Oil and gas price forecasting at a time of increased uncertainty

Integrating the latest forecasting techniques and broadening the scope of statistical models

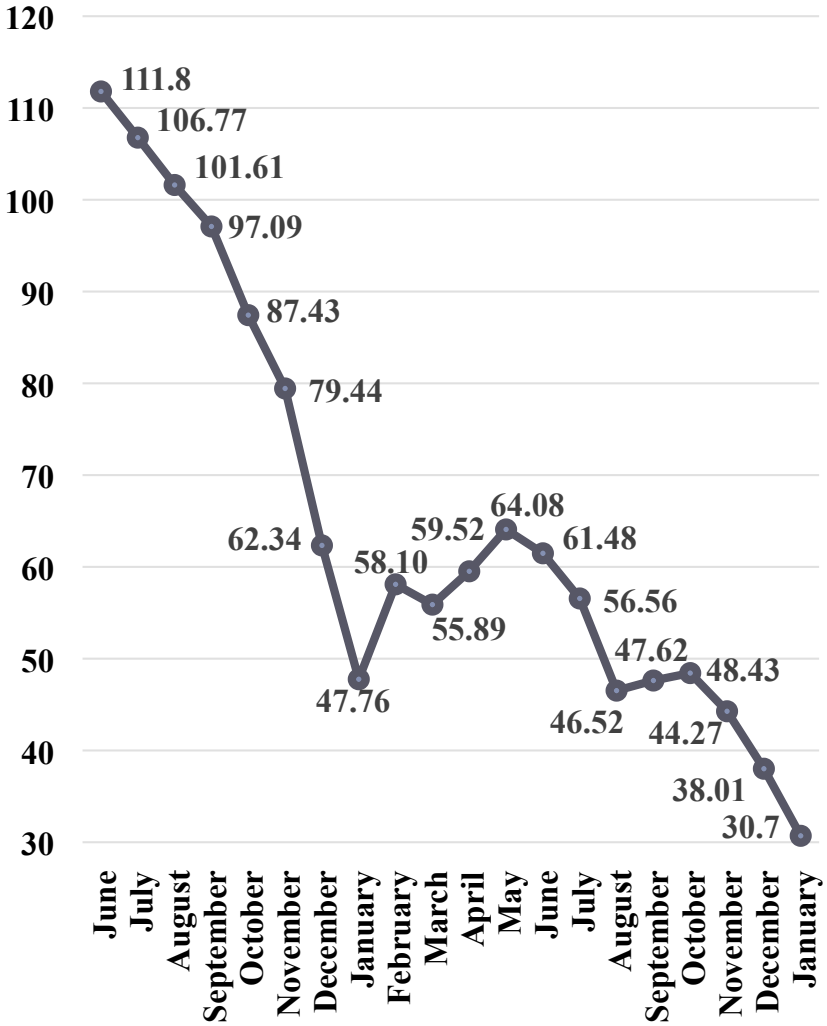
© Peter Kaznacheev, Khaznah Strategies Ltd. December 2016

The oil price collapse in 2014–16 was a classic "Black Swan event"

Various factors reinforced each other and culminated in an unanticipated market shift

