DBJ Discussion Paper Series, No.0905

Bagging and Forecasting in Nonlinear Dynamic Models

Mari Sakudo
(Research Institute of Capital Formation,
Development Bank of Japan,
and
Department of Economics, Sophia University)

December 2009

Discussion Papers are a series of preliminary materials in their draft form. No quotations, reproductions or circulations should be made without the written consent of the authors in order to protect the tentative characters of these papers. Any opinions, findings, conclusions or recommendations expressed in these papers are those of the authors and do not reflect the views of the Institute.

Bagging and Forecasting in Nonlinear Dynamic Models*

Mari Sakudo†
Development Bank of Japan,
and Sophia University

First Draft, September 2003
This Draft, December 2009

Abstract

This paper proposes new variants of point forecast estimators in Markov switching models (Hamilton, 1989) utilizing bagging (Breiman, 1996), and applies them to study real GNP in the U.S. The empirical and Monte Carlo simulation results on out-of-sample forecasting show that the bagged forecast estimators outperform the benchmark forecast estimator by Hamilton (1989) in the sense of the prediction mean squared error. The Monte Carlo experiments present that interactions between a Markov process for primitive states and an innovation affect the relative performance of the bagged forecast estimators, and that effectiveness of the bagging does not die out as sample size increases.

Keywords: Bagging; Bootstrap; Forecast; Regime Switching; Time Series

JEL Classifications: C13; C15; C53; E37

---

*I am truly indebted to Frank Diebold, Masayuki Hirukawa, Atsushi Inoue, Yuichi Kitamura, Kevin Song and seminar participants at Development Bank of Japan and University of Tokyo for helpful comments and suggestions. The earlier version of this paper is entitled "The Bootstrap and Forecasts in Markov Switching Models."

†*Address:* Development Bank of Japan, 9-3, Otemachi 1-chome, Chiyoda-ku, Tokyo, Japan 100-0004. Tel.: +81-3-3244-1241; fax: +81-3-3270-7084.

1 Introduction

The bagging (bootstrap aggregating) method (Breiman, 1996) has recently been applied in analysis of forecasting time series data (Inoue and Kilian, 2003, 2008; Lee and Yang, 2006; Stock and Watson, 2005). There, the presumed models are linear, or binary and quantile. In the existing literature on bagging and time series forecasting, genuine dynamics—autoregressive component—of the process of interest does not play a key role. Markov switching models (Hamilton, 1989) are widely used in analyses of time series economic and financial data (for example, Ang and Bekaert (2002); Garcia and Perron (1996); Guidolin and Timmermann (2006, 2007); Perez-Quiros and Timmermann (2000)), where the nonlinear dynamics is an important element. However, to the best of my knowledge, no attempt has been made to implement the bagging method in Markov switching models. This paper proposes new variants of point forecast estimators in Markov switching models utilizing the bagging method, and applies them to forecast nonlinear dynamics of real GNP in the United States.

In many economic and financial data, nonlinearity (Teräsvirta, 2006) or structural breaks (Stock and Watson, 1996) exist, suggesting that economic agents' forecasting must also be nonlinear (Clements and Hendry, 2006). Markov switching models (Hamilton, 1989)—a special case is the threshold autoregression (TAR) model (Tong, 1983) of observable shifts in regime—adapt nonlinearity by making inference on stochastic changes from the discrete-state process. The nonlinearity from shifts in regimes is determined by the interaction between the data and the Markov chain. As the sample size is often small in forecasting application, Markov switching models provide us flexible nonlinearity in a parsimonious way.

Another key ingredient of forecasting estimators in this paper is bagging. Breiman (1996) introduces bagging in the field of machine learning in random sample settings. The bagging method uses the bootstrap (Efron, 1979; Horowitz, 2000), but is a method to improve the accuracy of the estimators or predictors themselves—rather than to approximate distributions and access the accuracy of parameter estimators or predictions. In the context of cross sections, the existing literature on bagging studies estimators for nonlinear models of smooth functions (Friedman and Hall, 2000), indicator predictors (Bühlmann and Yu, 2001), and predictive risk estimators for discontinuous loss functions (Kitamura, 2001).

Recently, bagging has been analyzed in time series context. Lee and Yang (2006) study bagging binary and quantile predictors, where the improvement of accuracy in predictions comes from non-differentiabilities of objective functions. Inoue and Kilian (2003, 2008) and Stock and Watson (2005) apply bagging to construct forecasts among many possible predictors in linear dynamic models. In their cases, bagging improves the accuracy of forecasts, due to variance reduction in the part of predictors other than deterministic and autoregressive components. In contrast, this paper proposes bagged forecast estimators in nonlinear dynamic models of stochastic regime switching. Improvement in accuracy of forecasts comes from nonlinear dynamics of the process of interest.[1] In other words, it adopts the idea of bagging in nonlinear models in cross section context by Friedman and Hall (2000) to dynamic forecasting situations.

To implement bagging, we resample data by the bootstrap, construct a forecast—or an estimate, or an objective function—for each bootstrap sample, and take an average over the bootstrap replications. In the existing literature on bagging in time series forecasting, in order to capture the dependence structure of the data, Lee and Yang (2006) and Inoue and Kilian (2003, 2008) conduct the blocks bootstrap (Gonçalves and White, 2004; Hall and Horowitz, 1996). Stock and Watson (2005) implement the parametric bootstrap assuming a normal distribution to the error term. The model is reduced to a linear model with independent innovations.

In contrast, I implement bootstrapping in nonlinear dynamic models in which the nonlinearity comes from the changes in non-independent stochastic components. The nonlinear dynamic model in which I implement the bootstrap can be reduced to an autoregressive moving average (ARMA) model with serially dependent innovations. Direct draws of the serially dependent innovations in the ARMA model are not feasible. Conducting the nonparametric bootstrap without using information on the assumed model structure is not desirable. Therefore, for nonlinear dynamic models of stochastic regime shifts, to implement the bootstrap using the model information, values of the stochastic states also need to be drawn from the estimated probability distribution. I propose new variants of the parametric bootstrap and the residual bootstrap as I explain in detail later.

Empirical and Monte Carlo simulation results on out-of-sample point forecasting show that the bagged forecast estimators outperform the benchmark forecast estimator by Hamilton (1989) in terms of the prediction mean squared error. Monte Carlo results present that interactions between a Markov process for primitive states and an innovation have an influence on the relative performance of the bagged forecast estimators. Moreover, the experimental results present that the accuracy improvement by utilizing the bagging does not

[1] One recent paper by Li (2009) proposes bootstrapping in the TAR models. Both his paper and my paper study forecasting in nonlinear dynamic models. However, Li (2009) uses the bootstrap to approximate the distribution of predictors and improve confidence intervals of prediction, whereas I use bootstrap aggregating to improve the forecasts themselves.

die out even in large sample. A possible reason is that the nonlinearity, which comes from regime shifts, is stochastic in Markov switching models. The uncertainty about nonlinearity exists regardless of sample size; hence, even in large sample. The bagging reduces the variance of forecast that stems from stochastic nonlinearity.

This paper is organized as follows. The next section describes the model. Section 3 explains the bootstrap sampling procedure, the forecast estimators using the bagging, and the forecast evaluation method. Section 4 explains Monte Carlo simulation procedure, and presents Monte Carlo evidence on the numerical performance of the bagged forecast estimators in Markov switching models. In section 5, I estimate the business cycle model on real GNP in the United States—the first difference obeys a nonlinear stationary process, construct forecasts, and present the results. Section 6 concludes.

2 Model

A process of $\tilde{y}_t$ obeys a special case of an ARIMA model, in order to study forecasts as in Hamilton (1989) for comparison.

$$\Delta\tilde{y}_t = c_{s_t} + \phi_1\Delta\tilde{y}_{t-1} + \phi_2\Delta\tilde{y}_{t-2} + \cdots + \phi_r\Delta\tilde{y}_{t-r} + \epsilon_t, \tag{1}$$

where $\phi_j, j = 1, ..., r$ are parameters, $\Delta\tilde{y}_t = \tilde{y}_t - \tilde{y}_{t-1}$, an innovation of ϵ_t is i.i.d.$(0, \sigma^2)$, and s_t represents the unobserved state at date t. The constant term c_{s_t} depends on the state at date t, s_t. For simplicity, non-dependency of ϕ_j on the state conditions for $j = 1, ..., r$, is assumed.

Furthermore, the process of $\tilde{y}_t$ is the sum of a Markov trend process of ARIMA(1,1,0), n_t, and an ARIMA(r,1,0) process without drift, $\tilde{z}_t$.

$$\tilde{y}_t = n_t + \tilde{z}_t, \tag{2}$$

where $\tilde{z}_t - \tilde{z}_{t-1} = \phi_1(\tilde{z}_{t-1} - \tilde{z}_{t-2}) + \cdots + \phi_r(\tilde{z}_{t-r} - \tilde{z}_{t-r-1}) + \epsilon_t$, $n_t - n_{t-1} = \alpha_1 s_t^* + \alpha_0$, and ϵ_t is independent of n_{t+j} for all j. Here, s_t^*'s are primitive states that follow a first-order Markov process. α_0 and α_1 are parameters. The primitive state, s_t^*, is such that

$$\begin{aligned} &Prob(s_t^* = 1|s_{t-1}^* = 1) = p \\ &\text{and} \\ &Prob(s_t^* = 0|s_{t-1}^* = 0) = q. \end{aligned} \tag{3}$$

To simplify notations, let $y_t \equiv \Delta\tilde{y}_t$ and $z_t \equiv \Delta\tilde{z}_t$. Detailed explanation on the model is in the appendix.

