The new benchmark for forecasts of the real price of crude oil

Amor Aniss Benmoussa\textsuperscript{1} \hspace{5mm} Reinhard Ellwanger\textsuperscript{1} \hspace{5mm} Stephen Snudden\textsuperscript{2}

\textsuperscript{1}Bank of Canada \hspace{5mm} \textsuperscript{2}Wilfrid Laurier University

1st IAEE Online Conference

\textit{Wilfrid Laurier University}

The views expressed in this paper are those of the authors and no responsibility for them should be attributed to the Bank of Canada.
What is a good forecast?

- Suppose we want to evaluate a forecast of $X_{t+h}$
- Common practice: compare accuracy of your forecast against
  
  \[ no-change \text{ benchmark} \equiv X_t \]

- $X_t$ is optimal forecast under the random walk hypothesis
  
  $\rightarrow$ accuracy-improvements over NCB imply that
  
  - series is predictable in general
  - our forecast is more useful than “naive” approach

- **But: this approach not informative for temporally aggregated data**
  
  - lower-frequency series constructed by *averaging* or summation
    (e.g., real commodity prices, interest rates, ...)

The new benchmark

- Random walk hypothesis applies to high-frequency series
  - aggregated data are predictable by construction (Working 1960)
  - conventional NCB is not the optimal forecast
    → improvements over conventional NCB not informative about RWH

- New no-change benchmark:
  
  *last high-frequency observation*

  - optimal under random walk hypothesis
    → restores original interpretation of comparison with NCB
  - $\approx$ 45% improvement in MSPE for monthly/quarterly averages of daily data for 1-step-ahead prediction
Application to the real price of crude oil

- Real price of crude oil typically based on averaged data (Kilian 2009; Baumeister & Hamilton 2019)

- Existing literature: model-based forecasts beat NCB (Baumeister & Kilian 2012, 2014, 2015; Alquist et al. 2013; Snudden, 2018; Funk, 2018; Garratt et al., 2019)

- New no-change benchmark based on monthly closing price of oil
  - $\approx 40\%$ improvements in accuracy at 1-step-ahead prediction
  - using closing prices for estimation improves traditional models
  - but: most models do not beat the new benchmark

$\rightarrow$ oil prices are more difficult to predict than previously thought
Intuition under the RW null hypothesis

Forecaster’s goal is to predict $X_{t+h}$ given time $t$ information

$$X_t \equiv \frac{1}{n} \sum_{i=1}^{n} y_{t,i}$$

$t =$ month; $i =$ day of month; $n =$ # of days in month

Null hypothesis: daily observations follow random walk

$$y_{t,i} = y_{t,i-1} + \epsilon_{t,i}, \quad \text{for } i = 1, \cdots, n$$

$\epsilon_{t,i}$ is a mean-zero, iid error term with variance $\sigma_{\epsilon}^2$

- What is the optimal forecast in this setting?
- What are the consequences of using $X_t$ to evaluate forecasts?
The optimal forecast under the RWH

▶ Optimal forecast in MSPE terms is the conditional expectation

\[ E_t(X_{t+h}) = E(X_{t+h}|y_{1,1}; \ldots; y_{t,n}; Z_1; \ldots; Z_t) \]

▶ RWH: conditional expectation of each future daily obs

\[ E_t(y_{t+h,i}) = y_{t,n} \quad \text{for all } h > 0, \ i = 1, \ldots, n \]

⇒ conditional expectation of each future average obs

\[ E_t(X_{t+h}) = E_t\left(\frac{1}{n} \sum_{1}^{n} y_{t+h,n}\right) = y_{t,n} \]

▶ New benchmark is the optimal forecast under the RWH:

last high-frequency observation in period \( t \) \( \equiv y_{t,n} \)
Comparing the no-change forecasts $X_t$ and $y_{t,n}$

What is the MSPE of $y_{t,n}$ relative to the conventional NCB $X_t$?

MSPE for conventional benchmark:

$$E \left[ (X_t - X_{t+h})^2 \right] = \left( (h - 1) \cdot n + \frac{(n + 1) \cdot (2n + 1)}{6n} + \frac{(n - 1) \cdot (2n - 1)}{6n} \right) \sigma_e^2$$

MSPE for new benchmark:

$$E \left[ (y_{t,n} - X_{t+h})^2 \right] = \left( (h - 1) \cdot n + \frac{(n + 1) \cdot (2n + 1)}{6n} \right) \sigma_e^2$$

$y_{t,n}$ has lower MSPE than $X_t$ under the RWH

All forecasting gains occur at the 1-step ahead prediction

- forecast gain is constant, forecast error increases with $h$
  $$\Rightarrow$$ relative improvements largest for short horizons
MSPE ratios of the two NCBs

Theoretical MSPE improvements for monthly data

% MSPE improvement

Forecast horizon (months)
Consequences for forecast comparisons

- Conventional NCB is not the optimal forecast under the RWH
- Moreover, $\Delta X_t$ is autocorrelated (Working 1960)

$$X_t - X_{t-1} = \frac{1}{n} \left[ \sum_{i=1}^{n} i \cdot \epsilon_{t-1,i} + \sum_{j=1}^{n} (n + 1 - j) \cdot \epsilon_{t,j} \right]$$

