Estimating Re-fitting Frequencies for Short-term Energy Models

Janice Lent and Rebecca George
U.S. Energy Information Administration(EIA)

ESMD Seasonal Workshop
November 20, 2019 | Washington, DC

Acknowledgments

The authors thank the following EIA colleagues for their contributions to this research and presentation:

- Tim Hess
- Bruce Bawks

Outline

- Background
  - EIA's *Short-term Energy Outlook* (STEO) publication
  - EIA's Regional Short-term Energy Modeling (RSTEM) system.
- Research
  - Method of estimating appropriate model re-fitting frequencies
  - Results

Background

- EIA publishes monthly short-term (one to two years) forecasts of U.S. energy supply, demand, trade, prices, and other energy-related series.
- EIA's monthly *Short-Term Energy Outlook* (STEO) forecasts are based on a large collection of time series models, processed by EIA's Regional Short-term Energy Modeling (RSTEM) system.
- The RSTEM System
  - An integrated system including over 440 econometric regression and ARMA time series models
  - Comprises modules specific to energy topics (e.g., petroleum product prices, natural gas demand, electricity generation)
  - Model coefficients are estimated in the E-views software

STEO data tables display data at both a quarterly and monthly frequency

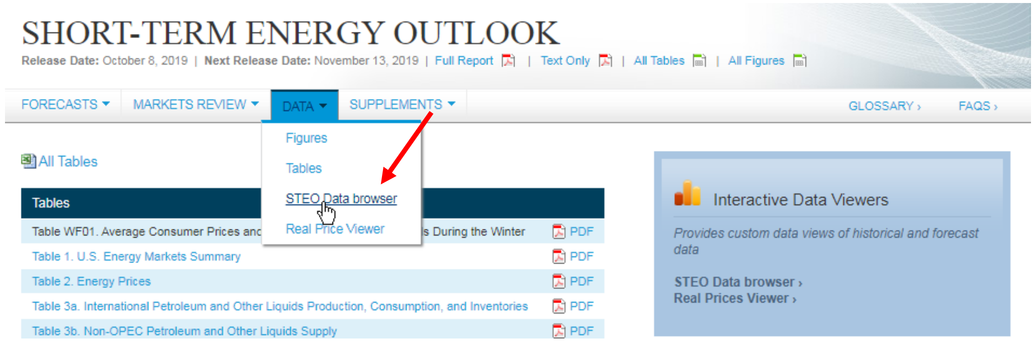
Table of Contents

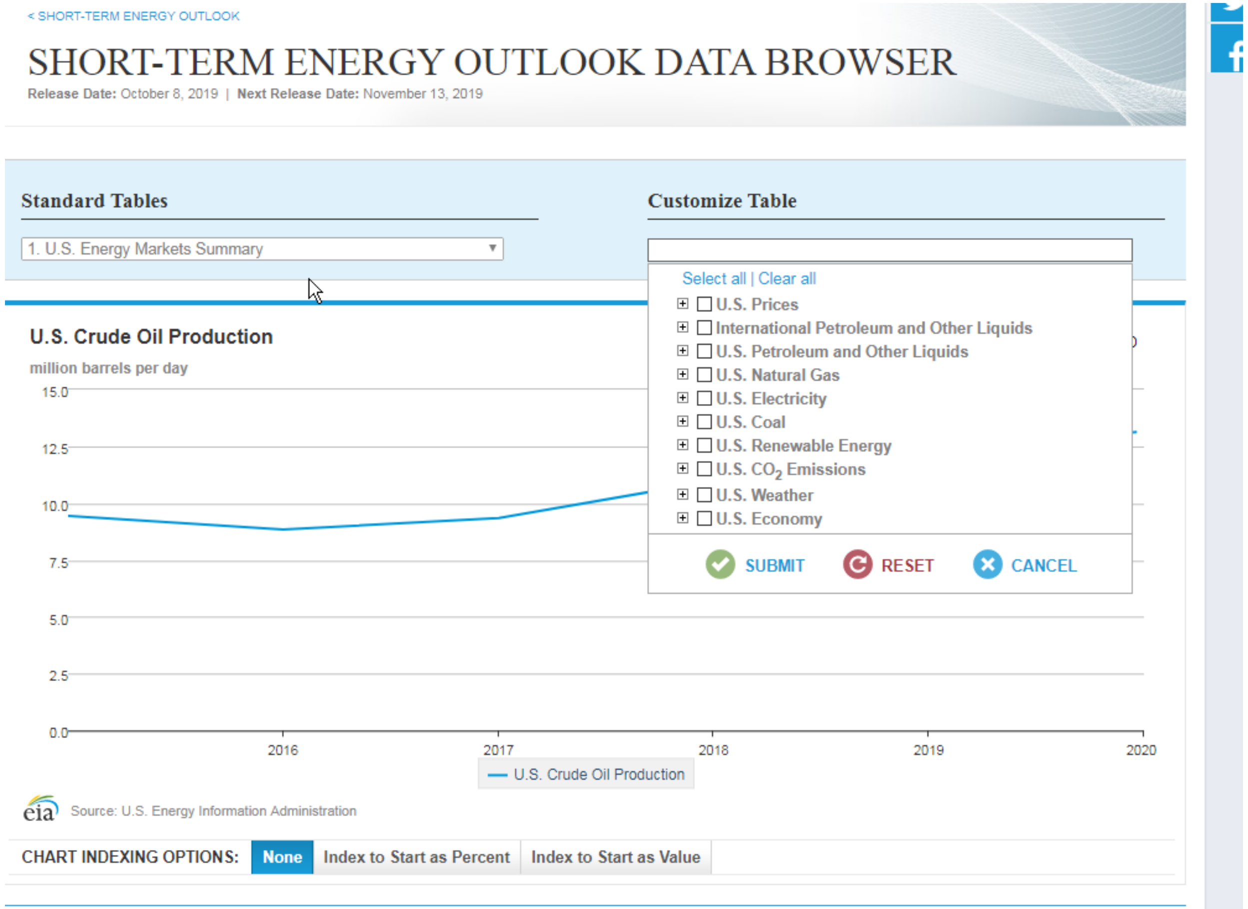
Table 1. U.S. Energy Markets Summary

U.S. Energy Information Administration | Short-Term Energy Outlook - October 2019

