Bias Correction and Out-of-Sample Forecast Accuracy

Kim, Hyeongwoo and Durmaz, Nazif

Auburn University

May 2009

Bias Correction and Out-of-Sample Forecast Accuracy[*]

Hyeongwoo Kim[†] and Nazif Durmaz[‡]

Auburn University

May 2009

Abstract

The least squares (LS) estimator suffers from significant downward bias in autoregressive models that include an intercept. By construction, the LS estimator yields the best in-sample fit among a class of linear estimators notwithstanding its bias. Then, why do we need to correct for the bias? To answer this question, we evaluate the usefulness of the two popular bias correction methods, proposed by Hansen (1999) and So and Shin (1999), by comparing their out-of-sample forecast performances with that of the LS estimator. We find that bias-corrected estimators overall outperform the LS estimator. Especially, Hansen's grid bootstrap estimator combined with a rolling window method performs the best.

Keywords: Small-Sample Bias, Grid Bootstrap, Recursive Mean Adjustment, Out-of-Sample Forecast, Diebold-Mariano Test

JEL Classification: C53

1   Introduction

It is a well-known statistical fact that the least squares (LS) estimator for autoregressive (AR) processes suffers from serious downward bias in the persistence coefficient when the stochastic process includes a non-zero intercept and/or deterministic time trend. The bias can be substantial especially when the stochastic process is highly persistent (Andrews, 1993).

Since the seminal work of Kendall (1954), an array of bias-correction methods has been put forward. To name a few, Andrews (1993) proposed a method to obtain the exactly median-unbiased estimator for an AR(1) process with Gaussian errors. Andrews and Chen (1994) extends the work of Andrews (1993) to get approximately median-unbiased estimator for higher order AR(p) processes. Hansen (1999) developed a nonparametric bias correction method, the grid bootstrap (GT), which is robust to distributional assumptions. The GT method has been actively employed by many researchers, among others, Kim and Ogaki (2009), Steinsson (2008), Karanasos *et al.* (2006), and Murray and Papell (2002).

An alternative approach has been also proposed by So and Shin (1999) who develop the recursive mean adjustment (RMA) estimator that belongs to a class of (approximately) mean-unbiased estimators. The RMA estimator is computationally convenient to implement yet powerful and used in the work of Choi *et al.* (2008), Sul *et al.* (2005), Taylor (2002), and Cook (2002), for instance.

By construction, the LS estimator provides the best in-sample fit among the class of linear estimators notwithstanding its bias.[1] A natural question then arises: Why do we need to correct for the bias? We attempt to find an answer by comparing the out-of-sample forecast performances of the bias-correction methods with that of the LS estimator. We apply the GT and the RMA approaches along with the LS estimator for quarterly commodity price indices for the period of 1974.QI to 2008.QIII, obtained from the Commodity Research Bureau (CRB). We find that both bias correction methods overall outperform the LS estimator.

---

[1] Recall that the LS estimator is obtained by minimizing the sum of squared residuals.

Especially, Hansen's GT estimator combined with a rolling window method performed the best.

Organization of the paper is as follows. In Section 2, we explain the source of bias and how each method corrects for biases. We also briefly explain how we evaluate the relative forecast performances. Section 3 reports our major empirical findings and Section 4 concludes.

2 Bias-Correction Methods

We start with a brief explanation of the source of the bias in the LS estimator for an autoregressive process. Consider the following AR(1) process.

$$y_t = c + \rho y_{t-1} + \varepsilon_t, \tag{1}$$

where $|\rho| < 1$ and ε_t is a white noise process. Note that estimating ρ by the LS estimator is equivalent to estimating the following.

$$(y_t - \bar{y}) = \rho (y_{t-1} - \bar{y}) + \varepsilon_t, \tag{2}$$

where $\bar{y} = T^{-1} \sum_{j=1}^{T} y_j$. The LS estimator for ρ is unbiased only when $E\left[\varepsilon_t | (y_{t-1} - \bar{y})\right] = 0$. This exogeneity assumption, however, is clearly violated because ε_t is correlated with y_j, for $j = t, t+1, \cdots, T$, thus with $\bar{y}$. Therefore, the LS estimator for AR processes with an intercept creates the mean-bias. The bias has an analytical representation, and as Kendall (1954) shows, the LS estimator $\hat{\rho}_{LS}$ is biased downward.

