Modelforecast af mælkepris

Formål

I forlængelse af tidligere arbejde med evaluering af Seges' prisprognoser var formålet med dette projekt at udvikle en eller flere statistiske modeller til brug som led i processen omkring prisprognoser, herunder særligt mælkeprisen.

Metode

Som første skridt udvalgtes en bruttoliste på ca. 75 variable, som ud fra erfaring eller teoretiske argumenter havde potentiale til at kunne forklare den danske mælkepris på kvartalsbasis: Historiske priser, sæsonindikatorer samt en række udbuds- og efterspørgselsindikatorer.

Selve den statistiske analyse blev foretaget og dokumenteret i det statistiske programmeringssprog R med udgangspunkt i pakken "eDMA" udviklet af Catania & Nonejad (2016)[1]. Denne pakke tillader efficient udnyttelse af computerkraft ved brug af metoden Dynamic Model Averaging (DMA), som er beregningsmæssigt meget omfangsrig.

I grove træk tager Dynamic Model Averaging en mængde forklarende variable og konstruerer alle de mulige kombinationer ved at in- eller ekskludere de enkelte variable i modellen. Dette giver således 2^m mulige modeller, hvor m er antallet af forklarende variable; fra en model uden variable til at inkludere samtlige variable og alle kombinationer der imellem.

Disse mange modeller estimeres så rekursivt ved trin for trin at tilføje én ekstra periodes data og bruge den enkelte model til at forecaste prisen i næste periode. På den måde lærer modellen gennem datasættet hvilke modeller, som generelt klarer sig bedst, og vil så, når et overordnet forecast skal produceres, vægte de bedste modellers bud højest. Som nævnt er denne en udregningsmæssigt meget tung metode, og for hver ekstra variabel fordobles mængden af udregninger – dette sætter en naturlig begrænsning på, hvor mange variable, man kan anvende samtidig.

I praksis var det altså ikke muligt anvende metoden på samtlige 75 variable samtidig. Derfor blev Principal Component Analysis (PCA) anvendt til at identificere grupper af variable, som varierer sammen. Ved at udvælge det mest "typiske" variabel fra hver gruppe, blev de øvrige i første omgang siet fra, så en mindre gruppe med højt potentiale var tilbage. Sideløbende blev korrelationen mellem hver variabel og mælkeprisen lagget med 1-10 kvartaler udregnet for at finde stærke positive eller negative korrelationer, hvilket ville indikere potentiale.

Med udgangspunkt i resultaterne fra PCA og korrelationer, blev et sæt variable udvalgt til DMA. På baggrund af de af pakken oplyste sandsynligheder for at en givent variabel var inkluderet i den optimale model, blev dårligt performende variable frasorteret og nye tilføjet for at forbedre modellens performance målt både på spredningen af forecasts versus den faktiske pris, og hvorvidt forecasts ramte den rigtige retning for prisstigninger eller -fald.

[1] Catania, Leopoldo, and Nonejad, Nima. "Dynamic model averaging for practitioners in economics and finance: The eDMA package." *arXiv preprint arXiv:1606.05656* (2016).

Af totalt 15 afprøvede grupper af variable, blev den bedst performende udvalgt til at levere estimater på 1-6 kvartalers horisont fremadrettet.

Resultater

Primært skaber modellens prognose et ”alt andet lige” udgangspunkt for prognose-processen, der fremadrettet som supplement til SEGES’ ekspertise ventes at kunne forbedre nøjagtigheden af prisprognoser for mælk. I sagens natur kan modellen ikke selv tage højde for uforudsete, eksterne chok.

Undervejs er gjort et betydeligt grundarbejde med at opbygge en bagvedliggende kode, der er fleksibel og klar til brug fremadrettet. Det vil således kun være nødvendigt at fodre modellen med opdateret data for at kunne bruge den fremadrettet. Dertil kan metoden og den udviklede kode relativt nemt tilpasses til prognoser for produkter som svinekød, hvede, soya, mm., så længe relevant data er tilgængelig. Som antydet ovenfor, vil den største nøjagtighed forventeligt være til at opnå for markeder, som ikke oplever mange store chok.

Som det ses af Appendiks B, er modellens nøjagtighed som forventeligt større, jo kortere ud i fremtiden, den laver prognoser for. Således ligger en gennemsnitlig 1-kvartals prognose ca. 10 øre over eller under den faktiske pris, imens det for 6-kvartalers prognose er ca. 40 øre. Forskellene mellem SEGES’ og modellens performance er tilpas lille til, at den kan tilskrives tilfældigheder samt den ufuldkomne dækning i SEGES’ prognoser, særligt på 5-6 kvartalers horisont.