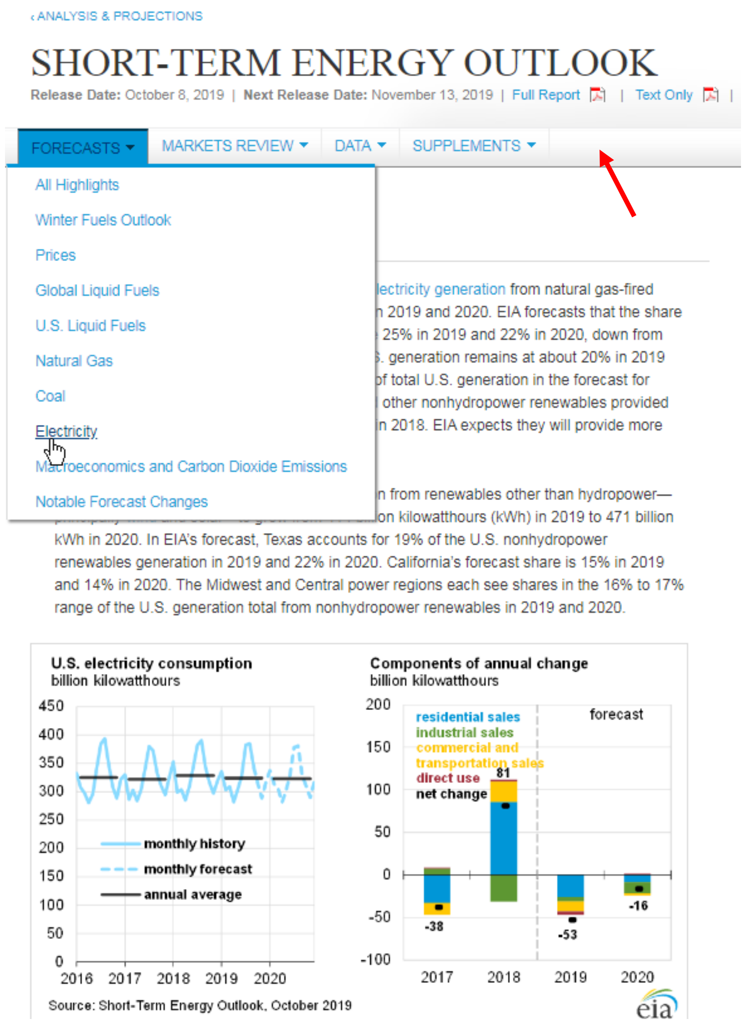
		2015												2016											
		Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
	Energy Supply																								
COPRPUS	Crude Oil Production (a) (million barrels per day)	9.38	9.51	9.58	9.66	9.47	9.35	9.44	9.41	9.48	9.40	9.32	9.26	9.20	9.06	9.09	8.87	8.82	8.65	8.65	8.68	8.53	8.83	8.90	8.80
NGPRPUS	Dry Natural Gas Production (billion cubic feet per day)	73.44	73.81	74.14	75.21	74.12	73.95	74.19	74.27	74.74	74.19	73.88	73.89	73.56	74.60	73.76	73.71	72.87	72.17	72.76	72.18	71.71	71.42	72.02	71.21
CLPRPUS_TON	Coal Production (million short tons)	86.6	72.3	81.5	75.2	70.4	66.9	76.5	82.6	77.7	75.7	68.6	63.0	60.6	57.3	55.3	48.2	53.1	59.5	61.8	68.2	65.1	68.7	67.1	63.3
	Energy Consumption																								
PATCPUSX	Liquid Fuels (million barrels per day)	19.26	19.66	19.34	19.25	19.32	19.85	20.13	19.94	19.43	19.49	19.13	19.59	19.06	19.85	19.73	19.34	19.33	19.85	19.78	20.27	19.76	19.65	19.66	19.98
NGTCPUS	Natural Gas (billion cubic feet per day)	100.48	104.47	83.59	66.93	59.94	63.33	66.70	66.22	63.38	64.11	74.97	83.49	99.73	91.46	76.01	69.46	63.41	66.69	70.54	71.24	64.92	62.10	71.98	92.46
CLTCPUS_TON	Coal (b) (million short tons)	76.9	72.3	63.6	53.2	61.9	73.8	81.4	78.6	69.4	58.4	53.6	54.9	66.7	55.2	44.6	43.4	49.3	67.6	78.6	78.2	66.6	59.0	52.5	69.5

