Forecasting Canadian GDP

Evaluating Point and Density Forecasts in Real-Time

Frédérick Demers

Research Department
Bank of Canada

Bank of Canada Workshop
Forecasting Short-term Economic Development and the Role of Econometric Models
October 25 and 26, 2007
Outline

1. Motivation, Data, and Notation
2. Forecasting Models and Set-Up of Experiment
3. Forecast Evaluation
4. Conclusion
Part I

Motivation, Data, and Notation
Motivation (what this paper does)

- Evaluate point and density forecasts in real time
- Compare linear and nonlinear univariate models
  - Clements and Krolzig (1998): Nonlinear models fit US GDP well in-sample, but don’t forecast that well out of sample
  - Clements and Smith (2002): Nonlinear models provide better density forecasts
Can we *robustify* linear models by using less time-information?
  - We know it works well for point forecasts
  - Does it work for density forecasts?

Compare various forecasting strategies (time-information, or *limited-memory* estimators)

*Real-time vs. revised* data
Motivation (what this paper doesn’t do)

- Account for parameter uncertainty in analytic expressions
- Multivariate models, e.g.:
  - Output and unemployment: Clements and Smith (2000)
  - Output and inflation
  - Money and inflation – see Shaun’s paper
- Relax the Gaussianity assumption for marginal distributions
  - Few conclusive examples for GDP
  - Yet, some predictive densities will not be Gaussian
- No quantile estimation
Why Investigate Interval/Density Forecasts?

- Natural generalization of point (conditional-mean) forecasts
- Common in finance (VaR) or weather forecasting
- But most macroeconomic forecasts are reported as point
  - ...seems odd when econometrics is about inference
- Notable exceptions:
  - *Fan Charts* from Bank of England and Riksbank
  - Increasing number of statements about recession probability
  - *Survey of Professional Forecasters*
Point forecasts provide little information about the likelihood of the possible outcomes.

While discussing *risks* without the associated likelihood is not very informative.

- Ask your insurance broker...
Some Useful Literature

- Predicting recessions: Kling (1987) and Zellner, Hong, and Min (1991)
- Comprehensive review: Corradi and Swanson (2005)
Data Set

- Real GDP at market prices, seasonally adjusted
- Sample: 1961Q1 - 2006Q4
- Forecast Period: 1990Q1 - 2006Q4
- Real-time vintages of GDP are used
- results based on real-time data compared with those based revised data
Initial vs. Final Estimates of Quarterly Real GDP Growth

Q/Q Real GDP Growth

% Time

-1.6 -0.8 0.0 0.8 1.2 1.6 2.0

Real-Time
Revised
Let $Y_t$ denote the log of real GDP times 100 with $t = 1, \ldots, T$

And $y_{t+h} = Y_{t+h} - Y_t$ denotes $h$-step ahead change of $Y_t$

with $h = 1, \ldots, H$

The usual first difference will be $y_t = Y_t - Y_{t-1}$

Finally

$$y_{t+h} \equiv \hat{Y}_{t+h} + \hat{\epsilon}_{t+h}$$
Part II

Forecasting Models and Set-Up of Experiment
Benchmark Linear Models

- Unconditional:

\[
UNC = S^{-1} \sum_{j=t-S+1}^{t} (Y_j - Y_{j-h})
\]

\[
\varepsilon_{t+h} \sim N(0, \sigma^2_{\varepsilon})
\]

where $S$ is a sample-size of interest

- AR($p$):

\[
y_{t+1} = \alpha + \phi(L)y_t + \varepsilon_{t+1}
\]

\[
\phi(L) = \phi_1L - \ldots - \phi_pL^p
\]

\[
\varepsilon_t \sim i.i.d. N(0, \sigma^2_{\varepsilon})
\]
Smooth-Transition Switching AR Models

- Exponential smooth transition AR, ESTAR

\[ y_{t+1} = \alpha_1 + \phi_1(L)y_t + \omega_t (\alpha_2 + \phi_2(L)y_t) + \varepsilon_{t+1} \]

\[ \omega_t = 1 - \exp(-\gamma(y_{t-d} - \mu)^2) \]

