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

Motivation (what this paper does)

- 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
  - Hansen (2006) and Wu (2006)
- 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*

Why Investigate Interval/Density Forecasts?

- 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...

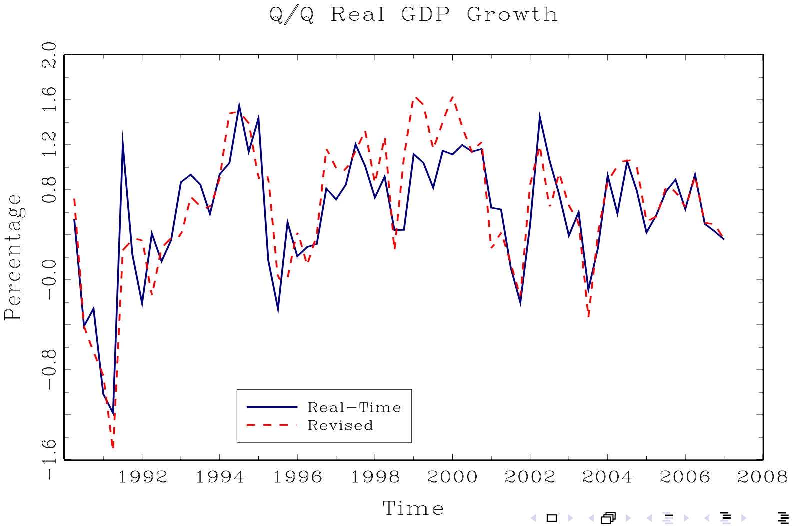
Some Useful Literature

- Introduction of principles to economics: Dawid (1984)
- Predicting recessions: Kling (1987) and Zellner, Hong, and Min (1991)
- Review: Tay and Wallis (2000)
- Applications: Clements (2004), Galbraith and van Norden (2007)
- 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



Notation

- Let Y_t denote the log of real GDP times 100 with $t = 1, ..., T$
- And $y_{t+h} = Y_{t+h} - Y_t$ denotes h-step ahead change of Y_t with $h = 1, ..., H$
- The usual first difference will be $y_t = Y_t - Y_{t-1}$
- Finally

$$y_{t+h} \equiv \hat{y}_{t+h} + \hat{\varepsilon}_{t+h}$$

Part II

Forecasting Models and Set-Up of Experiment

Benchmark Linear Models

- Unconditional:

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

$$\varepsilon_{t+h} \sim N(0, \sigma_\varepsilon^2)$$

where S is a sample-size of interest

- AR(p):

$$\begin{aligned} y_{t+1} &= \alpha + \phi(L) y_t + \varepsilon_{t+1} \\ \phi(L) &= \phi_1 L - ... - \phi_p L^p \\ \varepsilon_t &\sim i.i.d.N(0, \sigma_\varepsilon^2) \end{aligned}$$

Smooth-Transition Switching AR Models

- Exponential smooth transition AR, ESTAR

$$\begin{aligned} y_{t+1} &= \alpha_1 + \phi_1(L)y_t + \omega_t\left(\alpha_2 + \phi_2(L)y_t\right) + \varepsilon_{t+1} \\ \omega_t &= 1 - \exp(-\gamma(y_{t-d} - \mu)^2) \end{aligned}$$

- Logistic smooth transition AR, LSTAR:

$$\begin{aligned} y_{t+1} &= \alpha_1 + \phi_1(L)y_t + \omega_t\left(\alpha_2 + \phi_2(L)y_t\right) + \varepsilon_{t+1} \\ \omega_t &= \frac{1}{1 + \exp(-\gamma(y_{t-d} - \mu))} \end{aligned}$$

- 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, ω_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:

$$\varepsilon_t \sim i.i.d.N(0, \sigma_\varepsilon^2)$$

Markov-Switching AR Models with Heteroscedasticity

- MS models with state-dependent variance are also examined

$$\varepsilon_t \sim i.i.d.N(0, \sigma^2_{s_t})$$

- 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: $S = EXP$
  - Various sample sizes are compared: $S = 30, 40, 50, 60, 70, 80$
  - The so-called *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 $h > 1$

h-Step Forecasts with Smooth-Transition Models

- Need to draw pseudo-random value for ε_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 $\sigma^2 = E(\varepsilon_t^2)$