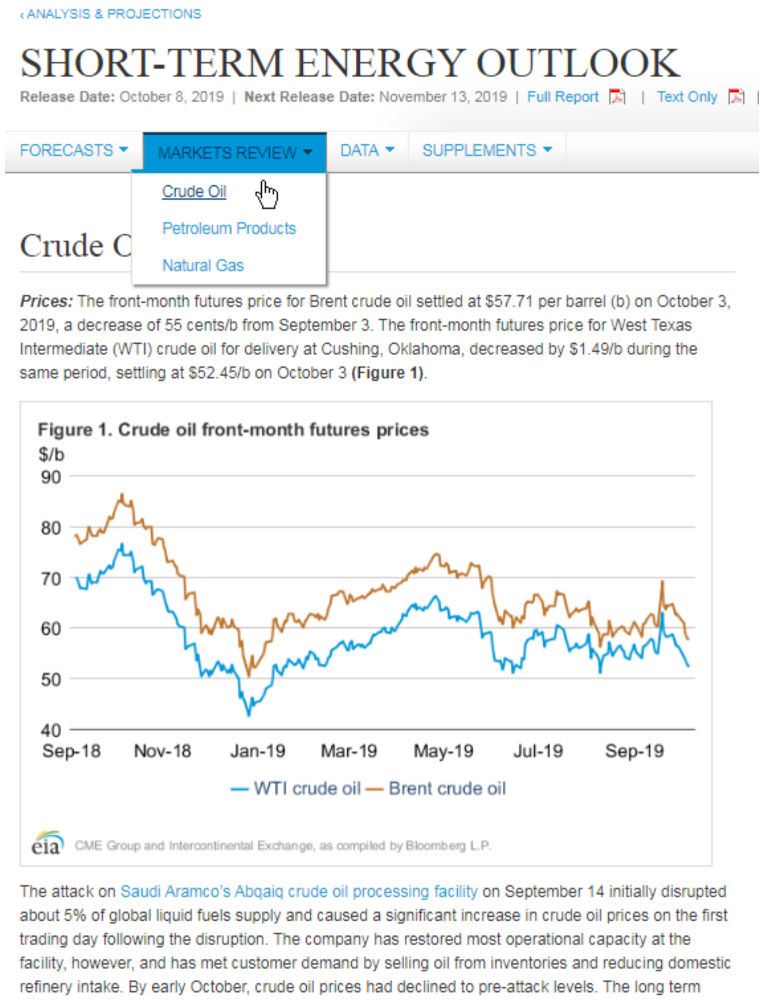
Data are also accessible through a custom data browser





STEO provides forecast notes and energy market analysis





EIA publications contributing data to STEO

- Petroleum Supply Monthly – https://www.eia.gov/petroleum/supply/monthly/
- Petroleum Marketing Monthly – https://www.eia.gov/petroleum/marketing/monthly/
- Weekly Petroleum Status Report – https://www.eia.gov/petroleum/supply/weekly/
- International Energy Statistics – https://www.eia.gov/beta/international/data/browser/
- Natural Gas Monthly – https://www.eia.gov/naturalgas/monthly/
- Weekly Natural Gas Storage Report – http://ir.eia.gov/ngs/ngs.html
- Quarterly Coal Report – https://www.eia.gov/coal/production/quarterly/
- Electric Power Monthly – https://www.eia.gov/electricity/monthly/

External data sources used in STEO

- Weather – National Oceanic and Atmospheric Administration
- Oil and natural gas spot prices – Thomson Reuters
- U.S. macroeconomics – IHS Markit
- Global macroeconomics – Oxford Economics
- Wholesale electricity prices – S&P Global Intelligence and PJM
- Data from other U.S. statistical agencies, e.g., Census Bureau, BLS

Exogenous models and forecasts used in STEO

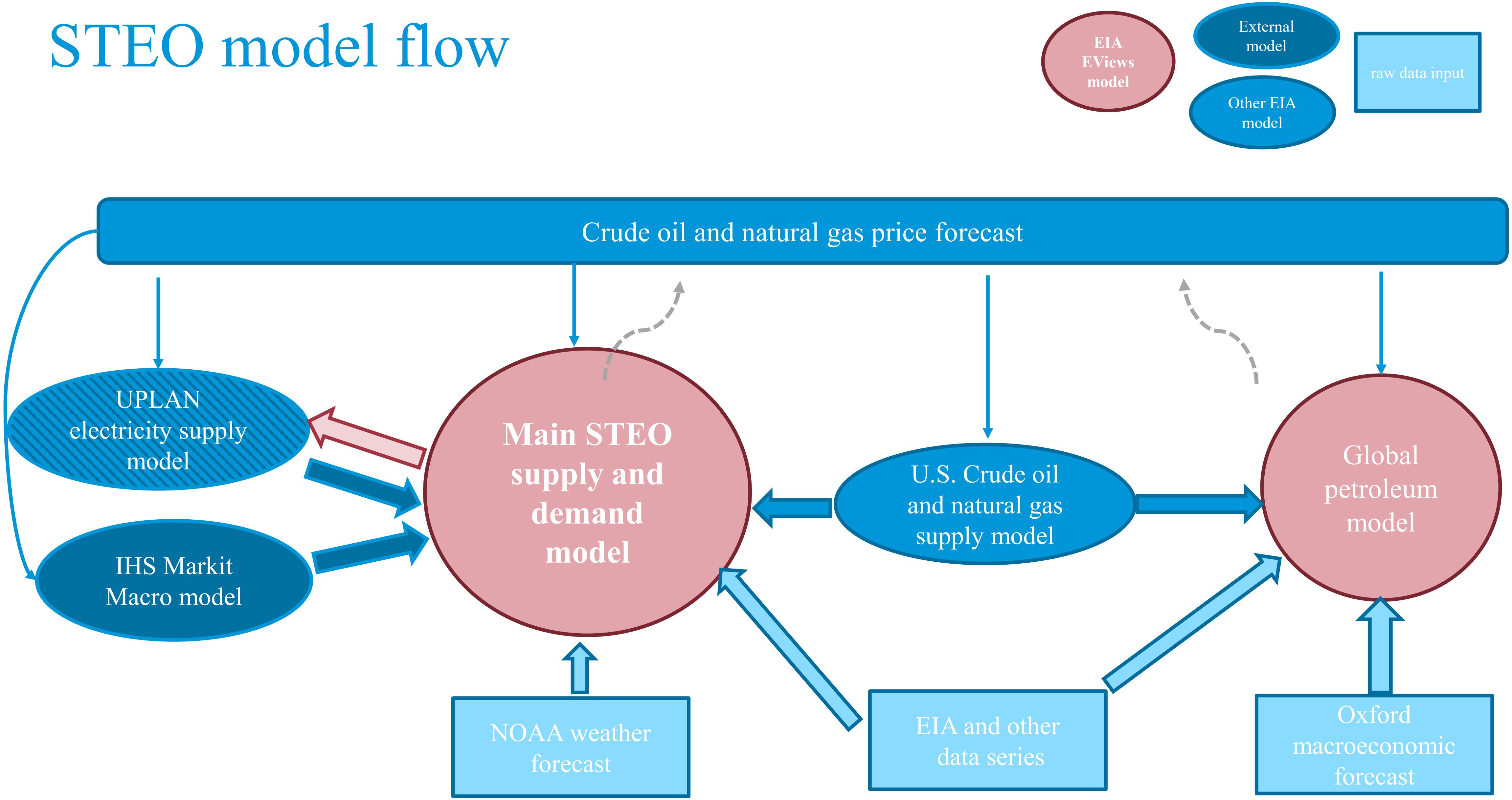
- IHS Macroeconomic model
- UPLAN electricity supply model
- EIA crude oil and natural gas supply model
- NOAA weather forecast
- Oxford global macroeconomic forecast

