Forecasting Canadian GDP: Evaluating Point and Density Forecasts in Real-Time

Frédérick Demers*

Research Department
Bank of Canada

October 2007
(*Preliminary and Incomplete - Do not Quote*)

*Presented at the Bank of Canada workshop *Forecasting Short-term Economic Development and the Role of Econometric Models*, October 25 and 26, 2007, Ottawa. Address: Research Department, Bank of Canada, 234 Wellington St., Ottawa, Ontario, K1A OG9, Canada (fdemers@bank-banque-canada.ca).

Contents

1 Introduction

To mitigate the effect of uncertainty about the future state of the economy, agents form expectations to predict what will happen. To a large extent, however, the future remains difficult to predict as the spectrum events which may materialize is large. As a way to address this issue, economic forecasts should take the form of probability distributions over a range of possible events in order to provide a description of the uncertainty faced by agents (Tay and Wallis, 2000). In contrast, point forecasts, the most common way by which forecasters report their predictions, are silent about the degree of uncertainty associated to events of interest despite being described as the most likely outcome.

This paper thus concerns the evaluation of point and density forecasts of real gross domestic product (GDP) growth for Canada. The ability of linear and nonlinear univariate models is compared for horizons up to two years. The paper also examines the impact of using all data available by considering a limited-memory estimator approach as a potential way to robustify linear models against structural changes (see, e.g., Giacomini and White, 2006; Clark and McCracken, 2004). More specifically, this paper concerns the usefulness of nonlinear functional forms to predict the distribution of GDP growth. The analysis documented in this paper is performed using vintages of data on real GDP so as to better reflect a real-time forecasting environment.

The rest of the paper is organized as follows. Section 2 discusses the data and presents the forecasting models. Section 3 presents the design of the forecast experiment. Section 4 presents and discusses the empirical results of the forecast evaluation. Section 5 briefly concludes.

2 The Forecasting Experiment

2.1 Data and notation

Data used for this analysis are the seasonally adjusted, real GDP at market prices and span from 1961Q1 until 2006Q4. The forecast evaluation is performed using vintages from 1990Q1. Let Y_t denote the logarithm of GDP multiplied by 100, and the quarterly growth as $y_t = Y_t - Y_{t-1}$, with $t = 1, ..., T$. Figures 1 compares initial and final quarter-over-quarter growth rates of GDP. In this paper, we are not interested in predicting the sequence of quarterly growth rates. We focus instead on predicting the more important forecast of the h-quarter growth in output, namely $y_{t+h}^h = \Sigma_{i=1}^h y_{t+i} = Y_{t+h} - Y_t$. By considering this quantity rather than the sequence of growth rates, we can more directly analyze the uncertainty around GDP forecasts at various horizons.

In practice, forecasters are confronted with the problem of predicting a variable which is measured with error and subject to quarterly revisions. Hence, even the information set at time t when the prediction is made is uncertain.[1] As a result, the resulting forecast error is composed of two parts: one due to a genuine forecast error, and the other due to revisions. Whether revisions can be predicted is outside the scope of this paper, but Jacobs and van Norden (2007) and Cunningham *et al* (2007) provide interesting insights on how this can be done.

2.2 Forecasting models

2.2.1 Unconditional forecasts

The usefulness of a forecasting model is generally analyzed by determining whether a particular model, or information set, is more informative than knowing nothing about the data except its unconditional behaviour, or the mean. To illustrate the potential variability in the process generating output growth, Table 1 reports summary statistics for different subperiods and compares the results using revised and real-time data over the recent period. The data generating process (DGP) of output growth appears unstable over time as suggested by the substantial heterogeneity across sub-period estimates of the mean, variance, skewness, and kurtosis. The effect of using revised data is mainly felt through the mean and variance of the process, with minor discrepancies on the skewness and kurtosis estimates.

[1] The behaviour of revisions to Canadian GDP growth has been recently examined by Demers (2007).

Table 1

Summary Statistics of Output Growth (annual rate)

Sample / Statistics	mean		variance		skewness		kurtosis	
1961Q1-2006Q4	real-time	revised	real-time	revised	real-time	revised	real-time	revised
	–	5.175	–	12.830	–	0.321	–	2.307
1970Q1-1979Q4								
	–	3.983	–	13.486	–	0.542	–	2.688
1980Q1-1989Q4								
	–	2.591	–	16.478	–	-0.184	–	1.859
1990Q1-1999Q4								
	2.275	2.576	5.966	8.185	-1.007	-0.865	3.873	3.694
2000Q1-2006Q4								
	2.641	2.694	2.395	2.614	-0.199	-0.549	2.816	3.070
1990Q1-2006Q4								
	2.426	2.625	4.471	5.821	-1.030	-0.935	4.575	4.513

2.2.2 AR(p)

The first model considered is the AR(p) model:

$$y_t = \alpha + \phi(L)y_{t-1} + \varepsilon_t, \tag{1}$$

where $\phi(L)$ is the operator in the lag polynomial, with, for instance, $\phi(L) = \phi_1 L - ... - \phi_p L^p$; α is the constant; and $\varepsilon_t \sim i.i.d.N(0, \sigma^2)$. This model is labeled as AR.

2.2.3 Markov switching AR(p)

Hamilton (1989) proposed a model of real US GNP which accommodates the expansion-recession features of the business cycle. We use a slight variant of Hamilton's mean-switching model proposed by Hansen (1992). Rather than considering a mean-switching process, we instead use an intercept-switching model, namely:

$$y_t = \alpha_{s_t} + \sum_{j=1}^{p} \phi_j y_{t-j} + \varepsilon_t, \tag{2}$$

where $\varepsilon_t \sim i.i.d.N(0, \sigma^2)$. This specification is labeled as MSI.

We also consider the case where the autoregressive parameter, ϕ, switches. This specification is labeled as MSIAR. Finally, Eq. (2) is generalized such that the innovations is a also a mixture process, or state-dependent, namely $\varepsilon_t \sim i.i.d.N(0, \sigma^2_{s_t})$. This specification is labeled as MSIH or MSIHAR, depending on whether the intercept switches or not, such that Eq. (2) can be fully generalized as

$$y_t = \alpha_{s_t} + \sum_{j=1}^{p} \phi_{s_t,j} y_{t-j} + \sigma_{s_t} \varepsilon_t, \tag{3}$$

with $\varepsilon_t \sim i.i.d.N(0, 1)$.

The switching mechanism is unobserved and the variable, s_t, is a stochastic process governed by a discrete time, ergodic, first order autoregressive M-state Markov chain with transition probabilities, or mixing weights, $\Pr[s_t = j | s_{t-1} = i] = p_{i,j}$ for $\forall i, j \in \{1, ..., M\}$, and transition matrix, $\mathbf{P}$:

$$\mathbf{P} = \begin{pmatrix} p_{1,1} & \cdots & p_{M,1} \\ \vdots & \ddots & \vdots \\ p_{1,M} & \cdots & p_{M,M} \end{pmatrix}.$$

Denoting ξ_t as the filtered probability of being in either regime at time t, the markovian properties allow easy derivation of values for ξ_{t+h} using $\mathbf{P}$, so that the regimes can be weighted appropriately (see Hamilton, 1994, for further technical details). The vector of population parameter, $\theta_t = \{\alpha_1, ..., \alpha_M, \phi_{1,1}, ..., \phi_{p,M}, \sigma_1, ..., \sigma_M, p_{i,j}\}'$, can thus be estimated using (constrained) maximum likelihood techniques, and is indexed by the subscript t to reflect that they are obtained using information up to time t and may change over time.

2.2.4 Smooth transition AR(p)

Another family of non-linear models proposed by Teräsvirta (1994) are the exponential smooth transition and logistic smooth transition AR(p) models, denoted respectively as ESTAR and LSTAR. The ESTAR model has the following form:

$$\begin{aligned} y_t &= \alpha_1 + \phi_1(L) y_{t-1} + w_t (\alpha_2 + \phi_2(L) y_{t-1}) + \varepsilon_t \\ w_t &= 1 - \exp(-\gamma (y_{t-d} - \mu)^2); \end{aligned} \tag{4}$$

whereas the LSTAR model is written as follows:

$$\begin{aligned} y_t &= \alpha_1 + \phi_1(L) y_{t-1} + w_t (\alpha_2 + \phi_2(L) y_{t-1}) + \varepsilon_t \\ w_t &= \frac{1}{1 + \exp(-\gamma (y_{t-d} - \mu))}, \end{aligned} \tag{5}$$

where $\varepsilon_t \sim i.i.d.N(0, \sigma^2)$, d is a delay parameter, and $p \geq d \geq 1$. The parameter $\gamma > 0$ determines the shape of the transition function, w_t. The vector of population parameter to be estimated is thus, $\theta_t = \{\alpha_1, \alpha_2, \phi_{1,1}, ..., \phi_{1,p}, \phi_{2,1}, ..., \phi_{2,p}, \gamma, \mu\}'$, and the estimate for σ is obtained by concentrating the likelihood function, $\mathcal{L}$.