There is no analytical representation of the median-bias. Monte Carlo simulations, however, can easily demonstrate that the LS estimator produces significant median-bias for ρ when ρ gets close to unity (see Hansen, 1999).

3

When ε_t is serially correlated, it is convenient to express (1) as follows.

$$y_t = c + \rho y_{t-1} + \sum_{j=1}^{k} \beta_j \Delta y_{t-j} + u_t, \tag{3}$$

where u_t is a white noise process that generates ε_t.[2]

For Hansen's (1999) GT method, we define the following grid-t statistic.

$$t_N(\rho_i) = \frac{\hat{\rho}_{LS} - \rho_i}{se(\hat{\rho}_{LS})},$$

where $\hat{\rho}_{LS}$ is the LS point estimate for ρ, $se(\hat{\rho}_{LS})$ denotes the corresponding LS standard error, and ρ_i is one of M fine grid points in the neighborhood of $\hat{\rho}_{LS}$. Implementing LS estimations for B bootstrap samples at each of M grid points, we obtain the $\alpha\%$ quantile function estimates, $\hat{q}_{N,\alpha}^*(\rho_i) = \hat{q}_{N,\alpha}^*(\rho_i, \varphi(\rho_j))$, where φ denotes nuisance parameters such as βs that are functions of ρ_i. After smoothing quantile function estimates, the (approximately) median-unbiased estimate is obtained by,

$$\hat{\rho}_G = \rho_i \in R, \text{ s.t. } t_N(\rho_i) = \tilde{q}_{N,50\%}^*(\rho_i),$$

where $\tilde{q}_{N,50\%}^*(\rho_i)$ is the smoothed 50% quantile function estimates obtained from $\hat{q}_{N,\alpha}^*$.[3] To correct for median-bias in β_j estimates, we treat other βs as well as ρ as nuisance parameters and follow the procedures described above.

So and Shin's (1999) RMA estimator utilizes demeaning variables using the partial mean instead of the global mean $\bar{y}$. Rather than implementing the LS for (2), the RMA estimator is obtained by the LS estimator for the following regression equation.

$$(y_t - \bar{y}_{t-1}) = \rho (y_{t-1} - \bar{y}_{t-1}) + \eta_t,$$

---

[2]When the stochastic process is of higher order than AR(1), exact bias-correction is not possible because the bias becomes random due to the existence of nuisance parameters. For higher order AR(p) models, the RMA and the GT methods yield approximately mean- and median-unbiased estimators, respectively.

[3]We used the Epanechinikov kernel $K(u) = 3(1 - u^2)/4I(|u| \leq 1)$, where $I(\cdot)$ is an indicator function.

where $\bar{y}_{t-1} = (t-1)^{-1} \sum_{j=1}^{t-1} y_j$ and $\eta_t = \varepsilon_t - (1-\rho)(t-1)^{-1} \sum_{j=1}^{t-1} y_j$. Note that the error term η_t is independent of $(y_{t-1} - \bar{y}_{t-1})$, which results in bias reduction for the RMA estimator $\hat{\rho}_R$. For a higher order AR process such as (3), the RMA estimator can be obtained by treating βs as nuisance parameters as in Hansen's (1999) GT method.

We use a conventional method proposed by Diebold and Mariano (1995) to evaluate the out-of-sample forecast accuracy of each bias-correction method relative to that of the LS estimator. Let $y_{t+h|t}^1$ and $y_{t+h|t}^2$ denote two competing (out-of-sample) h-step forecasts given information set at time t. The forecast errors from the two models are,

$$\varepsilon_{t+h|t}^1 = y_{t+h} - y_{t+h|t}^1, \ \ \varepsilon_{t+h|t}^2 = y_{t+h} - y_{t+h|t}^2$$

For the Diebold-Mariano test, define the following function.