Analyst judgment is used in STEO forecasts

EIA analysts adjust forecasts to account for known or likely future events that historical data do not reflect. Examples:

- Regulatory changes
- Weather disruptions
- Supply disruptions
- Pipeline or fuel distribution constraints

STEO model flow



Maintaining RSTEM model coefficients

- RSTEM includes over 440 regression and ARMA models, many with seasonal dummy variables to account for seasonality.
- Although EIA's previous goal was to re-estimate coefficients for all RSTEM models at least once every year, resource constraints sometimes prevented annual re-estimation.
- An examination in 2019 revealed that 156 (35%) of the models were most recently re-estimated prior to 2017.
- Question: How frequently should RSTEM model coefficients be re-estimated?

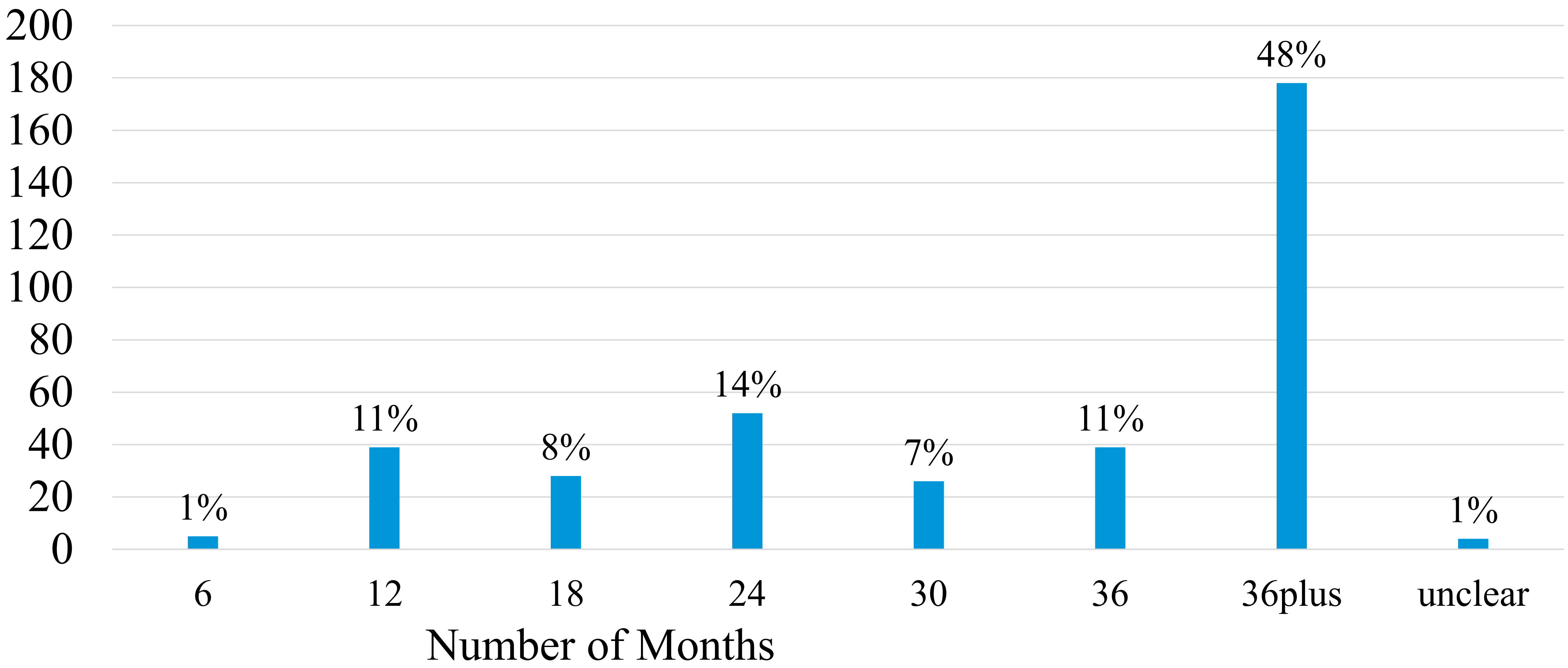
Model re-fitting frequency research

- Goal: Develop easy-to-follow guidelines for RSTEM model refitting frequency
- Method:
  - Fit the models using data from many “sliding window” sample periods.