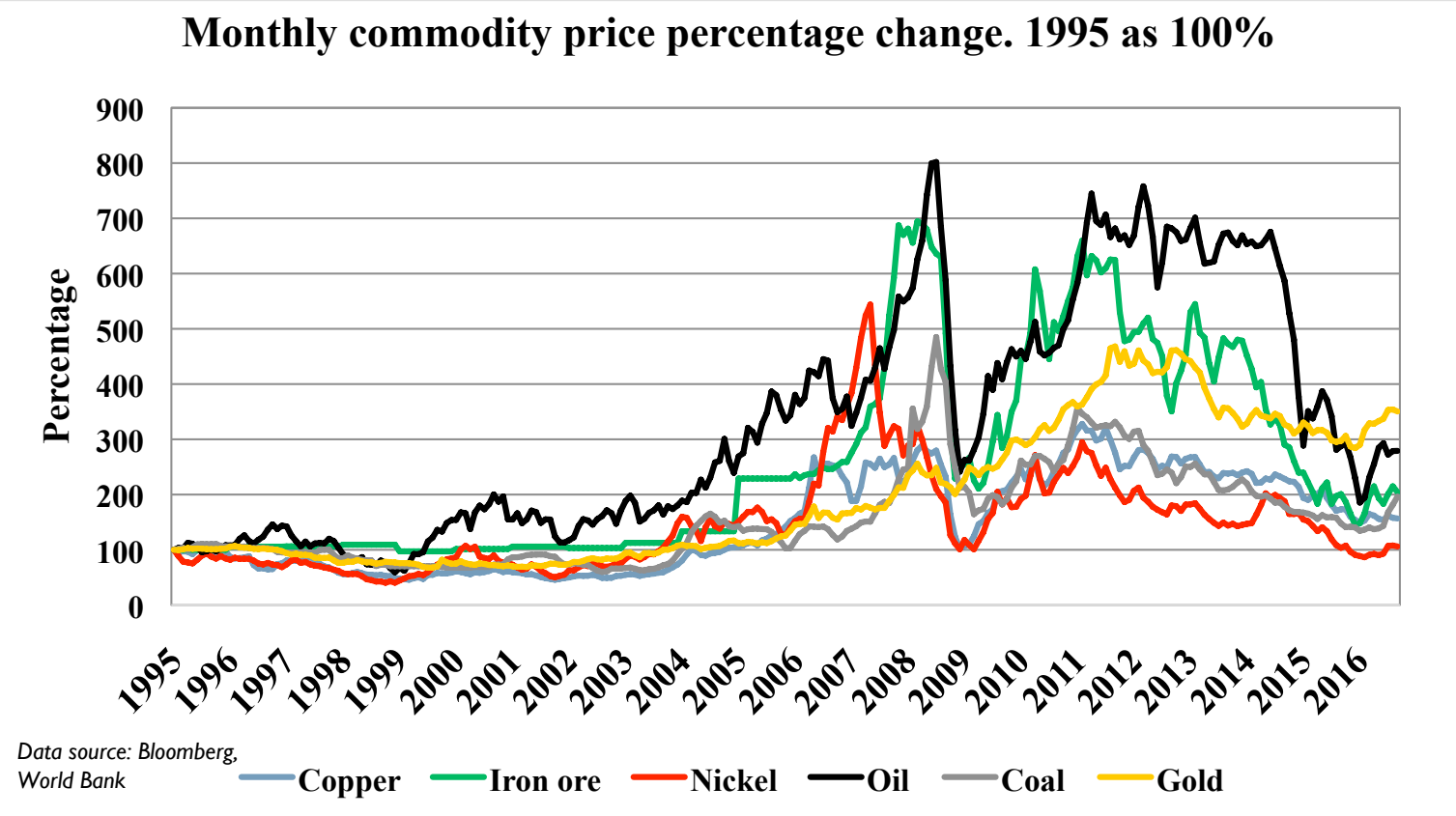
- The shale revolution significantly increased the supply of oil
- As U.S. became a leader in oil production, the political risk premium in the price decreased
- A stronger dollar added downward pressure on the price
- Economic growth slowed down in emerging markets – most importantly, in China
- The world made a big leap towards efficient energy consumption
- A surge in investment during high oil prices increased production and crude inventories
- OPEC's ability to influence the price dropped substantially
- To some degree, mineral commodities fell out of fashion among institutional investors