Density Forecasting with AR(p) Models

- Because the underlying process, Y_t, is $I(1)$, the h-step forecast-error variance, denoted as $\hat{\Omega}_h$, depends on σ^2, σ_h^2, and h
- But we estimate models based upon $h = 1$, so σ_h^2 and $\hat{\Omega}_h$ must be derived for $h > 1$
- Recall that for a stationary AR(1) process the h-step variance is

$$\sigma_h^2 = \sigma^2 \frac{1 - \phi^{2h}}{1 - \phi^2}$$

Density Forecasting with AR(p) Models

- The h-step error, $\varepsilon_{t+h} = Y_{t+h} - Y_t$, is a cumulative process
  - N.B. ε_{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 Forecasting with Switching Models

- Density forecasts are constructed the same as linear models when $\sigma = \sigma_t$
- The p.d.f. of ε_{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?

Evaluating Density Forecasts: Overview

- g
- Let $\mathbb{F}_{t+h}$ 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 $\mathbb{F}_{t+h}$
- Can consider the probability integral transform (p.i.t.):

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

  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

Evaluating Density Forecasts: Testing Strategies

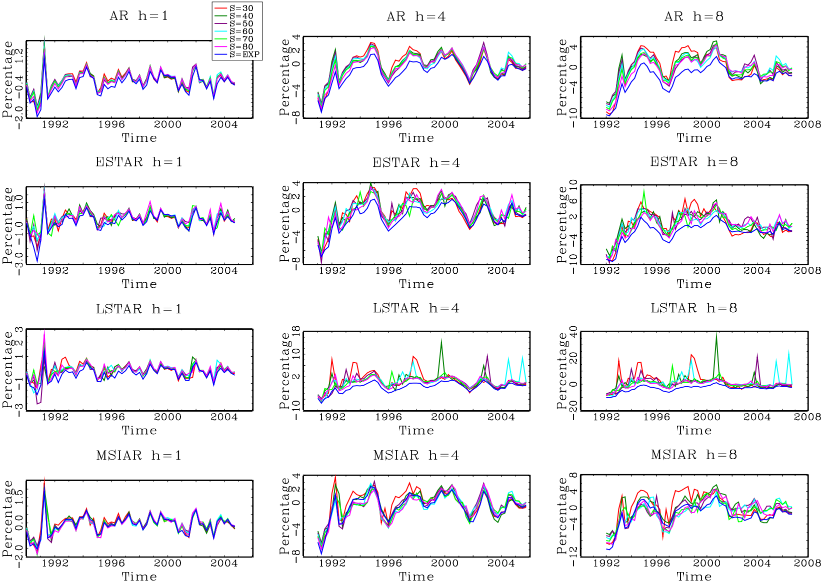
- When the predicted density, $\mathbb{F}_{t+h}$, correspond to the underlying density, F_{t+h}, then

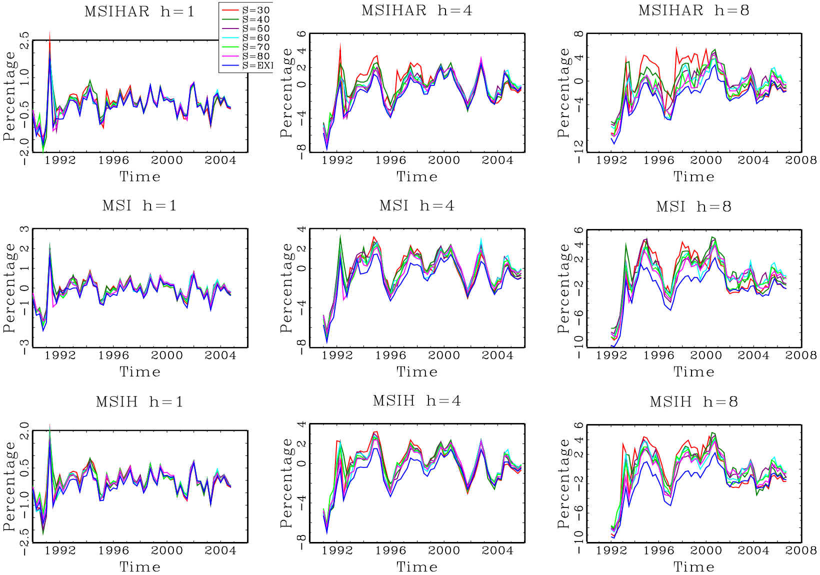
$$z_{t+h} \sim i.i.d.U[0,1]$$

- Which means testing that $\mathbb{F} - F = 0$
- Or that z_{t+h} departs from the 45° line
- N.B. when $h > 1$, the i.i.d. assumption will in general be invalid
  - Inference?

