	0.81	1.02	1.08
SqP	**0.73***	**0.70***	**0.77***	**0.88**	**0.78**	**0.78**	**0.78***	**0.87**	**0.99**	**0.74**	**0.96**	**1.02**

Santrauka

Mokslinė problema ir jos aktualumas

Dėl nuolat didėjančios duomenų apimties bei prieinamumo pastaruoju metu stebima stipri statistinių bei mašininio mokymosi metodų pažanga, gebančių tokio tipo duomenis modeliuoti. Įprastas tyrėjų tikslas – į atliekamas analizes įtraukti kuo daugiau prieinamos informacijos. Tačiau aiškinamųjų kintamųjų skaičiui augant sparčiau nei stebėjimų skaičiui, klasikiniai statistiniai metodai, pvz., mažiausiųjų kvadratų, didžiausiojo tikėtinumo ar Bajeso metodai su neinformatyviais aprioriniais skirstiniais susiduria su parametrų vertinimo problemomis (angl. *curse of dimensionality*, žr., pvz., Giannone ir kt. (2021)). Šiuo tikslu išvystyta aibė specifinių metodų, gebančių efektyviai panaudoti didelio matavimo duomenų masyvus. Literatūroje galima išskirti dvi tokių metodų kryptis, kurios skiriasi prielaidomis dėl modelių parametrų struktūrų: formuluojama, jog modelių parametrai yra *retos* arba *tankios* struktūros (Ng (2013), Chernozhukov ir kt. (2017)).

Tyrėjas, darydamas *tankios* modelio parametrų struktūros prielaidą siekia analizėje efektyviai panaudoti visą prieinamą didelio matavimo informaciją, net jeigu kiekvieno individualaus tyrime naudojamo kintamojo svoris yra nykstamai mažas. Galimų metodų specifika skiriasi priklausomai nuo taikymų srities, tačiau tarp dažniausiai sutinkamų literatūroje galima išskirti faktorių modelius, *Ridge* regresiją, svertinį mažiausiųjų kvadratų metodą (WALS) ir kt. (Tikhonov (1963), Hoerl ir Kennard (1970), Magnus ir De Luca (2016), Andreini ir kt. (2020)). Minėtieji metodai dažnai naudojami taikymuose (žr., pvz., Stock ir Watson (2002), Diebold (2003), De Mol ir kt. (2006) and Stock ir Watson (2009)).

Kitavertus, darant *retos* struktūros prielaidą yra tariama, jog tik nedidelis turimų kintamųjų poaibis yra esminis tiriamajam uždaviniui spręsti. Svarbiausias uždavinys yra tokių signalų atpažinimas ir įvertinimas. Atpažinus pagrindinius aiškinančiuosius kintamuosius likusi informacija iš analizės yra pašalinama užnulinant atitinkamus modelio parametrus. Vienas populiariausių retos struktūros metodų – *Least Absolute Shrinkage Selection Operator* (LASSO, Tibshirani (1996)), kurio esminis privalumas yra gebėjimas efektyviai vienu metu atlikti tiek reikšmingų kintamųjų atranką, tiek ir modelio parametrų įvertinimą. Dėl efektyvumo bei paprastumo LASSO dažnai sutinkamas literatūroje sprendžiant aibę įvairių uždavinių, yra itin mėgstamas tyrėjų bei akademikų. Papildomai, išvystyta daugybė metodo plėtinių bei modifikacijų, sprendžiančių bei koreguojančių įvairius trūkumus (žr., Adaptive LASSO (Zou (2006)), Relaxed LASSO (Meinshausen (2007)), Square-Root LASSO (Belloni ir kt. (2011))). Svarbiausi retos struktūros metodai išsamiai apžvelgiami 1 disertacijos skyriuje.

Retumo prielaida gali būti pagrįsta tais atvejais, kai tyrėjas tikisi, jog (santykinai, pagal imties dydį) nedidelio kintamųjų rinkinio gali pakakti siekiant tiksliai nusakyti tam tikrą stebimą ar modeliuojamą procesą. Pavyzdžiui, Bai ir Ng (2008) pastebi, jog per dideli aukšto dažnio duomenų rinkiniai kai kuriais atvejais gali neigiamai veikti prognozes. Kartu, Bulligan ir kt. (2015) teigia, jog retos struktūros prielaida tam tikrais atvejais gerokai pagerina trumpalaikių prognozių tikslumą. Kitavertus, tikslaus retumo prielaida (angl. *exact sparsity*), tikėtina, praktikoje yra per griežta – dažnai pakanka tarti, jog visi kintamieji turi nenulinį poveikį stebimam ar modeliuojamam procesui, tačiau statistiškai įmanoma atrinkti nedidelį kintamųjų poaibį, kurio pakaktų užtikrintai aproksimuoti modeliuojamą signalą. Tokia prielaida literatūroje apibrėžiama kaip apytikslio retumo (angl. *approximate sparsity*), plačiau apžvelgiama disertacijos 1.2 poskyryje.

Dėl paskutiniu metu sparčiai augančio retų struktūrų metodų populiarumo literatūroje sutinkama vis daugiau svarbių taikymų. Pavyzdžiui, Shibata (1980), Ing (2007) nagrinėja modelių atrankos problematiką dirbant su autoregresinių laiko eilučių modeliais; taip pat, Belloni ir Chernozhukov (2011), Javanmard ir Montanari (2014), Zhang ir Zhang (2014), Caner ir Kock (2018), Belloni ir kt. (2018), Gold ir kt. (2020), Ning ir kt. (2020), Guo ir kt. (2021) pristato rezultatus taikant daugiamates tiesines bei instrumentinių kintamųjų regresijas statistinių hipotezių testavimui. Svarbu pastebėti, jog minėtuose taikymuose apytikslio retumo prielaida yra kertinė.

Literatūroje, lyginančioje retų bei tankių struktūrų metodų taikymus, neretai sutinkama rezultatų, rodančių, jog negalime bet kuriam sprendžiamam modeliavimo uždaviniui rasti vieno universaliai geriausio metodo. Akcentuojama, jog joks metodas nėra tobulas, o gaunamų rezultatų tikslumas stipriai priklauso nuo nestebimo duomenis generuojančio proceso struktūros (žr., pvz., Giannone ir kt. (2021) apžvalgą). Šie ir kiti panašūs pastebėjimai atvėrė kelią naujai tyrimų krypčiai, apjungiančiai populiarius retų bei tankių struktūrų metodus, sprendžiančiai kiekvieno iš jų trūkumus bei apribojimus (žr., *Targeted Diffusion Index* modelius (Bulligan ir kt. (2015)), LAVA (Chernozhukov ir kt. (2017)), *Spike-and-Slab priors* (Giannone ir kt. (2021)), *Fast Best Subset Selection* (Hazimeh ir Mazumder (2020))). Šios idėjos plačiau apžvelgiamos disertacijos 3 skyriuje, kuriame lyginami populiarūs retų bei tankių struktūrų metodai atliekant pseudo-realaus laiko eksperimentus vertinant išankstines JAV bei ES BVP išlaidų komponenčių prognozes.

Tikėtina, jog minėtieji rezultatai kiekvieno tyrimo atveju priklauso nuo to, kiek tinkama tyrėjų daroma prielaida dėl nestebimo duomenis generuojančio proceso struktūros. Natūralu, jog suformavus klaidingą retos struktūros prielaidą tankiam procesui tyrimo rezultatai gali netenkinti tyrėjo poreikių. Tačiau nėra lengva atsakyti, kaip ir kada tyrėjas galėtų spręsti, kurią prielaidą konkrečiu atveju reikėtų taikyti. Giannone ir kt. (2021) siūlo retumą suprasti kaip neapibrėžtumą ir tyrimo metu nedaryti jokių esminių prielaidų dėl parametrų struktūros. Kitaip tariant, reti metodai turėtų būti taikomi tik turint stiprių įrodymų, jog retumo prielaida gali būti išpildyta. Kitais atvejais tyrėjas turėtų leisti metodui išmokti bei įvertinti, ar minėtoji prielaida išpildoma remiantis turimais duomenimis. Tuo tikslu, Giannone ir kt. (2021) pristato universalius metodus, gebančius prisitaikyti prie duomenų struktūros ir atliekančius parametrų vertinimą nepriklausomai nuo tyrėjo daromų prielaidų dėl modelio parametrų struktūros (žr., pvz., Chernozhukov ir kt. (2017), Cevid ir kt. (2020), Giannone ir kt. (2021)).

Universalių metodų lankstumas neretai reikalauja papildomų skaičiavimo resursų bei neigiamai atsispindi modelių prognozių dinamikoje. Šiuo atžvilgiu, ideali alternatyva būtų leisti statistiniam testui įvertinti, ar turimi duomenys suteikia pagrindą tyrėjui formuluoti retos struktūros prielaidą. Tačiau literatūros, nagrinėjančios retumo testavimą daugiamačių tiesinių modelių kontekste, yra palyginti mažai. Svarbūs rezultatai gauti Dicker (2014) bei Dicker (2016) sprendžiant signalo bei triukšmo santykio (angl. *signal-to-noise ratio*, (SNR)) vertinimo problemą. Papildomai, minėtasis uždavinys prie tam tikrų papildomų prielaidų yra glaudžiai susijęs su retumo testavimo uždaviniu, ką pastebi Carpentier ir Verzelen (2019) bei Carpentier ir Verzelen (2021), pasiūlydami tikslaus retumo testus daugiamačiams tiesiniams modeliams. Minėtos idėjos disertacijoje plačiau atskleidžiamos 1.4 poskyryje, kurias išplečia bei papildo 2 skyriuje pristatyti pagrindiniai disertacijos rezultatai.

Tyrimo objektas

Disertacijoje nagrinėjama statistika $\|\mathbb{X}'Y\|_2^2$ bei su ja susiję skirstiniai. Taip pat, tyrinėjami retos struktūros tiesinės regresijos modeliai makroekonominių laiko eilučių kontekste. Vertinant modelių prognozavimo tikslumą, nagrinėjamos išankstinės bei trumpalaikės kintamųjų prognozės.

Tikslas ir uždaviniai

Retų struktūrų prielaida yra labai svarbi tam tikrais taikymų atvejais daugiamačių tiesinių modelių kontekste. Šios prielaidos pagrįstumas disertacijoje nagrinėjamas iš teorinės bei iš empirinės pusių. Tuo tikslu, suformuluojami du pagrindiniai disertacijos tikslai.

Pirmasis tikslas, analizuojant retumo prielaidą iš teorinės pusės, papildyti bei išplėsti svarbius SNR bei retumo testavimo literatūroje žinomus rezultatus. Antrasis, apžvelgti retumo prielaidą iš empirinės pusės, lyginant populiarius retos bei tankios struktūrų metodus bei jų prognozavimo tikslumą.

Šiems tikslams pasiekti, pirma, disertacijoje suformuluojami tikslus bei asimptotinis statistikos $\|X'Y\|_2^2$ skirstiniai nagrinėjant konkrečią Kac–Murdock–Szegő (KMS) tipo duomenų kovariacinę struktūrą. Šis uždavinys bei jo sprendimas gali būti svarbus tolesniems tyrimams bei rezultatams SNR bei retumo testavimo literatūroje. Pastebime, jog tiriamoji statistika $\|X'Y\|_2^2$ yra kertinė tiek Dicker (2014), tiek Carpentier ir Verzelen (2019) darbų rezultatuose.

Antra, atliekami pseudo-realaus laiko eksperimentai, vertinant išankstines bei trumpalaikes JAV bei didžiųjų Europos ekonomikų BVP išlaidų komponenčių prognozes. Tikslus išankstinis makroekonominių kintamųjų prognozavimas yra praktikoje svarbus uždavinys, puikiai tinkamas lyginti daugiamačius tiesinės regresijos modelius.

Trečia, pasiūloma bei apžvelgiama LASSO metodo modifikacija, kuria LASSO apjungiamas su pagrindinių komponenčių metodu (LASSO-PC), kombinuojant retų bei tankių struktūrų metodus. Potenciali pasiūlyto metodo nauda analizuojama tiriant prognozių tikslumą bei lyginant gautus rezultatus su kitų eksperimentuose nagrinėjamų metodų rezultatais.

Tyrimų metodika

Sudarant statistikos pasiskirstymo funkcijas naudojamas *variance-gamma* skirstinys bei jo savybės. Įrodymuose taikomi žinomi centrinės ribinės teoremos rezultatai, kumuliančių metodas bei kiti bendros tikimybių teorijos ir matematinės statistikos metodai. Empirinėje dalyje dalis teorinių faktų tikrinama taikant Monte Carlo simuliacijas. Sudarant pseudo-realiuosius eksperimentus taikomi įvairūs laiko eilučių analizės metodai.

