



# AALBORG UNIVERSITY

## DENMARK

AALBORG UNIVERSITET

P6-PROJEKT

## Makroøkonomiske prognoser og store datamængder

*Forfattere:*

Jesper H. JAKOBSEN  
Kristin JOHANNESSEN  
Louise N. CHRISTENSEN  
Trine GRAFF

*Vejleder:*

Johannes TANG KRISTENSEN

27. maj 2016



Institut for Matematiske Fag  
Fredrik Bajers Vej 7G  
DK-9220 Aalborg Ø

## AALBORG UNIVERSITET

### STUDENTERRAPPORT

**Titel:**

Makroøkonomiske prognoser og store data-mængder

**Tema:**

Prædiktation af vækst af realt BNP

**Projektperiode:**

Forårssemestret 2016

**Projektgruppe:**

G4-117

**Deltagere:**

Jesper Haar Jakobsen  
Kristin Johannesen  
Louise N. Christensen  
Trine Graff

**Vejledere:**

Johannes Tang Kristensen

**Oplagstal:** 7**Sidetal:** 85**Afleveringsdato:**

27. maj 2016

**Synopsis:**

Dette projekt omhandler prædiktation af væksten for USA's realt BNP et kvartal frem. For at kunne udføre prædiktationen introduceres først den nødvendige teori for lineære regression med henholdsvis tværsnitsdata og tidsrække data samt generelle begreber indenfor tidsrækkeanalyse. Heri beskrives OLS, ridge regression, lasso samt adaptive lasso, som sidst i rapporten anvendes til at prædiktere væksten af USA's realt BNP et kvartal frem. I teorien antages, at designmatricen har fuld rang, dette er dog ikke tilfældet, da datasættet indeholder langt flere forklarende variable end observationer. Derfor fravælges manuelt nogle forklarende variable, således at teorien kan anvendes. Vi anvender metoderne på dette datasæt og tilsidst anvendes lasso på det fulde datasæt, selvom dette går udover teorien. Herudfra kan vi konkludere, at lasso udvælger variable, som er bedre til prædiktation, end variablene som er udvalgt manuelt.

# SUMMARY

---

This project deals with linear regression for macroeconomic variables in a time series context. The focus is a setting where the number of explanatory variable is large compared to the number of observations, which complicate the use of ordinary least square estimation. However the theory is still presented under the assumption that the designmatrix has full rank, so the case, where the number of explanatory variables is bigger than the number of observation, is not covered in the theory.

In chapter 1 the many nice properties that the OLS estimator enjoys in the right setting are presented and an explanation why complications arises in the setting that we are considering is given. Therefore alternative fitting procedures for linear regression are considered. We introduce ridge regression, lasso and the adaptive lasso in a cross-sectional setting, and we prove that the adaptive lasso enjoys the oracle properties under certain assumptions.

Basic timeseries theory are introduced in chapter 2, as we move away from the cross-sectional setting and consider linear regression in a timeseries context in chapter 3. We consider the autoregressive model of order  $p$ , and show that the estimator estimated by ordinary least squares will be asymptotically normally distributed. Furthermore we argue for, that the assumptions made for the setting, in which the adaptive lasso enjoys the oracle properties, will be satisfied by an autoregressive model.

In chapter 4 we use the different fitting procedures presented in the theory to make one-step-ahead forecasts of the quarterly percentage growth in the gross domestic product. The data analysed in this chapter is quarterly data made publicly available by the Federal Reserve Bank of St. Louis. The simple autoregressive model with parameters estimated by OLS is fitted to the data at first, and used as a benchmark model throughout the chapter, in the sense that it is used as a reference for the performance of the other models. Since the presented theory covers the case with a designmatrix of full rank, a reduction of the available data is made, so that this assumption is reasonable. Linear models is then fitted to the reduced data by ridge regression, lasso and the adaptive lasso procedure, and an analysis of the residuals from the last fitted model is carried out in each case. The adaptive lasso is tried fitted to the data with weights chosen by both ordinary least squares and the lasso.

We also try reducing the data by the estimates given by the lasso used on the full dataset, in the sense that an appropriate penalty parameter is chosen, so that the number of parameters, the lasso estimates to be different from zero, will be less than the number of observations. A linear model containing the variables from this reduced dataset is then fitted to the data by the adaptive lasso. This procedure is not covered in the theory, it yields, however, superior results compared to the case, in which the variable selection was done manually.



# FORORD

---

Dette projekt er udarbejdet i foråret 2016 på sjette semester på bacheloruddannelsen i Matematik-Økonomi på Aalborg Universitet.

*Makroøkonomiske prognoser og store datamængder* er projektets overemne. Formålet med projektet er, at danne prognoser for en makrovariabel, i dette tilfælde BNP, vha. metoder der kan håndtere et stort antal prædiktorer.

Projektet er skrevet i L<sup>A</sup>T<sub>E</sub>X, og referencer er lavet ved at anvende B<sub>I</sub>B<sub>T</sub><sub>E</sub>X

Vi anbefaler, at projektet læses i den angivne rækkefølge, og den tiltænkte målgruppe for denne rapport forventes at have samme viden som en sjette semester, matematik-økonomi studerende.

Igennem projektet vil der være en firkant,  $\square$ , hvor der afsluttes et bevis.

Desuden vil der til sidst i projektet være en litteraturliste, tabelliste, figurliste, samt et appendiks, der vil indeholde forskellige definitioner, sætninger, grafer, variableerne i datasættet og R-koder.

Afslutningsvis vil vi rette en stor tak til vores vejleder Johannes Tang Kristensen for god vejledning igennem projektperioden.

---

Jesper H. Jakobsen  
<jhja12@student.aau.dk>

---

Kristin Johannesen  
<kjohan13@student.aau.dk>

---

Louise N. Christensen  
<lnch13@student.aau.dk>

---

Trine Graff  
<tgraff13@student.aau.dk>



# INDHOLD

---

<b>Summary</b>	<b>iii</b>
<b>Forord</b>	<b>v</b>
<b>1 Lineær regression med tværsnitsdata</b>	<b>3</b>
1.1 Mindste kvadraters metode . . . . .	3
1.2 Ridge regression . . . . .	7
1.3 Lasso . . . . .	8
1.4 Orakelegenskaber . . . . .	10
1.5 Adaptive lasso . . . . .	11
<b>2 Generel tidsrækkeanalyse</b>	<b>15</b>
2.1 Stationære tidsrækker . . . . .	16
2.2 Estimation af korrelation . . . . .	18
<b>3 Lineære regression med tidsrække data</b>	<b>21</b>
3.1 Autoregressiv model . . . . .	21
3.1.1 OLS estimatoren for en AR(1) model . . . . .	23
3.1.2 Den asymptotiske fordeling af OLS estimatoren for en AR( $p$ ) model . . . . .	24
3.1.3 ACF og PACF for en AR( $p$ ) model . . . . .	29
3.2 Residualanalyse . . . . .	31
3.3 Forecasting . . . . .	32
<b>4 Modellering</b>	<b>35</b>
4.1 Datasættet . . . . .	35
4.2 Benchmark model . . . . .	38
4.3 Forbedring af prædiktions . . . . .	41
4.4 Sammenligning . . . . .	48
4.5 Lasso variabel udvælgelse og adaptive lasso . . . . .	50
<b>5 Konklusion</b>	<b>53</b>
<b>A Generelle statistiske begreber</b>	<b>55</b>
<b>B Gennemsnitlige kvadrerede fejl</b>	<b>57</b>
<b>C Informationskriterier</b>	<b>59</b>
<b>D Data</b>	<b>61</b>
<b>E R-koder</b>	<b>69</b>
E.0.1 Benchmark prædiktions . . . . .	69
E.0.2 Prædiktions med AR( $p_{AL}$ ) . . . . .	71
E.0.3 Ridge prædiktions . . . . .	72
E.0.4 Lasso prædiktions . . . . .	72
E.0.5 Adaptive lasso med OLS vægte prædiktions . . . . .	73
E.0.6 Adaptive lasso med lasso vægte prædiktions . . . . .	74

F De 10 hyppigste variable for hver metode	77
--	----



# INDLEDNING

---

I dette projekt arbejder vi med makroøkonomiske variable, og ønsker herudfra at prædiktere væksten af USA's realt BNP et kvartal frem. Bruttonationalproduktet, som forkortes BNP, er et udtryk for størrelsen af et lands økonomi og velstand. Vækst af realt BNP er den procentvis-ændring i landets BNP justeret i forhold til inflation.

Udfordringen ved at prædiktere makroøkonomiske nøglevariable, som f.eks. BNP, kan være den store mængde af tilgængelig data. Ofte er antallet af potentielle variable meget større end antallet af tidsobservationer. Dette skaber problemer, hvis man ønsker at anvende den klassiske mindste kvadraters metode på en lineær regressionsmodel,  $y = X\beta + \epsilon$ . Problemet kan løses ved at udvælge en delmængde af variablene, som man mener har størst indflydelse på nøglevariablen og fravælge variable af mindre betydning. Til det formål kigger vi på lasso og adaptive lasso.

I første kapitel beskrives lineær regression med tværsnitsdata, hvor OLS, ridge regression, lasso samt adaptive lasso introduceres. Hertil opstilles nogle generelle overordnede antagelser. Disse benyttes blandt andet til at udlede den asymptotiske fordeling af OLS estimatoren og til at bevise at adaptive lasso opfylder såkaldte orakelegenskaber.

I næste kapitel introduceres nogle generelle begreber indenfor tidsrække analyse, som anvendes igennem rapporten.

Teorien for lineær regression med tidsrækkedata introduceres herefter med henblik på en autoregressiv model, hvor vi anvender de samme generelle antagelser, blot i forhold til tidsrækkedata, til at udlede den asymptotiske fordeling af OLS estimatoren for en AR(1) model. For den asymptotiske fordeling af OLS estimatoren for en AR( $p$ ) model vil vi gå mere i dybden med disse antagelser og give specifikke sætninger og en proposition som benyttes i dette bevis.

I sidste kapitel anvendes teorien til prædiktions af væksten af USA's realt BNP et kvartal frem. Datasættet, som vi arbejder med, er hentet fra Federal Reserve Bank of St. Louis, og er kvartalsdata. Det indeholder 227 observationer og 257 variable i perioden 1. marts 1959 til 1. september 2015, hvoraf USA's realt BNP er én af disse variable.

Vi vil ikke stå til regnskab for undladte beviser, hvor der henvises til andre kilder.



# 1 LINEÆR REGRESSION MED TVÆRSNITSDATA

---

En multipel lineær regressionsmodel skrives generelt på formen

$$y_i = \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \epsilon_i, \quad i = 1, \dots, n,$$

som for alle  $n$  observationer kan skrives på matrix-vektor form

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

hvor  $\mathbf{y} = (y_1, \dots, y_n)^T$  er responsvariablen,  $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_p]$  er designmatricen, hvor  $\mathbf{x}_j = (x_{1j}, \dots, x_{nj})^T$ ,  $j = 1, \dots, p$  er prædiktorer, og  $\boldsymbol{\beta}$  er en  $p \times 1$  vektor bestående af koefficienter, mens  $\boldsymbol{\epsilon}$  er en  $n \times 1$  vektor bestående af fejld. Følgende antagelser vil blive anvendt igennem kapitlet.

## Antagelser 1.1

- a) Designmatricen  $\mathbf{X}$  har fuld rang.
- b)  $E[\boldsymbol{\epsilon}|\mathbf{X}] = \mathbf{0}$
- c)  $\text{Var}[\boldsymbol{\epsilon}|\mathbf{X}] = \sigma^2 I_n$
- d) Stabilitet af prædiktorer:  $\frac{1}{n} \mathbf{X}^T \mathbf{X} \xrightarrow{p} E[\mathbf{X}^T \mathbf{X}] = \mathbf{Q}$ , hvor  $\mathbf{Q}$  eksisterer og er en positiv definit matrix.
- e) Ortogonalitets betingelse:  $\frac{1}{n} \mathbf{X}^T \boldsymbol{\epsilon} \xrightarrow{p} E[\mathbf{X}^T \boldsymbol{\epsilon}] = \mathbf{0}$ .
- f)  $\frac{1}{\sqrt{n}} \mathbf{X}^T \boldsymbol{\epsilon} \xrightarrow{d} \mathbf{W} = N(\mathbf{0}, \sigma^2 \mathbf{Q})$ .

Bemærk, at antagelse 1.1.e) er indeholdt i antagelse 1.1.f).

Antagelse 1.1.d) og 1.1.e) er en anvendelse af store tals lov, som siger at den empiriske middelværdi konvergerer i sandsynlighed mod den sande middelværdi, når  $n \rightarrow \infty$ . Antagelse 1.1.f) er en anvendelse af den centrale grænseværdi sætning, som løst sagt siger, at en stokastisk variabel under givne betingelser konvergerer i fordeling mod en normalfordeling, når  $n \rightarrow \infty$ .

Ovenstående antagelser er meget generelle, og vi vil i dette kapitel blot antage at de er opfyldte uden at redegøre for hvornår og i hvilken kontekst, de er gældende. Senere i rapporten i forbindelse med lineær regression med tidsrækkedata vil vi gå mere i dybden med dette.

## 1.1 Mindste kvadraters metode

*I denne sektion introduceres mindste kvadraters metode, også kaldet OLS, som er en simpel metode til at estimere de ukendte parametre i en lineær regressionsmodel. Til sidst udledes OLS estimatorens asymptotiske fordeling. Afsnittet er skrevet ud fra Christian Heij & Paul de Boer & Philip Hans Frases & Teun Kloek & Herman K. van Dijk (2004) s. 191-197.*

Intuitionen bag mindste kvadraters metode er at minimere summen af de kvadrerede residualer

(SSR). Residualerne er differencen mellem de observerede værdier og de fittede værdier. SSR er defineret ved

$$\text{SSR}(\beta_j) = \sum_{i=1}^n \left( y_i - \sum_{j=1}^p x_{ij}\beta_j \right)^2.$$

OLS estimatoren findes da udfra

$$\hat{\beta}^{\text{OLS}} = \min_{\beta} \left\{ \sum_{i=1}^n \left( y_i - \sum_{j=1}^p x_{ij}\beta_j \right)^2 \right\}.$$

### Sætning 1.1

Under antagelse 1.1.a) er løsningen til OLS er givet ved

$$\hat{\beta}^{\text{OLS}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}.$$

*Bevis.* SSR skrives på matrix-vektor form

$$\begin{aligned} \text{SSR}(\beta) &= (\mathbf{y} - \mathbf{X}\beta)^T (\mathbf{y} - \mathbf{X}\beta) \\ &= \mathbf{y}^T \mathbf{y} - \beta^T \mathbf{X}^T \mathbf{y} - \mathbf{y}^T \mathbf{X}\beta + \beta^T \mathbf{X}^T \mathbf{X}\beta \\ &= \mathbf{y}^T \mathbf{y} - 2\beta^T \mathbf{X}^T \mathbf{y} + \beta^T \mathbf{X}^T \mathbf{X}\beta. \end{aligned}$$

Vi differentierer  $\text{SSR}(\beta)$  og sætter dette udtryk lig 0

$$\frac{\partial \text{SSR}(\beta)}{\partial \beta} = -2\mathbf{X}^T \mathbf{y} + 2\mathbf{X}^T \mathbf{X}\beta.$$

OLS estimatet er dermed givet ved

$$\hat{\beta}^{\text{OLS}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}.$$

□

En vigtig og ønsket egenskab for OLS estimatoren er, at den er unbiased.

### Proposition 1.1

Under antagelse 1.1.a) og 1.1.b) er OLS estimatoren,  $\hat{\beta}^{\text{OLS}}$ , unbiased.

*Bevis.* Af loven om itererede forventning, korollar A.1, finder vi, at

$$\begin{aligned} E[\hat{\beta}^{\text{OLS}}] &= E[(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T (\mathbf{X}\beta + \epsilon)] \\ &= \beta + E[(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \epsilon] \\ &= \beta + E\left[E[(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \epsilon | \mathbf{X}]\right] \\ &= \beta + E[(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T E[\epsilon | \mathbf{X}]] \\ &= \beta. \end{aligned}$$

Af definition A.1 er  $\hat{\beta}^{\text{OLS}}$  unbiased.

□

**Proposition 1.2**

Under antagelse 1.1.a), 1.1.d) og 1.1.e) er OLS estimatoren,  $\hat{\beta}^{\text{OLS}}$ , svagt konsistent.

*Bevis.* OLS estimatoren kan omskrives til

$$\begin{aligned}\hat{\beta}^{\text{OLS}} &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \\ &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T (\mathbf{X} \beta + \epsilon) \\ &= \beta + (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \epsilon,\end{aligned}$$

som yderligere kan omskrives til

$$\hat{\beta}^{\text{OLS}} = \beta + \left( \frac{1}{n} \mathbf{X}^T \mathbf{X} \right)^{-1} \left( \frac{1}{n} \mathbf{X}^T \epsilon \right). \quad (1.1)$$

Af antagelse 1.1.d) og 1.1.e) har vi, at

$$\hat{\beta}^{\text{OLS}} \xrightarrow{p} \beta + \mathbf{Q}^{-1} \cdot \mathbf{0} = \beta.$$

Dermed er  $\hat{\beta}^{\text{OLS}}$  svagt konsistent ifølge definition A.3. □

Hvis datasættet, som der arbejdes med, består af mange observationer, er konsistente estimators ønskede, men dette er ikke tilstrækkeligt for statistisk inferens. For at kunne bestemme konfidensintervaller eller udføre hypotesetests, kræves en asymptotisk fordeling.

**Sætning 1.2: Asymptotisk fordeling af OLS estimatoren**

Under antagelse 1.1.a), 1.1.d) og 1.1.f) er den asymptotiske fordeling af  $\hat{\beta}^{\text{OLS}}$  givet ved

$$\sqrt{n}(\hat{\beta}^{\text{OLS}} - \beta) \xrightarrow{d} N(\mathbf{0}, \sigma^2 \mathbf{Q}^{-1}).$$

*Bevis.* For at udlede den asymptotiske fordeling af OLS estimatoren, bestemmer vi først grænsefordelingen. Hertil omskrives (1.1) til

$$\sqrt{n}(\hat{\beta}^{\text{OLS}} - \beta) = \left( \frac{1}{n} \mathbf{X}^T \mathbf{X} \right)^{-1} \left( \frac{1}{\sqrt{n}} \mathbf{X}^T \epsilon \right). \quad (1.2)$$

Af antagelse 1.1.d) ved vi, at første del i (1.2) konvergerer i sandsynlighed mod  $\mathbf{Q}^{-1}$ , mens antagelse 1.1.f) giver, at anden del konvergerer i fordeling mod  $N(\mathbf{0}, \sigma^2 \mathbf{Q})$ . Af Slutskys sætning A.1 finder vi, at grænsefordelingen for OLS estimatoren er

$$\sqrt{n}(\hat{\beta}^{\text{OLS}} - \beta) \xrightarrow{d} N(\mathbf{0}, \sigma^2 \mathbf{Q}^{-1} \mathbf{Q} \mathbf{Q}^{-1}) = N(\mathbf{0}, \sigma^2 \mathbf{Q}^{-1}).$$

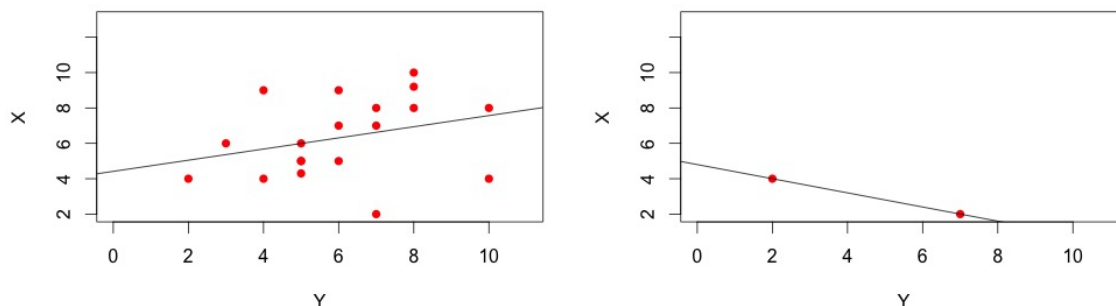
Heraf ses det at OLS estimatoren er rod- $n$ -konsistent. Den asymptotiske fordeling er dermed

$$\hat{\beta}^{\text{OLS}} \sim N[\beta, \sigma^2 n^{-1} \mathbf{Q}^{-1}].$$

□

Hvis antallet af forklarende variable  $p$  er approksimativt lig med antallet af observationer  $n$ , da bør OLS ikke anvendes. I dette tilfælde vil metoden estimere koefficienter, som stemmer fuldstændig overens med datasættet. Et eksempel er vist på figur 1.1 med  $p = 1$  (plus en skæring), hvor vi skelner mellem to tilfælde. I første tilfælde har vi 19 observationer, dvs.  $n >$

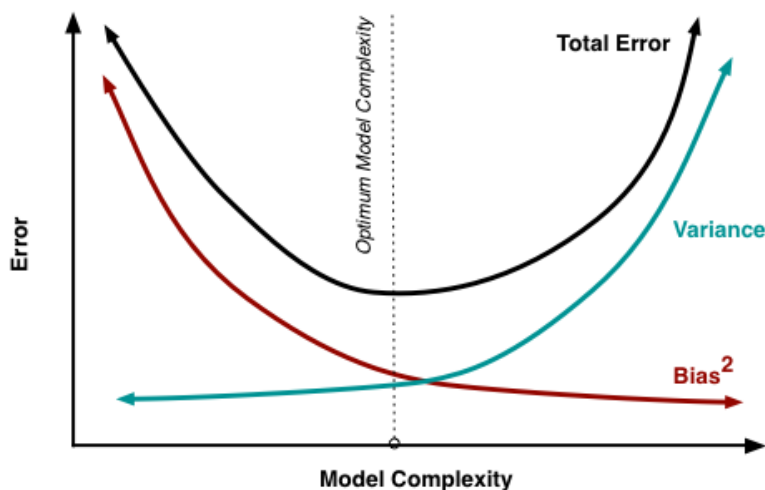
$p$ , og regressionslinjen approksimeres herpå. I andet tilfælde har vi 2 observationer, hvorved regressionslinjen vil fitte datasættet perfekt.



Figur 1.1: Til venstre illustreres OLS med  $n = 19$ , hvor  $\hat{\beta}_0 = 4.4173$  og  $\hat{\beta}_1 = 0.3153$  og til højre illustreres OLS med  $n = 2$ , hvor  $\hat{\beta}_0 = 4.8$  og  $\hat{\beta}_1 = -0.4$ .

En models prædiktionssevne af uobserverede værdier kan måles ved den gennemsnitlige kvadrerede fejl (MSE). I appendiks B, ses at MSE er en funktion af modellens bias og varians. For at minimere MSE skal vi derfor minimere varians og bias. Vi ønsker, at finde et tradeoff mellem disse, som giver den mindste MSE.

Dette tradeoff illustreres på figur 1.2. Heraf ses at en kompleks model med mange forklarende variable vil have en høj varians, men lav bias. Derfor ønsker vi, at mindske modellens kompleksitet, hvilket vil reducere variansen, dog på bekostning af en stigning i bias, men med henblik på at opnå en lavere forventet MSE.



Figur 1.2: Indflydelse af bias og varians på den totale fejl (Scott Fortmann-Roe, 2016).

Af proposition 1.1 ved vi, at OLS estimatoren er unbiased, forudsat at designmatricen har fuld rang. Hvis antallet af observationer er langt større end antallet af forklarende variable, da vil OLS estimaterne yderligere have lavere varians. Men hvis ikke designmatricen har fuld rang, da vil OLS ikke have en entydig løsning og variansen vil være uendelig.

Vi ser derfor på nye metoder, som kan forbedre dette tradeoff mellem bias og varians.

## 1.2 Ridge regression

Dette afsnit er skrevet ud fra s. 215-219 i Gareth James & Daniela Witten & Trevor Hastie & Robert Tibshirani (2013).

I dette afsnit introduceres ridge regression som blot er OLS underlagt en restriktion på parameter-værdierne. Ridge regression estimatoren findes ud fra

$$\min_{\beta} \left\{ \sum_{i=1}^n \left( y_i - \sum_{j=1}^p x_{ij} \beta_j \right)^2 \right\}, \quad \text{underlagt at } \sum_{j=1}^p \beta_j^2 \leq t^2, \quad (1.3)$$

som kan omskrives til et Lagrange problem

$$\hat{\beta}^R = \min_{\beta} \left\{ \sum_{i=1}^n \left( y_i - \sum_{j=1}^p x_{ij} \beta_j \right)^2 + \lambda \sum_{j=1}^p \beta_j^2 \right\}, \quad (1.4)$$

hvor  $\lambda \geq 0$  er en såkaldt strafparameter, som bestemmes separat. Der er en en-til-en korrespondance mellem det betingede problem (1.3) og Lagrange problemet (1.4). Første led i (1.4) svarer til OLS, som finder estimerede koefficienter ved at minimere SSR, mens værdien af andet led,  $\lambda \sum_j \beta_j^2$ , er lille når  $\beta_1, \dots, \beta_p$  er tæt på nul, og har dermed effekten at mindske estimerede koefficienter af  $\beta_j$  mod nul.

### Sætning 1.3

Under antagelse 1.1.a) er løsningen til ridge regression er givet ved

$$\hat{\beta}^R = (\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^T \mathbf{y}.$$

*Bevis.* Vi omskriver objektfunktionen af minimeringsproblemet (1.4) til matrix-vektor form

$$(\mathbf{y} - \mathbf{X}\beta)^T (\mathbf{y} - \mathbf{X}\beta) + \lambda \beta^T \beta.$$

Vi differentierer udtrykket og sætter dette lig 0

$$\frac{\partial (\mathbf{y} - \mathbf{X}\beta)^T (\mathbf{y} - \mathbf{X}\beta) + \lambda \beta^T \beta}{\partial \beta} = -2\mathbf{X}^T \mathbf{y} + 2\mathbf{X}^T \mathbf{X} \beta + 2\lambda \beta.$$

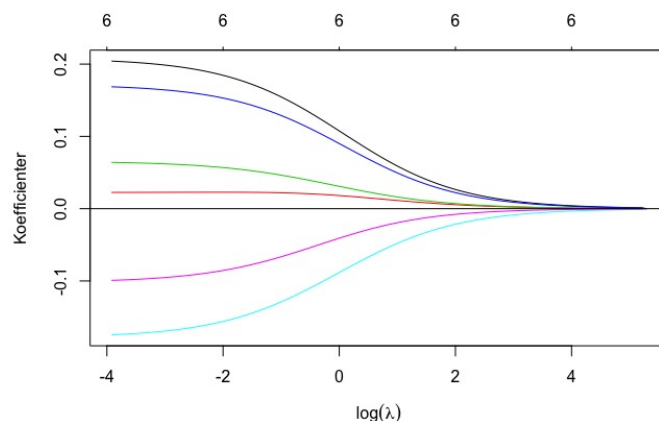
Ridge regression estimatet er dermed givet ved

$$\hat{\beta}^R = (\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^T \mathbf{y}.$$

□

Forskellen mellem  $\hat{\beta}^R$  og  $\hat{\beta}^{\text{OLS}}$ , er blot tilføjelsen af strafparametren  $\lambda$  på diagonalen af  $\mathbf{X}^T \mathbf{X}$ . Der gælder, at  $\hat{\beta}^R \rightarrow \hat{\beta}^{\text{OLS}}$  når  $\lambda \rightarrow 0$  og  $\hat{\beta}^R \rightarrow \mathbf{0}$  når  $\lambda \rightarrow \infty$ . I forhold til OLS, som blot finder én vektor med estimerede koefficienter, vil ridge regression give en vektor med estimerede koefficienter for hvert  $\lambda$ .

På figur 1.3 illustreres løsningsstien for ridge regression, hvor det ses, at de estimerede koefficienter konvergerer mod 0, når  $\lambda$  går mod uendelig.



Figur 1.3: Løsningssti af ridge regression med 6 simulerede variable.

Fordelen ved ridge regression i forhold til OLS ses i tradeoff mellem bias og varians. Når  $\lambda$  stiger, da vil fleksibiliteten af de estimerede parametre falde, hvilket medfører et fald i variansen, men en stigning i bias. Som nævnt har OLS estimererne, som svarer til ridge regression med  $\lambda = 0$ , høj varians, men ingen bias.

### 1.3 Lasso

I denne sektion introduceres metoden *Least Absolute Shrinkage and Selection Operator (lasso)*. Lasso udfører estimation og variabel udvælgelse samtidig. Afsnittet er baseret på s. 219-227 i Gareth James & Daniela Witten & Trevor Hastie & Robert Tibshirani (2013).