⇒ improvements over no-change forecast $X_t$ are expected even under the random walk null hypothesis
⇒ improvements over no-change forecast $y_{t,n}$ are evidence against the random walk null hypothesis

- Using $y_{t,n}$ as a benchmark
  - maintains original spirit of comparisons with the NCB
  - more difficult to achieve when HF observations are persistent
Application: Forecasting the real price of crude oil

- Forecast of the monthly real price of crude oil in standard setting (Baumeister & Kilian 2012, 2014, 2015; Alquist et al. 2013)
  - real oil price is deflated average of daily nominal prices
    \[ \bar{p}_t^r = \frac{1}{n} \sum_{i=1}^{n} \frac{p_{t,i}}{CPI_t} \]
  - goal: forecast \( \bar{p}_{t+h}^r \) given month \( t \) information

- The new benchmark: series of monthly closing prices
  \[ p_{t,closing}^r = \frac{p_{t,n}}{CPI_t} \]
Econometric models

1. Univariate time series models
   - AR (log-level and percent changes)
   - ARMA(1,1)
   - ARFI

2. VAR models (Kilian & Murphy 2014)
   - percent change in global crude oil production
   - real economic activity indicator (Kilian 2009)
   - real price of oil
   - change in above-ground global crude oil inventories
     > unrestricted least-squares estimation
     > Gaussian BVAR with prior variance (Giannone, Lenza & Primiceri 2010)

3. Futures price curve

4. Equal-weight forecast combination
Implementation

Following Baumeister & Kilian (2012):

- Real-time data
- Out-of-sample forecasts with expanding window
  - estimation period starts 1973M2
- All models estimated with average prices

- Criteria: MSPE ratio and directional accuracy
  - displayed relative to conventional NCB
  - improvements over the new NCB are in **bold**
  - standard tests for inference
    (Diebold & Mariano 1995, Pesaran and Timmermann 2009)
Baseline results: Real WTI prices, real-time data

<table>
<thead>
<tr>
<th>Months Ahead</th>
<th>Last WTI Close Price</th>
<th>BVAR(12)</th>
<th>VAR(12)</th>
<th>AR(12)</th>
<th>AR(12) %Δ</th>
<th>ARFI(1)</th>
<th>ARMA(1,1)</th>
<th>Futures Curve</th>
<th>Model Averaging</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.61***</td>
<td>0.97</td>
<td>1.01</td>
<td>0.94</td>
<td>0.95*</td>
<td>0.93</td>
<td>0.92</td>
<td>1.00</td>
<td>0.91</td>
</tr>
<tr>
<td>3</td>
<td>0.89**</td>
<td>1.00</td>
<td>1.00</td>
<td>0.97</td>
<td>0.99</td>
<td>0.96</td>
<td>0.95</td>
<td>0.97</td>
<td>0.92</td>
</tr>
<tr>
<td>6</td>
<td>0.95**</td>
<td>1.05</td>
<td>1.04</td>
<td>1.00</td>
<td>1.04</td>
<td>0.99</td>
<td>0.95</td>
<td>0.96</td>
<td>0.95</td>
</tr>
<tr>
<td>12</td>
<td>0.96**</td>
<td>1.10</td>
<td>1.11</td>
<td>1.00</td>
<td>1.10</td>
<td>1.00</td>
<td>0.94</td>
<td><strong>0.85</strong></td>
<td>0.94</td>
</tr>
<tr>
<td>24</td>
<td>0.99</td>
<td>1.08</td>
<td>1.06</td>
<td>1.02</td>
<td>1.19</td>
<td>1.04</td>
<td>0.96</td>
<td><strong>0.82</strong></td>
<td>0.91</td>
</tr>
</tbody>
</table>

**MSPE Ratios**

**Success Ratios**

- **Note:** *****, **, and * denote significant improvement over the average no-change forecast at the 1%, 5%, and 10% level. **Bold values indicate significant improvements over the last closing price no-change forecast at the 5 percent level.**
Evolution of baseline WTI Real price forecasts: Real-time data

**Note:** Dynamic, recursive, out-of-sample forecasts 1992M1–2018M12. The forecast criteria reported include the recursive MSPE expressed as a ratio relative to the monthly average no-change forecast. All forecast criteria are evaluated in the levels of the real price of oil. The first 30 months are dropped to reduce starting-point effects.
Interpreting baseline results

- Large(!) forecasting gains for short forecast horizons

- Actual vs. predicted MSPE ratios (RWH for daily data)

<table>
<thead>
<tr>
<th>Horizon (months)</th>
<th>1</th>
<th>3</th>
<th>6</th>
<th>12</th>
<th>24</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theoretical</td>
<td>0.54</td>
<td>0.88</td>
<td>0.95</td>
<td>0.97</td>
<td>0.99</td>
</tr>
<tr>
<td>Empirical (revised data)</td>
<td>0.60</td>
<td>0.89</td>
<td>0.95</td>
<td>0.96</td>
<td>0.99</td>
</tr>
</tbody>
</table>