Naujumas

Disertacijoje pasiūlomas LASSO bei pagrindinių komponenčių metodo derinys (LASSO-PC), kurio efektyvumas bei papildoma nauda įvertinama atlikus pseudo-realaus laiko eksperimentus, vertinant išankstines JAV bei kai kurių Europos šalių BVP komponenčių prognozes. Gauti rezultatai rodo, jog pasiūlytas metodas kai kuriais atvejais geba pagerinti prognozių tikslumą, lyginant su populiariais retų bei tankių struktūrų metodais.

Sudarytų eksperimentų rezultatai gali būti naudingi realiuose taikymuose kaip išsami metodų apžvalga išankstinio vertinimo uždaviniuose.

Disertacijoje išvedami tikslus bei asimptotinis statistikos $\|\mathbb{X}'Y\|_2^2$ skirstiniai. Gauti asimptotiniai rezultatai reikalauja tik $\|\beta\|_2 < \infty$, kuomet panašūs rezultatai literatūroje reikalauja tikslaus modelio koeficientų retumo arba priklauso nuo tam tikrų aproksimacijų tikslumo.

Disertacijoje išskiriama *variance-gamma* skirstinio panaudojimo svarba gautų rezultatų išvedimuose. Pirma, disertacijos rezultatai papildo žinomas *variance-gamma* skirstinio savybes. Antra, pasirinktas skirstinys yra patogus dirbant su normaliai pasiskirsčiusių atsitiktinių dydžių sandaugomis, todėl dalis gautų rezultatų gali būti pritaikyti formuluojant asimptotinius panašių struktūrų statistikų skirstinius, pvz., keičiant ℓ_2-normą kita.

Darbo struktūra

Disertacija sudaryta iš įvado, apžvalginio skyriaus, glaustai pristatančio esmines darbe naudojamas sąvokas bei metodologijas, jų svarbą ir žinomus trūkumus, bei dviejų skyrių, kuriuose pateikiami pagrindiniai disertacijos rezultatai. Taip pat, pateikiamos išvados bei literatūros sąrašas. Papildomai, dalis rezultatų pateikiami priede, kuris yra išskirtas į dvi dalis. Priedo A dalyje pateikiami ilgesni techniniai įrodymai ir tarpiniai rezultatai, tuo tarpu priedo B dalyje pateikiamos techninės pseudorealaus laiko eksperimentų detalės bei papildoma medžiaga. Disertacija parašyta anglų kalba.

Darbo apžvalga ir svarbiausi rezultatai

Darbo **įvade** trumpai supažindinama su disertacijoje nagrinėjama problematika, įvedama retumo sąvoka ir pristatomi pagrindiniai uždaviniai bei svarbiausi literatūroje stebimi rezultatai, galiojant retumo prielaidai. Retų struktūrų prielaida išskiriama kaip alternatyva tankioms struktūroms bei akcentuojama, jog tyrėjui neretai *a priori* nėra žinoma, kuria prielaida reikėtų remtis tyrimo metu, siekiant optimalių tyrimo rezultatų. Vienas iš galimų sprendimo būdų – analizėje naudoti metodus, gebančius prisitaikyti prie modeliuojamo uždavinio nepriklausomai nuo modelio parametrų struktūros, arba slypinčią struktūrą įvertinti remiantis analizuojamais duomenimis. Kitas galimas variantas – taikyti statistinį testą. Literatūros apie statistinį struktūrų testavimą nėra daug, perspektyvūs rezultatai glaudžiai siejasi su signalo bei triukšmo santykio vertinimo uždaviniais. Teorinė disertacijos rezultatų nauda siejasi su tam tikrų literatūroje sutinkamų rezultatų išplėtimu bei papildymu. Praktiniai poreikiai nagrinėjami atliekant populiarių tankių bei retų struktūrų metodų palyginimą vertinant išankstines bei trumpalaikes makroekonominių rodiklių prognozes.

Pirmame, **retumo daugiamatėje tiesinėje regresijoje**, skyriuje pateikiamos pagrindinės disertacijoje bei literatūroje sutinkamos sąvokos ir naudojamos prielaidos. Apžvalgoje įvedami svarbūs disertacijoje naudojami žymėjimai, taip pat trumpai aptariami skirtingi apytikslio retumo apibrėžimai.

Disertacijoje nagrinėjamas daugiamatis tiesinis regresijos modelis

$$Y = \mathbb{X}\beta + \varepsilon, \tag{1}$$

čia $Y = (y_1, \ldots, y_n)' \in \mathbb{R}^{n\times 1}$ yra vektorius sudarytas iš n priklausomo kintamojo stebėjimų, $\mathbb{X} = (X_1, \ldots, X_n)'$ yra $n \times p$ matrica sudaryta iš p-mačių aiškinančiųjų kintamųjų, $\beta = (\beta_1, \ldots, \beta_p)' \in \mathbb{R}^{p\times 1}$ yra nežinomų modelio parametrų vektorius, bei $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)' \in \mathbb{R}^{n\times 1}$ yra modelio liekanų vektorius.

Pažymima, jog šioje disertacijoje pagrindinis dėmesys yra skiriamas apytikslio retumo sąvokai, pristatytai Belloni ir Chernozhukov (2011). T.y., tariama, jog tikrieji modelio parametrai β tenkina:

$$|\beta_j| = Aj^{-\alpha}, \quad j = 1, \ldots, p, \quad \alpha \geq 1, \tag{2}$$

čia parametrų išrikiavimas bei narys $A > 0$ nėra esminiai. Įprastai reikalaujama, jog $\alpha \geq 1$, tačiau priklausomai nuo kintamųjų kovariacijų matricos Σ struktūros, šį reikalavimą galima atlaisvinti.

Retos struktūros metodų apžvalgoje pristatomi šie modeliai:

(i) LASSO (Tibshirani (1996)):

$$\hat{\beta}_{\text{LASSO}} = \arg\min_{\beta} \, (Y - \mathbb{X}\beta)'(Y - \mathbb{X}\beta) + \lambda ||\beta||_1, \tag{3}$$

čia hiperparametras $\lambda \in (0, \infty)$ yra fiksuotas.

(ii) Adaptive LASSO (Zou (2006)):

$$\hat{\beta}_{\text{adaLASSO}} = \arg\min_{\beta}(Y - \mathbb{X}\beta)'(Y - \mathbb{X}\beta) + \lambda w'|\beta|, \tag{4}$$

čia laikoma, jog $|\beta| := (|\beta_1|, \ldots, |\beta_p|)'$ bei $w = (w_1, \ldots, w_p)'$ žymi specifiškai parinktą svorių vektorių. Reikalavimai svorių parinkimui aptariami disertacijoje.

(iii) Relaxed LASSO (Meinshausen (2007)):

$$\begin{aligned} \hat{\beta}_{\mathrm{RL}} &= \arg\min_{\beta} n^{-1}\big(Y - \mathbb{X}\{\beta \cdot \mathbb{1}_{\mathcal{M}_\lambda}\}\big)'\big(Y - \mathbb{X}\{\beta \cdot \mathbb{1}_{\mathcal{M}_\lambda}\}\big) \\ &+ \ \phi\lambda||\beta||_1, \end{aligned} \tag{5}$$

čia $\lambda \in [0, \infty)$ bei $\phi \in (0, 1]$ yra hiperparametrai, $\mathbb{1}_{\mathcal{M}_\lambda}$ žymi indikatoriaus funkciją, įgyjančią vienetą tik tiems kintamiesiems, kuriuos LASSO metodas atrenka kaip reikšmingus prie fiksuotos λ reikšmės.

(iv) Square-Root LASSO (Belloni ir kt. (2011)):

$$\hat{\beta}_{\mathrm{sqrtLASSO}} = \arg\min_{\beta} n^{-1/2}\left((Y - \mathbb{X}\beta)'(Y - \mathbb{X}\beta)\right)^{1/2} + \lambda||\beta||_1, \tag{6}$$

čia $\lambda \in (0, \infty)$. Svarbu, jog literatūroje yra žinoma optimali parametro reikšmė: $\lambda = \sqrt{2\log(pn)/n}$.

(v) Fast Best Subset Selection (Mazumder ir kt. (2022)):

$$\hat{\beta}_{\mathrm{L0Lq}} = \arg\min_{\beta} \frac{1}{2}(Y - \mathbb{X}\beta)'(Y - \mathbb{X}\beta) + \lambda_0||\beta||_0 + \lambda_q||\beta||_q^q, \tag{7}$$

čia $q \in \{1, 2\}$ nurodo pasirinktos baudos tipą, hiperparametras $\lambda_0 \in [0, \infty)$ kontroliuoja atrinktų reikšmingų kintamųjų skaičių modelyje, tuo tarpu $\lambda_q \in [0, \infty)$ kontroliuoja modelio parametrų reikšmių sumažinimą (angl. *shrinkage*).

Pristatytų metodų privalumai, trūkumai bei svarbios savybės išsamiai apžvelgiamos disertacijoje.

Skyrius užbaigiamas glausta SNR literatūros apžvalga, jos sąryšiu su retos struktūros tiesiniais modeliais bei aktualiais retumo testavimo rezultatais. Išskiriamas vienas svarbus SNR literatūros rezultatas kaip esminė motyvacija 2 disertacijos skyriuje pristatytiems rezultatams.

Antrame, **asimptotinio normalumo tiesinėje regresijoje su apytiksliai reta struktūra**, skyriuje pristatomi pagrindiniai disertacijos rezultatai. Tiriamas (1) tiesinis modelis, tačiau papildomai tariama, jog $X_i = (X_{1,i}, \ldots, X_{p,i})'$ turi normalųjį pasiskirstymą su nuliniu vidurkiu bei kovariacine matrica Σ, t.y., $X_i \stackrel{d}{=} \mathcal{N}_p(0, \Sigma)$. Taip pat, daroma prielaida, jog Σ turi šią formą:

$$\Sigma = (\varrho^{|i-j|})_{i,j=1}^{p} = \begin{bmatrix} 1 & \varrho & \ldots & \varrho^{p-1} \\ \varrho & 1 & \ldots & \varrho^{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ \varrho^{p-1} & \varrho^{p-2} & \ldots & 1 \end{bmatrix}, \tag{8}$$

čia $0 < |\varrho| < 1$, bei $\Sigma = I_p$ jeigu $\varrho = 0$ (čia ir kitur I_p žymi $p \times p$ vienetinę matricą). Ši kovariacinė matrica vadinama Kac–Murdock–Szegő (KMS) matrica, pasiūlyta Kac ir kt. (1953). Kaip AR(1) proceso autokoreliacijų matrica, ji yra teigiamai apibrėžta ir disertacijoje pasirinkta dėl itin plačios galimų taikymų srities literatūroje, bei dėl žinomų spektrinių matricos savybių (Fikioris (2018)). Tinkamai parinkta ši struktūra gali ganėtinai tiksliai aproksimuoti plačią šeimą kovariacijos struktūrų (žr., pvz., Yang ir kt. (2021)). Taip pat, šiame skyriuje tariama, jog $\varepsilon := (\varepsilon_1, \ldots, \varepsilon_n)' \in \mathbb{R}^{n \times 1} \stackrel{d}{=} \mathcal{N}(0, \sigma_\varepsilon^2 I_n)$ yra nestebimos n.v.p. paklaidos su $\mathbb{E}\varepsilon_i = 0$, $\mathrm{Var}(\varepsilon_i) = \sigma_\varepsilon^2 > 0$.

Pagrindiniai šio disertacijos skyriaus rezultatai – tikslaus bei asimptotinio statistikos $\|\mathbb{X}'Y\|_2^2$ skirstinių išvedimas. Daroma prielaida, jog modelio kintamųjų kovariaciją nusako (8) KMS struktūra. Papildomai, tariama, jog $p, n \to \infty$ bei $p/n \to c \in (0, \infty)$. Įprastai panašūs rezultatai literatūroje gaunami pasitelkiant atsitiktinių matricų teoriją bei *Wishart* skirstinius (žr., Dicker (2014), Dicker ir Erdogdu (2016), Carpentier ir Verzelen (2019), Carpentier ir Verzelen (2021)), tačiau šiame disertacijos skyriuje pristatomi rezultatai remiasi *variance-gamma* (VG) skirstiniu bei jo savybėmis. Pasirinktas išvedimo būdas leidžia užrašyti tikslią statistikos pasiskirstymo funkciją bet kurioms p, n, ϱ reikšmėms, kas kartu atlaisvina kai kuriuos modelio parametrams daromus reikalavimus. Papildomai, gauta forma gali būti pritaikoma kitiems panašių struktūrų duomenims, pavyzdžiui, keičiant ℓ_2-normą kita.