3 Bagged Forecasts

3.1 The Bootstrap

First of all, ML estimates are obtained as in Hamilton (1989). For the bootstrap data, random draws for the independent innovations are obtained by either (1) the residual bootstrap, or (2) the parametric bootstrap with a normally distributed error.

Let $\mathcal{Y}\tau = (y_\tau', y_{\tau-1}, \ldots, y_{-m})'$ be a vector containing all observations obtained through date τ. The state s_t is a first-order $J(= 2^{r+1})$ state Markov chain. Let $P(s_t = j|\mathcal{Y}_\tau; \theta)$ denote a probability of the state s_t conditional on data obtained through date τ given the population parameters θ. Collect these conditional probabilities in a $(J \times 1)$ vector of $\xi_{t|\tau}$. Let $\hat{X}$ denote an estimate of X.

3.1.1 Residual Bootstrap

I calculate residuals from estimates of the parameters in the first estimation and from observed data.

$$\hat{\epsilon}_t = y_t - \hat{E}(c_{s_t}|\mathcal{Y}_\tau) - \hat{\phi}_1 y_{t-1} - \cdots - \hat{\phi}_r y_{t-r}, \qquad t = r+1, \ldots, T \tag{4}$$

In the above formula, I use the parameter estimates of ϕ, α's, and the estimated state probabilities. The negative of the second term on the right hand side is $\hat{E}(c_{s_t}|\mathcal{Y}_\tau) = (\hat{c}_1, \ldots, \hat{c}_J)'\hat{\xi}_{t|\tau}$, where $\tau = T$ or $\tau = t$. In other words, $\hat{\xi}_{t|\tau}$ is either the vector of estimated smoothed probabilities, $\hat{\xi}_{t|T}$, or the vector of estimated inferred probabilities, $\hat{\xi}_{t|t}$. $\hat{c}_j$, for $j = 1, \ldots, J$, are estimated constant terms for each state, s_t, in Equation (1).

Then, I repeat the following procedure B times, for $b = 1, \ldots, B$. The subscript b denotes the bootstrap replication, from 1 to B.

1. Draw bootstrapped residuals $\{e_t^b\}_{t=r+1}^T$ with replacement from the original residuals $\{\hat{\epsilon}_t\}_{t=r+1}^T$.

2. For $t = r+1, \ldots, T$, construct blocks of η_t's such that

$$\eta_t \equiv \begin{pmatrix} y_{t-r} \\ \vdots \\ y_t \end{pmatrix}. \tag{5}$$

   Assume that the distribution puts equal probability mass to each block. Draw an initial block of $(y_1^b, \ldots, y_r^b)'$ from the density distribution of the blocks.

3. For $t = 1 + r, \ldots, T$, draw state at date t, s_t^b, from the estimated probabilities, $\hat{\xi}_{t|\tau}$: either from the estimated smoothed probabilities, $\hat{\xi}_{t|T}$, or from the estimated inferred probabilities, $\hat{\xi}_{t|t}$.

4. Starting with the initial bootstrap observation block, $(y_1^b, \ldots, y_r^b)'$, construct the bootstrap sample of y_t^b's recursively by

$$y_t^b = \hat{E}^b(c_{s_t}|\mathcal{Y}_\tau) + \hat{\phi}_1 y_{t-1}^b + \cdots + \hat{\phi}_r y_{t-r}^b + \hat{\epsilon}_t^b, t = r+1, ..., T, \tag{6}$$

where $\hat{E}^b(c_{s_t}|\mathcal{Y}_\tau) = (\hat{c}_1, \ldots, \hat{c}_J)' s_t^b$. τ is T when the estimated smoothed state probabilities are used to draw states, and is t when the estimated inferred probabilities are used. Here, the bootstrap sample starts at date $t = r + 1$ rather than $t = 1$. As I utilize the estimated smoothed probabilities that are obtained from date $r + 1$ to date T, I set the time period from 1 to r as an initial time block.

3.1.2 Parametric Bootstrap with a Normally Distributed Error

For each bootstrap replication, $b = 1, \ldots, B$, I draw the bootstrap errors, $\{e_t^b\}_{t=r+1}^T$, from the normal distribution with mean zero and the estimated variance from the first estimation, $\hat{\sigma}^2$. The rest of the procedure is the same as in the case of the residual bootstrap above.

3.2 Forecasting Using Bagging

One-step-ahead optimal point forecast based on observable variables[2] is a sum of products of forecasts conditional on states and inferred state probability at date $T + 1$, $\xi_{T+1|T}$.

$$E(y_{T+1}|\mathcal{Y}_T; \theta) = \textstyle\sum_{j=1}^J P(s_{T+1} = j|\mathcal{Y}_T; \theta) E(y_{T+1}|s_{T+1} = j, \mathcal{Y}_T; \theta) \tag{7}$$

After resampling data by the bootstrap, I estimate parameters and state probabilities for each bootstrap replication, $b = 1, \ldots, B$. For each bootstrap sample of $\mathcal{Y}_T$, $b = 1, \ldots, B$, the forecast estimator is

$$\hat{E}^b(y_{T+1}|\mathcal{Y}_T, \theta) = \textstyle\sum_{j=1}^J \hat{P}^b(s_{T+1} = j|\mathcal{Y}_T; \theta) \hat{E}^b(y_{T+1}|s_{T+1} = j, \mathcal{Y}_T; \theta). \tag{8}$$

In Equation (8), I calculate the estimates of inferred probabilities of states at date $T + 1$ by

$$\hat{\xi}_{T+1|T}^b = \hat{\Lambda}^b \hat{\xi}_{T|T}^b, \tag{9}$$

[2]We can consider longer than one-step-ahead point forecasts and see how the forecast horizon affects relative performance of the bagged forecast estimator. In this paper, I analyze one-step-ahead forecasting as a start.

where Λ denotes the $(J \times J)$ transition matrix of state s_t. I obtain the estimates of conditional forecasts for $j = 1, \ldots, J$, by

$$\hat{E}^b(y_{T+1}|s_{T+1} = j, \mathcal{Y}_T; \theta) = \hat{c}_j^b + \hat{\phi}_1^b y_T^b + \cdots + \hat{\phi}_r^b y_{T-r}^b. \tag{10}$$

Taking an average of the maximum likelihood forecast estimators over the bootstrap samples, the one-step-ahead bagged forecast estimator is defined as follows.[3]

$$\hat{E}(y_{T+1}|\mathcal{Y}_T; \theta) \approx \tfrac{1}{B} \textstyle\sum_{b=1}^{B} \hat{E}^b(y_{T+1}|\mathcal{Y}_T; \theta) \tag{11}$$

The expression on the right hand side in Equation (11) is a Monte Carlo estimate of the bagged forecast estimator, which approaches the true bagged forecast as B goes to infinity. In estimation, I set the number of the bootstrap replications at 100.

3.3 Forecast Evaluation

The updating scheme I adopt is rolling horizon—rolling forecasts are constructed from only the most recent fixed interval of the past data. The length of the data window, which is the sample size for estimating and forecasting, T, remains the same as date is updated.

Let $\hat{y}_{t+h}$ denote $\hat{E}(y_{t+h}|\mathcal{Y}_t; \theta)$, an h-step-ahead forecasting estimator, where t is the time the forecast is made. An optimal forecast here implicitly uses squared difference between a forecast and outcome as a loss function. The risk used is the expected loss—the expectation is taken over the outcome of Y_{t+h} given $\mathcal{Y}_t$ and holding parameters θ fixed, as its value is unknown at the time the forecast is made:

$$E[(y_{t+h} - \hat{y}_{t+h})^2|\mathcal{Y}_t; \theta]. \tag{12}$$

The classical approach to forecasting uses the same risk on evaluating the forecast. In contrast, I use the population prediction mean squared error (PMSE) conditional on the outcome variable at the forecasting date:

$$E[(y_{t+h} - \hat{y}_{t+h})^2|y_{t+h}, \mathcal{Y}_t; \theta]. \tag{13}$$

Explanation on the PMSE and its sample analogue used is in the appendix.

[3]For simplicity, I abuse notation and use the same notation E for the expectation with respect to the bootstrap probability measure.

4 Monte Carlo Experiments

4.1 Monte Carlo Procedure

In Monte Carlo simulation, I set the number of lags in autoregression at one, $r = 1$, or at four, $r = 4$ as in Hamilton (1989). For simplicity, I explain Monte Carlo simulation procedure in the case of one lag. Thus, the process of y_t, that is, $\Delta\tilde{y}_t$, is rewritten as

$$\begin{aligned} &y_t = c_{s_t} + \phi_1 y_{t-1} + \epsilon_t, \\ &\text{where} \\ &c_{s_t} = \alpha_1 s_t^* + \alpha_0 - \phi_1(\alpha_1 s_{t-1}^* + \alpha_0). \end{aligned} \tag{14}$$

The primitive state at date 0 of s_0^* has a Bernoulli distribution such that the probability of state 1 is $p_{s_0^*}$. I assume that the initial primitive state probability is ergodic. I also assume that an initial observation y_0 is normally distributed, $y_0 \sim N(m_0, \sigma_0^2)$, where the mean is m_0, and σ_0^2 is the variance. The mean of an initial observation is defined as the expectation of a constant term associated with the initial state, $m_0 \equiv p_{s_0^*} \cdot \mu_0 + (1 - p_{s_0^*}) \cdot \mu_1$.