$$d_t = L(\varepsilon_{t+h|t}^1) - L(\varepsilon_{t+h|t}^2),$$

where $L(\varepsilon_{t+h|t}^j)$, $j = 1, 2$ is a loss function.[4] To test the null of equal predictive accuracy, $H_0 : Ed_t = 0$, the Diebold-Mariano statistic (DM) is defined as,

$$\text{DM} = \frac{\bar{d}}{\sqrt{\widehat{Avar}(\bar{d})}}$$

where $\bar{d}$ is the sample mean loss differential,

$$\bar{d} = \frac{1}{T - T_0} \sum_{t=T_0+1}^{T} d_t,$$

$\widehat{Avar}(\bar{d})$ is the asymptotic variance of $\bar{d}$,

$$\widehat{Avar}(\bar{d}) = \frac{1}{T - T_0} \sum_{j=-q}^{q} k(j, q) \hat{\mathbf{\Gamma}}_j,$$

---

[4]One may use either the squared error loss function, $(\varepsilon_{t+h|t}^j)^2$, or the absolute error loss function, $|\varepsilon_{t+h|t}^j|$.

$k(\cdot)$ denotes a kernel function where $k(\cdot) = 0$, $j > q$, and $\hat{\Gamma}_j$ is j^{th} autocovariance function estimate.[5] Under the null, DM has the standard normal distribution asymptotically.

3   Empirical Results

We use quarterly commodity price indices, CRB Spot Index and its six sub-indices, obtained from the Commodity Research Bureau (CRB) for the period of 1974 to 2008.[6] We noticed a structural break of these series in 1973, the year of the demise of the Bretton Woods system (see Figure 1). Since our main objective is to evaluate relative forecast performances of competing estimators, we use observations starting from 1974.Q1 instead of using a dummy variable for the Bretton Woods era.

Table 1 reports our estimates for the persistence parameter in (3). We find that both the RMA and the GT methods yield significant bias-corrections. For example, the ρ estimate for the Spot Index increases from 0.950 (LS) to 0.969 (RMA) and 0.975 (GT). This is far from being negligible because corresponding half-life estimates are 3.378, 5.503, and 6.844 years, respectively. Note also that median-unbiased estimates by the GT are not restricted to be less than one, because the GT is based on the local-to-unity framework and allows even mildly explosive processes.[7]

We evaluate the out-of-sample forecasting ability of the three estimators, the LS, the RMA, and the GT, with two alternative forecasting methods. First, we utilize first 69 out of 139 observations to obtain h-step ahead forecasts. Then, we keep forecasting recursively by adding one observation in each iteration until we forecast the last observation. Second, we obtain h-step ahead forecasts using first 69 observations, then keep forecasting with a

---

[5]Following Andrews and Monahan (1992), we use the quadratic spectral kernel with automatic bandwidth selection for our analysis.

[6]In order to reduce noise in the data, we converted monthly frequency raw data to quarterly data by taking end-of-period values. Alternatively, one may use quarterly averages. Averaging time series data, however, creates time aggregation bias as pointed by Taylor (2001).

[7]When the true data generating process is I(1), one may use AR models with differenced variables, then correct for biases. Median/Mean bias for such models, however, tends to be small, because differenced variables often exhibit much weaker persistence. Since we are interested in evaluating the usefulness of bias-corrected estimators, we do not consider such models.

rolling window by adding and dropping one observation in each iteration, maintaining 69 observations, until we reach the end of full sample. We report our results in Tables 2 and 3.

Overall, we find that both bias-correction methods outperform the LS estimator with an exception of the Textile Sub-Index. No matter what methods are employed, the ratios of root mean squared prediction errors (RMSPE), LS/RMA and LS/GT, are mostly greater than one, which implies higher prediction precision of these methods relative to the LS estimator. For example, 4-period (1 year) ahead out-of-sample forecasts for the Spot index by the LS, RMA, and GT with the recursive method yield 0.104, 0.099, and 0.102 RMSPEs, respectively (see Table 2). Because the ratio LS/RMA (1.050) is greater than LS/GT (1.018) and both ratios are greater than 1, the RMA performs the best and the LS is the worst for this case. The corresponding Diebold-Mariano statistic shows that the RMA outperforms the LS at the 5% significance level. The evidence of superior performance of the GT is weaker than the RMA because corresponding p-value is 0.185, that is, significant only at the 20% significance level. When we use the rolling window method for 4-period ahead Spot Index forecasts, the grid bootstrap works the best and the LS performs the worst. The GT is superior to the LS at the 1% significance level, while the RMA outperforms the LS at the 5% level.