Pagrindiniai statistikos $\|\mathbb{X}'Y\|_2^2$ asimptotinį normalumą nusakantys rezultatai suformuluojami 1 bei 2 teoremose. Tuo tikslu įvedami šie pažymėjimai:

$$\kappa_{1,p} := \sum_{k=1}^{p} \sum_{l=1}^{p} \beta_k \beta_l \varrho^{|k-l|}, \tag{9}$$

$$\kappa_{2,p} := \sum_{k=1}^{p}\Big(\sum_{l=1}^{p}\beta_l\varrho^{|k-l|}\Big)^2, \tag{10}$$

$$\kappa_{3,p} := \sum_{k,l,j,j'=1}^{p}\beta_j\beta_{j'}\varrho^{|k-j|}\varrho^{|l-j'|}\varrho^{|k-l|}. \tag{11}$$

Svarbu pastebėti, kad, galiojant prielaidai $\sum_{j=1}^{\infty}\beta_j^2 < \infty$, egzistuoja šios ribos:

$$\kappa_i = \lim_{p\to\infty}\kappa_{i,p}, \quad i = 1,2,3. \tag{12}$$

Taip pat, $\kappa_{i,p} \geq 0$, $i = 1,2,3$, $\forall p$.

Žemiau pateikiamos pagrindinių šiame skyriuje įrodytų rezultatų formuluotės.

1 teorema. *Tarkime, jog turime* (1) *modelį su* (8) *kovariacijų struktūra. Tegul* $n\to\infty$ *ir* $p = p_n$ *tenkina*

$$p\to\infty, \qquad \frac{p}{n}\to c\in(0,\infty). \tag{13}$$

Taip pat tarkime, jog β_j *tenkina*

$$\sum_{j=1}^{\infty}\beta_j^2 < \infty. \tag{14}$$

Tada

$$\frac{\|\mathbb{X}'Y\|_2^2 - n^2\kappa_{2,p} - pn(\kappa_{1,p}+\sigma_\varepsilon^2)}{n^{3/2}} \xrightarrow{d} \mathcal{N}(0,s^2), \tag{15}$$

čia dispersija s^2 *yra:*

$$s^2 = 4\kappa_2^2 + 4(\kappa_1+\sigma_\varepsilon^2)\,(2\kappa_2 c+\kappa_3) + 2c(\kappa_1+\sigma_\varepsilon^2)^2\Big(c+\frac{1+\varrho^2}{1-\varrho^2}\Big). \tag{16}$$

2 teoremoje patikslinami 1 teoremos rezultatai, įtraukiant papildomų reikalavimų modelio parametrams β_j. Tai leidžia modifikuoti statistiką centruojantį narį (15), keičiant $\kappa_{i,p}$ narių reikšmes ribinėmis, $i = 1,2$.

2 teorema. *Tegul galioja 1 teoremos prielaidos. Taip pat tarkime, kad*

$\sum_{j=p+1}^{\infty} \beta_j^2 = o(p^{-1/2})$ *bei* $\sup_{j \geq 1} |\beta_j| j^{\alpha} < \infty$ *su* $\alpha > 1/2$. *Tada*

$$\frac{\|\mathbb{X}'Y\|_2^2 - n^2\big(\kappa_2 + c(\kappa_1 + \sigma_\varepsilon^2)\big)}{n^{3/2}} \stackrel{d}{\to} \mathcal{N}(0, s^2), \tag{17}$$

čia dispersija s^2 *apibrėžta* (16) *lygtimi.*

Papildomai, suformuluojama išvada, kurioje nagrinėjamas $\varrho = 0$ atvejis, t.y., tariama, kad $\Sigma = I_p$. Šiuo tikslu įvedamas žymėjimas:

$$\beta(x) \;:=\; \sum_{j=1}^{\infty} \beta_j^2 x^j, \quad |x| \leq 1.$$

Išvadoje suformuluotas rezultatas seka iš 2 teoremos, pastebint, kad tokiu atveju $\kappa_i = \beta(1)$, $i = 1, 2, 3$.

3 išvada. *Tarkime, jog turime* (1) *modelį su kovariacijų matrica* $\Sigma = I_p$. *Tegul galioja* (13) *ir* (14) *prielaidos. Taip pat tarkime, kad* $\sum_{j=p+1}^{\infty} \beta_j^2 = o(p^{-1/2})$ *bei* $\sup_{j \geq 1} |\beta_j| j^{\alpha} < \infty$ *su* $\alpha > 1/2$. *Tada*

$$\frac{\|\mathbb{X}'Y\|_2^2 - n^2(\beta(1)(1 + c) + c\sigma_\varepsilon^2)}{n^{3/2}} \stackrel{d}{\to} \mathcal{N}(0, s^2), \tag{18}$$

čia

$$s^2 \;=\; 2\beta(1)^2\big(4 + 5c + c^2\big) + 4\beta(1)\sigma_\varepsilon^2\big(1 + 3c + c^2\big) + 2\sigma_\varepsilon^4(c + c^2). \tag{19}$$

Pagrindinių šiame skyriuje pristatytų rezultatų įrodymui svarbus *variance-gamma* (VG) skirstinys. Žemiau pristatomos pagrindinės šio skirstinio savybės.

VG skirstinys yra nusakomas parametrų $r > 0$, $\theta \in \mathbb{R}$, $\sigma > 0$ ir $\mu \in \mathbb{R}$, bei turi šią tankio funkciją:

$$\begin{aligned} f^{\mathrm{VG}}(x) \;&=\; \frac{1}{\sigma\sqrt{\pi}\Gamma(r/2)} \mathrm{e}^{\theta(x-\mu)/\sigma^2} \left(\frac{|x-\mu|}{2\sqrt{\theta^2+\sigma^2}}\right)^{(r-1)/2} \\ &\times\; K_{(r-1)/2}\left(\frac{\sqrt{\theta^2+\sigma^2}}{\sigma^2}\,|x-\mu|\right), \end{aligned} \tag{20}$$

čia $x \in \mathbb{R}$, $K_\nu(x)$ yra modifikuota antros rūšies Bessel funkcija. Atsitiktinį dydį Q su tankiu (20) žymime $Q \stackrel{d}{=} \mathrm{VG}(r, \theta, \sigma, \mu)$. Tegul

$\Gamma(a,b)$, $a>0$, $b>0$, žymi gama skirstinį su tankiu:

$$f^{\mathrm{G}}(x) \;=\; \frac{b^a}{\Gamma(a)}\,x^{a-1}\mathrm{e}^{-bx}, \quad x>0.$$

Tada

$$Q \;\stackrel{d}{=}\; \mu+\theta W_r+\sigma\sqrt{W_r}U, \tag{21}$$

čia $W_r \stackrel{d}{=} \Gamma(r/2,1/2)$, $U \stackrel{d}{=} \mathcal{N}(0,1)$, W_r ir U yra nepriklausomi. Charakteristinė atsitiktinio dydžio $Q \stackrel{d}{=} \mathrm{VG}(r,\theta,\sigma,\mu)$ funkcija turi šią formą (žr., pvz., Madan ir kt. (1998), Kotz ir kt. (2001)):

$$\varphi_Q(t) \;=\; \frac{\mathrm{e}^{\mathrm{i}\mu t}}{(1+\sigma^2 t-2\mathrm{i}\theta t)^{r/2}}, \quad t\in\mathbb{R}. \tag{22}$$

Atsitiktiniams dydžiams, turintiems VG pasiskirstymą, galioja šios savybės:

1. Jei $Q_1 \stackrel{d}{=} \mathrm{VG}(r_1,\theta,\sigma,\mu_1)$ ir $Q_2 \stackrel{d}{=} \mathrm{VG}(r_2,\theta,\sigma,\mu_2)$ yra nepriklausomi atsitiktiniai dydžiai, tada

$$Q_1+Q_2 \;\stackrel{d}{=}\; \mathrm{VG}(r_1+r_2,\theta,\sigma,\mu_1+\mu_2).$$

2. Jei $Q \stackrel{d}{=} \mathrm{VG}(r,\theta,\sigma,\mu)$, tada bet kuriam $a>0$

$$aQ \;\stackrel{d}{=}\; \mathrm{VG}(r,a\theta,a\sigma,a\mu).$$

4 teiginys. (i) *Jei* $(\xi_1,\xi_2)' \stackrel{d}{=} \mathcal{N}_2(0,\Sigma)$ *su* $\Sigma=\begin{pmatrix}\sigma_1^2 & \varrho\sigma_1\sigma_2\\ \varrho\sigma_1\sigma_2 & \sigma_2^2\end{pmatrix}$, *tada*

$$\xi_1\xi_2 \;\stackrel{d}{=}\; \mathrm{VG}(1,\varrho\sigma_1\sigma_2,\sqrt{1-\varrho^2}\sigma_1\sigma_2,0).$$

(ii) *Jei* $(\xi_{1j},\xi_{2j})'$, $j=1,\ldots,n$, *yra n.v.p. atsitiktiniai vektoriai turintys* $\mathcal{N}_2(0,\Sigma)$ *pasiskirstymą, tada*

$$\sum_{j=1}^{n}\xi_{1j}\xi_{2j} \;\stackrel{d}{=}\; \mathrm{VG}(n,\varrho\sigma_1\sigma_2,\sqrt{1-\varrho^2}\sigma_1\sigma_2,0)$$

ir

$$\sum_{j=1}^{n} \xi_{1j}\xi_{2j} \quad \overset{d}{=} \quad \sigma_1\sigma_2(\varrho W_n + \sqrt{1-\varrho^2}\sqrt{W_n}U),$$

čia $W_n \overset{d}{=} \Gamma(n/2, 1/2)$ *ir* $U \overset{d}{=} \mathcal{N}(0,1)$ *yra nepriklausomi atsitiktiniai dydžiai.*

(iii) *Tarkime, kad* $(\xi_{1j}^{(1)}, \ldots, \xi_{1j}^{(p)}, \xi_{2j})'$, $j = 1, \ldots, n$, *yra n.v.p. kopijos* $(\xi_1^{(1)}, \ldots, \xi_1^{(p)}, \xi_2)' \overset{d}{=} \mathcal{N}_{p+1}(0, \Sigma^{(p)})$ *ir tegul* $\varrho^{(kl)} := \mathrm{Corr}(\xi_1^{(k)}, \xi_1^{(l)})$, $\varrho^{(k)} := \mathrm{Corr}(\xi_1^{(k)}, \xi_2)$, $(\sigma_1^{(k)})^2 := \mathrm{Var}(\xi_1^{(k)})$, $\sigma_2^2 := \mathrm{Var}(\xi_2)$, $k, l = 1, \ldots, p$. *Tada*

$$\begin{pmatrix} \sum_{j=1}^{n} \xi_{1j}^{(1)}\xi_{2j} \\ \vdots \\ \sum_{j=1}^{n} \xi_{1j}^{(p)}\xi_{2j} \end{pmatrix} \overset{d}{=} \begin{pmatrix} \sigma_1^{(1)}\sigma_2(\varrho^{(1)}W_n + \sqrt{1-(\varrho^{(1)})^2}\sqrt{W_n}U_1) \\ \vdots \\ \sigma_1^{(p)}\sigma_2(\varrho^{(p)}W_n + \sqrt{1-(\varrho^{(p)})^2}\sqrt{W_n}U_p) \end{pmatrix},$$

čia $(U_1, \ldots, U_p)' \overset{d}{=} \mathcal{N}_p(0, \Sigma_U)$, $\Sigma_U = (\sigma_U^{(kl)})$ *su*

$$\sigma_U^{(k,l)} \quad = \quad \mathbb{E}U_kU_l \quad = \quad \frac{\varrho^{(kl)} - \varrho^{(k)}\varrho^{(l)}}{\sqrt{1-(\varrho^{(k)})^2}\sqrt{1-(\varrho^{(l)})^2}}, \quad k, l = 1, \ldots, p. \quad (23)$$

Disertacijoje pateiktas 4 teiginio įrodymas, kuriame remiamasi žemiau suformuluota 5 lema.