Given assumed parameter values of $\{\phi_1, p, q, \alpha_0, \alpha_1, \sigma, p_{s_0^*}, \sigma_0\}$, I generate the process of $y_t, t = 1, \ldots, T$, as follows. First, I draw an initial primitive state, s_0^*, from the Bernoulli distribution of $(p_{s_0^*}, 1 - p_{s_0^*})'$. An initial value of the process, y_0, is drawn from the normal distribution of $N(m_0, \sigma_0^2)$.

Second, I recursively draw primitive states, s_t^*, for $t = 1, \ldots, T$. Primitive state, s_t^*, is drawn from Bernoulli distribution of $(p, 1-p)'$ if primitive state at the previous date is in state 1 (i.e. $s_{t-1}^* = 1$), and is drawn from Bernoulli distribution of $(q, 1-q)'$ if primitive state at the previous date is in state 0 (i.e. $s_{t-1}^* = 0$). Then, state is defined as $s_t \equiv (I\{s_t^* = 1\}, I\{s_t^* = 0\}, I\{s_{t-1}^* = 1\}, I\{s_{t-1}^* = 0\})', t = 1, \ldots, T$, where $I\{A\}$ is an indicator function that takes 1 if A is true and 0 otherwise.

Third, I draw errors of ϵ_t's from the i.i.d. normal distribution of $N(0, \sigma^2)$ and iterate on $y_t = c_{s_t} + \phi_1 y_{t-1} + \epsilon_t$, for $t = 1, \ldots, T$.

4.2 Results

Two benchmark parameter settings are as below. I perturb a presumed parameter value or a presumed setting from the benchmark setting and see how forecast performance changes across different presumed settings. The first setting corresponds to the paper by Hamilton (1989). The second is a simplest setting.

- *Parameter setting as in Hamilton (1989):*
  $r = 4, \alpha_0 + \alpha_1 = 1.1647, \alpha_0 = -0.3577, \phi_1 = 0.014, \phi_2 = -0.58, \phi_3 = -0.247, \phi_4 = -0.213, \sigma = 0.796, \sigma_0 = 0.796, p = 0.9049, q = 0.755$. Note that $p_{s_0^*} = 0.72037636$ by the assumption that the initial primitive state probability is ergodic.

- *Base parameter setting:*
  $r = 1, \alpha_0 + \alpha_1 = 1.2, \alpha_0 = -0.4, \phi_1 = 0.1, \sigma = 0.7, \sigma_0 = 0.7, p = 0.85, q = 0.7$. Note that $p_{s_0^*} = 0.66666667$ by the assumption that the initial primitive state probability is ergodic.

Monte Carlo simulation results on out-of-sample one-step-ahead point forecasts show that the bagged forecast estimators dominate the benchmark forecast estimator in terms of the PMSE since the bagging reduces the variance. The tables of all the results are in the appendix.

First, magnitude of the improvement by the bagging depends on uncertainty in the data generating process. Table 7 compares the benchmark forecasts and the bagged forecasts that use the parametric bootstrap with smoothed probability across different standard deviations of the independent innovation term. The advantage to using the bagging becomes small when the standard deviation is small. Note that 'Percent difference in PMSE' takes a negatively large value when the bagging improves forecasts significantly.

Second, Tables 3 and 9 show the results across different Markov processes for the primitive state. If both states are less persistent (for example, if Markov transition probability of staying in the same state as at a previous date is 0.2) magnitude of the PMSE improvement becomes small.[4] Note that persistence of states tends to be high in real data. For instance, each Markov transition probability is larger than 0.7 in a study of real GNP in the U.S. by Hamilton (1989). It is around 0.9 in an analysis of stock returns by Guidolin and Timmermann (2007).

Uncertainty from the Markov process and the independent innovations determines nonlinearity of the observed process. Hence, relative performance of the bagged forecast estimators

[4]Overall, as the Markov transition probability of state 1 conditional on state 1 at the previous date, p, is larger, PMSE improvement by the bagging increases. Note that an absolute value of the constant term is larger in state 1 than in state 0 in these examples. High persistence of the state that generates a large constant increases PMSE improvement by the bagging. However, given small values of p (for example, $p = 0.2$) the improvement is larger as another Markov transition probability of state 0 conditional on state 0 at previous date is larger. That is, if state 1 is less persistent, the PMSE improvement by the bagging is larger as persistence of state 0 increases.

across different Markov transition probabilities is interrelated to magnitude of uncertainty from the independent innovations. If the standard deviation of the independent innovation term is smaller than that in the base parameter setting (for example, $\sigma = 0.5$) the relative performance of the bagged forecast estimators depends more on the Markov transition probability of states as in Table 11.[5]

Third, percentage improvement in the PMSE when using the bagging is similar across different parameter values of coefficient of the lagged dependent variable as in Table 8. Note that non-dependency of coefficient on the state conditions is assumed in the model.

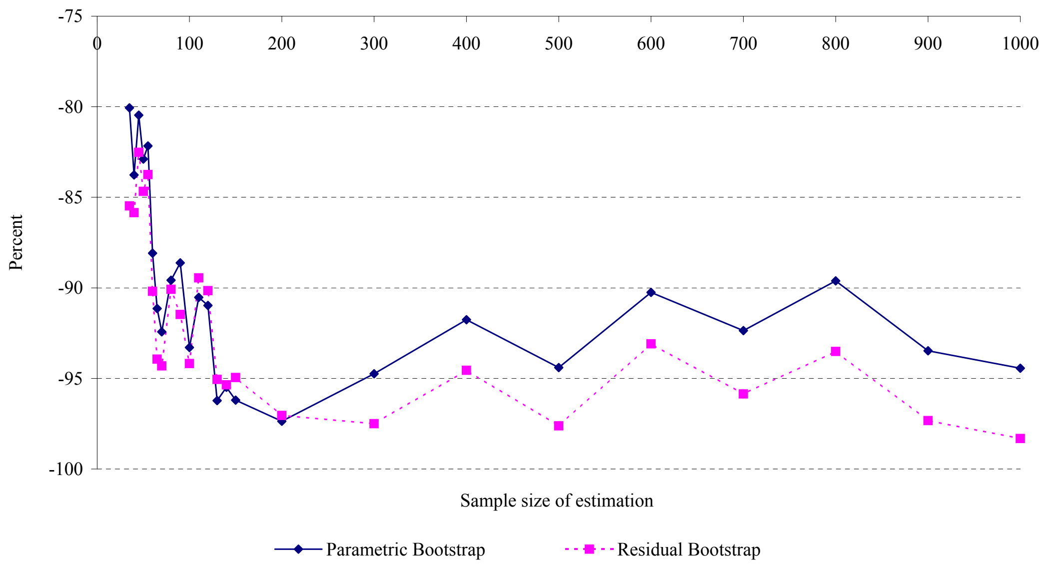

Figure 1: Percent difference in PMSE by sample size

Fourth, Figure 1 shows relative performance of the bagged forecast estimators by sample size of estimation.[6] Monte Carlo simulation results based on Markov switching models show the bagging is also effective in large sample. A possible reason is that the nonlinearity comes from regime shifts and is stochastic in Markov switching models. The uncertainty about

[5]If the standard deviation of the independent innovation is larger (for example, $\sigma = 0.9$) the magnitude of PMSE improvement by the bagging does not vary much across different Markov transition probabilities as in Table 12. Table 13 shows the results in smaller sample, $T = 35$. Table 10 compares the benchmark and five bagging methods for different Markov transition probabilities of states.

[6]In Figure 1, plots are at sample sizes $T = 35, 40, 50, 60, \ldots, 140, 150, 200, 300, \ldots, 1000$. Sample sizes for forecast evaluation M are set to 100. I use smoothed probability of states to randomly draw states for both the parametric and the residual bootstrap, and to construct original residuals for the residual bootstrap. In Tables 4, 5 and 6, I set long horizons and compare the performance of the benchmark and five bagging methods for sample sizes $T = 40, 150$, and 500, respectively.

nonlinearity exists regardless of sample size; hence, even in large sample. It implies that the nonlinear forecasts are unstable even in large sample. The bagging reduces variances of forecasts that stem from stochastic nonlinearity. Hence, as long as a Markov process for primitive states generates nonlinearity, the bagging improves forecasts.

5 Application: Postwar Real GNP in the U.S.

I apply the above methods to the real GNP quarterly data in the United States. The data come from Business Conditions Digest. The sample period is from the second quarter of year 1951 to the fourth quarter of year 1984. The level of GNP is measured at an annual rate in 1982 dollars. I let x_t be the real GNP. $\tilde{y}_t = \log(x_t)$. For computational convenience, I multiply $\triangle\tilde{y}_t$ by 100 in the estimation: $y_t = 100 \times \triangle\tilde{y}_t$. The variable of y_t is 100 times the first difference in the log of real GNP. I set $r = 4$ as in Hamilton (1989) and study out-of-sample one-step-ahead point forecasts.