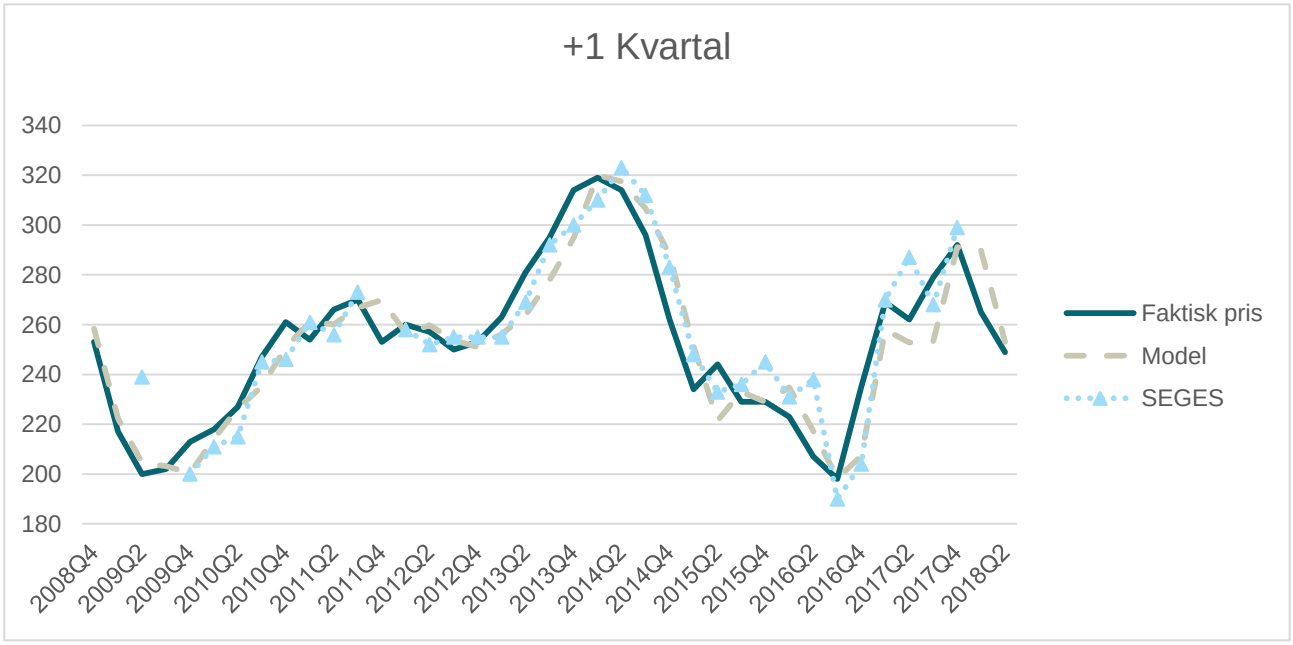
Ved inspektion af graferne i Appendiks A ses det, at modellen særligt har det svært i perioden 2013 til primo 2015, hvor prisen først stiger og så falder meget dramatisk. Det samme gælder dog for SEGES’ historiske prognoser, hvilket antyder at udsvingene har skyldtes uforudsigelige chok. I praksis indtræf meget store udsving i den kinesiske efterspørgsel og sidenhen Ruslands importembargo mod fødevarer fra EU mv.

