Appendix A Regression Analysis

1. The Authority's monitoring team has developed two regression price models. The purpose of these models is to understand the drivers of the wholesale spot price and if outcomes are indicative of effective competition.

Weekly Model

2. The weekly model is an updated version of the model published in https://www.ea.govt.nz/assets/dms-assets/27/27142Quarterly-Review-July-2020.pdf, Section 8, pg. 21-25

3. The regression equation is

$$\begin{aligned}\log(P_t - \theta_t) = \beta_0 &+ \beta_1(Storage_t - Seasonal.mean.storage_i) \\ &+ \beta_2(Demand_t - Ten.year.mean.demand_t) + \beta_3 Wind.generation_t \\ &+ \beta_4 \log(Gas.price_t) + \beta_5 Generation.HHI_t \\ &+ \beta_6 Ratio.of.adjusted.offer.to.generation_t + \beta_7 Dummy.gas.supply.risk_t\end{aligned}$$

where P_t is the PPI and trend adjusted weekly average spot prices; t =week 1,…,52 for each year; i = spring, summer, autumn, and winter

Daily Model

4. The daily model estimates the daily average spot price based on daily storage, demand, gas price, wind generation, the HHI for generation (as a measure of competition in generation), the ratio of offers to generation (a measure of excess capacity in the market), a dummy variable for the period since the 2018 unplanned Pohokura outage started, and the weekly carbon price (mapped to daily). The units for the raw data are as following: storage and demand are GWh, spot price is \$/MWh, gas price is \$/PJ, and wind generation is MW, carbon price is in New Zealand Units traded under NZ ETS, \$/tonne.

5. We used the Augmented Dicky-Fuller (ADF) to test all variables to see if they are stationary. If not, we tested the first difference and then the second difference using the ADF test until the variable was stationary. The first difference of a time series is the series of changes from one period to the next. For example, if the storage is not stationary, we use $storage_t - storage_{t-1}$.

6. We fitted the data using a dynamic regression model with Autoregressive with five lags (AR(5)). Dynamic regression is a method to transform ARIMAX (Autoregressive Integrated Moving Average with covariates model) and make the coefficients of covariates interpretable.

7. Once we dropped the insignificant variables; the ratio of offers to generation, the dummy variable for 2018 and carbon price, we got the following model[1], where diff is the first difference:

$$y_t = \beta_0 - \beta_1\left(storage_t - 20.year.mean.storage_{dayofyear}\right) + \beta_2 diff(demand_t) - \beta_3\, wind.generation_t + \beta_4\, gas.price_t - \beta_5 diff(generation\ HHI_t) + \beta_6 dummy + \eta_t$$

$$\eta_t = \varphi_1\eta_1 - \varphi_2\eta_2 + \varphi_3\eta_3 + \varphi_4\eta_4 + \varphi_5\eta_5 + \varepsilon_t$$

8. ε_t, the residuals of ARMA errors (from AR(5)), should not significantly different from white noise. Ideally, we expect the ARIMA errors are purely random, and are not correlated with each other (show no systematic pattern). ARIMA errors equals y_t minus the estimate $\hat{y}$ with their five time lags.

---

[1] Updated, $diff(storage_t)$ has been replaced with $\left(storage_t - 20.year.mean.storage_{dayofyear}\right)$