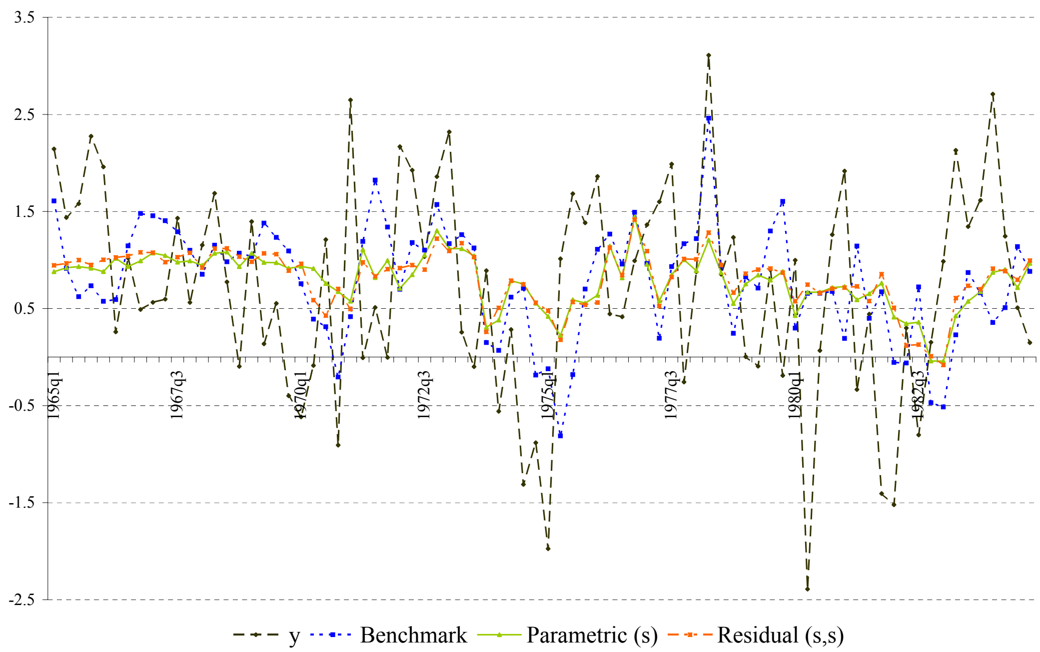

Figure 2: Out-of-sample one-step-ahead forecast estimates: $T = 55$

Figure 2 compares (1) forecasts by the benchmark estimator, (2) bagged forecasts using the parametric bootstrap with smoothed state probability for random draws, and (3) bagged forecasts using the residual bootstrap with smoothed probability for random draws and original residuals. Table 14, Table 15, Table 16, Table 17, and Table 18 in the appendix show the results of means, biases, variances, and PMSE for $T = 45, 55, 65, 75$, and 85, respectively. Due to the variance reduction, the bagging improves forecasts in the sense of the prediction mean squared error. Overall, the magnitude of improvement in the prediction mean squared error by the bagging is larger in smaller sample. However, note that sample sizes for forecast evaluation are not identical across these tables.

6 Conclusion

This paper proposes new variants of point forecast estimators in Markov switching models utilizing bagging. To construct the bagged forecast estimators, I implement the parametric bootstrap and the residual bootstrap to nonlinear dynamic models in which the nonlinearity comes from the changes in non-independent stochastic components.

I conduct Monte Carlo experiments to compare performance of the bagged forecast estimators with that of the benchmark forecast estimator in the sense of the PMSE. The Monte Carlo simulation results show that interactions between a Markov process for primitive states and independent innovations affect the relative performance of the bagged forecast estimators. First, the advantage to using the bagging becomes small when uncertainty of the independent innovations is small. Second, if all primitive states are less persistent, magnitude of the PMSE improvement by the bagging becomes small. Third, if uncertainty from the independent innovations is smaller, relative performance of the bagged forecasts depends more on the Markov transition probabilities. Fourth, the Monte Carlo simulations present that the bagged forecast estimators dominate the benchmark forecast estimator even in large sample. A possible reason is that the nonlinearity which comes from regime shifts, is stochastic in Markov switching models. The nonlinear forecasts are unstable regardless of sample size as long as the Markov process produces nonlinearity. Hence, the bagging reduces forecast variances that stem from stochastic nonlinearity.

I also apply the bagged forecast estimators to study nonstationary time series of postwar U.S. real GNP as in Hamilton (1989), where the first difference obeys a nonlinear stationary process. The empirical evidence on out-of-sample forecasting presents that the bagged forecast estimators outperform the benchmark forecast estimators by Hamilton (1989) in the sense of the prediction mean squared error, due to the variance reduction.

References

Ang, A. and G. Bekaert (2002, April). Regime switches in interest rates. *Journal of Business and Economic Statistics 20*(2), 163–182.

Breiman, L. (1996). Bagging predictors. *Machine Learning 26*, 123–140.

Bühlmann, P. and B. Yu (2001). Analyzing bagging.

Clements, M. P. and D. F. Hendry (2006). Forecasting with breaks. In G. Elliott, C. W. Granger, and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, Volume I, Chapter 12, pp. 605–657. Elsevier B.V.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics 7*, 1–26.

Friedman, J. H. and P. Hall (2000). On bagging and nonlinear estimation.

Garcia, R. and P. Perron (1996). An analysis of the real interest rate under regime shifts. *The Review of Economics and Statistics 78*(1), 111–125.

Gonçalves, S. and H. White (2004). Maximum likelihood and the bootstrap for nonlinear dynamic models. *Journal of Econometrics 119*, 199–219.

Guidolin, M. and A. Timmermann (2006). An econometric model of nonlinear dynamics in the joint distribution of stock and bond returns. *Journal of Applied Econometrics 21*, 1–22.

Guidolin, M. and A. Timmermann (2007). International asset allocation under regime switching, skew and kurtosis preferences.

Hall, P. and J. L. Horowitz (1996, July). Bootstrap critical values for tests based on generalized-method-of-moments estimators. *Econometrica 64*(4), 891–916.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica 57*, 357–384.

Horowitz, J. L. (2000). The bootstrap.

Inoue, A. and L. Kilian (2003). Bagging time series models.

Inoue, A. and L. Kilian (2008, June). How useful is bagging in forecasting economic time series? A case study of U.S. consumer price inflation. *Journal of the Amerian Statistical Association 103*(482), 511–522.

Kitamura, Y. (2001). Predictive inference and the bootstrap.

Lee, T.-H. and Y. Yang (2006). Bagging binary and quantile predictors for time series. *Journal of Econometrics 135*, 465–497.

Li, J. (2009). Bootstrap prediction intervals for threshold autoregressive models.

Perez-Quiros, G. and A. Timmermann (2000, June). Firm size and cyclical variations in stock returns. *The Journal of Finance LV*(3), 1229–1262.

Stock, J. H. and M. W. Watson (1996, January). Evidence on structural instability in macroeconomic time series relations. *Journal of Business and Economic Statistics 14*(1), 11–30.

Stock, J. H. and M. W. Watson (2005). An empirical comparison of methods for forecasting using many predictors.

Teräsvirta, T. (2006). Forecasting economic variables with nonlinear models. In G. Elliott, C. W. Granger, and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, Volume I, Chapter 8, pp. 413–457. Elsevier.

Tong, H. (1983). *Threshold Models in Non-linear Time Series Analysis.* New York, Berlin, Heidelberg, Tokyo: Springer-Verlag.

A Appendix: Model

Two primitive states are assumed for simplicity. The state at date t, s_t, depends on the primitive states at date from t to rth lags (s_t^*,...,s_{t-r}^*) and is independent of past observations on $\tilde{y}_t$. The constant term in Equation (1) becomes $c_{s_t} = \alpha_1 s_t^* + \alpha_0 - \phi_1(\alpha_1 s_{t-1}^* + \alpha_0) - \cdots - \phi_r(\alpha_1 s_{t-r}^* + \alpha_0)$. To normalize, I set $\alpha_1 > 0$. In the estimation, I assume normality of the innovation term ϵ_t. Let $\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_r L^r$. I assume the stability condition that all the roots of $\phi(z) = 0$ are greater than 1 in absolute value.

A.1 Discussion of the Model

$\tilde{z}_t$ is a driftless ARIMA(r,1,0) process whose first difference contains i.i.d.$(0, \sigma^2)$ innovations. In contrast, n_t is an ARIMA(1,1,0) process whose first difference contains the non-independent white noise innovations.

A.1.1 Primitive State Process – Serial Dependence, Non-normality, and Strict Stationarity

By the assumption of the first-order Markov process, the primitive state process is strictly stationary, and is an AR(1) process with unusual probability distribution of the innovation sequence.

$$s_t^* = (1-q) + \lambda s_{t-1}^* + \upsilon_t, \tag{15}$$

where $\lambda = -1 + p + q$. Thus, conditional on $s_{t-1}^* = 1$, $\upsilon_t = 1 - p$ with probability p, and $\upsilon_t = -p$ with probability $1-p$. Conversely, conditional on $s_{t-1}^* = 0$, $\upsilon_t = q - 1$ with probability q, and $\upsilon_t = q$ with probability $1 - q$.