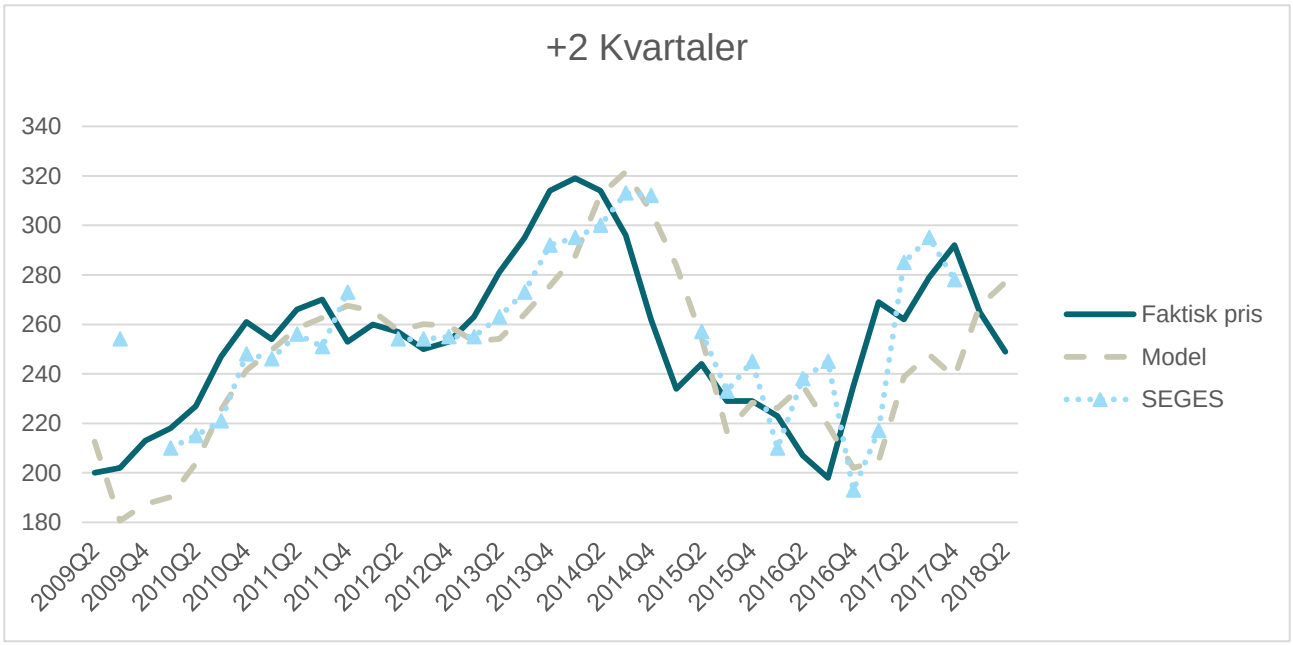
At modellen selv kan producere prognoser med stort set samme nøjagtighed som SEGES’ eksperter skaber potentiale for, at kombinationen af modellens ”neutrale” prognose og SEGES’ ekspertviden fremadrettet kan skabe endnu bedre prognoser.

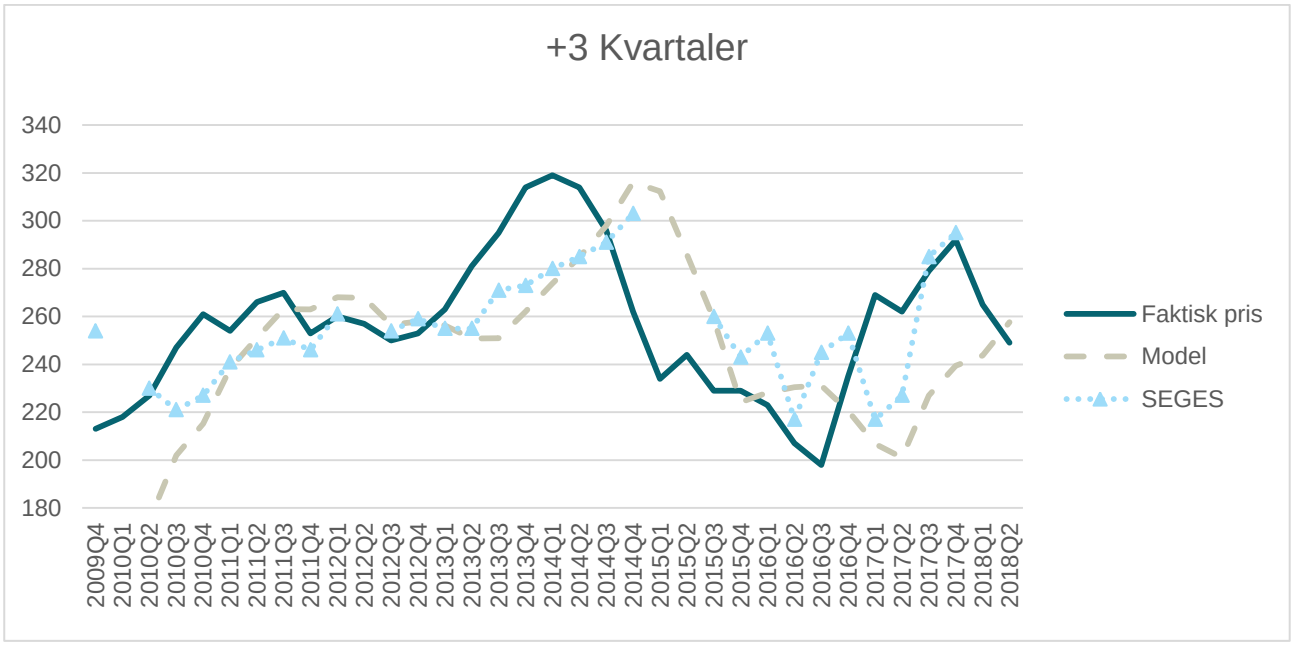
Appendiks A – Prognose vs. faktisk pris i øre