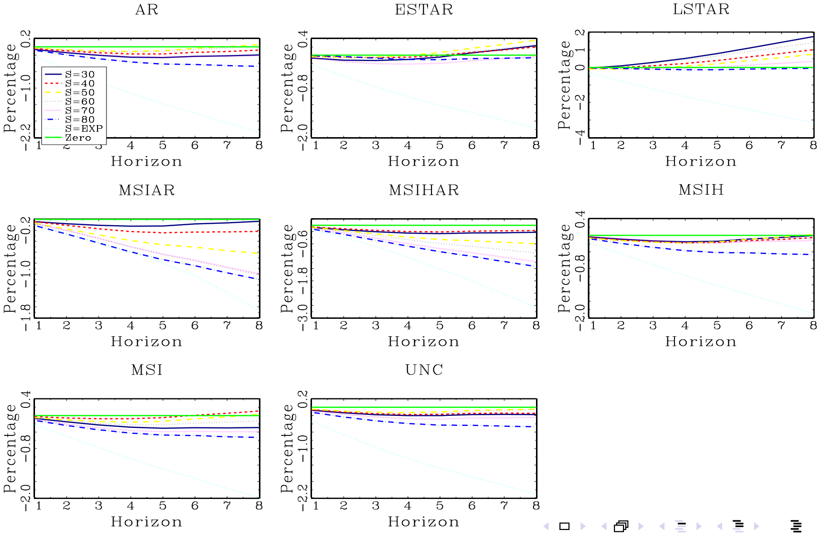
Evaluating Density Forecasts: Testing Strategies

- 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)

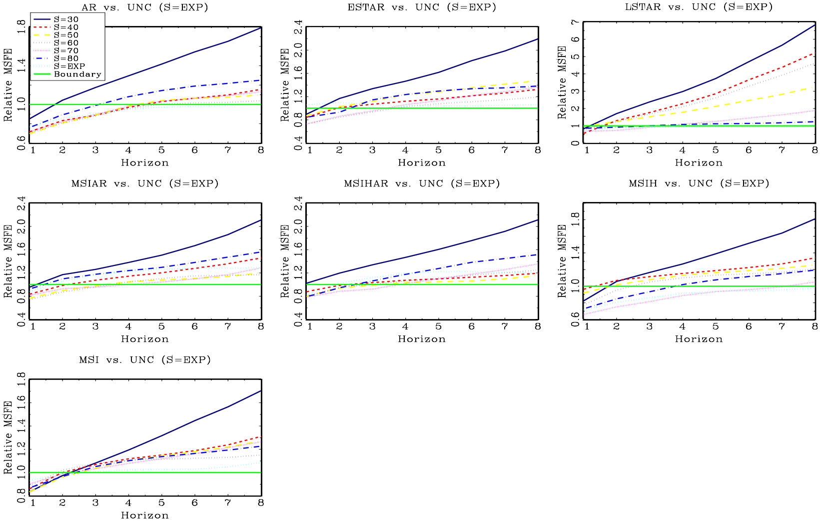

AR h=1
AR h=4
AR h=8
ESTAR h=1
ESTAR h=4
ESTAR h=8
LSTAR h=1
LSTAR h=4
LSTAR h=8
MSIAR h=1
MSIAR h=4
MSIAR h=8
S=30
S=40
S=50
S=60
S=70
S=80
S=EXP
Percentage
Time
1992
1996
2000
2004
2008


MSIHAR h=1
S=30
S=40
S=50
S=60
S=70
S=80
S=EXI
Percentage
Time
MSIHAR h=4
Percentage
Time
MSIHAR h=8
Percentage
Time
MSI h=1
Percentage
Time
MSI h=4
Percentage
Time
MSI h=8
Percentage
Time
MSIH h=1
Percentage
Time
MSIH h=4
Percentage
Time
MSIH h=8
Percentage
Time