Straffeddets,  $\lambda \sum \beta_j^2$ , i (1.4) vil mindske alle koefficienter mod nul, men ridge regression vil aldrig sætte koefficienterne lig nul medmindre  $\lambda = \infty$ . Dermed inkluderer ridge regression alle  $p$  prædiktorer i den endelige model. Men ofte har nogle af de forklarende variable ingen indflydelse på responsvariablen, og vi ønsker derfor en metode som fjerner disse for at mindske kompleksiteten af modellen. Lasso vil generelt ikke inkludere alle prædiktorer i den endelige model. Lasso opfylder følgende

$$\min_{\beta} \left\{ \sum_{i=1}^n \left( y_i - \sum_{j=1}^p x_{ij} \beta_j \right)^2 \right\}, \text{ underlagt at } \sum_{j=1}^p |\beta_j| \leq t, \quad (1.5)$$

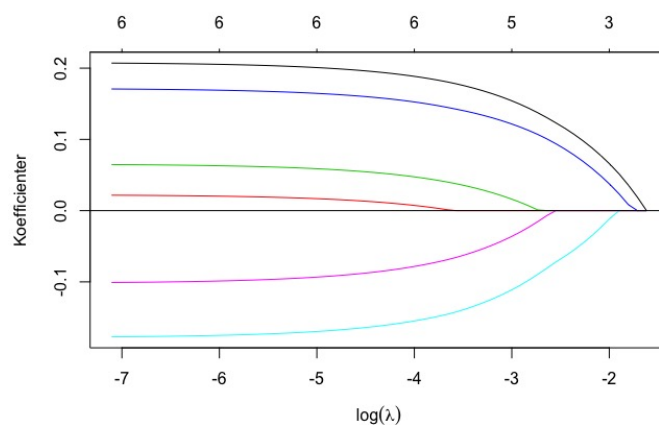
som også omskrives til et Lagrange problem

$$\hat{\beta}^L = \min_{\beta} \left\{ \sum_{i=1}^n \left( y_i - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^p |\beta_j| \right\}, \quad (1.6)$$

hvor  $\lambda \geq 0$ . Der er en en-til-en korrespondance mellem det betingede problem (1.5) og Lagrange problemet (1.6). Forskellen mellem ridge regression og lasso er, at  $\beta_j^2$  erstattes med  $|\beta_j|$  i straffeddets. Der gælder, at  $\hat{\beta}^L \rightarrow \hat{\beta}^{\text{OLS}}$ , når  $\lambda \rightarrow 0$ , men hvis  $\lambda$  er tilstrækkelig stor, da vil  $\hat{\beta}^L = \mathbf{0}$ .

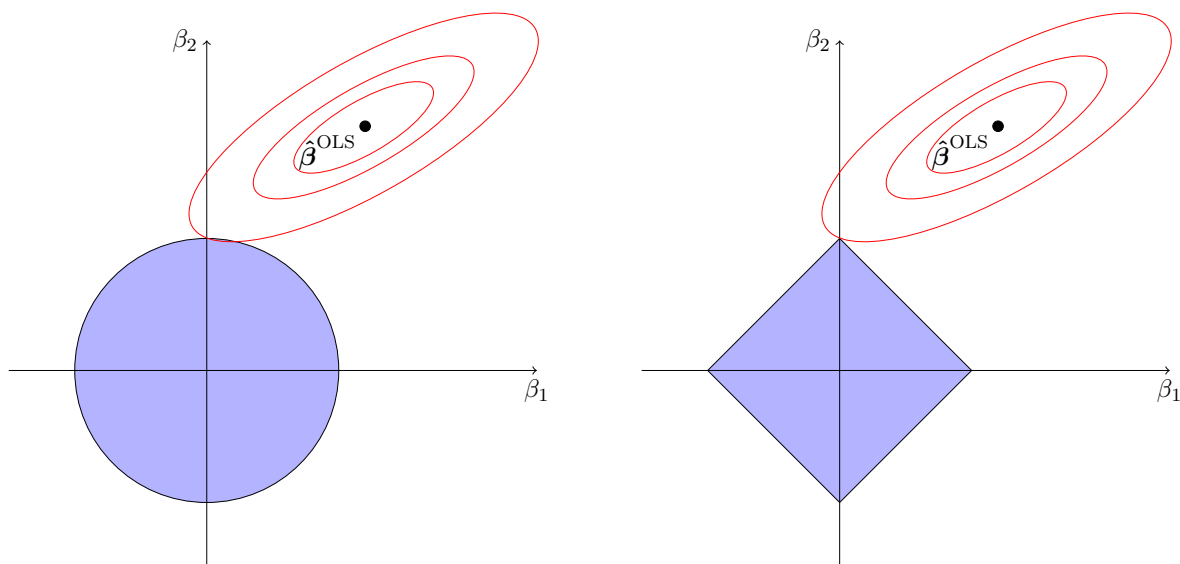
På figur 1.4 illustreres løsningsstien for lasso, hvor det ses, at de estimerede koefficienter ikke blot konvergerer mod, men er lig 0, når  $\lambda$  er tilstrækkelig stor. Dermed siges lasso at udføre variabel udvælgelse.





Figur 1.4: Løsningssti af lasso med 6 simulerede variable.

Variabel udvælgelsen for ridge regression og lasso illustreres på figur 1.5.



Figur 1.5: Konturer for SSR og betingelsesområderne for ridge regression (venstre) og lasso (højre). De blå arealer er betingelsesområderne  $|\beta_1| + |\beta_2| \leq t$  og  $\beta_1^2 + \beta_2^2 \leq t^2$ , mens de røde ellipser er konturkurver for SSR. Konturkurverne har centrum i OLS estimatoren,  $\hat{\beta}^{\text{OLS}}$ .

For  $p = 2$  er området hvori betingelsen for ridge regression er opfyldt givet ved cirklen  $\beta_1^2 + \beta_2^2 \leq t^2$ , mens betingelsesområdet for lasso er givet ved diamanten  $|\beta_1| + |\beta_2| \leq t$ . De røde ellipser omkring  $\hat{\beta}^{\text{OLS}}$  er konturkurverne for SSR, dvs. SSR er konstant i en given ellipse. Værdien af SSR stiger, som ellipsen udvides fra  $\hat{\beta}^{\text{OLS}}$ .

Ligningerne (1.3) og (1.5) indikerer at løsningen for lasso og ridge regression er givet ved det første punkt, hvor konturkurverne rammer betingelsesområdet. Siden ridge regression har et cirkulært betingelsesområde, vil skæringen med konturkurverne generelt ikke forekomme direkte på en akse. Modsat ridge regression har lassos betingelsesområde hjørner i hver akse, hvilket betyder, at hvis løsningen forekommer i et hjørne, da vil en af parameterne  $\beta_j$  være lig 0.

Hvis  $t$  er tilstrækkelig stor, så vil betingelsesområderne indeholde  $\hat{\beta}^{\text{OLS}}$  og derfor vil ridge regression og lasso estimaterne være ens med OLS estimaterne.

I figur 1.5 har vi blot betragtet det simple tilfælde hvor  $p = 2$ . Når  $p = 3$  da vil betingelsesområdet for ridge regression være en kugle, mens betingelsesområdet for lasso vil være en polydron.

## 1.4 Orakelegenskaber

Vi foretrækker en metode, som har orakelegenskaber. Lad  $\mathcal{A} = \{j : \beta_j^* \neq 0\}$ , hvor  $\beta_j^*$  betegner koefficienten af  $x_j$  i den sande model og antag at  $|\mathcal{A}| = p_0 < p$ . Så defineres orakelegenskaberne som følgende.

### Definition 1.1: Orakelegenskaber

En metode siges at have orakelegenskaberne, hvis den opfylder følgende:

- Variabel udvælgelsen er konsistent, dvs. for

$$\mathcal{A}_n = \{j : \hat{\beta}_j \neq 0\} \text{ og } \mathcal{A} = \{j : \beta_j^* \neq 0\},$$

gælder der, at  $\lim_{n \rightarrow \infty} P(\mathcal{A}_n = \mathcal{A}) = 1$ .

- Estimatoren er asymptotisk normalfordelt, dvs.

$$\sqrt{n}(\hat{\beta}_{\mathcal{A}} - \beta_{\mathcal{A}}^*) \xrightarrow{d} N(\mathbf{0}, \Sigma_I^*),$$

hvor  $\beta_{\mathcal{A}}^* = \{\beta_j^*, j \in \mathcal{A}\}$  og  $\Sigma_I^*$  er kovarians matricen, hvis vi kender den sande model.

Lad

$$\mathbf{Q} = \begin{bmatrix} \mathbf{Q}_{11} & \mathbf{Q}_{12} \\ \mathbf{Q}_{21} & \mathbf{Q}_{22} \end{bmatrix},$$

være en positiv definit matrix, hvor  $\mathbf{Q}_{11}$  er en  $p_0 \times p_0$  matrix. Så kan det udledes af proposition 1.3, at lasso ikke er konsistent i variabel udvælgelse i alle tilfælde og dermed ikke opfylder orakelegenskaberne.

### Proposition 1.3

Hvis  $\frac{\lambda_n}{\sqrt{n}} \rightarrow \lambda_0 \geq 0$ , da vil  $\hat{\beta}^L \xrightarrow{p} \arg \min Z$ , hvor

$$Z(\mathbf{u}) = (\mathbf{u} - \beta^*)^T \mathbf{Q}(\mathbf{u} - \beta^*) + \lambda_0 \sum_{j=1}^p |u_j|.$$

*Bevis.* Beviset undlades, men der henvises til s. 1358 i Keith Knight & Wenjiang Fu (2000).  $\square$

Hvis  $\lambda_0 = 0$ , da gælder ifølge proposition 1.3, at  $\hat{\beta}^L \xrightarrow{p} \hat{\beta}^*$ , da straffeddets forsvinder og  $\mathbf{Q}$  er positiv definit, og dermed er  $\hat{\beta}^L$  svagt konsistent. Men da straffeddets forsvinder, medfører det også, at lasso asymptotisk ingen variabel udvælgelse har. Hvis  $\lambda_0 > 0$  kan det ikke udledes fra proposition 1.3, om estimatoren er konsistent. Det angives i (Hui Zou, 2006), at dette kun vil være tilfældet, når en given betingelse er opfyldt. En metode, der konsekvent udfører konsistent variabeludvælgelse, kan dog opnås med en simpel tilføjelse til lasso estimatet. Denne metode kaldes adaptive lasso.

## 1.5 Adaptive lasso

Dette afsnit er baseret på s. 1419-1421 i (Hui Zou, 2006).

Adaptive lasso er en simpel udvidelse af lasso. Ideen bag adaptive lasso er, at i stedet for at straffe alle koefficienter i modellen ligeligt, udnyttes viden om data til at tildele koefficienterne individuelle straffævgte. Derved fås den vægtede lasso

$$\min_{\beta} \left\{ \sum_{i=1}^n \left( y_i - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^p w_j |\beta_j| \right\},$$

hvor  $\mathbf{w} \in \mathbb{R}^p$  er en kendt vektor og  $w_j \geq 0$ .

Adaptive lasso er blot en vægtet lasso, hvor vægtene  $\mathbf{w}$  er valgt således at metoden opfylder orakelegenskaberne. Lad  $\hat{\beta}$  være en rod- $n$ -konsistent estimator til  $\beta^*$ , tag  $\gamma > 0$  og definer vægtene  $\mathbf{w} = \frac{1}{|\hat{\beta}|^\gamma}$ , da er adaptive lasso estimatorne,  $\hat{\beta}^{\text{AL}}$ , givet ved

$$\hat{\beta}^{\text{AL}} = \min_{\beta} \left\{ \sum_{i=1}^n \left( y_i - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^p \frac{1}{|\hat{\beta}_j|^\gamma} |\beta_j| \right\}.$$

Den aktive mængde for adaptive lasso estimatet betegnes  $\mathcal{A}_n^{\text{AL}} = \{j : \hat{\beta}_j^{\text{AL}} \neq 0\}$ . OLS estimatoren,  $\hat{\beta}^{\text{OLS}}$ , samt lasso estimatoren,  $\hat{\beta}^{\text{L}}$ , er begge rod- $n$ -konsistent estimators og kan derfor anvendes til at bestemme vægtene  $\mathbf{w}$ .

### Sætning 1.4

Antag  $\frac{\lambda_n}{\sqrt{n}} \rightarrow 0$  og  $\lambda_n n^{\frac{\gamma-1}{2}} \rightarrow \infty$ . Under antagelse 1.1.d) og 1.1.f), må adaptive lasso estimatorne opfylde følgende:

- Udvælgelsen af variabler er konsistent:  $\lim_{n \rightarrow \infty} P(\mathcal{A}_n^{\text{AL}} = \mathcal{A}) = 1$
- Asymptotisk normalitet:  $\sqrt{n} (\hat{\beta}_{\mathcal{A}}^{\text{AL}} - \beta_{\mathcal{A}}^*) \xrightarrow{d} N(\mathbf{0}, \sigma^2 \mathbf{Q}_{11}^{-1})$ .

*Bevis.* Først bevises asymptotisk normalitet. Lad  $\beta = \beta^* + \frac{\mathbf{u}}{\sqrt{n}}$  og

$$\Psi_n(\mathbf{u}) = \left\| \mathbf{y} - \sum_{j=1}^p \mathbf{x}_j \left( \beta_j^* + \frac{u_j}{\sqrt{n}} \right) \right\|^2 + \lambda_n \sum_{j=1}^p \hat{w}_j \left| \beta_j^* + \frac{u_j}{\sqrt{n}} \right|.$$

Lad  $\hat{\mathbf{u}}^{(n)} = \arg \min \Psi_n(\mathbf{u})$ , da er  $\hat{\beta}^{\text{AL}} = \beta^* + \frac{\hat{\mathbf{u}}^{(n)}}{\sqrt{n}}$  eller  $\hat{\mathbf{u}}^{(n)} = \sqrt{n} (\hat{\beta}^{\text{AL}} - \beta^*)$ . Lad  $V(\mathbf{u})^{(n)} = \Psi_n(\mathbf{u}) - \Psi_n(\mathbf{0})$ , da gælder, at

$$V(\mathbf{u})^{(n)} = \left\| \mathbf{y} - \sum_{j=1}^p \mathbf{x}_j \left( \beta_j^* + \frac{u_j}{\sqrt{n}} \right) \right\|^2 + \lambda_n \sum_{j=1}^p \hat{w}_j \left| \beta_j^* + \frac{u_j}{\sqrt{n}} \right| - \left\| \mathbf{y} - \sum_{j=1}^p \mathbf{x}_j \beta_j^* \right\|^2 - \lambda_n \sum_{j=1}^p \hat{w}_j |\beta_j^*|.$$

Vi opdeler ligningen og ser først på leddene hvori strafparametrene indgår

$$\lambda_n \sum_{j=1}^p \hat{w}_j \left| \beta_j^* + \frac{u_j}{\sqrt{n}} \right| - \lambda_n \sum_{j=1}^p \hat{w}_j |\beta_j^*| = \lambda_n \sum_{j=1}^p \hat{w}_j \left( \left| \beta_j^* + \frac{u_j}{\sqrt{n}} \right| - |\beta_j^*| \right).$$

Vi ser herefter på de to resterende led

$$\left\| \mathbf{y} - \sum_{j=1}^p \mathbf{x}_j \left( \beta_j^* + \frac{u_j}{\sqrt{n}} \right) \right\|^2 - \left\| \mathbf{y} - \sum_{j=1}^p \mathbf{x}_j \beta_j^* \right\|^2,$$

som kan skrives på matrix-vektor form

$$\begin{aligned} \left( \mathbf{y} - \mathbf{X}\beta^* - \frac{\mathbf{X}\mathbf{u}}{\sqrt{n}} \right)^2 - (\mathbf{y} - \mathbf{X}\beta^*)^2 &= (\mathbf{y} - \mathbf{X}\beta^*)^2 + \left( \frac{\mathbf{X}\mathbf{u}}{\sqrt{n}} \right)^2 - 2(\mathbf{y} - \mathbf{X}\beta^*)^T \left( \frac{\mathbf{X}\mathbf{u}}{\sqrt{n}} \right) - (\mathbf{y} - \mathbf{X}\beta^*)^2 \\ &= \frac{\mathbf{u}^T (\mathbf{X}^T \mathbf{X}) \mathbf{u}}{n} - 2\epsilon^T \left( \frac{\mathbf{X}\mathbf{u}}{\sqrt{n}} \right) \\ &= \mathbf{u}^T \left( \frac{1}{n} \mathbf{X}^T \mathbf{X} \right) \mathbf{u} - 2 \frac{\epsilon^T \mathbf{X}}{\sqrt{n}} \mathbf{u}. \end{aligned}$$

Vi får så, at

$$\begin{aligned} V(\mathbf{u})^{(n)} &= \mathbf{u}^T \left( \frac{1}{n} \mathbf{X}^T \mathbf{X} \right) \mathbf{u} - 2 \frac{\epsilon^T \mathbf{X}}{\sqrt{n}} \mathbf{u} + \lambda_n \sum_{j=1}^p \hat{w}_j \left( \left| \beta_j^* + \frac{u_j}{\sqrt{n}} \right| - |\beta_j^*| \right) \\ &= \mathbf{u}^T \left( \frac{1}{n} \mathbf{X}^T \mathbf{X} \right) \mathbf{u} - 2 \frac{\epsilon^T \mathbf{X}}{\sqrt{n}} \mathbf{u} + \frac{\lambda_n}{\sqrt{n}} \sum_{j=1}^p \hat{w}_j \sqrt{n} \left( \left| \beta_j^* + \frac{u_j}{\sqrt{n}} \right| - |\beta_j^*| \right). \end{aligned} \quad (1.7)$$

Af antagelse 1.1.d) har vi, at der for første led i (1.7) gælder, at  $\frac{1}{n} \mathbf{X}^T \mathbf{X} \xrightarrow{p} \mathbf{Q}$ , mens det for andet led følger af antagelse 1.1.f), at  $\frac{\epsilon^T \mathbf{X}}{\sqrt{n}} \xrightarrow{d} \mathbf{W} = N(\mathbf{0}, \sigma^2 \mathbf{Q})$ . Derfor ser vi nu blot på sidste led i (1.7).

Hvis  $\beta_j^* \neq 0$ , da har vi, at  $\hat{w}_j \xrightarrow{p} |\beta_j^*|^{-\gamma}$ . Yderligere har vi, at

$$\lim_{n \rightarrow \infty} \frac{\left| \beta_j^* + \frac{u_j}{\sqrt{n}} \right| - |\beta_j^*|}{\frac{u_j}{\sqrt{n}}} = \frac{d}{d\beta_j^*} |\beta_j^*| = \text{sign}(\beta_j^*),$$

hvoraf der gælder, at  $\lim_{n \rightarrow \infty} \sqrt{n} \left( \left| \beta_j^* + \frac{u_j}{\sqrt{n}} \right| - |\beta_j^*| \right) = u_j \text{sign}(\beta_j^*)$ . Af Slutskys sætning A.1 har vi, at

$$\frac{\lambda_n}{\sqrt{n}} \hat{w}_j \sqrt{n} \left( \left| \beta_j^* + \frac{u_j}{\sqrt{n}} \right| - |\beta_j^*| \right) \xrightarrow{p} 0.$$

Hvis  $\beta_j^* = 0$ , da har vi  $\sqrt{n} \left( \left| \beta_j^* + \frac{u_j}{\sqrt{n}} \right| - |\beta_j^*| \right) = |u_j|$ . Da  $\gamma > 0$  kan vægtene omskrives til

$$\hat{w}_j = \left( \frac{1}{|\hat{\beta}_j|} \right)^\gamma = \left( \frac{\sqrt{n}}{\sqrt{n} |\hat{\beta}_j|} \right)^\gamma = \frac{n^{\gamma/2}}{(\sqrt{n} |\hat{\beta}_j|)^\gamma},$$

hvor  $\hat{\beta}_j$  er rod- $n$ -konsistent. Heraf har vi, at  $\frac{\lambda_n}{\sqrt{n}} \hat{w}_j = \frac{\lambda_n}{\sqrt{n}} \frac{n^{\gamma/2}}{(\sqrt{n} |\hat{\beta}_j|)^\gamma} = \lambda_n n^{\frac{\gamma-1}{2}} \frac{1}{|\sqrt{n} \hat{\beta}_j|^\gamma}$ , hvor  $\sqrt{n} \hat{\beta}_j = O_p(1)$  og vi ved at  $\lambda_n n^{\frac{\gamma-1}{2}} \rightarrow \infty$ , da har vi, at  $\frac{\lambda_n}{\sqrt{n}} \hat{w}_j |u_j| \rightarrow \infty$ . Af Slutsky sætning ser vi, at  $V^{(n)}(\mathbf{u}) \xrightarrow{d} V(\mathbf{u})$  for alle  $\mathbf{u}$ , hvor

$$V(\mathbf{u}) = \begin{cases} \mathbf{u}_{\mathcal{A}}^T \mathbf{Q}_{11} \mathbf{u}_{\mathcal{A}} - 2\mathbf{u}_{\mathcal{A}}^T \mathbf{W}_{\mathcal{A}} & \text{hvis } u_j = 0, \forall j \notin \mathcal{A}, \\ \infty & \text{hvis } \exists u_j \neq 0, j \notin \mathcal{A}. \end{cases}$$

Da funktionen  $V^{(n)}$  er konveks, og  $(\mathbf{Q}_{11}^{-1}\mathbf{W}_{\mathcal{A}}, 0)^T$  er et entydig minimum af  $V$ , følger det af Hui Zou (2006) at  $\arg \min V^{(n)} \rightarrow \arg \min V$ . Derfor får vi

$$\hat{\mathbf{u}}_{\mathcal{A}}^{(n)} \xrightarrow{d} \mathbf{Q}_{11}^{-1}\mathbf{W}_{\mathcal{A}} \quad \text{og} \quad \hat{\mathbf{u}}_{\mathcal{A}^c}^{(n)} \xrightarrow{d} \mathbf{0}. \quad (1.8)$$

Vi observerer, at  $\mathbf{W}_{\mathcal{A}} = N(\mathbf{0}, \sigma^2 \mathbf{Q}_{11})$ .

Herefter vil vi bevise at udvælgelsen af variable er konsistent. For alle  $j \in \mathcal{A}$ , giver den asymptotiske normalitet at  $\hat{\beta}_j^{\text{AL}} \xrightarrow{p} \beta_j^*$ , dvs.  $P(j \in \mathcal{A}_n^{\text{AL}}) \rightarrow 1$ . Derfor vil vi vise at  $\forall j' \notin \mathcal{A}$ , da vil  $P(j' \in \mathcal{A}_n^{\text{AL}}) \rightarrow 0$ .

Vi betragter  $j' \in \mathcal{A}_n^{\text{AL}}$  således at  $\hat{\beta}_{j'}^{\text{AL}} \neq 0$ . Af første ordens betingelser har vi, at

$$2\mathbf{x}_{j'}^T (\mathbf{y} - \mathbf{X}\hat{\beta}^{\text{AL}}) = \lambda_n \hat{w}_{j'} \left| \text{sign}(\hat{\beta}_{j'}^{\text{AL}}) \right|,$$

som er ækvivalent med

$$2 \frac{\mathbf{x}_{j'}^T (\mathbf{y} - \mathbf{X}\hat{\beta}^{\text{AL}})}{\sqrt{n}} = \frac{\lambda_n}{\sqrt{n}} \hat{w}_{j'}.$$

Vi fandt, at  $\frac{\lambda_n}{\sqrt{n}} \hat{w}_{j'} \xrightarrow{p} \infty$ . Vi har da, at

$$\begin{aligned} 2 \frac{\mathbf{x}_{j'}^T (\mathbf{y} - \mathbf{X}\hat{\beta}^{\text{AL}})}{\sqrt{n}} &= 2 \frac{\mathbf{x}_{j'}^T (\mathbf{X}\beta^* + \epsilon - \mathbf{X}\hat{\beta}^{\text{AL}})}{\sqrt{n}} \\ &= 2 \frac{\mathbf{x}_{j'}^T \mathbf{X} (\beta^* - \hat{\beta}^{\text{AL}})}{\sqrt{n}} + 2 \frac{\mathbf{x}_{j'}^T \epsilon}{\sqrt{n}} \\ &= 2 \frac{\mathbf{x}_{j'}^T \mathbf{X} \sqrt{n} (\beta^* - \hat{\beta}^{\text{AL}})}{n} + 2 \frac{\mathbf{x}_{j'}^T \epsilon}{\sqrt{n}}. \end{aligned}$$

Af (1.8) og Slutskys sætning A.1, ved vi at  $2 \frac{\mathbf{x}_{j'}^T \mathbf{X} \sqrt{n} (\beta^* - \hat{\beta}^{\text{AL}})}{n}$  konvergerer i fordeling mod en normalfordeling og  $2 \frac{\mathbf{x}_{j'}^T \epsilon}{\sqrt{n}} \xrightarrow{d} N(\mathbf{0}, 4\|\mathbf{x}_{j'}\|^2 \sigma^2)$ . Dvs.

$$P(j' \in \mathcal{A}_n^{\text{AL}}) \leq P\left(2\mathbf{x}_{j'}^T (\mathbf{y} - \mathbf{X}\hat{\beta}^{\text{AL}}) = \lambda_n \hat{w}_{j'}\right) \rightarrow 0.$$

□



## 2 GENEREL TIDSRÆKKEANALYSE

---

I dette afsnit introduceres generelle begreber indenfor tidsrækkeanalyse, som anvendes igennem rapporten. Afsnittet er baseret på s. 17-26 og s. 106 og s. 141 i Robert H. Shumway & David S. Stoffer (2011).

En tidsrække  $x_t$  er en følge af stokastiske variable, som observeres på forskellige tidspunkter  $t$ . Den statistiske model, som beskriver udviklingen i denne tidsrække, kaldes en stokastisk proces. De grundlæggende egenskaber ved en stokastisk proces defineres i dette kapitel.

### Definition 2.1: Middelfunktion

Middelværdifunktionen er defineret som

$$\mu_{xt} = E[x_t] = \int_{-\infty}^{\infty} x f_t(x) dx,$$

forudsat at denne eksisterer, hvor  $f_t$  er tæthedsfunktionen af  $x_t$ .

I beskrivelsen af en stokastisk variabel har det også interesse, hvor meget den afviger fra dens middelværdi, dette måles ved variansfunktionen.

### Definition 2.2: Variansfunktion

Variansfunktionen er defineret som

$$\text{Var}[x_t] = E[x_t - \mu_{xt}]^2.$$

For at måle afhængigheden mellem punkterne i en tidsrække introduceres autokovariansfunktionen og autokorelationsfunktionen.

Autokovariansfunktionen måler den lineære afhængighed mellem to værdier  $x_s$  og  $x_t$  i samme tidsrække, hvor  $s \neq t$ .

### Definition 2.3: Autokovariansfunktion

Autokovariansfunktion er givet ved

$$\gamma_x(s, t) = \text{Cov}[x_s, x_t] = E[(x_s - \mu_{xs})(x_t - \mu_{xt})],$$

for hele tal  $s, t \in \mathbb{Z}_+$ .

Der gælder, at  $\gamma_x(s, t) = \gamma_x(t, s)$  for alle  $s$  og  $t$ . Hvis  $\gamma_x(s, t) = 0$ , da er  $x_s$  og  $x_t$  ikke lineært relateret, men der kan stadig forekomme en afhængigheds struktur mellem dem. Hvis  $x_s$  og  $x_t$  følger en bivariat normalfordeling, da sikrer  $\gamma_x(s, t) = 0$  uafhængighed mellem disse. Hvis  $s = t$ , reduceres autokovariansfunktionen til variansfunktionen

$$\gamma_x(t, t) = E[(x_t - \mu_{xt})^2] = \text{Var}[x_t].$$

Ud fra værdien af autokovariansfunktionen er det svært at sige hvor stor afhængigheden er, da dette afhænger af hvilken enhed variablen måles i, derfor foretrækkes et mål som måler afhængigheden mellem -1 og 1, dertil anvendes autokorrelationsfunktionen.

**Definition 2.4: Autokorrelationsfunktion (ACF)**

Autokorrelationsfunktion er givet ved

$$\rho(s, t) = \frac{\text{Cov}[x_t, x_s]}{\sqrt{\text{Var}[x_t]\text{Var}[x_s]}} = \frac{\gamma(s, t)}{\sqrt{\gamma(s, s)\gamma(t, t)}},$$

for hele tal  $s, t \in \mathbb{Z}_+$ .

**Proposition 2.1**

Der gælder, at  $-1 \leq \rho(s, t) \leq 1$ .

*Bevis.* Vi lader  $Z = x_t + tx_s$ , hvor der gælder, at

$$\text{Var}[x_t + tx_s] = \text{Var}[x_t] + t^2(\text{Var}[x_s]) + 2t\text{Cov}[x_t, x_s] \geq 0.$$

Vi har en kvadratisk ulighed, som er større end eller lig med 0, hvilket betyder at vi enten må have én real rod eller to imaginære rødder, som medfører at diskriminanten må være mindre eller lig med 0. Dvs.

$$\begin{aligned} (2\text{Cov}[x_t, x_s])^2 - 4\text{Var}[x_t]\text{Var}[x_s] &\leq 0 \\ (2\text{Cov}[x_t, x_s])^2 &\leq 4\text{Var}[x_t]\text{Var}[x_s] \\ |2\text{Cov}[x_t, x_s]| &\leq 2\sqrt{\text{Var}[x_t]\text{Var}[x_s]} \\ -2\sqrt{\text{Var}[x_t]\text{Var}[x_s]} &\leq |2\text{Cov}[x_t, x_s]| \leq 2\sqrt{\text{Var}[x_t]\text{Var}[x_s]} \\ -1 &\leq \frac{\text{Cov}[x_t, x_s]}{\sqrt{\text{Var}[x_t]\text{Var}[x_s]}} \leq 1. \end{aligned}$$

□

## 2.1 Stationære tidsrækker

Stationaritet er en ønsket egenskab for en tidsrække. Der findes to typer stationaritet, streng- og svag stationaritet. Hvis en tidsrække har endelig varians, da gælder, at streng stationaritet medfører svag stationaritet.

