AARHUS UNIVERSITY
BUSINESS AND SOCIAL SCIENCES
DEPARTMENT OF ECONOMICS AND BUSINESS

CREATES
Center for Research in Econometric Analysis of Time Series

Bagging Weak Predictors

Manuel Lukas and Eric Hillebrand

CREATES Research Paper 2014-1

Department of Economics and Business
Aarhus University
Fuglesangs Allé 4
DK-8210 Aarhus V
Denmark

Email: oekonomi@au.dk
Tel: +45 8716 5515

Bagging Weak Predictors*

Manuel Lukas† and Eric Hillebrand

Aarhus University, Department of Economics and Business,
and CREATES‡

February 17, 2014

Abstract

Relations between economic variables can often not be exploited for forecasting, suggesting that predictors are weak in the sense that estimation uncertainty is larger than bias from ignoring the relation. In this paper, we propose a novel bagging predictor designed for such weak predictor variables. The predictor is based on an in-sample test for predictive ability. Our predictor shrinks the OLS estimate not to zero, but towards the null of the test which equates squared bias with estimation variance. We derive the asymptotic distribution and show that the predictor can substantially lower the MSE compared to standard *t*-test bagging. An asymptotic shrinkage representation for the predictor is provided that simplifies computation of the estimator. Monte Carlo simulations show that the predictor works well in small samples. In the empirical application, we find that the new predictor works well for inflation forecasts.

Keywords: Inflation forecasting, bootstrap aggregation, estimation uncertainty, weak predictors.
JEL classification: C32, E37.

*The authors wish to thank Atsushi Inoue and Michael McCracken for helpful comments and suggestions.
†Corresponding author: mlukas@creates.au.dk. Address: Fuglesangs Allé 4, 8210 Aarhus V, Denmark.
‡CREATES (Center for Research in Econometric Analysis of Time Series) is funded by the Danish National Research Foundation under grant number DNRF78.

1 Introduction

A frequent finding in pseudo out-of-sample forecasting exercises is that including predictor variables does not improve forecasting performance, even though the predictor variables are significant in in-sample regressions. For example, there is a large literature on forecast failure with economic predictor variables for forecasting inflation (see, e.g., Atkeson and Ohanian, 2001; Stock and Watson, 2009) and forecasting exchange rates (see, e.g, Meese and Rogoff, 1983; Cheung et al., 2005). Including predictor variables suggested by economic theory, or selected by in-sample regressions, typically does not help to consistently out-perform simple time series models across different sample splits and model specifications. Forecasting failure can be attributed to estimation variance and parameter instability. In this paper, we focus exclusively on the former. These two causes of forecast failure are, however, often interrelated in practice. If we are unwilling to specify the nature of instability, it is common practice to use a short rolling window for estimation to deal with parameter instability. While a short estimation window can better adapt to changing parameters, it increases estimation variance compared to using all data. In this sense, estimation variance can result from the attempt to accommodate parameter instability, such that our results are relevant for both kinds of forecast failure.

This paper is concerned with reducing estimation variance by bagging pre-test estimators when predictor variables have weak forecasting power. Modeling weak predictors in the framework of Clark and McCracken (2012) (CM henceforth) considers a non-vanishing bias-variance trade-off. CM propose an in-sample test for predictive ability, i.e., a test of whether bias reduction or estimation variance will prevail when including a predictor variable. Based on this test, we propose a novel bagging estimator that is designed to work well for predictors with non-zero coefficient of known sign. Under the null of the CM test, the parameter is not equal to zero, but equal to a value for which squared bias from omitting the predictor variable is equal to estimation variance. In our bagging scheme, we set the parameter equal to this value instead of zero whenever we fail to reject the null. For this, knowledge of the coefficient's sign is necessary. We derive the asymptotic distribution of the estimator and show that for a wide range of parameter values, asymptotic mean-squared error is superior to bagging a standard t-test. The improvements can be substantial and are not sensitive to the choice of the critical value, which is a remaining tuning parameter. We obtain forecast improvements if the data-generating parameter is small but non-zero. If the data-generating parameter is indeed zero, however, our estimator has a large bias and is therefore imprecise.

Bootstrap aggregation, *bagging*, was proposed by Breiman (1996) as a method to improve forecast accuracy by smoothing instabilities from modeling strategies that involve hard-thresholding and pre-testing. With bagging, the modeling strategy is applied repeatedly to bootstrap samples of the data, and the final prediction is obtained by averaging over the predictions from the bootstrap samples. Bühlmann and Yu (2002) show theoretically how bagging reduces variance of predictions and can thus lead to improved accuracy. Stock and Watson (2012) derive a shrinkage representation for bagging a hard-threshold variable selection based on the t-statistic. This representation shows that standard t-test bagging is asymptotically equivalent to shrinking the unconstrained coefficient estimate to zero. The degree of shrinkage depends on the value of the t-statistic.

Bagging is becoming a standard forecasting technique for economic and financial variables. Inoue and Kilian (2008) consider different bagging strategies for forecasting US inflation with many predictors, including bagging a factor model where factors are included if they are significant in a preliminary regression. They find that forecasting performance is similar to other forecasting methods, such as shrinkage methods and forecast combination. Rapach and Strauss (2010) use bagging to forecast US unemployment changes with 30 predictors. They apply bagging to a pre-test strategy that uses individual t-statistics to select variables, and find that this delivers very competitive forecasts compared to forecast combinations of univariate benchmarks. Hillebrand and Medeiros (2010) apply bagging to lag selection for heterogeneous autoregressive models of realized volatility, and they find that this method leads to improvements in forecast accuracy.

Our method requires a sign restriction in order to impose the null. We focus on a single predictor variable, because in this case, intuition and economic theory can be used to derive sign restrictions. For models with multiple correlated predictors, sign restrictions are harder to justify. In the literature, bagging has been applied for reducing variance from imposing sign restrictions on parameters. A hard-threshold estimator with sign restriction sets the estimate to zero if the sign restriction is violated. Gordon and Hall (2009) consider bagging the hard-threshold estimator and show analytically that bagging can reduce variance. Sign restrictions arise naturally in predicting the equity premium, see Campbell and Thompson (2008) for a hard-threshold, and Pettenuzzo et al. (2013) for a Bayesian approach. Hillebrand et al. (2013) analyze the bias-variance trade-off from bagging positive constraints on coefficients and the equity premium forecast itself, and they find empirically that bagging helps improving the forecasting performance.

The remainder of the paper is organized as follows. In Section 2, the bagging estimator for weak predictors is presented and asymptotic properties are analyzed. Monte Carlo results for small samples are presented in Section 3. In Section 4, the estimator is applied to Phillips curve-based inflation forecasting. Concluding remarks are given in Section 5.

2 Bagging Predictors

Let y be the target variable we wish to forecast h-steps ahead, for example inflation. Let T be the sample size. At time t, we forecast $y_{t+h,T}$ using the scalar variable $x_{t,T}$ as predictor and a model estimated on the available data. We model $x_{t,T}$ as a weak predictor that may or may not improve forecasting accuracy,

$$y_{t+h,T} = \mu + (T^{-1/2}b)x_{t,T} + u_{t+h,T}, \tag{1}$$

where μ is an intercept. We assume that the sign of b is known. Without loss of generality, we assume that b is strictly positive, i.e., $sign(b) = 1$. Let $\beta_T = T^{-1/2}b$. We require that the model (1) satisfies the following assumption.

Assumption 1 (Assumption 3 in Clark and McCracken (2012))
Let $U_{t,T} = (x_{t,T}u_{t+h,T}, x_{t,T}^2)$. *(a)* $T^{-1}\sum_{t=1}^{[rT]} U_{t,T}U_{t-l,T}' \Rightarrow r\Omega_l$, *where* $\Omega_l = \lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} \mathbb{E}[U_{t,T}U_{t-l,T}]$ *for all* $l \geq 0$ *and (b)* $\omega_{11}(l) = 0$ *for all* $l \geq h$, *where* $\omega_{11}(l)$ *is the top-left element of* Ω_l. *(c)* $\sup_{T\geq 1, s\leq T} \mathbb{E}[|U_{s,T}|^{2q}] < \infty$ *for some* $q > 1$. *(d)* $U_{t,T} - \mathbb{E}[U_{t,T}]$ *is a zero mean triangular array satisfying Theorem 1 of de Jong (1997).*

In (a) of Assumption 1 we require asymptotic mean square stationarity. In (b) we require the errors u_{t+h} to follow an MA$(h-1)$ process, which accounts for the overlapping nature of errors when forecasting multiple steps ahead. Finite second moments are ensured by (c), and (d) provides a central limit theorem (CLT).

For a given sample of length T and a given forecast horizon h, we consider two forecasting models, the unrestricted model (UR) that includes the predictor variable x_t, and the restricted model (RE) that contains only an intercept. Let $\hat{\mu}_T^{RE}$ and $(\hat{\mu}_T^{UR}, \hat{\beta}_T)'$ be the OLS parameter estimates from the restricted model and the unrestricted model, respectively. The forecasts for y_{t+h} from the unrestricted and restricted models are denoted

$$\hat{y}_{t+h,T}^{UR} = \hat{\mu}_T^{UR} + \hat{\beta}_T x_{t,T}, \tag{2}$$

and

$$\hat{y}^{RE}_{t+h,T} = \hat{\mu}^{RE}_T, \tag{3}$$

respectively.

In practice, we are often not certain whether to include the weak predictor x_t in the forecast model or not, i.e., whether RE or UR yields more accurate forecasts. In such a situation, it is common to use a pre-test estimator. Typically, the t-statistic $\hat{\tau}_T = T^{1/2}\hat{\beta}_T\hat{\sigma}^{-1}_{\infty,T}$ is used to decide whether or not to include the variable. Here, $\hat{\sigma}^2_{\infty,T}$ is a consistent estimator of the asymptotic variance of $\hat{\beta}_T$, $\sigma^2_{\infty,T} < \infty$. Let $\mathbf{I}(.)$ denote the indicator function that takes value 1 if the argument is true and 0 otherwise. The one-sided pre-test estimator is

$$\hat{\beta}^{PT}_T = \hat{\beta}_T \mathbf{I}(\hat{\tau}_T > c), \tag{4}$$

for some critical value c, for example 1.64 for a one-sided test at the 5% level. We focus on one-sided testing because we assume that the sign of β is known.

The hard-threshold indicator function involved in the pre-test estimator introduces estimation uncertainty, and it is not well designed to improve forecasting performance. Bootstrap aggregation (bagging) can be used to smooth the hard-threshold and thereby improve forecasting performance (see Bühlmann and Yu, 2002; Breiman, 1996). The bagging version of the pre-test estimator is defined as

$$\hat{\beta}^{BG}_T = \frac{1}{B}\sum_{b=1}^{B} \hat{\beta}^*_b \mathbf{I}(\hat{\tau}^*_b > c), \tag{5}$$

where $\hat{\beta}^*_b$ and $\hat{\tau}^*_b$ are calculated from bootstrap samples, and B is the number of bootstrap replications.

The bagging estimator and the underlying t-statistic pre-test estimator are based on a test for $\beta = 0$. We use the estimated value if this null hypothesis can be rejected at some pre-specified significance level, e.g., 5%. However, this test does not directly address the actual decision problem whether or not including x_t improves the predictive accuracy for the given sample size. Rather, it is a test for whether the coefficient is zero or not.

