Aalto-yliopisto
Perustieteiden
korkeakoulu

Identifying determinants of district heating prices for forecasting (presentation of complete work)

Jessica Norrbäck

01.12.2017

Instructor (ohjaaja): *M.Sc. Vilma Virasjoki*
Supervisor (valvoja): *Prof. Ahti Salo*

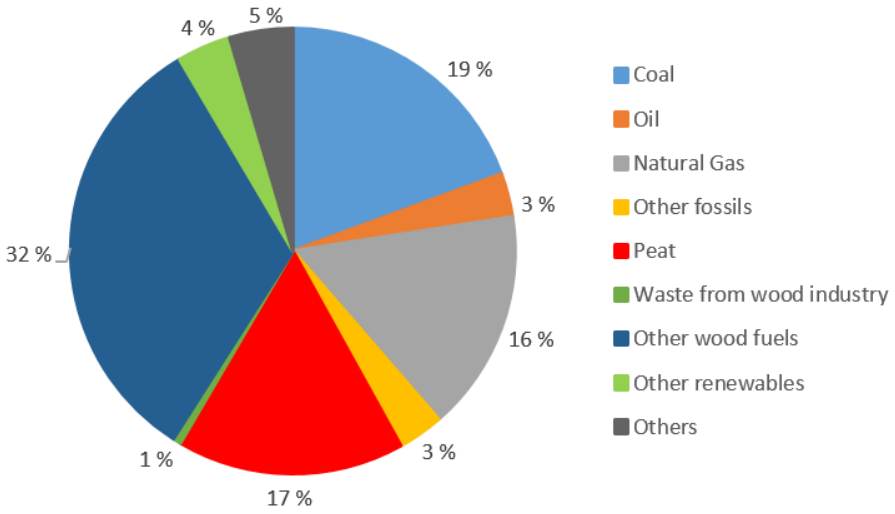
Systeemianalyysin
laboratorio

Background (1/2)

- Heating stands for approximately 67 % of the total energy consumption in buildings
- Nearly 50 % of all buildings are heated with district heating (DH)
- The Finnish heating market is unregulated and competitive
  - Pricing of DH is supervised by the competition authorities and based on the antitrust legislation
  - In some areas customers are obliged to connect to the DH network

Background (2/2)

- The price of DH varies in different municipalities depending on factors such as:
  - Size of DH system
  - Investments made
  - Age of production facility
  - Structure of municipality
  - Fuels used



Fuels used in DH production in 2015.
Data source: Tilastokeskus

Objectives

1. Study factors infuencing the DH price
2. Make a forecast of the DH price for the period 1.1.2015-1.7.2017 based on historical data
3. Compare different methods for making the forecast

Methods

- Methods
  - Multiple linear regression model
  - ARIMA model
  - Naive method
- Software
  - Excel
  - R

Multiple linear regression model

- The dependent variable Y is presented as a linear function of independent variables $X_1, X_2, \ldots, X_n$

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ni} + \epsilon_i \qquad (1)$$

- Assumptions:
  - The model is given by Eq. (1)
  - There may exist no exact linear relationship between two or more of the independent variables
  - The expected value of the error term equals to zero, the errors corresponding to different observations are uncorrelated and the error variable follows a normal distribution

ARIMA(*p*,*d*,*q*) model

- An ARIMA(*p*,*d*,*q*) model is obtained by combining differencing (I) with an autoregressive (AR) and a moving average (MA) model
- Homogenous nonstationary time series can be made stationary by differencing

$$w_t = \Delta^d y_t \tag{2}$$

- Autoregressive model of order *p*

$$y_t = \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \ldots + \Phi_p y_{t-p} + \delta + \epsilon_t \tag{3}$$

- Moving average model of order *q*

$$y_t = \mu + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \ldots - \theta_q \epsilon_{t-q} \tag{4}$$

Factors correlating with the DH price

Lag (months)	Woodchips	Peat	Natural gas	Coal	Electricity price	Consumer price index
0			0.945	0.806	0.861	0.977
1	**0.939**	**0.896**	0.898	0.744	0.894	0.972
2			0.907	0.729	0.896	0.971
3			0.923	0.773	0.886	0.974
4	0.908	0.847	0.934	0.837	0.898	0.977
5			0.941	0.800	0.917	0.979
6			0.945	0.806	0.914	0.980
7	0.896	0.883	0.937	0.801	0.927	0.987
8			0.925	0.777	0.928	0.982
9			0.946	0.829	0.920	0.979
10	0.867	0.851	0.951	0.829	0.921	0.974
11			0.919	0.851	0.937	0.973
12			**0.957**	**0.860**	0.932	0.968

Final multiple linear regression model

- Regression model chosen based on t-tests on individual regression coefficients, partial F-tests and adjusted R^2 value
- $Model_{2c}$ chosen as final model