- Logistic smooth transition AR, LSTAR:

\[ y_{t+1} = \alpha_1 + \phi_1(L)y_t + \omega_t (\alpha_2 + \phi_2(L)y_t) + \varepsilon_{t+1} \]

\[ \omega_t = \frac{1}{1 + \exp(-\gamma(y_{t-d} - \mu))} \]

Where
- \( \varepsilon_t \sim i.i.d. N(0, \sigma_{\varepsilon}^2) \)
- \( d \) is a delay parameter with \( p \geq d \geq 0 \)
- \( \gamma(> 0) \) determines the shape of transition function, \( \omega_t \)
Markov-Switching AR Models

- The intercept switching AR, MSI:
  \[ y_{t+1} = \alpha_{s_t} + \phi(L)y_t + \varepsilon_{t+1} \]

- The intercept switching and AR-coefficient switching, MSIAR:
  \[ y_{t+1} = \alpha_{s_t} + \phi_{s_t}(L)y_t + \varepsilon_{t+1} \]

- Processes are homoscedastic:
  \[ \varepsilon_t \sim i.i.d.N(0, \sigma^2_{\varepsilon}) \]
Markov-Switching AR Models with Heteroscedasticity

- MS models with state-dependent variance are also examined
  \[ \varepsilon_t \sim i.i.d. N(0, \sigma_{s_t}^2) \]
- MSIH and MSIHAR
Nonlinear Models and Forecast Distribution

- Forecast distribution can depart from normality
  - Will generate excess skewness - asymmetric risks
  - Will generate excess kurtosis - recession/boom
- Although the marginal distributions are normal
Linear and Nonlinear Univariate Forecasting Models

- Unconditional forecast
- AR
- ESTAR
- LSTAR
- MSI, MSIH
- MSIAR, MSIHAR
Rolling vs. Expanding Schemes

Expanding window:
- Add an observation to the sample at each iteration

Rolling window:
- Roll the sample forward at each iteration: \( S = \text{EXP} \)
- Various sample sizes are compared: \( S = 30, 40, 50, 60, 70, 80 \)
- The so-called \textit{limited-memory} estimator

The rolling approach is advantageous if uncertain about homogeneity of DGP (Giacomini and White, 2006; Clark and McCracken, 2004)
Lag Selection

- For the AR model, lags are selected by AIC at each period
  - The maximum lag is 4
- For the ESTAR, LSTAR, and MS models, a single lag is used
  - Computationally cumbersome otherwise
  - No *insanity filter*
  - But a few conditional statements about numerical convergence
Forecasting $h$-step Ahead

- Values for $y_{t+h}$ are obtained by recursion (or iteration)
  - i.e., the *iterated* forecast method, not the *direct*
- Analytic expressions can be used for the AR and MS models
- Stochastic simulations are necessary for smooth-transition models when $h > 1$
Need to draw pseudo-random value for $\varepsilon_t$

- Easy to do when we draw from Gaussian
  - Or when we bootstrap
- I choose to draw from the Gaussian to emphasize on model specification
  - 1000 replications
  - Each point and density forecast is the average over the simulated values (when $h > 1$)
Interval/Density Forecasting with AR Models

- Estimate parameters (intercept and AR parameters)
- Obtain an estimate of $\sigma^2 = E(\varepsilon_t^2)$
Because the underlying process, $Y_t$, is $I(1)$, the $h$-step forecast-error variance, denoted as $\hat{\Omega}_h$, depends on $\sigma^2$, $\sigma^2_h$, and $h$.

But we estimate models based upon $h = 1$, so $\sigma^2_h$ and $\hat{\Omega}_h$ must be derived for $h > 1$.

Recall that for a stationary AR(1) process the $h$-step variance is

$$\sigma^2_h = \sigma^2 \frac{1 - \phi^{2h}}{1 - \phi^2}$$
The $h$-step error, $\varepsilon_{t+h} = Y_{t+h} - Y_t$, is a cumulative process.

- N.B. $\varepsilon_{t+h}$ is (at most) a $MA(h-1)$ process

Hence $\hat{\Omega}_h$ increases at rate $O(h)$, in contrast to $O(1)$ when the underlying process is $I(0)$.