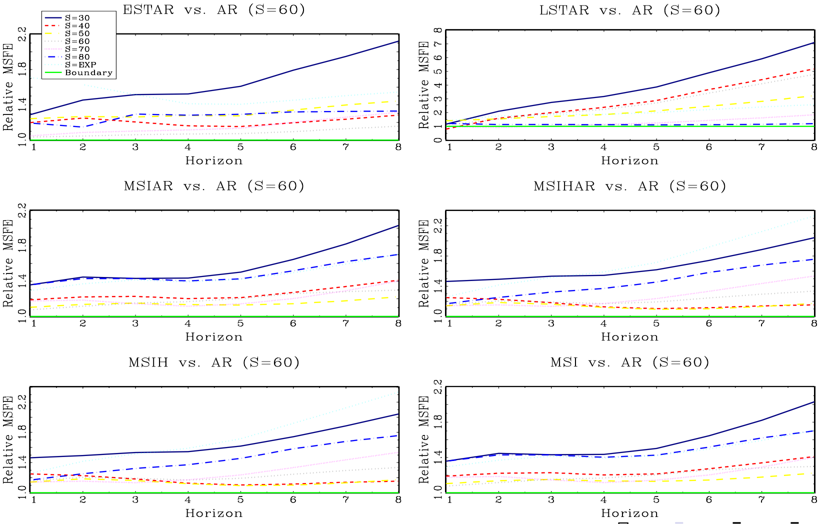
Using More Time-Information Leads to Biased Predictions



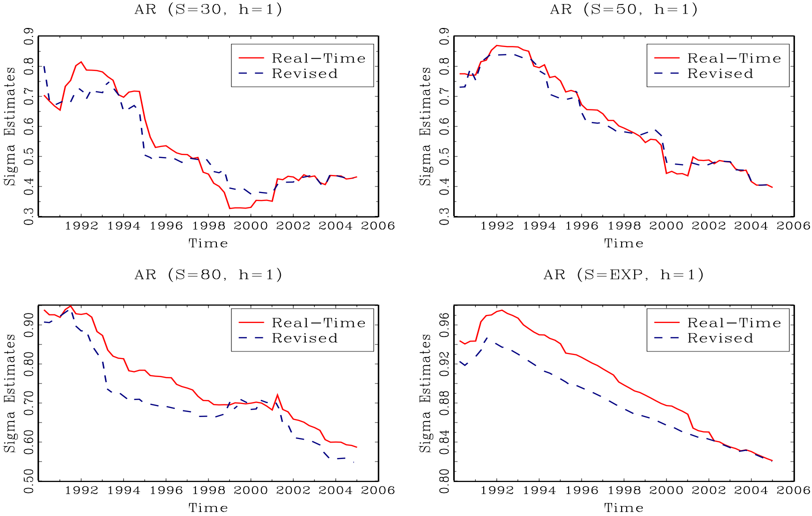
Variance Ratios: Limited Information at Long Horizons



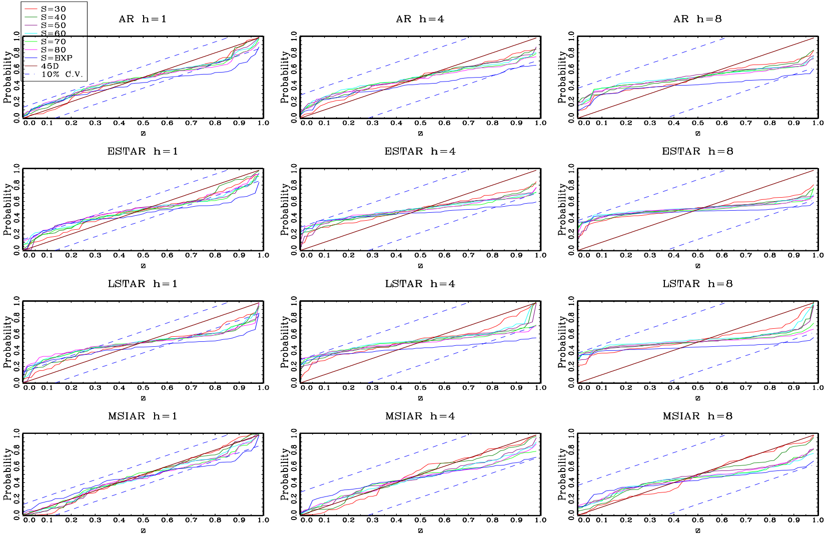
Relative MSEs: Bias Makes a Big Difference