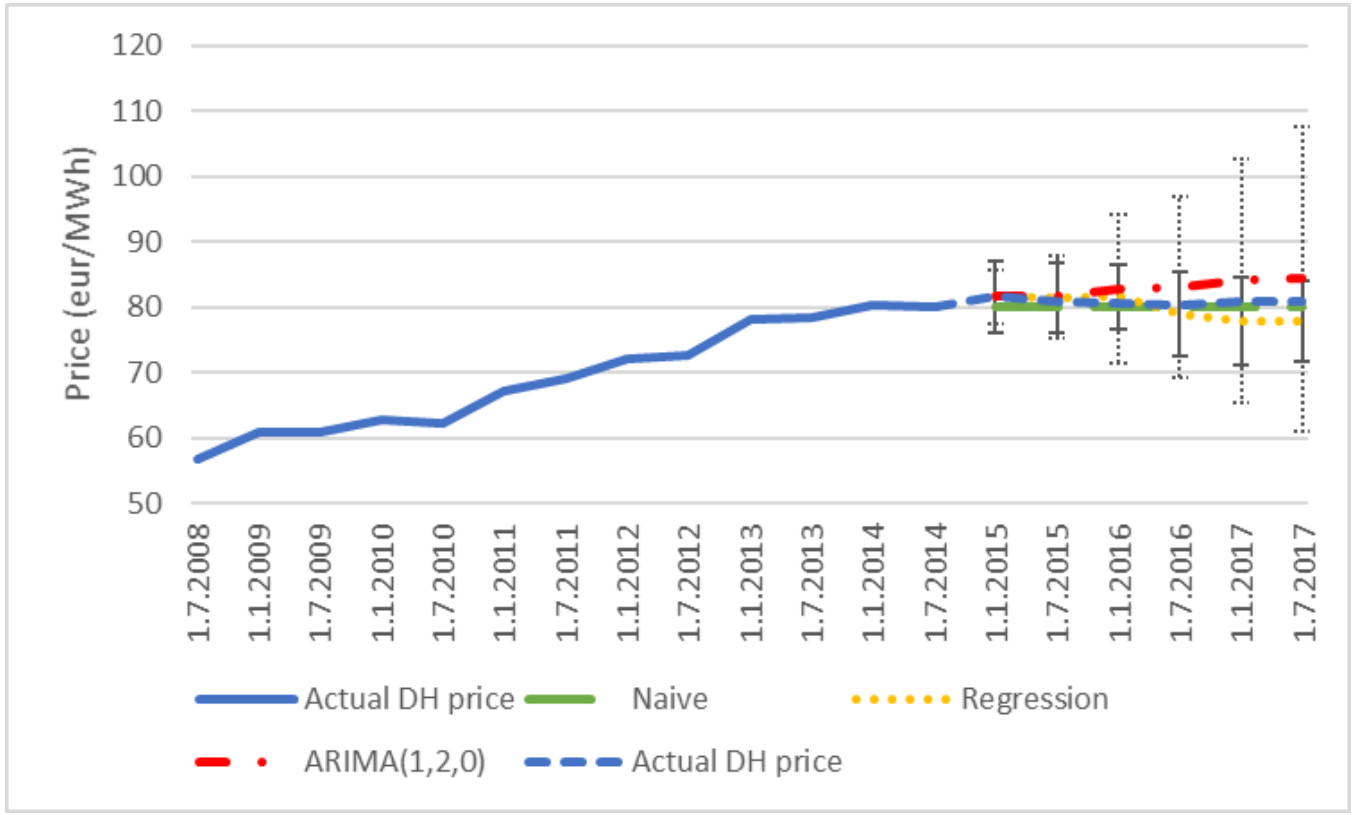
Name	Independent variables	Adjusted R^2
$Model_{1a}$	**Woodchips**	0.870
$Model_{1b}$	**Natural gas**	0.909
$Model_{2a}$	**Woodchips**, Peat	0.895
$Model_{2b}$	**Woodchips**, Coal	0.882
$Model_{2c}$	**Natural gas**, Woodchips	0.926
$Model_{2d}$	**Natural gas**, Peat	0.925
$Model_{2e}$	**Natural gas**, Coal	0.915

Final ARIMA(*p*,*d*,*q*) model

- Parameters *p*,*d* and *q* chosen using R's auto.arima() – function
- ARIMA(1,2,0) chosen as final model

Potential models
ARIMA(1,1,0) with drift
ARIMA(0,1,0)
ARIMA(1,2,0)

Comparison of models (1/2)



Comparison of models (2/2)

Date	Regression	ARIMA	Naive
1.1.2015	0%	0.02%	-1.85%
1.7.2015	0.81%	0.9%	-0.95%
1.1.2016	1.13%	2.65%	-0.73%
1.7.2016	-1.64%	3.37%	-0.3%
1.1.2017	-3.71%	4.05%	-0.9%
1.1.2017	-3.82%	4.41%	-0.99%

Conclusions

- Fuel prices, electricity price and CPI highly correlated with the DH price, however, not all can be included in a multiple linear regression model
- Naive method gives the best forecast for the period 1.1.2015-1.7.2015 and will most likely also give a good short-term forecast
- For an accurate long-term forecast other methods must be used and more factors taken into account
  - Thorough analysis on components of DH price
  - Competitive situation
  - Impact of climate change

Sources

- Energiateollisuus: Kaukolämmön hintatilasto.
- Nuorikivi. (2009), Kaukolämmön hinnoittelumallit.
- Pindyck & Rubinfeld. (1981), Econometric Models and Economic Forecasts.
- DH price data: material bank of Finnish Energy
  - 1.1.1999-1.1.2017, data for every half year
- Data of fuel prices: webpages of Tilastokeskus
  - Peat: 1.9.1999-1.3.2017, every third month
  - Woodchips: 1.3.2007-1.3.2017, every third month
  - Imported fuels (coal and natural gas): 1.1.1990-1.6.2017, every month
- Electricity price data: webpages of the Energy Authority
  - 1.1.1997-1.9.2017, every month