$\hat{\Omega}_h$ can be approximated by

$$h\hat{\sigma}_h^2(1 + h/T)$$
Density forecasts are constructed the same as linear models when $\sigma = \sigma_t$.

The p.d.f. of $\varepsilon_{t+h}$ is normal although the p.d.f. of $y_{t+h}$ is not.

When $\sigma \neq \sigma_t$, the forecast-error distribution will vary over time.

And will not be normal at each $t$ due to the mixture process.
Part III

Forecast Evaluation
Evaluating Point Forecasts

- Compute the bias
- Variance
- Mean Squared Error (MSE)

\[ MSE = \varepsilon'_{t+h} \varepsilon_{t+h}/(P - h) \]
Evaluating Density Forecasts: Some Background

The idea:
- Determine whether a predicted density function is *identical* to some distribution of interest
- A forecasting model is judged as *good* or *bad* based on the probabilities it predicts (Dawid, 1984)
  - Model don’t need to agree with economic theory
- Are the probabilities well calibrated?
Let $F_{t+h}$ denote the empirical distribution function of the process $\hat{y}_{t+h}$.

We want to know whether the realizations $\{y_{t+h}\}_{t=1}^S$ are drawn from $F_{t+h}$.

Can consider the probability integral transform (p.i.t.):

$$z_{t+h} = \int_{-\infty}^{y_{t+h}} F(u)du,$$

where $F_t$ is the (unobserved) density governing the process.

Or, the probability of observing values no greater than the realizations.

Densities need not be constant over time.
When the predicted density, $F_{t+h}$, correspond to the underlying density, $F_{t+h}$, then

$$z_{t+h} \sim i.i.d. U[0, 1]$$

Which means testing that $F - F = 0$

Or that $z_{t+h}$ departs from the $45^\circ$ line

N.B. when $h > 1$, the i.i.d. assumption will in general be invalid

  • Inference?
Evaluating Density Forecasts: Testing Strategies

- Can be done using Kolmogorov-Smirnov or Cramer-von-Mises GoF
- Alternative strategy: take the inverse normal CDF transformation of $z_t$, $z_t^*$, and use normality tests on $z_t^*$ (Berkowitz, 2001)
Forecast Evaluation

Frédéric Demers  
Forecasting Canadian GDP
Using More Time-Information Leads to Biased Predictions

Frédéric Demers
Forecasting Canadian GDP
Variance Ratios: Limited Information at Long Horizons

Frédéric Demers  
Forecasting Canadian GDP
Relative MSEs: Bias Makes a Big Difference

- **ESTAR vs. AR (S=60)**
- **LSTAR vs. AR (S=60)**
- **MSIAR vs. AR (S=60)**
- **MSIHAR vs. AR (S=60)**
- **MSIH vs. AR (S=60)**
- **MSI vs. AR (S=60)**
Real-time vs. Revised Estimates of $\Omega_1$ for AR

AR ($S=30$, $h=1$)

AR ($S=50$, $h=1$)

AR ($S=80$, $h=1$)

AR ($S=\text{EXP}$, $h=1$)
Empirical Cumulative Density Function of $z_{t+h}$
Empirical Cumulative Density Function of $z_{t+h}$
Cramer-von-Mises Test Results

- AR CvM
- ESTAR CvM
- LSTAR CvM
- MSI CvM
- MSIH CvM
- MSIHAR CvM

Frédérick Demers  
Forecasting Canadian GDP
Doornick-Hansen Normality Test Results

- AR DH
- ESTAR DH
- LSTAR DH
- MSI DH
- MSIIH DH
- MSIAR DH
- MSIHAR DH

Frédéric Demers  Forecasting Canadian GDP
Part IV

Conclusion
Limited Information Content for all Models

- We can’t predict too far out!
- Too much time information tends to lead to biased forecasts
  - And bias can be large
  - MSE or Variance ratio?
- Possible to robustify linear model point and density forecasts against structural changes
- Nonlinearities do matter for point and density forecasts
Smaller $S$ forecast better with revised data for short horizons

- Information content (Galbraith and Tkacz, 2007) looks better with revised data
  - More models are informative at long horizons

- Uncertainty looks smaller with real-time data (in absolute terms)

- Nonlinear look worse (higher MSE) with revised data
Thank You!