Brent crude monthly average USD per barrel, June 2014 – January 2016



Price uncertainty and a new approach to forecasting

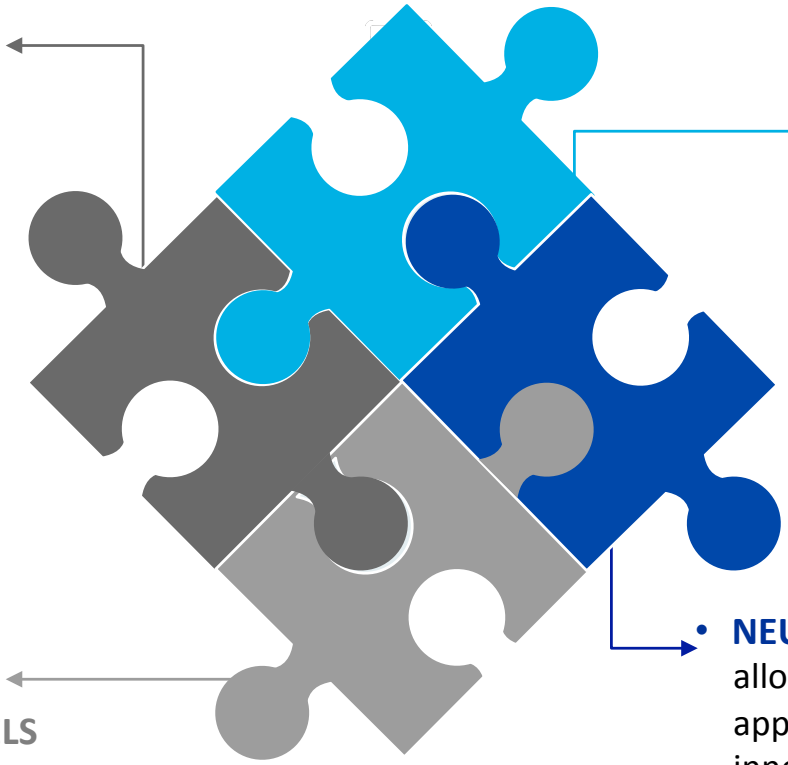
- In recent years, there has been significant instability in commodity markets. Prices of most metals and coal fell sharply in 2011, three years before the oil price plunge.
- Before the price drop, there was an overall tendency within oil companies not to attribute too much importance to structural forecasting methods – corporations largely relied on in-house expertise and scenario building.
- The price uncertainty created demand for deeper market analysis and new approaches to price forecasting.
- To meet that demand we developed an *integral price forecasting ©* method which combines four major techniques. We apply it to crude oil, regional gas markets, prices of petrol, diesel and jet fuel and also other commodities.
- Apart from the actual forecast, one of the key products of our models is a detailed map of main price drivers.



Our *integral price forecasting* © consolidates four major methods

We build multi-factor econometric models with an auto-regressive component (ARIMA or GARCH); identify optimal models based on opinions of industry experts; and selectively apply artificial neural networks and other innovative algorithms in order to improve forecast accuracy.

- **MULTI-FACTOR MODELS** determine the influence of various external factors on the commodity price.
- Such factors include industry fundamentals – supply, demand and inventories – along with financial market and macroeconomic data (misc. indices, GDP growth etc.).
- Vector auto-regressions (VAR) are a subset of such statistical models.



- **EXPERTS' FORECASTS** combine market insights with extensive industry experience.
- Methods include: consensus forecasting based on several experts' opinions, the Delphi method (relying on a panel of experts), and scenario forecasting.

- **AUTO-REGRESSIVE MODELS** are based on price time series.
- They forecast the future value of a commodity based on historic patterns of its price fluctuations.
- Major techniques include: ARIMA, ARCH and GARCH models.

- **NEURAL NETWORKS AND BIG DATA** allow to improve forecasting accuracy by applying machine learning and other innovative algorithms.
- In addition, forecasts can utilise studies of commodity market sentiment based on semantic analysis of news reports and "refined text mining".