Another interesting finding is that a long memory is not necessarily good because forecast performance seems better with the rolling window method. It is easy to see the RMSPEs for each estimator are much smaller when we employ the rolling window strategy rather than the recursive method.[8] Especially, Hansen's GT estimator combined with the rolling window method performs the best because the associated RMSPEs are the smallest in majority cases.

4   Concluding Remarks

This paper evaluates relative forecast performances of two bias-correction methods, the RMA and the GT, to the LS estimator without bias-correction. When an intercept or an intercept

---

[8]We implemented same analysis for the sample period of 1974.Q1 to 2005.Q4 to see whether recent persistent movements of commodity indices significantly affected our results. We found very similar results.

and linear time trend are included in AR models, the LS estimator for the slope coefficient is downward-biased. Despite the bias, the LS estimator provides the best in-sample fit among a class of linear estimators. We attempt to find some justification of using these bias-correction methods by comparing the out-of-sample forecast accuracy of the methods with that of the LS estimator. Using the CRB Spot Index and its six sub-indices, we find that both methods overall outperform the LS estimator. Especially, Hansen's GT performs the best when it is combined with the rolling window strategy.

References

ANDREWS, D. W. K. (1993): "Exactly Median-Unbiased Estimation of First Order Autoregressive/Unit Root Models," *Econometrica*, 61, 139–165.

ANDREWS, D. W. K., AND H.-Y. CHEN (1994): "Approximately Median-Unbiased Estimation of Autoregressive Models," *Journal of Business and Economic Statistics*, 12, 187–204.

ANDREWS, D. W. K., AND J. C. MONAHAN (1992): "An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator," *Econometrica*, 60, 953–966.

CHOI, C.-Y., N. C. MARK, AND D. SUL (2008): "Bias Reduction in Dynamic Panel Data Models by Common Recursive Mean Adjustment," manuscript.

COOK, S. (2002): "Correcting Size Distortion of the Dickey-Fuller Test via Recursive Mean Adjustment," *Statistics and Probability Letters*, 60, 75–79.

DIEBOLD, F. X., AND R. S. MARIANO (1995): "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics*, 13, 253–263.

HANSEN, B. E. (1999): "The Grid Bootstrap and the Autoregressive Model," *Review of Economics and Statistics*, 81, 594–607.

KARANASOS, M., S. H. SEKIOUA, AND N. ZENG (2006): "On the Order of Integration of Monthly US Ex-ante and Ex-post Real Interest Rates: New Evidence from over a Century of Data," *Economics Letters*, 90, 163–169.

KENDALL, M. G. (1954): "Note on Bias in the Estimation of Autocorrelation," *Biometrika*, 41, 403–404.

KIM, H., AND M. OGAKI (2009): "Purchasing Power Parity and the Taylor Rule," Ohio State University Department of Economics Working Paper No. 09-03.

MURRAY, C. J., AND D. H. PAPELL (2002): "The Purchasing Power Parity Persistence Paradigm," *Journal of International Economics*, 56, 1–19.

NG, S., AND P. PERRON (2001): "Lag Length Selection and the Construction of Unit Root Tests with Good Size and Power," *Econometrica*, 69, 1519–1554.

SO, B. S., AND D. W. SHIN (1999): "Recursive Mean Adjustment in Time-Series Inferences," *Statistics and Probability Letters*, 43, 65–73.

STEINSSON, J. (2008): "The Dynamic Behavior of the Real Exchange Rate in Sticky-Price Models," *American Economic Review*, 98, 519–533.