5 lema. *Tegul* $(\xi_1^{(1)}, \ldots, \xi_1^{(p)}, \xi_2)'$ *yra* $(p+1) \times 1$ *atsitiktiniai vektoriai, turintys* $\mathcal{N}_{p+1}(0, \Sigma^{(p)})$ *pasiskirstymą, bei tegul* $\varrho^{(kl)} := \mathrm{Corr}(\xi_1^{(k)}, \xi_1^{(l)})$, $\varrho^{(k)} := \mathrm{Corr}(\xi_1^{(k)}, \xi_2)$, $(\sigma_1^{(k)})^2 := \mathrm{Var}(\xi_1^{(k)})$, $\sigma_2^2 := \mathrm{Var}(\xi_2)$, $k, l = 1, \ldots, p$. *Tada*

$$\begin{pmatrix} \xi_1^{(1)}\xi_2 \\ \vdots \\ \xi_1^{(p)}\xi_2 \end{pmatrix} \overset{d}{=} \begin{pmatrix} \sigma_1^{(1)}\sigma_2\big(\varrho^{(1)}W_1 + \sqrt{1-(\varrho^{(1)})^2}\sqrt{W_1}U_1\big) \\ \vdots \\ \sigma_1^{(p)}\sigma_2\big(\varrho^{(p)}W_1 + \sqrt{1-(\varrho^{(p)})^2}\sqrt{W_1}U_p\big) \end{pmatrix},$$

čia $W_1 \overset{d}{=} \Gamma(1/2, 1/2)$, *bei* $(U_1, \ldots, U_p)'$ *yra, nepriklausomai nuo* W_1, *nulinį vidurkį turintys normalaus pasiskirstymo atsitiktiniai vektoriai su kovariacijų matrica, nusakyta* (23) *lygtimi.*

5 lemos įrodymas pateiktas disertacijoje. Toliau pristatomos kelios pagalbinės lemos.

6 lema. *Tegul* $V = (V_1, \ldots, V_p)' \stackrel{d}{=} \mathcal{N}_p(0, \Sigma_V^{(p)})$*, čia* $\Sigma_V^{(p)}$ *yra teigiamai apibrėžta kovariacijos matrica bei* $\operatorname{tr}((\Sigma_V^{(p)})^2) = o(p^2)$*,* $p \to \infty$*. Tada*

$$\frac{1}{p}\sum_{k=1}^{p}\left(V_k^2 - \mathbb{E}V_k^2\right) \ \stackrel{\mathbb{P}}{\to}\ 0, \quad kai \ \ p \to \infty. \tag{24}$$

Papildomai, jeigu $p^{-1}\operatorname{tr}(\Sigma_V^{(p)}) \to 1$*, tada*

$$\frac{1}{p}\sum_{k=1}^{p} V_k^2 \ \stackrel{\mathbb{P}}{\to}\ 1, \quad kai \ \ p \to \infty. \tag{25}$$

6 lemos įrodymas pateiktas disertacijoje.

7 lema. *Tarkime, kad* $\tilde{Z}_1, \tilde{Z}_2, \ldots$ *yra n.v.p.* $\mathcal{N}(0,1)$ *atsitiktiniai dydžiai. Bet kuriam* $p \in \mathbb{N}$ *apibrėžkime*

$$\zeta_j^{(p)} \ := \ \nu_j^{(p)}(\tilde{Z}_j^2 - 1) + \gamma_j^{(p)}\sqrt{p}\tilde{Z}_j, \quad j = 1, \ldots, p, \tag{26}$$

čia $\nu_j^{(p)}$*,* $j = 1, \ldots, p$*, yra teigiami skaliarai, taip pat* $\gamma_j^{(p)}$*,* $j = 1, \ldots, p$*, yra realūs skaliariai, tenkinantys*

$$\sum_{j=1}^{p}(\nu_j^{(p)})^3 \ = \ o\Bigg(\Bigg(\sum_{j=1}^{p}\operatorname{Var}\big(\zeta_j^{(p)}\big)\Bigg)^{3/2}\Bigg), \tag{27}$$

$$p\sum_{j=1}^{p}(\gamma_j^{(p)})^2\nu_j^{(p)} \ = \ o\Bigg(\Bigg(\sum_{j=1}^{p}\operatorname{Var}\big(\zeta_j^{(p)}\big)\Bigg)^{3/2}\Bigg), \tag{28}$$

čia $\operatorname{Var}(\zeta_j^{(p)}) = 2(\nu_j^{(p)})^2 + p(\gamma_j^{(p)})^2$*. Tada, kai* $p \to \infty$*,*

$$\frac{\sum_{j=1}^{p}\zeta_j^{(p)}}{\sqrt{\sum_{j=1}^{p}\operatorname{Var}(\zeta_j^{(p)})}} \ \stackrel{d}{\to}\ \mathcal{N}(0,1). \tag{29}$$

7 lemos įrodymas pateiktas disertacijoje.

Prieš pristatant 2 teoremos įrodymą, disertacijoje suformuluojama lema, užtikrinanti $\kappa_{1,p}$ bei $\kappa_{2,p}$ narių $\mathcal{O}(p^{-1/2})$ konvergavimo greitį prie papildomų reikalavimų modelio parametrams β_j.

8 lema. *Tarkime, kad* $\sum_{j=p+1}^{\infty}\beta_j^2 = o(p^{-1/2})$ *bei* $\sup_{j\geq 1}|\beta_j|j^{\alpha} < \infty$*,* $\alpha > 1/2$ *ir* $|\varrho| < 1$*. Tada*

1. $\kappa_1 \ = \ \kappa_{1,p} + o(p^{-1/2})$*,*

2. $\kappa_2 \;=\; \kappa_{2,p} + o(p^{-1/2})$.

8 lemos įrodymas pateiktas disertacijoje.

Žemiau pateikti keli papildomi rezultatai, leidžiantys išvesti alternatyvias κ_1, κ_2 bei κ_3 narių išraiškas. Šiuo tikslu įvedamos pagalbinės funkcijos $\beta(\cdot)$ bei $b(\cdot)$, nusakytos 9 apibrėžimu. Tuomet, tarus, jog 1 teoremos sąlygos yra išpildytos bei β_j parametrų struktūra yra žinoma, κ_1, κ_2 bei κ_3 išraiškoms rasti pakanka įvertinti šiuos narius: $\beta(1), \beta(\varrho), \beta(\varrho^2)$ ir $b_1(\varrho), b_2(\varrho)$. Minėtosios išraiškos suformuluotos 10 lemoje.

9 apibrėžimas. *Tegul* $\sum_{j=1}^{\infty} \beta_j^2 < \infty$ *ir* $|\varrho| \leq 1$. *Apibrėžkime*

$$\beta(\varrho) \;:=\; \sum_{j=1}^{\infty} \beta_j^2 \varrho^j, \tag{30}$$

$$b_1(\varrho) \;:=\; \sum_{j'=2}^{\infty} \sum_{j=1}^{j'-1} \beta_j \beta_{j'} \varrho^{j'-j}, \tag{31}$$

$$b_2(\varrho) \;:=\; \sum_{j=2}^{\infty} \sum_{j'=1}^{j-1} \beta_j \beta_{j'} \varrho^{j+j'}, \tag{32}$$

bei

$$\beta^{(1)}(\varrho) \;:=\; \varrho \frac{\mathrm{d}\beta(\varrho)}{\mathrm{d}\varrho} \;=\; \sum_{j=1}^{\infty} j \beta_j^2 \varrho^j, \tag{33}$$

$$b_1^{(1)}(\varrho) \;:=\; \varrho \frac{\mathrm{d}b_1(\varrho)}{\mathrm{d}\varrho} \;=\; \sum_{j'=2}^{\infty} \sum_{j=1}^{j'-1} \beta_j \beta_{j'} \varrho^{j'-j} (j'-j), \tag{34}$$

$$b_2^{(1)}(\varrho) \;:=\; \varrho \frac{\mathrm{d}b_2(\varrho)}{\mathrm{d}\varrho} \;=\; \sum_{j'=2}^{\infty} \sum_{j=1}^{j'-1} \beta_j \beta_{j'} \varrho^{j'+j} (j'+j), \tag{35}$$

$$b^{(2)}(\varrho) \;:=\; \varrho^2 \frac{\mathrm{d}^2 b_1(\varrho)}{\mathrm{d}\varrho^2} + b_1^{(1)}(\varrho) \;=\; \sum_{j'=2}^{\infty} \sum_{j=1}^{j'-1} \beta_j \beta_{j'} \varrho^{j'-j} (j'-j)^2. \tag{36}$$

10 lema. *Tegul galioja 1 teoremos sąlygos. Tarkime, jog* κ_1, κ_2 *ir* κ_3 *yra nusakytos* (9)–(12) *lygtimis. Tada, pasinaudojant 9 apibrėžimo žymėjimais, yra teisingos šios lygybės:*

1. $\kappa_1 \;=\; \beta(1) \;+\; 2b_1(\varrho)$,

$$\textit{2. } \kappa_2 = \beta(1)\frac{1+\varrho^2}{1-\varrho^2} - \beta(\varrho^2)\frac{1}{1-\varrho^2}$$
$$+ 2\Big(b_1^{(1)}(\varrho) + b_1(\varrho)\frac{1+\varrho^2}{1-\varrho^2} - b_2(\varrho)\frac{1}{1-\varrho^2}\Big),$$

$$\textit{3. } \kappa_3 = \frac{1}{1-\varrho^2}\big(3b_1^{(1)}(\varrho)(1+\varrho^2) - 2\big(b_2^{(1)}(\varrho) + \beta^{(1)}(\varrho^2)\big)\big) + b^{(2)}(\varrho)$$
$$+ \frac{1}{(1-\varrho^2)^2}(1+4\varrho^2+\varrho^4)(\beta(1)+2b_1(\varrho))$$
$$- \frac{1}{(1-\varrho^2)^2}(1+3\varrho^2)(\beta(\varrho^2)+2b_2(\varrho)).$$

10 lemos įrodymas pateiktas disertacijoje.

Likusioje 2 disertacijos skyriaus dalyje teoriniai rezultatai iliustruojami atliekant Monte Carlo imitacinius eksperimentus. Daroma prielaida, jog duomenis generuojantis procesas yra apytiksliai retos struktūros, t.y., galioja $\beta_j = j^{-1}, j \geq 1$.

Norint pritaikyti 2 teoremos rezultatus, pasinaudojama 10 lemoje gautomis išraiškomis. Šiuo tikslu yra randamos pagalbinės išraiškos: $\beta(1), \beta(\varrho), \beta(\varrho^2)$ ir $b_1(\varrho), b_2(\varrho)$.

Apibrėžkime realaus dilogaritmo funkciją (žr., pvz., Morris (1979)):

$$\mathrm{Li}_2(x) = -\int_0^x \frac{\log(1-u)}{u}\,\mathrm{d}u, \quad x \leq 1, \quad x \in \mathbb{R}. \tag{37}$$

(Čia ir kitur laikoma, jog $\int_0^x = -\int_x^0$ jeigu $x \leq 0$.) Galiojant $|x| \leq 1$, realus dilogaritmas turi šią išraišką:

$$\mathrm{Li}_2(x) = \sum_{k=1}^{\infty} \frac{x^k}{k^2}. \tag{38}$$

Tuomet,

$$\beta(1) = \sum_{j=1}^{\infty} \frac{1}{j^2} = \frac{\pi^2}{6}, \qquad \beta(\varrho) = \sum_{j=1}^{\infty} \frac{\varrho^j}{j^2} = \mathrm{Li}_2(\varrho). \tag{39}$$

Kadangi

$$\frac{\mathrm{d}}{\mathrm{d}\varrho}\mathrm{Li}_2(\varrho) = -\frac{\log(1-\varrho)}{\varrho}, \tag{40}$$

turime, jog

$$\beta^{(1)}(\varrho) \;=\; -\log(1-\varrho). \tag{41}$$

Analogiškai gaunamos likusios reikalingos išraiškos:

$$\begin{aligned}
b_1(\varrho) &= \frac{\log^2(1-\varrho)}{2} + \mathrm{Li}_2(\varrho), && (42)\\
b_1^{(1)}(\varrho) &= -\frac{\log(1-\varrho)}{1-\varrho}, && (43)\\
b^{(2)}(\varrho) &= \frac{\varrho - \varrho\log(1-\varrho)}{(1-\varrho)^2}, && (44)\\
b_2(\varrho) &= \frac{1}{2}\big(\log^2(1-\varrho) - \mathrm{Li}_2(\varrho^2)\big), && (45)\\
b_2^{(1)}(\varrho) &= \log\big(1-\varrho^2\big) - \frac{\varrho\log(1-\varrho)}{1-\varrho}. && (46)
\end{aligned}$$

Pasitelkus (39)–(46) bei 10(i) lemos rezultatus, gaunama ši κ_1 išraiška:

$$\kappa_1 \;=\; \frac{\pi^2}{6} \;+\; \log^2(1-\varrho) \;+\; 2\mathrm{Li}_2(\varrho). \tag{47}$$

Analogiškai, naudojant 10(ii) lemą, išvedama κ_2 išraiška:

$$\begin{aligned}
\kappa_2 &= \frac{1+\varrho^2}{1-\varrho^2}\Big(\frac{\pi^2}{6} + 2\mathrm{Li}_2(\varrho)\Big) \;-\; \frac{2\log(1-\varrho)}{1-\varrho} \;+\; \log^2(1-\varrho)\frac{\varrho^2}{1-\varrho^2}\\
&= \frac{1}{1-\varrho^2}\big((1+\varrho^2)\kappa_1 - \log^2(1-\varrho) - 2(1+\varrho)\log(1-\varrho)\big). && (48)
\end{aligned}$$

Taip pat, naudojant 10(iii) lemą, išvedama κ_3 išraiška:

$$\begin{aligned}
\kappa_3 &= \frac{1}{(1-\varrho^2)^2}\Big((1+4\varrho^2+\varrho^4)\Big(\frac{\pi^2}{6} + 2\mathrm{Li}_2(\varrho)\Big) + \log^2(1-\varrho)\varrho^2(1+\varrho^2)\\
&\quad - (3-\varrho+4\varrho^2)(1+\varrho)\log(1-\varrho) + \varrho(1+\varrho)^2\Big)\\
&= \frac{1}{(1-\varrho^2)^2}\Big((-1+\varrho+2\varrho^2)(1+\varrho)\log(1-\varrho) + \varrho(1+\varrho)^2 - 2\varrho^4\kappa_1\Big)\\
&\quad + \kappa_2\,\frac{1+3\varrho^2}{1-\varrho^2}. && (49)
\end{aligned}$$

Apytikslio retumo prielaida bei (47)–(49) išraiškos leidžia pritaikyti 2 teoremą ir apibendrinti šio poskyrio rezultatus žemiau pateikta 11 išvada.