These various specifications have one important thing in common in that they assume conditional normality—i.e., ε_t is Gaussian. Meanwhile, mixture models approximate general density functions and translate into processes with varying degrees of excess skewness and kurtosis (Timmermann, 2000). By comparing a range of linear and nonlinear specifications to perform point and density forecasts, we will thus determine if departure from linearity is important when making such predictions.

3 Design of the Out-of-Sample Forecasting Exercise

3.1 Set-up and benchmark

To perform the pseudo out-of-sample forecast evaluation, forecasts are computed using two strategies, namely the *expanding* and *rolling* approach. For the expanding scheme, the models are recursively estimated using information until time t. On the other hand, the rolling approach consists of rolling a fixed sample forward at each iteration. The rolling-window approach is also synonymous to using the so-called "limited-memory" estimator (Giacomini and White, 2006). When the DGP is homogeneous over time, the expanding window should work well, all else equal, since the econometrician is using all the available information. In contrast, if the DGP is unstable, using *old* information will bias the estimates, whereas the rolling window, which discards old, uninformative data, will reduce (or eliminate) the bias. This, however, comes at the cost of a decrease in precision—the so-called bias-variance trade-off. Under this scheme, data that are no longer seen as informative are excluded and parameters are estimated using the most recent information.

To reflect the fact that the generating process is possibly unstable, a set of possible benchmark forecasts is considered in the following way. Let $\bar{y}^h_{t+h,S}$ denote the unconditional mean of the change in output over h periods calculated using various amount of time information, S, or the length of the rolling window used for the estimation, where $S = 30, 40, 50, 60, 70, 80$, or expanding ($exp$), such that $\bar{y}^h_{t+h,S}$ is approximated using

$$\bar{y}^h_{t+h,S} = \frac{1}{S} \sum_{j=t-S+1}^{t} (Y_j - Y_{j-h}) \,. \tag{6}$$

The quantity $\bar{y}^h_{t+h,S}$ is thus the average cumulative change in output over h period for the S most recent data, at time t. The scheme of the expanding-window amounts to using all available time-information when making the forecast at time t, thus starting with $j = 1$. From the set of benchmark forecasts, the one yielding the most accurate predictions (i.e., smallest loss) will thus be the forecast to beat. This benchmark model is labeled as UNC.

For the AR model, the optimal lag, p, is chosen using Akaike's information criterion (AIC). Because of significant numerical cumbersomeness—the problem of identification—when estimating nonlinear models, p (and d for the smooth transition models) is fixed to 1 in the forecasting exercise. M is fixed to 2, again for the purpose of the forecasting exercise—estimating 3-regime models with fewer than 100 observations could be difficult. Finally, constrained maximum likelihood estimates are used to ensure that $|\hat{\phi}| < 1$ and $\hat{\sigma} \geq 0$, with '$\hat{}$' denoting an estimate. When some elements of θ_t lie on their boundary, maximization of $\mathcal{L}$ was attempted using different starting values. If after ten attempts the estimation failed to provide a satisfactory estimate of θ_t, the forecast is carried out using $\hat{\theta}_{t-1}$ instead. This was necessary in only a few occasions. An "insanity filter" was not applied although some nonlinear forecasts were a bit odd (Swanson and White, 1995).

Forecasts are evaluated for horizon $h = 1, 2, 4, 8$, and $h_{\max} = 8$. The out-of-sample forecast experiment runs from 1990Q1 until 2004Q4.

Because this paper strictly is concerned with the evaluation of point and density forecasting performance, specification tests for linearity or nonlinearities are not implemented.

3.2 Point forecasts

Obtaining point forecasts from AR(p) and Markov-switching models is trivial. For the AR model, the point forecast is derived by simply iterating forward Eq. (1). For the Markov-switching models, the point forecast is obtained in a similar way for each regime, but each regime is weighted appropriately—i.e., using ξ_{t+h}.

In contrast, obtaining point forecasts with smooth-transition models is not possible using closed-form expressions when $h > 1$, and thus requires numerical integration, or simulation methods (Teräsvirta *et al*, 2005). To emphasize on the incidence of using a particular functional (i.e., linear vs. nonlinear) form when forecasting, multi-step forecasts are obtained by drawing vectors of εs of length $h_{\max}$ from the Gaussian distribution.[2] One thousand replications are used and averaged out to obtain $\hat{y}^h_{t+h}$.

[2]A natural alternative would be to bootstrap the residuals, but this would take us away from looking at the incidence of the normality assumption, which is a key objective here.

3.3 Density forecasts

Multi-step density forecasts are obtained in a similar fashion as the point forecasts, but another quantity of interest is necessary however when $h > 1$. While variance estimates of the one-step-ahead forecast-error can be estimated directly from the models, or by concentration of the likelihood function, estimates of the multi-step-ahead forecast-error variance, denoted as $\hat{\Omega}_{t+h}$, must be derived. Let $\hat{\sigma}^2_{t+h}$ denote the h-step-ahead Newey-West estimate of the forecast error variance for the process $\{Y_{t+h} - Y_{t+h-1}\}$, $\hat{\sigma}^2_{t+h}$ is obtained using the following expression for an AR(1), ignoring parameter uncertainty (Clements and Hendry, 1998, ch. 4):

$$\hat{\sigma}^2_{t+h} = \hat{\sigma}^2 \frac{\left(1+\phi^{2h}\right)}{\left(1-\phi^2\right)}.$$

As h increases, $\hat{\sigma}^2_{t+h}$ approaches the unconditional variance of y_t.

Then the variance of the forecast error for the partial-sum process y_{t+h} can be obtained using

$$\hat{\Omega}_{t+h} = h\hat{\sigma}^2_{t+h}\left(1+\frac{h}{T}\right).$$

Denoting $f_t(y_{t+h}|I_t)$ as the h-step-ahead density forecast, or conditional density distribution, of the true density $p_t(y_{t+h}|I_t)$ at time t, the sequence $\hat{\Omega}_{t+h}$ is used to construct sequences of vectors of $f_t(y_{t+h}|I_t)$, where I_t is the information set known at time t.

It is important to note that the MS models with constant variance generate Gaussian density forecast at each period t, although the density is time varying due to the mixture of distributions embedded in the model. In contrast, when the distribution of ε_t depends upon the state, the resulting density at time t is non-Gaussian.

For the smooth-transition models, forecast densities are obtained by taking the average of the simulated densities at each period.

4 Forecast Evaluation

4.1 Point forecasts

Denoting v_{t+h} as the out-of-sample forecast error at time $t+h$, it is important to emphasis on the difference between the residuals of the models, ε_t, and the resulting ($h =$) 1-step out-of-sample forecast error. Although v_{t+1} and ε_{t+1} may coincide, they will have different properties in general. For instance, the addition of a constant term to a linear regression ensures that $E\left(\varepsilon_t\right) = 0$, however nothing prevents v_{t+1} to have a non-zero mean as forecasts can be biased in an unstable environment.

Since accuracy of the point forecasts is examined using the mean squared forecast error (MSFE) criterion, and because the MSFE combines the (squared) forecast-error bias and the forecast-error variance, these two quantities of interest are examined separately.

4.1.1 Results

To illustrate the impact of using different amounts of time-information (S) when forecasting GDP, figures 2 and 3 plot the forecast errors at different horizons for all the time-series forecasting models with the S varying. The interesting thing to note from these figures is that the choice of S matters more for long-horizon forecasts than for short horizons. In other words, the choice of S has a smaller impact on predicting the short-term dynamics of output growth than for predicting its trend (i.e., the unconditional mean of GDP growth), which is possibly time-varying.

More concretely, Figure 4 plots the MSFE of each model relative to the unconditional (UNC) forecast's MSFE based on a rolling window of 60 observations, which is found to provide the lowest MSFE for nearly all horizons. When this ratio is less than one, a model is said to improve the accuracy relative to the benchmark forecasting device. Whereas when the ratio is above one, the model is declared non-informative relative to the benchmark. While nearly all model specifications improve the forecast accuracy relative the unconditional forecast at horizons of one or two quarters, most models fail to further improve the forecast accuracy at the one-year horizon, at which point an uninformed, unconditional forecast proves superior to the models. Another interesting feature that comes out of Figure 4 is that the MSFEs of each specification tend to be clustered for very short term forecasts, although by varying S the MSFEs can be inflated. For instance, the MSFE is inflated by over 25 per cent when $h = 1$ in the case of the AR forecasts, depending on the choice of S. On the other hand, the models behave very differently for longer horizons: the MSFE can easily be twofold by varying S within a model specification. This result clearly shows how using different forecasting strategies can affect accuracy, given a particular model specification.