The innovation term, υ_t, is a white noise. That is, it is zero mean and serially uncorrelated: $E(\upsilon_t | s_{t-j}^*) = 0$, hence, $E(\upsilon_t) = 0$, $E(\upsilon_t s_{t-j}^*) = 0$ and $E(\upsilon_t \upsilon_{t-j}) = 0$ for $j = 1, 2, \ldots$. However, the innovation term, υ_t, is not independent of lagged value of the primitive state. For instance, $E(\upsilon_t^2 \mid s_{t-1}^*)$ depends on s_{t-1}^* : $E(\upsilon_t^2 \mid s_{t-1}^* = 1) = p(1-p)$ and $E(\upsilon_t^2 \mid s_{t-1}^* = 0) = q(1-q)$.

First, the innovation term, υ_t, cannot be normally distributed, whereas I suppose normality of the innovations in the other AR component, ϵ_t's, in the estimation. Second, the serial dependence of the system $\tilde{y}_t$ comes only from the innovation term in the Markov process, υ_t. It is because the other innovation term of ϵ_t is i.i.d. and is independent of n_{t+j} for all j. The serial dependence of υ_t generates a nonlinear process for the observed series $\tilde{y}_t$. The best forecast is nonlinear because of this serial dependence.

Since the innovation term, υ_t, is a white noise, the primitive state process is covariance stationary when $|p + q - 1| < 1$. Moreover, the primitive state process is strictly stationary because the joint distribution of the processes depends only on their relative distance. The primitive state is strictly stationary when $|p + q - 1| = 1$, too. But, in this case, the primitive state is degenerated to be a constant, that is, a non-stochastic process, and the process of $\tilde{y}_t$ becomes a usual ARIMA process whose first difference has i.i.d. innovations. The condition such that $|p + q - 1| > 1$ cannot be the case, because $0 \leq p, q \leq 1$ hold.

A.1.2 I(1) Process

Although the process of $\tilde{y}_t(=\log(x_t))$ contains the special non-independent innovations, it has the same property as the usual I(1) process. Information about the economy at an initial point has no effect on the long run growth rate, and a permanent effect on the level. To see it, let's consider Beveridge-Nelson decomposition.

$$\begin{aligned}&\Delta\tilde{y}_t = \\ &\alpha_0 + \alpha_1 s_t^* + z_t \\ &= \alpha_0 + \alpha_1 \tfrac{1-q}{1-\lambda} + (\tfrac{1}{1-\lambda L}\upsilon_t + \tfrac{1}{\phi(L)}\epsilon_t),\end{aligned} \tag{16}$$

where $\phi(L) = 1-\phi_1 L-\phi_2 L^2-\cdots-\phi_r L^r$. There exist $\psi(L) = 1+\psi_1 L+\psi_2 L^2+\cdots$ and a white noise u_t such that $(\frac{1}{1-\lambda L}\upsilon_t + \frac{1}{\phi(L)}\epsilon_t) = \psi(L)u_t$. The u_t is a white noise process, because the innovation terms ϵ_t and υ_t are white noises and are independent of each other. I assume one summability condition $\sum_{j=0}^{\infty} j|\psi_j| < \infty$. Using the following identity $\psi(L) = \psi(1) + \Delta\beta(L)$, where $\Delta = 1 - L$, $\beta(L) = \sum_{j=0}^{\infty}\beta_j L^j$, $\beta_j = -(\psi_{j+1} + \psi_{j+2} + \cdots)$, the process of $\Delta\tilde{y}_t$ can be rewritten as

$$\Delta\tilde{y}_t = \delta + \psi(1)u_t + \eta_t - \eta_{t-1}, \tag{17}$$

where $\delta = \alpha_0 + \alpha_1\frac{1-q}{1-\lambda}$. The process of $\tilde{y}_t$ is obtained as

$$\tilde{y}_t = \delta \cdot t + \textstyle\sum_{i=1}^{t} u_i + \eta_t + (\tilde{y}_0 - \eta_0). \tag{18}$$

By the one summability assumption, $\beta(L)$ is absolutely summable. Therefore, the third term, η_t, is zero-mean covariance stationary. The first component, $\delta \cdot t$, is linear time trend. The second component, $\sum_{i=1}^{t} u_i$, is stochastic trend—that is, driftless random walk. The last component, $\tilde{y}_0 - \eta_0$, is the initial condition. Here, implicitly, I assume that u_t is a stationary martingale difference sequence such that $0 < E(u_t^2) < \infty$ and $\frac{1}{T}\sum_{t=1}^{T} E(u_t^2|u_{t-1}, u_{t-2}, \cdots)$ converges in probability to $E(u_t^2)$.[7]

Both components, n_t and $\tilde{z}_t$, contain stochastic trend, stationary process and the initial condition. The n_t term contains the linear time trend. While information about the economy at date 0 has no effect on the long run growth rate of $\tilde{y}_t$, it does exert a permanent effect on the level. The permanent effect of the information differences between the states comes only from the Markov trend term. Hence, when I compare certain knowledge about the initial

[7]Because the two innovation terms, ϵ_t and υ_t, are independent, I can analyze them separately. Since ϵ_t is i.i.d., the assumptions for the property of the above stationary martingale difference sequence hold for its component. As for υ_t, as t goes to infinity $E(\upsilon_t^2) = \pi p(1-p) + (1-\pi)q(1-q)$, where $\pi = \frac{1-q}{2-p-q}$, and the assumptions for υ_t also seem to hold.

states, I can focus on the effect of n_t. The effect of the difference between these knowledge sets is $E[\tilde{y}_t|s_0^*=1] - E[\tilde{y}_t|s_0^*=0] \longrightarrow \alpha_1\frac{\lambda}{1-\lambda}$ as $t \longrightarrow \infty$. $\frac{E_0 x_t|\pi_0=1}{E_0 x_t|\pi_0=0} \longrightarrow \frac{\kappa-(-1+p+q)}{\kappa-\exp(\alpha_1)(-1+p+q)}$ as $t \longrightarrow \infty$, where κ is a solution of $\kappa(q+p\exp(\alpha_1)-\kappa) = \exp(\alpha_1)(-1+p+q)$.

B Appendix: PMSE

The population prediction mean squared error is population variance plus squared bias of the forecast estimator. Let M denote a sample size for forecast evaluation. To simplify the notation, $E[.|y_{t+h}]$ denotes $E[.|y_{t+h},\mathcal{Y}_t;\theta]$, and $E[.]$ denotes $E[.|\mathcal{Y}_t;\theta]$ in this section. The PMSE conditional on the target variable of y_{t+h} is

$$\begin{aligned}&E[(y_{t+h}-\hat{y}_{t+h})^2|y_{t+h}]\\&=E[\{\hat{y}_{t+h}-E(\hat{y}_{t+h}|y_{t+h})\}^2|y_{t+h}]+[E(\hat{y}_{t+h}-y_{t+h}|y_{t+h})]^2\\&=E[\{\hat{y}_{t+h}-E(\hat{y}_{t+h})\}^2]+[E(\hat{y}_{t+h})-y_{t+h}]^2\\&=Var(\hat{y}_{t+h})+[Bias(\hat{y}_{t+h}|y_{t+h})]^2.\end{aligned} \tag{19}$$

The derivation from the second line to the third line in Equation (19) comes from the fact that the forecast of $\hat{y}_{t+h}$ is a function of the current and past data of $\mathcal{Y}_t$, and does not depend on the target variable. The evaluation method removes randomness of the target variable and accesses the PMSE from the forecasting errors only.

To obtain the sample PMSE corresponding to Equation (19), I calculate the sample prediction variance by

$$\begin{aligned}&\hat{Var}(\hat{y}_{t+h})\\&=\tfrac{1}{M}\textstyle\sum_{t=T}^{T+M-1}[\hat{y}_{t+h}-\tfrac{1}{M}\sum_{t=T}^{T+M-1}\hat{y}_{t+h}]^2.\end{aligned} \tag{20}$$

The sample prediction bias is obtained by

$$\begin{aligned}&B\hat{i}as(\hat{y}_{t+h}|y_{t+h})\\&=\tfrac{1}{M}\textstyle\sum_{t=T}^{T+M-1}(\hat{y}_{t+h}-y_{t+h}).\end{aligned} \tag{21}$$

C Appendix: Monte Carlo Simulation

A few notes on the Monte Carlo simulation are as follows. First, in the Monte Carlo procedure, I use the same random draws of standard normal and uniform distributions across different parameter settings. Given presumed parameter values, I obtain simulated data of the longest sample size once, and use a part of them in the analysis for a smaller sample size. In other words, across different total sample sizes of simulated data, simulated data at the overlapping dates are the same in an identical parameter setting.

Second, the first 499 simulated data are discarded. In other words, for estimation, forecasting, and the forecast evaluation, I use observations from the 500th simulated data.

Third, due to computational burden, for each combination of a sample size of estimation, a sample size of forecast evaluation, and parameter setting, I conduct Monte Carlo simulation once. Instead of many simulations, I let a time horizon be very long—that is, the sample size of forecast evaluation is large, in some simulation experiments.

In the following tables, 'Variance' is sample prediction variance of a forecast from Equation (20). 'Bias' is sample prediction bias of a forecast from Equation (21). 'Mean' is sample prediction mean of a forecast: $\frac{1}{M}\sum_{t=T}^{T+M-1}\hat{y}_{t+h}$. 'Difference in PMSE' denotes the percentage difference between the prediction mean squared error using the bagging and that in the benchmark: $100 \times \frac{PMSE_{bagging} - PMSE_{benchmark}}{PMSE_{benchmark}}$.