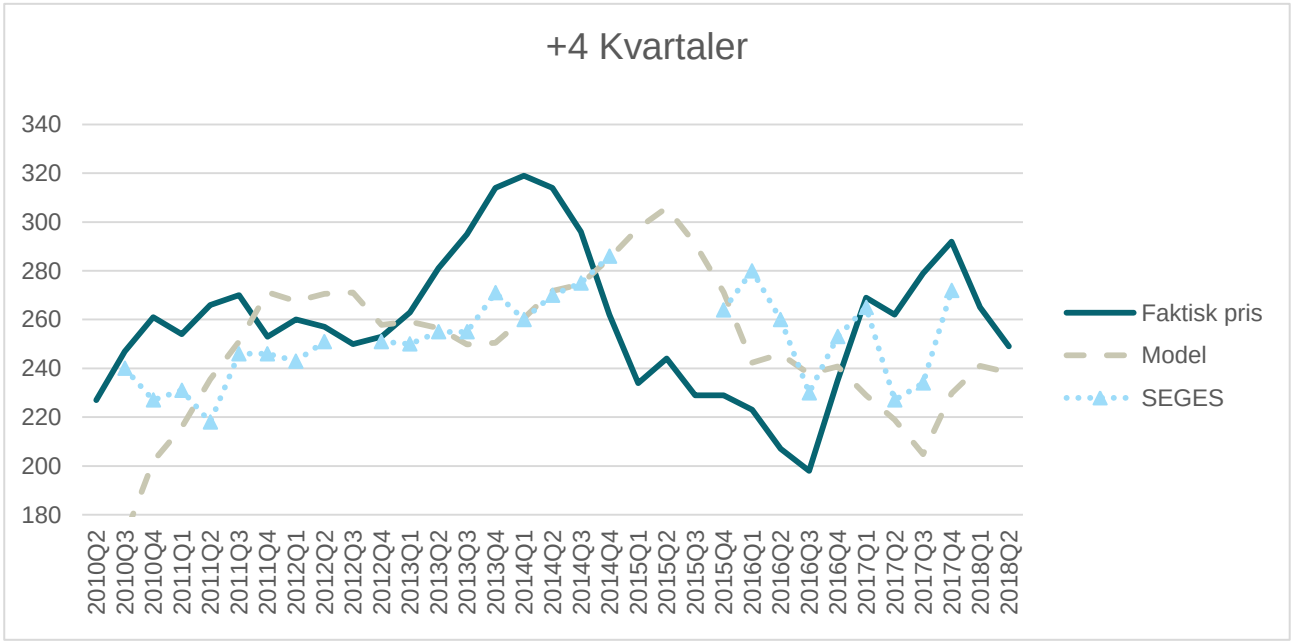
Model-kurven består af out-of-sample prædiktioner fra modellen frembragt ved at afskære datasættet. Således repræsenterer de, hvad modellen havde forudsagt med det på daværende tidspunkt tilgængelige data.

