MONASH University

MONASH BUSINESS SCHOOL

ISSN 1440-771X

Department of Econometrics and Business Statistics

http://business.monash.edu/econometrics-and-business-statistics/research/publications

Bagging Weak Predictors

Eric Hillebrand, Manuel Lukas, and Wei Wei

April 2020

Working Paper 16/20

AACSB ACCREDITED EFMD EQUIS ACCREDITED ASSOCIATION OF MBAS ACCREDITED

Bagging Weak Predictors

Eric Hillebrand[a], Manuel Lukas[a], and Wei Wei * [a,b]

[a] *CREATES, Department of Economics and Business Economics, Aarhus University, Denmark*
[b] *Department of Econometrics and Business Statistics, Monash University, Australia*

April 10, 2020

Abstract

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

Keywords: Inflation forecasting, bootstrap aggregation, estimation uncertainty, weak predictors, shrinkage methods.

JEL classification: C32, E37.

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

*Email: wei.wei2@monash.edu. Address: 900 Dandenong Road, Caulfield East, VIC 3145, Australia

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, often 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 uncertainty and the instability of a variable's predictive power. 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 or model instability. While a short estimation window can better adapt to changing parameters, it increases the variance from estimation 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 accounting for estimation variance in finite sample using pre-test estimators and reducing estimation variance by bagging. Clark and McCracken (2012) (CM henceforth) 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. Jin et al. (2014) justify the validity of bagging in reducing mean squared error in the presence of time series dependence, and their application in forecasting equity premium shows that bagging helps to improve the mean squared forecast error for "unstable" predictors. Jordan et al. (2017) apply bagging to forecast equity premium in G7 and a broad set of Asian countries and they find bagging generally improves 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 to reduce the variance from imposing hard-threshold sign restrictions on parameters. Gordon and Hall (2009) consider bagging the hard-threshold estimator and show analytically that bagging can reduce variance. In predicting the equity premium, sign restrictions arise somewhat naturally, see Campbell and Thompson (2008) for an illustration using 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.

For macroeconomic variables, economic theory can provide support for sign restrictions. For example, in a New Keynesian model, inflation is positive related to the output gap (the deviation of output from its potential level)—a relationship often referred to as the New Keynesian Phillips curve (Galí, 2015). Meanwhile, the negative relationship between unemployment and inflation—the traditional Phillips curve—also seems quite robust when changes in inflation expectation is taken into consideration, see for example Watson (2014) and Ball and Mazumder (2019). The gain from using activity variables to forecast inflation, however, remains nebulous due to estimation uncertainty and parameter instability.

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 and T 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

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

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 as explained in the introduction.

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

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

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 that whether 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

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

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 from those of standard significance tests 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 the estimation variance of $\hat{\beta}_T$ to be equal to the 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

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

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 this estimator indeed improves the asymptotic mean-squared error (AMSE) and for which parameter configurations. 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)

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

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

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 (PT) and the bagging (BG) 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 PT and BG, 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 $\Phi(\cdot)$.

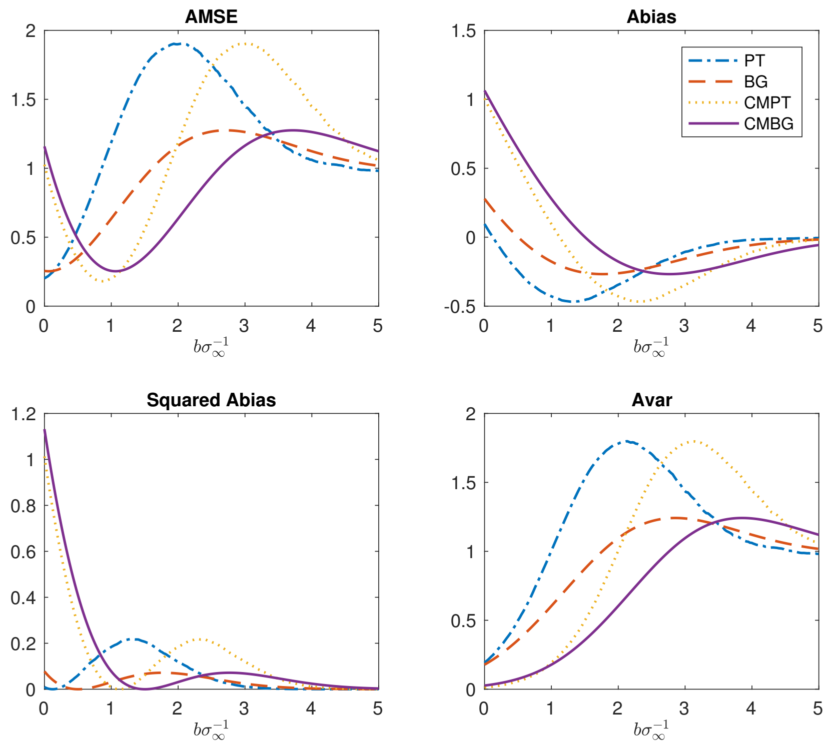

Figure 1: Comparison of asymptotic mean-squared error (AMSE), asymptotic bias (Abias), asymptotic square bias (Abias square), and asymptotic variance (Avar) as a function of $b\sigma_\infty^{-1}$ for 5% significance level .