  - Examine changes in regression and time series coefficients computed using data from different time periods.
  - Determine the length of time that resulted in a statistically significant change in model coefficients, indicating a change in a cointegrated relationship.
  - Discuss the statistical test results with subject matter experts to help determine patterns and develop guidelines.

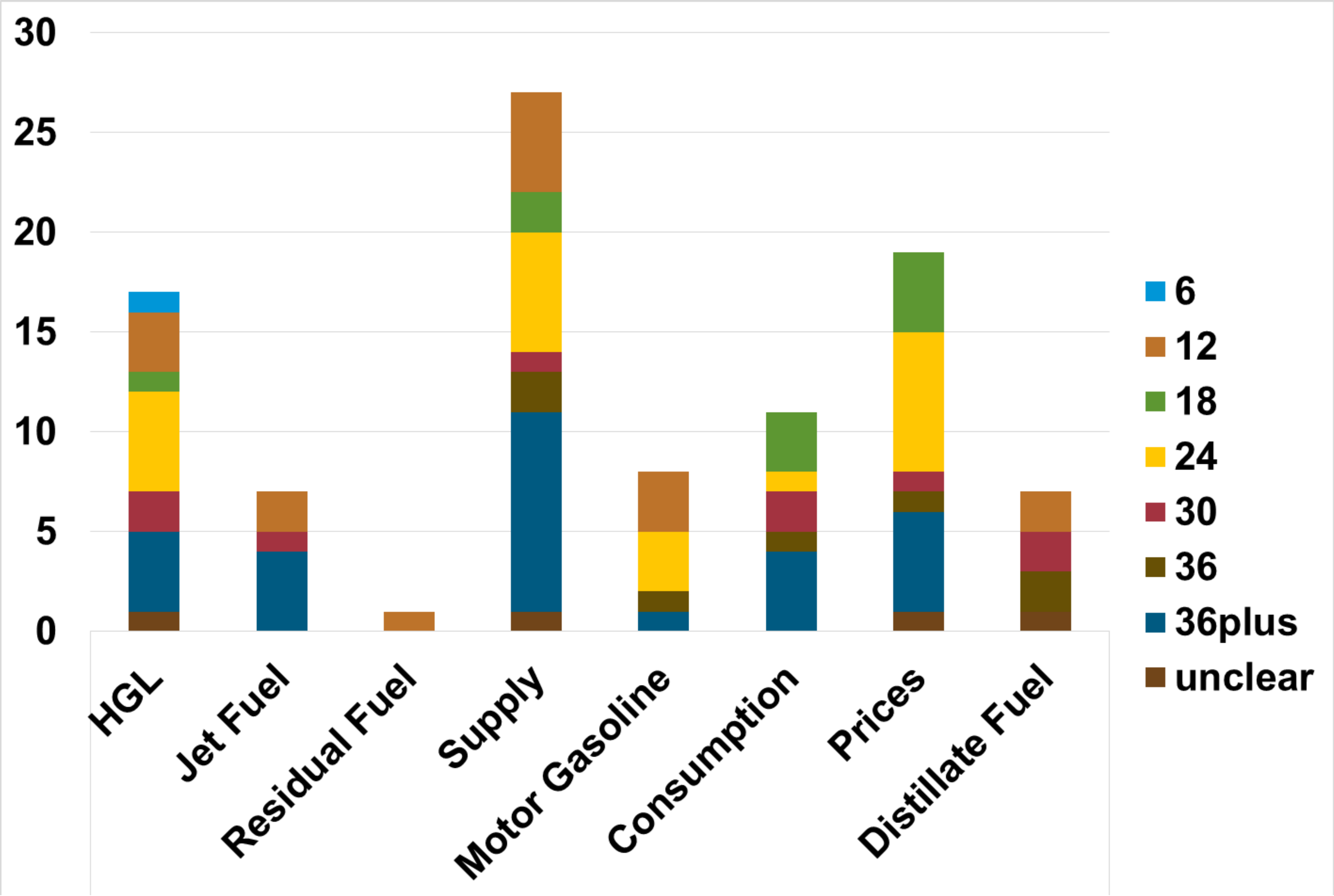
Research method

- For each of the 442 RSTEM models, we estimated coefficients 145 or more times using rolling sample periods.
- We estimated each model $n \geq 30$ times on data from different 10-year time windows starting k months apart, where $k \in \{6,12,18,24,30,36\}$.
- We stored coefficients after each re-estimation, obtaining 442 matrices of coefficients. Only 371 models had enough data to support the research.
- Accounting for correlations, we estimated empirical t-statistics to test the hypothesis of a significant difference ($\alpha = 0.05$) between the model coefficients from time windows starting k months apart.
- Details and formulas are in the supplementary slides.