Rather than comparing the forecast accuracy with the best unconditional forecasts, Figure 5 compares the MSFEs relative to the best AR forecasts, which is obtained from a rolling window of 60 observations. We see that the point-forecast accuracy of the AR model is relatively good when compared to nonlinear alternatives. The reason why this is the case is because nonlinear models tend to produce forecasts that are slightly more biased than those obtained from the AR model. Figure 6 plots the bias of the different model. As expected, the models which are estimated using large amount of time-information tend to be more biased than those relying on smaller Ss, a problem which becomes more severe as h increases. For longer horizons, biases can cause serious problems when predicting trend growth.

Turning to the forecast-error variance by factoring out the bias, Figure 7 plots the ratios of variances of the nonlinear models relative to the UNC forecasts based on $S = exp$. While most models are informative in the short-run, few models remain informative about GDP growth at the one-year horizon, a result consistent with the findings of Galbraith and Tkacz (2007). In contrast, Figure 8 plots the ratios of variances of the nonlinear models relative to the AR forecasts with $S = 60$. In this case, a few nonlinear forecasts (i.e., mainly from the MSIH and LSTAR specifications) outperform the best linear forecasts.

Because nonlinear models produce biased forecasts, they tend to generate higher MSFEs than the linear AR model. When the biased is taken into account, however, we then notice that they generate smaller forecast error variance. This result could be due to the absence of nonlinearities over the forecast period.

4.2 Density forecasts

In contrast to the point-forecasts evaluation or the examination of the conditional mean, the evaluation of density forecasts is a generalized approach to discriminate between forecasting models in their ability to characterize the unconditional density of a stochastic process of interest. In this subsection we turn to the analysis of density forecasts.

The approach discussed in Dawid (1984) and Diebold, Gunther, and Tay (1998) is followed. They suggest using probability integral transform to evaluate model-based economic forecasts. Let $\{y_t\}_{t=1}^{T}$ denote the sequence of realization of the process (i.e., output growth); with $\{p_t(y_t|I_t)\}_{t=1}^{T}$ denoting the sequence of conditional densities governing y_t with $I_t = \{y_{t-1}, y_{t-2}, ...\}$; and $\{f_t(y_t|I_t)\}_{t=1}^{T}$ is the corresponding sequence of density forecast. Note that the subscript t signifies that the density function can be time varying.

Because $p_t(y_t|I_t)$ is unobserved, even *ex post*, we cannot directly compare the predicted densities with the data generating process densities. The strategy proposed by Diebold *et al* thus consists of evaluating the forecast densities through the probability integral transform (pit). The sequence of pits, denoted as $\{z_t\}_{t=1}^{T}$, is the cumulative density function which corresponds to $p_t(y_t|I_t)$ evaluated at y_t, or:

$$z_t = \int_{-\infty}^{y_t} p_t(u)\mathrm{d}u. \tag{7}$$

If the predicted densities correspond to the true densities, then $\{z_t\}_{t=1}^{T} \sim i.i.d.U[0,1]$ without the need to impose any distributional assumptions about the underlying process being predicted.

To analyze z_t, the hypothesis of uniformity can be verified by plotting the empirical distribution of z_t against the 45° line. Formal tests of $i.i.d.U[0,1]$ can also be applied under certain conditions, for instance the Kolmogorov-Smirnov or Cramer-von Mises goodness-of-fit test statistics. Because goodness-of-fit tests lack power, Berkowitz (2001) suggests to examine z_t by taking the inverse normal CDF transformation instead, z_t^*, such that $z_t^* \sim i.i.d.N$. The normality test due to Doornick and Hansen (1994) is used.

When $h = 1$, the $i.i.d.$ assumption about z_t or z_t^* is a natural hypothesis to examine. When $h > 1$, however, z_t and z_t^* will exhibit dependence which will distort the uniformity of normality tests that rely upon the absence of independence.

To perform the inference and account for the effects of the dependence due to the overlapping of the $h(> 1)$-step-ahead forecasts, the finite-sample distribution of the Kolmogorov-Smirnov (KS), Cramer-von Mises (CvM), and Doornick-Hansen (DH) tests is simulated. For each test, 10 000 replications are used to generate the distribution under the null hypothesis of interest. Normal. For simplicity, the effect of parameter estimation is ignored, although this may be an issue for inference.

4.2.1 Results

Figures 9 and 10 plot the empirical cumulative density functions (CDF) of the z_{t+h} values for the different forecasting models against the theoretical 45° line. The CDFs are plotted for various S at horizon 1, 4 and 8. The 10 per cent critical values of the Kolmogorov-Smirnov are plotted alongside: when the empirical CDF lies outside the confidence interval, the null hypothesis that $z_{t+h} \sim i.i.d.U$ is rejected, implying that the predicted densities do not match the ones which generated the data. It is important to note that z_{t+h} could depart from the 45° line simply because the forecasts are biased. Departure from the 45° line could be caused by higher order moments as well. From figures 9 and 10, we see that a number of models fail to match the density. Even at the one-quarter horizon, the choice of S has a large influence on the ability of a model to match the density of GDP growth. The MSIHAR and the MSIH seem to produce the worse density forecasts at all horizons. On the other hand, the MSI model performs reasonably well at the one-quarter ahead only. The MSIAR model with rolling-windows, in contrast, can generate good density forecasts at all horizons with a number of CDFs lying close to the ideal 45° line. From a distance point of view, the CDF from best MSIAR lies closer to 45° line than the one from the best AR.

The results for the Cramer-von-Mises test are summarized in Figure 11. The CvM test statistic at each forecast horizon, of each models with varying S, are plotted alongside the simulated 10 per cent critical values. Overall, these results do not contradict the conclusions drawn earlier from the KS test results: for many functional form, it is possible to find a S which generates forecast densities that match the data. For all functional forms, using old time information deteriorates the ability of a model to match the density of the GDP growth data. In general, either the KS or CvM tests find little rejection of the null hypothesis, whether this is because of the low power of these tests is open question.

An alternative, more powerful testing strategy is to test whether the transformed process z^*_{t+h} is normally distributed. Figure 12 summarizes the Doornick-Hansen normality test results based on the simulated critical values for a confidence level of 10 per cent. If the test statistics, DH, lies below the 10 per cent critical value, we conclude that the null hypothesis that the z^*s are normally distributed cannot be rejected. These results generally comfort the earlier results based on the goodness-of-fit statistics.

5 Conclusion [*to be completed*]

References

Berkowitz, J. 2001. "Testing Density Forecasts with Applications to Risk Management." *Journal of Business and Economic Statistics*

Clark, T.E. and M.W. McCracken. 2004. "Improving Forecast Accuracy by Combining Recursive and Rolling Forecasts." Kansas City Federal Reserve, Working Paper No. 04–10.

Clements, M.P. and D.F. Hendry. 1998. *Forecasting Economic Time Series.* Cambridge, Mass.: Cambridge University Press.

Clements, M.P. and J. Smith. 2000. "Evaluating the Forecast Densities of Linear and Non-linear Models: Applications to Output Growth and Unemployment." *Journal of Forecasting* 19: 255–76.

Clements, M.P., and H.-M. Krolzig. 2003. "Business Cycle Asymmetries: Characterization and Testing Based on Markov-Switching Autoregressions." *Journal of Business and Economic Statistics* 21: 196–211.

Cunningham, A., J. Eklund, J. C. Jeffrey, G. Kapetanios, and V. Labhard. 2007. "A State Space Approach to Extracting the Signal from Uncertain Data." Bank of England, forthcoming Working Paper.

Dawid, A.P. 1984. "Statistical Theory: The Prequential Approach." *Journal of the Royal Statistical Society* 13: 124–41.

Demers, F. 2007. "National Income Revisions: Are they Important and can they be Predicted?" Bank of Canada, forthcoming Discussion Paper.

Diebold, F.X., T. Gunther, and A. Tay. 1998. "Evaluating Density Forecasts with Applications to Financial Risk and Management." *International Economic Review* 39: 863–83.

Doornick, J. A. and H. Hansen. 1994. "An Omnibus Test for Univariate and Multivariate Normality." Nuffield College, Oxford. Working paper.

Galbraith, J. W. and G. Tkacz .2007. "How Far Can Forecasting Models Forecast? Forecast Content Horizons for Some Important Macroeconomic Variables." Bank of Canada Working Paper No. 2007–1.