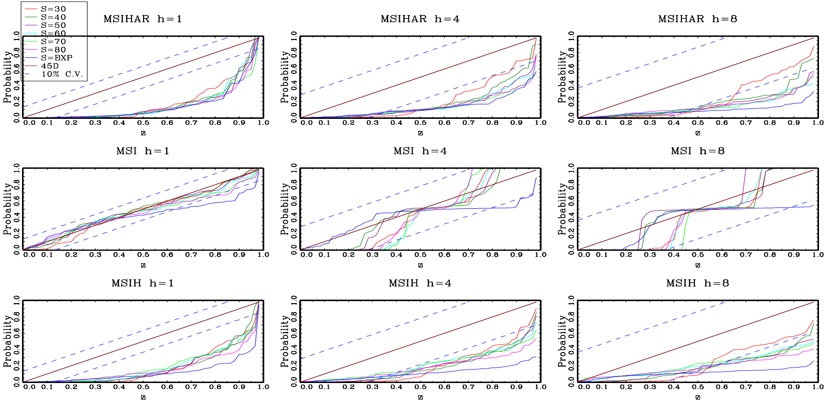
Real-time vs. Revised Estimates of Ω_1 for AR



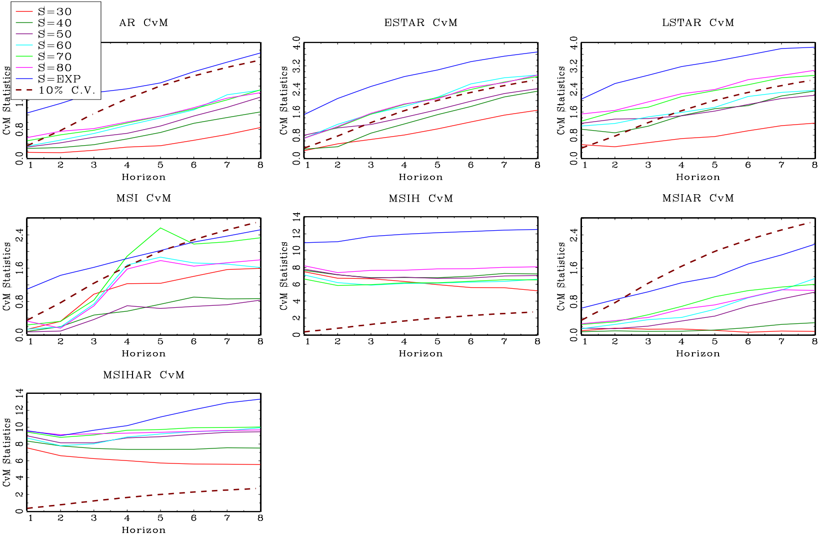
Empirical Cumulative Density Function of z_{t+h}



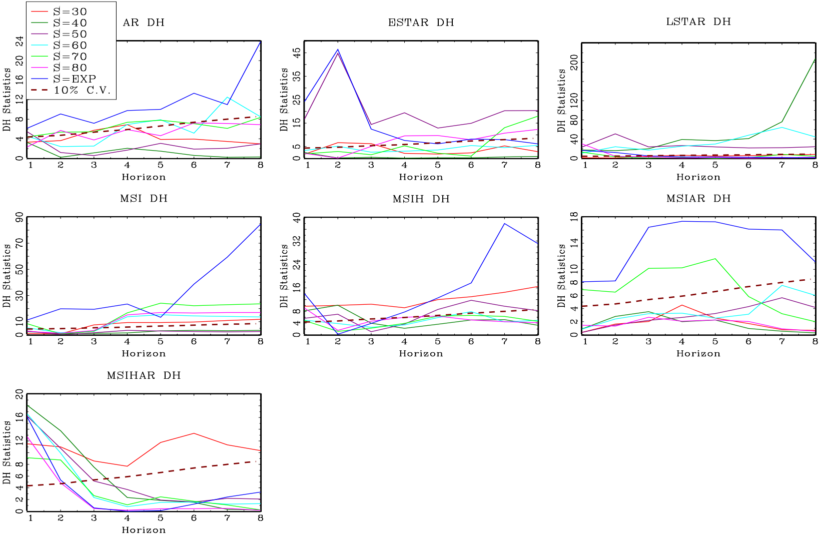
Empirical Cumulative Density Function of z_{t+h}



Cramer-von-Mises Test Results



Doornick-Hansen Normality Test Results



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

With Revised Data?

- 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!