Challenges and responses: key features of our price models

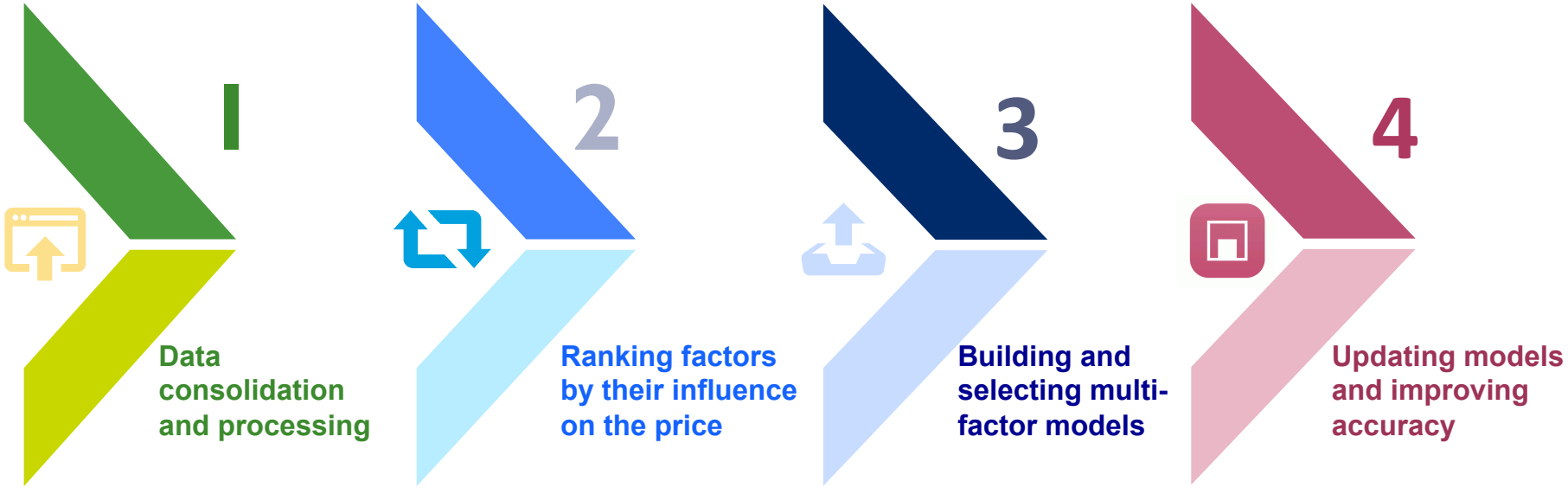
Main problems

- **Overreliance on auto-regressive models** (ARIMA, ARCH, GARCH) and less interest in structural multi-factor models.
- **Insufficient number of variables** evaluated for multi-factor models – the initial stage usually covers 5-6 basic factors (supply, demand, inventories, GDP growth etc.).
- Multi-factor models often include **factors without time lags** which leads to "forecasting with forecasts" instead of known data.
- Oil price analysis is developed within three relatively **disconnec-ted groups of forecasters**: traders, oil companies' economists, and academic researchers.
- There is demand for **deep structural analysis** and **integrating various methods.**

Our solutions and major steps

- Models should cover a very broad set of external factors – **from 50 to 100 variables** at the stage of initial evaluation. A sample list of factors is included in this presentation.
- All **models include an auto-regressive component** – they analyse previous oil price fluctuations with an ARIMA or GARCH element.
- We build **models for different time frequencies**: annual, quarterly, monthly and, sometimes, daily (when necessary).
- To get maximum accuracy we **evaluate a very large number of models – at least 5000** for one commodity. The whole process consists of four stages (see next slide).
- We select influencing factors with time lags so that our **models forecast based on actual known data** – not predictions.
- The final selection of models is made **in consultation with industry experts** – this integrates quantitative and qualitative analysis.
- The mean absolute percentage error – **MAPE of our models is low** – in the range of 3 to 7%. **Artificial neural networks** and other AI algorithms can be used to further improve accuracy.
- In addition to the actual forecast, one of the key products of our models is **a detailed map of main price drivers**.

Four stages of building multi-factor price models



Summary of activity	Collecting data by factors for the longest available time interval. Organising data by time frequency: annual, quarterly, monthly and daily.	Building pair regressions to determine the influence of each factor on the price according to its coefficients, p-value and R^2.	Building and evaluating 5000+ models for one commodity. Selecting optimal models based on econometric criteria and knowledge of the industry.	Updating models based on newest data; further improving model accuracy using neural networks and other AI algorithms.
Delive-rables	Identifying a list of factors which can influence the price. Creating a dataset for the selected factors.	Ranking factors by their significance and identifying those most relevant for the multi-factor model.	Final selection of forecasting models in consultation with a panel of industry experts.	Using selected models to make forecasts, build scenarios and identify major clusters of factors affecting the price.

Data evaluation for the crude price models. A brief overview

Over 100 factors identified for initial evaluation. They fall into three categories

Industry factors, global and by country (not a full list)

- Oil production
- Oil consumption
- Inventory change
- Proved reserves
- Storage capacity
- U.S. oil rigs in operation
- Gasoline production
- Refining capacity

Macroeconomic factors, global and by country (not a full list)

Major exporters and importers:

- GDP growth (+ global)
- Industrial production index (IP)
- Consumer price index (CPI)
- US Fed and ECB interest rates
- Purchasing Managers Index (PMI)
- Consumer Sentiment Index (CSI)
- Composite Leading Indicators (CLI)