Clark and McCracken (2012) (CM henceforth) propose an asymptotic in-sample test for predictive ability for weak predictors to address this problem. The null hypothesis is

$$H_{0,CM} : \lim_{T\to\infty} T\mathbb{E}[(y_{t+h,T} - \hat{y}^{RE}_{t+h,T})^2] = \lim_{T\to\infty} T\mathbb{E}[(y_{t+h,T} - \hat{y}^{UR}_{t+h,T})^2], \tag{6}$$

i.e, that the predictive accuracies of the restricted and the unrestricted model are asymptotically equal as measured by mean-squared error.

Clark and McCracken (2012) show that for the data-generating process equal to model (1) and under Assumption 1, the asymptotic distribution under the null (6) is:

$$\hat{\tau}_T \to_d \mathcal{N}(sign(b),1). \tag{7}$$

As we have assumed $sign(b) = 1$, the distribution converges to a normal distribution with mean and variance equal to 1. The asymptotic distribution is non-central, because under the null the coefficient is not zero. The critical values are different than for the standard significance test and depend on the sign of b. More importantly, the null hypothesis for the CM test is not $\beta = 0$. Therefore, we cannot set $\beta = 0$ if the CM test does not reject equal predictive ability of restricted and unrestricted model (RE and UR). Instead, in that case, we require estimation variance for $\hat{\beta}_T$ to be equal to squared bias for the restricted model, such that the MSE for estimation of the coefficient of $x_{t,T}$ is the same for RE and UR. This can be achieved by setting the coefficient to

$$\beta_{0,CM} = \sqrt{\mathrm{var}[\hat{\beta}_T]} = \sqrt{T^{-1}\hat{\sigma}^2_{\infty,T}} = T^{-1/2}\hat{\sigma}_{\infty,T}. \tag{8}$$

Note that we utilized the sign restriction on b to identify the sign of $\beta_{0,CM}$.

This results in the following pre-test estimator based on the CM test, which we call CMPT (Clark-McCracken Pre-Test).

$$\hat{\beta}_T^{CMPT} = \hat{\beta}_T \mathbf{I}(\hat{\tau} > c) + T^{-1/2}\hat{\sigma}_{\infty,T}\mathbf{I}(\hat{\tau} \leq c), \tag{9}$$

where, in general, c is different from the c used in the standard pre-test estimator (4), because the distributions of the test statistics differ.

The bagging version of the CMPT estimator (9) is

$$\hat{\beta}_T^{CMBG} = \frac{1}{B}\sum_{b=1}^{B}\left[\hat{\beta}_b^* \mathbf{I}(\hat{\tau}_b^* > c) + T^{-1/2}\hat{\sigma}_{\infty,T}\mathbf{I}(\hat{\tau}_b^* \leq c)\right]. \tag{10}$$

The first term in the sum is exactly the standard bagging estimator, except that the critical value c differs. The critical values for $CMBG$ come from the normal distribution $\mathcal{N}(sign(b),1)$, while critical values for standard bagging come from the standard normal distribution. The second term in the sum of (10) stems from the cases where the null is not rejected for bootstrap replication b. Note that we do not re-estimate the variance under the null, $\hat{\sigma}^2_{\infty,T}$, for every bootstrap sample. The main reason to apply bagging are hard-thresholds, which are not involved in the estimation of $\hat{\sigma}^2_{\infty,T}$, such that there is no obvious reason for bagging the variance estimator.

2.1 Asymptotic Distribution and Mean-Squared Error

We have proposed an estimator that is based on the CM test that better reflects our goal of improving forecast accuracy. In this section, we derive the asymptotic properties of this estimator to see if, and for which parameter configurations, this estimator indeed improves the asymptotic mean-squared error (AMSE). The asymptotic distribution for bagging estimators has been analyzed for bagging t-tests by Bühlmann and Yu (2002), and for sign restrictions by Gordon and Hall (2009).

Assumption 2 (Bühlmann and Yu (2002), A1)

$$T^{1/2}(\hat{\beta}_T - \beta_T) \xrightarrow{d} \mathcal{N}(0,\sigma_\infty^2), \tag{11}$$

$$\sup_{v\in\mathbb{R}} |\mathbb{P}^*[T^{1/2}(\hat{\beta}_T^* - \hat{\beta}_T) \leq v] - \Phi(v/\sigma_\infty)| = o_p(1), \tag{12}$$

where $\mathbb{P}^*$ *is the bootstrap probability measure.*

In fact, in the triangular array considered here, the CLT in equation (11) follows from Assumption 1. We restate it explicitly to make clear that the asymptotic framework is identical to Bühlmann and Yu (2002). The second part assumes that the bootstrap distribution converges to the asymptotic distribution of the CLT. Under Assumption 1, with a local-to-zero coefficient as in model (1), Bühlmann and Yu (2002) derive the asymptotic distribution for two-sided versions of the pre-test and the bagging estimators. The one-sided versions considered in this paper follow immediately as special cases. Let $\phi(.)$ denote the pdf and $\Phi(.)$ the cdf of a standard normal variable.

Proposition 1 (Special case of Bühlmann and Yu (2002), Proposition 2.2)

Under Assumption 2 for model (1)

$$T^{1/2}\hat{\sigma}_{\infty,T}^{-1}\hat{\beta}_T^{PT} \xrightarrow{d} (Z+b)\mathbf{I}(Z+b>c), \tag{13}$$

$$T^{1/2}\hat{\sigma}_{\infty,T}^{-1}\hat{\beta}_T^{BG} \xrightarrow{d} (Z+b)\Phi(Z+b-c)+\phi(c-Z-b), \tag{14}$$

where Z is a standard normal random variable.

The proposition follows immediately from Bühlmann and Yu (2002). The asymptotic distributions depend on b and c. For the pre-test estimator, the indicator function enters the asymptotic distribution. The distribution of the bagging estimator, on the other hand, contains smooth functions of b and c. Bühlmann and Yu (2002) show how for certain values of b and c, this reduces the variance of the estimator substantially. We adapt this proposition to derive the asymptotic distributions of the estimators CMPT, equation (9), and CMBG, equation (10).

Proposition 2

Under Assumption 2 and model (1)

$$T^{1/2}\hat{\sigma}_{\infty,T}^{-1}\hat{\beta}_T^{CMPT} \xrightarrow{d} (Z+b)\mathbf{I}(Z+b>c)+\mathbf{I}(Z+b\leq c), \tag{15}$$

and

$$T^{1/2}\hat{\sigma}_{\infty,T}^{-1}\hat{\beta}_T^{CMBG} \xrightarrow{d} (Z+b)\Phi(Z+b+c)+\phi(Z+b-c)+1-\Phi(Z+b-c), \tag{16}$$

where Z is a standard normal variable.

The proof of the proposition is given in the appendix. The asymptotic distributions are similar to those of the pre-test and bagging estimators (BG and PT), but involve extra terms due to the different null hypothesis. For CMPT, the extra term is simply an indicator function, and for CMBG it involves the standard normal cdf $\Phi(\cdot)$.

Figures 1 and 2 show asymptotic mean-squared error, asymptotic bias, asymptotic squared bias, and asymptotic variance of the pre-test and bagging estimators for test levels 5% and 1%, respectively. These quantities are functions of $b\sigma_\infty^{-1}$, which we will refer to as b_σ in the following for simplicity. Note that critical values for the t-test and the CM-test differ. The results for the two different levels are qualitatively identical. The effect of choosing a smaller significance level is that the critical values increase, and the effects from pre-testing become more pronounced. For the asymptotic mean-squared error (AMSE), we get the usual picture for PT and PTBG (see Bühlmann and Yu, 2002). Bagging improves the AMSE compared to pre-testing for a wide range of values of b_σ, except at the extremes. CMBG compares similarly to CMPT, but shifted towards the right compared to BG and PT. When looking at any given value b_σ, there are striking differences between the estimators based on the CM-test and the ones based on the t-test. Both CMPT and CMBG do not perform well for

b_σ close to zero, but the AMSE decreases as b_σ increases, before starting to slightly increase again. For values of b_σ from around 0.5 to 3, CMBG performs better than BG. For values larger than 3 the estimators PT, BG, and CMBG perform similarly and get closer as b_σ increases. Thus, the region where CMBG does not perform well are values of b_σ below 0.5.

The asymptotic biases for CMPT and CMBG are largest at $b_\sigma = 0$. For all estimators, the bias can be both positive or negative, depending on b_σ. Bagging can reduce bias compared to the corresponding pre-test estimation, in particular in the region where the pre-test estimator has the largest bias. CMPT and CMBG have very low variance for b_σ close to zero, because the CM test almost never rejects for these parameters. However, as the null hypothesis is not close to the true b_σ in this region, CMPT and CMBG are very biased. As b increases slightly, CMBG has the lowest asymptotic variance for b_σ up to around 3.

Figures 1 and 2 about here.

The asymptotic results show that imposing a different null hypothesis dramatically changes the characteristics of the estimators. The estimator based on the CM test is not intended to work for b_σ very close to zero. In this case, the standard pre-test estimator has much better properties. For larger b_σ, the CM-based estimators give substantially better forecasting results. The results highlight that the estimator will be useful for relations that are not expected to be zero, but too small to exploit with an unrestricted model.

2.2 Asymptotic Shrinkage Representation

Stock and Watson (2012) provide an asymptotic shrinkage representation of the BG estimator. This representation is given by

$$\hat{\beta}_T^{BGA} = \hat{\beta}_T \left[1 - \Phi(c - \hat{\tau}_T) + \hat{\tau}_T^{-1} \phi(c - \hat{\tau}_T)\right] \tag{17}$$

and Stock and Watson (2012, Theorem 2) show under general conditions that $\hat{\beta}_T^{BG} = \hat{\beta}_T^{BGA} + o_P(1)$. This allows computation without bootstrap simulation. While bootstrapping can improve test properties, bagging can improve forecasts even without actual resampling. There is no reason to suspect that the estimator based on the asymptotic distribution will be inferior to the standard bagging estimator. Therefore, we consider a version of the bagging estimators that samples from the asymptotic, rather than the empirical, distribution of $\hat{\beta}_T$. We can find closed form solutions for estimators that do not require bootstrap simulations.

Proposition 3 (Asymptotic Shrinkage representation)
Apply CMBG with the asymptotic distribution of $\hat{\beta}$ under Assumption 2, then

$$\hat{\beta}_T^{CMBGA} = \hat{\beta}_T \left[1 - \Phi(c - \hat{\tau}_T) + \hat{\tau}_T^{-1}\phi(c - \hat{\tau}_T) + \hat{\tau}_T^{-1}\Phi(c - \hat{\tau}_T)\right] \tag{18}$$

The proof of the proposition is given in the appendix. The representation is very similar to BGA in equation (17), with an extra term for the contribution for the CM null. Note that we can express $\hat{\beta}_T^{CMBGA}$ as the OLS estimator $\hat{\beta}_T$ multiplied by a function that depends on the data only through the t-statistic $\hat{\tau}_T$, just like $\hat{\beta}_T^{BGA}$.

Figure 3 about here.