11 išvada. *Tarkime, kad turime* (1) *modelį su* (8) *kintamųjų kovariacijos matrica. Taip pat tarkime, kad* $\beta_j := j^{-1}$, $j = 1, \ldots, p$. *Tegul* $p = p_n$ *tenkina*

$$p \to \infty, \qquad \frac{p}{n} \to c \in (0, \infty).$$

Tada

$$\frac{\|\mathbb{X}'Y\|_2^2 - n^2\big(\kappa_2 + c(\kappa_1 + \sigma_\varepsilon^2)\big)}{n^{3/2}} \ \stackrel{d}{\to} \ \mathcal{N}(0, s^2), \tag{50}$$

čia

$$s^2 \ = \ 4\kappa_2^2 + 4(\kappa_1 + \sigma_\varepsilon^2)\,(2\kappa_2 c + \kappa_3) + 2c(\kappa_1 + \sigma_\varepsilon^2)^2\Big(c + \frac{1 + \varrho^2}{1 - \varrho^2}\Big), \tag{51}$$

bei κ_1, κ_2 *ir* κ_3 *nusakyti* (47), (48) *bei* (49) *lygtimis, atitinkamai.*

2 disertacijos skyrius užbaigiamas atliekant Monte Carlo imitacinius eksperimentus. Generuojama 1000 nepriklausomų statistikos $\|\mathbb{X}'Y\|_2^2$ replikų pagal 11 išvados prielaidas. Analizuojamos šios parametrų reikšmės: $p = 100, 500, 1000, 1500, 2000, 3000$, $c = 1, 2, 5, 10$, $\sigma_\varepsilon^2 = 1, 2, 4, 10$. Pagrindiniai rezultatai pristatyti disertacijoje, 2.1–2.9 paveiksluose, kuriuose iliustruojamas gautų empirinių skirstinių panašumas į ribinius skirstinius, nusakytus 11 išvadoje. Paveiksluose vaizduojamos empirinės pasiskirstymo funkcijos (žymimos CDF), empirinės tankio funkcijos (žymimos PDF) bei atitinkami kvantilių grafikai (žymimi *Q-Q plots*), padedantys aiškiau iliustruoti empirinių skirstinių elgesį bei jų skirtumus.

Rezultatai rodo, kad mažoms parametro ϱ reikšmėms empiriniai skirstiniai yra arti teorinio ribinio skirstinio net ir palyginti mažoms parametrų p, n reikšmėms, bei aukštoms c reikšmėms. ϱ reikšmei augant stebimas lėtesnis konvergavimas. Prie vidutinių parametrų $\varrho, c, \sigma_\varepsilon^2$ reikšmių tik su palyginti aukštomis p reikšmėmis stebimas adekvatus empirinių skirstinių panašumas į ribinius. Analogiški rezultatai stebimi ir augant kitoms parametrų reikšmėms, jeigu likę parametrai yra atitinkamai kontroliuojami. Pavyzdžiui, mažinant parametrų σ_ε^2 arba c reikšmes bei atitinkamai didinant ϱ galima išlaikyti norimą skirstinių panašumo lygį. Galiausiai, aukštoms parametrų $\varrho, c, \sigma_\varepsilon^2$ reikšmėms stebimi didžiausi nukrypimai nuo ribinio skirstinio, leidžiantys teigti, kad tokiais atvejais reikalinga itin daug stebėjimų n arba kintamųjų p norint kuo panašesnių į ribinį skirstinį rezultatų.

Trečiame, **retų struktūrų: išankstinio JAV bei ES BVP komponenčių vertinimo**, skyriuje pristatomi pseudo-realaus laiko išankstinių prognozių vertinimo eksperimentai, prognozuojant JAV bei didžiųjų Europos šalių BVP išlaidų komponentes. Pagrindinis sudarytų eksperimentų tikslas – praktiškai palyginti populiarius retos struktūros metodus su žinomais tankios struktūros metodais, dažnai sutinkamais išankstinio vertinimo literatūroje. Papildomai, šiame skyriuje pasiūloma bei išsamiai apžvelgiama LASSO-PC modifikacija, apjungianti LASSO metodo variantus su pagrindinių komponenčių metodu.

Praktiniam metodų palyginimui disertacijoje pasirinkta spręsti išankstinio makroekonominių rodiklių prognozavimo uždavinius. Tikslus šių uždavinių sprendimas yra svarbus tiek institucijoms, tiek ir įvairiems ekonominiams agentams, kadangi savalaikis makroekonominės situacijos nustatymas gali reikšmingai lemti svarbius politinius sprendimus ar veiklos bei rizikos strategiją. Ši svarba ypatingai išryškėja ekonominių neramumų laikais, pavyzdžiui, 2007–2008 m. finansinės krizės metu.

Pagrindinis dėmesys šiame skyriuje skiriamas nacionalinių sąskaitų rodikliams. Oficialiai statistikos institucijos BVP bei jo komponenčių įverčius publikuoja su reikšmingu uždelsimu po atitinkamo ketvirčio pabaigos. Pavyzdžiui, pirmieji BVP įverčiai JAV bei ES yra publikuojami praėjus vienam mėnesiui po atitinkamo ketvirčio pabaigos stebint tik ekonomikos pasiūlos dalį (BVP gamybos komponentes), tuo tarpu paklausos dalis (BVP išlaidų komponentės) skelbiamos net su dviejų mėnesių vėlavimu.

Apžvelgiant išankstinio vertinimo literatūrą disertacijoje pastebima, jog empiriniuose taikymuose labiausiai paplitę tankių struktūrų metodai. Kadangi praktikoje yra prieinama visa aibė aukšto dažnio informacijos su palyginti nedideliu publikavimo uždelsimu, tyrėjai tikisi visą šią informaciją panaudoti išankstinio vertinimo modeliuose. Šiuo tikslu dažniausiai naudojami įvairūs faktorių modeliai, gebantys apibendrinti didžiulius informacijos kiekius latentinėje erdvėje bei atpažinti pagrindinius duomenų rinkinio signalus.

Kitavertus, makroekonominių indikatorių modeliavimo uždavinių specifika leidžia daryti retų struktūrų prielaidą – tarti, kad iš visos turimos didelio matavimo informacijos galima atpažinti kelis svarbiausius rodiklius, kurių pakaktų tinkamai įvertinti išankstinę ekonomikos situaciją. Minėtoji prielaida šiame disertacijos skyriuje tikrinama empiriškai, bandant atsakyti į klausimą, ar retų struktūrų metodais galima atpažinti bei įvertinti pagrindinius indikatorius, tinkamus išankstiniam BVP išlaidų komponenčių vertinimui. Taip pat, ar pasitelkus šiuos indikatorius gautos trumpalaikės prognozės bus tikslesnės lyginant su populiariausių tankių struktūrų metodų prognozėmis.

Papildomai, šiame skyriuje pristatomas bei apžvelgiamas LASSO bei pagrindinių komponenčių metodo derinys LASSO-PC, apjungiantis tiek retų, tiek ir tankių struktūrų metodus. Bai ir Ng (2008) pastebi, jog dirbant su makroekonominiais duomenimis kintamųjų atranka prieš atliekant bet kokią analizę gali būti labai svarbi, padedanti išvalyti turimą duomenų imtį nuo triukšmo. Tai savo ruožtu gali pridėti papildomo prognozavimo tikslumo, ypač dirbant su faktorių modeliais. Disertacijoje siūloma naudoti LASSO, Adaptive LASSO arba Square-Root LASSO pradinei kintamųjų atrankai dėl jų asimptotinių kintamųjų atrankos savybių dirbant su didelio matavimo duomenimis.

Toliau tekste laikoma, jog $\mathbb{X}^\lambda \in \mathbb{R}^{n\times q}$ žymi atrinktą reikšmingų kintamųjų poaibį naudojant pasirinktą LASSO atmainą, čia λ yra atitinkamas modelio baudos hiperparametras, $0 < q \leq n$. Taip pat, tariama, jog duomenys normalizuoti. Disertacijoje suformuluojamas šis LASSO-PC algoritmas:

1. Pasinaudojant LASSO ar kuria jo modifikacija atrinkti reikšmingų kintamųjų poaibį $\mathbb{X}^\lambda \in \mathbb{R}^{n\times q}$ iš duomenų rinkinio $\mathbb{X} \in \mathbb{R}^{n\times p}$, tam tikrai fiksuotai hiperparametro λ reikšmei, čia $0 < q \leq n$.

2. Įvertinti imties $\mathbb{X}^\lambda$ pagrindines komponentes, sudarant $F = \mathbb{X}^\lambda L$, čia $L \in \mathbb{R}^{q\times q}$ yra transformacijos matrica bei $F \in \mathbb{R}^{n\times q}$ žymi pagrindinių komponenčių matricą.

3. Sudaryti kiekvieno atrinkto reikšmingojo kintamojo $\mathbb{X}^\lambda$ prognozę pritaikant ARIMA laiko eilučių metodus. Šioje disertacijoje prognozuojami mėnesinio dažnumo duomenys, gautos prognozės agreguojamos į ketvirtinį dažnumą.

4. Gautas prognozes agreguoti į jų faktorių išraišką pasinaudojant transformacijos matrica L, t.y., sudaryti $\hat{F}^*_{n+h} = \hat{X}^\lambda_{n+h} L$, kur $h = 1, 2, \ldots$.

5. Naudojant LASSO įvertinti modelį $Y_n = \hat{F}_n\beta + \varepsilon_n$. Papildomai, analizę galima praplėsti į vertinamo modelio specifikaciją įtraukiant ir duomenų matricą $\mathbb{X}_n^\lambda$. Tokiu atveju, vertinti modelį $Y_n = (\hat{F}_n, X_n^\lambda)\beta + \varepsilon_n$.

6. Sudaryti galutines prognozes $\hat{Y}_{n+h} = \hat{F}_{n+h}\hat{\beta}$.

Kadangi 3 disertacijos skyriuje analizuojami makroekonominiai duomenys, tikėtina, jog atrinkti reikšmingieji kintamieji pasižymės stipriomis tarpusavio koreliacijomis. Papildomai, gali būti stebimos sudėtingos netiesinės priklausomybės. Pasiūlyta LASSO-PC modifikacija stipriai paremta Relaxed LASSO metodo idėjomis. Pastarojo formuluotė leidžia atskirti kintamųjų atrankos bei parametrų vertinimo procedūras į dvi dalis, siekiant tikslesnių modelio parametrų įverčių. LASSO-PC metodas taip pat reikalauja atskiro kintamųjų atrankos žingsnio. Tačiau parametrų vertinimo procedūra papildoma šiuo žingsniu: prieš vertinant modelio parametrus, atrinkti reikšmingieji kintamieji yra transformuojami į pagrindines komponentes. Pastarasis žingsnis motyvuojamas idėja, jog dalis atrinktų kintamųjų gali būti vedami to paties ekonominio signalo. Teisingai šiuos signalus atpažinus bei modeliuojant juos tiesiogiai, tikėtina, jog būtų galima pagerinti modeliuojamo proceso prognozių tikslumą. Parametrų vertinimui pasinaudojama Adaptive LASSO idėjomis, suformuluojama:

$$\begin{aligned}\hat{\beta}_{\text{LASSO-PC}} &= \arg\min_{\beta} (Y - \mathbb{X}^\lambda LL'\beta)'(Y - \mathbb{X}^\lambda LL'\beta) + \lambda||L'\beta||_1 \\ &= \arg\min_{\tilde{\beta}} (Y - F\tilde{\beta})'(Y - F\tilde{\beta}) + \lambda||\tilde{\beta}||_1, \end{aligned} \tag{52}$$

čia $LL' = I$ galioja bet kuriam $\tilde{q} \leq q$ pagal konstrukciją dėl pagrindinių komponenčių metodo, $\tilde{\beta} = L'\beta$; visi parametrai gali būti efektyviai įvertinti naudojant LARS algoritmą. Taigi, metodas leidžia sutraukti arba pašalinti silpnus signalus latentinėje erdvėje.