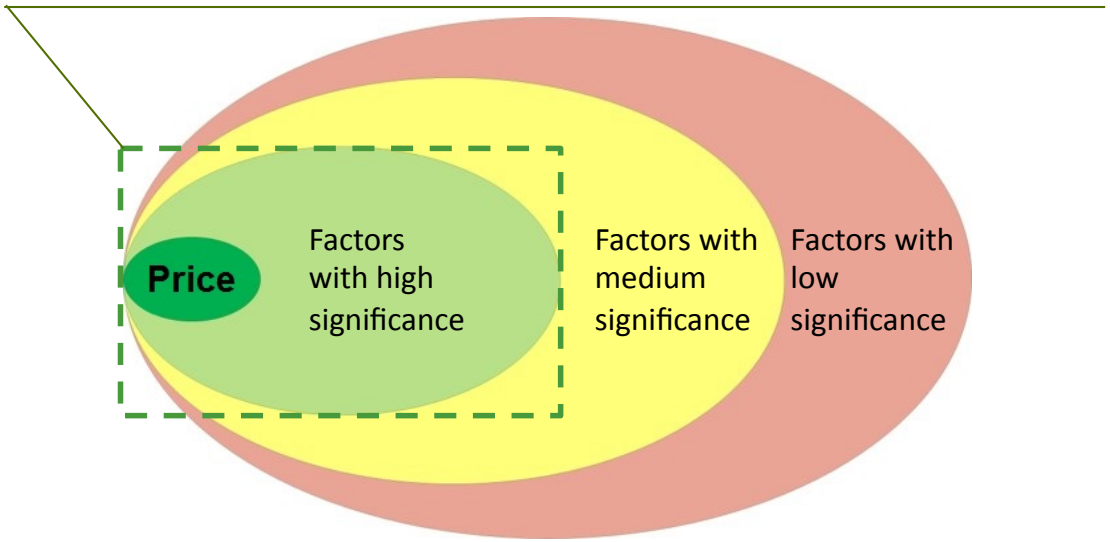
Stock, commodity and fin. market factors (not a full list)

- MSCI World Index
- S&P 500
- GSCI (Goldman Sachs Commodity Index)
- RICI (Rogers Internat-l Commodity Index)
- Exchange rates
- VIX (Volatility Index)

Factors ranked by their influence on the price

1. Evaluating factors in three main areas effecting the oil price: 1) financial market data; 2) industry data; 3) global and regional macroeconomic data.
2. Building pair regressions to determine the individual degree of influence by each factor on the oil price: coefficients, p-value, R^2.
3. Dividing all factors into 3 groups according to their influence on the oil price.

Factors that have the greatest degree of influence on the oil price



Main results of the crude price models

Over 5000 models built and evaluated. Two products: forecasts and a map of price drivers

5000+ MODELS EVALUATED

The selection of models is based on:

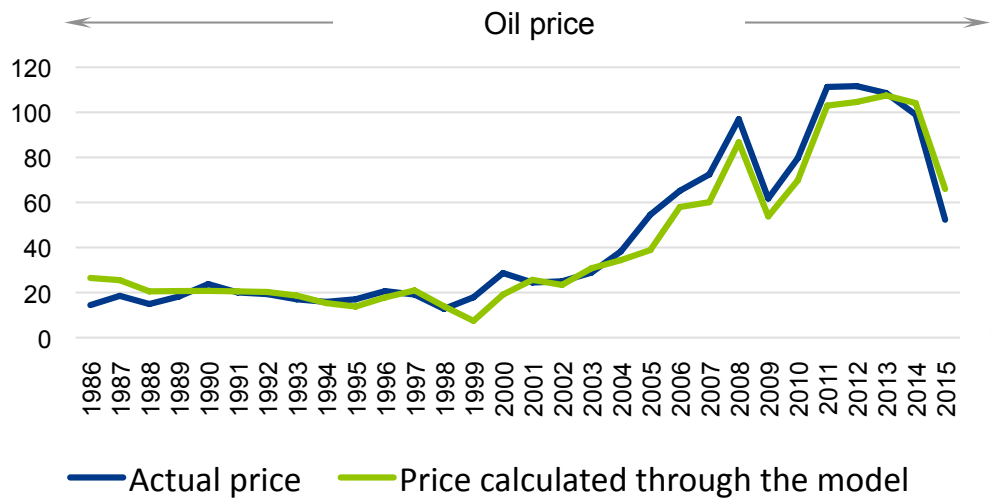
1. *The significance of factors*
2. *Predictive and explanatory strength*
3. *Economic logic (experts' opinions)*

Selected models are used in two ways:

Product 1: formulas for calculating the price forecast

Product 2: a map of major price drivers

Mean absolute percent. error (MAPE) of optimal price models is as low as 3.6%. Forecasts for years, quarters, months and days



Five clusters of crude price drivers were identified. Each of them contains specific indicators and indices