Figure 3 plots BGA and CMBGA against the OLS estimate $\hat{\beta}$. The vertical deviation from the 45° line indicates the degree and direction of shrinkage applied by the estimator to the OLS estimate $\hat{\beta}$. This reveals the main difference between BGA and CMBGA. Rather than shrinking towards zero, CMBGA shrinks towards σ_β, which makes a substantial difference for b close to 0. For larger $\hat{\beta}$, the CMBGA, and thus CMBG, shrink more heavily downwards than BGA.

3 Monte Carlo Simulations

The asymptotic analysis suggests that our modified bagging estimator can yield significant improvements in MSE for the estimation of β. This section uses Monte Carlo simulations to investigate the performance for the prediction of $y_{t+h,T}$ using the estimators presented above in small samples. In our linear model (1), lower MSE for estimation of β can be expected to translate directly into lower MSE for prediction of $y_{t+h,T}$.

For the Monte Carlo simulations, we generate data from the following model, which is designed to resemble the empirical application of inflation forecasting:

$$\begin{aligned} y_{t+h,T} &= \mu + \beta_T x_t + u_{t+h} \\ u_{t+h} &= \epsilon_{t+h} + \theta_1 \epsilon_{t+h-1} + \cdots + \theta_{h-1}\epsilon_{t+1} \\ x_t &= \phi x_{t-1} + v_t \\ \epsilon_t &\sim \mathcal{N}(0, \sigma_\epsilon^2) \\ v_t &\sim \mathcal{N}(0, \sigma_v^2). \end{aligned} \tag{19}$$

We allow for serially correlated errors in the form of an MA(h-1) model. The choice of AR(1) for x_t is guided by the model for the monthly unemployment change series selected by AIC.

The predictor variable x_t is a weak predictor with $\beta_T = T^{-1/2}b$. We consider values $b \in \{\sigma_\infty, 0, 1, 2, 4\}$. For $b = \sigma_\infty$, the asymptotic standard deviation of $\hat{\beta}$, the performance of restricted and unrestricted model are asymptotically identical. We consider the forecasting methods discussed above. Table 1 presents an overview of all these methods.

Table 1 about here.

We are interested in the small-sample properties and consider sample sizes $T \in \{25, 50, 200\}$. Furthermore, we set $\mu = 0.1$ and $\phi = 0.66$, which we take from the our empirical example, i.e., monthly changes in unemployment. Additionally, we consider $\phi = 0.9$ to investigate the behavior for more persistent processes. Finally, we consider the forecast horizons $h = 1$ and $h = 6$. The MA coefficients are set to $\theta_i = 0.4^{i-1}$ for $1 \leq i \leq h-1$, and 0 otherwise. The critical values are taken from the respective asymptotic distribution of both tests for significance levels 5% and 1%. We run 10,000 Monte Carlo simulations, and use 299 bootstrap replications for bagging.

Columns 2 through 9 of Tables 2-5 show the MSE for the different estimators listed in Table 1. The last two columns show the rejection frequencies for the t-test and CM test. The MSE is reported in excess of $\text{var}[u_{t+h}]$, which does not depend on the forecasting model, such that the true model with known parameters will have MSE of zero.

For different values of $b\sigma_\infty^{-1}$, we get the overall patterns expected from the asymptotic results for all parameter configurations, sample sizes T, persistence parameters ϕ, and forecast horizons h. For $b\sigma_\infty^{-1} = 0$ the restricted model is correct. Forecast errors of the restricted model stem only from mean estimation. The CM-based methods perform worst, as the null hypothesis $b\sigma_\infty^{-1} = 1$ is incorrect, and the CM-test rejects very infrequently. The null of the t-test-based pre-test estimator is correct and is imposed whenever the test fails to reject, which happens frequently under all parameter configurations. This allows PT and its bagging version to achieve a lower MSE than the unrestricted model.

For $b\sigma_\infty^{-1} = 0.5$, the predictor is still so weak that the unrestricted model always performs best. The difference between using t-tests and CM-tests is not as large as it is for $b\sigma_\infty^{-1} = 0$. Setting $b\sigma_\infty^{-1} = 1$ (or equivalently $b = \sigma_\infty$) imposes that the unrestricted and restricted methods asymptotically have the same MSE for estimation of β. For $T = 25$, however, the restricted model has substantially lower MSE than the unrestricted model for the prediction of y_{t+h}. The difference disappears as the sample size grows. The rejection frequency for the CM-tests is fairly close to the nominal size for $h = 1$. For $h = 6$ the test is over-sized in small samples. Despite these small sample issues of the test, the CM-based estimators work well when $b\sigma_\infty^{-1} = 1$ even for $T = 25$ with $\phi = 0.66$ in Tables 2 and 4. For $\phi = 0.9$, shown in Tables 3 and 5, CM-test and t-test-based estimators perform very similarly for $T = 25$.

For $b\sigma_\infty^{-1} = 2$, the CM-based method is able to improve the MSE, even though the null hypothesis is not precisely true. The magnitude of the improvement depends on the persistence parameter ϕ, critical value, and sample size. For $b\sigma_\infty^{-1} = 4$ the coefficient is large enough such that the unrestricted model dominates. All other models except RE provide very similar performance. Both CM and t-test reject very frequently, such that the different null hypotheses are less important.

Our Monte Carlo simulations confirm that the asymptotic properties carry over to the small sample behavior of the estimators and the resulting forecasts. The bagging version of the CM-test can be expected to perform well when bias is not too small relative to the estimation uncertainty, i.e., $b\sigma_\infty^{-1}$ is not close to zero. If bias is much smaller than estimation uncertainty, then methods that shrink towards zero dominate. Our estimators will work well if the predictors is weak but the coefficient is strictly bigger than zero.

Tables 2 through 5 about here.

4 Application to CPI Inflation Forecasting

Inflation is a key macroeconomic variable, measuring changes in consumer price levels. Clearly, these price levels depend on the demand and supply for production and consumer goods. Thus, one would expect them to be linked negatively to unemployment and positively to industrial production. While economists and the media pay attention to such variables to assess inflationary pressure, the variables do not help to forecast inflation more accurately than univariate models. Cecchetti et al. (2000) find that using popular candidate variables as predictors fails to provide more accurate forecasts for US inflation, and that the relationship between inflation and some of the predictors is of the opposite sign as one would expect. Thus, they conclude that single predictor variables provide unreliable inflation forecasts.

Atkeson and Ohanian (2001) consider more complex autoregressive distributed-lags models for inflation forecasting and conclude that none of the models outperforms a random walk model. Stock and Watson (2007) argue that the relative performance of inflation forecasting methods depends crucially on the time period considered. Not only does the relative performance of forecasting methods change over time, but coefficients in the models are also likely to be time-varying. Stock and Watson (2009) go so far as to call it the consensus that including macroeconomic variables in models does not improve inflation forecasts over univariate benchmarks that do not utilize information other than past inflation.

We denote inflation by

$$\pi_t^h = \ln(P_{t+h}/P_t). \tag{20}$$

where P_t is the level of the US consumer price index (CPI, All Urban Consumers: All Items). We specify our models in terms of changes in inflation and aim to forecast these changes for different forecast horizons h. We define the change in inflation as $\Delta\pi_t^h = h^{-1}\pi_t^h - \pi_{t-1}^1$, i.e, the change of average inflation over the next h month compared to the most recent inflation rate. The forecast models are the specified as

$$\Delta\pi_t^h = \mu + \beta x_t + \epsilon_{t+h}, \tag{21}$$

where x_t is some predictor variable. For example, with a forecast horizon of 6 months ($h = 6$), we forecast the change in average inflation over the next 6 months compared to the current month's inflation. Figure 4 shows the target variable $\Delta\pi_t^h$ for different forecast horizons h. Even at the longest forecast horizon of 12 months, where we are forecasting annual inflation, the series is not very persistent. The estimation methods used to determine the parameters are the same as the ones used for the Monte Carlo simulations and are summarized in Table 1.

Figure 4 about here.

As predictor variables x_t, we use unemployment changes (UNEMP) and growth in industrial production (INDPRO). These transformation ensure that the predictors variables are stationary. Both variables are seasonally adjusted. We use monthly data for the period 1:1948–7:2013 from the latest data vintage available from St. Louis Fed's FRED[1], on August 21st, 2013.

For multiple-horizon forecasts, we choose a direct forecasting approach. Thus, the test statistics and parameter estimates depend on the forecasting horizon and can differ. For all forecast horizons, we use a short estimation window to allow for parameter instability. We use estimation window lengths of 24 and 60 months, which are reasonable sample sizes as we use only one predictor variable.

Bagging is conducted using a block bootstrap with block-length optimally chosen by the method of Politis and White (2004), applying the correction of Patton et al. (2009). For multiple-month forecasts ($h > 1$), we calculate standard errors using the method of Newey and West (1987) to account for serial correlation.

In Table 6, we show the MSE results for the pseudo out-of-sample forecasting exercise. The maximal out-of-sample period depends on the estimation window length m and the forecast horizon h. For example, for $m = 24$ and $h = 6$ we forecast inflation over 3:1953–7:2013 (725 observations) and for $m = 60$ and $h = 6$ over 3:1956–7:2013 (689 observations).

Table 6 about here.

[1] URL: http://research.stlouisfed.org/fred2

The first observation, in line with the existing literature on inflation forecasting, is that the restricted model is very hard to beat. The unrestricted model never performs better for the estimation window of 24, and for 60 it only performs better for the one-year ahead forecast ($h = 12$). The relative performance of the forecasting methods depends on the forecast horizon h. We apply the model confidence set of Hansen et al. (2011) to the resulting loss series and indicate whether results are significant.

The forecasting performance shows that the CM-based bagging predictor can indeed improve forecasting accuracy. Overall CMBGA and CMBG fare very well compared to standard bagging, BG and BGA, and the unrestricted model. Compared to the restricted model, which imposes a zero coefficient, we only improve for $h = 12$ significantly. The different critical values have only a minor effect on the performance of the predictors.

The performance differences between the bootstrap and the asymptotic versions of the bagging estimators are small. Thus, the asymptotic versions BGA and CMBGA offer computationally attractive alternatives to the bootstrap-based predictors BG and CMBG.

Figures 5 and 6 display the time series of coefficients from unrestricted estimation and CMBGA for $m = 24$ and $m = 60$, respectively. For $m = 24$, the coefficients from unrestricted estimation are very volatile and frequently change sign for both predictor variables. CMBGA imposes the sign restriction by construction and shrinks the coefficients heavily towards the null hypothesis, which results in much less volatile coefficients. For $m = 60$, the coefficients from unrestricted estimation are more stable, and sign changes of the coefficients are less frequent. CMBGA again shrinks the coefficients substantially and imposes the sign restriction.

Figures 5 and 6 about here.

Overall the proposed method works well for inflation forecasting, although the random walk benchmark, i.e., the restricted model, can only be beaten significantly at the one-year forecasting horizon. In line with previous literature, we find that inflation is hard to forecast and the unpredictable component remains large compared to the part that is predictable using either industrial production or unemployment. Even though our method improves the accuracy of the forecast, the total gains for prediction of inflation are very modest.

