



Aalto-yliopisto
Perustieteiden
korkeakoulu

Projection Predictive Method for Bayesian Model Selection in Retail Time Series Forecasting

Elias Ylä-Jarkko

21.1.2022

Ohjaaja: FM *Ella Tamir*

Valvoja: *Prof. Fabricio Oliveira*

Työn saa tallentaa ja julkistaa Aalto-yliopiston avoimilla verkkosivuilla. Muilta osin kaikki oikeudet pidätetään.

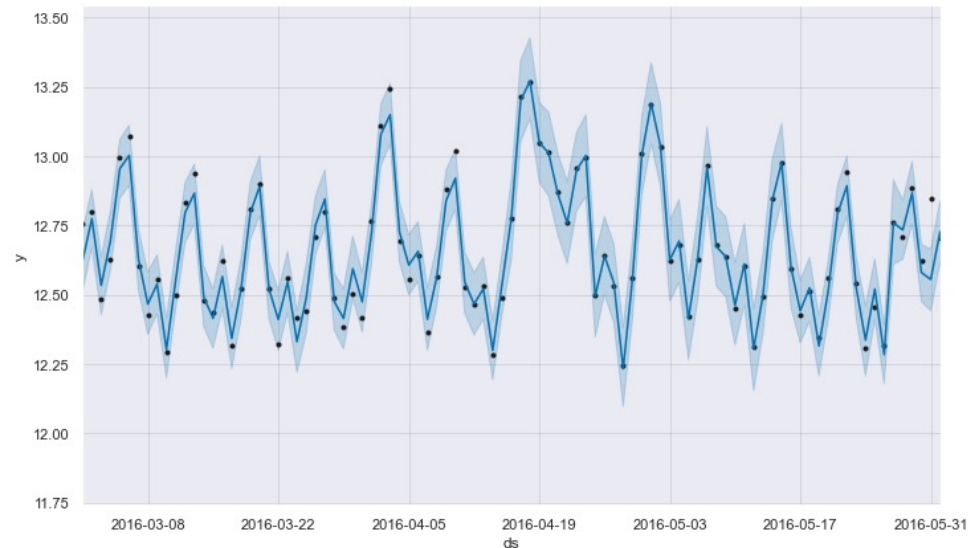
Motivaatio

- Myynnin selittäminen tarjoaa taloudellista hyötyä
 - Johtuuko päivän myyntimäärä markkinoinnista, säästä, perjantaista, tms.
- Tarkempi malli kasvattaa hyötyä ja vähentää riskiä
- Muuttujien määrän vähentäminen voi lisätä tarkkuutta
 - Parantamalla ennustuskykyä
 - Lyhentämällä laskenta-aikaa
 - Helpottamalla päätöksentekoa

(Chan D. & Perry M., 2017), (Piironen, J. 2019)

Menetelmä

- Ennustetaan equadorilaisen kauppaketjun myyntiä
- Muuttujina kausivaihtelu, öljyn, maissin ja perunoiden hinta, lämpötila, sademäärä, alennukset
- Referenssimalli: kaikki muuttujat mukana
- Sovitetaan yksinkertaisempia malleja referenssimallin ennusteeseen



Menetelmä yksityiskohtaisemmin

- Valitsee relevantimpia muuttujia
- Stabiilimpi

(Piironen & Vehtari, 2017)

- Formaalisti: jokaista muuttujamäärää kohden valitaan se alimalli, joka eroaa täydestä mallista vähiten.

$$\begin{aligned}\boldsymbol{\theta}_{\perp} &= \arg \min_{\boldsymbol{\theta} \in \Theta} \text{KL}(p(\tilde{y} | \boldsymbol{\theta}_*) \parallel p(\tilde{y} | \boldsymbol{\theta})) \\ &= \arg \max_{\boldsymbol{\theta} \in \Theta} \mathbb{E}_{\tilde{y} | \boldsymbol{\theta}_*} (\log p(\tilde{y} | \boldsymbol{\theta})).\end{aligned}$$