SUL, D., P. C. B. PHILLIPS, AND C.-Y. CHOI (2005): "Prewhitening Bias in HAC Estimation," *Oxford Bulletin of Economics and Statistics*, 67, 517–546.

TAYLOR, A. M. (2001): "Potential Pitfalls for the Purchasing-Power-Parity Puzzle? Sampling and Specification Biases in Mean-Reversion Tests of the Law of One Price," *Econometrica*, 69, 473–498.

TAYLOR, R. (2002): "Regression-Based Unit Root Tests with Recursive Mean Adjustment for Seasonal and Nonseasonal Time Series," *Journal of Business and Economic Statistics*, 20, 269–281.

Table 1. Persistence Parameter Estimation Results

Index	ρ_L	CI	ρ_R	CI	ρ_G	CI
Spot	0.950	[0.856,0.972]	0.969	[0.872,0.985]	0.975	[0.910,1.022]
Livestock	0.933	[0.770,0.966]	0.972	[0.795,0.986]	0.990	[0.875,1.044]
Fats&Oil	0.933	[0.776,0.965]	0.951	[0.800,0.985]	0.997	[0.864,1.049]
Foodstuff	0.952	[0.813,0.976]	0.977	[0.836,0.993]	1.008	[0.890,1.049]
Raw Industrials	0.940	[0.847,0.966]	0.969	[0.863,0.979]	0.955	[0.907,1.009]
Textiles	0.917	[0.807,0.951]	0.947	[0.824,0.967]	0.932	[0.874,1.003]
Metals	0.963	[0.870,0.981]	0.974	[0.887,0.993]	0.996	[0.929,1.024]

Index	HL_L	CI	HL_R	CI	HL_G	CI
Spot	3.378	[1.114,6.102]	5.503	[1.265,11.47]	6.844	[1.837, ∞]
Livestock	2.499	[0.663,5.010]	6.102	[0.755,12.29]	17.24	[1.298, ∞]
Fats&Oil	2.499	[0.683,4.864]	3.449	[0.777,11.47]	57.68	[1.185, ∞]
Foodstuff	3.523	[0.837,7.133]	7.447	[0.967,24.70]	∞	[1.487, ∞]
Raw Industrials	2.801	[1.044,5.010]	5.503	[1.176,8.165]	3.764	[1.775, ∞]
Textiles	2.000	[0.808,3.449]	3.182	[0.895,5.164]	2.461	[1.287, ∞]
Metals	4.596	[1.244,9.033]	6.578	[1.445,24.70]	43.24	[2.353, ∞]

Note: i) The number of lags (k) was chosen by the general-to-specific rule as recommended by Ng and Perron (2001). ii) ρ_L, ρ_R, and ρ_G denote the least squares (LS), recursive mean adjustment (RMA, So and Shin 1999), and grid bootstrap (GT, Hansen 1999) estimates for persistence parameter, respectively. iii) 95% confidence intervals (CI) were constructed by 10,000 nonparametric bootstrap simulations for the LS and RMA estimators, and by 10,000 nonparametric bootstrap simulations on 30 grid points over the neighborhood of the LS estimate for the GT estimator. iv) HL_L, HL_R, and HL_G denote the corresponding half-lives in years, calculated by $(\ln(0.5)/\ln(\rho))/4$.