5 Conclusion

Bootstrap aggregation (bagging) is typically applied to t-tests of whether coefficients are significantly different from zero. In finite samples, a significantly non-zero coefficient is not sufficient to guarantee that including the predictor improves forecast accuracy. Instead, estimation variance has to be taken into account and weighed against bias from excluding the predictor.

We propose a novel bagging estimator that is based on the in-sample test for predictive ability of Clark and McCracken (2012), which addresses the bias-variance trade-off. We show that this estimator performs well when bias and variance are of similar magnitude. This is achieved by shrinking the coefficient towards an estimate of the estimation variance rather than shrinking towards zero. In order to find this shrinkage target, the sign of the coefficient has to be known. Thus, the method is appropriate for predictor variables for which theory postulates the sign of the relation, as is often the case for economic variables.

The new bagging estimator is shown to have good asymptotic properties, dominating the standard bagging estimator if bias and estimation variance are of similar magnitude. If, however, the data-generating coefficient is very close to zero, such that the forecasting power of the predictor is completely dominated by estimation uncertainty, the new estimator is very biased.

In this paper, we have been concerned with improving accuracy of a single predictor variable when predictive power is diluted by estimation variance. Using single predictors for forecasting is important, as many inflation predictors, for example, are considered individually (cf. Cecchetti et al., 2000). Econometric forecasting models, however, typically include multiple correlated predictor variables. In this context, our estimator could be applied to the individual predictor variables, just as standard bagging is applied in this context by, e.g., Inoue and Kilian (2008). The drawbacks of applying our estimator in this context to each predictor is that, first, it is harder to motivate sign restrictions on coefficients and, second, covariances are ignored when assessing the estimation uncertainty. The second issue can be fixed by using orthogonal factors instead of the original predictors, which makes it potentially even harder to find credible sign restrictions. The extension to multivariate specifications is left to future research.

References

Atkeson, A. and L. E. Ohanian (2001): "Are Phillips Curves Useful for Forecasting Inflation?" *Federal Reserve Bank of Minneapolis Quarterly Review*, 25, 2–11.

Breiman, L. (1996): "Bagging predictors," *Machine Learning*, 24, 123–140.

Bühlmann, P. and B. Yu (2002): "Analyzing bagging," *The Annals of Statistics*, 30, 927–961.

Campbell, J. Y. and S. B. Thompson (2008): "Predicting excess stock returns out of sample: Can anything beat the historical average?" *Review of Financial Studies*, 21, 1509–1531.

Cecchetti, S. G., R. S. Chu, and C. Steindel (2000): "The unreliability of inflation indicators," *Federal Reserve Bank of New York: Current Issues in Economics and Finance.*, 4.

Cheung, Y.-W., M. D. Chinn, and A. G. Pascual (2005): "Empirical exchange rate models of the nineties: Are any fit to survive?" *Journal of International Money and Finance*, 24, 1150–1175.

Clark, T. E. and M. W. McCracken (2012): "In-sample tests of predictive ability: A new approach," *Journal of Econometrics*, 170, 1–14.

de Jong, R. M. (1997): "Central limit theorems for dependent heterogeneous random variables," *Econometric Theory*, 13, 353–367.

Gordon, I. R. and P. Hall (2009): "Estimating a parameter when it is known that the parameter exceeds a given value," *Australian & New Zealand Journal of Statistics*, 51, 449–460.

Hansen, P. R., A. Lunde, and J. M. Nason (2011): "The Model Confidence Set," *Econometrica*, 79, 453–497.

Hillebrand, E., T.-H. Lee, and M. C. Medeiros (2013): "Bagging Constrained Equity Premium Predictors," in *Essays in Nonlinear Time Series Econometrics (Festschrift for Timo Teräsvirta)*, ed. by N. Haldrup, M. Meitz, and P. Saikkonen, Oxford University Press (forthcoming).

Hillebrand, E. and M. C. Medeiros (2010): "The benefits of bagging for forecast models of realized volatility," *Econometric Reviews*, 29, 571–593.

Inoue, A. and L. Kilian (2008): "How useful is bagging in forecasting economic time series? A case study of US consumer price inflation," *Journal of the American Statistical Association*, 103, 511–522.

Meese, R. A. and K. Rogoff (1983): “Empirical exchange rate models of the seventies: do they fit out of sample?” *Journal of International Economics*, 14, 3–24.

Newey, W. K. and K. D. West (1987): “A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix,” *Econometrica*, 55, 703–708.

Patton, A., D. N. Politis, and H. White (2009): “Correction to "Automatic block-length selection for the dependent bootstrap" by D. Politis and H. White,” *Econometric Reviews*, 28, 372–375.

Pettenuzzo, D., A. Timmermann, and R. Valkanov (2013): “Forecasting Stock Returns under Economic Constraints,” *CEPR Discussion Papers No. 9377*.

Politis, D. N. and H. White (2004): “Automatic block-length selection for the dependent bootstrap,” *Econometric Reviews*, 23, 53–70.

Rapach, D. E. and J. K. Strauss (2010): “Bagging or combining (or both)? An analysis based on forecasting US employment growth,” *Econometric Reviews*, 29, 511–533.

Stock, J. H. and M. W. Watson (2007): “Why has US inflation become harder to forecast?” *Journal of Money, Credit and Banking*, 39, 3–33.

——— (2009): “Phillips Curve Inflation Forecasts,” in *Understanding Inflation and the Implications for Monetary Policy: A Phillips Curve Retrospective*, ed. by J. Fuhrer, Y. Kodrzycki, J. Little, and G. Olivei, MIT Press, 99–186.

——— (2012): “Generalized shrinkage methods for forecasting using many predictors,” *Journal of Business & Economic Statistics*, 30, 481–493.

A Proofs

A.1 Proof of Proposition 2

The proof follows Bühlmann and Yu (2002), Proposition 2.2. We suppress the subscripts T for the sample sizes in the proofs to reduce notational clutter. From Assumption 2 and $\beta_T = T^{-1/2}b$, we get

$$T^{1/2}\hat{\sigma}_\infty^{-1}\hat{\beta} \xrightarrow{d} Z + b,\ Z \sim \mathcal{N}(0,1).$$

For $\hat{\beta}_{CMPT}$,

$$\begin{aligned} T^{1/2}\hat{\sigma}_\infty^{-1}\hat{\beta}^{CMPT} &= T^{1/2}\hat{\sigma}_\beta^{-1}\hat{\beta}\mathbf{I}(\hat{\tau} > c) + T^{1/2}\hat{\sigma}_\beta^{-1}T^{-1/2}\hat{\sigma}_\infty\mathbf{I}(\hat{\tau} \leq c) \\ &= T^{1/2}\hat{\sigma}_\infty^{-1}\hat{\beta}\mathbf{I}(\hat{\tau} > c) + \mathbf{I}(\hat{\tau} \leq c). \end{aligned}$$

Then by the continuous mapping theorem, because the right-hand side is continuous except for a single point of measure zero,

$$T^{1/2}\hat{\sigma}_\infty^{-1}\hat{\beta}^{CMPT} \xrightarrow{d} (Z+b)\mathbf{I}(Z+b > c) + \mathbf{I}(Z+b \leq c).$$

Next consider the bagged version

$$T^{1/2}\hat{\sigma}_\infty^{-1}\hat{\beta}^{CMBG} = \frac{1}{B}\sum_{b=1}^{B}\left[T^{1/2}\hat{\sigma}_\infty^{-1}\hat{\beta}_b^*\mathbf{I}(\hat{\tau}_b^* > c) + \mathbf{I}(\hat{\tau}_b^* \leq c)\right]. \tag{22}$$

From Assumption 2, part 2, we get

$$T^{1/2}(\hat{\beta}_b^* - \hat{\beta}) \xrightarrow{d^*} \mathcal{N}(0,\sigma_\infty^2),$$

where $\xrightarrow{d^*}$ denotes converges in distribution w.r.t. the bootstrap measure $\mathbb{P}^*$. That is,

$$\begin{aligned} T^{1/2}\sigma_\infty^{-1}\hat{\beta} &\xrightarrow{d} Z + b,\ Z \sim \mathcal{N}(0,1), \\ T^{1/2}\sigma_\infty^{-1}\hat{\beta}_b^* &\xrightarrow{d^*} W{\sim}|_Z\ \mathcal{N}(Z+b,1), \end{aligned}$$

where $W{\sim}|_Z$ denotes the distribution of W conditional on Z. Then,

$$\begin{aligned} \frac{1}{B} & \sum_{b=1}^{B}\left[T^{1/2}\hat{\sigma}_\infty^{-1}\hat{\beta}_b^*\mathbf{I}(\hat{\tau}_b^* > c) + \mathbf{I}(\hat{\tau}_b^* \leq c)\right], \\ \xrightarrow{d^*} & \ \mathbb{E}_W\left[W\mathbf{I}(W > c) + \mathbf{I}(W \leq c)|Z\right], \\ = & \ \mathbb{E}_W[W|Z] - \mathbb{E}_W\left[W\mathbf{I}(W \leq c)|Z\right] + \mathbb{E}_W\left[\mathbf{I}(W \leq c)|Z\right], \\ = & \ Z + b - \mathbb{E}_W\left[W\mathbf{I}(W \leq c)|Z\right] + \Phi(c - Z - b). \end{aligned}$$

For $x \sim \mathcal{N}(m,1)$, we have (Eqn. (6.3) in Bühlmann and Yu, 2002),

$$\mathbb{E}[x\mathbf{I}(x \leq k)] = m\Phi(k-m) - \phi(k-m),$$

and thus

$$\begin{aligned} & Z + b - \mathbb{E}_W\left[W\mathbf{I}(W \leq c)|Z\right] + \Phi(c - Z - b), \\ = \; & Z + b - (Z+b)\Phi(c-Z-b) + \phi(c-Z-b) + 1 - \Phi(Z+b-c), \\ = \; & Z + b - (Z+b)(1 - \Phi(Z+b-c)) + \phi(c-Z-b) + 1 - \Phi(Z+b-c), \\ = \; & (Z+b)\Phi(Z+b-c) + \phi(Z+b-c) + 1 - \Phi(Z+b-c), \end{aligned}$$

which completes the proof.