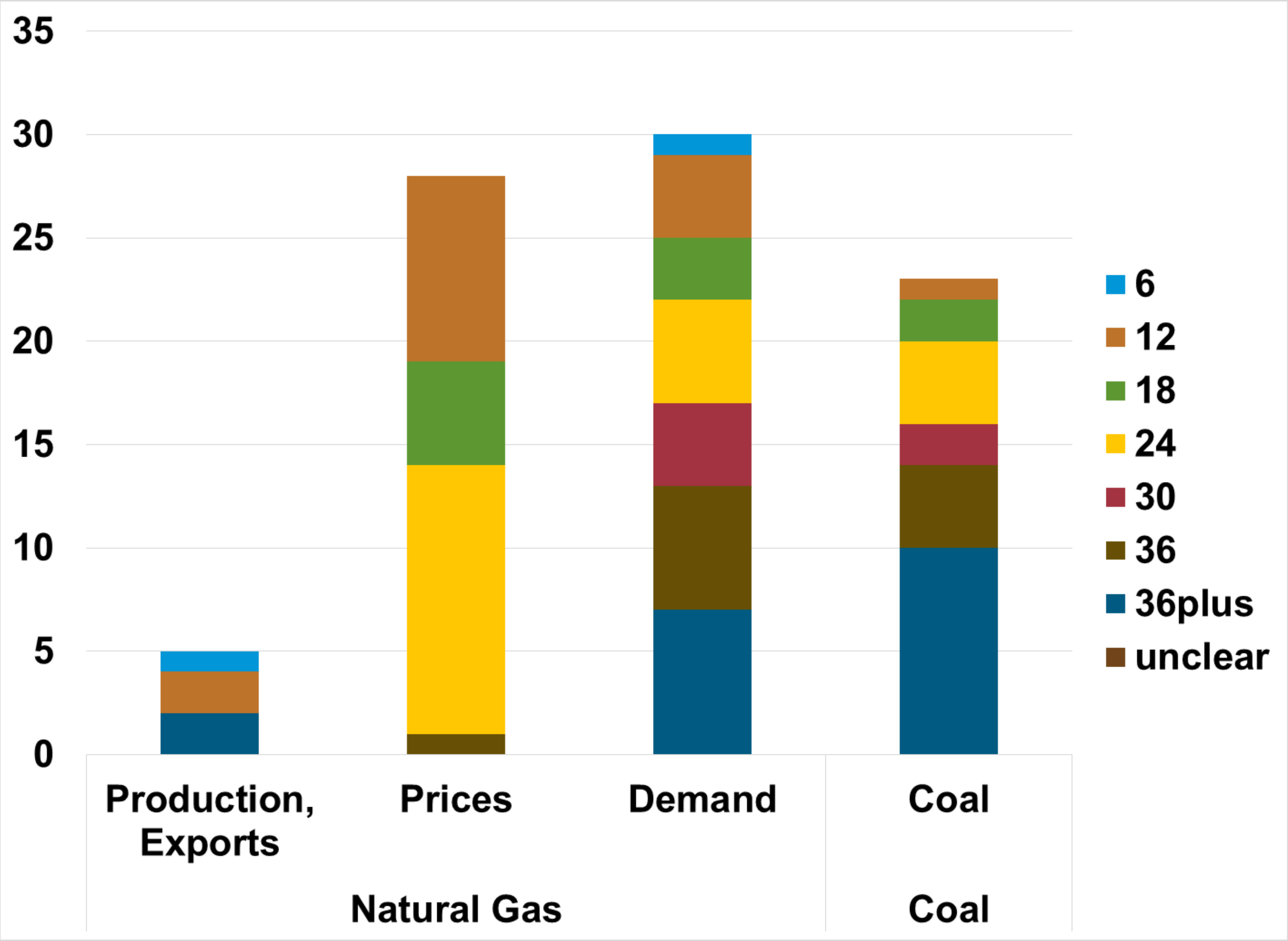
Summary of Results: Number of models (of 371) to be re-estimated at various intervals