Giacomini, R. and H. White. 2006. "Tests of Conditional Predictive Ability." *Econometrica* 75: 1545–78.

Hamilton, J.D. 1989. "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle." *Econometrica* 57: 357–384.

__________. 1994. *Time Series Analysis.* New Jersey: Princeton University Press.

Hansen, B.E. 1992. "The Likelihood Ratio Test Under Nonstandard Conditions: Testing the Markov Switching Model of GNP." *Journal of Applied Econometrics* 7: 561–82.

Jacobs J. and S. van Norden. 2007. "Modeling Data Revisions: Measurement Error and Dynamics of "True" Values." HEC mimeo.

Timmermann, A. 2000. "Moments of Markov Switching Models." *Journal of Econometrics* 96: 75–111.

Tay A. S. and K. F. Wallis. 2000. "Density Forecasting: A Survey." *Journal of Forecasting* 19: 235–54.

Swanson, N. R. and H. White. 1995. "A Model-Selection Approach to Assessing the Information in Term Structure using Linear Models versus Artificial Neural Networks." *Journal of Business and Economic Statistics* 13: 265–75.

Teräsvirta, T. 1994. "Specification, Estimation, and Evaluation of Smooth Transition Autoregressive Models." *Journal of American Statistical Association* 89: 208–18.

Teräsvirta, T., D. van Dijk, and M. C. Medeiros. 2005. "Linear Models, Smooth Transition Autoregressions, and Neural Network for Forecasting Macroeconomic Time Series: A re-examinination." *International Journal of Forecasting* 21: 755–74.

Figure 1: **Data**

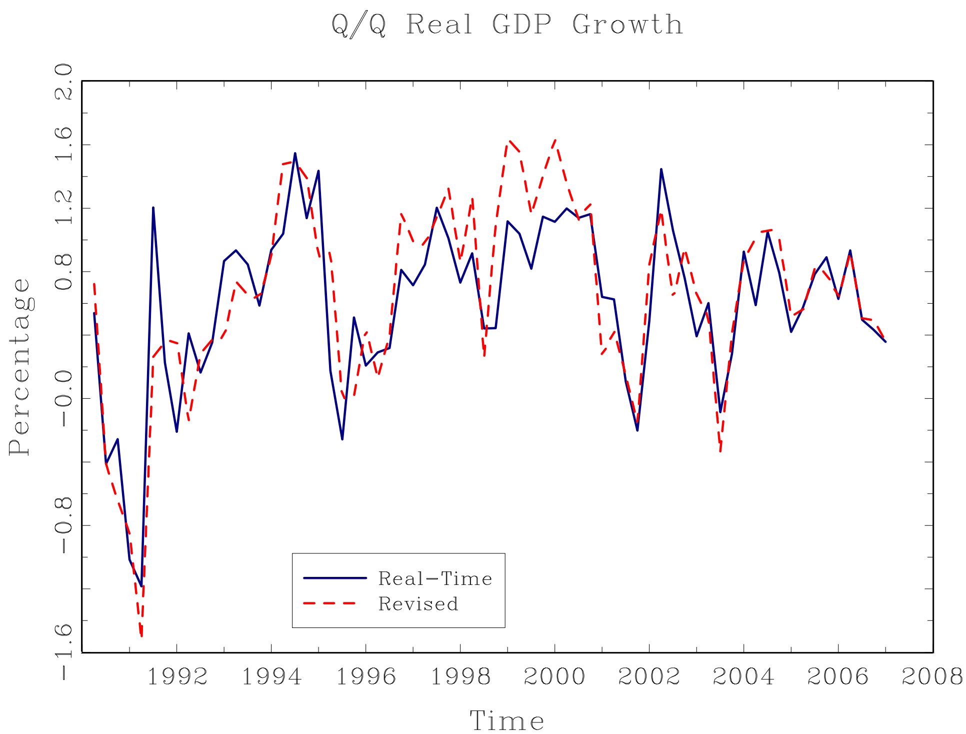

Figure 2: **Forecast Error with Different Amount of Time Information in Real-Time**

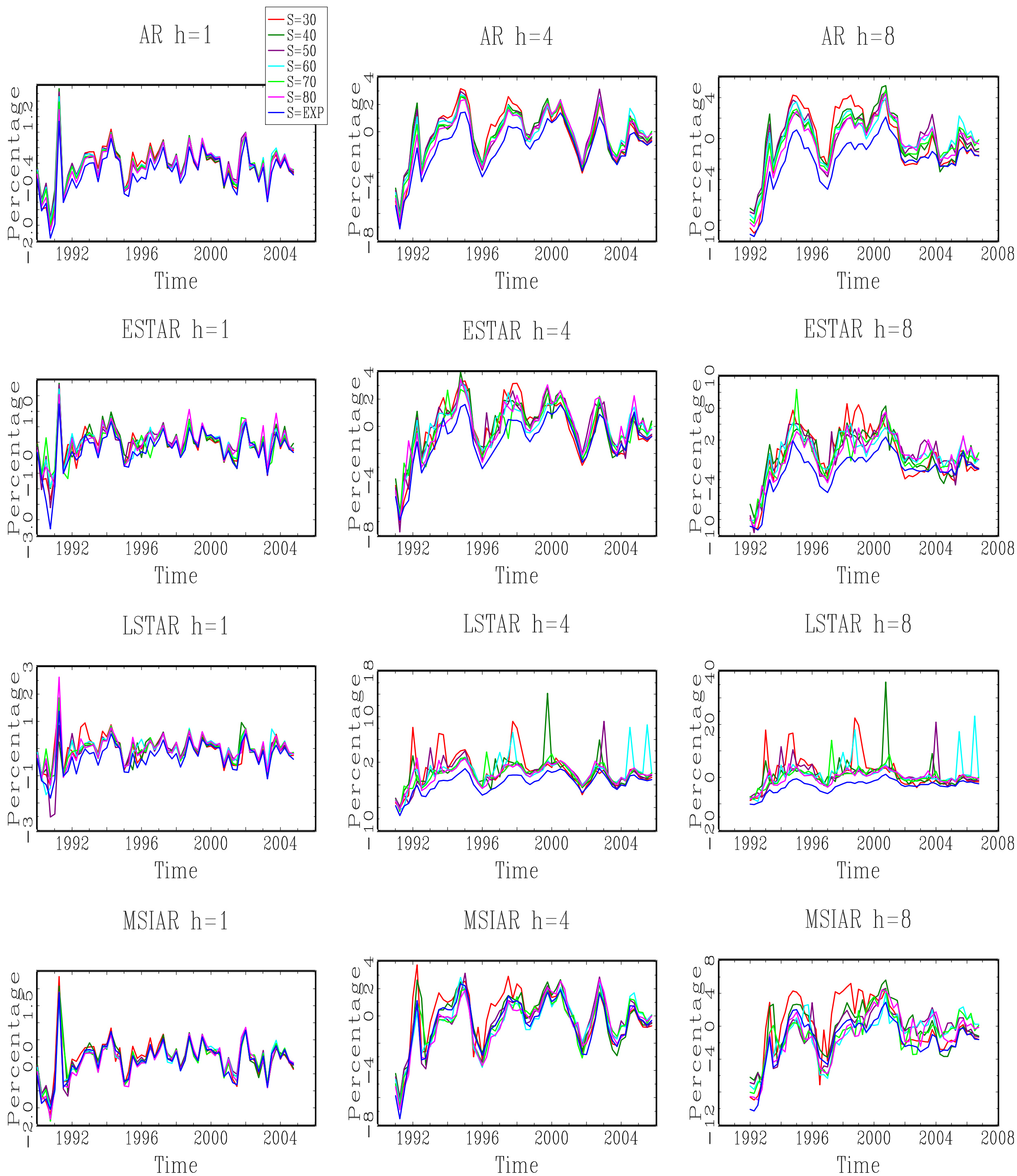

Figure 3: **Forecast Error with Different Amount of Time Information in Real-Time**