A.2 Proof of Proposition 3

Let $\beta_A \sim \mathcal{N}(\hat{\beta}, T^{-1}\hat{\sigma}^2_\infty)$, the random variable sampled from the asymptotic distribution of the OLS estimation with fixed $\hat{\beta}$ and $\hat{\sigma}_\infty$. Then, by the same arguments employed in the proof of Proposition 2,

$$\begin{aligned} \hat{\beta}^{BGA} &= \mathbb{E}[\beta_A \mathbf{I}(T^{1/2}\hat{\sigma}^{-1}_\infty \beta_A \; > c)] \\ &= \hat{\beta} - \mathbb{E}[\beta_A \mathbf{I}(T^{1/2}\hat{\sigma}^{-1}_\infty \beta_A \; \leq c)] \\ &= \hat{\beta} - T^{-1/2}\hat{\sigma}_\infty \mathbb{E}[T^{1/2}\hat{\sigma}^{-1}_\infty \beta_A \mathbf{I}(T^{1/2}\hat{\sigma}^{-1}_\infty \beta_A \; \leq c)] \\ &= \hat{\beta} - \hat{\beta}\Phi(c - T^{1/2}\hat{\sigma}^{-1}_\infty \hat{\beta}) + T^{-1/2}\hat{\sigma}_\infty \phi(c - T^{1/2}\hat{\sigma}^{-1}_\infty \hat{\beta}). \end{aligned}$$

With $\hat{\tau} = T^{1/2}\hat{\sigma}^{-1}_\infty \hat{\beta}$ we get

$$\begin{aligned} \hat{\beta}^{BGA} &= \hat{\beta}\left[1 - \Phi(c - \hat{\tau})\right] + T^{-1/2}\hat{\sigma}_\infty \phi(c - \hat{\tau}), \\ &= \hat{\beta}\left[1 - \Phi(c - \hat{\tau}) + \hat{\tau}^{-1}\phi(\hat{\tau} - c)\right]. \end{aligned}$$

We proceed along the same lines for $\hat{\beta}^{CMBGA}$:

$$\begin{aligned} \hat{\beta}^{CMBGA} &= \mathbb{E}[\beta_A \mathbf{I}(T^{1/2}\hat{\sigma}^{-1}_\infty \beta_A > c) + T^{-1/2}\hat{\sigma}_\infty \mathbf{I}(T^{1/2}\hat{\sigma}^{-1}_\infty \beta_A \leq c)] \\ &= \mathbb{E}[\beta_A \mathbf{I}(T^{1/2}\hat{\sigma}^{-1}_\infty \beta_> c)] + \mathbb{E}[T^{-1/2}\hat{\sigma}^*_\infty \mathbf{I}(T^{1/2}\hat{\sigma}^{-1}_\infty \beta_A \leq c)] \\ &= \hat{\beta}^{BGA} + T^{-1/2}\hat{\sigma}_\infty \mathbb{E}[\mathbf{I}(T^{1/2}\hat{\sigma}^{-1}_\infty \beta_A \leq c)] \\ &= \hat{\beta}^{BGA} + T^{-1/2}\hat{\sigma}_\infty \Phi(c - \hat{\tau}), \end{aligned}$$

which gives the desired result:

$$\hat{\beta}^{CMBGA} = \hat{\beta}\left[1 - \Phi(c - \hat{\tau}) + \hat{\tau}^{-1}\phi(c - \hat{\tau})\right] + T^{-1/2}\hat{\sigma}_\infty \Phi(c - \hat{\tau}).$$

Tables and Figures

Table 1: Forecasting methods for Monte Carlo and empirical application

Name	Method	Formula
RE	Restricted Model	$\hat{y}_{t+h} = \hat{\mu}_{t,h}$
UR	Unestricted Model	$\hat{y}_{t+h} = \hat{\mu}_{t,h} + \hat{\beta}_{t,h} x_t$
PT	Pre-Test t-test	$\hat{y}_{t+h} = \hat{\mu}_{t,h} + \mathbf{I}(\hat{\tau} > c)\hat{\beta}_{t,h} x_t$
BG	Bagging t-test	$\hat{y}_{t+h} = \hat{\mu}_{t,h} + \frac{1}{B}\sum_{b=1}^{B} \hat{\beta}_b^* \mathbf{I}(\hat{\tau}_b^* > c) x_t$
BGA	Asymptotic BG	$\hat{y}_{t+h} = \hat{\mu}_{t,h} + \hat{\beta}_{t,h}\left[1 - \Phi(c - \hat{\tau}) + \hat{\tau}^{-1}\phi(c - \hat{\tau})\right] x_t$
CMPT	Pre-Test CM-test	$\hat{y}_{t+h} = \hat{\mu}_{t,h} + \left(\hat{\beta}_{t,h}\mathbf{I}(\hat{\tau} > c) + T^{-1/2}\hat{\sigma}_{\infty}\mathbf{I}(\hat{\tau} \leq c)\right) x_t$
CMBG	Bagging CM-test	$\hat{y}_{t+h} = \hat{\mu}_{t,h} + \left(\frac{1}{B}\sum_{b=1}^{B} \hat{\beta}_b^* \mathbf{I}(\hat{\tau}_b^* > c) + T^{-1/2}\hat{\sigma}_{\infty,\beta}\mathbf{I}(\hat{\tau}_b^* \leq c)\right) x_t$
CMBGA	Asymptotic CMBG	$\hat{y}_{t+h} = \hat{\mu}_{t,h} + \left(\hat{\beta}_{t,h}\left[1 - \Phi(c - \hat{\tau}) + \hat{\tau}^{-1}\phi(c - \hat{\tau}) + \hat{\tau}^{-1}\Phi(c - \hat{\tau})\right]\right) x_t$

Note: $\hat{\mu}_{t,h}$ and $\hat{\beta}_{t,h}$ are the OLS estimates that depend on the forecast horizon.

Table 2: Monte Carlo Results for $\phi = 0.66$ and $c_{0.95}$

	MSE								Rejection %	
	RE	UR	PT	PTBG	PTBGA	CM	CMBG	CMBGA	t-test	CM-test
Panel 1: $b\sigma_\infty^{-1} = 0$										
h=1										
$T = 25$	6.84	21.74	11.55	12.63	12.54	21.60	23.39	23.19	5.90	0.60
$T = 50$	4.69	9.41	6.19	6.42	6.41	10.07	10.70	10.69	4.95	0.70
$T = 200$	0.74	1.78	0.97	1.03	1.01	1.73	1.88	1.87	4.60	0.55
h=6										
$T = 25$	19.43	42.22	28.92	30.66	29.81	39.33	42.17	41.16	14.00	6.10
$T = 50$	13.07	22.34	15.97	16.60	16.20	21.15	22.65	22.17	9.80	2.55
$T = 200$	2.60	4.60	3.10	3.34	3.22	4.84	5.18	5.07	5.85	1.00
Panel 2: $b\sigma_\infty^{-1} = 0.5$										
h=1										
$T = 25$	9.36	20.96	15.72	13.95	13.73	15.64	17.50	17.28	13.20	2.70
$T = 50$	6.63	10.69	8.18	7.50	7.48	7.60	8.24	8.23	11.95	2.15
$T = 200$	0.45	1.35	0.89	0.68	0.67	0.63	0.81	0.80	13.85	1.90
h=6										
$T = 25$	27.22	46.59	38.00	35.50	35.39	35.07	37.03	36.43	25.15	11.90
$T = 50$	13.82	21.23	17.03	16.26	15.82	16.12	17.59	16.93	20.00	6.75
$T = 200$	3.87	5.62	4.60	4.41	4.22	4.16	4.66	4.47	14.70	2.45
Panel 3: $b\sigma_\infty^{-1} = 1$										
h=1										
$T = 25$	17.67	23.81	22.89	18.13	18.00	14.43	15.48	15.50	23.80	6.10
$T = 50$	9.30	10.07	10.71	8.48	8.42	6.80	7.18	7.17	24.40	4.85
$T = 200$	2.06	2.26	2.43	1.89	1.87	1.35	1.44	1.43	27.05	5.25
h=6										
$T = 25$	44.20	46.14	45.87	39.96	40.41	36.61	36.49	36.24	42.60	23.40
$T = 50$	17.97	20.42	20.02	17.08	16.81	13.84	14.60	13.91	34.90	14.75
$T = 200$	4.85	4.80	5.28	4.36	4.22	3.47	3.77	3.55	26.40	7.10
Panel 4: $b\sigma_\infty^{-1} = 2$										
h=1										
$T = 25$	48.44	23.89	32.83	24.03	23.81	20.79	17.56	17.57	53.75	21.35
$T = 50$	21.81	9.94	14.16	10.49	10.42	9.58	7.76	7.73	58.50	22.75
$T = 200$	4.87	1.83	2.76	2.03	2.01	1.97	1.47	1.47	63.30	25.80
h=6										
$T = 25$	96.54	46.97	56.01	46.86	47.80	44.48	40.44	40.79	70.95	49.05
$T = 50$	46.09	21.85	27.92	22.41	22.79	21.54	19.01	18.89	66.25	38.45
$T = 200$	11.64	5.16	7.14	5.48	5.57	5.48	4.62	4.51	64.60	29.75
Panel 5: $b\sigma_\infty^{-1} = 4$										
h=1										
$T = 25$	149.03	21.10	26.21	24.11	23.82	29.24	22.90	22.78	93.65	74.20
$T = 50$	74.98	9.74	11.08	10.76	10.65	12.84	10.92	10.83	96.55	83.70
$T = 200$	18.29	2.74	2.84	2.90	2.90	3.30	3.05	3.04	98.80	90.15
h=6										
$T = 25$	302.49	40.50	44.53	42.16	42.87	51.46	42.16	45.52	96.95	87.95
$T = 50$	147.60	21.55	22.86	22.23	22.68	25.83	22.42	23.12	98.30	89.15
$T = 200$	34.15	4.31	4.57	4.45	4.64	5.59	4.49	4.83	98.40	87.50

Notes: MSE calculated in excess of var$[u_t]$, and multiplied by 100.

Table 3: Monte Carlo Results for $\phi = 0.9$ and $c_{0.95}$

	MSE								Rejection %	
	RE	UR	PT	PTBG	PTBGA	CM	CMBG	CMBGA	t-test	CM-test
Panel 1: $b\sigma_{\infty}^{-1} = 0$										
h=1										
$T = 25$	6.82	39.64	22.50	23.47	23.49	40.09	43.48	43.14	6.98	0.80
$T = 50$	3.60	13.89	7.72	8.07	8.03	13.36	14.46	14.38	5.64	0.74
$T = 200$	0.91	2.20	1.31	1.36	1.36	2.26	2.43	2.43	5.46	0.44
h=6										
$T = 25$	21.89	84.13	55.36	57.93	56.25	78.06	83.07	81.65	15.20	6.45
$T = 50$	13.49	35.24	23.79	25.76	24.68	34.48	37.33	36.31	9.50	3.00
$T = 200$	3.24	6.39	4.09	4.50	4.27	6.46	7.06	6.85	5.90	0.62
Panel 2: $b\sigma_{\infty}^{-1} = 0.5$										
h=1										
$T = 25$	12.17	41.66	27.36	26.12	25.94	32.45	36.03	35.62	10.64	1.62
$T = 50$	5.35	13.74	9.70	8.84	8.77	9.95	11.07	11.03	10.86	1.58
$T = 200$	1.40	2.47	1.87	1.69	1.68	1.73	1.90	1.90	11.98	1.84
h=6										
$T = 25$	24.19	77.86	56.40	53.55	52.63	58.47	62.08	61.08	23.90	11.25
$T = 50$	15.02	34.25	25.23	24.92	23.84	25.05	28.46	27.17	18.80	6.65
$T = 200$	3.66	5.99	4.77	4.60	4.28	4.48	5.20	4.87	14.02	2.50
Panel 3: $b\sigma_{\infty}^{-1} = 1$										
h=1										
$T = 25$	17.49	37.97	30.36	25.42	25.43	23.88	26.93	26.92	17.24	3.36
$T = 50$	8.18	12.51	11.21	8.77	8.68	7.18	8.13	8.07	20.36	3.76
$T = 200$	2.07	2.37	2.36	1.83	1.82	1.31	1.42	1.41	23.86	4.24
h=6										
$T = 25$	43.83	84.33	70.23	63.89	63.15	68.45	70.06	70.60	35.15	18.00
$T = 50$	24.15	30.77	28.57	24.50	23.59	21.09	23.42	22.05	31.65	12.90
$T = 200$	5.71	6.25	6.52	5.46	5.17	4.22	4.78	4.36	27.00	7.38
Panel 4: $b\sigma_{\infty}^{-1} = 2$										
h=1										
$T = 25$	52.56	39.39	46.45	34.30	34.30	26.47	26.01	26.06	38.94	12.62
$T = 50$	24.42	14.11	18.03	13.28	13.16	10.53	9.51	9.45	47.92	16.88
$T = 200$	5.57	2.47	3.51	2.60	2.58	2.42	1.94	1.93	59.18	23.00
h=6										
$T = 25$	128.12	82.76	91.29	73.93	74.94	65.95	63.16	62.30	58.20	37.95
$T = 50$	59.63	30.62	38.18	29.77	29.92	26.82	24.87	24.02	58.40	32.85
$T = 200$	13.22	5.93	8.41	6.17	6.22	5.74	4.98	4.70	59.84	27.30
Panel 5: $b\sigma_{\infty}^{-1} = 4$										
h=1										
$T = 25$	175.35	39.32	56.76	43.18	42.70	44.32	33.49	33.57	76.24	47.18
$T = 50$	83.22	13.60	18.01	15.59	15.38	18.25	13.91	13.79	88.14	64.32
$T = 200$	17.05	2.08	2.43	2.37	2.35	2.93	2.39	2.37	96.22	81.52
h=6										
$T = 25$	404.67	72.41	89.23	74.60	76.49	79.20	68.59	69.56	88.25	72.60
$T = 50$	205.15	33.91	41.26	35.47	36.85	41.92	33.88	35.09	92.05	75.40
$T = 200$	44.01	6.15	6.91	6.34	6.74	8.38	6.28	6.85	96.78	81.46