Table 2. Recursive Out-of-Sample Forecast Results

Index	h	RMSPE$_L$	RMSPE$_R$	RMSPE$_G$	LS/RMA	LS/GT	DM$_R$	DM$_G$
Spot	1	0.045	0.044	0.045	1.031	1.004	1.183 (0.237)	0.180 (0.857)
	2	0.066	0.063	0.064	1.059	1.033	1.808 (0.071)	1.310 (0.190)
	3	0.084	0.078	0.081	1.065	1.029	2.555 (0.011)	1.544 (0.122)
	4	0.104	0.099	0.102	1.050	1.018	2.421 (0.015)	1.324 (0.185)
	6	0.141	0.138	0.139	1.026	1.012	1.456 (0.145)	0.917 (0.359)
Livestock	1	0.082	0.079	0.081	1.035	1.012	1.561 (0.119)	1.182 (0.237)
	2	0.118	0.110	0.115	1.066	1.025	2.598 (0.009)	2.585 (0.010)
	3	0.128	0.124	0.127	1.035	1.012	2.064 (0.039)	1.450 (0.147)
	4	0.144	0.138	0.142	1.039	1.011	2.683 (0.007)	1.839 (0.066)
	6	0.178	0.172	0.174	1.034	1.021	1.810 (0.070)	2.027 (0.043)
Fats&Oil	1	0.110	0.109	0.110	1.003	0.995	0.397 (0.692)	-0.360 (0.719)
	2	0.159	0.157	0.156	1.013	1.018	1.712 (0.087)	1.780 (0.075)
	3	0.174	0.173	0.172	1.008	1.011	1.294 (0.196)	1.543 (0.123)
	4	0.193	0.192	0.192	1.001	1.003	0.230 (0.818)	0.458 (0.647)
	6	0.245	0.246	0.247	0.994	0.992	-1.082 (0.279)	-1.608 (0.108)
Foodstuff	1	0.063	0.062	0.062	1.027	1.029	1.521 (0.128)	1.113 (0.266)
	2	0.090	0.088	0.087	1.032	1.040	2.172 (0.030)	3.458 (0.001)
	3	0.105	0.103	0.103	1.017	1.022	1.532 (0.125)	2.116 (0.034)
	4	0.124	0.122	0.121	1.015	1.020	1.326 (0.185)	1.864 (0.062)
	6	0.157	0.156	0.156	1.003	1.004	0.299 (0.765)	0.559 (0.576)
Raw	1	0.049	0.047	0.048	1.028	1.009	1.053 (0.292)	0.721 (0.471)
Industrials	2	0.076	0.072	0.074	1.057	1.021	1.800 (0.072)	1.444 (0.149)
	3	0.097	0.092	0.095	1.056	1.023	2.639 (0.008)	1.642 (0.101)
	4	0.122	0.118	0.121	1.036	1.010	2.235 (0.025)	0.963 (0.335)
	6	0.162	0.157	0.159	1.030	1.015	1.980 (0.048)	1.339 (0.181)
Textiles	1	0.037	0.037	0.037	0.993	0.989	-0.450 (0.653)	-0.935 (0.350)
	2	0.056	0.056	0.056	0.997	0.999	-0.115 (0.908)	-0.072 (0.943)
	3	0.074	0.075	0.074	0.990	0.994	-0.532 (0.595)	-0.448 (0.654)
	4	0.089	0.091	0.090	0.978	0.985	-1.776 (0.076)	-1.962 (0.050)
	6	0.109	0.113	0.112	0.964	0.973	-2.240 (0.025)	-2.417 (0.016)
Metals	1	0.087	0.085	0.086	1.020	1.014	1.878 (0.060)	0.612 (0.540)
	2	0.139	0.135	0.134	1.031	1.034	2.296 (0.022)	1.283 (0.199)
	3	0.187	0.181	0.178	1.033	1.046	3.540 (0.000)	3.078 (0.002)
	4	0.226	0.223	0.221	1.016	1.024	2.565 (0.010)	2.102 (0.036)
	6	0.309	0.303	0.301	1.019	1.025	2.546 (0.011)	2.458 (0.014)