C.1 Hamilton Parameter Setting

Table 1: Monte Carlo Simulation

T=40, M=1000, r=4. Parameter values are the same as in Hamilton (1989).

	Benchmark	Bagging				
Type of Bootstrap		Parametric	Parametric	Residual	Residual	Residual
State probability for draws		smoothed	inferred	smoothed	smoothed	inferred
State probability for residuals				smoothed	inferred	inferred
Mean	0.97333	0.99554	0.99361	0.99264	0.99014	0.99966
Bias	-0.04650	-0.02430	-0.02622	-0.02720	-0.02969	-0.02018
Variance	1.90515	0.09806	0.22470	0.43796	0.41203	0.57132
PMSE	1.90731	0.09865	0.22539	0.43870	0.41291	0.57173
Difference in PMSE, %		-94.83	-88.18	-77.00	-78.35	-70.02

Table 2: Monte Carlo Simulation

T=150, M=1000, r=4. Parameter values are the same as in Hamilton (1989).

	Benchmark	Bagging				
Type of Bootstrap		Parametric	Parametric	Residual	Residual	Residual
State probability for draws		smoothed	inferred	smoothed	smoothed	inferred
State probability for residuals				smoothed	inferred	inferred
Mean	1.00589	1.02557	1.02134	1.02364	1.03167	1.03876
Bias	-0.01389	0.00580	0.00157	0.00386	0.01190	0.01899
Variance	1.65749	0.03942	0.35748	0.33684	0.33735	0.66216
PMSE	1.65768	0.03946	0.35748	0.33686	0.33749	0.66252
Difference in PMSE, %		-97.62	-78.44	-79.68	-79.64	-60.03

Table 3: Monte Carlo Simulation

Parametric Bootstrap with Normally Distributed Error

Smoothed state probability is used for random draws. T=150, M=100, r=4.

Parameter values except Markov transition probability of primitive states are the same as in Hamilton (1989).

Markov transition	p	0.8	0.8	0.8	0.8	0.6	0.4	0.2	0.6	0.6
probability	q	0.8	0.6	0.4	0.2	0.8	0.8	0.8	0.6	0.4
Mean	Benchmark	0.9013	0.9090	0.9054	0.8944	0.5919	0.2913	-0.0034	0.5891	0.5967
	Bagging	0.9320	0.9286	0.9345	0.9331	0.5790	0.2919	-0.0239	0.5599	0.5470
Bias	Benchmark	0.0284	0.0362	0.0325	0.0215	0.0083	0.0122	0.0372	0.0055	0.0130
	Bagging	0.0591	0.0557	0.0616	0.0602	-0.0046	0.0128	0.0167	-0.0238	-0.0366
Variance	Benchmark	2.4813	2.5130	2.5222	2.5638	3.2080	2.9963	1.5979	3.2144	3.2057
	Bagging	0.0859	0.0911	0.0878	0.0905	0.0657	0.1585	0.0885	0.0738	0.0781
PMSE	Benchmark	2.4821	2.5143	2.5232	2.5643	3.2081	2.9964	1.5993	3.2144	3.2058
	Bagging	0.0894	0.0942	0.0916	0.0941	0.0657	0.1587	0.0888	0.0743	0.0795
Percent difference in PMSE		-96.40	-96.26	-96.37	-96.33	-97.95	-94.70	-94.45	-97.69	-97.52

0.6	0.4	0.2	0.5	0.4	0.4	0.2	0.3	0.2
0.2	0.6	0.6	0.5	0.4	0.2	0.4	0.3	0.2
0.6467	0.2913	0.0470	0.4438	0.3098	0.5948	0.2247	0.4089	0.2895
0.6466	0.2885	-0.0234	0.4197	0.3260	0.5689	0.2049	0.4190	0.2460
0.0478	0.0122	0.0571	0.0276	0.0154	0.0416	-0.0240	0.0079	-0.0201
0.0478	0.0093	-0.0133	0.0035	0.0317	0.0157	-0.0438	0.0181	-0.0636
3.3533	2.9963	1.7444	3.4526	2.9967	2.1525	1.7698	1.6513	2.6947
0.1067	0.1603	0.4679	0.0748	0.2164	0.1370	0.7224	0.6031	1.7945
3.3556	2.9964	1.7476	3.4534	2.9969	2.1543	1.7704	1.6514	2.6951
0.1090	0.1604	0.4680	0.0748	0.2174	0.1373	0.7243	0.6034	1.7986
-96.75	-94.65	-73.22	-97.83	-92.75	-93.63	-59.09	-63.46	-33.27

C.2 Base Parameter Setting

Table 4: Monte Carlo Simulation

T=40, M=1000, r=1. Parameter values are the same as in the base parameter set.

	Benchmark	Bagging				
Type of Bootstrap		Parametric	Parametric	Residual	Residual	Residual
State probability for draws		smoothed	inferred	smoothed	smoothed	inferred
State probability for residuals				smoothed	inferred	inferred
Mean	1.00206	0.95953	0.84674	0.96408	1.07841	0.96204
Bias	0.03957	-0.00296	-0.11575	0.00159	0.11593	-0.00044
Variance	0.10253	0.01660	0.04021	0.01494	0.01769	0.02446
PMSE	0.10410	0.01660	0.05361	0.01494	0.03113	0.02446
Difference in PMSE, %		-84.05	-48.50	-85.65	-70.10	-76.50

Table 5: Monte Carlo Simulation

T=150, M=1000, r=1. Parameter values are the same as in the base parameter set.

	Benchmark	Bagging				
Type of Bootstrap		Parametric	Parametric	Residual	Residual	Residual
State probability for draws		smoothed	inferred	smoothed	smoothed	inferred
State probability for residuals				smoothed	inferred	inferred
Mean	0.99503	0.95755	0.85249	0.96150	1.06749	0.95948
Bias	0.03601	-0.00146	-0.10653	0.00248	0.10848	0.00047
Variance	0.08557	0.00767	0.02710	0.00433	0.00499	0.01328
PMSE	0.08687	0.00767	0.03845	0.00434	0.01676	0.01328
Difference in PMSE, %		-91.17	-55.73	-95.01	-80.71	-84.71

Table 6: Monte Carlo Simulation

T=500, M=1000, r=1. Parameter values are the same as in the base parameter set.

	Benchmark	Bagging				
Type of Bootstrap		Parametric	Parametric	Residual	Residual	Residual
State probability for draws		smoothed	inferred	smoothed	smoothed	inferred
State probability for residuals				smoothed	inferred	inferred
Mean	0.98926	0.95265	0.84617	0.95525	1.06519	0.95389
Bias	0.03412	-0.00250	-0.10897	0.00010	0.11004	-0.00126
Variance	0.07954	0.00577	0.02506	0.00242	0.00125	0.01155
PMSE	0.08070	0.00577	0.03694	0.00242	0.01336	0.01155
Difference in PMSE, %		-92.84	-54.23	-97.00	-83.45	-85.69

Table 7: Monte Carlo Simulation

Parametric Bootstrap with Normally Distributed Error

Smoothed state probability is used for random draws. T=150, M=100, r=1.

Parameter values except standard diviation of innovation terms are the same as the base parameter set.

Standard deviation		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Mean	Benchmark	0.96479	0.96729	0.97496	0.98353	0.99491	1.00803	1.01359	0.98411	0.96952
	Bagging	0.96515	0.96741	0.97109	0.97378	0.97624	0.97531	0.97402	0.97256	0.97945
Bias	Benchmark	0.00399	0.00570	0.01258	0.02035	0.03094	0.04326	0.04803	0.01775	0.00236
	Bagging	0.00435	0.00582	0.00871	0.01060	0.01226	0.01054	0.00846	0.00620	0.01229
Variance	Benchmark	0.01156	0.01567	0.02078	0.03782	0.04743	0.05985	0.07250	0.09688	0.14911
	Bagging	0.01114	0.01156	0.01162	0.01122	0.00730	0.00396	0.00277	0.01859	0.03744
PMSE	Benchmark	0.01158	0.01570	0.02094	0.03823	0.04839	0.06172	0.07481	0.09720	0.14911
	Bagging	0.01116	0.01159	0.01170	0.01134	0.00745	0.00407	0.00284	0.01863	0.03759
Percent difference in PMSE		-3.66	-26.15	-44.15	-70.35	-84.61	-93.40	-96.20	-80.83	-74.79

		1	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8
		0.96817	0.96954	0.97123	0.97353	0.97589	0.97969	0.98173	0.97953	0.98022
		0.98274	0.98556	0.98727	0.99122	0.99563	1.00224	1.00036	1.00757	1.00988
		0.00022	0.00080	0.00169	0.00320	0.00476	0.00777	0.00902	0.00602	0.00591
		0.01480	0.01682	0.01773	0.02088	0.02450	0.03031	0.02764	0.03406	0.03558
		0.21215	0.28113	0.35383	0.43160	0.51767	0.61719	0.72927	0.86473	0.99604
		0.04556	0.04803	0.05076	0.05913	0.07774	0.09722	0.12422	0.15037	0.18083
		0.21215	0.28113	0.35383	0.43161	0.51769	0.61725	0.72935	0.86477	0.99607
		0.04578	0.04831	0.05107	0.05957	0.07834	0.09814	0.12499	0.15153	0.18209
		-78.42	-82.82	-85.57	-86.20	-84.87	-84.10	-82.86	-82.48	-81.72

Table 8: Monte Carlo Simulation

Parametric Bootstrap with Normally Distributed Error

Smoothed state probability is used for random draws. T=150, M=100, r=1.