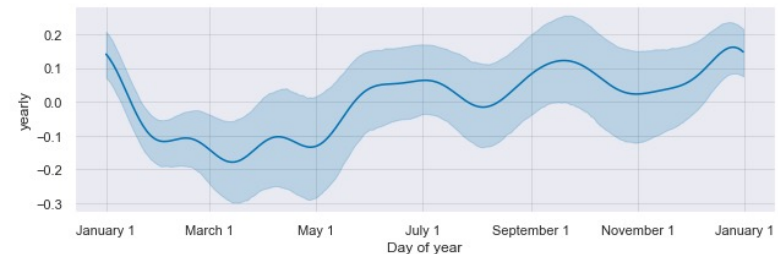
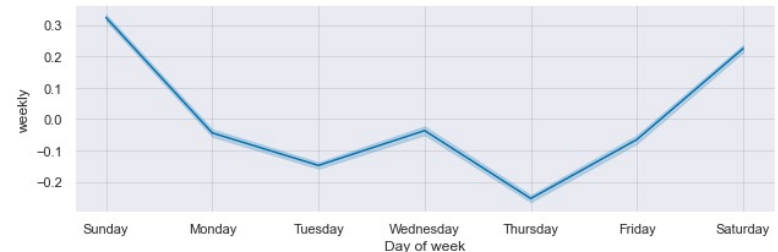
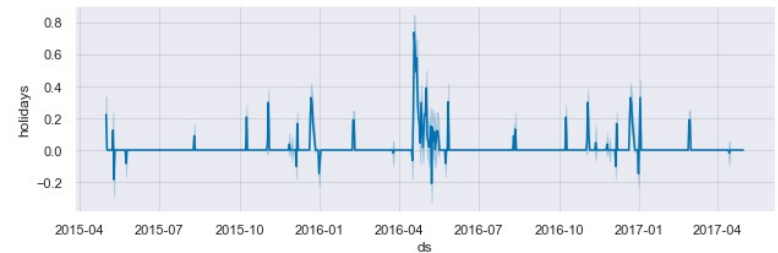
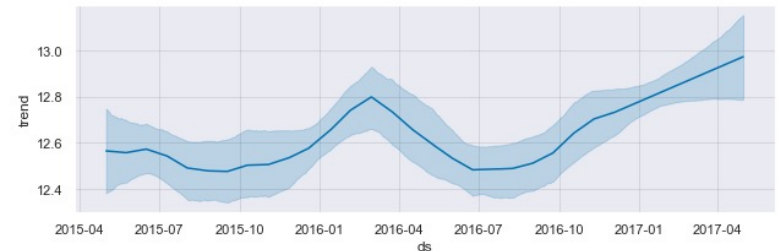
- KL-divergenssin minimoiminen on ekvivalenttia täyden mallin ennusteisiin MLE-sovittamisen kanssa.

Työkalut

- Python:lle ja R:lle toteutettu aikasarjamallintamisen kirjasto
- Taustalla Stan
- Sovittaa automaattisesti jaksollisuuksia, paloittaisen trendin ja yksittäisiä lomapäiviä

$$y \sim N(\text{trendi} + X * \beta, \sigma)$$

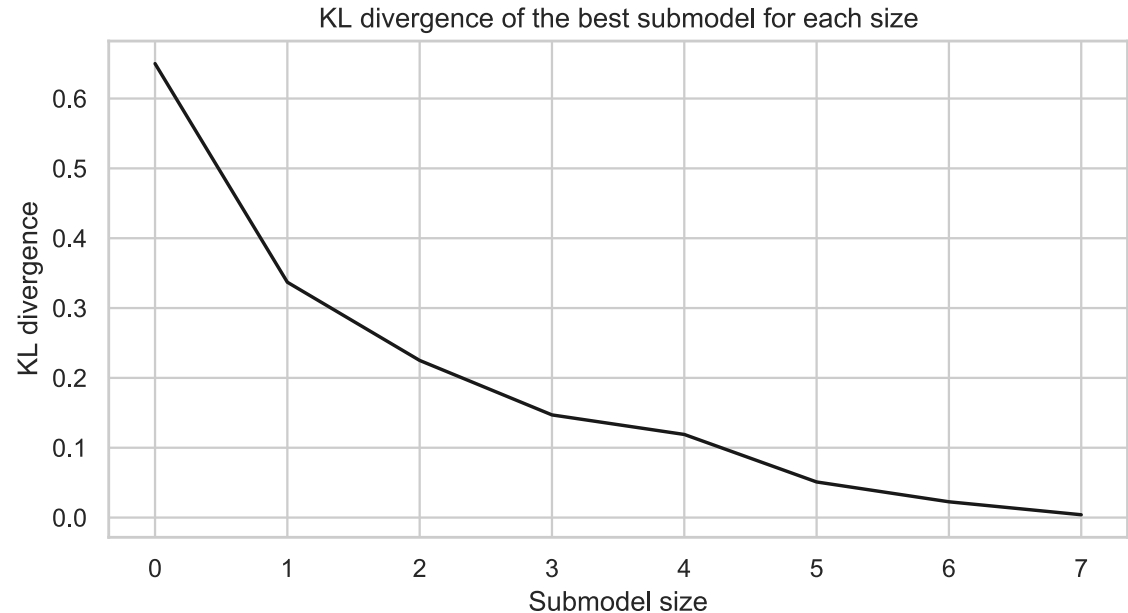
β = ”muuttujat ja fourier–kertoimet”