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

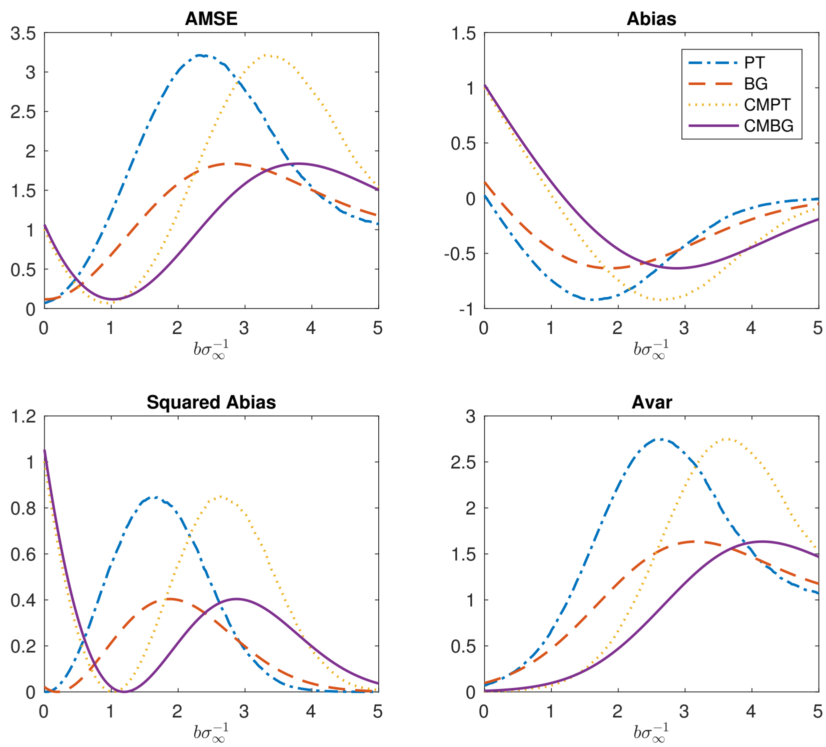

Figure 2: Comparison of asymptotic mean-squared error (AMSE), asymptotic bias (Abias), asymptotic square bias (Abias square), and asymptotic variance (Avar) as a function of $b\sigma_\infty^{-1}$ for 1% significance level .

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.

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 $\hat{\beta}_T^{BGA}$.

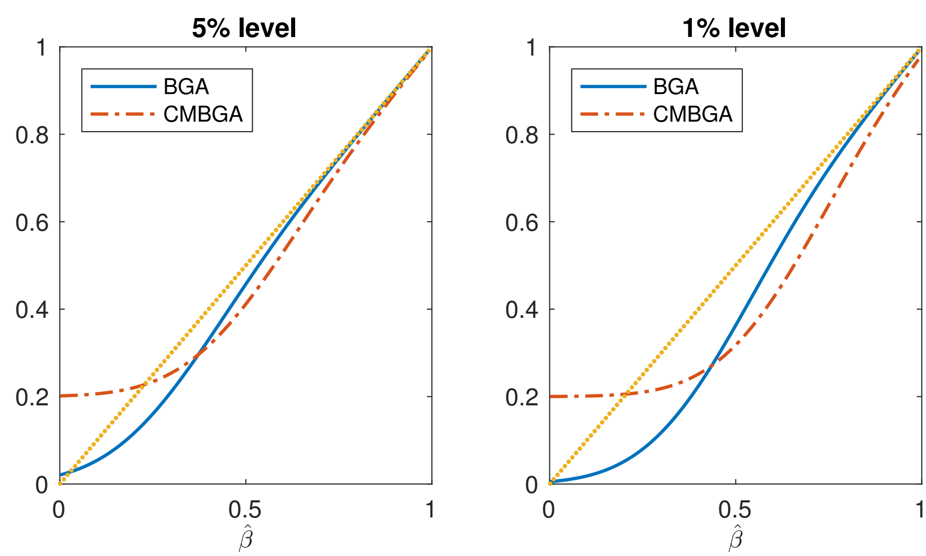

Figure 3: Shrinkage of slope parameter for $\sigma_\beta = 0.2$. Dotted line is 45° line.

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. Figure 3 also demonstrates the differences between the significant levels, with lower significant level providing heavier shrinkage factor.

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 resembles many empirical applications in macro 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 predictor variable x_t is a weak predictor with $\beta_T = T^{-1/2} b$. We consider values $b \in \{0,0.5,1,2,4\}\sigma_\infty$. For $b = \sigma_\infty$, the asymptotic standard deviation of $\hat{\beta}$, the performance of restricted and unrestricted model are asymptotically identical. Forecasts from the unrestricted model are given by $\hat{y}^{UR}_{t+h,T} = \hat{\mu} + \hat{\beta} x_t$, where $\hat{\mu}$ and $\hat{\beta}$ are the OLS estimates for the given h and T. We apply the forecasting methods discussed in Section 2 and obtain forecasts from method M as $\hat{y}^M_{t+h,T} = \hat{\mu} + \hat{\beta}^M x_t$. Table 1 presents an overview of all these methods.