Parameter values except the coefficient of the first lagged dependent variable are the same as the base parameter set.

Coefficient of first lagged dependent variable		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Mean	Benchmark	1.01359	1.01391	1.01470	1.01465	1.01328	1.00941	1.00019	0.97755	0.91712
	Bagging	0.97402	0.97230	0.97309	0.97676	0.98195	0.98832	1.00382	1.03180	1.10991
Bias	Benchmark	0.04803	0.04908	0.05103	0.05268	0.05380	0.05383	0.05174	0.04591	0.04013
	Bagging	0.00846	0.00747	0.00941	0.01479	0.02247	0.03274	0.05537	0.10017	0.23291
Variance	Benchmark	0.07250	0.10157	0.14775	0.22327	0.34661	0.55203	0.90801	1.54245	2.54553
	Bagging	0.00277	0.00378	0.00523	0.00755	0.01161	0.01835	0.03048	0.05942	0.14045
PMSE	Benchmark	0.07481	0.10397	0.15035	0.22604	0.34951	0.55493	0.91068	1.54456	2.54714
	Bagging	0.00284	0.00383	0.00532	0.00777	0.01212	0.01942	0.03354	0.06945	0.19470
Percent difference in PMSE		-96.20	-96.31	-96.46	-96.56	-96.53	-96.50	-96.32	-95.50	-92.36

Table 9: Monte Carlo Simulation

Parametric Bootstrap with Normally Distributed Error

Smoothed state probability is used for random draws. T=150, M=100, r=1.

Parameter values except Markov transition probability of primitive states are the same as in the base parameter set.

Markov transition probability										
	p	0.8	0.8	0.8	0.8	0.6	0.4	0.2	0.6	0.6
	q	0.8	0.6	0.4	0.2	0.8	0.8	0.8	0.6	0.4
Mean	Benchmark	0.9158	0.9158	0.9158	0.9158	0.5620	0.2569	-0.0413	0.5620	0.5620
	Bagging	0.8855	0.8862	0.8868	0.8868	0.5712	0.2461	-0.0698	0.5699	0.5695
Bias	Benchmark	0.0462	0.0462	0.0462	0.0462	-0.0036	0.0113	0.0491	-0.0036	-0.0036
	Bagging	0.0160	0.0167	0.0172	0.0173	0.0056	0.0005	0.0206	0.0043	0.0040
Variance	Benchmark	0.0621	0.0621	0.0621	0.0621	0.1502	0.1164	0.0969	0.1502	0.1502
	Bagging	0.0089	0.0088	0.0089	0.0089	0.0687	0.0555	0.0080	0.0696	0.0694
PMSE	Benchmark	0.0642	0.0642	0.0642	0.0642	0.1502	0.1166	0.0993	0.1502	0.1502
	Bagging	0.0092	0.0091	0.0092	0.0092	0.0687	0.0555	0.0084	0.0696	0.0694
Percent difference in PMSE		-85.73	-85.81	-85.60	-85.61	-54.24	-52.36	-91.54	-53.65	-53.76

Markov transition probability								
	p	0.6	0.4	0.2	0.4	0.4	0.2	0.2
	q	0.2	0.6	0.6	0.4	0.2	0.4	0.2
Mean	Benchmark	0.5916	0.2569	-0.0069	0.2775	0.5875	0.2339	0.3315
	Bagging	0.5976	0.2464	-0.0433	0.2638	0.5480	0.2308	0.3129
Bias	Benchmark	0.0101	0.0113	0.0515	0.0159	0.0539	0.0203	0.0539
	Bagging	0.0160	0.0008	0.0151	0.0022	0.0145	0.0173	0.0353
Variance	Benchmark	0.1102	0.1164	0.1100	0.1010	0.0772	0.2958	0.3862
	Bagging	0.0415	0.0553	0.0137	0.0452	0.0116	0.1522	0.2360
PMSE	Benchmark	0.1103	0.1166	0.1127	0.1012	0.0801	0.2962	0.3891
	Bagging	0.0418	0.0553	0.0140	0.0453	0.0119	0.1525	0.2372
Percent difference in PMSE		-62.15	-52.59	-87.62	-55.29	-85.20	-48.51	-39.03

Table 10: Monte Carlo Simulation

Parameter values except Markov transition probability of primitive states are the same as the base parameter set.
T=150, M=100, r=1.

		Markov transition probability of primitive states									
		p=0.6, q=0.8					p=0.8,q=0.6				
Type of Bootstrap		Parametric	Parametric	Residual	Residual	Residual	Parametric	Parametric	Residual	Residual	Residual
state probability for draws		smoothed	inferred	smoothed	smoothed	inferred	smoothed	inferred	smoothed	smoothed	inferred
primitive state probability for residuals				smoothed	inferred	inferred			smoothed	inferred	inferred
Mean	Benchmark	0.5620	0.5620	0.5620	0.5620	0.5620	0.9158	0.9158	0.9158	0.9158	0.9158
	Bagging	0.5712	0.5949	0.5692	0.5468	0.5712	0.8862	0.8178	0.8890	0.9612	0.8919
Bias	Benchmark	-0.0036	-0.0036	-0.0036	-0.0036	-0.0036	0.0462	0.0462	0.0462	0.0462	0.0462
	Bagging	0.0056	0.0293	0.0037	-0.0188	0.0056	0.0167	-0.0518	0.0194	0.0916	0.0224
Variance	Benchmark	0.1502	0.1502	0.1502	0.1502	0.1502	0.0621	0.0621	0.0621	0.0621	0.0621
	Bagging	0.0687	0.0713	0.0673	0.0666	0.0686	0.0088	0.0220	0.0042	0.0031	0.0109
PMSE	Benchmark	0.1502	0.1502	0.1502	0.1502	0.1502	0.0642	0.0642	0.0642	0.0642	0.0642
	Bagging	0.0687	0.0721	0.0673	0.0669	0.0686	0.0091	0.0247	0.0046	0.0115	0.0114
Percent difference in PMSE		-54.24	-51.97	-55.19	-55.45	-54.29	-85.81	-61.59	-92.81	-82.08	-82.20

Table 11: Monte Carlo Simulation

Parametric Bootstrap with Normally Distributed Error
Standard deviation of innovation terms is 0.5.
Smoothed state probability is used for random draws. T=150, M=100, r=1.
Parameter values except standard deviation of innovation terms and Markov transition probability of primitive states are the same as in the base parameter set.

Markov transition	p	0.8	0.8	0.8	0.8	0.6	0.4	0.2	0.6	0.6
probability	q	0.8	0.6	0.4	0.2	0.8	0.8	0.8	0.6	0.4
Mean	Benchmark	0.9152	0.9152	0.9152	0.9152	0.5628	0.2534	-0.0517	0.5628	0.5628
	Bagging	0.8941	0.8941	0.8941	0.8941	0.5607	0.2441	-0.0701	0.5607	0.5607
Bias	Benchmark	0.0473	0.0473	0.0473	0.0473	-0.0012	0.0095	0.0403	-0.0012	-0.0012
	Bagging	0.0261	0.0262	0.0261	0.0261	-0.0033	0.0001	0.0220	-0.0033	-0.0033
Variance	Benchmark	0.0438	0.0438	0.0438	0.0438	0.0971	0.0747	0.0634	0.0971	0.0971
	Bagging	0.0043	0.0043	0.0043	0.0043	0.0540	0.0363	0.0083	0.0540	0.0540
PMSE	Benchmark	0.0460	0.0460	0.0460	0.0460	0.0971	0.0748	0.0650	0.0971	0.0971
	Bagging	0.0050	0.0050	0.0050	0.0050	0.0540	0.0363	0.0088	0.0540	0.0540
Percent difference in PMSE		-89.23	-89.23	-89.23	-89.23	-44.36	-51.42	-86.43	-44.36	-44.36

0.6	0.4	0.2	0.4	0.4	0.2	0.2
0.2	0.6	0.6	0.4	0.2	0.4	0.2
0.5962	0.2534	-0.0237	0.2736	0.5750	0.2306	0.3237
0.5907	0.2441	-0.0426	0.2632	0.5517	0.2314	0.3173
0.0162	0.0095	0.0364	0.0137	0.0430	0.0186	0.0477
0.0107	0.0001	0.0174	0.0032	0.0198	0.0195	0.0413
0.0570	0.0747	0.0679	0.0609	0.1365	0.2670	0.3702
0.0203	0.0363	0.0142	0.0249	0.0694	0.1982	0.2998
0.0572	0.0748	0.0692	0.0611	0.1384	0.2674	0.3725
0.0204	0.0363	0.0145	0.0249	0.0698	0.1986	0.3016
-64.37	-51.42	-79.00	-59.23	-49.54	-25.72	-19.05

Table 12: Monte Carlo Simulation

Parametric Bootstrap with Normally Distributed Error
Standard deviation of innovation terms is 0.9.
Smoothed state probability is used for random draws. T=150, M=100, r=1.