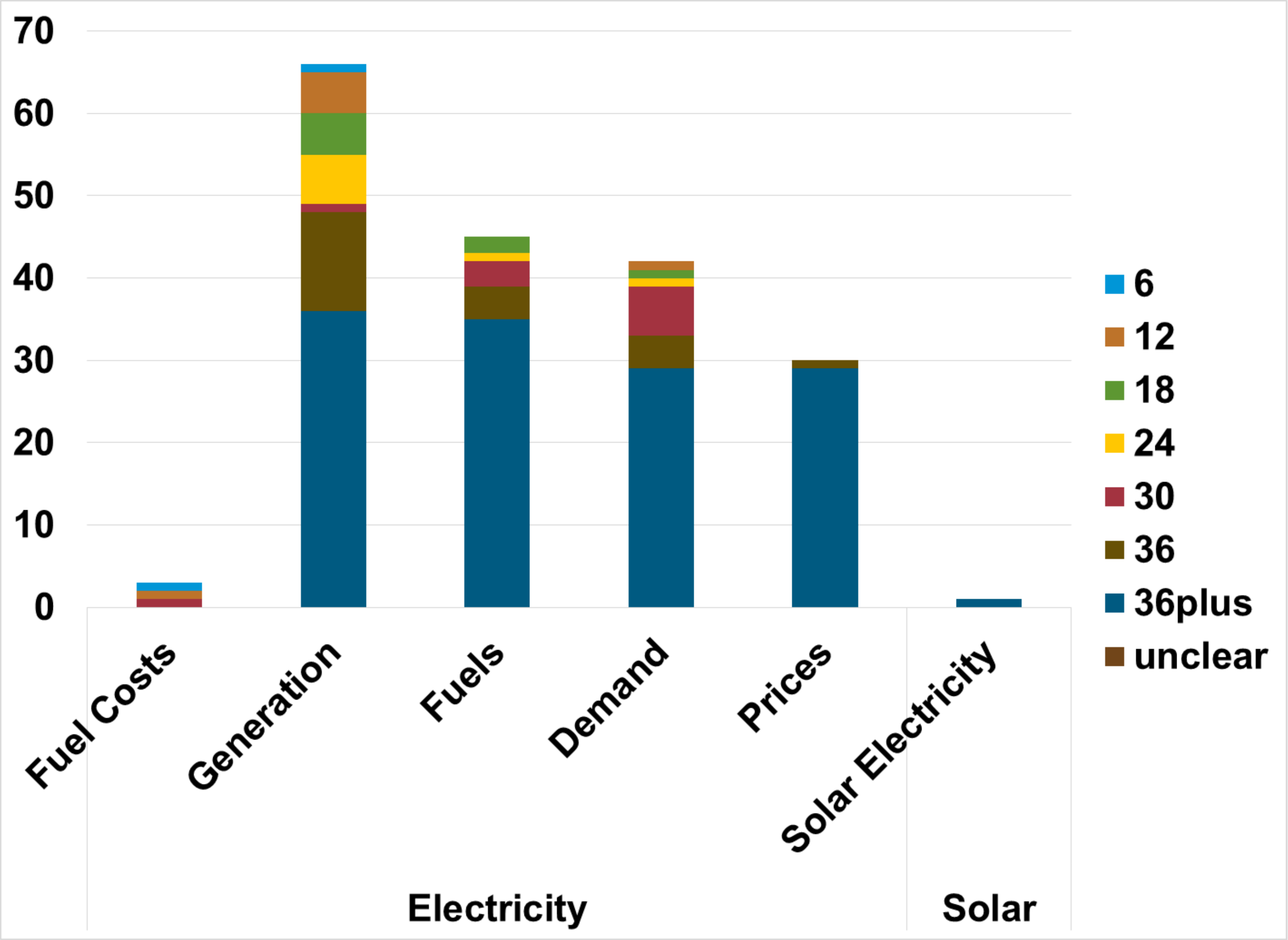
Numbers of petroleum-related models by category



Numbers of natural gas and coal-related models by category



Numbers of electricity-related models by category



Results summary

- Based on our criterion, most of the RSTEM models analyzed only need refitting every three years.
- Of the remaining models,
  - 12% need re-estimation at least once a year;
  - 22% need re-estimation at least every one to two years
- However, models related to hydrocarbon gas liquids (HGL) and several models related to natural gas production need re-estimation every six months.

Research-based recommendations

- **General:** When structural changes have occurred or appear likely to occur in energy markets (as currently in HGL markets), related RSTEM models need re-fitting every 6 to 12 months.

- **Electricity Prices and Demand:** Most of the related models fell into the three-year update category. These may be given lower priority, as resource constraints require.

- **Electricity Generation:** Models related to electricity generation from solar, wind, biomass, and geothermal sources need more frequent re-fitting (one to two years).

Research-based recommendations (2)

- **Natural Gas:** Models related to natural gas production should be re-estimated at least once a year, and models related to natural gas spot prices (Henry Hub) may need more frequent re-fitting.
- **Petroleum:** Models related to hydrocarbon gas liquids (HGL), gasoline, distillate, and jet fuel need refitting every 6 to 12 months.
- **Coal:** Models related to coal production or coal stocks need annual re-estimation. While models related to coal imports and exports appeared in the three-year re-fitting category, recent increases in coal exports may indicate that these models need more frequently re-estimation.

Summary

- Forecasts in EIA's STEO publication are generated by a complex modeling system that requires substantial resources to maintain.
- In order to appropriately allocate analysts' time, we developed recommendations for regression and time-series model re-estimation frequencies.
- We developed a "sliding window" method that was useful for detecting changes in market equilibrium relationships.
- Results indicated that structural changes in energy markets necessitated more intense model maintenance. Some models, however, may be re-estimated once every three years.