Tulokset

Muuttujien tärkeysjärjestys:

1. Perunoiden hinta
2. Keskimääräinen lämpötila
3. Alin lämpötila
4. Öljyn hinta
5. Alennusten lukumäärä
6. Sademäärä
7. Korkein lämpötila
8. Keltaisen maissin hinta

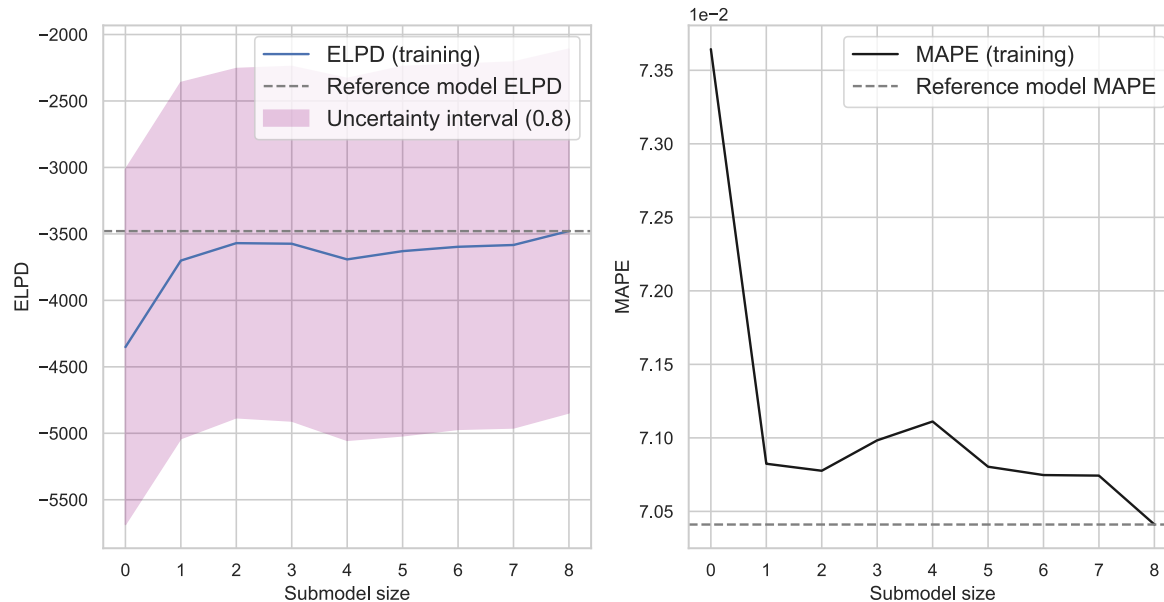


Huomioita:

- Maissin ja perunoiden hinta on vahvasti kääntäen verrannollinen
- Sademäärä on melko tasainen ympäri vuoden, koska Ecuador sijaitsee päiväntasaajalla