Note: i) Out-of-sample forecasting was recursively implemented by sequentially adding one additional observation from 69 initial observations toward 139 total observations. ii) The number of lags (k) was chosen by the general-to-specific rule recommended by Ng and Perron (2001). iii) h denotes the forecast horizon (quarters). iv) RMSPE$_L$, RMSPE$_R$, and RMSPE$_G$ denote the root mean squared prediction errors (RMSPE) for the Least Squares (LS), Recursive Mean Adjustment (RMA), and grid bootstrap (GT) estimators, respectively. v) LS/RMA and LS/GT are RMSPE$_L$/RMSPE$_R$ and RMSPE$_L$/RMSPE$_G$, respectively. vi) DM$_R$ and DM$_G$ denote Diebold-Mariano (1995) asymptotic test statistics for the pairs of estimators, LS-RMA and LS-GT. Null hypothesis is equal prediction accuracy. p-values from an asymptotic standard normal distribution are in parenthesis.
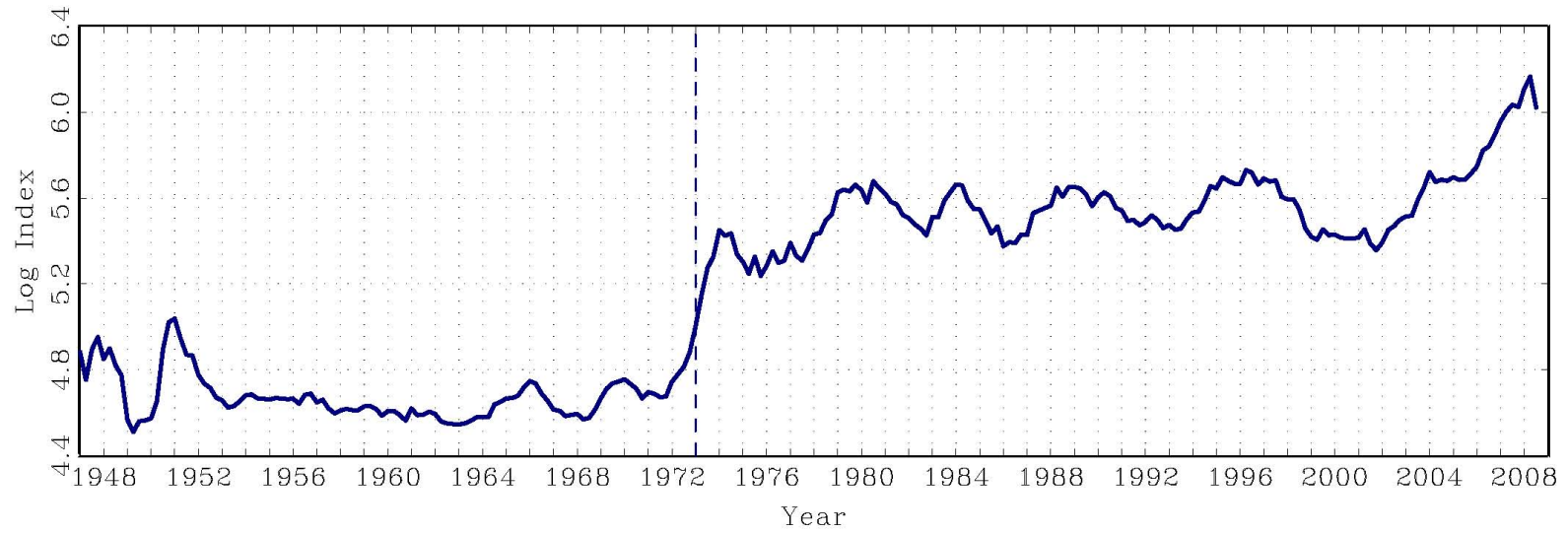
Table 3. Rolling Window Out-of-Sample Forecast Results