Parameter values except standard deviation of innovation terms and Markov transition probability of primitive states are the same as in the base parameter set.

Markov transition	p	0.8	0.8	0.8	0.8	0.6	0.4	0.2	0.6	0.6	0.6	0.4	0.2	0.4	0.4	0.2	0.2
probability	q	0.8	0.6	0.4	0.2	0.8	0.8	0.8	0.6	0.4	0.2	0.6	0.6	0.4	0.2	0.4	0.2
Mean	Benchmark	0.8964	0.8964	0.8964	0.8964	0.5615	0.2559	-0.0662	0.5615	0.5615	0.5883	0.2559	-0.0357	0.2770	0.5695	0.2214	0.3114
	Bagging	0.8904	0.8910	0.8909	0.8909	0.5793	0.2525	-0.0654	0.5781	0.5795	0.6034	0.2512	-0.0406	0.2683	0.5528	0.2288	0.3074
Bias	Benchmark	0.0252	0.0252	0.0252	0.0252	-0.0057	0.0088	0.0227	-0.0057	-0.0057	0.0051	0.0088	0.0212	0.0138	0.0343	0.0062	0.0323
	Bagging	0.0192	0.0198	0.0198	0.0197	0.0121	0.0054	0.0234	0.0110	0.0123	0.0202	0.0041	0.0163	0.0052	0.0177	0.0136	0.0283
Variance	Benchmark	0.1183	0.1183	0.1183	0.1183	0.2355	0.2037	0.1560	0.2355	0.2355	0.2008	0.2037	0.1747	0.1916	0.0632	0.2839	0.3751
	Bagging	0.0464	0.0459	0.0459	0.0459	0.0996	0.0861	0.0091	0.1023	0.0952	0.0748	0.0864	0.0132	0.0790	0.0129	0.1000	0.1711
PMSE	Benchmark	0.1189	0.1189	0.1189	0.1189	0.2355	0.2038	0.1565	0.2355	0.2355	0.2008	0.2038	0.1752	0.1918	0.0643	0.2840	0.3762
	Bagging	0.0468	0.0463	0.0463	0.0463	0.0997	0.0861	0.0096	0.1024	0.0953	0.0752	0.0864	0.0135	0.0790	0.0132	0.1002	0.1719
Percent difference in PMSE		-60.66	-61.11	-61.05	-61.07	-57.66	-57.76	-93.86	-56.52	-59.52	-62.53	-57.60	-92.31	-58.81	-79.53	-64.72	-54.30

Table 13: Monte Carlo Simulation

Parametric Bootstrap with Normally Distributed Error

Smoothed state probability is used for random draws. T=35, M=100, r=1.

Parameter values except Markov transition probability of primitive states are the same as in the base parameter set.

Markov transition probability																
	p	0.8	0.8	0.8	0.8	0.6	0.4	0.2	0.6	0.6	0.6	0.4	0.2	0.4	0.4	0.2	0.2
	q	0.8	0.6	0.4	0.2	0.8	0.8	0.8	0.6	0.4	0.2	0.6	0.6	0.4	0.2	0.4	0.2
Mean	Benchmark	0.9611	0.9607	0.9607	0.9607	0.5874	0.2861	-0.0281	0.5874	0.5874	0.6300	0.2861	0.0232	0.2865	0.6421	0.2384	0.3738
	Bagging	0.9146	0.9156	0.9183	0.9188	0.5784	0.2947	-0.0455	0.5776	0.5787	0.6234	0.2941	0.0000	0.2946	0.5964	0.2381	0.3448
Bias	Benchmark	0.0074	0.0069	0.0069	0.0069	-0.0304	-0.0436	-0.0378	-0.0304	-0.0304	-0.0358	-0.0436	-0.0185	-0.0433	0.0243	-0.0594	-0.0040
	Bagging	-0.0391	-0.0381	-0.0355	-0.0349	-0.0394	-0.0351	-0.0552	-0.0401	-0.0391	-0.0423	-0.0357	-0.0417	-0.0351	-0.0214	-0.0597	-0.0330
Variance	Benchmark	0.0738	0.0741	0.0741	0.0741	0.1765	0.1155	0.0978	0.1765	0.1765	0.1140	0.1155	0.1273	0.1152	0.0847	0.3005	0.4409
	Bagging	0.0179	0.0171	0.0168	0.0169	0.0886	0.0514	0.0243	0.0867	0.0864	0.0433	0.0529	0.0476	0.0527	0.0283	0.1892	0.3238
PMSE	Benchmark	0.0739	0.0741	0.0741	0.0741	0.1774	0.1174	0.0992	0.1774	0.1774	0.1152	0.1174	0.1276	0.1171	0.0853	0.3040	0.4409
	Bagging	0.0195	0.0186	0.0180	0.0181	0.0901	0.0526	0.0274	0.0883	0.0879	0.0451	0.0541	0.0493	0.0540	0.0288	0.1928	0.3249
Percent difference in PMSE		-73.65	-74.93	-75.69	-75.57	-49.20	-55.22	-72.43	-50.24	-50.46	-60.86	-53.91	-61.34	-53.92	-66.26	-36.60	-26.32

D Appendix: Application

Table 14: Real GNP in the U.S.

T (sample size of estimation) =45, M (sample size of forecast evaluation) =90, r (# lags) =4.

	Benchmark	Bagging				
Type of Bootstrap		Parametric	Parametric	Residual	Residual	Residual
State probability for draws		smoothed	inferred	smoothed	smoothed	inferred
State probability for residuals				smoothed	inferred	inferred
Mean	0.67980	0.75803	0.74745	0.73797	0.75698	0.74212
Bias	-0.06623	0.01200	0.00143	-0.00806	0.01095	-0.00391
Variance	0.58287	0.12983	0.13386	0.16172	0.14801	0.14678
PMSE	0.58726	0.12997	0.13386	0.16179	0.14813	0.14679
Difference in PMSE, %		-77.87	-77.21	-72.45	-74.78	-75.00

Table 15: Real GNP in the U.S.

T (sample size of estimation) =55, M (sample size of forecast evaluation) =80, r (# lags) =4.

	Benchmark	Bagging				
Type of Bootstrap		Parametric	Parametric	Residual	Residual	Residual
State probability for draws		smoothed	inferred	smoothed	smoothed	inferred
State probability for residuals				smoothed	inferred	inferred
Mean	0.78937	0.79537	0.81182	0.81094	0.78056	0.81182
Bias	0.08034	0.08635	0.10280	0.10192	0.07153	0.10280
Variance	0.32766	0.07483	0.08242	0.08353	0.08296	0.08242
PMSE	0.33411	0.08229	0.09299	0.09391	0.08807	0.09299
Difference in PMSE, %		-75.37	-72.17	-71.89	-73.64	-72.17

Table 16: Real GNP in the U.S.

T (sample size of estimation) =65, M (sample size of forecast evaluation) =70, r (# lags) =4.

	Benchmark	Bagging				
Type of Bootstrap		Parametric	Parametric	Residual	Residual	Residual
State probability for draws		smoothed	inferred	smoothed	smoothed	inferred
State probability for residuals				smoothed	inferred	inferred
Mean	0.72385	0.75869	0.77920	0.75370	0.73408	0.72989
Bias	0.08960	0.12445	0.14496	0.11945	0.09983	0.09564
Variance	0.17170	0.06269	0.04530	0.06836	0.08738	0.04970
PMSE	0.17973	0.07818	0.06631	0.08262	0.09735	0.05885
Difference in PMSE, %		-56.50	-63.11	-54.03	-45.84	-67.26

Table 17: Real GNP in the U.S.

T (sample size of estimation) =75, M (sample size of forecast evaluation) =60, r (# lags) =4.

	Benchmark	Bagging				
Type of Bootstrap		Parametric	Parametric	Residual	Residual	Residual
State probability for draws		smoothed	inferred	smoothed	smoothed	inferred
State probability for residuals				smoothed	inferred	inferred
Mean	0.71622	0.71737	0.76592	0.71105	0.69325	0.74687
Bias	0.09622	0.09738	0.14592	0.09105	0.07325	0.12687
Variance	0.13394	0.05238	0.05322	0.06364	0.06490	0.04660
PMSE	0.14319	0.06186	0.07452	0.07193	0.07027	0.06270
Difference in PMSE, %		-56.80	-47.96	-49.77	-50.93	-56.21

Table 18: Real GNP in the U.S.

T (sample size of estimation) =85, M (sample size of forecast evaluation) =50, r (# lags) =4.

	Benchmark	Bagging				
Type of Bootstrap		Parametric	Parametric	Residual	Residual	Residual
State probability for draws		smoothed	inferred	smoothed	smoothed	inferred
State probability for residuals				smoothed	inferred	inferred
Mean	0.73161	0.74922	0.78627	0.73870	0.71612	0.75162
Bias	0.12444	0.14205	0.17910	0.13153	0.10895	0.14445
Variance	0.15789	0.07382	0.03621	0.07660	0.08826	0.04156
PMSE	0.17338	0.09400	0.06829	0.09390	0.10013	0.06243
Difference in PMSE, %		-45.78	-60.61	-45.84	-42.25	-63.99