Notes: MSE calculated in excess of var$[u_t]$, and multiplied by 100.

Table 4: Monte Carlo Results for $\phi = 0.66$ and $c_{0.99}$

	MSE								Rejection %	
	RE	UR	PT	PTBG	PTBGA	CM	CMBG	CMBGA	t-test	CM-test
Panel 1: $b\sigma_{\infty}^{-1} = 0$										
h=1										
$T = 25$	7.26	21.02	10.54	11.06	10.91	21.38	22.26	21.98	1.42	0.16
$T = 50$	2.93	8.09	3.67	3.88	3.86	8.12	8.48	8.42	1.20	0.04
$T = 200$	0.63	1.67	0.74	0.78	0.78	1.71	1.76	1.76	1.04	0.10
h=6										
$T = 25$	18.45	42.04	26.89	27.70	26.98	37.10	39.17	37.99	8.15	3.45
$T = 50$	10.35	21.45	12.85	13.90	13.24	19.93	20.95	20.23	3.95	1.45
$T = 200$	3.39	5.39	3.62	3.83	3.69	5.52	5.73	5.62	1.56	0.30
Panel 2: $b\sigma_{\infty}^{-1} = 0.5$										
h=1										
$T = 25$	11.02	21.62	14.24	13.30	13.21	15.37	16.26	16.03	3.74	0.72
$T = 50$	5.62	10.08	6.73	6.29	6.25	6.66	7.06	7.02	3.62	0.36
$T = 200$	1.42	2.26	1.62	1.48	1.48	1.49	1.57	1.57	3.06	0.26
h=6										
$T = 25$	21.72	42.23	32.86	30.26	30.34	30.34	31.96	31.05	16.95	7.65
$T = 50$	10.73	18.75	13.89	13.21	12.76	13.12	14.29	13.48	8.90	3.10
$T = 200$	2.84	4.46	3.22	3.13	2.94	3.01	3.35	3.15	4.42	0.60
Panel 3: $b\sigma_{\infty}^{-1} = 1$										
h=1										
$T = 25$	17.39	22.84	20.74	16.60	16.57	12.09	12.92	12.71	9.08	1.54
$T = 50$	8.18	9.41	9.42	7.33	7.34	4.94	5.24	5.22	9.96	1.34
$T = 200$	1.75	1.89	2.04	1.51	1.49	0.88	0.95	0.94	9.46	0.94
h=6										
$T = 25$	41.53	45.77	45.05	38.09	39.21	33.44	33.88	33.13	28.55	14.70
$T = 50$	20.87	22.21	22.59	19.34	19.22	15.60	16.46	15.50	17.80	7.60
$T = 200$	4.52	4.81	4.98	4.17	4.02	2.97	3.29	3.03	9.88	1.98
Panel 4: $b\sigma_{\infty}^{-1} = 2$										
h=1										
$T = 25$	47.77	23.19	38.34	26.24	26.28	19.64	16.40	16.56	33.86	10.12
$T = 50$	20.74	9.30	16.73	11.02	10.98	8.37	6.61	6.65	33.70	9.18
$T = 200$	5.27	2.15	4.22	2.79	2.76	2.36	1.81	1.80	37.70	10.20
h=6										
$T = 25$	84.60	40.83	57.31	42.06	44.63	39.57	34.09	34.60	52.35	32.35
$T = 50$	48.41	22.29	34.13	24.36	25.91	23.47	19.73	20.04	50.10	26.25
$T = 200$	11.17	4.88	8.70	5.77	6.05	5.30	4.36	4.25	41.00	13.84
Panel 5: $b\sigma_{\infty}^{-1} = 4$										
h=1										
$T = 25$	148.82	22.75	40.96	31.12	30.87	40.47	27.96	28.23	80.40	53.04
$T = 50$	71.20	9.13	14.79	12.52	12.37	18.16	12.44	12.46	88.04	62.20
$T = 200$	17.51	2.31	3.05	2.92	2.90	4.40	3.20	3.18	93.68	70.42
h=6										
$T = 25$	345.55	49.56	64.55	54.44	57.74	67.39	54.12	57.22	93.55	81.45
$T = 50$	154.77	20.70	29.62	23.07	25.63	32.96	23.24	26.19	92.10	75.20
$T = 200$	37.91	4.95	6.22	5.36	5.95	8.45	5.59	6.52	95.06	76.14

Notes: MSE calculated in excess of var$[u_t]$, and multiplied by 100.

Table 5: Monte Carlo Results for $\phi = 0.9$ and $c_{0.99}$

	MSE								Rejection %	
	RE	UR	PT	PTBG	PTBGA	CM	CMBG	CMBGA	t-test	CM-test
Panel 1: $b\sigma_\infty^{-1} = 0$										
h=1										
$T = 25$	7.70	40.13	20.26	21.18	20.98	40.48	42.39	41.94	1.28	0.08
$T = 50$	3.76	13.34	6.46	6.74	6.69	13.41	13.90	13.83	1.10	0.08
$T = 200$	0.93	2.30	1.15	1.19	1.19	2.23	2.31	2.30	0.96	0.04
h=6										
$T = 25$	21.81	81.10	48.04	50.87	48.61	106.76	110.36	108.59	5.62	2.00
$T = 50$	10.76	32.52	18.37	19.82	18.55	32.27	34.25	33.05	3.78	1.02
$T = 200$	2.04	5.17	2.56	2.92	2.64	5.09	5.46	5.25	1.66	0.18
Panel 2: $b\sigma_\infty^{-1} = 0.5$										
h=1										
$T = 25$	12.23	40.64	22.78	21.74	21.49	30.68	32.76	32.24	3.00	0.26
$T = 50$	6.25	14.07	9.09	8.60	8.58	10.67	11.27	11.19	3.46	0.46
$T = 200$	1.22	2.26	1.48	1.34	1.34	1.39	1.49	1.48	3.48	0.24
h=6										
$T = 25$	32.00	80.32	54.94	53.58	51.93	61.73	67.38	63.71	14.35	7.50
$T = 50$	16.48	35.75	25.92	25.07	23.81	27.35	29.73	28.22	11.10	3.35
$T = 200$	3.49	5.95	4.23	4.21	3.86	4.04	4.58	4.22	4.58	0.64
Panel 3: $b\sigma_\infty^{-1} = 1$										
h=1										
$T = 25$	17.16	39.57	27.37	24.15	24.13	24.46	26.54	26.08	5.86	0.84
$T = 50$	9.25	13.95	11.92	9.85	9.84	7.70	8.30	8.24	7.88	0.94
$T = 200$	1.86	2.14	2.19	1.67	1.66	1.07	1.14	1.13	7.90	0.88
h=6										
$T = 25$	54.06	82.65	69.21	61.55	60.74	72.97	73.95	73.73	18.10	8.26
$T = 50$	23.57	32.24	28.70	24.88	23.95	19.93	22.05	20.35	14.76	4.86
$T = 200$	5.52	6.48	6.27	5.49	5.15	3.80	4.43	3.92	10.82	1.98
Panel 4: $b\sigma_\infty^{-1} = 2$										
h=1										
$T = 25$	46.31	36.34	43.88	31.21	31.50	21.13	20.94	20.74	17.50	4.28
$T = 50$	23.56	13.75	20.90	14.05	14.03	8.79	7.97	7.96	24.50	6.28
$T = 200$	5.31	2.26	4.28	2.79	2.76	2.16	1.72	1.71	33.78	8.60
h=6										
$T = 25$	127.54	81.29	100.74	75.29	77.74	64.03	58.88	59.03	39.90	21.04
$T = 50$	59.66	33.76	47.63	34.44	35.36	28.31	25.72	25.16	34.58	15.44
$T = 200$	14.60	6.58	11.28	7.47	7.87	6.66	5.71	5.48	37.94	12.38
Panel 5: $b\sigma_\infty^{-1} = 4$										
h=1										
$T = 25$	173.15	36.79	73.40	48.12	48.62	49.19	33.34	34.30	57.28	29.28
$T = 50$	80.02	13.95	27.18	19.35	19.18	22.39	15.46	15.48	71.10	42.02
$T = 200$	17.42	2.14	3.57	2.98	2.96	4.42	3.03	3.00	87.96	62.18
h=6										
$T = 25$	420.58	82.77	133.75	94.18	102.51	105.78	82.30	87.08	76.10	58.90
$T = 50$	206.64	31.99	56.46	36.95	42.37	48.94	34.13	37.89	80.70	60.90
$T = 200$	43.69	5.67	9.06	6.31	7.53	10.86	6.23	7.65	88.52	64.06

Notes: MSE calculated in excess of var$[u_t]$, and multiplied by 100.

Table 6: Out-of-sample inflation forecasting: MSE relative to restricted model.