**Definition 2.5: Streng stationaritet**

En tidsrække er streng stationær, hvis den statistiske fordeling af

$$\{x_{t_1}, x_{t_2}, \dots, x_{t_k}\}$$

er den samme som den statistiske fordeling af

$$\{x_{t_1+h}, x_{t_2+h}, \dots, x_{t_k+h}\}$$

for alle  $t_1 < t_2 < \dots < t_k$  og alle  $h \in \mathbb{Z}$ , dvs.

$$P\{x_{t_1} \leq c_1, \dots, x_{t_k} \leq c_k\} = P\{x_{t_1+h} \leq c_1, \dots, x_{t_k+h} \leq c_k\}$$

er den samme for alle  $h$ , hvor  $c_1, \dots, c_k$  er reelle tal.



**Definition 2.6: Svag stationaritet**

En tidsrække er svag stationær, hvis

- Middelværdifunktionen,  $\mu_t$ , er konstant og dermed uafhængig af tiden  $t$ .
- Autokovariansfunktionen,  $\gamma(s, t)$ , kun afhænger af  $s$  og  $t$  igennem deres differens  $|s - t|$ .

Når der nævnes begrebet stationaritet i rapporten, menes svagt stationaritet.

At tage første differensen af en tidsrække er en ofte anvendt metode til at fjerne en lineær trend og dermed opnå stationaritet. Dette skyldes, at en tidsrække i mange tilfælde kan ses som værende sammensat af to komponenter,  $x_t = \mu_t + y_t$ , hvor  $\mu_t = \beta_0 + \beta_1 t$  er et ikke-stationært komponent og  $y_t$  er et stationært komponent, hvor  $E[y_t] = 0$ . Hvis vi tager første differensen af denne tidsrække får vi, at

$$\nabla x_t = x_t - x_{t-1} = \beta_1 + y_t - y_{t-1} = \beta_1 + \nabla y_t,$$

som er stationær, da trenden er fjernet. Hvis  $\mu_t$  er et  $k$ 'te grads polynomium,  $\mu_t = \sum_{j=0}^k \beta_j t^j$ , da vil  $\nabla^k y_t$  være stationær.

Hvis variansen af data er en stigende funktion af tiden transformeres data for at stabilisere variansen. Lad

$$x_t = (1 + p_t)x_{t-1},$$

hvor  $p_t$  er den procentvise ændring af tidsrækken fra periode  $t - 1$  til  $t$ , som kan være negativ. Vi tager logaritmen på begge sider, og får

$$\begin{aligned}\log(x_t) &= \log(1 + p_t) + \log(x_{t-1}) \\ \nabla \log(x_t) &= \log(1 + p_t).\end{aligned}$$

Hvis den procentvise ændring  $p_t$  er relativ lille, da vil  $\log(1 + p_t) \approx p_t$  og dermed vil

$$\nabla \log(x_t) \approx p_t,$$

være en relativ stabil proces. Vi kalder  $\nabla \log(x_t)$  for vækstrate.

Da autokovariansfunktionen for en stationær tidsrække kun afhænger af differencen mellem to tidspunkter, gælder der

$$\gamma_x(t + h, t) = \text{Cov}[x_{t+h}, x_t] = \text{Cov}[x_h, x_0] = \gamma(h, 0),$$

hvor vi ser bort fra andet argument. Vi betegner derfor autokovariansfunktionen og ACF, som hhv.  $\gamma_x(h)$  og  $\rho_x(h)$  som defineres som følgende.

**Definition 2.7: Autokovariansfunktion af en stationær tidsrække**

Autokovariansfunktionen for en stationær tidsrække er givet ved

$$\gamma_x(h) = \text{Cov}[x_{t+h}, x_t] = E[(x_{t+h} - \mu_{x(t+h)})(x_t - \mu_{xt})].$$

**Definition 2.8: ACF af en stationær tidsrække**

Autokorrelationsfunktionen for en stationær tidsrække er givet ved

$$\rho_x(h) = \frac{\gamma_x(t+h, t)}{\sqrt{\gamma_x(t+h, t+h)\gamma_x(t, t)}} = \frac{\gamma_x(h)}{\gamma_x(0)}.$$

Den lineære afhængighed mellem  $x_{t+h}$  og  $x_t$  som ACF måler, tager ikke hensyn til om afhængigheden mellem variablene skyldes begge variables lineære afhængighed af mellemliggende variable. For at finde korrelationen mellem  $x_{t+h}$  og  $x_t$ , hvor den lineære afhængighed af  $\{x_{t+1}, \dots, x_{t+h-1}\}$  på hver er fjernet, benyttes den partielle autokorrelation,  $\phi_{hh}$ . Lad  $\hat{x}_{t+h}$  betegne den lineære OLS regression af  $x_{t+h}$  på  $\{x_{t+h-1}, x_{t+h-2}, \dots, x_{t+1}\}$ . På samme måde betegner  $\hat{x}_t$  regressionen af  $x_t$  på  $\{x_{t+1}, x_{t+2}, \dots, x_{t+h-1}\}$ . Bemærk, at  $(x_{t+h} - \hat{x}_{t+h})$  og  $(x_t - \hat{x}_t)$  er ukorreleret med  $\{x_{t+1}, \dots, x_{t+h-1}\}$ .

**Definition 2.9: Den partielle autokorrelationsfunktion (PACF)**

Den partielle autokorrelationsfunktion af en stationær tidsrække,  $x_t$ , betegnes  $\phi_{hh}$  og er for  $h = 1, 2, \dots$ , givet ved

$$\phi_{11} = \text{Corr}[x_{t+1}, x_t] = \rho(1),$$

og

$$\phi_{hh} = \text{Corr}[x_{t+h} - \hat{x}_{t+h}, x_t - \hat{x}_t], \quad h \geq 2.$$

Et eksempel på en stationær tidsrække er hvid støj. Hvid støj er en simpel stokastisk proces, da den er uafhængig af tiden.

**Definition 2.10: Hvid støj**

Lad  $w_t$  være en følge af ukorrelerede og identisk fordelte stokastiske variable med middelværdi 0 og varians  $\sigma_w^2$ . Den stokastiske proces  $w_t$  kaldes hvid støj.

**Definition 2.11: En lineær proces**

En lineær proces,  $x_t$ , defineres til at være en linarkombination af en hvid støj  $w_t$  og er givet ved

$$x_t = \mu_x + \sum_{j=-\infty}^{\infty} \psi_j w_{t-j}, \quad \text{hvor} \quad \sum_{j=-\infty}^{\infty} |\psi_j| < \infty. \quad (2.1)$$

## 2.2 Estimation af korrelation

For en given stikprøve defineres den empiriske middelværdi, den empiriske autokovariansfunktion og den empiriske autokorrelationsfunktion, hvor vi antager, at den observerede tidsrække er stationær.

**Definition 2.12: Den empiriske middelværdi**

Den empiriske middelværdi er givet ved

$$\bar{x} = \frac{1}{n} \sum_{t=1}^n x_t,$$

hvor  $n$  er antallet af observationer.

**Definition 2.13: Den empiriske autokovariansfunktion**

Den empiriske autokovariansfunktion er defineret som

$$\hat{\gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-h} (x_{t+h} - \bar{x})(x_t - \bar{x}),$$

hvor  $\hat{\gamma}_x(-h) = \hat{\gamma}_x(h)$  for  $h = 0, 1, \dots, n-1$ .

**Definition 2.14: Den empiriske ACF**

Den empiriske ACF er givet ved

$$\hat{\rho}(h) = \frac{\hat{\gamma}_x(h)}{\hat{\gamma}_x(0)}.$$



### 3 LINEÆRE REGRESSION MED TIDSRÆKKEDATA

---

Lad  $\mathbf{y} = (y_1, \dots, y_n)^T$  være en tidsrække og lad  $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_p]$  være designmatricen, hvor  $\mathbf{x}_j = (x_{1j}, \dots, x_{nj})^T$ ,  $j = 1, \dots, p$  er prædiktorer. Da elementerne i designmatricen  $\mathbf{X}$  afhænger af tiden, betegnes designmatricen  $\mathbf{X}_t$  for tidsrækkedata. For at modellere  $\mathbf{y}$  som en funktion af  $\mathbf{X}_t$  betragtes en lineær regressionsmodel

$$\mathbf{y} = \mathbf{X}_t \boldsymbol{\phi} + \mathbf{w},$$

hvor  $\boldsymbol{\phi}$  er en  $p \times 1$  vektor af parametre som skal estimeres og  $\mathbf{w}$  er en  $n \times 1$  vektor af fejled. Vi vil referere til følgende antagelser igennem kapitlet ved anvendelse.

#### Antagelser 3.1

- a) Designmatricen  $\mathbf{X}_t$  har fuld rang.
- b)  $E[\mathbf{w}|\mathbf{X}_t] = \mathbf{0}$
- c)  $\text{Var}[\mathbf{w}|\mathbf{X}_t] = \sigma^2 I_n$
- d) Stabilitet af prædiktorer:  $\frac{1}{n} \mathbf{X}_t^T \mathbf{X}_t \xrightarrow{p} E[\mathbf{X}_t^T \mathbf{X}_t] = \mathbf{Q}$ , hvor  $\mathbf{Q}$  eksisterer og er en positiv definit matrix.
- e) Ortogonalitets betingelse:  $\frac{1}{n} \mathbf{X}_t^T \mathbf{w} \xrightarrow{p} E[\mathbf{X}_t^T \mathbf{w}] = \mathbf{0}$ .
- f)  $\frac{1}{\sqrt{n}} \mathbf{X}_t^T \mathbf{w} \xrightarrow{d} \mathbf{W} = N(\mathbf{0}, \sigma^2 \mathbf{Q})$ .

Under disse antagelser, som er ækvivalent med antagelserne 1.1, gælder samme resultater for OLS estimererne og adaptive lasso estimererne, som vi viste i kapitel 1, for tidsrækkedata. Hidtil har vi blot antaget at antagelserne er gældende, men vi vil i dette kapitel bestemme, hvornår antagelserne er opfyldt og i hvilken kontekst for en  $AR(p)$ .

#### 3.1 Autoregressiv model

*I dette afsnit introduceres den generelle teori for autoregressive modeller. Afsnittet er baseret på s. 84-86 i Robert H. Shumway & David S. Stoffer (2011).*

En autoregressiv model er en lineær regressionsmodel, hvor designmatricen, som betegnes  $\mathbf{y}_{t-1}$  for en autoregressiv model, blot indeholder laggede værdier af tidsrækken. En autoregressiv model er dermed givet ved

$$y_t = \mathbf{y}_{t-1}^T \boldsymbol{\phi} + w_t,$$

hvor  $\mathbf{y}_{t-1} = (y_{t-1}, \dots, y_{t-p})$ . Modellen er simpel, da der ikke indgår andre tidsrækker i modellen.

**Definition 3.1: En autoregressiv model af orden  $p$** 

En autoregressiv model af orden  $p$ , som betegnes  $AR(p)$ , er givet ved

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + w_t, \quad t = p+1, \dots, n, \quad (3.1)$$

hvor  $y_t$  er stationær og  $\phi_1, \phi_2, \dots, \phi_p$  er konstanter ( $\phi_p \neq 0$ ) og  $w_t \sim \text{iid}(0, \sigma_w^2)$  hvis ikke andet er angivet.

Bemærk, at middelværdien af  $y_t$  i ligning (3.1) er nul. Hvis middelværdien,  $\mu_y$ , af  $y_t$  ikke er nul, da erstattes  $y_t$  med  $y_t - \mu_y$  i (3.1), og vi får

$$y_t - \mu_y = \phi_1 (y_{t-1} - \mu_y) + \phi_2 (y_{t-2} - \mu_y) + \dots + \phi_p (y_{t-p} - \mu_y) + w_t,$$

eller

$$y_t = \alpha + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + w_t,$$

hvor  $\alpha = \mu_y(1 - \phi_1 - \dots - \phi_p)$ .

**Definition 3.2**

Polynomiet for en autoregressive model er defineret ved

$$\phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p, \quad \phi_p \neq 0,$$

hvor  $z$  er et komplekst tal.

Herefter introduceres begrebet kausalitet.

**Definition 3.3: Kausal**

En  $AR(p)$  model siges at være kausal, hvis tidsrækken  $y_t$  kan skrives som en en-sidet lineær proces

$$y_t = \sum_{j=0}^{\infty} \psi_j w_{t-j},$$

hvor  $\sum_{j=0}^{\infty} |\psi_j| < \infty$ . Vi sætter  $\psi_0 = 1$ .

En  $AR(p)$  er kausal hvis og kun hvis

$$\phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p = 0 \quad \Rightarrow \quad |z| > 1,$$

dvs. at rødderne ligger uden for enhedscirklen.

Polynomiet for en  $AR(1)$  model,  $y_t = \phi y_{t-1} + w_t$ , er givet ved  $\phi(z) = 1 - \phi z$ , hvoraf vi finder roden  $z = \frac{1}{\phi}$ . For at en  $AR(1)$  model er kausal, skal der derfor gælde, at  $|\frac{1}{\phi}| > 1$ , dvs.  $|\phi| < 1$ . Da kausalitet medfører stationaritet, er dette netop også betingelsen for at en  $AR(1)$  model er stationær, som vises i nedenstående proposition.

**Proposition 3.1**

Hvis  $|\phi| < 1$ , da er en  $AR(1)$  model,  $y_t = \phi y_{t-1} + w_t$ , stationær.

*Bevis.* Hvis vi itererer baglæns  $k$  gange, da får vi

$$\begin{aligned} y_t &= \phi y_{t-1} + w_t \\ &= \phi(\phi y_{t-2} + w_{t-1}) + w_t \\ &= \phi^2 y_{t-2} + \phi w_{t-1} + w_t \\ &\vdots \\ &= \phi^k y_{t-k} + \sum_{j=0}^{k-1} \phi^j w_{t-j}. \end{aligned}$$

Hvis vi antager  $|\phi| < 1$ , da kan AR(1) modellen skrives som en lineær proces givet ved

$$y_t = \sum_{j=0}^{\infty} \phi^j w_{t-j},$$

da  $\phi^k y_{t-k} \rightarrow 0$ , når  $k \rightarrow \infty$ . Middelværdien af den autoregressive model af orden én er givet ved

$$E[y_t] = \sum_{j=0}^{\infty} \phi^j E[w_{t-j}] = 0,$$

og autokovariansfunktionen er givet ved

$$\begin{aligned} \gamma_y(h) &= \text{Cov}[y_{t+h}, y_t] = E \left[ \left( \sum_{j=0}^{\infty} \phi^j w_{t+h-j} \right) \left( \sum_{k=0}^{\infty} \phi^k w_{t-k} \right) \right] \\ &= E \left[ (w_{t+h} + \dots + \phi^h w_t + \phi^{h+1} w_{t-1} + \dots)(w_t + \phi w_{t-1} + \dots) \right] \\ &= \sigma_w^2 \sum_{j=0}^{\infty} \phi^{h+j} \phi^j = \sigma_w^2 \phi^h \sum_{j=0}^{\infty} \phi^{2j} = \frac{\sigma_w^2 \phi^h}{1 - \phi^2}, \quad h \geq 0. \end{aligned} \tag{3.2}$$

Da  $\gamma_y(h) = \gamma_y(-h)$  finder vi blot autokovariansfunktionen for  $h \geq 0$ . Dvs. hvis  $|\phi| < 1$ , da er AR(1) stationær ifølge definition 2.6.  $\square$

### 3.1.1 OLS estimatoren for en AR(1) model

OLS estimatoren af  $\phi$  for en AR(1) model,  $y_t = \phi y_{t-1} + w_t$ , er givet ved

$$\hat{\phi}^{\text{OLS}} = \frac{\sum_{t=2}^n y_{t-1} y_t}{\sum_{t=2}^n y_{t-1}^2}.$$

#### Proposition 3.2

Under antagelse 3.1.d) og 3.1.e) er OLS estimatoren for en AR(1) model,  $\hat{\phi}^{\text{OLS}}$ , svagt konsistent.

*Bevis.* Vi omskriver OLS estimatet til

$$\hat{\phi}^{\text{OLS}} = \phi + \frac{n^{-1} \sum_{t=2}^n y_{t-1} w_t}{n^{-1} \sum_{t=2}^n y_{t-1}^2}. \tag{3.3}$$

Af antagelse 3.1.d) har vi, at

$$\frac{1}{n} \sum_{t=2}^n y_{t-1}^2 \xrightarrow{p} Q, \tag{3.4}$$

for  $0 < Q < \infty$ , hvor  $Q = \gamma_y(0) = \frac{\sigma_w^2}{1-\phi^2}$ . Fra antagelse 3.1.e) ved vi, at

$$\frac{1}{n} \sum_{t=2}^n y_{t-1} w_t \xrightarrow{p} 0,$$

og dermed får vi, at

$$\hat{\phi}^{\text{OLS}} \xrightarrow{p} \phi + \frac{0}{Q} = \phi.$$

Af definition A.3 er OLS estimatoren for en AR(1) model svagt konsistent.  $\square$

### Proposition 3.3

Under antagelse 3.1.d) og 3.1.f) er den asymptotiske fordeling af OLS estimatet for en AR(1) model, givet ved

$$\sqrt{n}(\hat{\phi}^{\text{OLS}} - \phi) \xrightarrow{d} N(0, 1 - \phi^2).$$

*Bevis.* Først omskrives (3.3) til

$$\sqrt{n}(\hat{\phi}^{\text{OLS}} - \phi) = \frac{\frac{1}{\sqrt{n}} \sum_{t=2}^n y_{t-1} w_t}{\frac{1}{n} \sum_{t=2}^n y_{t-1}^2}.$$

Af antagelse 3.1.f) har vi, at

$$\frac{1}{\sqrt{n}} \sum_{t=2}^n y_{t-1} w_t \xrightarrow{d} N(0, \sigma_w^2 Q),$$

hvor  $Q = \gamma_y(0) = \frac{\sigma_w^2}{1-\phi^2}$ . Af antagelse 3.1.d) har vi, at

$$\frac{1}{n} \sum_{t=2}^n y_{t-1}^2 \xrightarrow{p} Q.$$

Vi finder da, at

$$\sqrt{n}(\hat{\phi}^{\text{OLS}} - \phi) \xrightarrow{d} N\left(0, \frac{\sigma_w^2 Q}{Q^2}\right) = N(0, 1 - \phi^2).$$

$\square$

### 3.1.2 Den asymptotiske fordeling af OLS estimatoren for en AR(p) model

*Dette afsnit er baseret på s. 533-536 i Robert H. Shumway & David S. Stoffer (2011).*

OLS estimatoren af  $\phi$  for en kausal AR(p),  $y_t = \mathbf{y}_{t-1}^T \phi + w_t$ , hvor  $\mathbf{y}_{t-1} = [y_{t-1}, \dots, y_{t-p}]^T$  og  $w_t \sim \text{iid}(0, \sigma_w^2)$ , er givet ved

$$\hat{\phi}^{\text{OLS}} = \left( \sum_{t=p+1}^n \mathbf{y}_{t-1} \mathbf{y}_{t-1}^T \right)^{-1} \sum_{t=p+1}^n \mathbf{y}_{t-1} y_t.$$

For at finde den asymptotiske fordeling erstattes den nedre grænse med én, og vi betragter istedet

$$\tilde{\phi} = \left( \sum_{t=1}^n \mathbf{y}_{t-1} \mathbf{y}_{t-1}^T \right)^{-1} \sum_{t=1}^n \mathbf{y}_{t-1} y_t, \quad (3.5)$$



hvilket ikke ændrer den asymptotiske fordeling, da estimerne er asymptotisk ækvivalente. Indledningsvis introduceres sætninger og en proposition, som anvendes til at udlede den asymptotiske fordeling for OLS estimatoren for en kausal  $AR(p)$ . Disse formuleres blot og bevises ikke.

### Sætning 3.1: Basic Approximation Theorem (BAT)

Lad  $\mathbf{x}_n$  for  $n = 1, 2, \dots$ , og  $\mathbf{y}_{mn}$  for  $m = 1, 2, \dots$ , være stokastiske  $p \times 1$  vektorer således at

1.  $\mathbf{y}_{mn} \xrightarrow{d} \mathbf{y}_m$  når  $n \rightarrow \infty$  for hvert  $m$ .
2.  $\mathbf{y}_m \xrightarrow{d} \mathbf{y}$  når  $m \rightarrow \infty$ .
3.  $\lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} P\{|\mathbf{x}_n - \mathbf{y}_{mn}| > \epsilon\} = 0$  for alle  $\epsilon > 0$ .

Da gælder, at  $\mathbf{x}_n \xrightarrow{d} \mathbf{y}$ .

*Bevis.* Beviset undlades, men der henvises til s. 514 i Robert H. Shumway & David S. Stoffer (2011).  $\square$

Sætning 3.1 giver muligheden for at benytte en approksimation  $\mathbf{y}_{mn}$  af en følge  $\mathbf{x}_n$  for at bevise konvergering i fordeling.

Grænseværdisætningen, som vi benytter, er formuleret for en  $M$ -afhængig tidsrække. En tidsrække er  $M$ -afhængig, hvis observationer med et tidsinterval større end  $M$  imellem sig er uafhængige, dvs. mængden af værdier  $y_s$ ,  $s \leq t$  er uafhængig af mængden af værdier  $y_s$ ,  $s \geq t + M + 1$ .

### Sætning 3.2: $M$ -afhængighed grænseværdisætning

Hvis  $y_t$  er en strengt stationær  $M$ -afhængig følge af stokastiske variable med middelværdi nul og autokovariansfunktion  $\gamma(\cdot)$ , og hvis

$$V_M = \sum_{u=-M}^M \gamma(u), \quad V_M \neq 0,$$

da vil

$$n^{-1/2} \bar{y}_n \xrightarrow{d} N(0, V_M),$$

hvor  $\bar{y}_n = n^{-1} \sum_{t=1}^n y_t$ .

*Bevis.* Beviset undlades, men der henvises til s. 516 i Robert H. Shumway & David S. Stoffer (2011).  $\square$

Den  $M$ -afhængige grænseværdisætning kan benyttes for en følge af stokastiske variable. OLS-estimatoren er dog en følge af stokastiske  $p \times 1$  vektorer, derfor benytter vi følgende sætning.

### Proposition 3.4

En følge af stokastiske vektorer er asymptotisk normalfordelt, dvs.,

$$\mathbf{y}_n \xrightarrow{d} N(\boldsymbol{\mu}_n, \boldsymbol{\Sigma}_n),$$

hvis og kun hvis

$$\mathbf{c}^T \mathbf{y}_n \xrightarrow{d} N(\mathbf{c}^T \boldsymbol{\mu}_n, \mathbf{c}^T \Sigma_n \mathbf{c}),$$

hvor  $\mathbf{c} \in \mathbb{R}^p$  og  $\Sigma_n$  er positiv definit.

I Robert H. Shumway & David S. Stoffer (2011) vises den asymptotiske fordeling for de empiriske autokovariansfunktioner, vi har dog kun brug for at de empiriske autokovariansfunktioner konvergerer til populations autokovariansfunktionerne.

### Sætning 3.3

Hvis  $y_t$  er en stationær lineær proces på formen (2.1), hvor fejllidene  $w_t$  er iid og  $E(w_t^4) = \eta\sigma_w^4 < \infty$ , da konvergerer de empiriske autokovariansfunktioner til populations autokovariansfunktionerne, dvs. for givet  $K$  gælder, at

$$\begin{pmatrix} \hat{\gamma}(0) \\ \hat{\gamma}(1) \\ \vdots \\ \hat{\gamma}(K) \end{pmatrix} \xrightarrow{p} \begin{pmatrix} \gamma(0) \\ \gamma(1) \\ \vdots \\ \gamma(K) \end{pmatrix}.$$

*Bevis.* Beviset undlades, men der henvises til s. 522 i Robert H. Shumway & David S. Stoffer (2011).  $\square$

Herefter er vi klar til at bestemme den asymptotiske fordeling af OLS estimatorerne for en  $AR(p)$ . I beviset refereres til ovenstående sætninger og proposition.

### Sætning 3.4: Asymptotiske fordeling af OLS estimatoren for en $AR(p)$ model

Lad  $y_t$  være en kausal  $AR(p)$  model, hvor fejllidene  $w_t$  er iid og  $E(w_t^4) = \eta\sigma_w^4$ . Da gælder, at

$$\sqrt{n}(\hat{\phi}^{\text{OLS}} - \phi) \xrightarrow{d} N(\mathbf{0}, \sigma_w^2 \mathbf{Q}_p^{-1}),$$

hvor  $\mathbf{Q}_p = \{\gamma(i-j)\}_{i,j=1}^p$  er en  $p \times p$  autokovariansmatrix for vektoren  $\mathbf{y}_{t-1} = [y_{t-1}, \dots, y_{t-p}]^T$ .

*Bevis.* Estimatoren (3.5) kan omskrives til

$$\begin{aligned} \tilde{\phi} &= \left( \sum_{t=1}^n \mathbf{y}_{t-1} \mathbf{y}_{t-1}^T \right)^{-1} \sum_{t=1}^n \mathbf{y}_{t-1} (\mathbf{y}_{t-1}^T \phi + w_t) \\ &= \phi + \left( \sum_{t=1}^n \mathbf{y}_{t-1} \mathbf{y}_{t-1}^T \right)^{-1} \sum_{t=1}^n \mathbf{y}_{t-1} w_t, \end{aligned}$$

således at

$$\begin{aligned} n^{1/2}(\tilde{\phi} - \phi) &= \left( n^{-1} \sum_{t=1}^n \mathbf{y}_{t-1} \mathbf{y}_{t-1}^T \right)^{-1} n^{-1/2} \sum_{t=1}^n \mathbf{y}_{t-1} w_t \\ &= \left( n^{-1} \sum_{t=1}^n \mathbf{y}_{t-1} \mathbf{y}_{t-1}^T \right)^{-1} n^{-1/2} \sum_{t=1}^n \mathbf{u}_t, \end{aligned} \tag{3.6}$$

hvor  $\mathbf{u}_t = \mathbf{y}_{t-1}w_t$ . Ifølge sætning 3.3 konvergerer de empiriske autokovariansfunktioner i sand-synlighed til populations autokovariansfunktionerne, derfor følger det, at

$$n^{-1} \sum_{t=1}^n \mathbf{y}_{t-1} \mathbf{y}_{t-1}^T \xrightarrow{p} E[\mathbf{y}_{t-1} \mathbf{y}_{t-1}^T] = \mathbf{Q}_p.$$

For en kontinuert funktion  $g(\mathbf{x})$  gælder der, at  $g(\mathbf{x}_n) \xrightarrow{p} g(\mathbf{x})$  hvis  $\mathbf{x}_n \xrightarrow{p} \mathbf{x}$ . Den inverse matrix funktion er kontinuert, derfor gælder for første del af (3.6), at

$$\left( n^{-1} \sum_{t=1}^n \mathbf{y}_{t-1} \mathbf{y}_{t-1}^T \right)^{-1} \xrightarrow{p} \mathbf{Q}_p^{-1}.$$

Vi mangler derfor at bevise at sidste del af (3.6),  $n^{-1/2} \sum_{t=1}^n \mathbf{u}_t$ , konvergerer i fordeling til en normalfordeling, hvorefter vi kan benytte Slutskys sætning til at bestemme den asymptotiske fordeling af (3.6).

For at kunne benytte den  $M$ -afhængige grænseværdisætning approksimeres den kausale  $y_t$  ved

$$y_t^m = \sum_{j=0}^m \psi_j w_{t-j},$$

og processen  $\mathbf{u}_t^m = w_t(y_{t-1}^m, y_{t-2}^m, \dots, y_{t-p}^m)^T$  defineres. Lad

$$x_{nm} = n^{-1/2} \sum_{t=1}^n \boldsymbol{\lambda}^T \mathbf{u}_t^m,$$

hvor  $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_p)^T$  er vilkårlige vektorer, være en approksimation af

$$S_n = n^{-1/2} \sum_{t=1}^n \boldsymbol{\lambda}^T \mathbf{u}_t.$$

Ifølge proposition 3.4 medfører en konvergering i fordeling af  $S_n$  en konvergering i fordeling af  $n^{-1/2} \sum_{t=1}^n \mathbf{u}_t$ . Derfor findes den asymptotiske fordeling af  $S_n$  vha. sætning 3.1.