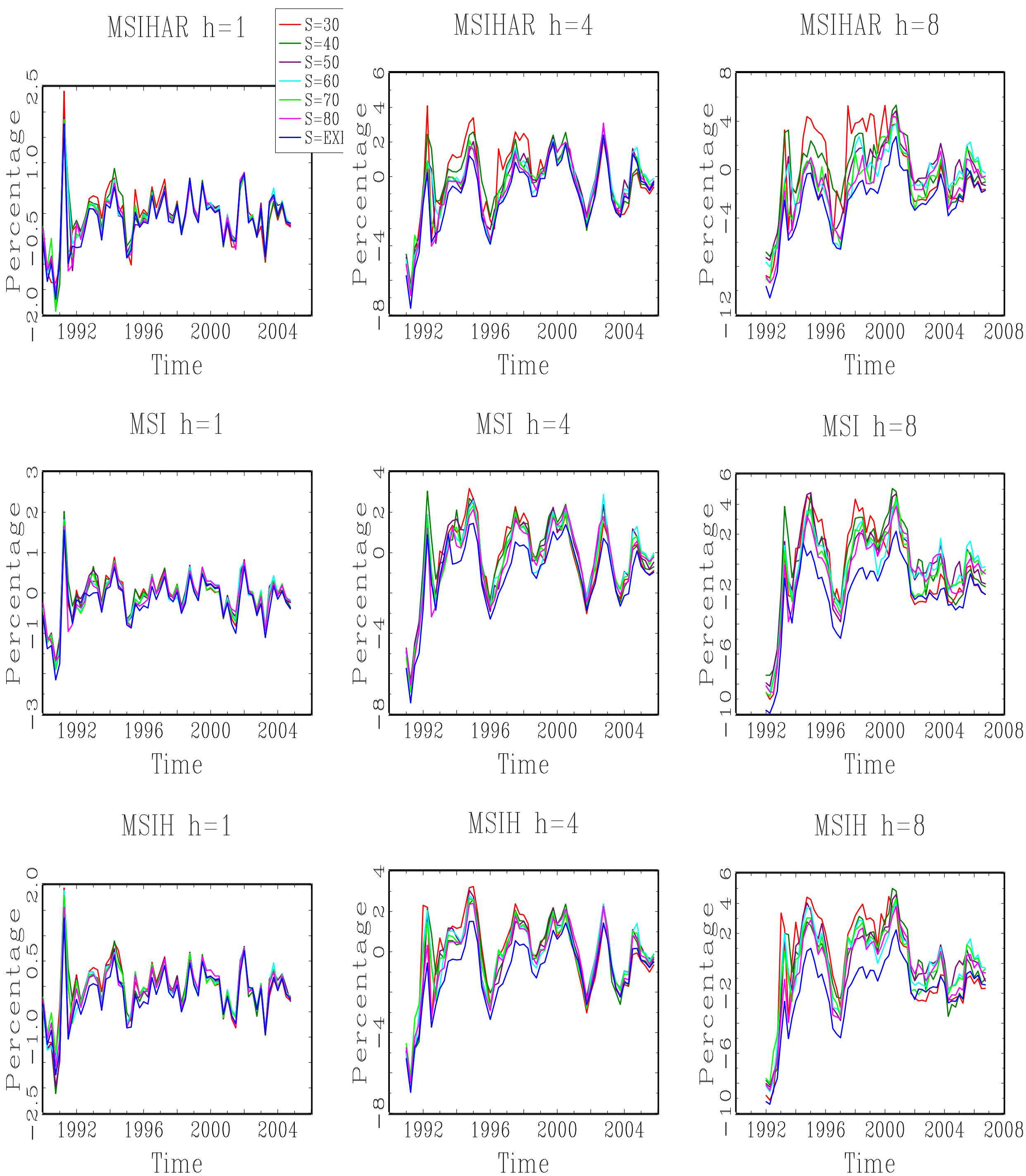

Figure 4: **Relative MSFEs**

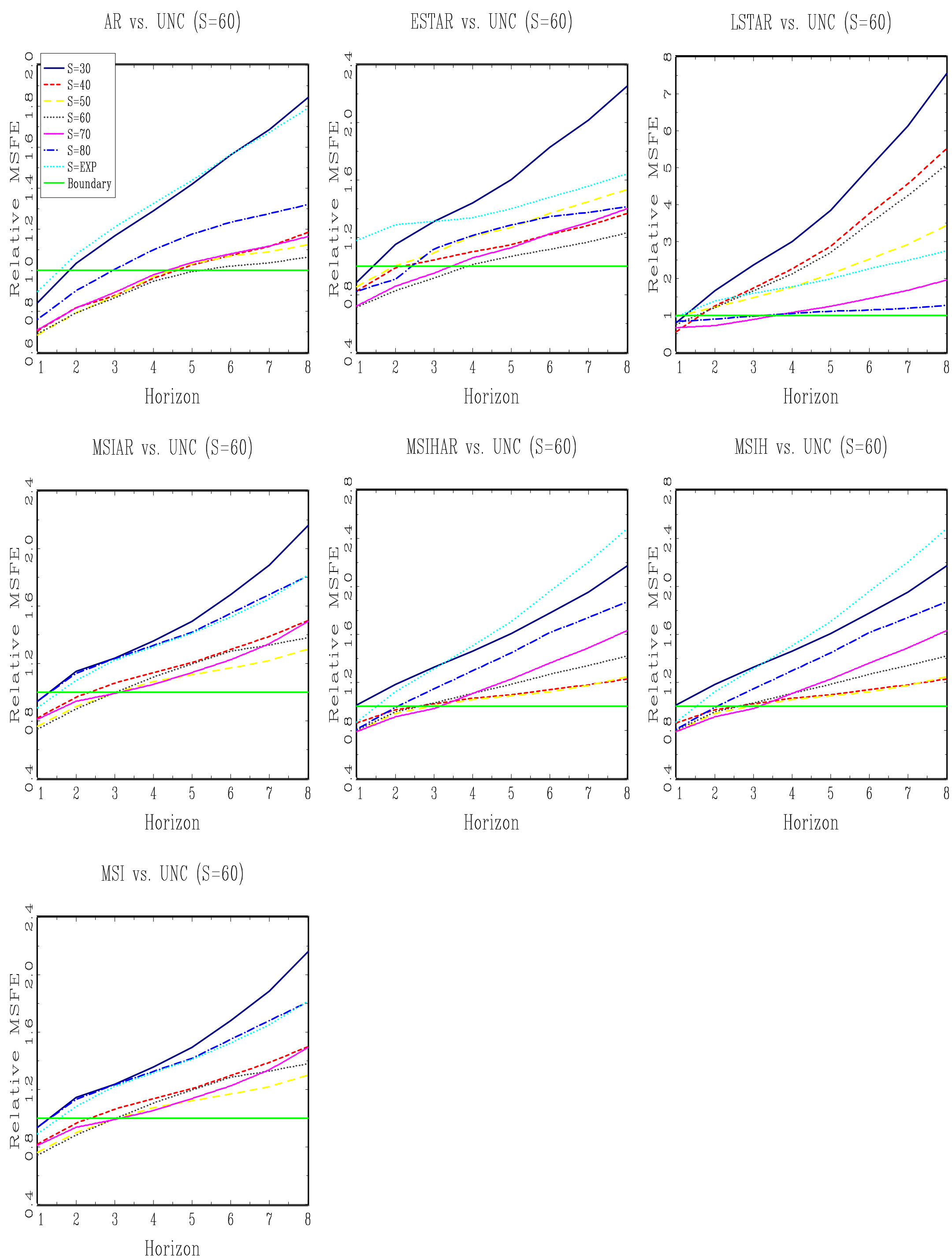

Figure 5: **Relative MSFEs**

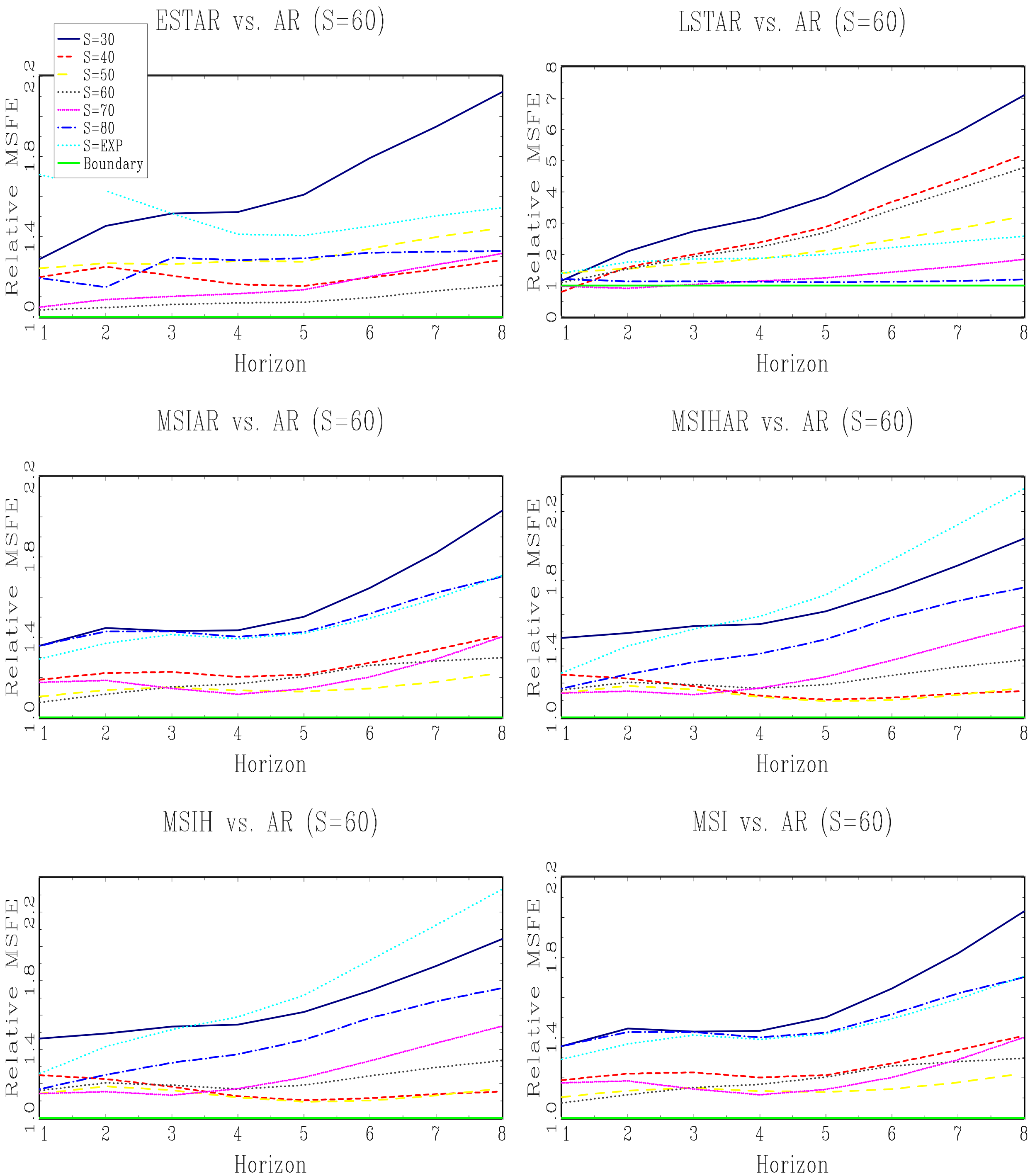