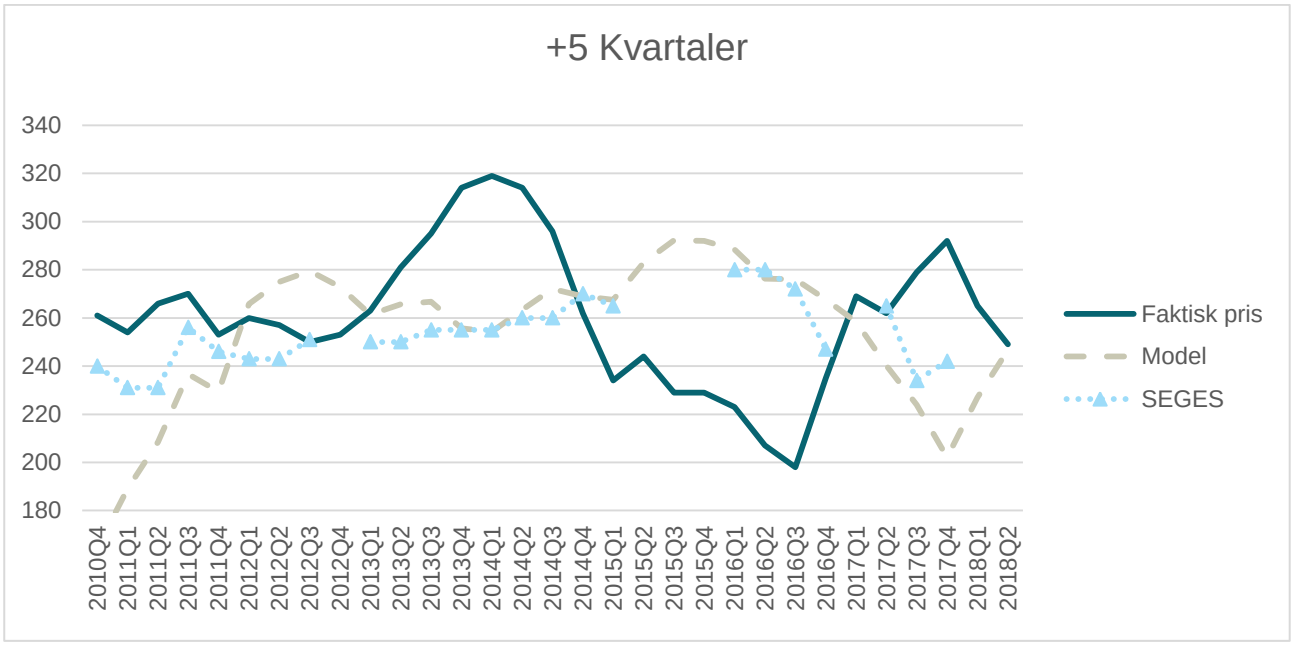
SEGES-kurven består af SEGES' officielle prognosepriser med den pågældende prognose-horisont. Hvor kurven har "huller", lavede SEGES ikke prognoser med den pågældende horisont.