Det ses umiddelbart, at  $\mathbf{u}_t^m$  er en  $(m+p)$ -afhængig proces, da fejllenede der indgår i processen spænder fra  $w_t$  til  $w_{t-p-m}$  og  $w_t$  er iid, så  $\mathbf{u}_{t+m+p+1}^m$  er uafhængig af  $\mathbf{u}_t^m$ . Hvis  $\boldsymbol{\lambda}^T \mathbf{u}_t^m$  også er en stationær proces kan vi benytte den  $M$ -afhængige grænseværdisætning. Middelværdien og kovariansmatricen af  $\mathbf{u}_t$  udregnes ved at benytte at  $w_t$  og  $\mathbf{y}_{t-1}$  er uafhængige, da  $w_t$  er iid og  $y_t$  er en kausal proces. Det medfører, at  $E[\mathbf{u}_t] = E[\mathbf{y}_{t-1}]E[w_t] = \mathbf{0}$ , da fejllenedes middelværdier er nul, og at

$$E[\mathbf{u}_t \mathbf{u}_t^T] = E[\mathbf{y}_{t-1} w_t w_t \mathbf{y}_{t-1}^T] = E[\mathbf{y}_{t-1} \mathbf{y}_{t-1}^T] E[w_t^2] = \sigma_w^2 \mathbf{Q}_p = \Gamma_p.$$

Derudover gælder der for  $h > 0$ , at

$$E[\mathbf{u}_{t+h} \mathbf{u}_t^T] = E[\mathbf{y}_{t+h-1} w_{t+h} w_t \mathbf{y}_{t-1}^T] = E[\mathbf{y}_{t+h-1} w_t \mathbf{y}_{t-1}^T] E[w_{t+h}] = 0.$$

En lignende udregning er gældende for  $h < 0$ . Herudfra ses det at  $\boldsymbol{\lambda}^T \mathbf{u}_t$  er en stationær proces. En lignende udledning kan laves for  $\boldsymbol{\lambda}^T \mathbf{u}_t^m$  ved igen at benytte, at den nuværende fejl er ukorreleret med de tidligere fejl.

Vi beviser nu, at  $x_{nm}$  og  $S_n$  opfylder de tre punkter fra sætning 3.1. Ved at benytte den  $M$ -afhængige grænseværdisætning på  $x_{nm}$ , når  $n \rightarrow \infty$  og  $m$  holdes fast, får vi, at  $x_{nm} \xrightarrow{d} x_m$ , hvor  $x_m$  er asymptotisk normalfordelt med middelværdi nul og kovariansmatrix  $\boldsymbol{\lambda}^T \Gamma_p^{(m)} \boldsymbol{\lambda}$ , hvor

$\Gamma_p^{(m)}$  er autokovariansen for  $\mathbf{u}_t^m$ . Dermed opfylder  $x_{nm}$  punkt et. Derudover vil  $\Gamma_p^{(m)} \rightarrow \Gamma_p$ , når  $m \rightarrow \infty$ , så  $x_m \xrightarrow{d} x$ , hvor  $x$  er en stokastisk normalfordelt variabel med middelværdi nul og autokovarians  $\boldsymbol{\lambda}^T \Gamma_p \boldsymbol{\lambda}$ . Derved er andet punkt i sætning 3.1 opfyldt. For at påvise at punkt tre er opfyldt, benytter vi at

$$E[(S_n - x_{nm})^2] = n^{-1} \sum_{t=1}^n \boldsymbol{\lambda}^T E[(\mathbf{u}_t - \mathbf{u}_t^m)(\mathbf{u}_t - \mathbf{u}_t^m)^T] \boldsymbol{\lambda},$$

konvergerer til nul når  $n, m \rightarrow \infty$ . Dette kan ses udfra komponenterne af  $\mathbf{u}_t - \mathbf{u}_t^m$ , som udgøres af

$$y_t - y_t^m = \sum_{j=m+1}^{\infty} \psi_j w_{t-j},$$

som konvergerer mod nul når  $m \rightarrow \infty$ . De tre punkter fra sætning 3.1 er altså opfyldt af  $x_{nm}$  og  $S_n$  og det følger deraf at

$$S_n \xrightarrow{d} N(0, \boldsymbol{\lambda}^T \Gamma_p \boldsymbol{\lambda}).$$

Deraf følger det fra proposition 3.4, at der for sidste del af (3.6) gælder, at

$$n^{-1/2} \sum_{t=1}^n \mathbf{u}_t \xrightarrow{d} N(\mathbf{0}, \sigma_w^2 \mathbf{Q}_p).$$

Da  $\tilde{\phi}$  og  $\hat{\phi}^{\text{OLS}}$  er asymptotisk ækvivalente får vi fra Slutskys sætning A.1, at

$$\sqrt{n}(\hat{\phi}^{\text{OLS}} - \phi) \xrightarrow{d} N(\mathbf{0}, \sigma^2 \mathbf{Q}_p^{-1} \mathbf{Q}_p \mathbf{Q}_p^{-1}).$$

□

I beviset for sætning 1.4, anvendes antagelse 1.1.d) og 1.1.f)

$$\frac{1}{n} \mathbf{X}^T \mathbf{X} \xrightarrow{p} E[\mathbf{X}^T \mathbf{X}] = \mathbf{Q}$$

og

$$\frac{1}{\sqrt{n}} \mathbf{X}^T \boldsymbol{\epsilon} \xrightarrow{d} \mathbf{W} = N(\mathbf{0}, \sigma^2 \mathbf{Q}),$$

til at bevise at adaptive lasso har orakelegenskaberne. Vi viste i beviset for sætning 3.4, vha. sætning 3.2 og 3.3, at disse antagelser er opfyldt for en autoregressiv model. Et lignende bevis for at adaptive lasso benyttet på en  $\text{AR}(p)$  model opfylder orakelegenskaberne kan derfor konstrueres. Et sådant kan findes på s. 4-7 i (Anders Bredahl Kock, 2015). Adaptive lasso kan derfor benyttes til at vælge relevante lags og estimere parametrene i en autoregressiv model.

I anvendelsen vil vi også inkludere andre forklarende variable end blot tidligere lags. Vi beviser ikke, at adaptive lasso vil besidde orakelegenskaberne i en sådan situation, det ville kræve andre versioner af grænseværdisætningen og store tals lov, hvorudfra antagelse 1.1.d) og 1.1.f) kan vises. Inklusionen af tidsafhængige forklarende variable, der ikke er tidligere lags, er meget lig tværssnitsdata og tilføjer ikke samme afhængighed, mellem responsvariablen og forklarende variable, som tidligere lags. Så når orakelegenskaberne er opfyldt for en  $\text{AR}(p)$ , er vi fortrøstningsfulde for, at inklusionen af yderligere variable er rimelig.

### 3.1.3 ACF og PACF for en AR( $p$ ) model

Dette afsnit er skrevet udfra s. 103-104 og s. 106-107 i Robert H. Shumway & David S. Stoffer (2011).

For at udregne ACF for en kausal AR( $p$ ) model

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + w_t,$$

multipliseres begge sider af modellen med  $y_{t-h}$ , hvorefter den forventede værdi af begge sider tages

$$E[y_{t-h}y_t] = \phi_1 E[y_{t-h}y_{t-1}] + \phi_2 E[y_{t-h}y_{t-2}] + \dots + \phi_p E[y_{t-h}y_{t-p}] + E[y_{t-h}w_t].$$

Da  $E[y_t] = 0$  og da  $y_t$  er kausal følger det af definition 3.3 at, for  $h > 0$

$$E(w_t y_{t-h}) = E\left(w_t \sum_{j=0}^{\infty} \psi_j w_{t-h-j}\right) = 0.$$

Dermed er autokovariansfunktionen givet ved

$$\gamma(h) = \phi_1 \gamma(h-1) + \phi_2 \gamma(h-2) + \dots + \phi_p \gamma(h-p), \quad h = 1, 2, \dots$$

Vi dividerer igennem med  $\gamma(0)$  og trækker venstresiden fra højresiden, og finder den homogene differensligning for tidsrækkens ACF

$$\rho(h) - \phi_1 \rho(h-1) + \phi_2 \rho(h-2) + \dots + \phi_p \rho(h-p) = 0, \quad h \geq p.$$

Lad  $z_1, \dots, z_r$  være rødderne af det tilhørende autoregressive polynomie  $\phi(z)$ , hver med multiplicitet  $m_1, \dots, m_r$ , hvor  $m_1 + \dots + m_r = p$ . Da er den generelle løsning af differensligningen givet ved

$$\rho(h) = z_1^{-h} P_1(h) + z_2^{-h} P_2(h) + \dots + z_r^{-h} P_r(h), \quad h \geq p,$$

hvor  $P_j(h)$  er et polynomie af grad  $m_j - 1$ . Da modellen er kausal ligger alle rødderne  $z_1, \dots, z_r$  udenfor enhedscirklen. Derfor vil ACF for en AR( $p$ ) model aftage eksponentielt mod 0, når  $h \rightarrow \infty$ .

Vi kan ikke bestemme ordenen af den autoregressive proces udfra ACF, da processen laver regression på tidligere værdier af tidsrækken, så for to tidspunkter i tidsrækken,  $y_t$  og  $y_{t+h}$ , vil der altid være en vis afhængighed som skyldes den lineære afhængighed for begge tidspunkter af  $\{y_{t+1}, \dots, y_{t+h-1}\}$ . Derfor benytter vi PACF til at bestemme ordenen, da den måler korrelationen mellem to tidspunkter, mens den tager højde for de mellemliggende variable.

PACF for en AR( $p$ ) proces vil være 0, for lags der er højere end  $p$ . Dette skyldes at den lineære regression af  $y_{t+h}$  på  $\{y_{t+1}, \dots, y_{t+h-1}\}$  for en AR( $p$ ) er givet ved

$$\hat{y}_{t+h} = \sum_{j=1}^p \phi_j y_{t+h-j}.$$

Da en AR( $p$ ) model er givet ved  $y_{t+h} = \sum_{j=1}^p \phi_j y_{t+h-j} + w_{t+h}$  bliver PACF for  $h > p$  dermed

$$\phi_{hh} = \text{Corr}(y_{t+h} - \hat{y}_{t+h}, y_t - \hat{y}_t) = \text{Corr}(w_{t+h}, y_t - \hat{y}_t) = 0.$$

Den sidste lighed fås, da modellen er kausal, så  $y_t - \hat{y}_t$ , hvor  $\hat{y}_t$  er regression af  $y_t$  på  $\{y_{t+1}, \dots, y_{t+h-1}\}$ , afhænger kun af mængden af tidligere fejl  $\{w_{t+h-1}, w_{t+h-2}, \dots\}$ . For  $h \leq p$  vil  $\phi_{pp}$  være forskellig

fra 0, dette illustrerer vi ved at benytte en AR(2) model som eksempel.  
For en AR(2) model er

$$\begin{aligned}y_{t+2} &= \phi_1 y_{t+1} + \phi_2 y_t + w_{t+2}, \\ \hat{y}_{t+2} &= \phi_1 y_{t+1}.\end{aligned}$$

Derved bliver

$$\phi_{22} = \text{Corr}(y_{t+2} - \hat{y}_{t+2}, y_t - \hat{y}_t) = \text{Corr}(\phi_2 y_t + w_{t+2}, y_t - \phi_1 y_{t+1}) \neq 0.$$

## 3.2 Residualanalyse

Vi har antaget at modellens residualer er uafhængige. Denne antagelse kan bekræftes, hvis residualerne er normalfordelte og autokorrelerede. Derfor introduceres to teste, som tester for normalitet og autokorrelation.

Evidensen imod nulhypotesen  $\mathcal{H}_0$  måles ved hjælp af  $p$ -værdien. En lav  $p$ -værdi giver et stærkt bevis imod  $\mathcal{H}_0$ , og vi afviser  $\mathcal{H}_0$ , hvis  $p \leq 0.05$ . Hvis  $p > 0.05$  accepteres  $\mathcal{H}_0$ . Dette betyder ikke, at  $\mathcal{H}_0$  er sand, men blot at den ikke afvises.

### Jarque-Bera test

*Dette afsnit er skrevet ud fra Poul Newbold and Thorne (2016).*

Jarque-Bera test er en statistisk teststørrelse, som kan anvendes til at bestemme om residualerne er normalfordelte. Den tester om residualernes skewness og kurtosis svarer til normalfordelingens.

#### Definition 3.4: Jarque-Bera test (JB)

Betragt  $\mathcal{H}_0$  : er normalfordelt imod  $\mathcal{H}_A$  : ikke normalfordelt. Jarque Bera statistisk er givet ved

$$JB = \frac{n - k + 1}{6} \left( S^2 + \frac{1}{4} (C - 3)^2 \right),$$

hvor  $n$  er antallet af observationer,  $S$  er den empiriske skewness,  $C$  er den empiriske kurtosis og  $p$  er antallet af prædiktorer.

Under  $\mathcal{H}_0$  kan det vises, at teststørrelsen er asymptotisk  $\chi^2(2)$ -fordelt, hvor 2 angiver antallet af frihedsgrader. Store værdier af JB er derfor kritiske for nulhypotesen, og nulhypotesen afvises hvis  $JB > 5.991465$ , da kun fem procent af data ligger over denne værdi i en  $\chi^2(2)$ -fordeling.

### Ljung Box test

*Dette afsnit er skrevet ud fra s. 150 i Robert H. Shumway & David S. Stoffer (2011).*

Ljung-Box er en statistisk teststørrelse, som kan anvendes til at bestemme om der er autokorrelation i residualerne.

#### Definition 3.5: Ljung Box test (LB)

Betragt  $\mathcal{H}_0$  :  $\rho(1) = \rho(2) = \dots = \rho(H) = 0$ , imod  $\mathcal{H}_A$  : der forekommer autokorrelation i de første  $H$  lags. Ljung Box testen er givet ved

$$LB(H) = n \sum_{h=1}^H \frac{n+2}{n-h} \hat{\rho}^2(h), \quad (3.7)$$

hvor  $n$  er antallet af observationer og  $\hat{\rho}^2(h)$  er sample autokorrelationerne af de standardiserede residualer.

Antallet af lags  $H$  i (3.7) vælges vilkårligt, men typisk  $H = 20$ .

Under  $\mathcal{H}_0$  følger  $LB(H)$  asymptotisk en  $\chi^2(H - p)$ -fordeling, hvor  $p$  er antallet af parametre estimeret i modellen, der fittes til data. Store værdier af  $LB(H)$  er derfor kritiske for nulhypotesen.

### 3.3 Forecasting

*Dette afsnit er baseret på s. 108-109 i (Robert H. Shumway & David S. Stoffer, 2011) og Diebold-Mariano testen, som introduceres i dette kapitel, er skrevet udfra (Mehmet Balcilar, 2016).*

Designmatricen noteres i dette afsnit som  $\mathbf{X}_n = [\mathbf{x}_1, \dots, \mathbf{x}_n]^T$ , hvor  $\mathbf{x}_t = (x_{t1}, \dots, x_{tp})$ , for  $t = 1, \dots, n$  er den  $t$ 'te række i designmatricen.

Målet med forecasting er, at prædiktere fremtidige værdier af en tidsrække,  $y_{n+h}$ , for  $h = 1, 2, \dots$ , baseret på data frem til  $n$ , dvs.  $\mathbf{X}_n$ . Prædiktionen med den mindste gennemsnitlige kvadrerede prædiktionsfejl af  $y_{n+h}$  er givet ved

$$y_{n+h|n} = E[y_{n+h} | \mathbf{X}_n].$$

I dette projekt ønsker vi blot, at prædiktere one-step-ahead, dvs.  $h = 1$ . Til one-step-ahead prædiktion betragtes modellen

$$y_{n+1} = \mathbf{x}_n \phi + w_{n+1},$$

hvor

$$E[y_{n+1} | \mathbf{X}_n] = \mathbf{x}_n \phi.$$

For at prædiktere opdeles datasættet i en initialtræningsmængde med  $t_0$  observationer og en testmængde med  $T = T_0 - t_0$  observationer. Observationerne i træningsmængden benyttes til at identificere modellen og estimere parametrene, mens observationerne i testmængden fungerer som en referenceramme, når vi vil vide hvor meget de forudsagte værdier afviger fra de observerede værdier.

En måde at vurdere, hvor god en model er til at prædiktere, er ved at anvende den gennemsnitlige kvadrerede fejl (MSE) på testmængden (se definition B.1). En lav værdi af MSE på testsættet indikerer, at de prædikterede værdier er tæt på de observerede værdier.

En anden måde at vurdere to konkurrerende prædiktioner i forhold til hinanden er ved anvendelse af Diebold-Mariano testen. Vi lader  $y_{n+1|n}^1$  og  $y_{n+1|n}^2$  være de to konkurrerende prædiktioner af  $y_{n+1}$ , hvor  $n = t_0, \dots, T_0$ . De tilsvarende prædiktionsfejl er defineret ved

$$\begin{aligned}\epsilon_{n+1|n}^1 &= y_{n+1} - y_{n+1|n}^1, \\ \epsilon_{n+1|n}^2 &= y_{n+1} - y_{n+1|n}^2.\end{aligned}$$

Vi måler nøjagtigheden af one-step-ahead forecastet i en tabsfunktion

$$L(\epsilon_{n+1|n}^i), \quad i = 1, 2.$$

Disse tabsfunktioner kan eksempelvis være

- $L(\epsilon_{n+1|n}^i) = (\epsilon_{n+1|n}^i)^2$
- $L(\epsilon_{n+1|n}^i) = |\epsilon_{n+1|n}^i|$ .

Diebold-Mariano testen er baseret på differensen af tabsfunktionerne

$$d_n = L(\epsilon_{n+1|n}^1) - L(\epsilon_{n+1|n}^2).$$



**Definition 3.6: Diebold-Mariano test (DM)**

Betragt  $\mathcal{H}_0 : E[d_n] = 0$  imod  $\mathcal{H}_A : E[d_n] \neq 0$ . Diebold-Mariano testen er defineret som følgende

$$DM = \frac{\bar{d}}{(\widehat{\text{avar}}(\bar{d}))^{1/2}} = \frac{\bar{d}}{(\widehat{\text{LRV}}_{\bar{d}}/T_0)^{1/2}},$$

hvor  $\bar{d} = \frac{1}{T} \sum_{n=t_0}^{T_0} d_n$  og  $\text{LRV}_{\bar{d}} = \gamma_0 + 2 \sum_{j=1}^{\infty} \gamma_j$ , hvor  $\gamma_j = \text{Cov}(d_n, d_{n-j})$ .

Det kan vises, at teststørrelsen DM er asymptotisk standard normalfordelt under nulhypotesen. Derfor afvises nulhypotesen, hvis  $|DM| > 1.96$ .



## 4 MODELLERING

---

Formålet med dette kapitel er at prædiktere væksten af USA's realt BNP et kvartal frem. I teorien antog vi, at designmatricen har fuld rang, dette er dog ikke tilfældet med datasættet, som vi arbejder med, da vi har langt flere forklarende variable end observationer. Derfor fravælges manuelt nogle forklarende variable således at teorien kan anvendes (mvs). Herpå findes en benchmark model samt øvrige modeller, som har til formål at prædiktere bedre end benchmark modellen.

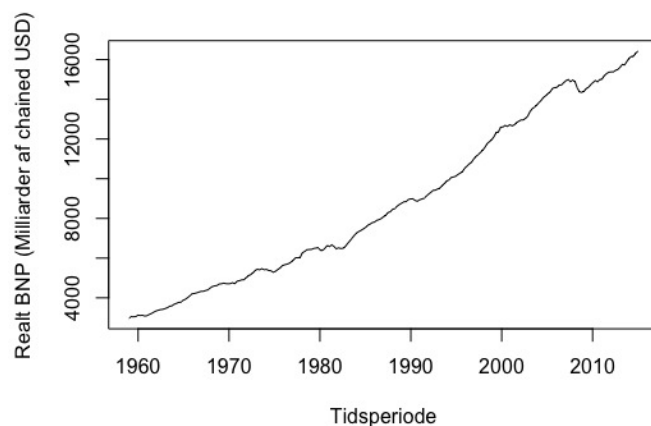
Herefter anvendes lasso til variabel udvælgelse på det fulde datasæt, hvor der ikke manuelt er fravalgt variable (lvs).

### 4.1 Datasættet

Datasættet, som vi arbejder med i dette projekt, er kvartalsdata som er hentet fra Federal Reserve Bank of St. Louis. Datasættet indeholder 227 observationer og 257 variable i perioden 1. marts 1959 til 1. september 2015, hvoraf USA's realt BNP er én af disse variable.<sup>1</sup>

#### Transformation af data

USA's realt BNP er opgjort i chained amerikanske dollars, som er et mål der udtrykker reelle priser, dvs. priser justeret i forhold til inflation. Tidsrækken for USA's realt BNP er vist i figur 4.1.



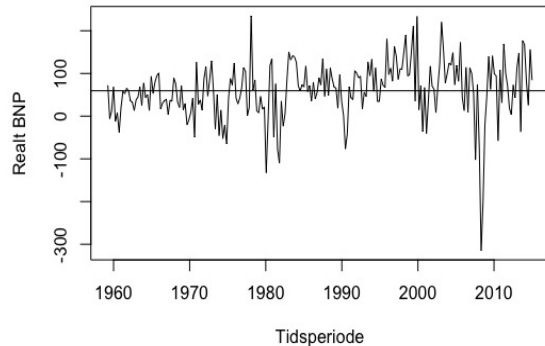
Figur 4.1: USA's realt BNP fra d. 1. marts 1959 til 1. september 2015.

Af figur 4.1 ses en opadgående trend for USA's realt BNP, hvilket betyder, at tidsrækken ikke er stationær. Vi tager derfor første differensen, som ses på figur 4.2. Her er den opadgående trend fjernet, men samtidig lader det til, at variansen af tidsrækken er en stigende funktion af tiden. Det skyldes, at vi betragter den absolutte ændring i realt BNP fra periode til periode i stedet for den procentviseændring og dermed vækst af realt BNP.

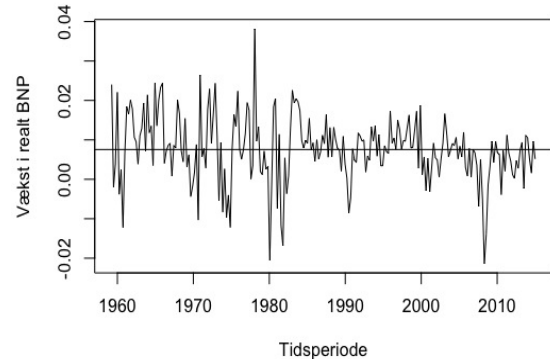
---

<sup>1</sup>[https://research.stlouisfed.org/econ/mccracken/fred-databases/FRED-QD\\_appendix\\_v6.pdf](https://research.stlouisfed.org/econ/mccracken/fred-databases/FRED-QD_appendix_v6.pdf)

Derfor tager vi yderligere logaritmen for at stabilisere variansen. Denne transformation ses på figur 4.3, som ser stationær ud.



Figur 4.2: Transformationen  $\nabla x_t$  af USA's realt BNP fra d. 1. marts 1959 til 1. september 2015.



Figur 4.3: Transformationen  $\nabla \log(x_t)$  af USA's realt BNP fra d. 1. marts 1959 til 1. september 2015.

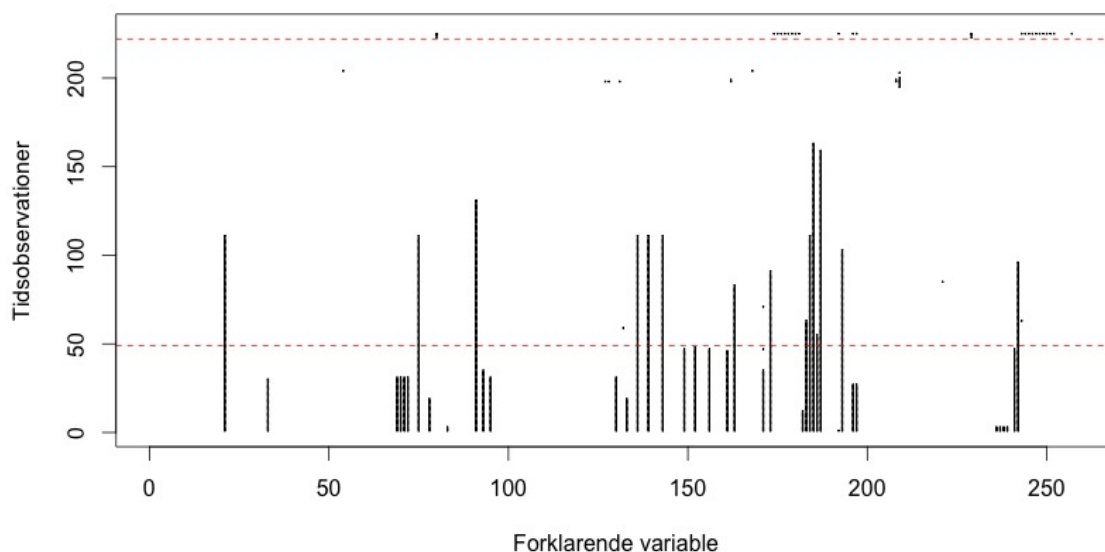
Som for USA's realt BNP foreslår Michael McCracken og Serena Ng. fra Federal Reserve Bank of St. Louis en transformation til hver af de resterende variable, således at disse også bliver stationære.

### Reducering af de forklarende variable

For at kunne anvende teorien, skal designmatricen have fuld rang, derfor reduceres antallet af forklarende variable. Variablerne, som fravælges, kan ses i appendiks D.

Først og fremmest ønsker vi at fjerne alle NAs i datasættet, som er ikke tilgængelige observationer. Ideen er at fjerne så mange NAs som muligt, men samtidig beholde flest muligt variable og observationer, da disse stadig kan have stor betydning i forhold til at prædiktere BNP. På figur 4.4 ses NAs for hver variabel til et givent tidspunkt.

De røde stiplede linjer indikerer tidsperioden, som vi mener, der fjerner flest NAs, men samtidig bevarer flest mulige variable og observationer. Derfor skal vi fremover blot se på 174 observationer i perioden d. 1 september 1971 til d. december 2014.



Figur 4.4: NAs for hver af de 257 variable i tidsperioden 1. september 1959 til 1. september 2015 svarende til 227 observationer. Den nedre røde stiplede linje markerer d. 1 september 1971, mens den øvre markerer d. 1. december 2014.

Derudover har vi valgt at se bort fra 15 variable, da disse har NAs længere fremme end d. 1. september 1971. De enkelte outliers erstattes med gennemsnittet af variabelens forrige værdier. Hermed har vi fjernet alle NAs. Datasættet består nu af 174 tidsobservationer og 242 variable i perioden d. 1 september 1971 til d. 1. december 2014.

For at reducere antallet af variable yderligere har vi lavet en subjektiv vurdering af variablerne i datasættet og fjernet de variable, som vi mener bedst kan udelades. Vi har fjernet variable, hvor vi har vurderet, at informationen i variabelen er indeholdt i større eller mindre grad i en anden variabel. Vi har for eksempel valgt at beholde variabelen “Consumption”, som er niveauet af privatpersoners forbrug og fjernet variablerne “Cons:Dur”, “Cons:Svc” og “Cons:NonDur”, som beskriver forbrug af goder i kategorierne: service, varige forbrugsgoder og ikke-varige forbrugsgoder. Efter reducere af variablerne indeholder datasættet 174 tidsobservationer og 127 variable i perioden d. 1 september 1971 til d. 1. december 2014, således at antallet af observationer er større end antallet af variable.

Datasættet opdeles herefter i en træningsmængde og en testmængde. Vi lader træningsmængden bestå af observationerne fra 1. september 1971 til 1. december 2004, og testmængden bestå af observationerne fra 1. marts 2005 til 1. december 2014 igennem hele kapitlet. Initialtrænings-sættet består dermed af de første 134 observationer, og testmængden består af de sidste 40 observationer.

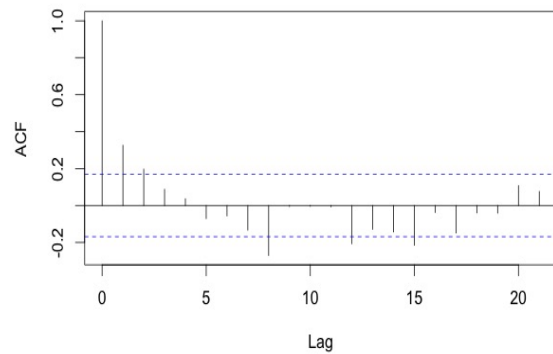
Vi prædikter som nævnt et kvartal frem, hvilket betyder, at træningsmængden udvides med én observation efter hver prædiktion. Denne fremgangsmåde fortsætter, indtil træningsmængden består af alle de givne observationer undtagen én. Således tager metoderne hensyn til nye informationer og tilpasser sig derefter. Modellens koefficienter revurderes derfor løbende efter hvert forecast. Vi udfører en residualanalyse af den sidst fittede model for hver metode, da denne har al information givet.