Figure 6: **Forecast Bias**

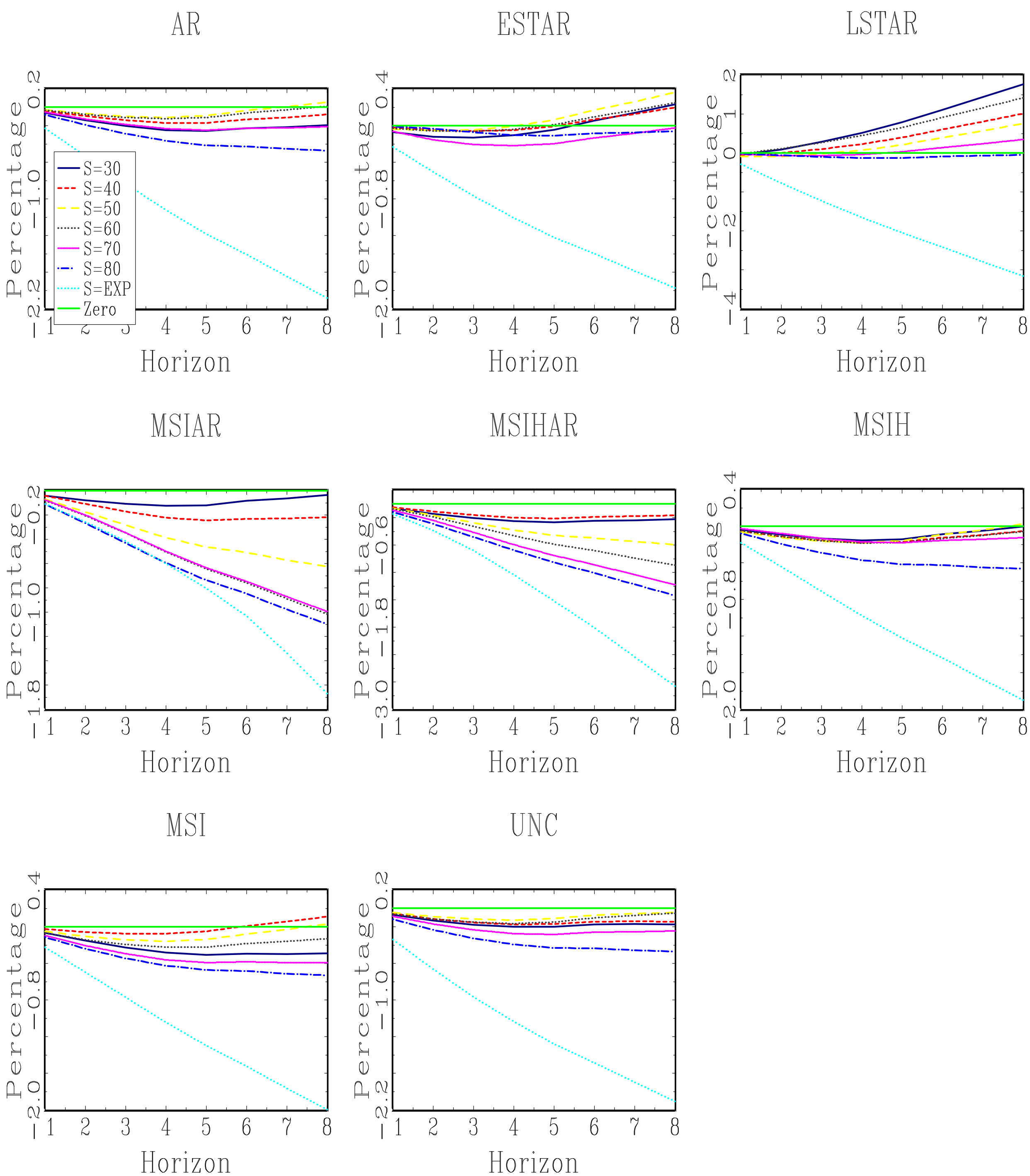

Figure 7: **Variance Ratios**

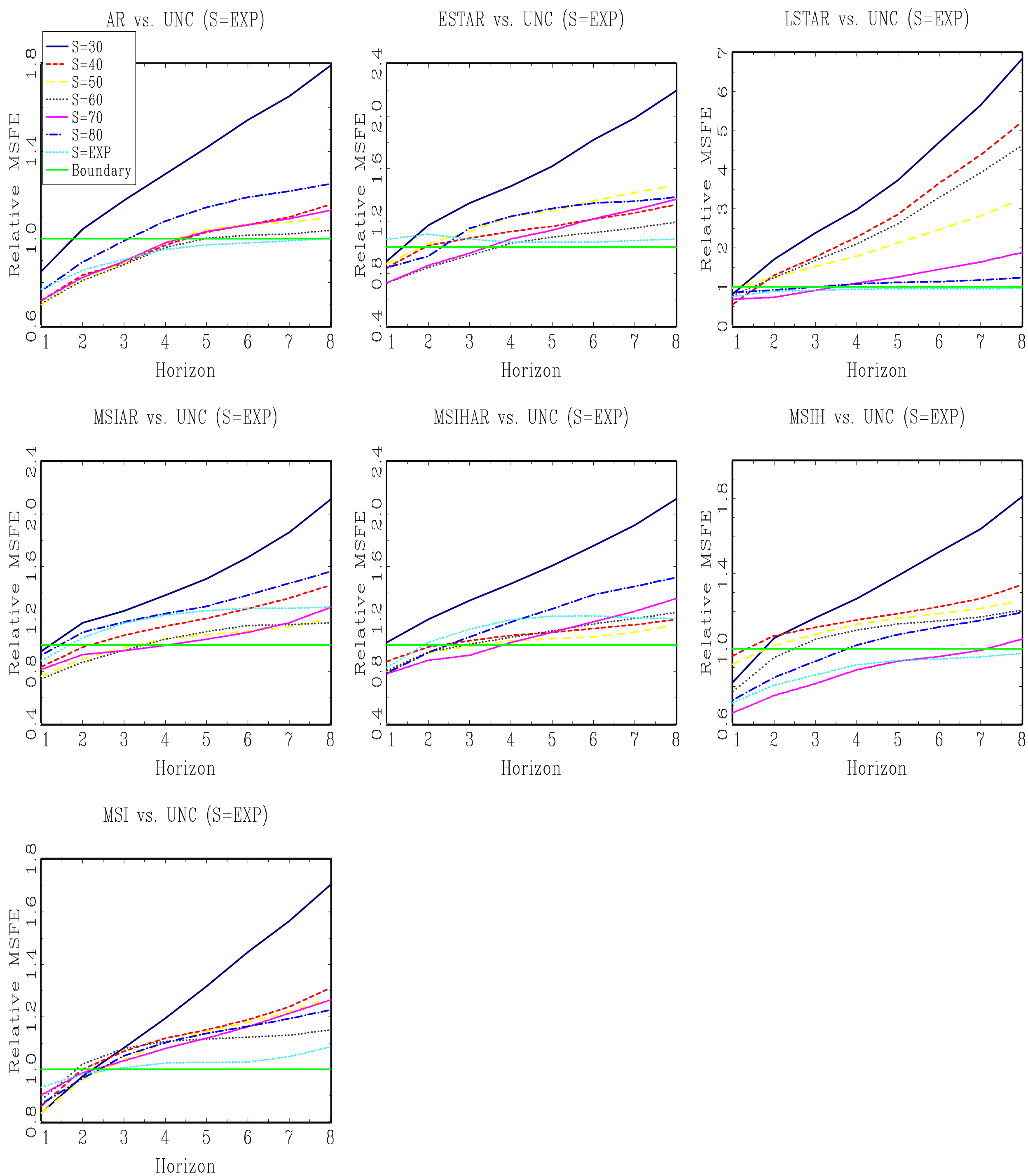

Figure 8: **Variance Ratios**