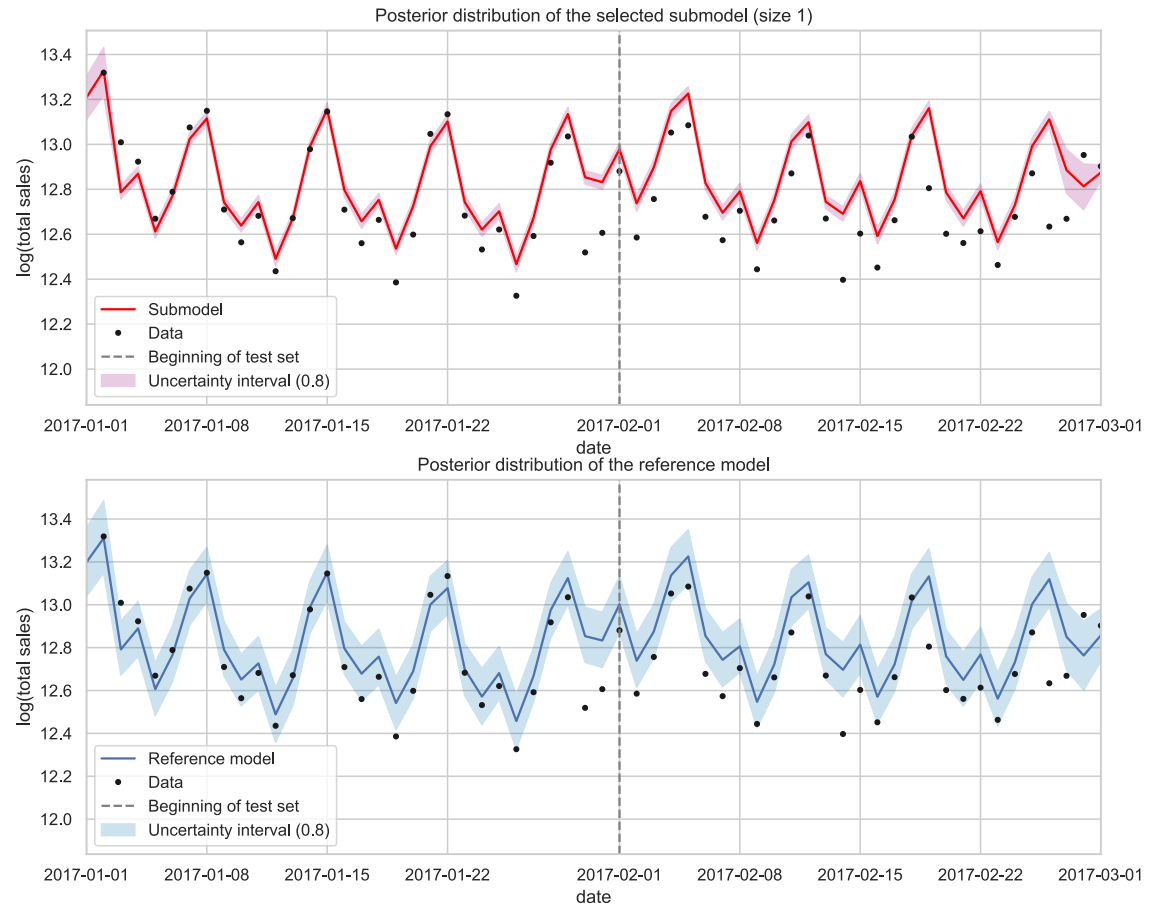
Tulokset

- Referenssimallin suorituskyky saavutetaan 95%:sti jo yhden muuttujan sisältävällä alimallilla (perunoiden hinta)



Tulokset

- Malliksi valitaan perunoiden hinnan sisältävä malli
- Valitun ja referenssimallin posteriorijakaumia vertaillaan oheisessa kuvassa



Johtopäätökset

1. Projection Predictive Method soveltuu vähittäiskaupan aikasarjamallintamiseen Prophet-kirjastolla
 2. Valitussa koejärjestelyssä kahdeksasta muuttujasta paras riitti muodostamaan tyydyttävän mallin
- Tulokset olivat melko herkkiä datan suhteen
 - Muuttujia voisi prosessoida enemmän ja täten luoda lisää muuttujavaihtoehtoja

Lähteet

- Chan, D., & Perry, M. (2017). Challenges and opportunities in media mix modeling.
- Piironen, J. (2019). Bayesian Predictive Inference and Feature Selection for High-Dimensional Data, *Aalto University publication series DOCTORAL DISSERTATIONS*, 84/2019
- Piironen, J., & Vehtari, A. (2017). Comparison of Bayesian predictive methods for model selection. *STATISTICS AND COMPUTING*, 27(3), 711-735. <https://doi.org/10.1007/s11222-016-9649-y>
- Piironen, J., Paasiniemi, M., Vehtari, A., (2020) Projective inference in high-dimensional problems: Prediction and feature selection, *Electronic Journal of Statistics*, 14/2020 2155–2197