## 4.2 Benchmark model

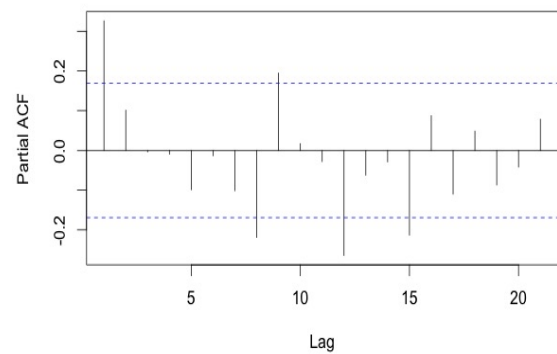
Vi har valgt en autoregressiv model som benchmark model, da den er simpel og kun anvender tidligere lags af BNP.

R-koderne for benchmark modellen kan findes i appendiks ??.

Ordenen af den autoregressive model kan bestemmes ud fra plots af ACF og PACF. På figur 4.5 og figur 4.6 ses den empiriske ACF og empiriske PACF for datasættet i træningsmængden. Plottet for den empiriske ACF kan ses som værende aftagende mod 0, og det taler derfor ikke imod, at vi benytter en autoregressive model. Ud fra plottet af den empiriske PACF kunne det vurderes, at den skærer af efter første lag, hvorefter en AR(1) model kan benyttes.



Figur 4.5: ACF for transformationen  $\nabla \log(x_t)$  af USA's realt BNP fra d. 1. september 1971 til d. 1. december 2004.



Figur 4.6: PACF for transformationen  $\nabla \log(x_t)$  af USA's realt BNP fra d. 1. september 1971 til d. 1. december 2004.

At vælge ordenen ud fra plots kan være en meget subjektiv vurdering, og vi benytter derfor informationskriterierne AIC, AICc og BIC til at vælge den endelige model. Vi antager, at væksten for realt BNP ikke er korreleret med BNP med mere end 2 år tilbage. I tabel 4.1 ses værdierne af AIC, AICc og BIC for AR( $p$ ) modeller, hvor  $p = 1, \dots, 8$ . Alle informationskriterierne foretrækker en AR(1) model.

$p$	BIC	AIC	AICc
1	-9.661045	-8.683556	-8.666908
2	-9.625442	-8.670462	-8.653028
3	-9.587318	-8.654849	-8.636352
4	-9.550823	-8.640863	-8.621022
5	-9.514432	-8.626983	-8.605508
6	-9.477222	-8.612283	-8.588877
7	-9.471855	-8.629426	-8.603785
8	-9.499291	-8.679373	-8.651184

Tabel 4.1: Værdier af BIC, AIC og AICc for AR( $p$ ), hvor  $p = 1, \dots, 8$  for væksten af realt BNP i perioden fra d. 1. september 1971 til d. 1. december 2004.

Det viser sig, at BIC konsekvent foretrækker en AR(1) model for hvert forecast, derfor forecastes observationerne fra 1. marts 2005 til 1. december 2014 alle ved

$$x_{n+1|n} = \hat{\phi}_0 + \hat{\phi}_1 x_n,$$

hvor  $\hat{\phi}_0$  og  $\hat{\phi}_1$  er OLS estimater for henholdsvis  $\phi_0$  og  $\phi_1$ . AIC vælger en AR(1) model til de 11 første forecasts, hvorefter en AR(8) model foretrækkes, de sidste 29 observationer forecasts derfor ved

$$x_{n+1|n} = \hat{\phi}_0 + \sum_{i=1}^8 \hat{\phi}_i x_{n+1-i}.$$

AICc laver samme skift fra en AR(1) model til en AR(8) model, men først efter 15 forecasts.

I tabel 4.2 ses det, at modellen udvalgt ved hjælp af BIC præsterer den mindste MSE. BIC valgte konsekvent en AR(1) og derfor vælges en AR(1) som benchmark.

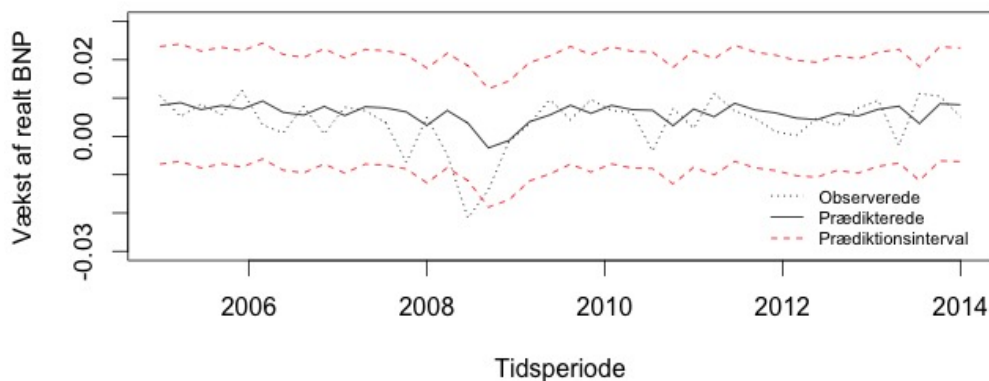
	MSE
AR( $p_{\text{BIC}}$ )	$4.339064 \cdot 10^{-05}$
AR( $p_{\text{AIC}}$ )	$4.918886 \cdot 10^{-05}$
AR( $p_{\text{AICc}}$ )	$4.999394 \cdot 10^{-05}$

Tabel 4.2: MSE af AR( $p_{\text{BIC}}$ ), AR( $p_{\text{AIC}}$ ) og AR( $p_{\text{AICc}}$ ) i perioden fra 1. marts 2005 til 1. december 2014.

På figur 4.7 illustreres de observerede værdier samt de prædikterede værdier med AR( $p_{\text{BIC}}$ ), hvor ordenen er valgt ud fra BIC, samt prædiktionsintervallet herfor. Intervallet findes ud fra

$$x_{n+1|n} \pm 1.96\sqrt{\hat{\sigma}_p},$$

hvor  $\hat{\sigma}_p$  betegner den estimerede varians af de rå residualer, og det antages, at prædiktionerne er normalfordelte.

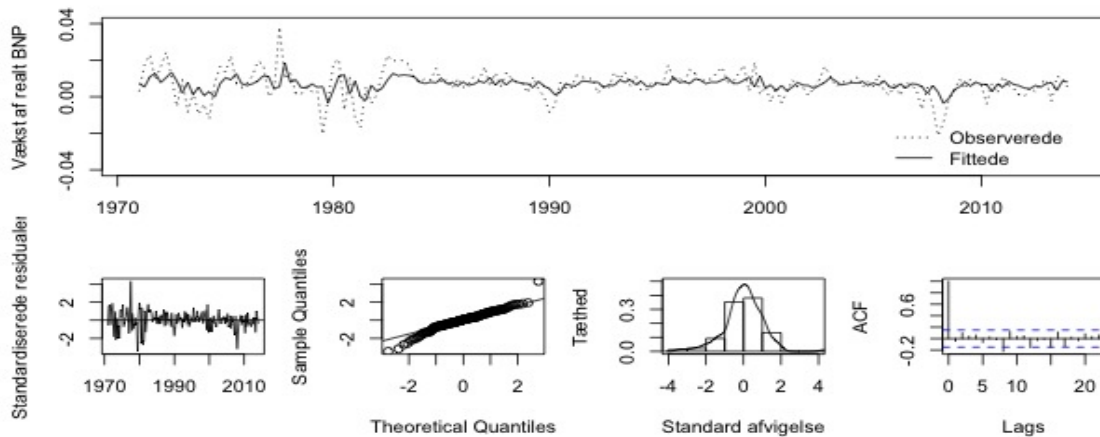


Figur 4.7: De givne observationer og prædiktionerne af disse modelleret med AR( $p_{\text{BIC}}$ ) samt prædiktionsintervallet i perioden fra 1. marts 2005 til 1. december 2014.

Herefter analyseres de standardiserede residualer for benchmark modellen fittet på alle 173 tidsobservationer

$$\hat{y}_{t+1} = 0.004 + 0.373y_t, \quad t = 1, \dots, n-1,$$

som bør være uafhængige. Derfor undersøger vi, om de standardiserede residualer er normalfordelte og ukorrelerede.



Figur 4.8: Analyse af de standardiserede residualer for benchmark modellen.

Af figur 4.8 tydes, at de standardiserede residualer tilnærmelsesvis følger en normalfordeling med en middelværdi omkring 0. Ud fra qq-plottet ses dog, at de standardiserede residualer afviger fra den rette linje. Derudover ses der ingen autokorrelation i de standardiserede residualer. Dette kan testes med Jarque-Bera testen og Ljung Box testen.

	Jarque-Bera testen	Ljung Box testen			
		$h = 5$	$h = 10$	$h = 15$	$h = 20$
$P$ -værdi	$6.726 \cdot 10^{-10}$	0.3051	0.05394	0.03363	0.03525

Tabel 4.3:  $P$ -værdi fra Jarque-Bera testen og Ljung Box testen på de standardiserede residualer af benchmark modellen.

Af tabel 4.3 kan vi dog ikke konkludere, at de standardiserede residualer for benchmark modellen er uafhængige som ønsket, da de standardiserede residualer afvises at være normalfordelte og ukorrelerede for  $h = 15$  og  $h = 20$ .

De mest ekstreme outliers som ses på qq-plottet identificeres til henholdsvis d. 1. juni 1977 og d. 1. juni 1979, som formentlig følger af oliekrisen i 1970'erne og den økonomiske krise i 1980'erne. Vi har forsøgt at fjerne disse outliers, men dette har ikke resulteret i at de standardiserede residualer er normalfordelte dog er autokorrelationen væk, hvor den mindste  $p$ -værdi er 0.1586 for  $h = 5$ .



### 4.3 Forbedring af prædiktion

I denne sektion ønsker vi at forbedre vores benchmark model. Derfor anvendes metoderne, som er beskrevet i teoriafsnittet, på datasættet. Til sidst sammenlignes modellerne, hvor vi herudfra diskuterer hvilken model, der er bedst til at prædiktere væksten af realt BNP et kvartal frem. Til de nedenstående modeller bruges pakken glmnet i R (skrevet af Hastie, T. Tibshirani, R. et al.), som finder løsninger til koefficienterne med en følge af relevante strafparametre,  $\lambda$ . Løsningen med den mindste BIC bruges i modellen til at prædiktere med. Med samme procedure som tidligere opdeles datasættet i en træningsmængde og en testmængde.

#### AR( $p$ ) med adaptive lasso

Ordenen i en autoregressiv model kan også bestemmes udfra adaptive lasso, hvor vægtene bestemmes udfra OLS (se appendiks E.0.2).

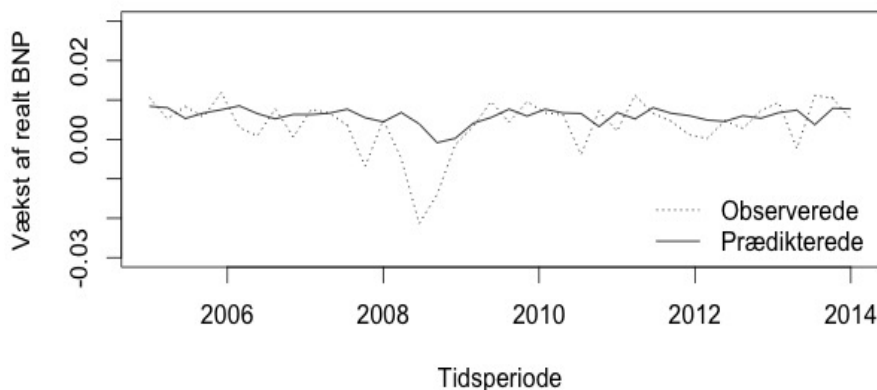
I de første 16 forecasts bruges en AR model, som indeholder skæring og lag 1, 2 og 8

$$x_{n+1|n} = \hat{\phi}_0 + \hat{\phi}_1 x_n + \hat{\phi}_2 x_{n-1} + \hat{\phi}_9 x_{n-8},$$

og til de resterende 24 forecasts anvendes en AR(1) model

$$x_{n+1|n} = \hat{\phi}_0 + \hat{\phi}_1 x_n.$$

Som det ses på figur 4.9, sker der en drastisk ændring i BNP omkring 2008. Derfor kan finanskrisen være skyld i dette skrift mellem de to autoregressive modeller.

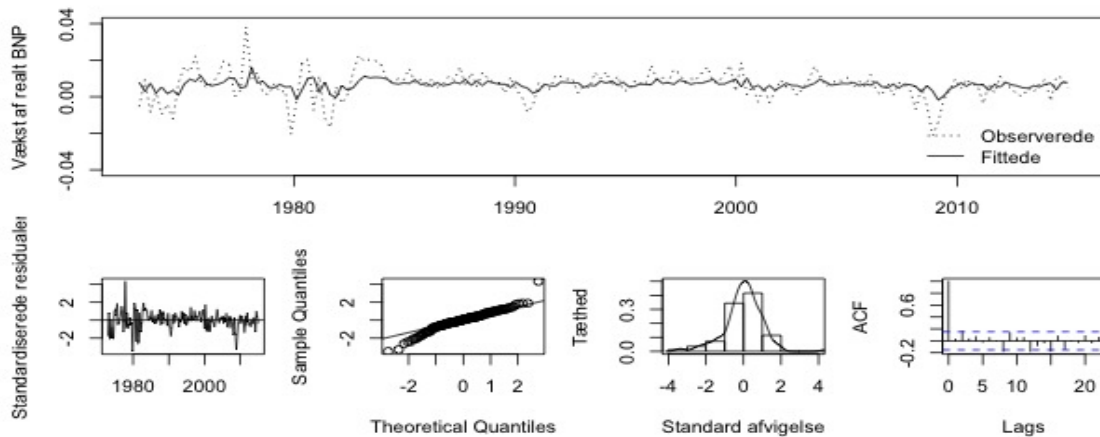


Figur 4.9: Prædiktion af væksten for USA's realt BNP med en autoregressiv model, hvor ordenen bestemmes udfra adaptive lasso.

Den sidst fittede model er givet ved

$$\hat{y}_{t+1} = 0.005 + 0.298y_t, \quad t = 1, \dots, n-1. \quad (4.1)$$

Vi udfører herefter residualanalyse på denne model, som ses i figur 4.10.



Figur 4.10: Analyse af de standardiserede residualer af (4.1).

	Jarque-Bera testen	Ljung Box testen			
$P$ -værdi	$3.462 \cdot 10^{-12}$	$h = 5$	$h = 10$	$h = 15$	$h = 20$
		0.3762	0.1618	0.03548	0.0465

Tabel 4.4:  $P$ -værdi fra Jarque-Bera testen og Ljung Box testen på de standardiserede residualer af (4.1).

Af tabel 4.4 ses, at de standardiserede residualer afvises at være normalfordelte, og igen er de standardiserede residualer ukorrelerede for  $h = 5$  og  $h = 10$ , men korrelerede for  $h = 15$  og  $h = 20$ .

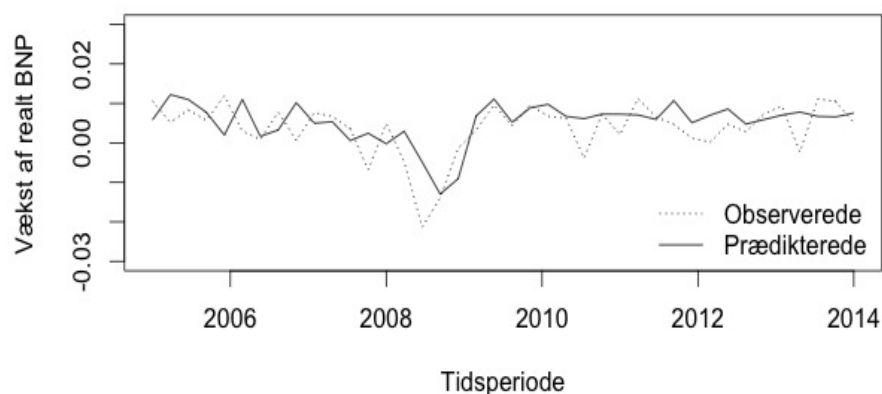
Igen fjernes de voldsommeste outliers, som hænder d. 1. september 1977 og d. 1. september 1979. Men dette forbedrer ikke de standardiserede residualer.

Vi ønsker nu at inkludere flere forklarende variable end blot lags i modellerne. Dermed lader vi de forklarende variable bestå af både laggede værdier af BNP og de øvrige forklarende variable, som vi har udvalgt manuelt. Da den partielle autokovariansfunktion for BNP skærer af efter første lag, hvilket ses på figur 4.6, og benchmark modellen blot anvender et lag, vælger vi blot at anvende et enkelt lag af BNP.

## Ridge regression

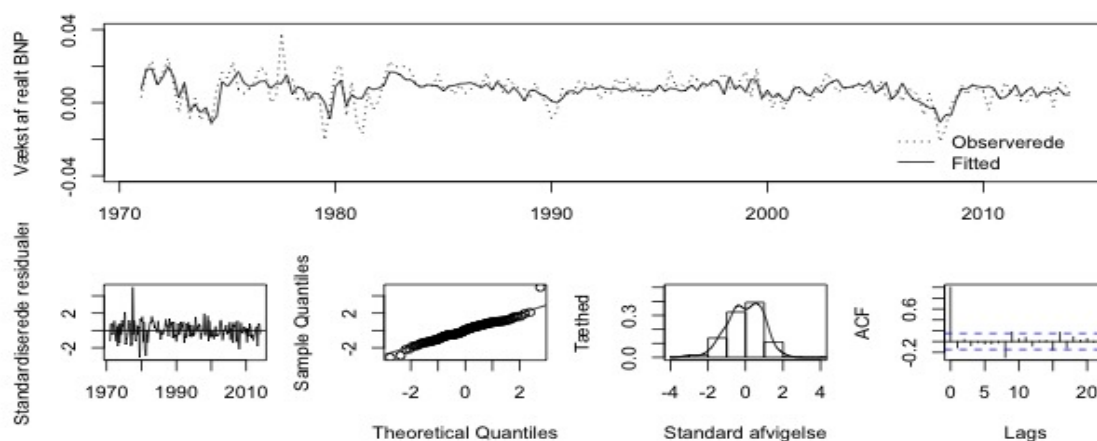
Som nævnt tidligere anvender ridge regression alle de forklarende variable, som er udvalgt manuelt.

I figur 4.12 ses ét-kvartalsprædiktation foretaget med en model, hvor estimationen af parametrene er foretaget ved ridge regression (se appendiks E.0.3).



Figur 4.11: Prædiktion af væksten for USA's reelt BNP med ridge regression.

Igen ser vi på de standardiserede residualer for den sidst fittede model i testmængden.



Figur 4.12: Analyse af de standardiserede residualer af modellen udvalgt ved ridge regression.

	Jarque-Bera testen
$P$ -værdi	$4.885 \cdot 10^{-15}$

Tabel 4.5:  $P$ -værdi fra Jarque-Bera testen på de standardiserede residualer af modellen udvalgt ved ridge regression.

Af tabel 4.5 ses, at de standardiserede residualer afvises at være normalfordelte. Ljung Box testen kan ikke anvendes på de standardiserede residualer i dette tilfælde, da antallet af forklarende variable er større end antallet af lags, som vi har valgt at bruge i testen.

Outlieren, som fremkommer d. 1. juni 1977, afviger mest fra de øvrige residualer, og fjernes derfor, hvorved vi opnår, at de standardiserede residualer er normalfordelte, hvor  $p$ -værdien er

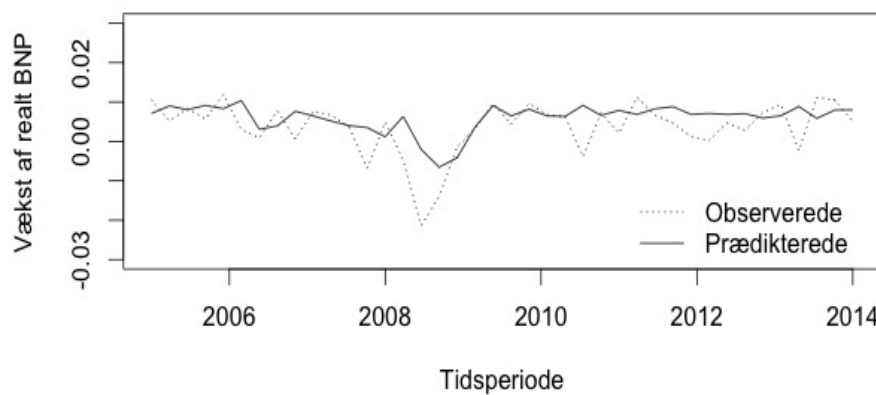
0.0682.

Modellen udvalgt udfra ridge regression er meget kompleks, da alle forklarende variable er inkluderet. Men ofte har alle variable ikke lige stor betydning på prædiktionen af BNP, derfor anvendes lasso som udfører variable udvælgelse.

### Lasso

På figur 4.13 vises ét-kvartalsprædiktion foretaget med en model, hvor estimationen af parametrene er foretaget med lasso (se appendiks E.0.4).

For hvert forecast bestemmer lasso, hvilke forklarende variable der fravælges i modellen. Derfor vil lasso generelt ikke inkludere alle forklarende variable i modellen.



Figur 4.13: Prædiktion af væksten for USA's reelt BNP med lasso.

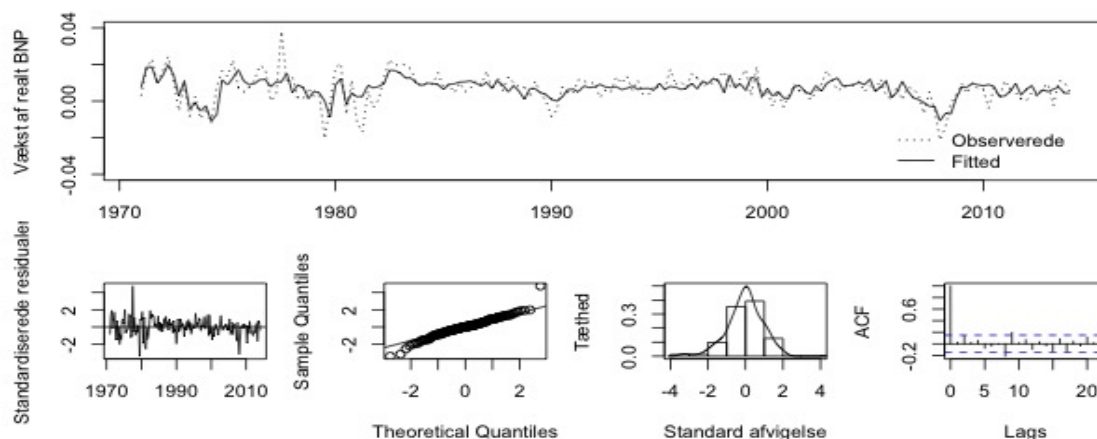
Spørgsmålet er så, hvordan lasso udvælger disse forklarende variable. Hvis eksempelvis vi har to højt korrelerede variable, vælger lasso blot den ene af disse, hvilket kan have en betydning for prædiktionsevnen af lasso. Dette problem opstår ikke i ridge regression, da den anvender alle variable.

I tabel F.1 ses hvilke variable, som lasso anvender hyppigst. Herudfra ses, at variablene "Consumption", "FixedInv" og "Hpermits" vælges i hvert forecast. Antallet af variable, som lasso medtager i hvert forecast, ændres omkring finanskrisen i 2008. Før og under krisen anvender lasso omtrent 12 forklarende variable, mens 6 forklarende variable anvendes efter krisen.

Den sidst tilpassede model i testmængden er givet ved

$$\hat{y}_{t+1} = 1.99 \cdot 10^{-3} + 0.22x_{t1} + 0.022x_{t2} + 0.01x_{t3} + 5.62 \cdot 10^{-5}x_{t4} + 0.25x_{t5}, \quad (4.2)$$

for  $t = 1, \dots, n-1$ , hvor  $x_{t1}$  er "PCECC96",  $x_{t2}$  er "FPIx",  $x_{t3}$  er "PERMIT",  $x_{t4}$  er "NAPMNOI" og  $x_{t5}$  er "HWIURATION", som netop også er variablene der forekommer hyppigst gennem alle forecasts.



Figur 4.14: Analyse af de standardiserede residualer for (4.2).

	Jarque-Bera testen	Ljung Box testen		
		$h = 10$	$h = 15$	$h = 20$
$P$ -værdi	$9.992 \cdot 10^{-15}$	0.02931	0.04429	0.04236

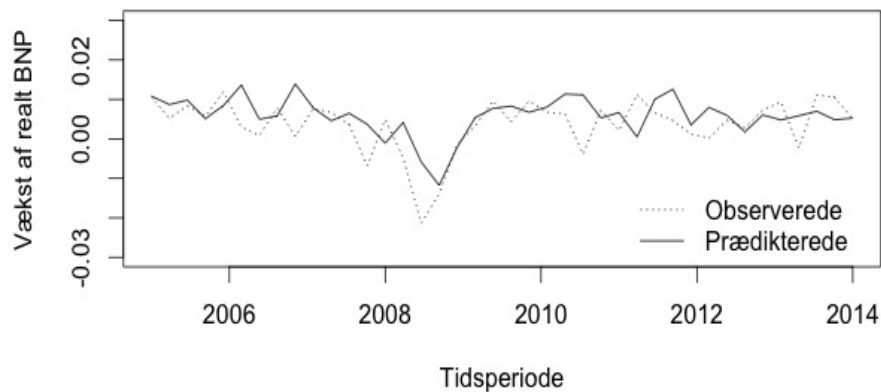
Tabel 4.6:  $P$ -værdi fra Jarque-Bera testen og Ljung Box testen på de standardiserede residualer af (4.2).

Igen kan vi af tabel 4.6 konkludere, at de standardiserede residualer ikke er normalfordelte og korreleret.

Vi fjerner outlierne, som opstår d. 1. juni 1977 og d. 1. juni 1979, hvilket resulterer i, at de standardiserede residualer er normalfordelte og ukorrelerede, og dermed uafhængige, da  $p$ -værdien for Jarque-Bera testen er 0.06301 og den mindste  $p$ -værdi for Ljung Box testen er 0.09441 for  $h = 5$ .

### Adaptive lasso

I figur 4.17 ses ét-kvartalsprædiktions foretaget med en model, hvor estimationen af parametrene er foretaget ved adaptive lasso, hvor vægtene bestemmes udfra OLS (se appendiks E.0.5).



Figur 4.15: Prædiktion af væksten for USA's reall BNP med adaptive lasso, hvor vægtene bestemmes ud fra OLS.

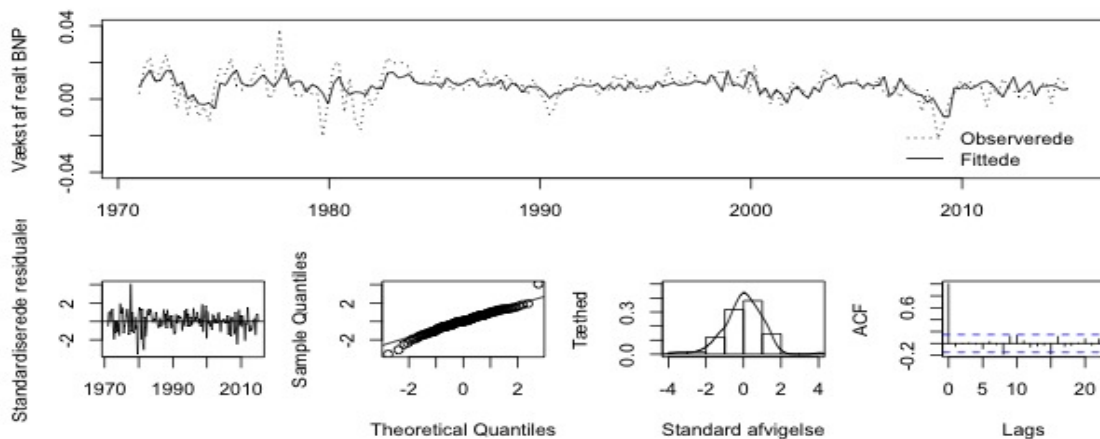
I tabel F.2 ses hvilke variable adaptive lasso med OLS vægte anvender oftest. Herudfra ses, at variablene “Output:NFB” og “Real HHW:WSA” vælges i hvert forecast.

Antallet af forklarende variable ændres ikke markant omkring finanskrisen med adaptive lasso med OLS vægte, som det var tilfældet med lasso. Adaptive lasso vælger forholdsvis de samme forklarende variable og antallet er omtrent uændret, som før krisen.

Den sidst fittede model er givet ved

$$\hat{y}_{t+1} = 0.08 + 0.40x_{t1} - 0.001x_{t2} - 0.16x_{t3} - 0.10x_{t4} - 0.03x_{t5} + 0.09x_{t6}, \quad (4.3)$$

for  $t = 1, \dots, n-1$ , hvor  $x_{t1}$  er “OUTNFB”,  $x_{t2}$  er “AWHMAN”,  $x_{t3}$  er “OPHBS”,  $x_{t4}$  er “UNLP-NBS”,  $x_{t5}$  er “LIABPIx” og  $x_{t6}$  er “TNWBSHNOx”, som netop også er variablene der forekommer hyppigst gennem alle forecasts.