Panel 1: $m = 24$

	$c_{0.99}$				$c_{0.95}$			
$h =$	1	3	6	12	1	3	6	12
RE	1*	1*	1*	1	1*	1*	1*	1
				INDPRO				
UR	1.095	1.116	1.072	1.142	1.095	1.116	1.072	1.142
PT	1.050	1.050	0.994*	0.995	1.046	1.076	1.005*	1.020
BGA	1.032	1.060	1.004	1.018	1.048	1.067	1.018	1.049
BG	1.036	1.057	1.009	1.015	1.053	1.067	1.025	1.045
CM	1.079	1.075	1.038	1.009	1.117	1.079	1.042	1.020
CMBGA	1.018	1.041*	0.993*	0.986	1.033	1.052	1.000*	1.005
CMBG	1.023	1.038*	0.999*	0.991	1.037	1.050	1.004*	1.008
				UNEMP				
UR	1.059	1.092	1.080	1.066	1.059	1.092	1.080	1.066
PT	1.004*	0.994*	1.007	0.973*	1.010*	1.042	1.020*	1.020
BGA	1.006	1.015*	1.008	0.980	1.019	1.035	1.026	0.991*
BG	1.006	1.025*	1.012	0.980	1.020	1.043	1.027	0.989*
CM	1.025	1.043	1.007*	0.981*	1.044	1.044*	1.015*	0.983*
CMBGA	0.991*	0.999*	0.988*	0.965*	1.001*	1.010*	0.999*	0.973*
CMBG	0.991*	1.011*	0.992*	0.967*	1.000*	1.021*	1.003*	0.973*

Panel 2: $m = 60$

	$c_{0.99}$				$c_{0.95}$			
$h =$	1	3	6	12	1	3	6	12
RE	1*	1*	1*	1	1*	1*	1*	1
				INDPRO				
UR	1.019	1.057	1.024	1.021	1.019	1.057	1.024	1.021
PT	1.001*	1.003*	1.002*	1.000	0.998*	1.005*	1.005*	0.999*
BGA	0.999*	1.014	1.000*	0.995*	1.002*	1.026	1.004*	0.997*
BG	0.999*	1.021	1.000*	0.995*	1.003*	1.032	1.004*	0.998*
CM	1.025	1.026	1.025	1.011	1.025	1.030	1.025	1.012
CMBGA	0.999*	1.003*	0.999*	0.995	1.000*	1.009*	1.000*	0.995*
CMBG	1.000*	1.010	0.999*	0.993*	1.000*	1.016*	0.999*	0.994*
				UNEMP				
UR	1.022	1.042	1.044	1.018	1.022	1.042	1.044	1.018
PT	0.998*	1.004*	1.012*	1.006	1.000*	1.022	1.043	1.003
BGA	1.000*	1.009	1.019	0.997	1.004	1.021	1.031	1.002
BG	1.001*	1.014	1.029	0.995*	1.005	1.025	1.037	1.000
CM	1.005	1.018	1.003*	0.997*	1.005*	1.022*	1.005*	1.001*
CMBGA	0.996*	1.000*	1.002*	0.991*	0.998*	1.006*	1.011*	0.994*
CMBG	0.997*	1.003*	1.009*	0.988*	0.999*	1.009*	1.021	0.992*

Notes: An asterisk (*) indicates that the model is included in 95% model confidence set (MCS). The MCS are computed for all methods with same m, c, and h, i.e., for every column in each panel. Thus, each MCS is computed for 15 models.

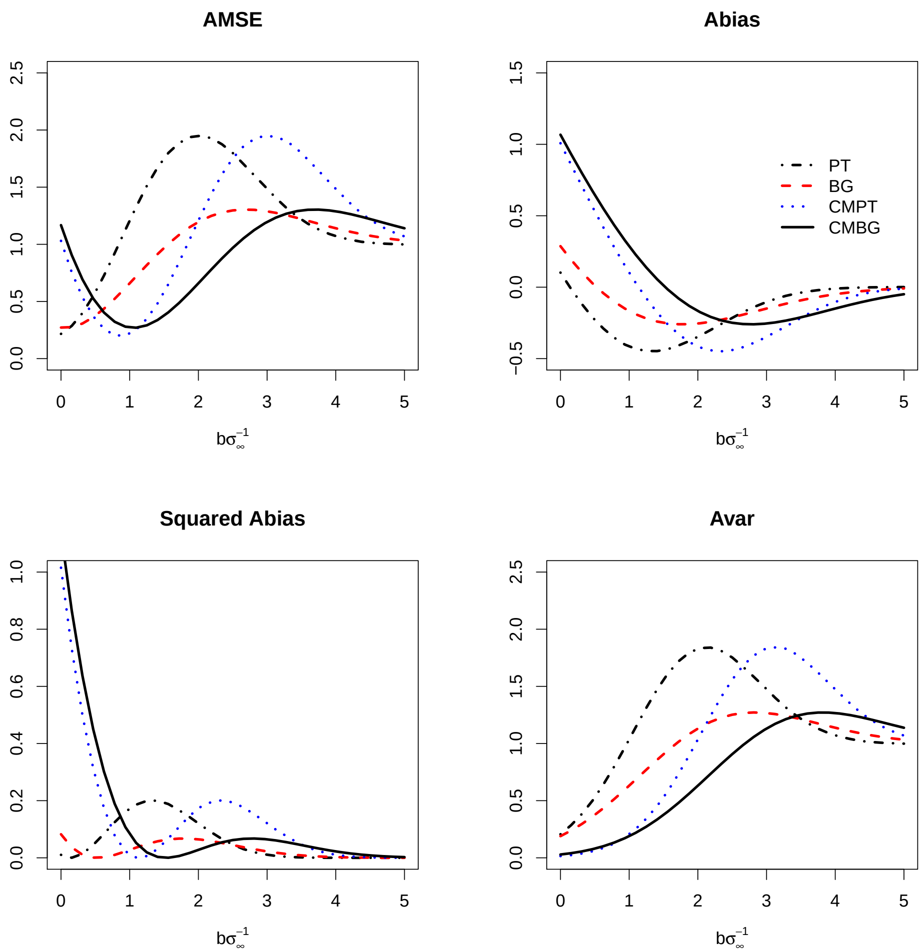


Figure 1: Comparison of asymptotic mean-squared error (AMSE), asymptotic bias (Abias), asymptotic square bias (Abias square), and asymptotic variance (Avar) as a function of $b\sigma_\infty^{-1}$ for 5% significance level .

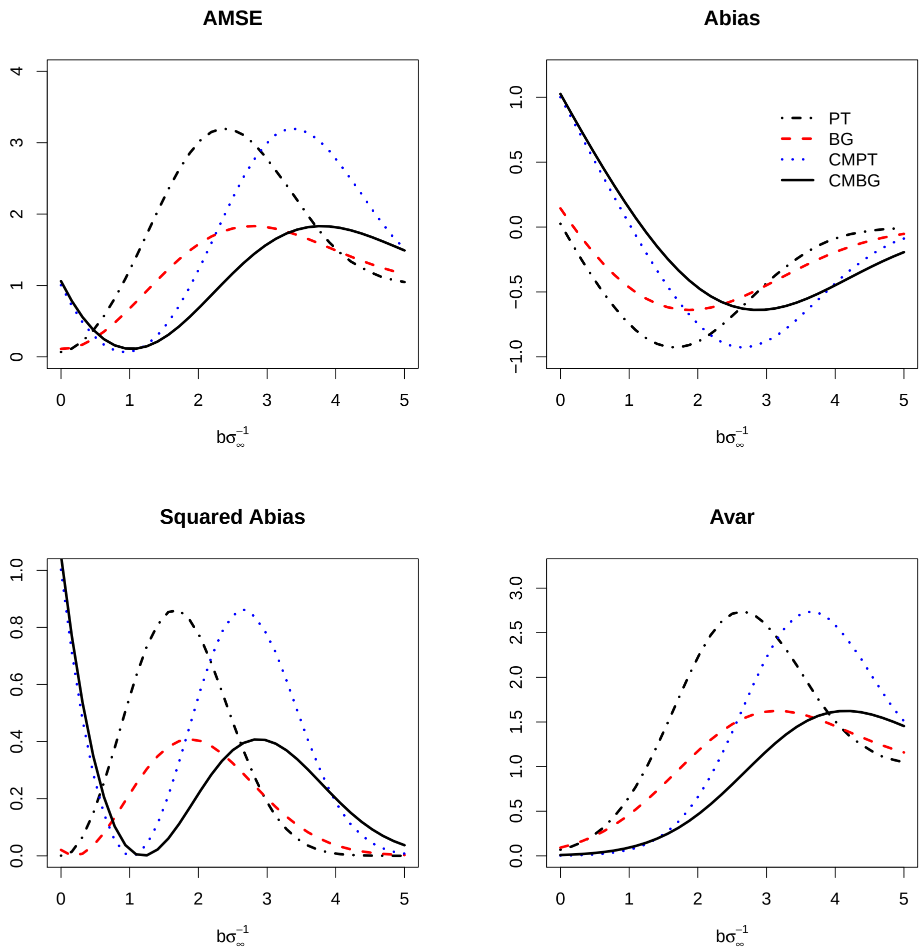


Figure 2: Comparison of asymptotic mean-squared error (AMSE), asymptotic bias (Abias), asymptotic square bias (Abias square), and asymptotic variance (Avar) as a function of $b\sigma_{\infty}^{-1}$ for 1% significance level .

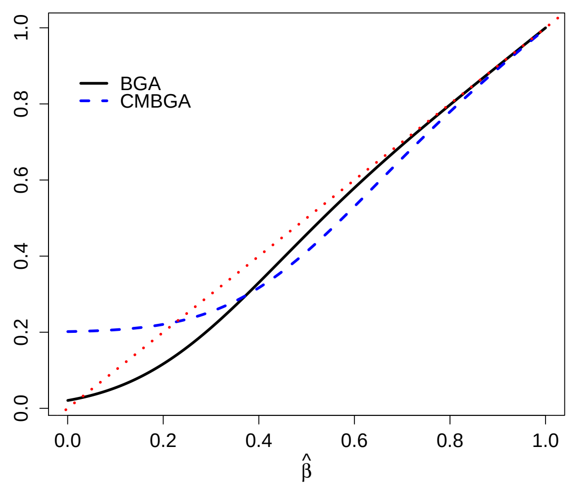


Figure 3: Shrinkage of slope parameter for $\sigma_\beta = 0.2$ *and* 5% *level. Dotted line is* 45° *line.*