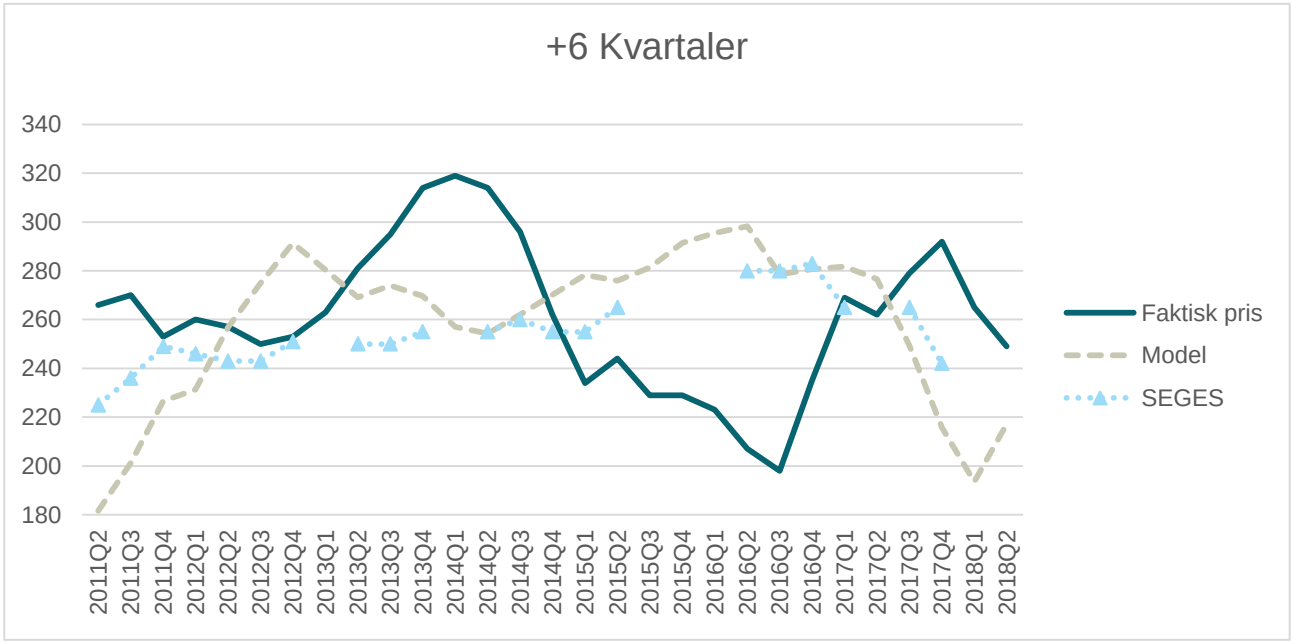





+3 Kvartaler
340
320
300
280
260
240
220
200
180
2009Q4
2010Q1
2010Q2
2010Q3
2010Q4
2011Q1
2011Q2
2011Q3
2011Q4
2012Q1
2012Q2
2012Q3
2012Q4
2013Q1
2013Q2
2013Q3
2013Q4
2014Q1
2014Q2
2014Q3
2014Q4
2015Q1
2015Q2
2015Q3
2015Q4
2016Q1
2016Q2
2016Q3
2016Q4
2017Q1
2017Q2
2017Q3
2017Q4
2018Q1
2018Q2
Faktisk pris
Model
SEGES

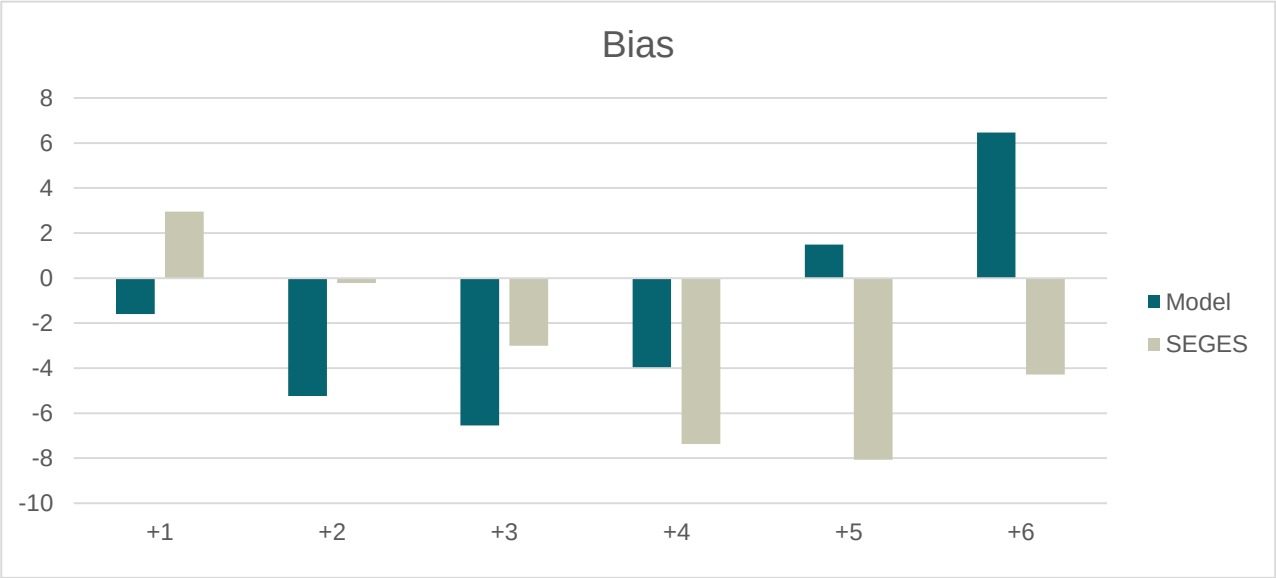

+4 Kvartaler
340
320
300
280
260
240
220
200
180
2010Q2
2010Q3
2010Q4
2011Q1
2011Q2
2011Q3
2011Q4
2012Q1
2012Q2
2012Q3
2012Q4
2013Q1
2013Q2
2013Q3
2013Q4
2014Q1
2014Q2
2014Q3
2014Q4
2015Q1
2015Q2
2015Q3
2015Q4
2016Q1
2016Q2
2016Q3
2016Q4
2017Q1
2017Q2
2017Q3
2017Q4
2018Q1
2018Q2
Faktisk pris
Model
SEGES


+5 Kvartaler
340
320
300
280
260
240
220
200
180
2010Q4
2011Q1
2011Q2
2011Q3
2011Q4
2012Q1
2012Q2
2012Q3
2012Q4
2013Q1
2013Q2
2013Q3
2013Q4
2014Q1
2014Q2
2014Q3
2014Q4
2015Q1
2015Q2
2015Q3
2015Q4
2016Q1
2016Q2
2016Q3
2016Q4
2017Q1
2017Q2
2017Q3
2017Q4
2018Q1
2018Q2
Faktisk pris
Model
SEGES