Figur 4.16: Analyse af de standardiserede residualer for (4.3).

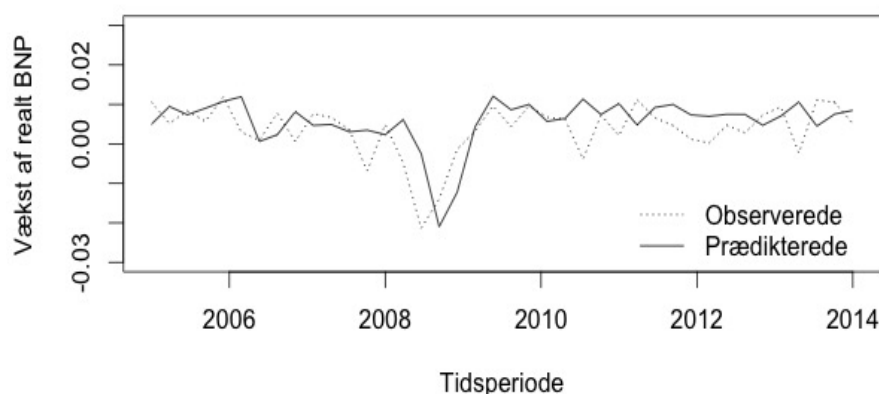
	Jarque-Bera testen	Ljung Box testen		
$P$ -værdi	$3.789 \cdot 10^{-12}$	$h = 10$	$h = 15$	$h = 20$
		0.1464	0.273	0.3192

Tabel 4.7:  $P$ -værdi fra Jarque-Bera testen og Ljung Box testen på de standardiserede residualer for (4.3).

Af tabel 4.7 konkluderes, at de standardiserede residualer ikke er normalfordelte, men ukorrelerede. Outlierne, som fjernes i dette tilfælde, fremkommer igen d. 1. juni 1977 og d. 1. juni 1979. Hermed bliver de standardiserede residualer normalfordelte og ukorrelerede og dermed uafhængige, da  $p$ -værdien for Jarque-Bera testen er 0.8433 og den mindste  $p$ -værdi for Ljung Box testen er 0.5704 for  $h = 5$ .

Vi har valgt også at se på adaptive lasso med lasso vægte, da antallet af forklarende variable ikke er så stort i forhold til antallet af observationer, således at OLS estimaterne nok ikke er særligt gode grundet overfitting.

I figur 4.18 ses ét-kvartalsprædiktions foretaget med en model, hvor estimationen af parametrene er foretaget ved adaptive lasso, hvor vægtene bestemmes ud fra lasso (se appendiks E.0.6).



Figur 4.17: Prædiktions af væksten for USA's reall BNP med adaptive lasso, hvor vægtene bestemmes ud fra lasso.

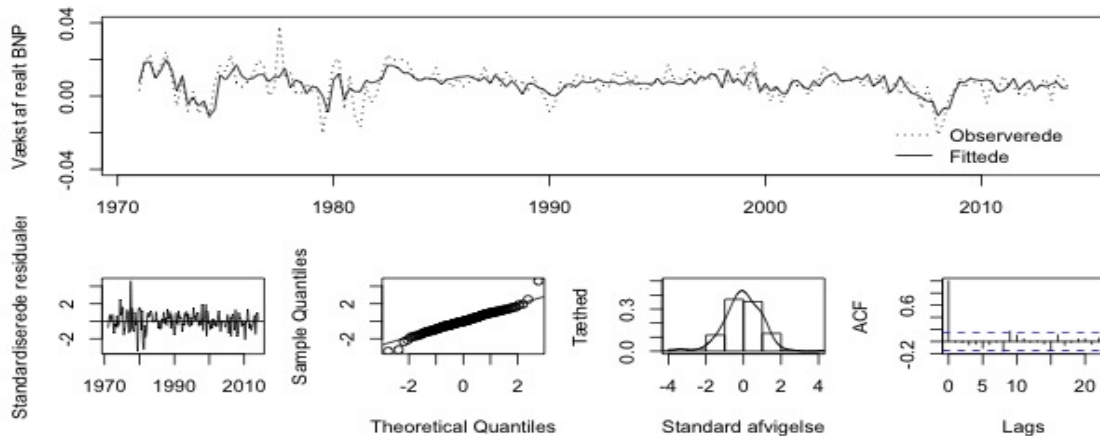
I tabel F.3 ses hvilke variable adaptive lasso, hvor vægtene bestemmes ud fra lasso, anvender hyppigst. Herudfra ses, at variablene "FixedInv" og "Hpermits" vælges i hvert forecast.

Igen medfører finanskrisen et fald i antallet af forklarende variable, som adaptive lasso med lasso vægte udvælger.

Modellen, som anvendes i sidste observation i testmængden, er givet ved

$$\hat{y}_{t+1} = -0.0075 + 0.293x_{t1} + 0.0343x_{t2} + 0.0267x_{t3} + 0.0002x_{t4}, \quad (4.4)$$

for  $t = 1, \dots, n-1$ , hvor  $x_{t1}$  er "PCEC96",  $x_{t2}$  er "FBIx",  $x_{t3}$  er "PERMIT" og  $x_{t4}$  er "NAPMNOI", som netop også er variablene der forekommer hyppigst gennem alle forecasts.



Figur 4.18: Analyse af de standardiserede residualer for (4.4).

	Jarque-Bera testen	Ljung Box testen			
		$h = 5$	$h = 10$	$h = 15$	$h = 20$
$P$ -værdi	$2.2 \cdot 10^{-16}$	0.001593	$1.094 \cdot 10^{-5}$	$3.676 \cdot 10^{-5}$	$2.542 \cdot 10^{-5}$

Tabel 4.8:  $P$ -værdi fra Jarque-Bera testen og Ljung Box testen på de standardiserede residualer af (4.4).

Af tabel 4.8, afvises de standardiserede residualer at være normalfordelte og ukorrelerede. Hvis vi fjerner outlieren, som finder sted d. 1. juni 1977, er de standardiserede residualer normalfordelte, men ikke ukorrelerede, da  $p$ -værdien for Jarque-Bera testen er 0.2418 og den mindste  $p$ -værdi for Ljung Box testen er 0.0047 for  $h = 20$ .

## 4.4 Sammenligning

Det ses tydeligt på figur 4.3, at finanskrisen i 2008 påvirkede BNP. I modellerne for hver metode, kan vi ofte se krisens påvirkning i forhold til antallet af forklarende variable, der anvendes til prædiktion af væksten af USA's realt BNP. Der ses en tendens til, at modellerne bruger flere forklarende variable før og under krisen end efter krisen. Derudover kan vi også se, at de fleste modeller har svært ved prædiktere selve udsvinget, som ofte forekommer med et par observationers forsinkelse.

Herefter ser vi på modellerne, som anvendes i sidste observation i testmængden, og undersøger om modellens standardiserede residualer uafhængige, som de bør være ifølge teorien. Men vi finder, at modellernes standardiserede residualer ikke er normalfordelte. Dette skyldes primært modellernes outliers, der hovedsagligt finder sted d. 1. juni 1977 og d. 1. juni 1979. Hvis vi fjerner disse outliers, da fandt vi, at de standardiserede residualer er normalfordelte og ukorrelerede og dermed uafhængige, som ønsket. Det bør bemærkes, at standardiserede residualer stadig kan være uafhængige, selvom de ikke er normalfordelte.

Herefter vurderes metodernes prædiktion ud fra MSE. Prædiktionen med mindst MSE prædikerer de observerede værdier bedst.



	MSE	Relativ MSE til benchmark
AR( $p_{\text{BIC}}$ ) (benchmark)	$4.339064 \cdot 10^{-5}$	1
AR( $p_{\text{AL}}$ )	$4.320997 \cdot 10^{-5}$	0.995
Ridge regression	$3.300393 \cdot 10^{-5}$	0.760
Lasso	$3.47645 \cdot 10^{-5}$	0.801
Adaptive lasso m. OLS vægte	$3.905933 \cdot 10^{-5}$	0.900
Adaptive lasso m. lasso vægte	$4.391141 \cdot 10^{-5}$	1.012

Tabel 4.9: Metodernes MSE på testmængden samt MSE relativ til benchmark modellen.

Af tabel 4.9 ses at ridge regression giver mindst MSE. Men modellen vil have høj kompleksitet, da alle forklarende variable er inkluderet, derfor foretrækkes lasso istedet.

Vi har også anvendt Diebold-Mariano testen på metodernes absolutte og kvadrerede prædiktionsfejl i forhold til benchmark modellen.

	$\left(\epsilon_{n+1 n}^i\right)^2$	$\left \epsilon_{n+1 n}^i\right $
AR( $p_{\text{AL}}$ )	0.4587	0.1118
Ridge regression	0.1567	0.3319
Lasso	0.1077	0.0803
Adaptive lasso m. OLS vægte	0.2589	0.3036
Adaptive lasso m. lasso vægte	0.5238	0.7276

Tabel 4.10:  $P$ -værdi fra DM testen.

Af tabel 4.10 ses at ingen model er signifikant bedre end benchmark modellen, da  $p$ -værdierne er større end 0.05.

## 4.5 Lasso variabel udvælgelse og adaptive lasso

Vi behandlede kun tilfældet, hvor designmatricen har fuld rang i teoriafsnittet, hvorfor vi slettede forklarende variable fra datasættet, for at kunne anvende den beskrevne teori.

I dette afsnit går vi ud over teorien, der er præsenteret i rapporten og anvender lasso på datasættet, som består af 174 tidsobservationer og 242 variable i perioden d. 1. september 1971 til d. 1. december 2014, dvs. det fulde datasæt uden NAs, for at foretage variabel udvælgelsen. For at benytte lasso til variabel udvælgelse, bestemmer vi på forhånd, hvor mange forklarende variable  $p_j$  vi ønsker at have tilbage i datasættet, som vi benytter adaptive lasso på. Ud fra det bestemmes  $\lambda_j$  således, at den tilsvarende lasso estimator

$$\hat{\beta}_{\lambda_j}^L = \min_{\beta} \left\{ \sum_{t=1}^n \left( Y_{t+1} - \sum_{j=1}^p \beta_j X_{jt} \right)^2 + \lambda_j \sum_{j=1}^p |\beta_j| \right\}$$

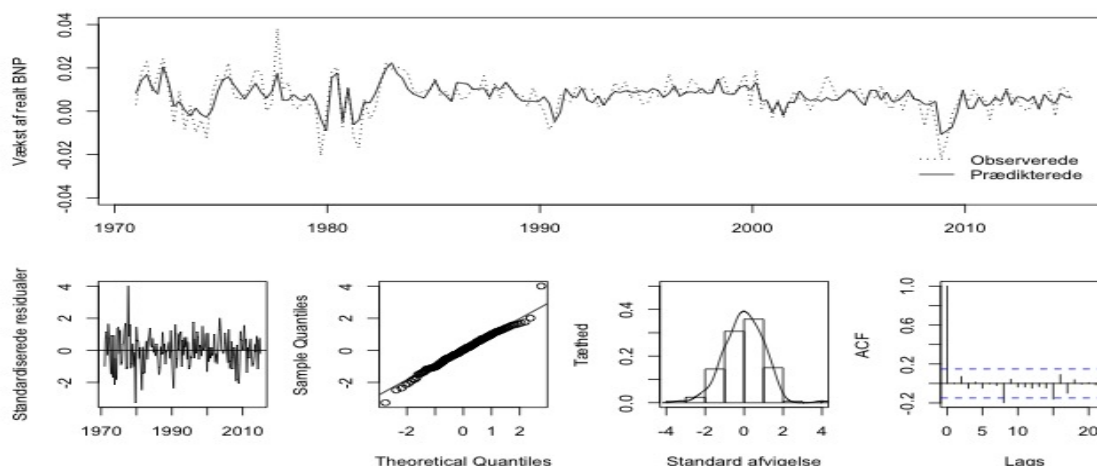
indeholder  $p_j$  indgange, der er forskellige fra nul. Vi har anvendt denne fremgangsmåde på det fulde datasæt, hvor vi har valgt antallet af forklarende variable til at være 10, 15, 20, 30 og 127. De 127 variable har vi valgt, da det er en lignende reducere af data, som vi selv foretog manuelt. De andre fire værdier er valgt for at øge forskellen i antal mellem observerede værdier og forklarende variable, så udgangspunktet for de OLS estimerede vægte i adaptive lasso forbedres.

Når vi benytter adaptive lasso på datasættet reduceret til 127 variable inkluderes følgende variable i næsten hvert eneste skridt af modellen: “A014RE1Q156NBEA”, “IMPGSC96”, “DPIC96”, “USMINE”, “USWTRADE”, “DNDGRG3Q086SBEA”, “DHUTRG3Q086SBEA”, “MZMREALx” og “CLAIMSx”. Forecasten fra denne model giver en MSE på  $3.564438 \cdot 10^{-5}$ , hvilket er bedre end da vi selv foretog variabel udvælgelsen som gav en MSE på  $3.909879 \cdot 10^{-5}$ . Da vi slettede variable ud fra vores egen vurdering valgte vi “A014RE1Q156NBEA”, “USMINE”, “USWTRADE”, “DNDGRG3Q086SBEA” og “DHUTRG3Q086SBEA” fra, da vi vurderede, at informationen i dem var indeholdt til en vis grad i andre variable. Det tyder altså på at informationen i disse variable ikke var tilstrækkeligt indeholdt i de variable vi valgte at beholde på bekostning af dem.

I tabel 4.11 ses det, at forecastene foretaget med modeller fitted med adaptive lasso, ser ud til at drage fordel af, at antallet af forklarende variable falder i forhold til antallet af observationer. For at sikre, at de forbedrede forecast ikke alene skyldes brugen af lasso, har vi lavet et forecast med en model fitted med lasso på hele datasættet, som i de fleste forecasts inkluderede mellem 7 og 12 forklarende variable i modellen, hvilket gav en MSE på  $3.191355 \cdot 10^{-5}$ . Kombinationen af lasso og adaptive lasso viser sig altså at kunne give bedre forecasts. Det bør dog også noteres, at forecastene for  $p = 127$  er bedre end for  $p = 30$ , så en reducere af  $p$  medfører nødvendigvis ikke et bedre forecast.

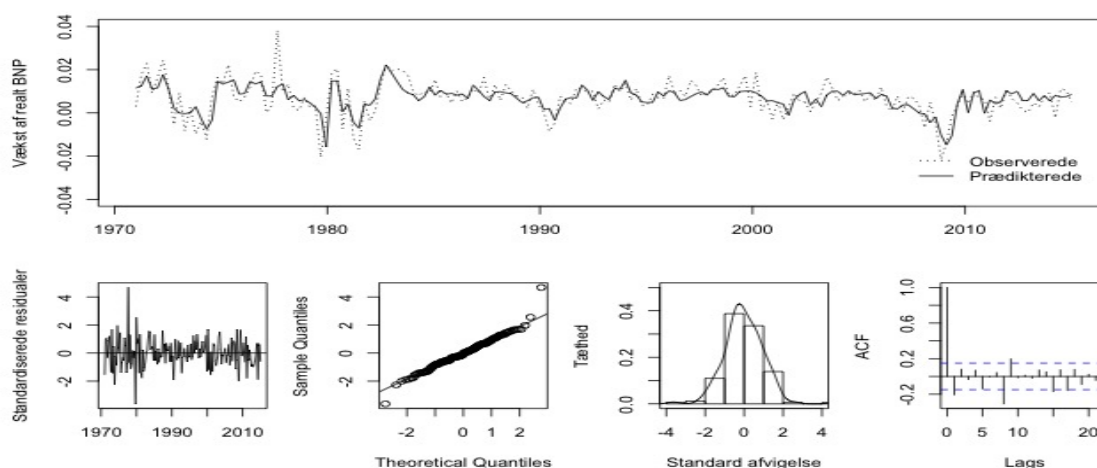
$p$	MSE	Relativ MSE til benchmark
10	$2.6205 \cdot 10^{-5}$	0.6039229
15	$2.7013 \cdot 10^{-5}$	0.6225531
20	$2.9819 \cdot 10^{-5}$	0.6872219
30	$4.0201 \cdot 10^{-5}$	0.9264946
127	$3.5644 \cdot 10^{-5}$	0.8214761

Tabel 4.11: MSE for modeller fitted med adaptive lasso modeller, hvor antallet af forklarende variable  $p$  er valgt ved hjælp af lasso.



Figur 4.19: Analyse af de standardiserede residualer for adaptive lasso med OLS vægte, hvor  $p = 127$ .

Når lasso reducerer antallet af variable til 10 inkluderer adaptive lasso udelukkende variablene “PCESVx”, “PRFIx”, “IPNMAT”, “LNS12032194”, “TOTRESNS”, “PERTMITS” og “NIKKEI25” i de sidste 10 forecasts. Af disse 7 valgte vi kun at beholde “TOTRESNS” og “NIKKEI25” i vores datasæt, da vi fjernede variable manuelt. Dette giver en betydeligt bedre MSE end da vi selv valgte variable, hvilket indikerer at vi ved manuel variabel udvælgelse har fjernet vigtig information fra datasættet.



Figur 4.20: Analyse af de standardiserede residualer for adaptive lasso med OLS vægte, hvor  $p = 10$ .

Residualanalysen for de to sidst fittede modeller, hvor  $p = 10$  og  $p = 127$  bekræfter ikke, at residualerne er uafhængige.

Når  $p = 10$ , kan det ses i tabel 4.12, at Jarque-Bera testen giver en  $p$ -værdi på  $1.202 \cdot 10^{-11}$ , så  $H_0$ -hypotesen om normalfordelte residualer afvises. Udfra qq-plottet i figur 4.20 ser det dog ud til, at bortset fra et par enkelte observerede residualer, så stemmer stikprøvekvartilerne meget godt overens med de teoretiske. Ved at fjerne de to standardiserede residualer med henholdsvis

den højeste og mindste værdi, giver Jarque-Bera testen en  $p$ -værdi på 0.785, hvilket indikerer normalfordelte variable. De to slettede observationer er begge fra sidst i 70'erne, hvor væksten i BNP ses at være meget volatil. I den sidst fittede model estimeres 8 parametre, derfor sammenlignes teststørrelsen fra Ljung box testen, hvor antallet af lags er valgt til 10, 15 og 20, med henholdsvis en  $\chi^2(2)$ -,  $\chi^2(7)$ - og  $\chi^2(12)$ -fordeling, hvilket giver  $p$ -værdier på  $1.276 \cdot 10^{-8}$ ,  $3.664 \cdot 10^{-7}$  og  $1.063 \cdot 10^{-6}$ . Det tyder altså på, at der er korrelation imellem variablene. Ved fjernelse af de to førnævnte variable fås en  $p$ -værdi på 0.0442.

Resultatet af residualanalysen for den sidst fittede model, hvor  $p = 127$  giver lignende resultater. Jarque-Bera testen giver en  $p$ -værdi på 0.0171, mens Ljung box testen giver 0.0004, så begge  $\mathcal{H}_0$ -hypoteser afvises.  $P$ -værdien af Ljung box teststørrelsen aflæses fra en  $\chi^2(3)$ -fordeling, da der estimeres 17 parametre i den sidste model. Ved at fjerne de to førnævnte observationer igen, giver Jarque-Bera testen og Ljung box testen henholdsvis 0.2784 og 0.0016 i  $p$ -værdi.

	Jarque-Bera testen	Ljung Box testen		
		$h = 10$	$h = 15$	$h = 20$
$p = 10$	$1.202 \cdot 10^{-11}$	$1.276 \cdot 10^{-8}$	$3.664 \cdot 10^{-7}$	$1.063 \cdot 10^{-6}$
$p = 127$	0.01714	—	—	0.0004285

Tabel 4.12:  $P$ -værdi fra Jarque-Bera testen og Ljung Box testen på de standardiserede residualer.

Hvis vi sammenligner de to forecasts med benchmark modellens ved hjælp af Diebold-Mariano testen, giver modellen, hvor  $p = 10$ ,  $p$ -værdier på 0.2551 og 0.1199 for henholdsvis den kvadrerede og den absolute tabsfunktion, hvorfor  $\mathcal{H}_0$ -hypotesen om, at modellernes forecasts er lige nøjagtige ikke forkastes. En lignende konklusion kan laves for  $p = 127$  modellen, som ses i tabel 4.13.

	$\left(\epsilon_{n+1 n}^i\right)^2$	$\left \epsilon_{n+1 n}^i\right $
$p = 10$	0.2551385	0.1198795
$p = 127$	0.4500845	0.2474051

Tabel 4.13:  $P$ -værdier fra DM testen.

## 5 KONKLUSION

---

En af antagelserne, der anvendes igennem teorien er, at designmatricen har fuld rang. Da datasættet fra Federal Reserve Bank of St. Louis indeholder 174 observationer og 242 variable har vi manuelt fravalgt nogle variable. For dette datasæt er metoden, der anvender ridge regression, den med mindst MSE. Denne er dog ikke betydelig lavere end lasso, så hvis en simpel model er ønsket, er lasso et bedre alternativt.

Af metoderne testet i rapporten er kombinationen af lasso til variable udvælgelse og derefter adaptive lasso til model tilpasing den mest lovende. Dette må derfor være en indikation på, at vi manuelt har fjernet vigtigt information fra datasættet. Dog skal det bemærkes, at ingen af modellerne, der er beskrevet, er signifikante bedre end benchmark modellen ifølge Diebold-Mariano testen.

Med hensyn til modellerne, der er identificeret fra metoderne, kan vi konkludere, at residualerne ikke er normalfordelte og ukorrelerede pga. ændringen af variansen i kriserne. Hvis dette er en ønsket egenskab, kunne man udvide modellen til at tage hensyn til en ændring i volatilitet. I kriser har væksten af reelt BNP en højere volatilitet end normalt. Denne ændring er ikke implementeret i modellerne, så det kunne tænkes, at vi burde have en model for volatiliteten. Uden at komme mere i dybden med det kunne to forslag til at modellere volatiliteten være GARCH( $p, q$ ), der tager hensyn til informationer om tidligere fejl og volatilitet, eller en markov model, hvor det antages, at der er forskellige volatilitetstilstande.



# A GENERELLE STATISTISKE BEGREBER

---

I dette appendiks defineres generelle statistiske begreber, som anvendes igennem rapporten.

## Definition A.1: Unbiased estimator

En estimator  $\hat{\theta} = \hat{\theta}(\mathbf{Y})$  siges at være unbiased, hvis

$$E[\hat{\theta}] = \theta.$$

## Definition A.2: Bias

Bias af en estimator defineres som

$$\text{bias}(\hat{\theta}) = E(\hat{\theta}) - \theta.$$

Løst sagt, siges en estimator at være konsistent, hvis følgen  $\theta_n(\mathbf{Y})$  af estimatorer for parameteren  $\theta$  konvergerer i sandsynlighed mod den sande værdi  $\theta$ . Ellers siges estimatoren at være inkonsistent. Konsistens er altså en asymptotisk egenskab.

## Definition A.3: Svag konsistens

Lad  $\{\hat{\theta}_n\}$  være en følge af estimatorer. Da siges  $\{\hat{\theta}_n\}$  at være konsistens for  $\theta$  hvis

$$\hat{\theta} \xrightarrow{p} \theta.$$

## Definition A.4: Rod-n-konsistent estimator

En estimator  $\hat{\theta}_n$  for  $\theta$  er rod-n-konsistent hvis

$$\hat{\theta}_n - \theta = O\left(\frac{1}{\sqrt{n}}\right).$$

## Sætning A.1: Slutsky's Theorem

Hvis  $X_n \xrightarrow{d} X$  og  $Y_n \xrightarrow{p} c$ , hvor  $X$  er en stokastisk variabel og  $c$  er en konstant, da gælder, at

$$\begin{aligned} X_n + Y_n &\xrightarrow{d} X + c \\ X_n Y_n &\xrightarrow{d} cX \\ \frac{X_n}{Y_n} &\xrightarrow{d} \frac{X}{c}, \quad \text{forudsat at } P(c = 0) = 0. \end{aligned}$$

Sætning A.1 tillader at finde grænsfordelingen af  $X_n$  og sandsynlighedsgrænsen af  $Y_n$  separat.

**Korollar A.1: Loven om itererede forventninger**

$$E[Y] = E[E[Y|X]].$$



## B GENNEMSNITLIGE KVADREREDE FEJL

---

Indledningsvis introduceres den gennemsnitlige kvadrerede fejl.

**Definition B.1: Gennemsnittet af kvadrerede fejl (MSE)**

Gennemsnittet af kvadrerede fejl er givet ved

$$\text{MSE} = \frac{\text{SSR}}{n}.$$

Vi ønsker, at finde en model med mindst mulig MSE, som vi nedenfor vil se, er en funktion af variansen og bias.

Vi lader  $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ , hvor  $\boldsymbol{\epsilon} \sim \text{iid}(0, \sigma_{\epsilon}^2)$ . Vi ønsker, at finde  $\hat{\boldsymbol{\beta}}$  som approksimerer den sande funktion  $\mathbf{y} = \mathbf{X}\boldsymbol{\beta}$  bedst muligt.

**Sætning B.1**

MSE er givet ved

$$E[(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})] = \sigma_{\epsilon}^2 + \text{Var}[\mathbf{X}\hat{\boldsymbol{\beta}}] + \text{Bias}[\mathbf{X}\hat{\boldsymbol{\beta}}]^2.$$

*Bevis.* Vi har, at  $\mathbf{X}\boldsymbol{\beta}$  er deterministisk og lad  $\boldsymbol{\epsilon}$  og  $\hat{\boldsymbol{\beta}}$  være uafhængige.

$$\begin{aligned} E[(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})] &= E[\mathbf{y}^T\mathbf{y} + \hat{\boldsymbol{\beta}}^T\mathbf{X}^T\mathbf{X}\hat{\boldsymbol{\beta}} - 2\hat{\boldsymbol{\beta}}^T\mathbf{X}^T\mathbf{y}] \\ &= E[\mathbf{y}^T\mathbf{y}] + E[\hat{\boldsymbol{\beta}}^T\mathbf{X}^T\mathbf{X}\hat{\boldsymbol{\beta}}] - E[2\hat{\boldsymbol{\beta}}^T\mathbf{X}^T\mathbf{y}] \\ &= \text{Var}[\mathbf{y}] + E[\mathbf{y}]^T E[\mathbf{y}] + \text{Var}[\mathbf{X}\hat{\boldsymbol{\beta}}] + E[\mathbf{X}\hat{\boldsymbol{\beta}}]^T E[\mathbf{X}\hat{\boldsymbol{\beta}}] - 2\mathbf{y}^T E[\mathbf{X}\hat{\boldsymbol{\beta}}] \\ &= \text{Var}[\mathbf{y}] + \text{Var}[\mathbf{X}\hat{\boldsymbol{\beta}}] + (\mathbf{X}\boldsymbol{\beta} - E[\mathbf{X}\hat{\boldsymbol{\beta}}])^T (\mathbf{X}\boldsymbol{\beta} - E[\mathbf{X}\hat{\boldsymbol{\beta}}]) \\ &= \sigma_{\epsilon}^2 + \text{Var}[\mathbf{X}\hat{\boldsymbol{\beta}}] + \text{Bias}[\mathbf{X}\hat{\boldsymbol{\beta}}]^2, \end{aligned}$$

da der gælder, at  $\text{Var}[\mathbf{y}] = E[(\mathbf{y} - E[\mathbf{y}])(\mathbf{y} - E[\mathbf{y}])^T] = E[(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T] = E[\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T] = \text{Var}[\boldsymbol{\epsilon}] + E[\boldsymbol{\epsilon}]^T E[\boldsymbol{\epsilon}] = \sigma_{\epsilon}^2$ .  $\square$



## C INFORMATIONSKRITERIER

---

I dette afsnit introduceres Akaikes informationskriterium og Bayesian informationskriterium, som kan anvendes til at udvælge den bedste model. Afsnittet er skrevet ud fra s. 52-53 i Robert H. Shumway & David S. Stoffer (2011).

Lad os betragte en regressions model med normalfordelte residualer,  $k$  koefficienter og maksimum likelihood estimatoren for variansen defineret som

$$\hat{\sigma}_p^2 = \frac{\text{SSR}_p}{n}, \quad (\text{C.1})$$

hvor  $\text{SSR}_p$  betegner summen af kvadrede fejl under modellen med  $p$  regressions koefficienter og  $n$  er antallet af observationer i tidsrækken. Modellen med den laveste AIC, AICc eller BIC vælges, da dette indikerer, at modellen giver en god tilnærmelse af data i forhold til modellens kompleksitet.

### Definition C.1: Akaikes informationskriterium (AIC)

$$\text{AIC} = \log \hat{\sigma}_p^2 + \frac{n + 2p}{n},$$

hvor  $\hat{\sigma}_p^2$  er givet ved (C.1),  $p$  er antallet af parametre i modellen og  $n$  er antallet af observationer i tidsrækken.