- Robustness: remarkably similar results for
  - various estimation / evaluation periods
  - different oil price series (Brent, RAC, nominal prices)
  - ex-post revised data, nominal price, quarterly data

- What about estimating models with closing prices?
Models estimated with last closing price

<table>
<thead>
<tr>
<th>Months Ahead</th>
<th>Last WTI Close Price</th>
<th>BVAR(12)</th>
<th>VAR(12)</th>
<th>AR(2)</th>
<th>AR(1) %Δ</th>
<th>ARFI(2)</th>
<th>ARMA(1,1)</th>
<th>Last Future Price Curve</th>
<th>Model Averaging</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.61***</td>
<td>0.62**</td>
<td>0.73*</td>
<td>0.57***</td>
<td>0.62***</td>
<td>0.57***</td>
<td>0.57***</td>
<td>0.60***</td>
<td>0.58***</td>
</tr>
<tr>
<td>3</td>
<td>0.89**</td>
<td>0.90</td>
<td>0.91</td>
<td>0.85</td>
<td>0.93*</td>
<td>0.86</td>
<td>0.86</td>
<td>0.88**</td>
<td>0.82*</td>
</tr>
<tr>
<td>6</td>
<td>0.95**</td>
<td>0.98</td>
<td>0.97</td>
<td>0.89</td>
<td>1.04</td>
<td>0.89</td>
<td>0.90</td>
<td>0.92*</td>
<td>0.87</td>
</tr>
<tr>
<td>12</td>
<td>0.96**</td>
<td>1.05</td>
<td>1.06</td>
<td>0.90</td>
<td>1.15</td>
<td>0.90</td>
<td>0.91</td>
<td><strong>0.84</strong></td>
<td>0.87</td>
</tr>
<tr>
<td>24</td>
<td>0.99</td>
<td>1.06</td>
<td>1.06</td>
<td>0.95</td>
<td>1.48</td>
<td>0.95</td>
<td>0.97</td>
<td><strong>0.80</strong></td>
<td>0.87</td>
</tr>
</tbody>
</table>

MSPE Ratios

<table>
<thead>
<tr>
<th>Months Ahead</th>
<th>Last WTI Close Price</th>
<th>BVAR(12)</th>
<th>VAR(12)</th>
<th>AR(2)</th>
<th>AR(1) %Δ</th>
<th>ARFI(2)</th>
<th>ARMA(1,1)</th>
<th>Last Future Price Curve</th>
<th>Model Averaging</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.71***</td>
<td>0.71***</td>
<td>0.66***</td>
<td>0.71***</td>
<td>0.73***</td>
<td>0.72***</td>
<td>0.71***</td>
<td>0.73***</td>
<td>0.70***</td>
</tr>
<tr>
<td>3</td>
<td>0.60***</td>
<td>0.58**</td>
<td>0.56**</td>
<td>0.57**</td>
<td>0.62***</td>
<td>0.60***</td>
<td>0.60***</td>
<td>0.60***</td>
<td>0.59***</td>
</tr>
<tr>
<td>6</td>
<td>0.56**</td>
<td>0.58**</td>
<td>0.58**</td>
<td>0.52</td>
<td>0.58**</td>
<td>0.51</td>
<td>0.52</td>
<td>0.55</td>
<td>0.55</td>
</tr>
<tr>
<td>12</td>
<td>0.59***</td>
<td>0.54</td>
<td>0.58**</td>
<td>0.54</td>
<td>0.59**</td>
<td>0.51</td>
<td>0.53</td>
<td><strong>0.63</strong></td>
<td>0.54</td>
</tr>
<tr>
<td>24</td>
<td>0.53</td>
<td>0.51</td>
<td>0.57*</td>
<td>0.58</td>
<td>0.56</td>
<td><strong>0.55</strong></td>
<td>0.54</td>
<td><strong>0.64</strong></td>
<td>0.56</td>
</tr>
</tbody>
</table>

**Note:** ***, **, and * denote significant improvement over the average no-change forecast at the 1%, 5%, and 10% level. **Bold values indicate significant improvements over the last closing price no-change forecast at the 5 percent level.**
Evolution of real-time closing-price WTI forecasts

Note: Dynamic, recursive, out-of-sample forecasts 1992M1–2018M12. The forecast criteria reported include the recursive MSPE expressed as a ratio relative to the monthly average no-change forecast. All forecast criteria are evaluated in the levels of the real price of oil. The first 30 months are dropped to reduce starting-point effects.
Take-away

- Aggregating higher-frequency data
  - introduces loss of information (e.g., Rossana & Seater 1995)
  - changes interpretation of standard forecast comparisons
  → use closing observations for estimation & forecast evaluation

- New benchmark for real price of crude oil changes assessments of models & oil price predictability:
  - daily oil prices are “random walkish”
  - real price of crude oil is predictable by construction
  - but most models do not beat the new benchmark

- Averaging can be desirable, but watch out for settings in which
  - series of interest is temporally aggregated
  - underlying data is persistent
Thank you!