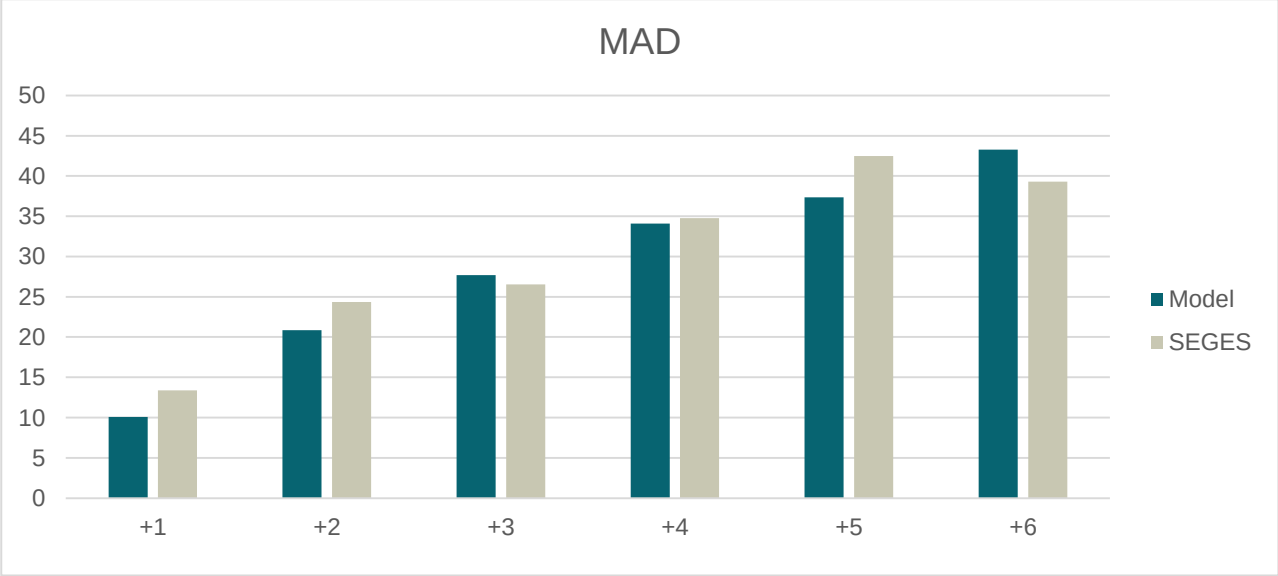

+6 Kvartaler
340
320
300
280
260
240
220
200
180
2011Q2
2011Q3
2011Q4
2012Q1
2012Q2
2012Q3
2012Q4
2013Q1
2013Q2
2013Q3
2013Q4
2014Q1
2014Q2
2014Q3
2014Q4
2015Q1
2015Q2
2015Q3
2015Q4
2016Q1
2016Q2
2016Q3
2016Q4
2017Q1
2017Q2
2017Q3
2017Q4
2018Q1
2018Q2
Faktisk pris
Model
SEGES

Appendiks B – Modelperformance vs. SEGES' historiske prognoser

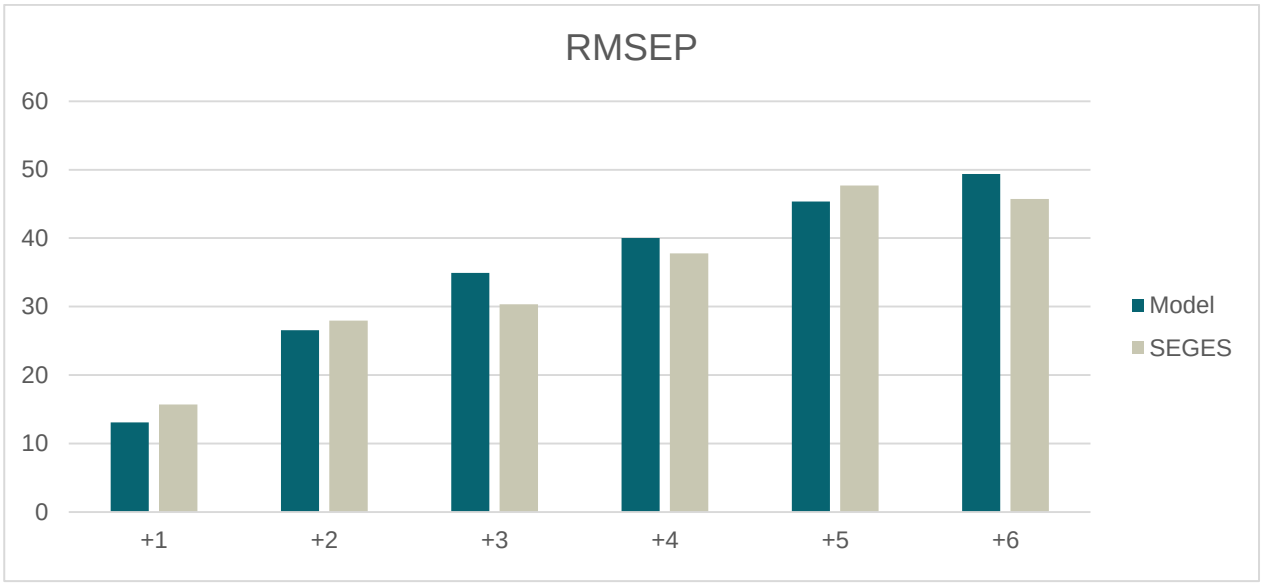
De første fem kvartalers prognoser er ikke medregnet her, da modellen naturligt har en tilpasningsperiode i starten, når den "lærer" processen at kende. Sammenligningsgrundlaget bliver progressivt ringere grundet "huller" i prognoserne fra SEGES.