Index	h	RMSPE$_L$	RMSPE$_R$	RMSPE$_G$	LS/RMA	LS/GT	DM$_R$	DM$_G$
Spot	1	0.045	0.044	0.044	1.006	1.010	0.328 (0.743)	0.473 (0.636)
	2	0.065	0.062	0.062	1.039	1.054	1.833 (0.067)	1.778 (0.075)
	3	0.079	0.076	0.074	1.046	1.066	2.296 (0.022)	3.348 (0.001)
	4	0.097	0.094	0.093	1.034	1.046	2.116 (0.034)	3.267 (0.001)
	6	0.134	0.130	0.129	1.032	1.043	1.633 (0.102)	2.648 (0.008)
Livestock	1	0.083	0.082	0.083	1.014	1.008	1.162 (0.245)	0.303 (0.762)
	2	0.119	0.115	0.112	1.030	1.058	2.046 (0.041)	2.145 (0.032)
	3	0.126	0.123	0.122	1.026	1.039	2.387 (0.017)	1.683 (0.092)
	4	0.140	0.138	0.135	1.020	1.036	1.531 (0.126)	2.075 (0.038)
	6	0.170	0.168	0.164	1.012	1.036	1.347 (0.178)	3.026 (0.002)
Fats&Oil	1	0.110	0.110	0.108	1.001	1.011	0.094 (0.925)	0.461 (0.645)
	2	0.158	0.158	0.153	1.005	1.037	0.433 (0.665)	1.338 (0.181)
	3	0.173	0.173	0.167	1.001	1.035	0.132 (0.895)	1.892 (0.058)
	4	0.192	0.193	0.188	0.994	1.018	-0.821 (0.411)	1.246 (0.213)
	6	0.248	0.253	0.251	0.980	0.989	-2.277 (0.023)	-1.354 (0.176)
Foodstuff	1	0.062	0.062	0.061	1.010	1.016	0.945 (0.345)	0.793 (0.428)
	2	0.085	0.082	0.080	1.034	1.069	2.068 (0.039)	2.226 (0.026)
	3	0.100	0.098	0.095	1.018	1.057	1.483 (0.138)	3.073 (0.002)
	4	0.117	0.115	0.111	1.016	1.055	1.373 (0.170)	2.824 (0.005)
	6	0.153	0.152	0.148	1.007	1.032	0.656 (0.512)	2.396 (0.017)
Raw Industrials	1	0.048	0.048	0.047	1.007	1.021	0.388 (0.698)	0.828 (0.408)
	2	0.078	0.076	0.074	1.014	1.049	0.516 (0.606)	1.417 (0.156)
	3	0.093	0.093	0.090	1.004	1.035	0.201 (0.841)	1.858 (0.063)
	4	0.120	0.119	0.116	1.007	1.034	0.522 (0.601)	2.206 (0.027)
	6	0.159	0.159	0.156	1.000	1.020	0.016 (0.987)	1.286 (0.198)
Textiles	1	0.037	0.037	0.037	1.017	1.002	0.745 (0.457)	0.213 (0.832)
	2	0.058	0.056	0.057	1.029	1.009	1.203 (0.229)	0.563 (0.573)
	3	0.074	0.073	0.074	1.010	0.999	0.482 (0.629)	-0.066 (0.947)
	4	0.087	0.088	0.088	0.990	0.991	-1.004 (0.316)	-1.408 (0.159)
	6	0.106	0.108	0.108	0.985	0.985	-1.211 (0.226)	-1.633 (0.103)
Metals	1	0.083	0.084	0.083	0.998	1.006	-0.127 (0.899)	0.282 (0.778)
	2	0.133	0.134	0.132	0.997	1.014	-0.111 (0.912)	0.454 (0.650)
	3	0.171	0.170	0.165	1.004	1.035	0.369 (0.712)	2.439 (0.015)
	4	0.215	0.215	0.210	1.003	1.028	0.440 (0.660)	2.909 (0.004)
	6	0.293	0.292	0.288	1.002	1.019	0.214 (0.831)	1.471 (0.141)

Note: i) Out-of-sample forecasting was implemented by sequentially adding one additional observation and dropping one observation in each iteration, maintaining 69 observations. ii) The number of lags (k) was chosen by the general-to-specific rule recommended by Ng and Perron (2001). iii) h denotes the forecast horizon (quarters). iv) RMSPE$_L$, RMSPE$_R$, and RMSPE$_G$ denote the root mean squared prediction errors (RMSPE) for the Least Squares (LS), Recursive Mean Adjustment (RMA), and grid bootstrap (GT) estimators, respectively. v) LS/RMA and LS/GT are RMSPE$_L$/RMSPE$_R$ and RMSPE$_L$/RMSPE$_G$, respectively. vi) DM$_R$ and DM$_G$ denote Diebold-Mariano (1995) asymptotic test statistics for the pairs of estimators, LS-RMA and LS-GT. Null hypothesis is equal prediction accuracy. p-values from an asymptotic standard normal distribution are in parenthesis.

Figure 1. CRB Historical Data

CRB Spot Index: 1947.I — 2008.III

CRB Spot Index: 1974.I — 2008.III