Hvis vi tilføjer uendelig mange parametre til modellen, da vil første led  $\log \hat{\sigma}_p^2$  blive vilkårligt lille, mens sidste led  $(n + 2p)/n$  øges. Hermed har vi en overtilpassede model. Ledet  $(n + 2p)/n$  er et såkaldt straffed, som forhindrer modellen i overfitting.

En lille variation fra AIC er AICc.

### Definition C.2: AICc

$$\text{AICc} = \log \hat{\sigma}_p^2 + \frac{n + p}{n - p - 2},$$

med samme notation som i definition C.1.

Forskellen mellem AIC, AICc og BIC er blot det såkaldte straffed.

### Definition C.3: Bayesian informationskriterium (BIC)

$$\text{BIC} = \log \hat{\sigma}_p^2 + \frac{p \log n}{n},$$

med samme notation som i definition C.1.

Straffedet for BIC er strengere end straffedet for AIC. For  $n \geq 8$  da vil  $\frac{p \log(n)}{n} > \frac{n+2p}{n}$ . Dermed foretrækkes BIC for mindre modeller end AIC.



## D DATA

---

I dette appendiks har vi vores data, som vi har fra Federal Reserve Bank of St. Louis. De variable, som er markeret med lyserød er dem vi har fjernet pga. NAs. Variablerne, som er markeret med grøn, har vi manuelt fravalgt. Datasættet er givet i grupperinger.

### NIPA

GDPC96	GDP	Real Gross Domestic Product, 3 Decimal (Billions of Chained 2009 Dollars)
PCECC96	Consumption	Real Personal Consumption Expenditures (Billions of Chained 2009 Dollars)
PCDGx	Cons:Dur	Real personal consumption expenditures: Durable goods (Billions of Chained 2009 Dollars), deflated using PCE
PCESVx	Cons:Svc	Real Personal Consumption Expenditures: Services (Billions of 2009 Dollars), deflated using PCE
PCNDx	Cons:NonDur	Real Personal Consumption Expenditures: Nondurable Goods (Billions of 2009 Dollars), deflated using PCE
GPDIC96	Investment	Real Gross Private Domestic Investment, 3 decimal (Billions of Chained 2009 Dollars)
FPIx	FixedInv	Real private fixed investment (Billions of Chained 2009 Dollars), deflated using PCE
Y033RC1Q027SBEAx	Inv:Equip&Software	Real Gross Private Domestic Investment: Fixed Investment: Nonresidential: Equipment (Billions of Chained 2009 Dollars), deflated using PCE
PNFIx	FixInv:NonRes	Real private fixed investment: Nonresidential (Billions of Chained 2009 Dollars), deflated using PCE
PRFIx	FixedInv:Res	Real private fixed investment: Residential (Billions of Chained 2009 Dollars), deflated using PCE
A014RE1Q156NBEA	Inv:Inventories	Shares of gross domestic product: Gross private domestic investment: Change in private inventories (Percent)
GCEC96	Gov:Spending	Real Government Consumption Expenditures & Gross Investment (Billions of Chained 2009 Dollars)
A823RL1Q225SBEA	Gov:Fed	Real Government Consumption Expenditures and Gross Investment: Federal (Percent Change from Preceding Period)
FGRECPTx	Real Gov Receipts	Real Federal Government Current Receipts (Billions of Chained 2009 Dollars), deflated using PCE
SLCEx	Gov:State&Local	Real government state and local consumption expenditures (Billions of Chained 2009 Dollars), deflated using PCE
EXPGSC96	Exports	Real Exports of Goods & Services, 3 Decimal (Billions of Chained 2009 Dollars)
IMPGSC96	Imports	Real Imports of Goods & Services, 3 Decimal (Billions of Chained 2009 Dollars)
DPIC96	Disp-Income	Real Disposable Personal Income (Billions of Chained 2009 Dollars)
OUTNFB	Output:NFB	Nonfarm Business Sector: Real Output (Index 2009=100)
OUTBS	Output:Bus	Business Sector: Real Output (Index 2009=100)
OUTMS	Output:Manuf	Manufacturing Sector: Real Output (Index 2009=100)
B020RE1Q156NBEA		Shares of gross domestic product: Exports of goods and services (Percent)
B021RE1Q156NBEA		Shares of gross domestic product: Imports of goods and services (Percent)

## Industrial Production

INDPRO	IP:Total index	Industrial Production Index (Index 2012=100)
IPFINAL	IP:Final products	Industrial Production: Final Products (Market Group) (Index 2012=100)
IPCONGD	IP:Consumer goods	Industrial Production: Consumer Goods (Index 2012=100)
IPMAT	IP:Materials	Industrial Production: Materials (Index 2012=100)
IPDMAT	IP:Dur gds materials	Industrial Production: Durable Materials (Index 2012=100)
IPNMAT	IP:Nondur gds materials	Industrial Production: Nondurable Materials (Index 2012=100)
IPDCONGD	IP:Dur Cons. Goods	Industrial Production: Durable Consumer Goods (Index 2012=100)
IPB51110SQ	IP:Auto	Industrial Production: Durable Goods: Automotive products (Index 2012=100)
IPNCONGD	IP:NonDur Cons God	Industrial Production: Nondurable Consumer Goods (Index 2012=100)
IPBUSEQ	IP:Bus Equip	Industrial Production: Business Equipment (Index 2012=100)
IPB51220SQ	IP:Energy Prds	Industrial Production: Consumer energy products (Index 2012=100)
TCU	Capu Tot	Capacity Utilization: Total Industry (Percent of Capacity)
CUMFNS	Capu Man.	Capacity Utilization: Manufacturing (SIC) (Percent of Capacity)
IPMANSICS		Industrial Production: Manufacturing (SIC) (Index 2012=100)
IPB51222S		Industrial Production: Residential Utilities (Index 2012=100)
IPFUELS		Industrial Production: Fuels (Index 2012=100)
NAPMPI		ISM Manufacturing: Production Index
NAPM		ISM Manufacturing: PMI Composite Index

## Employment and Unemployment

PAYEMS	Emp:Nonfarm	All Employees: Total nonfarm (Thousands of Persons)
USPRIV	Emp:Private	All Employees: Total Private Industries (Thousands of Persons)
MANEMP	Emp:mfg	All Employees: Manufacturing (Thousands of Persons)
SRVPRD	Emp:Services	All Employees: Service-Providing Industries (Thousands of Persons)
USGOOD	Emp:Goods	All Employees: Goods-Producing Industries (Thousands of Persons)
DMANEMP	Emp:DurGoods	All Employees: Durable goods (Thousands of Persons)
NDMANEMP	Emp:Nondur Goods	All Employees: Nondurable goods (Thousands of Persons)
USCONS	Emp:Const	All Employees: Construction (Thousands of Persons)
USEHS	Emp:Edu&Health	All Employees: Education & Health Services (Thousands of Persons)
USFIRE	Emp:Finance	All Employees: Financial Activities (Thousands of Persons)
USINFO	Emp:Infor	All Employees: Information Services (Thousands of Persons)
USPBS	Emp:Bus Serv	All Employees: Professional & Business Services (Thousands of Persons)
USLAH	Emp:Leisure	All Employees: Leisure & Hospitality (Thousands of Persons)
USSERV	Emp:OtherSvcs	All Employees: Other Services (Thousands of Persons)
USMINE	Emp:Mining/NatRes	All Employees: Mining and logging (Thousands of Persons)
USTPU	Emp:Trade&Trans	All Employees: Trade, Transportation & Utilities (Thousands of Persons)
USGOVT	Emp:Gov	All Employees: Government (Thousands of Persons)
USTRADE	Emp:Retail	All Employees: Retail Trade (Thousands of Persons)
USWTRADE	Emp:Wholesal	All Employees: Wholesale Trade (Thousands of Persons)
CES9091000001	Emp:Gov(Fed)	All Employees: Government: Federal (Thousands of Persons)
CES9092000001	Emp:Gov (State)	All Employees: Government: State Government (Thousands of Persons)
CES9093000001	Emp:Gov (Local)	All Employees: Government: Local Government (Thousands of Persons)
CE16OV	Emp:Total (HHSurve)	Civilian Employment (Thousands of Persons)
CIVPART	LF Part Rate	Civilian Labor Force Participation Rate (Percent)
UNRATE	Unemp Rate	Civilian Unemployment Rate (Percent)
UNRATESTx	Urate-ST	Unemployment Rate less than 27 weeks (Percent)
UNRATELTx	Urate-LT	Unemployment Rate for more than 27 weeks (Percent)
LNS14000012	Urate:Age16-19	Unemployment Rate - 16 to 19 years (Percent)
LNS14000025	Urate:Age>20 Men	Unemployment Rate - 20 years and over, Men (Percent)
LNS14000026	Urate:Age>20 Women	Unemployment Rate - 20 years and over, Women (Percent)
UEMPLT5	U:Dur<5wks	Number of Civilians Unemployed - Less Than 5 Weeks (Thousands of Persons)
UEMP5TO14	U:Dur5-14wks	Number of Civilians Unemployed for 5 to 14 Weeks (Thousands of Persons)
UEMP15T26	U:dur>15-26wks	Number of Civilians Unemployed for 15 to 26 Weeks (Thousands of Persons)
UEMP27OV	U:Dur>27wks	Number of Civilians Unemployed for 27 Weeks and Over (Thousands of Persons)
LNS13023621	U:Job losers	Unemployment Level - Job Losers (Thousands of Persons)
LNS13023557	U:LF Reenty	Unemployment Level - Reentrants to Labor Force (Thousands of Persons)
LNS13023705	U:Job Leavers	Unemployment Level - Job Leavers (Thousands of Persons)
LNS13023569	U:New Entrants	Unemployment Level - New Entrants (Thousands of Persons)

LNS12032194	Emp:SlackWk	Employment Level - Part-Time for Economic Reasons, All Industries (Thousands of Persons)
HOABS	EmpHrs:Bus Sec	Business Sector: Hours of All Persons (Index 2009=100)
HOAMS	EmpHrs:mfg	Manufacturing Sector: Hours of All Persons (Index 2009=100)
HOANBS	EmpHrs:nfb	Nonfarm Business Sector: Hours of All Persons (Index 2009=100)
AWHMAN	AWH Man	Average Weekly Hours of Production and Nonsupervisory Employees: Manufacturing (Hours)
AWHNONAG	AWH Privat	Average Weekly Hours Of Production And Nonsupervisory Employees: Total private (Hours)
AWOTMAN	AWH Overtime	Average Weekly Overtime Hours of Production and Nonsupervisory Employees: Manufacturing (Hours)
HWIx	HelpWnted	Help-Wanted Index
UEMPMEAN		Average (Mean) Duration of Unemployment (Weeks)
CES0600000007		Average Weekly Hours of Production and Nonsupervisory Employees: Goods-Producing
NAPMEI		ISM Manufacturing: Employment Index
HWIURATIO		Ratio of Help Wanted/No. Unemployed
CLAIMSx		Initial Claims

## Housing

HOUST	Hstarts	Housing Starts: Total: New Privately Owned Housing Units Started (Thousands of Units)
HOUST5F	Hstarts 5units	Privately Owned Housing Starts: 5-Unit Structures or More (Thousands of Units)
PERMIT	Hpermits	New Private Housing Units Authorized by Building Permits (Thousands of Units)
HOUSTMW	Hstarts:MW	Housing Starts in Midwest Census Region (Thousands of Units)
HOUSTNE	Hstarts:NE	Housing Starts in Northeast Census Region (Thousands of Units)
HOUSTS	Hstarts:S	Housing Starts in South Census Region (Thousands of Units)
HOUSTW	Hstarts:W	Housing Starts in West Census Region (Thousands of Units)
USSTHPI	Real Hprice:OFHEO	All-Transactions House Price Index for the United States (Index 1980 Q1=100)
SPCS10RSA	Real CS.10	S&P/Case-Shiller 10-City Composite Home Price Index (Index January 2000 = 100)
SPCS20RSA	Real CS.20	S&P/Case-Shiller 20-City Composite Home Price Index (Index January 2000 = 100)
PERMITNE		New Private Housing Units Authorized by Building Permits in the Northeast Census Region (Thousands, SAAR)
PERMITMW		New Private Housing Units Authorized by Building Permits in the Midwest Census Region (Thousands, SAAR)
PERMITS		New Private Housing Units Authorized by Building Permits in the South Census Region (Thousands, SAAR)
PERMITW		New Private Housing Units Authorized by Building Permits in the West Census Region (Thousands, SAAR)

## Inventories, Orders and Sales

CMRMTSPLx	MT Sales	Real Manufacturing and Trade Industries Sales (Millions of Chained 2009 Dollars)
RSAFSx	Ret. Sale	Real Retail and Food Services Sales (Millions of Chained 2009 Dollars), deflated by Core PCE
AMDNOx	Orders (DurMfg)	Real Manufacturers' New Orders: Durable Goods (Millions of 2009 Dollars), deflated by Core PCE
ACOGNOx	Orders(ConsGoods/Mat.)	Real Value of Manufacturers' New Orders for Consumer Goods Industries (Million of 2009 Dollars), deflated by Core PCE
AMDMUOx	UnfOrders(DurGds)	Real Value of Manufacturers' Unfilled Orders for Durable Goods Industries (Million of 2009 Dollars), deflated by Core PCE
ANDENOx	Orders(NonDefCap)	Real Value of Manufacturers' New Orders for Capital Goods: Nondefense Capital Goods Industries (Million of 2009 Dollars), deflated by Core PCE
NAPMSDI	VendPerf	ISM Manufacturing: Supplier Deliveries Index (lin)
INVCQRMTSPL	MT Invent	Real Manufacturing and Trade Inventories (Millions of 2009 Dollars)
NAPMNOI		ISM Manufacturing: New Orders Index
NAPMII		ISM Manufacturing: Inventories Index
BUSINVx		Total Business Inventories (Millions of Dollars)
ISRATIOx		Total Business: Inventories to Sales Ratio

## Prices

PCECTPI	PCED	Personal Consumption Expenditures: Chain-type Price Index (Index 2009=100)
PCEPILFE	PCED.LFE	Personal Consumption Expenditures Excluding Food and Energy (Chain-Type Price Index) (Index 2009=100)
GDPCTPI	GDP Defl	Gross Domestic Product: Chain-type Price Index (Index 2009=100)
GPDICTPI	GPDI Defl	Gross Private Domestic Investment: Chain-type Price Index (Index 2009=100)
IPDBS	BusSec Defl	Business Sector: Implicit Price Deflator (Index 2009=100)
DGDSRG3Q086SBEA	PCED.Goods	Personal consumption expenditures: Goods (chain-type price index)
DDURRG3Q086SBEA	PCED.DurGoods	Personal consumption expenditures: Durable goods (chain-type price index)
DSERRG3Q086SBEA	PCED.Serv	Personal consumption expenditures: Services (chain-type price index)
DNDGRG3Q086SBEA	PCED.NDurGoods	Personal consumption expenditures: Nondurable goods (chain-type price index)
DHCERG3Q086SBEA	PCED.HouseholdServ.	Personal consumption expenditures: Services: Household consumption expenditures (chain-type price index)
DMOTRG3Q086SBEA	PCED.MotorVec	Personal consumption expenditures: Durable goods: Motor vehicles and parts (chain-type price index)
DFDHRG3Q086SBEA	PCED.DurHousehold	Personal consumption expenditures: Durable goods: Furnishings and durable household equipment (chain-type price index)
DREQRG3Q086SBEA	PCED.Recreation	Personal consumption expenditures: Durable goods: Recreational goods and vehicles (chain-type price index)
DODGRG3Q086SBEA	PCED.OthDurGds	Personal consumption expenditures: Durable goods: Other durable goods (chain-type price index)
DFXARG3Q086SBEA	PCED.Food.Bev	Personal consumption expenditures: Nondurable goods: Food and beverages purchased for off-premises consumption (chain-type price index)
DCLORG3Q086SBEA	PCED.Clothing	Personal consumption expenditures: Nondurable goods: Clothing and footwear (chain-type price index)
DGOERG3Q086SBEA	PCED.Gas.Enrgy	Personal consumption expenditures: Nondurable goods: Gasoline and other energy goods (chain-type price index)
DONGRG3Q086SBEA	PCED.OthNDurGds	Personal consumption expenditures: Nondurable goods: Other nondurable goods (chain-type price index)
DHUTRG3Q086SBEA	PCED.Housing-Utilities	Personal consumption expenditures: Services: Housing and utilities (chain-type price index)
DHLCRG3Q086SBEA	PCED.HealthCare	Personal consumption expenditures: Services: Health care (chain-type price index)
DTRSRG3Q086SBEA	PCED.TransSvg	Personal consumption expenditures: Transportation services (chain-type price index)



DRCARG3Q086SBEA	PCED_RecServices	Personal consumption expenditures: Recreation services (chain-type price index)
DFSARG3Q086SBEA	PCED_FoodServ_Acc.	Personal consumption expenditures: Services: Food services and accommodations (chain-type price index)
DIFSRG3Q086SBEA	PCED_FIRE	Personal consumption expenditures: Financial services and insurance (chain-type price index)
DOTSRG3Q086SBEA	PCED_OtherServices	Personal consumption expenditures: Other services (chain-type price index)
CPIAUCSL	CPI	Consumer Price Index for All Urban Consumers: All Items (Index 1982-84=100)
CPILFESL	CPLLFE	Consumer Price Index for All Urban Consumers: All Items Less Food & Energy (Index 1982-84=100)
PPIFGS	PPI:FinGds	Producer Price Index by Commodity for Finished Goods (Index 1982=100)
PPIACO	PPI	Producer Price Index for All Commodities (Index 1982=100)
PPIFCG	PPI:FinConsGds	Producer Price Index by Commodity for Finished Consumer Goods (Index 1982=100)
PPIFCF	PPI:FinConsGds(Food)	Producer Price Index by Commodity for Finished Consumer Foods (Index 1982=100)
PPIIDC	PPI:IndCom	Producer Price Index by Commodity Industrial Commodities (Index 1982=100)
PPIITM	PPI:IntMat	Producer Price Index by Commodity Intermediate Materials: Supplies & Components (Index 1982=100)
NAPMPRI	NAPM com price	ISM Manufacturing: Prices Index (Index)
WPU0531	Real Price:NatGas	Producer Price Index by Commodity for Fuels and Related Products and Power: Natural Gas (Index 1982=100)
WPU0561	Real Price:Oil	Producer Price Index by Commodity for Fuels and Related Products and Power: Crude Petroleum (Domestic Production) (Index 1982=100)
OILPRICEx	Real Crudeoil Price	Real Crude Oil Prices: West Texas Intermediate (WTI) - Cushing, Oklahoma (2009 Dollars per Barrel), deflated by Core PCE
PPICRM		Producer Price Index: Crude Materials for Further Processing (Index 1982=100)
PPICMM		Producer Price Index: Commodities: Metals and metal products: Primary nonferrous metals (Index 1982=100)
CPIAPPSL		Consumer Price Index for All Urban Consumers: Apparel (Index 1982-84=100)
CPITRNSL		Consumer Price Index for All Urban Consumers: Transportation (Index 1982-84=100)
CPIMEDSL		Consumer Price Index for All Urban Consumers: Medical Care (Index 1982-84=100)
CUSR0000SAC		Consumer Price Index for All Urban Consumers: Commodities (Index 1982-84=100)
CUUR0000SAD		Consumer Price Index for All Urban Consumers: Durables (Index 1982-84=100)
CUSR0000SAS		Consumer Price Index for All Urban Consumers: Services (Index 1982-84=100)
CPIULFSL		Consumer Price Index for All Urban Consumers: All Items Less Food (Index 1982-84=100)
CUUR0000SA0L2		Consumer Price Index for All Urban Consumers: All items less shelter (Index 1982-84=100)
CUSR0000SA0L5		Consumer Price Index for All Urban Consumers: All items less medical care (Index 1982-84=100)
CUSR0000SEHC		CPI for All Urban Consumers: Owners' equivalent rent of residences (Index Dec 1982=100)

## Earnings and Productivity

AHETPIx	Real AHE:PrivInd	Real Average Hourly Earnings of Production and Nonsupervisory Employees: Total Private (2009 Dollars per Hour), deflated by Core PCE
CES2000000008x	Real AHE:Const	Real Average Hourly Earnings of Production and Nonsupervisory Employees: Construction (2009 Dollars per Hour), deflated by Core PCE
CES3000000008x	Real AHE:MFG	Real Average Hourly Earnings of Production and Nonsupervisory Employees: Manufacturing (2009 Dollars per Hour), deflated by Core PCE
COMPRMS	CPH:Mfg	Manufacturing Sector: Real Compensation Per Hour (Index 2009=100)
COMPRNFB	CPH:NFB	Nonfarm Business Sector: Real Compensation Per Hour (Index 2009=100)
RCPHBS	CPH:Bus	Business Sector: Real Compensation Per Hour (Index 2009=100)
OPHMFG	OPH:mfg	Manufacturing Sector: Real Output Per Hour of All Persons (Index 2009=100)
OPHNFB	OPH:nfb	Nonfarm Business Sector: Real Output Per Hour of All Persons (Index 2009=100)
OPHPBS	OPH:Bus	Business Sector: Real Output Per Hour of All Persons (Index 2009=100)
ULCBS	ULC:Bus	Business Sector: Unit Labor Cost (Index 2009=100)
ULCMFG	ULC:Mfg	Manufacturing Sector: Unit Labor Cost (Index 2009=100)
ULCNFB	ULC:NFB	Nonfarm Business Sector: Unit Labor Cost (Index 2009=100)
UNLPNBS	UNLPay:nfb	Nonfarm Business Sector: Unit Nonlabor Payments (Index 2009=100)
CES0600000008		Average Hourly Earnings of Production and Nonsupervisory Employees: Goods-Producing (Dollars per Hour)

## Interest Rates

FEDFUNDS	FedFunds	Effective Federal Funds Rate (Percent)
TB3MS	TB-3Mth	3-Month Treasury Bill: Secondary Market Rate (Percent)
TB6MS	TM-6MTH	6-Month Treasury Bill: Secondary Market Rate (Percent)
MED3	EuroDol3M	3-Month Eurodollar Deposit Rate (London) (Percent)
GS1	TB-1YR	1-Year Treasury Constant Maturity Rate (Percent)
GS10	TB-10YR	10-Year Treasury Constant Maturity Rate (Percent)
MORTG	Mort-30Yr	30-Year Conventional Mortgage Rate (Percent)
AAA	AAA Bond	Moody's Seasoned Aaa Corporate Bond Yield (Percent)
BAA	BAA Bond	Moody's Seasoned Baa Corporate Bond Yield (Percent)
BAA10YM	BAA.GS10	Moody's Seasoned Baa Corporate Bond Yield Relative to Yield on 10-Year Treasury Constant Maturity (Percent)
MORTG10YRx	MRTG.GS10	30-Year Conventional Mortgage Rate Relative to 10-Year Treasury Constant Maturity (Percent)
TB6M3Mx	tb6m.tb3m	6-Month Treasury Bill Minus 3-Month Treasury Bill, secondary market (Percent)
GS1TB3Mx	GS1.tb3m	1-Year Treasury Constant Maturity Minus 3-Month Treasury Bill, secondary market (Percent)
GS10TB3Mx	GS10.tb3m	10-Year Treasury Constant Maturity Minus 3-Month Treasury Bill, secondary market (Percent)
CPF3MTB3Mx	CP_TBbill Spread	3-Month Commercial Paper Minus 3-Month Treasury Bill, secondary market (Percent)
MED3TB3Mx	Ted_spr	3-Month Eurodollar Deposit Minus 3-Month Treasury Bill, secondary market (Percent)
GS5		5-Year Treasury Constant Maturity Rate
TB3SMFFM		3-Month Treasury Constant Maturity Minus Federal Funds Rate
T5YFFM		5-Year Treasury Constant Maturity Minus Federal Funds Rate
AAAFFM		Moody's Seasoned Aaa Corporate Bond Minus Federal Funds Rate
CP3M		3-Month AA Financial Commercial Paper Rate
COMPAPFF		3-Month Commercial Paper Minus Federal Funds Rate

## Money and Credit

AMBSLREALx	Real Mbase	St. Louis Adjusted Monetary Base (Billions of 1982-84 Dollars), deflated by CPI
IMFSLx	Real InsMMF	Real Institutional Money Funds (Billions of 2009 Dollars), deflated by Core PCE
MIREALx	Real m1	Real M1 Money Stock (Billions of 1982-84 Dollars), deflated by CPI
M2REALx	Real m2	Real M2 Money Stock (Billions of 1982-84 Dollars), deflated by CPI
MZMREALx	Real mzm	Real MZM Money Stock (Billions of 1982-84 Dollars), deflated by CPI
BUSLOANSx	Real C&Iloand	Real Commercial and Industrial Loans, All Commercial Banks (Billions of 2009 U.S. Dollars), deflated by Core PCE
CONSUMERx	Real ConsLoans	Real Consumer Loans at All Commercial Banks (Billions of 2009 U.S. Dollars), deflated by Core PCE
NONREVSLx	Real NonRevCredit	Total Real Nonrevolving Credit Owned and Securitized, Outstanding (Billions of Dollars), deflated by Core PCE
REALLNx	Real LoansRealEst	Real Real Estate Loans, All Commercial Banks (Billions of 2009 U.S. Dollars), deflated by Core PCE
REVOLSLx	Real RevolvCredit	Total Real Revolving Credit Owned and Securitized, Outstanding (Billions of 2009 Dollars), deflated by Core PCE
TOTALSLx	Real ConsuCred	Total Consumer Credit Outstanding, deflated by Core PCE
DRJWCIL	FRBSLO.Consumers	FRB Senior Loans Officer Opions. Net Percentage of Domestic Respondents Reporting Increased Willingness to Make Consumer Installment Loans
TOTRESNS		Total Reserves of Depository Institutions (Billions of Dollars)
NONBORRES		Reserves Of Depository Institutions, Nonborrowed (Millions of Dollars)
DTCOLNVHFM		Consumer Motor Vehicle Loans Outstanding Owned by Finance Companies (Millions of Dollars)
DTCTHFM		Total Consumer Loans and Leases Outstanding Owned and Securitized by Finance Companies (Millions of Dollars)
INVEST		Securities in Bank Credit at All Commercial Banks (Billions of Dollars)

## Household Balance Sheets

TABSHNOx	Real HHW:TASA	Real Total Assets of Households and Nonprofit Organizations (Billions of 2009 Dollars), deflated by Core PCE
TLBSHNOx	Real HHW:LiabSA	Real Total Liabilities of Households and Nonprofit Organizations (Billions of 2009 Dollars), deflated by Core PCE
LIABPIx	liab.PDISA	Liabilities of Households and Nonprofit Organizations Relative to Personal Disposable Income (Percent)
TNWBSHNOx	Real HHW:WSA	Real Net Worth of Households and Nonprofit Organizations (Billions of 2009 Dollars), deflated by Core PCE
NWPIx	W.PDISA	Net Worth of Households and Nonprofit Organizations Relative to Disposable Personal Income (Percent)
TARESAX	Real HHW:TA_RESAX	Real Assets of Households and Nonprofit Organizations excluding Real Estate Assets (Billions of 2009 Dollars), deflated by Core PCE
HNOREMQ027Sx	Real HHW:RESA	Real Real Estate Assets of Households and Nonprofit Organizations (Billions of 2009 Dollars), deflated by Core PCE
TFAABSHNOx	Real HHW:FinSA	Real Total Financial Assets of Households and Nonprofit Organizations (Billions of 2009 Dollars), deflated by Core PCE
CONSPI		Nonrevolving consumer credit to Personal Income

## Exchange Rates

TWEXMMTH	Ex rate:major	Trade Weighted U.S. Dollar Index: Major Currencies (Index March 1973=100)
EXUSEU	Ex rate:Euro	U.S. / Euro Foreign Exchange Rate (U.S. Dollars to One Euro)
EXSZUSx	Ex rate:Switz	Switzerland / U.S. Foreign Exchange Rate
EXJPUSx	Ex rate:Japan	Japan / U.S. Foreign Exchange Rate
EXUSUKx	Ex rate:UK	U.S. / U.K. Foreign Exchange Rate
EXCAUSx	EX rate:Canada	Canada / U.S. Foreign Exchange Rate

## Other

UMCSENTx	Cons. Expectations	University of Michigan: Consumer Sentiment (Index 1st Quarter 1966=100)
USEPUINDEXM	PolicyUncertainty	Economic Policy Uncertainty Index for United States

## Stock Markets

VXOCLSX	VXO	CBOE S&P 100 Volatility Index: VXO
NIKKEI225		Nikkei Stock Average
NASDAQCOM		NASDAQ Composite (Index Feb 5, 1971=100)
S&P 500		S&P's Common Stock Price Index: Composite
S&P: indust		S&P's Common Stock Price Index: Industrials
S&P: div yield		S&P's Composite Common Stock: Dividend Yield
S&P PE ratio		S&P's Composite Common Stock: Price-Earnings Ratio

## Non-Household Balance Sheets

GFDEGDQ188S	Federal Debt: Total Public Debt as Percent of GDP (Percent)
GFDEBTNx	Real Federal Debt: Total Public Debt (Millions of 2009 Dollars), deflated by PCE
TLBSNNCBx	Real Nonfinancial Corporate Business Sector Liabilities (Billions of 2009 Dollars), Deflated by Implicit Price Deflator for Business Sector IPDBS
TLBSNNCBBDIx	Nonfinancial Corporate Business Sector Liabilities to Disposable Business Income (Percent)
TTAABSNNCBx	Real Nonfinancial Corporate Business Sector Assets (Billions of 2009 Dollars), Deflated by Implicit Price Deflator for Business Sector IPDBS
TNWMVBSNNCBx	Real Nonfinancial Corporate Business Sector Net Worth (Billions of 2009 Dollars), Deflated by Implicit Price Deflator for Business Sector IPDBS
TNWMVBSNNCBBDIx	Nonfinancial Corporate Business Sector Net Worth to Disposable Business Income (Percent)
NNBTILQ027Sx	Real Nonfinancial Noncorporate Business Sector Liabilities (Billions of 2009 Dollars), Deflated by Implicit Price Deflator for Business Sector IPDBS
NNBTILQ027SBDIx	Nonfinancial Noncorporate Business Sector Liabilities to Disposable Business Income (Percent)
NNBTASQ027Sx	Real Nonfinancial Noncorporate Business Sector Assets (Billions of 2009 Dollars), Deflated by Implicit Price Deflator for Business Sector IPDBS
TNWSNNBx	Real Nonfinancial Noncorporate Business Sector Net Worth (Billions of 2009 Dollars), Deflated by Implicit Price Deflator for Business Sector IPDBS
TNWSNNBBDIx	Nonfinancial Noncorporate Business Sector Net Worth to Disposable Business Income (Percent)
CNCFx	Real Disposable Business Income, Billions of 2009 Dollars (Corporate cash flow with IVA minus taxes on corporate income, deflated by Implicit Price Deflator for Business Sector IPDBS)

## E R-KODER

De følgende koder er lavet til étkvartals prædiktions af vækst i BNP med de metoder beskrevet i projektet. Datasættet `cuttedData` har alle forklarende variable, mens `cuttedData_mvs` (*Manual Variable Selection*) er datasættet hvor vi selv har valgt variable fra. Begge består af 174 observationer. Initialtræningssættet er de første 134 observationer, og testsættet er de sidste 40 observationer. For mere detaljer henvises til det tilførende R-script.