Contact information

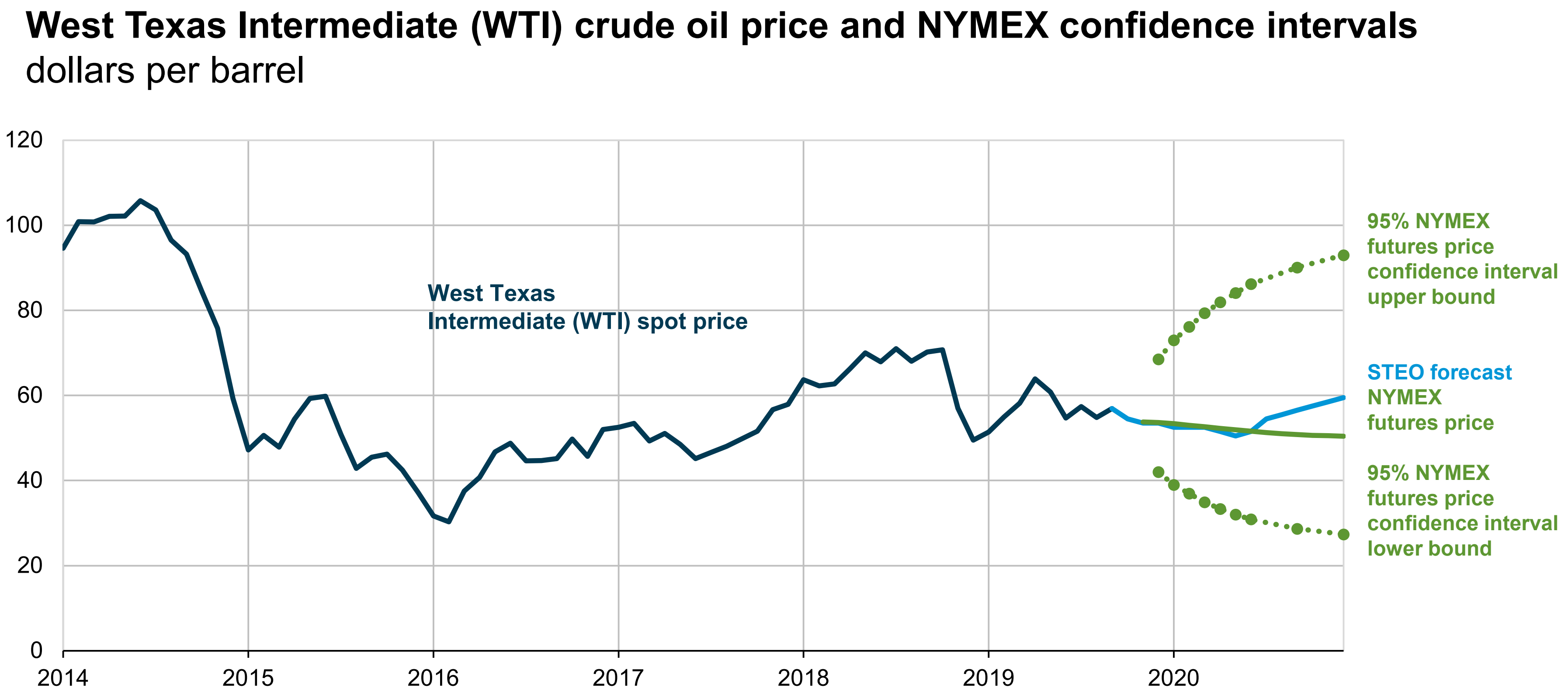
Janice Lent

U.S. Energy Information Administration

Janice.Lent@eia.gov

Supplementary Slides

STEO uses market indicators to contextualize forecasts



Note: Confidence interval derived from options market information for the five trading days ending Oct 3, 2019. Intervals not calculated for months with sparse trading in near-the-money options contracts.

Source: U.S. Energy Information Administration and CME Group

Research method step 1: Estimate average model coefficients

1. For each of the 442 RSTEM models, we estimated coefficients 145 or more times using rolling sample periods.

2. We estimated each model $n \geq 30$ times on data from different 10-year time windows starting k months apart, where $k \in \{6,12,18,24,30,36\}$.

3. We stored coefficients after each re-estimation, obtaining 442 matrices of coefficients. Only 371 models had enough data to support the research.

4. Notation: For $i = 1, \ldots, n$, let $\hat{\beta}_{t_i,t_{i+120}}$ denote the estimate of β computed from the i^{th} time window, and let $\bar{\hat{\beta}}_{t_1,t_2} = \frac{1}{n}\sum_{i=1}^{n} \hat{\beta}_{t_i,t_{i+120}}$.

Research method step 2: Compute variances and covariances

For each of value of $k \in \{6,12,18,24,30,36\}$, we estimated empirical variances and covariances of the average coefficients:

$$\hat{\sigma}^2_{\beta,t} = \frac{1}{n-1}\sum_{i=1}^{n}\left[\left(\hat{\beta}_{t_i,t_{i+120}} - \bar{\hat{\beta}}_{t_i,t_{i+120}}\right)^2\right],$$

$$\hat{\sigma}^2_{\beta,t+k} = \frac{1}{n-1}\sum_{i=1}^{n}\left[\left(\hat{\beta}_{t_i+k,t_{i+120}+k} - \bar{\hat{\beta}}_{t_i+k,t_{i+120}+k}\right)^2\right],$$

and

$$\hat{\sigma}_{\beta,t,t+k} = \frac{1}{n-1}\sum_{i=1}^{n}\left[\left(\hat{\beta}_{t_i,t_{i+120}} - \bar{\hat{\beta}}_{t_1,t_2}\right)\left(\hat{\beta}_{t_i+k,t_{i+120}+k} - \bar{\hat{\beta}}_{t_i+k,t_{i+120}+k}\right)\right].$$

Research method step 3: Estimate test statistics

1. We estimated the variance of the difference $\bar{\hat{\beta}}_{t_1,t_2} - \bar{\hat{\beta}}_{t_1+k,t_2+k}$ as

$$\hat{\sigma}^2_{\beta,k} = \hat{\sigma}^2_{\beta,t} + \hat{\sigma}^2_{\beta,t+k} - 2\hat{\sigma}_{\beta,t,t+k}.$$

2. The t-statistic for testing the hypothesis $\left|\bar{\hat{\beta}}_{t_i,t_{i+120}} - \bar{\hat{\beta}}_{t_i+k,t_{i+120}+k}\right| > 0$ is

$$t_{\beta,k} = \frac{\bar{\hat{\beta}}_{t_i,t_{i+120}} - \bar{\hat{\beta}}_{t_i+k,t_{i+120}+k}}{\sqrt{\hat{\sigma}^2_{\beta,k}}}.$$

3. We grouped the models based on the number of months k needed to generate a significant ($\alpha = 0.05$) difference in at least one coefficient.