Disertacijoje pasiūlytas LASSO-PC metodas skiriasi nuo Bai ir Ng (2008) pasiūlyto metodo tuo, kad reikšmingi faktoriai parenkami pasitelkiant LASSO metodus, o ne informacinius kriterijus. T.y., tiek reikšmingų faktorių atranka, tiek modelio parametrų vertinimas bei jų sutraukimas yra atliekamas kartu, kas gali teigiamai veikti modelio prognozių rezultatus bei jų tikslumą. Galutiniai modelio koeficientai $\hat{\beta}^* := \hat{\hat{\beta}}L$ suformuojami naudojant tą pačią kintamųjų aibę kaip ir klasikiniu (Relaxed arba Adaptive) LASSO atveju, tačiau koeficientų reikšmės koreguojamos priklausomai nuo įvertintų signalų stiprumo bei reikšmingumo latentinėje erdvėje.

Modelio įvertintų faktorių prognozes disertacijoje siūloma sudaryti remiantis individualių atrinktų kintamųjų X_j^λ, $j = 1, \ldots, q$, prognozėmis. Pažymėkime $X_t^\lambda = (X_{1,t}^\lambda, \ldots, X_{q,t}^\lambda)$ atrinktų kintamųjų vektorių laiko momentu $t \in \{1, \ldots, n\}$. Tuomet, kiekvienam $h > 0$, prognozė $\hat{F}_{n+h}^* = (\hat{F}_{1,n+h}, \ldots, \hat{F}_{q,n+h})$ konstruojama kaip:

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

čia matrica L yra žinoma bei $\hat{X}_{j,n+h}^\lambda$, kiekvienam $j = 1, \ldots, q$, yra prognozuoti naudojant ARIMA laiko eilučių metodus.

Toks prognozių agregavimo būdas gali sumažinti prognozių paklaidas formuojant galutinius faktorius F tais atvejais, kai mes negalime tiksliai įvertinti individualių faktorių dinamikos. Pirma, jeigu reikšmingus faktorius sudarantys transformacijų svoriai yra panašaus dydžio, minėtasis prognozių agregavimas ekvivalentus *bagging* procedūrai (*bootstrap aggregation*, Breiman (1996)). Antra, tikėtina, jog F_j generuojantis procesas gali būti sudėtingas, ilgos atminties procesas, kurį tiesiogiai įvertinti laiko eilučių metodais gali būti sunku, ypač dirbant su trumpomis istorinėmis eilutėmis. Kitavertus, tiesioginis individualių kintamųjų prognozių agregavimas suteikia papildomos laisvės suklysti, kadangi agreguotos prognozės gali išlaikyti teisingą galutinio signalo kryptį. Papildomai, Granger (1980) parodė, jog mažos eilės AR/ARMA procesų agregavimas tam tikrais atvejais gali generuoti gerokai sudėtingesnės dinamikos procesus. Taikymuose tikimasi, kad naudojant prognozių agregavimą tam tikrais atvejais pavyks atkurti sudėtingą F_j eilutės dinamiką, taip pagerinant modelio prognozių tikslumą. Šio metodo praktinė nauda apžvelgiama disertacijoje, 3.5 lentelėje.

Pristatytas LASSO-PC metodo idėjas galima toliau plėsti, siekiant geresnių prognozavimo rezultatų. Šiame disertacijos skyriuje pristatomos dvi galimos pagrindinių komponenčių metodo modifikacijų kryptys: koreguojant signalų dydžius (angl. *scale transformation*) bei komponenčių transformacijos kampus (angl. *angle transformation*).

Signalo dydžio transformacijos tikslas – modifikuoti algoritmo įvertinamas pagrindines komponentes santykinai padidinant pasirinktų vertinime naudojamų kintamųjų svorius. Klasikiniu pagrindinių komponenčių atveju visi naudojami kintamieji yra normalizuoti. Šiuo atveju, tikimasi, jog tinkamai padidinus svarbių kintamųjų svorius sudarant pagrindines komponentes pavyktų tiksliau atpažinti aktualius ekonominius signalus bei tiksliau prognozuoti dominantį procesą. Panašios idėjos aptariamos Stakėnas (2012), čia autoriai pastebi reikšmingą prognozių tikslumo pagerėjimą naudojant *Weighted PCA* bei *Generalized PCA* metodus, prognozuojant Lietuvos BVP.

Kita kryptis – pagrindinių komponenčių transformacijos kampų modifikacija, kuri būtų gaunama atsisakius komponenčių ortogonalumo sąlygos. Tai pasiekti galima praretinant (pvz., pritaikant *Sparse PCA* (Zou ir kt. (2006))) arba kitu būdų modifikuojant komponenčių transformacijos matricą L. Tikėtina, jog kai kuriais atvejais tokia modifikacija galėtų leisti tiksliau atpažinti duomenyse slypinčius signalus, kas galimai leistų pagerinti modelio prognozių tikslumą. Ši idėja motyvuojama žemiau pristatytu pavyzdžiu.

Pažymėkime poaibį kintamųjų $Z \subset \{X_j : j = 1, \ldots, q\}$, ortogonalių visiems kitiems atrinktiems analizėje dalyvaujantiems kintamiesiems. Tarkime, jog kintamųjų matrica $\mathbb{X}^\lambda$ yra perrikiuota taip, kad kintamųjų blokas $Z \in \mathbb{R}^{n\times q_0}$, $q_0 < q$, pilnai apima svarbų signalą, kurio dėka tyrėjas galėtų sėkmingai modeliuoti jį dominantį procesą. Taip pat, tarkime, jog šis blokas yra ortogonalus likusiam blokui $\bar{\mathbb{X}}$. Tuomet LASSO, dirbdamas su tokiu būdu generuotomis pagrindinėmis komponentėmis, mėgins įtraukti kuo daugiau aiškinančiosios informacijos į modeliuojamą uždavinį. Kadangi suformuotos komponentės turės šią struktūrą:

$$F \;=\; \mathbb{X}^\lambda L \;=\; (Z,\; \bar{\mathbb{X}})L,$$

$j-$oji komponentė turės šią struktūrą:

$$f_j = (Z,\; \bar{\mathbb{X}})L_j \;=\; (Z,\; \bar{\mathbb{X}})[\Lambda_{q\times q_0}, \Phi_{q\times(q-q_0)}]_j'.$$

Taigi, LASSO, rekonstruodamas Z bloke slypintį signalą įtrauks per didelį skaičių faktorių f_j į galutinį modelį, nes kiekvienas iš įtrauktų faktorių turės informacijos iš Z bloko. Dalis šių faktorių į modelį nebūtų įtraukti, jeigu atitinkami komponenčių transformacijos matricos nariai būtų lygūs nuliui – šiame pavyzdyje tai atitiktų bloką Φ. Tarkime, jog

$\mathcal{G} \subset \{1, \ldots, q\}$ žymi rinkinį indeksų, kuriems f_j buvo atrinktas kaip reikšmingas panaudojant LASSO tik dėl minėtųjų nenulinių svorių transformacijos matricoje Φ'_j. Tuomet, su kiekvienu tokiu papildomai įtrauktu faktoriumi f_r, $r \in \mathcal{G}$, į modelį kartu įtraukiama triukšmo, priklausančio nuo $\bar{\mathbb{X}}\Phi'_r$ lygio. Taigi, kuo daugiau tokių kintamųjų įtraukiama, tuo LASSO-PC būdu sudarytas modelis yra panašesnis į Relaxed LASSO modelį.

Todėl disertacijoje siūloma LASSO-PC procedūrą išplėsti papildomu žingsniu, modifikuojant transformacijos matricą L. Vienas būdas, modifikuoti transformacijos matricą L ją praretinant (pvz., pritaikant Sparse PCA). Kitas galimas būdas – apjungiant atrinktų reikšmingų kintamųjų matricą $\mathbb{X}^\lambda$ su įvertintų faktorių matrica F ir modeliuojant juos kartu, konstruojant $\bar{F} := (F, \mathbb{X}^\lambda)$. Šiuo būdu LASSO atliktų reikšmingų kintamųjų atranką bloke $\bar{F}$. Tokią procedūrą galima interpretuoti kaip tam tikrą transformacijos matricos L modifikaciją. Tokiu atveju, Sparse PCA, LASSO bei LASSO-PC procedūros atitiktų specifinius atvejus, priklausomai nuo įvertintų $\bar{F}_j$ koeficientų.

Likusioje 3 disertacijos skyriaus dalyje pristatomi atliktų pseudorealaus laiko eksperimentų rezultatai. Nagrinėjamos keturios pagrindinės BVP išlaidų komponentės: Bendrojo Pagrindinio Kapitalo Formavimas (BPKF), vartojimo išlaidos, importas bei eksportas. Tyrime naudojami duomenys yra sezoniškai išlyginti bei stacionarizuoti. Taip pat, kai kuriais atvejais atliktos kintamųjų transformacijos bei pašalintos adityvios išskirtys. Visos atliktos kintamųjų transformacijos išsamiai apžvelgtos disertacijos Priedo B.1 skyrelyje. Priklausomai nuo modeliuojamos šalies, tyrime naudojami kintamieji iš Eurostat arba FRED duomenų bazių, apimantys laikotarpį nuo 1990 iki 2019. JAV atveju, modeliuojant naudojama daugiau nei 2400 mėnesinio dažnumo indikatorių.

Siekiant, kad prognozavimo eksperimentai kuo tiksliau atspindėtų realią situaciją, kiekvienos tyrime atliktos iteracijos metu buvo imituojamas duomenų prieinamumas sudarant realistišką duomenų publikavimo tvarkaraštį. Tokio tvarkaraščio pavyzdys pateiktas 1 lentelėje. Kiekvieną eksperimento ketvirtį stebime bent 3 duomenų atnaujinimus, tačiau šiame tyrime pagrindiniai rezultatai buvo gauti tariant, jog prognozės sudaromos kiekvieno ketvirčio paskutinį mėnesį. Tokio pasirinkimo tikslas – kokybiškai atskirti trumpalaikes 1 ketvirčio prognozes nuo išankstinių.

Mėnesinio dažnumo kintamieji buvo agreguoti į ketvirtinius vidurki-

1 lentelė: Duomenų skelbimo tvarkaraštis bei prieinamumas pseudo-realaus-laiko eksperimentuose

Kintamasis	BPKF	Atbulinė, K1			Išankstinė, K2			Trumpalaikė, K3			Trumpalaikė, K4			Ketvirtis
		1	2	3	4	5	6	7	8	9	10	11	12	Mėnuo
Grupė	Pramonės gamyba													
	Apklausos													
	Palūkanų normos													
	Akcijų rinkos indeksai													
	Prekyba													

Legenda: Turima informacija: ▒ Prognozės mėnuo: ☐ Prognozuoti duomenys: ░

nant, trūkstamos reikšmės užpildomos ARIMA laiko eilučių metodais.

Eksperimentai atlikti simuliuojant 2005K1–2019K1 laikotarpį, sudaromos 4-ių tipų prognozės: praėjusio ketvirčio prognozė (angl. *backcast*), išankstinė einamojo ketvirčio prognozė (angl. *nowcast*), bei dvi trumpalaikės 1 ir 2 ketvirčių prognozės. Nagrinėtos BVP išlaidų komponentės šioms 5 šalims: JAV, Prancūzija, Italija, Ispanija bei Vokietija. Pagrindinės įžvalgos suformuluojamos aptariant JAV komponenčių prognozavimo rezultatus. Šios įžvalgos vėliau apibendrinamos, pasitelkiant Europos šalių rezultatus.

Aptariant eksperimentų rezultatus pagrindinis dėmesys skirtas BPKF komponentei, sudarytai iš investicijų skirtinguose ekonomikos sektoriuose bei pramonės šakose. Šią komponentę turėtų veikti ne tik signalai, stebimi ekonomikoje, bet ir investuotojų lūkesčiai. Taip pat, įprastai BPKF pasižymi didesne sklaida nei likusios BVP komponentės, todėl ši komponentė yra palyginti sunkiau prognozuojama. Priešingai nuo likusių BVP komponenčių, BPKF neturi mėnesinio dažnumo atitikmenų.