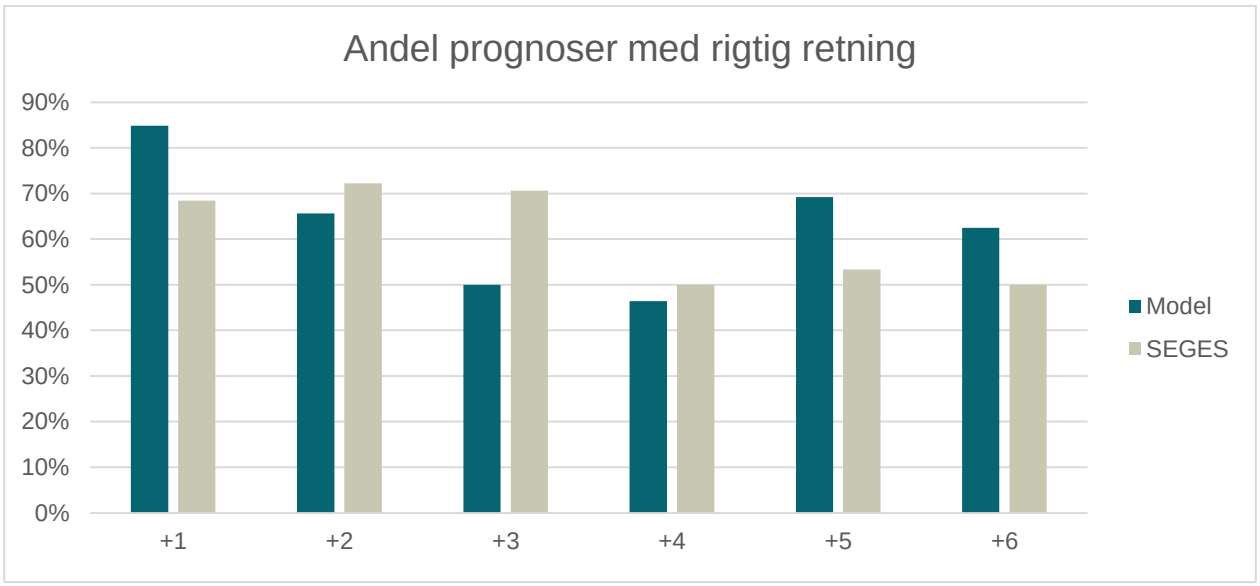
Bias er gennemsnit af alle residualer (forskel fra prognose til faktisk pris), og bør derfor være tæt på 0, hvis modellen ikke systematisk over- eller underskyder priserne.



Mean Absolute Deviation (MAD) er gennemsnittet af de absolutte værdier (uden at tage højde for fortegn) af residualerne. Den udtrykker således en gennemsnitlig +- afvigelse fra den faktiske pris



Root Mean Squared Error of Prediction (RMSEP) findes ved at kvadrere alle residualer, tage gennemsnittet af disse og derpå tage kvadratroden af gennemsnittet. Hvor MAD vægter alle afvigelser ligeligt, ”straffer” RMSEP store afvigelser højere.



Rigtig retning defineres binært med udgangspunkt i det kvartal, prognosen laves i, ved at sammenholde hvor vidt prognosen og den faktiske pris i det fremtidige kvartal ligger over eller under prisen i udgangspunktet. En rigtig retning giver 1 imens forkert giver 0; der tages således ikke højde for afstand mellem prognose og den faktiske pris.

For notatet:
Per Winterberg