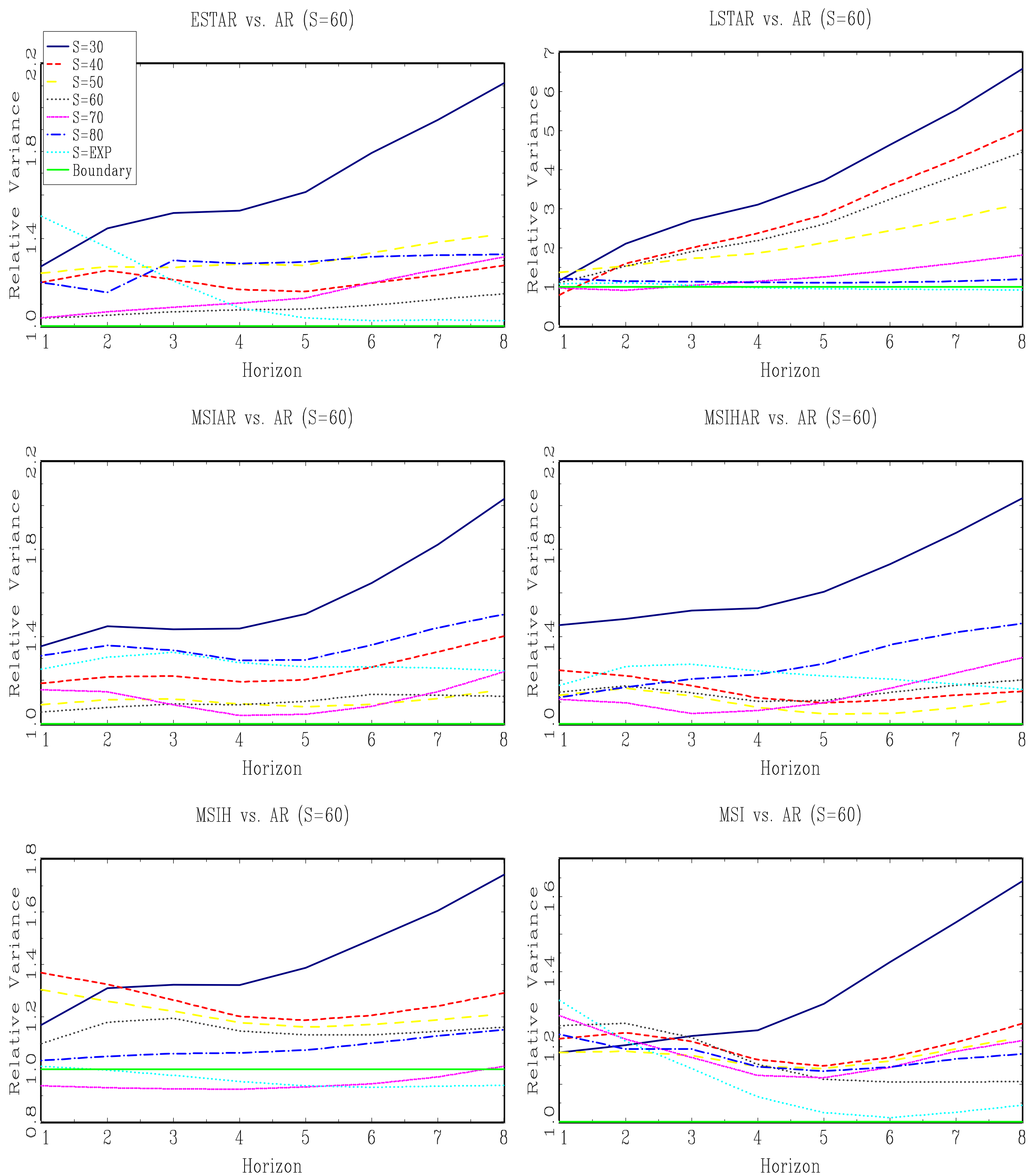

Figure 9: **CDFs of z_{t+h}-values From Forecasting Models**

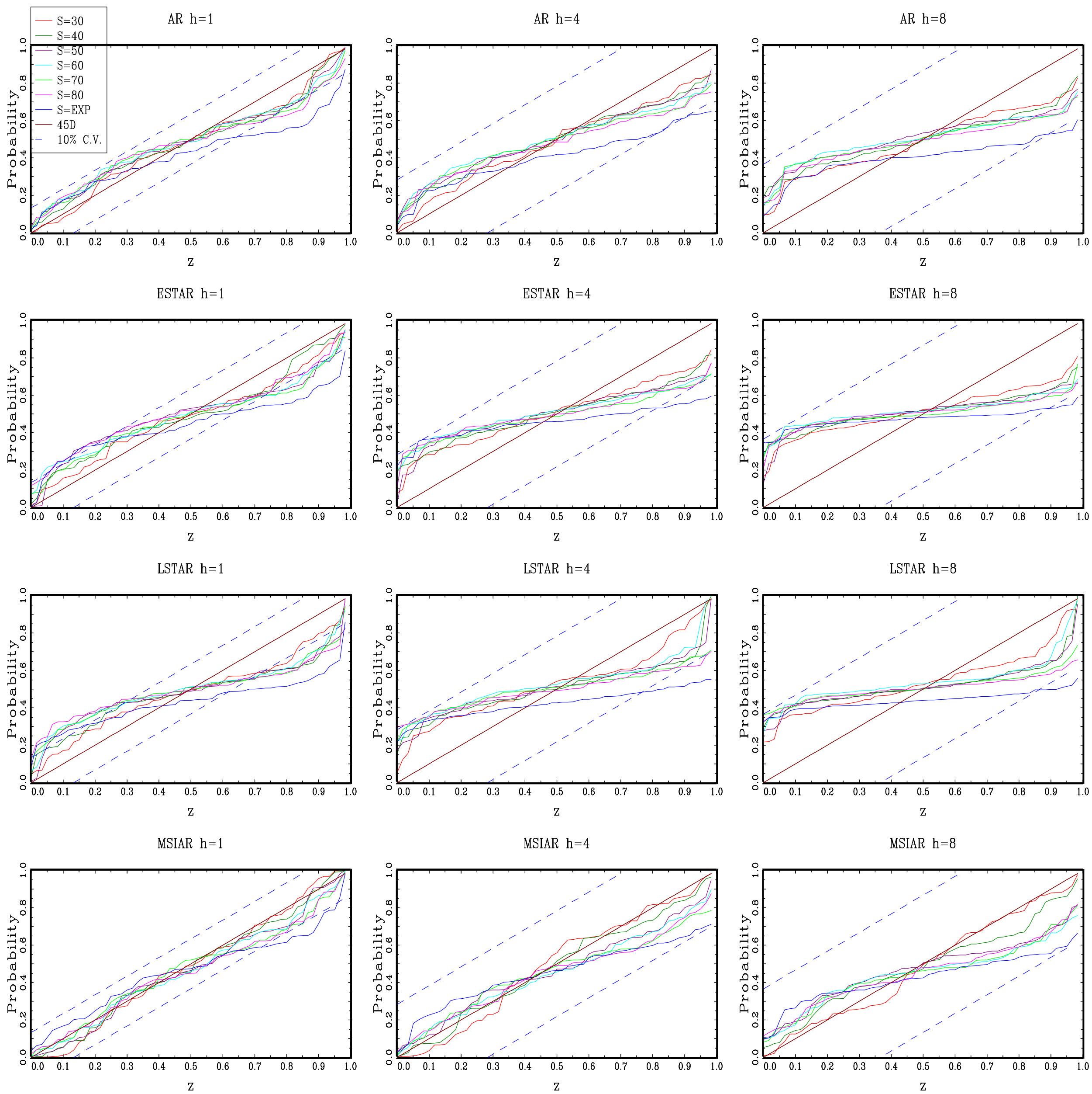

Figure 10: **CDFs of z_{t+h}-values From Forecasting Models**