Tyrime nagrinėti šie modeliai: Square-Root LASSO (lentelėse žymima kaip SqL), Relaxed LASSO (RL), Adaptive LASSO (AdL), LASSO (L), L0Lq modeliai (L0, L0L1, L0L2) bei pasiūlyta LASSO-PC modifikacija, kurioje pradinė kintamųjų atranka atlikta tam tikro LASSO varianto, pvz., Adaptive LASSO-PC (AdP), L0-PC (L0P), L0L1-PC (L0L1P), L0L2-PC (L0L2P), Square-Root LASSO-PC (SqP) ar LASSO-PC (LP). Visais atvejais „P“ raidė pridedama prie atitinkamų modelių žymėjimų, jeigu buvo papildomai taikyta pagrindinių komponenčių transformacija. Taip pat, LASSO-PC-X (lentelėse LPX) žymi modelį, kuriame kartu su pagrindinėmis komponentėmis įtraukta ir netransformuotų kintamųjų matrica.

Retų struktūrų metodai lyginti su išankstinio vertinimo literatūroje populiariais tankių struktūrų metodais: *diffusion index* modeliais (lentelėse žymima DI, Stock ir Watson (2002)) bei dinaminiais faktorių modeliais (DF, Giannone ir kt. (2008), Banbura ir kt. (2011)). Papildomai, naudota Bai ir Ng (2008) pasiūlyta dinaminių faktorių modifikacija (lentelėse žymima BNDF), kurioje reikšmingų kintamųjų atranka buvo atliekama taikant LASSO.

Tiriant JAV BPKF eksperimentų rezultatus, pastebima, jog dauguma retos struktūros metodų prognozuoja ganėtinai panašiai, įskaitant ir BNDF metodą. Silpnesnės prognozės gautos SW bei DF metodais, didžiausią įtaką tokiam rezultatui darė itin netikslios prognozės finansinės krizės laikotarpiu. Silpniausi prognozių rezultatai gauti ARMA modeliais, tačiau verta pastebėti, jog 2005K1–2007K4 laikotarpiu, ARMA generuotos 2 ketvirčių prognozės buvo tiksliausios tarp visų tyrime naudotų metodų. Panašūs rezultatai stebimi ir literatūroje, pvz., D'Agostino ir Giannone (2012) pastebi, jog ramaus bei stabilaus ekonominio augimo laikotarpiais net ir sudėtingi bei gerai prognozuojantys modeliai gali nesugebėti pagerinti paprastų AR modelio prognozių.

Kaip ir buvo tikėtasi literatūros apžvalgoje, RL bei SqL modeliai dažnu atveju prognozavo tiksliau nei LASSO. Tačiau AdL ir L0Lq metodų pranašumai nebuvo tokie ryškūs: visi L0Lq metodai gerokai pagerino 2 ketvirčių prognozių tikslumą, tačiau panašus efektas nebuvo stebimas išankstinėse bei 1 ketvirčio prognozėse. Vienas iš galimų paaiškinimų susijęs su reikšmingų atrenkamų kintamųjų skaičiumi – pavyzdžiui, AdL metodai vidutiniškai atrinkdavo 55 kintamuosius kaip reikšmingus, tuo tarpu RL bei SqL metodai atrinkdavo tarp 23–26 kintamųjų. Kartu, L0Lq metodai atrinkdavo 2–9 kintamuosius kaip reikšmingus, iš kurių rečiausius modelius sudarydavo L0 metodas, atrinkdamas vidutiniškai tik po 2 kintamuosius. Panašu, kad per didelis įtrauktos informacijos kiekis stipriai padidino modelio prognozių paklaidas. Tuo tarpu pernelyg reti modeliai sugebėjo įvertinti ilgalaikę dinamiką, tačiau, panašu, kad naudotos informacijos neužteko adekvačioms išankstinėms bei 1 ketvirčio prognozėms. Minėtieji skirtumai ypač pastebimi 2008K1–2014K1 laikotarpiu.

Pagrindinių komponenčių apjungimas su LASSO tam tikrais atvejais reikšmingai pagerino prognozių tikslumą. Lyginant atitinkamų modelių poras (LP su RL, SqP su SqL, AdP su AdL) nagrinėta LASSO-PC transformacija daugumoje atvejų pagerino prognozių tikslumą. Svarbu pastebėti, jog kai kuriais atvejais šie skirtumai buvo labai maži. Prognozės nebuvo pagerintos L0Lq metodams, tikėtina, dėl pernelyg mažo atrinktų kintamųjų skaičiaus.

Lyginant retų struktūrų modelius su tankių struktūrų faktorių modeliais, LP bei SqL metodai reikšmingai pagerino DF bei SW metodų išankstines prognozes (remiantis Diebold-Mariano (DM) testu, su 5% reikšmingumo lygmeniu). Panašiai, RL bei SqL metodų trumpalaikės 1 ir 2 ketvirčių prognozės buvo reikšmingai tikslesnės už DF bei SW modelius. Svarbu paminėti, jog bendrai tiksliausios išankstinės prognozės buvo sudarytos BNDF modelio.

Siekiant ištirti, kokią papildomą vertę prognozavimo tikslumui atneša pagrindinių komponenčių transformacijos taikymas LASSO-PC metodikoje, atliekamas išsamus palyginimas fiksavus atrenkamų reikšmingų kintamųjų skaičių į 10, 15, 20 bei 30. Tuomet, lyginamas atitinkamų RL bei LP modelių prognozavimo tikslumas. Pradinė kintamųjų atranka abiems modeliams yra identiška, todėl visus stebimus prognozių skirtumus lemia tik atliekama pagrindinių komponenčių transformacija. Eksperimentų rezultatai rodo, jog LP transformacija daugumoje atvejų pagerino tiek išankstinių, tiek ir trumpalaikių 1 bei 2 ketvirčių prognozių tikslumą. Remiantis DM testu su 5 % reikšmingumo lygmeniu, statistiškai reikšmingas prognozių tikslumo pagerėjimas stebėtas LP15 bei LP30 atvejais sudarant 1 ketvirčio prognozes. Kitavertus, išankstinių bei 2 ketvirčių prognozių tikslumo pagerėjimas nebuvo statistiškai reikšmingas.

Gauti rezultatai leidžia daryti išvadą, jog pritaikius LASSO-PC transformaciją nežymus prognozių tikslumo pagerėjimas stebimas ir sąlyginai mažoms kintamųjų imtims, tačiau nauda lengviau pastebima kintamųjų skaičiui augant. Atrinkus apie 30 kintamųjų skirtumai tampa ryškesni. Tikėtina, jog tokiu atveju atrinkti kintamieji suformuoja koreliuotų signalų grupes. Disertacijoje toks elgesys pastebimas ir 3.4.2.3 skirsnyje, analizuojant įvertintus retus signalus. Šios grupės, tikėtina, padeda tiksliau įvertinti ekonominius signalus konstruojant pagrindines komponentes. Savo ruožtu, koreliuojančios kintamųjų grupės neigiamai veikia RL metodo prognozes.

Atrinktų reikšmingų signalų grupės aktualios ir iš kokybinės pusės, tyrėjui siekiant geriau suprasti modeliuojamą procesą. Sekant eksperimento metu kiekvienu ketvirčiu susiformuojančias reikšmingų kintamųjų grupes galime nusakyti pagrindinių signalų dinamiką bei struktūrą, taip pat jos pokyčius po ekonominio šoko ar skirtingais verslo ciklo momentais. Pagrindinės įžvalgos trumpai pristatytos disertacijoje bei 3.2–3.5 paveiksluose, B.1–B.5 lentelėse.

Papildomai, atliekamas pagrindinių komponenčių prognozių agregavimo būdų palyginimas. Nagrinėti du scenarijai, sudarius specifinį post-LASSO modelį, fiksuojant reikšmingų kintamųjų poaibį naudojant LASSO, bei įvertinus pagrindines 5 šio poaibio komponentes. Komponenčių prognozės konstruotos dviem būdais: pirmasis, prognozuojant mėnesinius kintamuosius individualiai, taikant ARIMA laiko eilučių metodus, gautas prognozes agreguojant į ketvirtinį dažnumą, kurios vėliau naudojamos sudarant galutines pagrindinių komponenčių prognozes (šis metodas disertacijoje žymimas AggregatedPC); bei antrasis, prognozuojant šias pagrindines komponentes tiesiogiai, taikant ARIMA laiko eilučių metodus ketvirtinio dažnumo lygmenyje (DirectPC). Disertacijoje pateikiami palyginimo rezultatai. Esminės įžvalgos yra šios: tiek agreguotas, tiek ir tiesioginis metodas ganėtinai tiksliai sudarė išankstines prognozes. Tokių rezultatų galima buvo tikėtis, nes dalis informacijos einamuoju ketvirčiu jau yra žinoma. Didesnis skirtumas stebimas lyginant trumpalaikes 1 bei 2 ketvirčių prognozes. Panašu, jog tiesiogiai prognozuojant pagrindines komponentes nebuvo adekvačiai įvertinta signalo dinamika, o tai lėmė didesnes prognozių paklaidas.

Disertacijos skyrius užbaigiamas kartojant dalį pristatytų palyginimų, keičiant kai kuriuos pseudo-realaus-laiko eksperimentų parametrus. Pirma, vietoje plečiamo istorinių duomenų lango principo taikytas slenkančio lango principas. Antra, atliktas metodų palyginimas prognozuojant likusias BVP išlaidų komponentes: importą, eksportą bei vidaus vartojimo išlaidas. Lyginant su BKPF, išankstinis šių komponenčių prognozių vertinimas yra paprastesnis, nes kiekvienai iš jų egzistuoja atitinkami mėnesinio dažnumo indikatoriai (pvz., mėnesiniai importo, eksporto bei vartojimo rodikliai). Tikėtina, jog šie kintamieji bus esminiai, nusakantys išankstinių prognozių dinamiką. Trečia, analizė atliekama 4 Europos šalims: Italijai, Ispanijai, Prancūzijai ir Vokietijai. Dėl to kyla papildomas skirtumas, kad priklausomai nuo šalies, skiriasi prieinamos informacijos kiekiai bei istorinių eilučių ilgiai.

Apibendrinant visus eksperimentų rezultatus disertacijoje pristatomos šios išvados. Pirma, nei vienas iš taikytų metodų nebuvo universaliai geriausias. Iš tiesų, buvo stebima šalių bei indikatorių, kuriems tankių struktūrų metodai generavo tiksliausias prognozes. Kartu, vertinant Italijos vidaus vartojimo rodiklį, nei retų nei tankių struktūrų metodai nesugeneravo tikslesnių prognozių už ARIMA laiko eilučių metodus.

Antra, tam tikrais atvejais LASSO-PC metodo variantai generavo tiksliausias išankstines prognozes bei 1 ir 2 ketvirčių prognozes. Nors dalis šių rezultatų nebuvo statistiškai reikšmingi remiantis GW prognozių tikslumo testu, gautos išvados iš esmės sutapo su išvadomis, darytomis analizuojant JAV BPKF prognozių rezultatus. Svarbu pastebėti, jog tyrime taikytų BNDF bei DF metodų išankstinės prognozės buvo ruoštos taikant aibę išankstinio vertinimo literatūroje žinomų įrankių (pvz., kintamųjų agregavimo strategijos, *bridge* lygtys, nusakytos Bańbura ir kt. (2013)), kurie nebuvo naudojami nei vienu retos struktūros atveju. Šių įrankių dėka daugumoje atvejų stebimas labai geras DF bei BNDF metodų išankstinių prognozių tikslumas. Kitavertus, minėtųjų modifikacijų pranašumas sparčiai išnyksta sudarant trumpalaikes 1 ir 2 ketvirčių prognozes, kas leidžia tiksliau palyginti retų bei tankių struktūrų metodus.

Taip pat, disertacijoje daroma išvada, jog pasiūlyta LASSO-PC modifikacija stabiliai demonstravo prognozių tikslumo pagerėjimą prognozuojant BPKF rodiklius visoms tyrime dalyvavusioms šalims. Kitavertus, likusiems indikatoriams skirtumai nebuvo tokie ryškūs bei stabilūs. Modelių stabilumą disertacijoje iliustruoja 3.8 lentelė, kurioje pateikiamos tyrime nagrinėtų metodų pozicijų medianos. Pastebima, jog nei vienas metodas nebuvo visuomet tiksliausias, tačiau LP bei LPX metodai viso eksperimento metu generavo stabiliausias išankstines bei trumpalaikes 1 ir 2 ketvirčių prognozes. Likusių modelių rezultatai buvo labiau varijuojantys, indikuojantys mažesnį prognozių patikimumą bei stabilumą.

Išvados

Disertacijoje nagrinėta retų struktūrų daugiamatėje tiesinėje regresijoje problematika. Ypatingas dėmesys skirtas apytikslio retumo prielaidai tirti, kuri yra kertinė daugumoje retos struktūros metodų. Siekiant įvertinti praktinius retų struktūrų modelių pranašumus disertacijoje atlikti pseudo-realaus laiko eksperimentai, vertinant išankstines bei trumpalaikes makroekonominių rodiklių prognozes. Žemiau apibendrinami pagrindiniai disertacijos rezultatai.