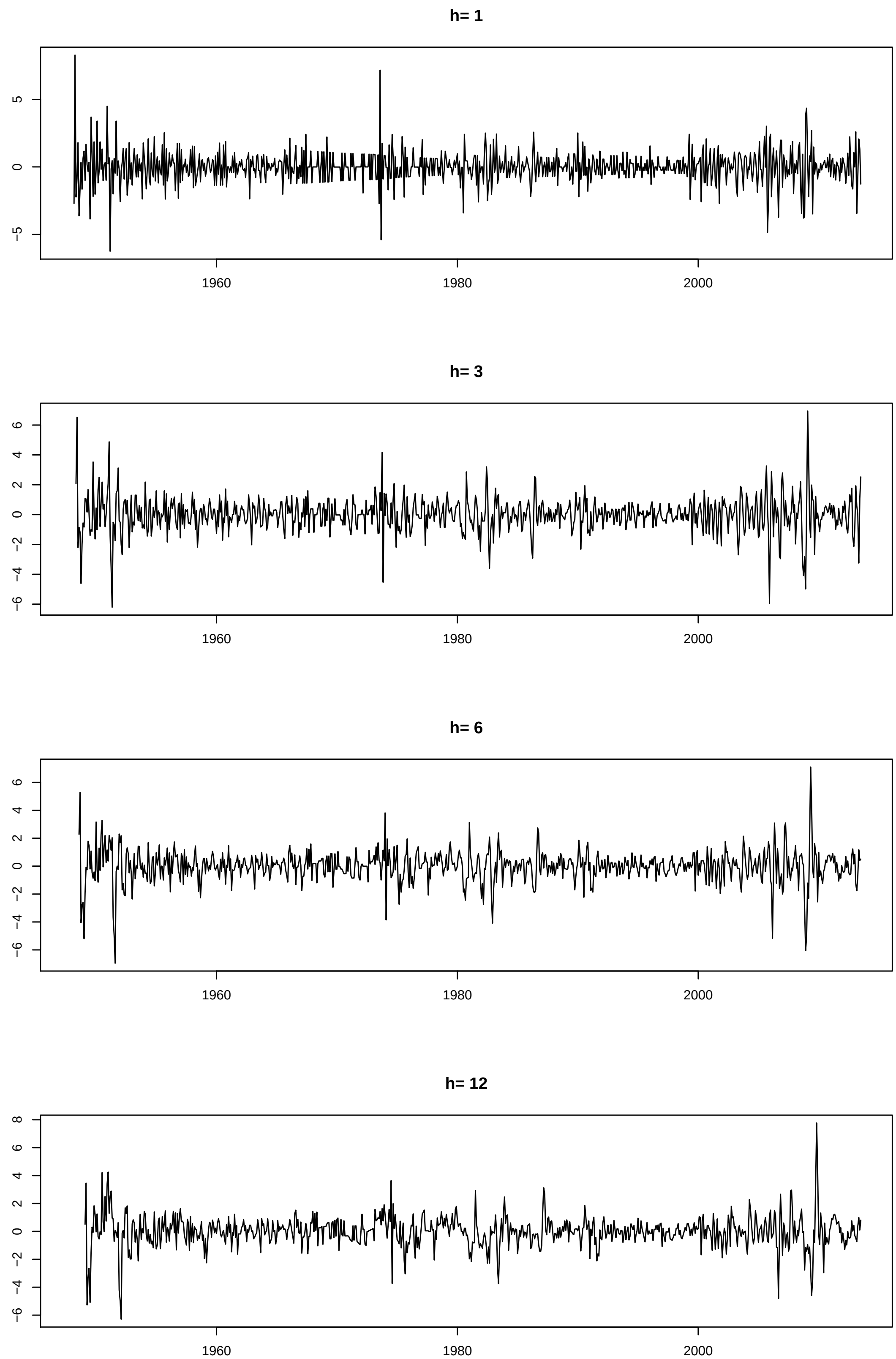
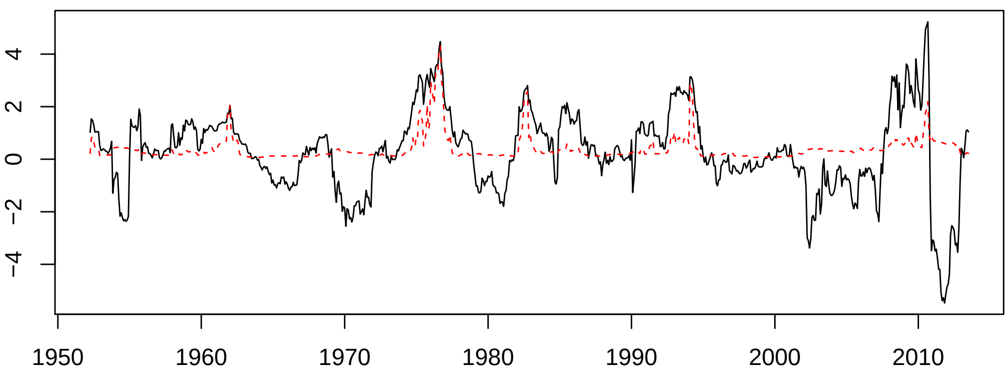
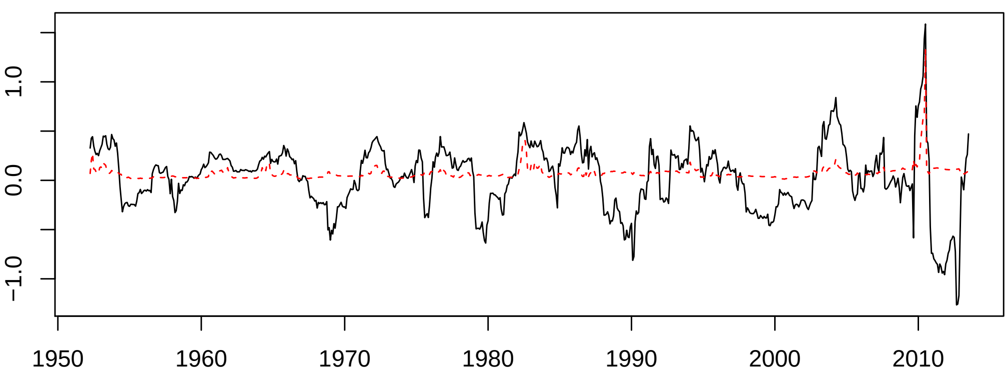


Figure 4: Time series of target variable $\Delta\pi_t^h$ at the different forecasting horizons h.

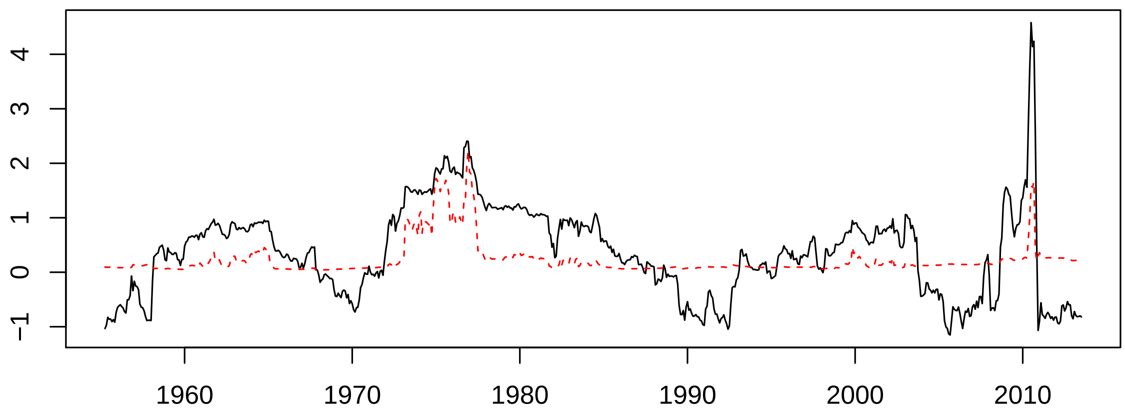


(a) Coefficients for unemployment changes (UNEMP).

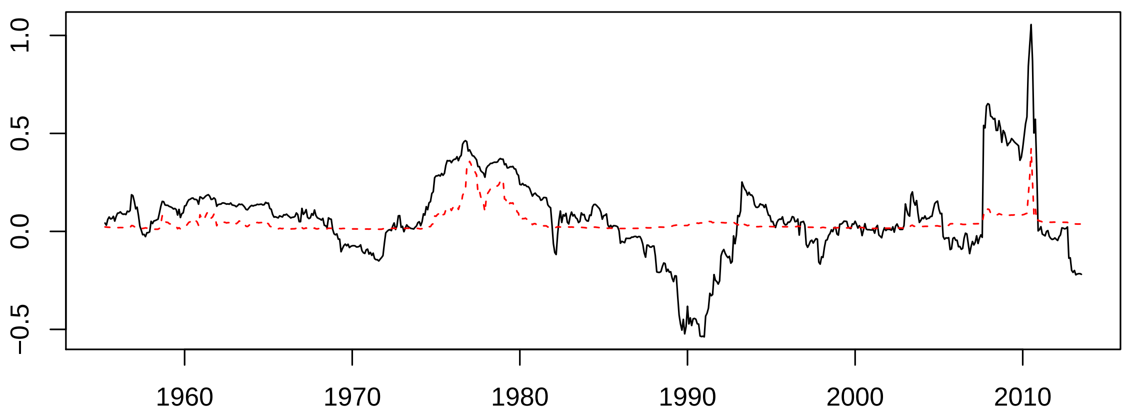


(b) Coefficients for industrial production growth (INDPRO).

Figure 5: Recursive coefficients for UR and CMBGA in forecast regressions of inflation changes on (a) unemployment changes and (b) industrial production growth. Forecast horizon $h = 12$ *and significance level 1%. Estimation window length* $m = 24$.



(a) Coefficients for unemployment changes (UNEMP).



(b) Coefficients for industrial production growth (INDPRO).

Figure 6: Recursive coefficients for UR and CMBGA in forecast regressions of inflation changes on (a) unemployment changes and (b) industrial production growth. Forecast horizon $h = 12$ and significance level 1%. Estimation window length $m = 60$.

Research Papers 2013

2013-36: Niels S. Hansen and Asger Lunde: Analyzing Oil Futures with a Dynamic Nelson-Siegel Model

2013-37: Charlotte Christiansen: Classifying Returns as Extreme: European Stock and Bond Markets

2013-38: Christian Bender, Mikko S. Pakkanen and Hasanjan Sayit: Sticky continuous processes have consistent price systems

2013-39: Juan Carlos Parra-Alvarez: A comparison of numerical methods for the solution of continuous-time DSGE models

2013-40: Daniel Ventosa-Santaulària and Carlos Vladimir Rodríguez-Caballero: Polynomial Regressions and Nonsense Inference

2013-41: Diego Amaya, Peter Christoffersen, Kris Jacobs and Aurelio Vasquez: Does Realized Skewness Predict the Cross-Section of Equity Returns?

2013-42: Torben G. Andersen and Oleg Bondarenko: Reflecting on the VPN Dispute

2013-43: Torben G. Andersen and Oleg Bondarenko: Assessing Measures of Order Flow Toxicity via Perfect Trade Classification

2013-44: Federico Carlini and Paolo Santucci de Magistris: On the identification of fractionally cointegrated VAR models with the F(d) condition

2013-45: Peter Christoffersen, Du Du and Redouane Elkamhi: Rare Disasters and Credit Market Puzzles

2013-46: Peter Christoffersen, Kris Jacobs, Xisong Jin and Hugues Langlois: Dynamic Diversification in Corporate Credit

2013-47: Peter Christoffersen, Mathieu Fournier and Kris Jacobs: The Factor Structure in Equity Options

2013-48: Peter Christoffersen, Ruslan Goyenko, Kris Jacobs and Mehdi Karoui: Illiquidity Premia in the Equity Options Market

2013-49: Peter Christoffersen, Vihang R. Errunza, Kris Jacobs and Xisong Jin: Correlation Dynamics and International Diversification Benefits

2013-50: Georgios Effraimidis and Christian M. Dahl: Nonparametric Estimation of Cumulative Incidence Functions for Competing Risks Data with Missing Cause of Failure

2013-51: Mehmet Caner and Anders Bredahl Kock: Oracle Inequalities for Convex Loss Functions with Non-Linear Targets

2013-52: Torben G. Andersen, Oleg Bondarenko, Viktor Todorov and George Tauchen: The Fine Structure of Equity-Index Option Dynamics

2014-01 Manuel Lukas and Eric Hillebrand: Bagging Weak Predictors