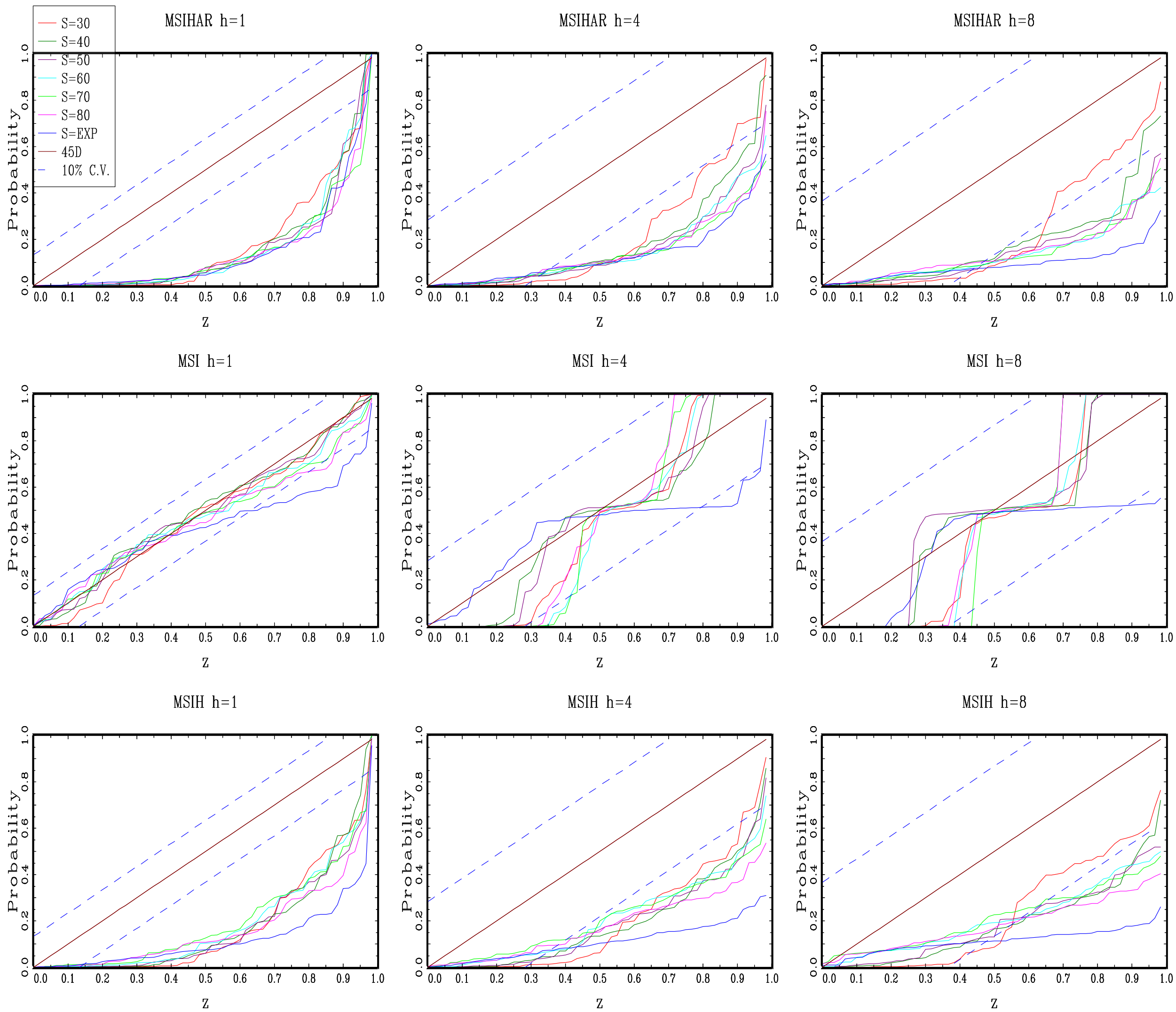

Figure 11: **Cramer-von-Mises (CvM) Test Results**

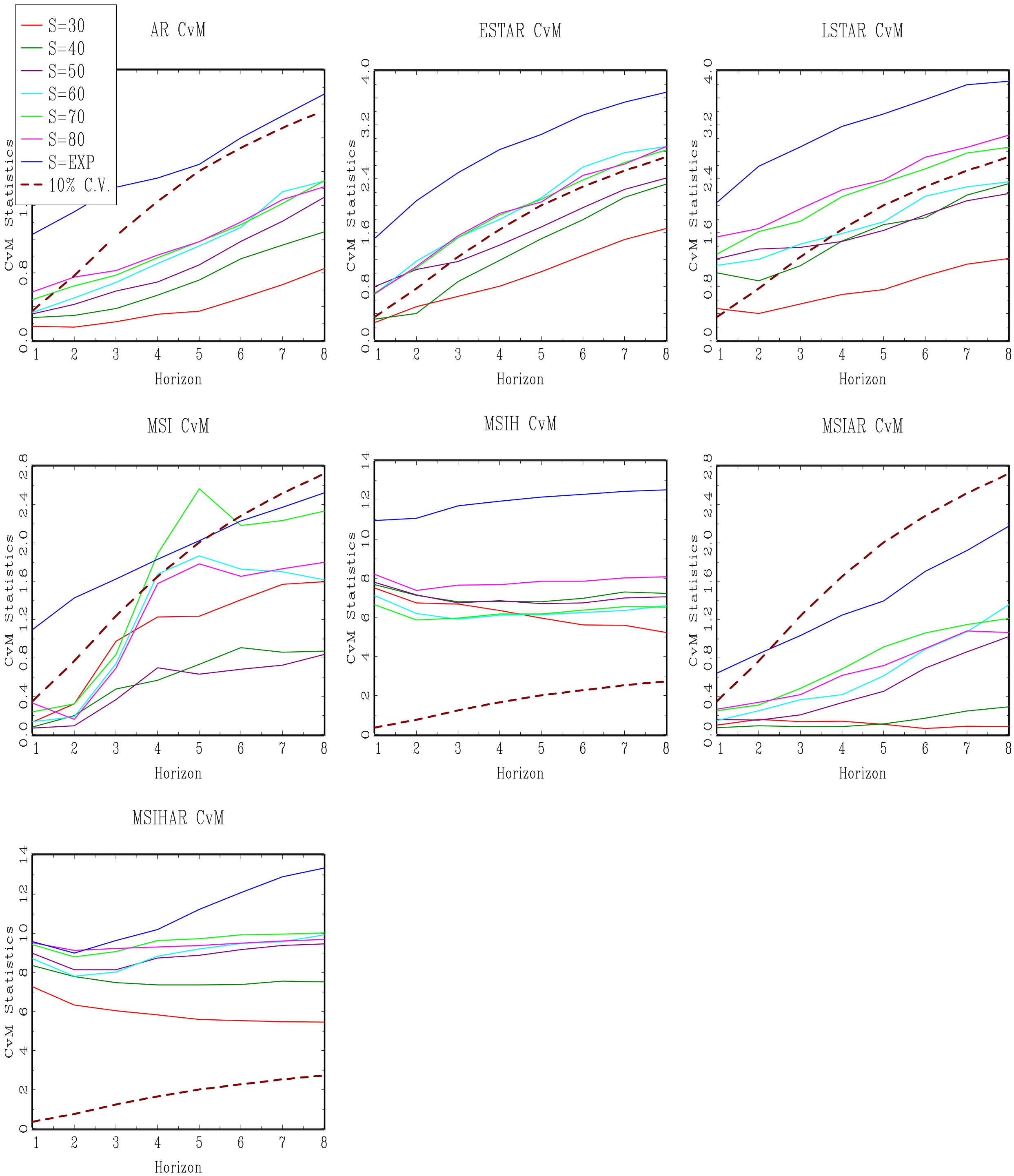

Figure 12: **Doornick-Hansen (DH) Test Results**