### E.0.1 Benchmark prædiktions

```
1 ## best order AR(p) model forward lag selection with AIC, AICc & BIC
2 AICCBIC = function(X_orig, pmax, AIC = F, BIC = F, AICc = F){
3   BIC_vektor = c(NA)
4   AICc_vektor = c(NA)
5   AIC_vektor = c(NA)
6
7   for(p in 1:pmax){
8     respons = X_orig[(pmax + 1):length(X_orig)]
9     n = length(respons)
10    Xlag = matrix(nrow = n, ncol = p)
11    for(j in 1:p){
12      for(i in 1:n){
13        Xlag[i, j] = X_orig[p+i-j]
14      }
15    }
16    design.matrix = cbind(rep(1, n), Xlag)
17    betahat = solve(t(design.matrix) %*% design.matrix) %*% t(design.matrix) %*%
      respons
18    estimerede.vaerdier = design.matrix %*% betahat
19    SSR = sum((respons - estimerede.vaerdier)^2)
20    sigma2hat = SSR/n
21    if(BIC){
22      BIC = log(sigma2hat) + (p * log(n)/n)
23      BIC_vektor[p] = BIC
24    }
25    if(AICc){
26      AICc = log(sigma2hat) + ((n + p)/(n - p - 2))
27      AICc_vektor[p] = AICc
28    }
29    if(AIC){
30      AIC = log(sigma2hat) + ((n + 2 * p)/n)
31      AIC_vektor[p] = AIC
32    }
33  }
34  if(BIC){
35    min_BIC = which(BIC_vektor == min(BIC_vektor))
36    return(min_BIC)
37  }
38  if(AICc){
39    min_AICc = which(AICc_vektor == min(AICc_vektor))
40    return(min_AICc)
41  }
42  if(AIC){
43    min_AIC = which(AIC_vektor == min(AIC_vektor))
44    return(min_AIC)
45  }
```

```

46 }
47 ## for different output purposes
48 betahat = function(X_orig, p, output){
49   respons = X_orig[(p+1):length(X_orig)]
50   n = length(respons)
51
52   Xlag = matrix(nrow = n, ncol = p)
53   for(j in 1:p){
54     for(i in 1:n){
55       Xlag[i, j] = X_orig[p+i-j]
56     }
57   }
58   design.matrix = cbind(rep(1, n), Xlag)
59   betahat = solve(t(design.matrix) %*% design.matrix) %*% t(design.matrix) %*%
     respons
60   estimerede.vaerdier = design.matrix %*% betahat
61   SSR = sum((respons - estimerede.vaerdier)^2)
62   sigma2hat = SSR/n
63   if(output == "betahat"){
64     return(betahat)
65   }
66   if(output == "sigma2hat"){
67     return(sigma2hat)
68   }
69   if(output == "estimerede.vaerdier"){
70     return(estimerede.vaerdier)
71   }
72 }
73
74 ## one-step-ahead forecast with benchmark AR(p) forward lag selection
75 forecast_AICCBIC = function(training_saet = BNP_training, X_orig = BNP, pmax, AIC
     = F, AICc = F, BIC = F, conf, PI){
76   if(AIC){
77     traenings_saet_forecast_AIC = c(NA)
78   }
79   if(AICc){
80     traenings_saet_forecast_AICc = c(NA)
81   }
82   if(BIC){
83     traenings_saet_forecast_BIC = c(NA)
84   }
85   upper_PI = c(NA)
86   lower_PI = c(NA)
87   for(i in length(training_saet):(length(X_orig)-1)){
88     training_saet1 = X_orig[1:i]
89     p = AICCBIC(training_saet1, pmax, AIC, BIC, AICc)
90     betahat1 = betahat(training_saet1, p, output = "betahat")
91     sigma2hat = betahat(training_saet1, p, output = "sigma2hat")
92     if(AIC){
93       traenings_saet_forecast_AIC[i-length(training_saet)+1] = c(1, training_
         saet1[i:(i-p+1)]) %*% betahat1
94     }
95     if(AICc){
96       traenings_saet_forecast_AICc[i-length(training_saet)+1] = c(1, training_
         saet1[i:(i-p+1)]) %*% betahat1
97     }
98     if(BIC){
99       traenings_saet_forecast_BIC[i-length(training_saet)+1] = c(1, training_
         saet1[i:(i-p+1)]) %*% betahat1
100     upper_PI[i-length(training_saet)+1] = traenings_saet_forecast_BIC[i-length(
         training_saet)+1] + 1.96 * sqrt(sigma2hat)
101     lower_PI[i-length(training_saet)+1] = traenings_saet_forecast_BIC[i-length(
         training_saet)+1] - 1.96 * sqrt(sigma2hat)

```

```

102 }
103 }
104 if(AIC & conf != TRUE){
105   return(traenings_saet_forecast_AIC)
106 }
107 if(AICc & conf != TRUE){
108   return(traenings_saet_forecast_AICc)
109 }
110 if(BIC & conf != TRUE){
111   return(traenings_saet_forecast_BIC)
112 }
113 if(PI == "upper" & conf == TRUE){
114   return(upper_PI)
115 }
116 if(PI == "lower" & conf == TRUE){
117   return(lower_PI)
118 }
119 }

```

### E.0.2 Prædiktion med AR( $p_{AL}$ )

```

1 ## one-step-ahead forecast AR(p) lags chosen with adaptive lasso
2 AR_AL_forecast = function(traening_saet = BNP_traening, X_orig = BNP, pmax,
3   output = FALSE){
4   out_of_sample_length = length(X_orig) - length(traening_saet)
5   forecast_AR_AL = c(NA)
6   for(k in 0:out_of_sample_length){
7     respons = X_orig[(pmax+1):(length(traening_saet)+k)]
8     n = length(respons)
9     Xlag = matrix(nrow = n, ncol = pmax)
10    for(j in 1:pmax){
11      for(i in 1:n){
12        Xlag[i,j] = X_orig[pmax+i-j]
13      }
14    }
15    betahat = solve(t(Xlag) %*% Xlag) %*% t(Xlag) %*% respons
16    weights = 1/abs(betahat)
17    fit = glmnet(Xlag, respons, alpha = 1, penalty.factor = weights, intercept =
18      TRUE)
19    design.matrix = cbind(rep(1,n), Xlag)
20    beta.matrix = data.matrix(coef(fit))
21    rownames(beta.matrix) = NULL
22    BIC_vektor = c(NA)
23    for(l in 1:dim(beta.matrix)[2]){
24      estimated.values = design.matrix %*% beta.matrix[,l]
25      p = length(which(beta.matrix[,l] != 0))
26      SSR = sum((respons - estimated.values)^2)
27      sigma2hat = SSR/n
28      BIC = log(sigma2hat) + (p * log(n)/n)
29      BIC_vektor[l] = BIC
30    }
31    lambda_best = which.min(BIC_vektor)
32    best_coef = beta.matrix[,lambda_best]
33    nonzero_coef_index = which(best_coef != 0)
34    nonzero_coef_index_wo_intercept = (nonzero_coef_index-1)[-1]
35    if(k < out_of_sample_length){
36      forecast_AR_AL[k+1] = c(1, X_orig[(length(BNP_traening)+k):(length(BNP_
37        traening)+k-pmax+1)]) %*% best_coef
38    }
39    if(k == 40 & output == "estimerede.vaerdier"){

```



```

37     estimerede.vaerdier = design.matrix %*% best_coef
38   }
39   cat("k = ", k, "non zero lags = ", nonzero_coef_index_wo_intercept, "\n")
40 }
41 if(output != "estimerede.vaerdier"){
42   return(forecast_AR_AL)
43 }
44 if(output == "estimerede.vaerdier"){
45   return(estimerede.vaerdier)
46 }
47 }

```

### E.0.3 Ridge prædiktio

```

1  ## one-step-ahead forecast ridge mvs
2  forecast_ridge_mvs = c(NA)
3  for(i in 0:40){
4    BNP_traening = BNP[2:(134+i)]
5    explanatory_variables = cuttedData_mvs[1:(133+i), ]
6    fit = glmnet(explanatory_variables, BNP_traening, alpha = 0, intercept = TRUE)
7    design.matrix = cbind(rep(1, dim(explanatory_variables)[1]), explanatory_
8      variables)
9    beta.matrix = data.matrix(coef(fit))
10   rownames(beta.matrix) = NULL
11   BIC_vektor = c(NA)
12   for(j in 1:dim(beta.matrix)[2]){
13     estimated.values = design.matrix %*% beta.matrix[,j]
14     n = length(BNP_traening)
15     p = dim(design.matrix)[2]
16     H = design.matrix %*% solve(t(design.matrix) %*% design.matrix + (length(BNP
17       _traening)/2)*fit$lambda[j] * diag(p)) %*% t(design.matrix)
18     df = matrix.trace(H)
19     SSR = sum((BNP_traening - estimated.values)^2)
20     sigma2hat = SSR/n
21     BIC = log(sigma2hat) + df * log(n)/n
22     BIC_vektor[j] = BIC
23   }
24   lambda_best = which.min(BIC_vektor)
25   best_coef = beta.matrix[,lambda_best]
26   if(i == 40){
27     estimerede.vaerdier = design.matrix %*% best_coef
28   }
29   if(i < 40){
30     forecast_ridge_mvs[(i+1)] = c(1, cuttedData_mvs[dim(explanatory_variables)
31       [1]+1, ]) %*% best_coef
32   }
33   cat("i = ", i, "\n")
34 }

```

### E.0.4 Lasso prædiktio

```

1  ## one-step-ahead forecast lasso mvs
2  forecast_lasso_mvs = c(NA)
3  for(i in 0:40){
4    BNP_traening = BNP[2:(134+i)]
5    explanatory_variables = cuttedData_mvs[1:(133+i), ]

```



```

6  fit = glmnet(explanatory_variables, BNP_traening, alpha = 1, intercept = TRUE)
7  design.matrix = cbind(rep(1, dim(explanatory_variables)[1]), explanatory_
    variables)
8  beta.matrix = data.matrix(coef(fit))
9  rownames(beta.matrix) = NULL
10 BIC_vektor = c(NA)
11 for(j in 1:dim(beta.matrix)[2]){
12   estimated.values = design.matrix %*% beta.matrix[,j]
13   n = length(BNP_traening)
14   p = length(which(beta.matrix[,j] != 0))
15   SSR = sum((BNP_traening - estimated.values)^2)
16   sigma2hat = SSR/n
17   BIC = log(sigma2hat) + p * log(n)/n
18   BIC_vektor[j] = BIC
19 }
20 lambda_best = which.min(BIC_vektor)
21 best_coef = beta.matrix[,lambda_best]
22 nonzero_coef_index = which(best_coef != 0)
23 col_used = nonzero_coef_index - 1
24 vars_used = cuttedData_mvs[,col_used]
25 colnames = colnames(vars_used)
26 if(i < 40){
27   forecast_lasso_mvs[(i+1)] = c(1, cuttedData_mvs[dim(explanatory_variables)
    [1]+1, ]) %*% best_coef
28 }
29 if(i == 40){
30   estimerede.vaerdier = design.matrix %*% best_coef
31 }
32 cat("i = ", i, "nonzero_coef_index = ", nonzero_coef_index, "\nvars_used = ",
    colnames, "\n")
33 }

```

### E.0.5 Adaptive lasso med OLS vægte prædiktion

```

1  ## one-step-ahead forecast adaptive lasso mvs OLS vægte
2  forecast_lasso_mvs = c(NA)
3  for(i in 0:40){
4    BNP_traening = BNP[2:(134+i)]
5    explanatory_variables = cuttedData_mvs[1:(133+i), ]
6    fit_lm = lm(BNP[2:(134+i)] ~ ., data = as.data.frame(cuttedData_mvs[1:(133+i),
    ]))
7    OLS = fit_lm$coefficients
8    weights = 1/abs(OLS)
9    fit = glmnet(explanatory_variables, BNP_traening, alpha = 1, intercept = TRUE,
    penalty.factor = weights)
10   design.matrix = cbind(rep(1, dim(explanatory_variables)[1]), explanatory_
    variables)
11   beta.matrix = data.matrix(coef(fit))
12   rownames(beta.matrix) = NULL
13   BIC_vektor = c(NA)
14   for(j in 1:dim(beta.matrix)[2]){
15     estimated.values = design.matrix %*% beta.matrix[,j]
16     n = length(BNP_traening)
17     p = length(which(beta.matrix[,j] != 0))
18     SSR = sum((BNP_traening - estimated.values)^2)
19     sigma2hat = SSR/n
20     BIC = log(sigma2hat) + p * log(n)/n
21     BIC_vektor[j] = BIC
22   }
23   lambda_best = which.min(BIC_vektor)

```

```

24 best_coef = beta.matrix[,lambda_best]
25 nonzero_coef_index = which(best_coef != 0)
26 vars = nonzero_coef_index - 1
27 vars_used = cuttedData_mvs[,vars]
28 colnames = colnames(vars_used)
29 if(i < 40){
30   forecast_lasso_mvs[(i+1)] = c(1, cuttedData_mvs[dim(explanatory_variables)
31     [1]+1, ]) %*% best_coef
32 }
33 if(i == 40){
34   estimerede.vaerdier = design.matrix %*% best_coef
35 }
36 cat("i = ", i, "nonzero_coef_index =", nonzero_coef_index, "\n colnames = ",
    colnames, "\n")
37 }

```

### E.0.6 Adaptive lasso med lasso vægte prædiktion

```

1 ## one-step-ahead forecast adaptive lasso mvs lasso vaegte
2 forecast_lasso_mvs_lasso = c(NA)
3 for(i in 0:40){
4   BNP_traening = BNP[2:(134+i)]
5   explanatory_variables = cuttedData_mvs[1:(133+i), ]
6   fit = glmnet(explanatory_variables, BNP_traening, alpha = 1, intercept = TRUE)
7   design.matrix = cbind(rep(1, dim(explanatory_variables)[1]), explanatory_
8     variables)
9   beta.matrix = data.matrix(coef(fit))
10  rownames(beta.matrix) = NULL
11  BIC_vektor = c(NA)
12  for(j in 1:dim(beta.matrix)[2]){
13    estimated.values = design.matrix %*% beta.matrix[,j]
14    n = length(BNP_traening)
15    p = length(which(beta.matrix[,j] != 0))
16    SSR = sum((BNP_traening - estimated.values)^2)
17    sigma2hat = SSR/n
18    BIC = log(sigma2hat) + p * log(n)/n
19    BIC_vektor[j] = BIC
20  }
21  lambda_best = which.min(BIC_vektor)
22  best_coef = beta.matrix[,lambda_best]
23  vars = which(best_coef != 0)
24  coef_wo_zeroes = best_coef[vars]
25  vars_first = vars - 1
26  new_expl_variables = cuttedData_mvs[1:(133+i),vars_first]
27  new_expl_variables_full = cuttedData_mvs[,vars_first]
28  weights = 1/abs(coef_wo_zeroes)
29  fit_AL = glmnet(new_expl_variables, BNP_traening, alpha = 1, penalty.factor =
30    weights, intercept = TRUE)
31  new_design.matrix = cbind(rep(1, dim(new_expl_variables)[1]), new_expl_
32    variables)
33  new_beta.matrix = data.matrix(coef(fit_AL))
34  rownames(new_beta.matrix) = NULL
35  new_BIC_vektor = c(NA)
36  for(j in 1:dim(new_beta.matrix)[2]){
37    estimated.values = new_design.matrix %*% new_beta.matrix[,j]
38    n = length(BNP_traening)
39    p = length(which(new_beta.matrix[,j] != 0))
40    SSR = sum((BNP_traening - estimated.values)^2)
41    sigma2hat = SSR/n
42    new_BIC = log(sigma2hat) + p * log(n)/n

```

```
40   new_BIC_vektor[j] = new_BIC
41 }
42 lambda_best = which.min(new_BIC_vektor)
43 new_best_coef = new_beta.matrix[,lambda_best]
44 nonzero_coef = which(new_best_coef != 0)
45 vars = nonzero_coef - 1
46 vars_used = new_expl_variables[,vars]
47 colnames = colnames(vars_used)
48 if(i < 40){
49   forecast_lasso_mvs_lasso[(i+1)] = c(1, new_expl_variables_full[(dim(new_expl
      _variables)[1]+1), i]) %% new_best_coef
50 }
51 if(i == 40){
52   estimerede.vaerdier = new_design.matrix %% new_best_coef
53 }
54 cat("i = ", i, "colnames = ", colnames, "\n")
55 }
```



## F DE 10 HYPPIGSTE VARIABLE FOR HVER METODE

---

### Lasso (mvs)

Variable	Popularitet
PCECC96	100%
FPIx	100%
PERMIT	100%
HWIURATIO	82.5%
NAPMNOI	60%
UNLPNBS	40%
M2REALx	40%
NAPMPI	40%
NIKKEI225	40%
NASDAQCOM	40%

Tabel F.1: De 10 hyppigste variable anvendt af lasso med OLS vægte til forecasting af testmængden.

### Adaptive lasso med OLS vægte (mvs)

Variable	Popularitet
OUTNFB	100%
TNWBSHNOx	100%
OPHPBS	45%
LIABPIx	45%
UNLPNBS	42.5%
AWHMAN	42.5%
PCECC96	32.5%
INDPRO	32.5%
OUTBS	27.5%
RCPHBS	22.5%

Tabel F.2: De 10 hyppigste variable anvendt af adaptive lasso med OLS vægte til forecasting af testmængden.

**Adaptive lasso med lasso vægte (mvs)**

Variable	Popularitet
FPIx	100%
PERMIT	100%
NAPMNOI	60%
UNLPNBS	40%
M2REALx	40%
NAPMPI	40%
NASDAQCOM	40%
TNWMVBSNNCBBDIx	40%
TABSHNOx	37.5%
PCECC96	32.5%

Tabel F.3: De 10 hyppigste variable anvendt af adaptive lasso med lasso vægte til forecasting af testmængden.

**Lasso som variable udvælgelse (lvs)**

Variable	Popularitet
PCECC96	100%
Y033RC1Q027SBEAx	100%
GS1TB3Mx	70%
LNS13023705	70%
HOUST	67.5%
PERMITS	62.5%
IPMAT	47.5%
PERMITNE	45%
REVOLSLx	37.5%
ULCBS	20%

Tabel F.4: De 10 hyppigste variable anvendt af lasso til forecasting af testmængden.

**Adaptive lasso med 10 lassovalgte variable (lvs)**

Variable	Popularitet
PCESVx	100%
PRFIx	100%
IPNMAT	100%
LNS12032194	100%
TOTRESNS	100%
PERMITS	100%
NIKKEI225	90%
PERMIT	75%
CPF3MTB3Mx	65%

Tabel F.5: De 9 hyppigste variable anvendt af adaptive lasso med OLS vægte til forecasting af testmængden.

**Adaptive lasso med 127 lassovalgte variable (lvs)**

Variable	Popularitet
USMINE	100%
USWTRADE	100%
MZMREALx	100%
DPIC96	97,5%
A014RE1Q156NBEA	95%
DHUTRG3Q086SBEA	95%
CLAIMSx	95%
IMPGSC96	80%
DNDGRG3Q086SBEA	67,5%
CIVPART	65%

Tabel F.6: De 10 hyppigste variable anvendt af adaptive lasso med OLS vægte til forecasting af testmængden.





# FIGURER

---

1.1	Til venstre illustreres OLS med $n = 19$ , hvor $\hat{\beta}_0 = 4.4173$ og $\hat{\beta}_1 = 0.3153$ og til højre illustreres OLS med $n = 2$ , hvor $\hat{\beta}_0 = 4.8$ og $\hat{\beta}_1 = -0.4$ . . . . .	6
1.2	Indflydelse af bias og varians på den totale fejl (Scott Fortmann-Roe, 2016). . . . .	6
1.3	Løsningssti af ridge regression med 6 simulerede variable. . . . .	8
1.4	Løsningssti af lasso med 6 simulerede variable. . . . .	9
1.5	Konturer for SSR og betingelsesområderne for ridge regression (venstre) og lasso (højre). De blå arealer er betingelsesområderne $ \beta_1  +  \beta_2  \leq t$ og $\beta_1^2 + \beta_2^2 \leq t^2$ , mens de røde ellipser er konturkurver for SSR. Konturkurverne har centrum i OLS estimatoren, $\hat{\beta}^{\text{OLS}}$ . . . . .	9
4.1	USA's reallt BNP fra d. 1. marts 1959 til 1. september 2015. . . . .	35
4.2	Transformationen $\nabla x_t$ af USA's reallt BNP fra d. 1. marts 1959 til 1. september 2015. . . . .	36
4.3	Transformationen $\nabla \log(x_t)$ af USA's reallt BNP fra d. 1. marts 1959 til 1. september 2015. . . . .	36
4.4	NAs for hver af de 257 variable i tidsperioden 1. september 1959 til 1. september 2015 svarende til 227 observationer. Den nedre røde stiplede linje markerer d. 1. september 1971, mens den øvre markerer d. 1. december 2014. . . . .	37
4.5	ACF for transformationen $\nabla \log(x_t)$ af USA's reallt BNP fra d. 1. september 1971 til d. 1. december 2004. . . . .	38
4.6	PACF for transformationen $\nabla \log(x_t)$ af USA's reallt BNP fra d. 1. september 1971 til d. 1. december 2004. . . . .	38
4.7	De givne observationer og prædiktioner af disse modelleret med $\text{AR}(p_{\text{BIC}})$ samt prædiktionsintervallet i perioden fra 1. marts 2005 til 1. december 2014. . . . .	39
4.8	Analyse af de standardiserede residualer for benchmark modellen. . . . .	40
4.9	Prædiktion af væksten for USA's reallt BNP med en autoregressiv model, hvor ordenen bestemmes udfra adaptive lasso. . . . .	41
4.10	Analyse af de standardiserede residualer af (4.1). . . . .	42
4.11	Prædiktion af væksten for USA's reallt BNP med ridge regression. . . . .	43
4.12	Analyse af de standardiserede residualer af modellen udvalgt ved ridge regression. . . . .	43
4.13	Prædiktion af væksten for USA's reallt BNP med lasso. . . . .	44
4.14	Analyse af de standardiserede residualer for (4.2). . . . .	45
4.15	Prædiktion af væksten for USA's reallt BNP med adaptive lasso, hvor vægtene bestemmes udfra OLS. . . . .	46
4.16	Analyse af de standardiserede residualer for (4.3). . . . .	46
4.17	Prædiktion af væksten for USA's reallt BNP med adaptive lasso, hvor vægtene bestemmes udfra lasso. . . . .	47
4.18	Analyse af de standardiserede residualer for (4.4). . . . .	48
4.19	Analyse af de standardiserede residualer for adaptive lasso med OLS vægte, hvor $p = 127$ . . . . .	51
4.20	Analyse af de standardiserede residualer for adaptive lasso med OLS vægte, hvor $p = 10$ . . . . .	51



# TABELLER

---

4.1	Værdier af BIC, AIC og AICc for $AR(p)$ , hvor $p = 1, \dots, 8$ for væksten af realt BNP i perioden fra d. 1. september 1971 til d. 1. december 2004. . . . .	38
4.2	MSE af $AR(p_{BIC})$ , $AR(p_{AIC})$ og $AR(p_{AICc})$ i perioden fra 1. marts 2005 til 1. december 2014. . . . .	39
4.3	$P$ -værdi fra Jarque-Bera testen og Ljung Box testen på de standardiserede residualer af benchmark modellen. . . . .	40
4.4	$P$ -værdi fra Jarque-Bera testen og Ljung Box testen på de standardiserede residualer af (4.1). . . . .	42
4.5	$P$ -værdi fra Jarque-Bera testen på de standardiserede residualer af modellen udvalgt ved ridge regression. . . . .	43
4.6	$P$ -værdi fra Jarque-Bera testen og Ljung Box testen på de standardiserede residualer af (4.2). . . . .	45
4.7	$P$ -værdi fra Jarque-Bera testen og Ljung Box testen på de standardiserede residualer for (4.3). . . . .	47
4.8	$P$ -værdi fra Jarque-Bera testen og Ljung Box testen på de standardiserede residualer af (4.4). . . . .	48
4.9	Metodernes MSE på testmængden samt MSE relativ til benchmark modellen. . .	49
4.10	$P$ -værdi fra DM testen. . . . .	49
4.11	MSE for modeller fitted med adaptive lasso modeller, hvor antallet af forklarende variable $p$ er valgt ved hjælp af lasso. . . . .	50
4.12	$P$ -værdi fra Jarque-Bera testen og Ljung Box testen på de standardiserede residualer. . . . .	52
4.13	$P$ -værdier fra DM testen. . . . .	52
F.1	De 10 hyppigste variable anvendt af lasso med OLS vægte til forecasting af testmængden. . . . .	77
F.2	De 10 hyppigste variable anvendt af adaptive lasso med OLS vægte til forecasting af testmængden. . . . .	77
F.3	De 10 hyppigste variable anvendt af adaptive lasso med lasso vægte til forecasting af testmængden. . . . .	78
F.4	De 10 hyppigste variable anvendt af lasso til forecasting af testmængden. . . . .	78
F.5	De 9 hyppigste variable anvendt af adaptive lasso med OLS vægte til forecasting af testmængden. . . . .	79
F.6	De 10 hyppigste variable anvendt af adaptive lasso med OLS vægte til forecasting af testmængden. . . . .	79



# LITTERATUR

---

- Anders Bredahl Kock (2. marts 2015). Supplementary proofs for "Consistent and Conservative Model Selection with adaptive Lasso in Stationary and Nonstationary Autoregressions. *Aarhus University and CRESTES*.
- Christian Heij & Paul de Boer & Philip Hans Franses & Teun Kloek & Herman K. van Dijk (2004). *Econometric Methods with Applications in Business and Economics*. Oxford University Press.
- Gareth James & Daniela Witten & Trevor Hastie & Robert Tibshirani (2013). *An Introduction to Statistical Learning with Applications in R*. Springer Science + Business Media New York.
- Hui Zou (2006). The Adaptive Lasso and Its Oracle Properties. *Journal of the American Statistical Association*, 101.
- Keith Knight & Wenjiang Fu (2000). Asymptotics for Lasso-type Estimators. *University of Toronto and Michigan State University*.
- Mehmet Balcilar (21. maj 2016). The Diebold-Mariano Statistic for Comparing Predictive Accuracy. <http://www.emu.edu.tr/mbalcilar/teaching2008/econ604/notes/DieboldMarino.pdf>.
- Poul Newbold, W. L. C. and Thorne, B. M. (27. maj 2016). A test of normality. <http://faculty.arts.ubc.ca/dwhistler/325ClassNotes/chapNorTestPRN.pdf>.
- Robert H. Shumway & David S. Stoffer (2011). *Time Series Analysis and Its Applications With R Examples*. Springer, Third edition.
- Scott Fortmann-Roe (25. maj 2016). Understanding the Bias-Variance Tradeoff. <http://scott.fortmann-roe.com/docs/BiasVariance.html>.