Pirmoje disertacijos dalyje nagrinėta $\|\mathbb{X}'Y\|_2^2$ statistika bei išvesti jos tikslus bei asimptotinis skirstiniai. Verta akcentuoti rezultatų išvedime naudoto *variance-gamma* skirstinio svarbą, kurį pasitelkus gautus rezultatus galima būtų plėtoti toliau, nagrinėjant panašių formų statistikas.

Monte Carlo imitacinių eksperimentų rezultatai leidžia daryti išvadą, kad empirinis statistikos $\|\mathbb{X}'Y\|_2^2$ skirstinys sąlyginai greitai artėja prie ribinio skirstinio disertacijoje tyrinėtais scenarijais. Tai rodo, jog gauti rezultatai gali būti pritaikomi tiek apytikslio retumo testavimo, tiek susijusiuose signalo bei triukšmo santykio vertinimo uždaviniuose.

Tolesniems tyrimams prasminga būtų rezultatus papildyti apibendrinant nagrinėjamo tiesinio modelio prielaidas, pavyzdžiui, leidžiant kintamiesiems įgyti sudėtingesnę tarpusavio priklausomybę. Papildomai, rezultatai galėtų būti išplėsti sudarant didžiausiojo tikėtinumo funkcijas ir jas panaudojant signalo bei triukšmo santykio ar parametrų β vertinimo uždaviniuose.

Antroje disertacijos dalyje tyrinėtas retų struktūrų metodų prognozavimo tikslumas lyginant su ARMA bei tankių struktūrų faktorių modeliais. Eksperimentų rezultatai rodo, jog visi nagrinėti LASSO metodai geba sudaryti adekvačias trumpalaikes prognozes, dažnu atveju tikslesnes už ARMA bei faktorių modelių. Taip pat, retų struktūrų metodai geba įvertinti bei atpažinti esmines retas struktūras modeliuojamuose duomenyse. Vidutinis įvertintų reikšmingųjų kintamųjų skaičius buvo santykinai mažas, kas leidžia teigti, jog retos struktūros prielaida gali būti išpildyta modeliuojant JAV bei Europos šalių makroekonominius rodiklius.

Papildomai, daugumoje tyrinėtų atvejų populiarios LASSO modifikacijos gebėjo reikšmingai pagerinti LASSO prognozių tikslumą. Kai kuriais atvejais disertacijoje pasiūlyta LASSO-PC modifikacija suteikė papildomo prognozavimo tikslumo, lyginant su Orakulo savybes turinčiais Adaptive LASSO bei Relaxed LASSO metodais. Šie rezultatai suteikia pagrindo tęsti tyrimus bei toliau plėtoti LASSO metodologiją, siekiant tikslesnio retų struktūrų atpažinimo bei papildomo prognozavimo tikslumo.

Literatūra

Andreini, P., C. Izzo ir G. Ricco (2020). Deep dynamic factor models. *arXiv preprint arxiv:2007.11887*.

Bai, J. ir 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 ir L. Reichlin (2011). Nowcasting. In M. P. Clements ir D. F. Hendry (Eds.), *Oxford Handbook on Economic Forecasting*. Oxford: Oxford University Press.

Bańbura, M., D. Giannone, M. Modugno ir L. Reichlin (2013). Now-Casting and the Real-Time Data Flow. In G. Elliott ir A. Timmermann (Eds.), *Handbook of Economic Forecasting*, Volume 2 of *Handbook of Economic Forecasting*, pp. 195–237. Elsevier.

Belloni, A. ir V. Chernozhukov (2011). High dimensional sparse econometric models: An introduction. In *Inverse Problems and High-Dimensional Estimation*, pp. 121–156. Springer.

Belloni, A., V. Chernozhukov, D. Chetverikov, C. Hansen ir K. Kato (2018). High-dimensional econometrics and regularized GMM. *arXiv preprint arXiv:1806.01888*.

Belloni, A., V. Chernozhukov ir L. Wang (2011). Square-root lasso: pivotal recovery of sparse signals via conic programming. *Biometrika 98*(4), 791–806.

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

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

Caner, M. ir A. B. Kock (2018). Asymptotically honest confidence regions for high dimensional parameters by the desparsified conservative lasso. *Journal of Econometrics 203*(1), 143–168.

Carpentier, A. ir N. Verzelen (2019). Adaptive estimation of the sparsity in the gaussian vector model. *The Annals of Statistics 47*(1), 93–126.

Carpentier, A. ir N. Verzelen (2021). Optimal sparsity testing in linear regression model. *Bernoulli 27*(2), 727–750.

Cevid, D., P. Bühlmann ir N. Meinshausen (2020). Spectral deconfounding via perturbed sparse linear models. *Journal of Machine Learning Research 21*(232), 1–41.

Chernozhukov, V., C. Hansen ir Y. Liao (2017). A lava attack on the recovery of sums of dense and sparse signals. *Annals of Statistics 45*(1), 39–76.

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

Dicker, L. H. (2014). Variance estimation in high-dimensional linear models. *Biometrika 101*(2), 269–284.

Dicker, L. H. (2016). Ridge regression and asymptotic minimax estimation over spheres of growing dimension. *Bernoulli 22*(1), 1–37.

Dicker, L. H. ir M. A. Erdogdu (2016). Maximum likelihood for variance estimation in high-dimensional linear models. In *Artificial Intelligence and Statistics*, pp. 159–167. PMLR.

Diebold, F. X. (2003). Big data dynamic factor models for macroeconomic measurement and forecasting. In L. H. M. Dewatripont ir S. Turnovsky (Eds.), *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress of the Econometric Society*, pp. 115–122.

D'Agostino, A. ir D. Giannone (2012). Comparing alternative predictors based on large-panel factor models. *Oxford bulletin of economics and statistics 74*(2), 306–326.

Fikioris, G. (2018). Spectral properties of Kac–Murdock–Szegő matrices with a complex parameter. *Linear Algebra and its Applications 553*, 182–210.

Giannone, D., M. Lenza ir G. E. Primiceri (2021). Economic predictions with big data: The illusion of sparsity. *Econometrica 89*(5), 2409–2437.

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

Gold, D., J. Lederer ir J. Tao (2020). Inference for high-dimensional instrumental variables regression. *Journal of Econometrics 217*(1), 79–111.

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

Guo, Z., D. Ćevid ir P. Bühlmann (2021). Doubly debiased lasso: High-dimensional inference under hidden confounding. *ArXiv preprint arXiv:2004.03758*.

Hazimeh, H. ir R. Mazumder (2020). Fast best subset selection: Coordinate descent and local combinatorial optimization algorithms. *Operations Research 68*(5), 1517–1537.

Hoerl, A. E. ir R. W. Kennard (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics 12*(1), 55–67.

Ing, C.-K. (2007). Accumulated prediction errors, information criteria and optimal forecasting for autoregressive time series. *The Annals of Statistics 35*(3), 1238–1277.

Javanmard, A. ir A. Montanari (2014). Confidence intervals and hypothesis testing for high-dimensional regression. *The Journal of Machine Learning Research 15*(1), 2869–2909.

Kac, M., W. Murdock ir G. Szegő (1953). On the eigen-values of certain Hermitian forms. *Journal of Linear Rational Mechanics and Analysis 2*, 767–800.

Kotz, S., T. Kozubowski ir K. Podgórski (2001). The laplace distribution and generalizations: A revisit with applications to communications, economics, engineering, and finance. Boston. Birkhäuser.

Madan, D. B., P. P. Carr ir E. C. Chang (1998). The variance gamma process and option pricing. *Review of Finance 2*(1), 79–105.

Magnus, J. R. ir G. De Luca (2016). Weighted-average least squares (WALS): a survey. *Journal of Economic Surveys 30*(1), 117–148.

Mazumder, R., P. Radchenko ir A. Dedieu (2022). Subset selection with shrinkage: Sparse linear modeling when the snr is low. *Operations Research*.

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

Morris, R. (1979). The dilogarithm function of a real argument. *Mathematics of Computation 33*(146), 778–787.

Ng, S. (2013). Variable selection in predictive regressions. In *Handbook of Economic Forecasting*, Volume 2, pp. 752–789. Elsevier B.V.

Ning, Y., S. Peng ir J. Tao (2020). Doubly robust semiparametric difference-in-differences estimators with high-dimensional data. *ArXiv preprint arXiv:2009.03151*.

Shibata, R. (1980). Asymptotically efficient selection of the order of the model for estimating parameters of a linear process. *The Annals of Statistics*, 147–164.

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. ir M. Watson (2002). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics 20*(2), 147–62.

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

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

Tikhonov, A. N. (1963). On the solution of ill-posed problems and the method of regularization. In *Doklady Akademii Nauk*, Volume 151, pp. 501–504. Russian Academy of Sciences.

Yang, Y., J. Zhou ir J. Pan (2021). Estimation and optimal structure selection of high-dimensional toeplitz covariance matrix. *Journal of Multivariate Analysis 184*, 104739.

Zhang, C.-H. ir S. S. Zhang (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 76*(1), 217–242.

Zou, H. (2006). The Adaptive Lasso and Its Oracle Properties. *Journal of the American Statistical Association 101*(476), 1418–1429.

Zou, H., T. Hastie ir R. Tibshirani (2006). Sparse Principal Component Analysis. *Journal of Computational and Graphical Statistics 15*(2), 265–286.

Trumpos žinios apie autorių

Saulius Jokubaitis 2011 m. įstojo į Vilniaus universitetą, Matematikos ir informatikos fakultetą, studijavo ekonometriją ir 2015 m. įgijo statistikos bakalauro laipsnį. Ekonometrijos studijas tęsė toliau, 2017 m. Vilniaus universiteto Matematikos ir informatikos fakultete įgijo statistikos magistro laipsnį. 2017–2021 m. studijavo Vilniaus universiteto doktorantūroje. 2019–2022 m. dalyvavo „Reassessment of the Optimal Currency Area theory in the persistently heterogeneous European Union“ (Euro4Europe) projekte, dirbo projekto jaunesniuoju mokslo darbuotoju. Nuo 2021 m. yra Vilniaus universiteto jaunesnysis asistentas. Studijuodamas palaikė ryšius su verslo atstovais: 2014–2017 m. bendradarbiavo su Euromonitor International, o nuo 2017 m. užsiima konsultacine veikla duomenų mokslo bei rizikos valdymo klausimais.

Aprobacija

Disertacijos rezultatai pristatyti šiose mokslinėse konferencijose:

- *59-th conference of Lithuanian Mathematical Society*, Vilnius, Lietuva, 2018 m. birželio 18–19 d.

- *12th International Vilnius Conference on Probability Theory and Mathematical Statistics and 2018 IMS Annual Meeting on Probability and Statistics (IMS-2018 Vilnius)*, Vilnius, Lietuva, 2018 m. liepos 2–6 d.

- *20th IWH-CIREQ-GW Macroeconometric Workshop: Micro Data and Macro Questions*, organizuota IWH, Halė, Vokietija, 2019 m. spalio 29–30 d.

- *13th International Conference of the ERCIM WG on Computational and Methodological Statistics, 14th International Conference on Computational and Financial Econometrics, Virtual Conference (CFE-CMS2020)*, Londonas, Jungtinė Karalystė, 2020 m. gruodžio 19–21 d.

- *61-th conference of Lithuanian Mathematical Society*, Vilnius, Lietuva, 2020 m. gruodžio 4 d.

- *7th International Conference on Time Series and Forecasting* (ITISE 2021), Gran Kanarija, Ispanija, 2021 m. liepos 19–21 d.

- *22nd European Young Statisticians Meeting (EYSM 2021)*, Atėnai, Graikija, 2021 m. rugsėjo 6–10 d.

Publikacijos

Disertacijos rezultatai publikuoti šiuose straipsniuose:

1. S. Jokubaitis, D. Celov ir R. Leipus, Sparse structures with LASSO through principal components: Forecasting GDP components in the short-run. *International Journal of Forecasting* (2021), 37(2), 759–776, https://doi.org/10.1016/j.ijforecast.2020.09.005

2. S. Jokubaitis ir R. Leipus, Asymptotic Normality in Linear Regression with Approximately Sparse Structure. *Mathematics* (2022), 10, no. 10: 1657. https://doi.org/10.3390/math10101657

NOTES

Vilnius University Press
9 Saulėtekio Ave., Building III, LT-10222 Vilnius
E-mail: info@leidykla.vu.lt, www.leidykla.vu.lt
bookshop.vu.lt, journals.vu.lt
Print run 20