Table 1: Forecasting methods

Name	Method	Formula
RE	Restricted Model	$\hat{\beta}^{\text{RE}} = 0$
UR	Unrestricted Model	$\hat{\beta}^{\text{UR}} = \hat{\beta}$
PT	Pre-Test t-test	$\hat{\beta}^{\text{PT}} = \mathbf{I}(\hat{\tau} > c)\hat{\beta}$
PTBG	Bagging t-test	$\hat{\beta}^{\text{BG}} = B^{-1}\sum_{b=1}^{B}\hat{\beta}^*_b\mathbf{I}(\hat{\tau}^*_b > c)$
PTBGA	Asymptotic BG	$\hat{\beta}^{\text{BGA}} = \hat{\beta}\left[1 - \Phi(c-\hat{\tau}) + \hat{\tau}^{-1}\phi(c-\hat{\tau})\right]$
CMPT	Pre-Test CM-test	$\hat{\beta}^{\text{CMPT}} = \hat{\beta}\mathbf{I}(\hat{\tau} > c) + T^{-1/2}\hat{\sigma}_\infty\mathbf{I}(\hat{\tau} \le c)$
CMBG	Bagging CM-test	$\hat{\beta}^{\text{CMBG}} = B^{-1}\sum_{b=1}^{B}\hat{\beta}^*_b\mathbf{I}(\hat{\tau}^*_b > c) + T^{-1/2}\hat{\sigma}_{\infty,\beta}\mathbf{I}(\hat{\tau}^*_b \le c)$
CMBGA	Asymptotic CMBG	$\hat{\beta}^{\text{CMBGA}} = \hat{\beta}\left[1 - \Phi(c-\hat{\tau}) + \hat{\tau}^{-1}\phi(c-\hat{\tau}) + \hat{\tau}^{-1}\Phi(c-\hat{\tau})\right]$

Note: $\hat{\beta}$ is the OLS estimates for a given forecast horizon and estimation window.

We are interested in the small-sample properties and consider sample sizes $T \in \{25, 50, 200\}$. Furthermore, we set $\mu = 0.1$ and $\phi \in \{0.66, 0.9\}$ to investigate the behavior of less versus more persistent processes. Finally, we consider the forecast horizons $h = 1$ and $h = 6$. The MA coefficients are set to $\theta_i = 0.4^i$ for $1 \le i \le h-1$, 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 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 allows PT and its bagging version to achieve a lower MSE than the unrestricted model for all ϕ and significance levels.

For $b\sigma_\infty^{-1} = 0.5$, the predictor is still so weak that the restricted model always performs best, but the difference between using t-tests and CM-tests is not as large as it is for $b\sigma_\infty^{-1} = 0$. For $b\sigma_\infty^{-1} = 1$ (or equivalently $b = \sigma_\infty$), the unrestricted and restricted methods have the similar MSE as predicted. 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 $b\sigma_\infty^{-1} = 1$ and $\phi = 0.66$ in Tables 2 and 4. For $\phi = 0.9$, shown in Tables 3 and 5, CM and t-test-based estimators perform similarly for $T = 25$, but CM-based methods have the lowest MSE among all methods with bigger sample size.

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.

4 Application to CPI Inflation Forecasting

Inflation is a key macroeconomic variable, measuring changes in price levels. Since price levels depend on the demand and supply for production and consumer goods, one would expect them to be linked negatively to unemployment and positively to industrial production. While economists and the media pay attention to unemployment/industrial production for assessing inflationary pressure, whether models that use these activity variables—often referred to as Phillips curve models—produce more accurate inflation forecast than univariate models such as integrated moving average models remains controversial. 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 multivariate model outperforms a random walk. Stock and Watson (2007) divide the data into two sample periods, 1970-1984 and 1983-2004. They show that the random walk model substantially outperforms multivariate ones in the post-1984 period. They propose an unobserved component with stochastic volatility model which can be interpreted as an integrated moving average (MA) with time-varying MA parameter. Since then, models that account for time-varying parameters became benchmarks in inflation forecasting. For example, Faust and Wright (2013) conclude "any good forecast must account for a slowly varying local mean". Pettenuzzo and Timmermann (2017) compare the inflation and GDP growth forecasts from several models that allow parameter instability and they find that the model with time varying parameter and stochastic volatility offers the largest gain over a constant parameter model, especially in terms of density forecast and at short horizons up to one year.

Parameter and model instability is even more crucial for Phillips curve models. Stock and Watson (2009) argue that the relative performance of Phillips curve models depends on the time period considered; there are notable periods when these models do well, but on average they do not improve upon an univariate model. Rossi (2013) states as a stylized fact that the predictive content of some variables is substantial but sporadic. Facing these instabilities, shrinkage methods such as bagging, ridge regression and Lasso regression become increasingly popular in forecasting macroeconomic variables, see for example Kim and Swanson (2014). Carrasco and Rossi (2016) show that ridge and Bayesian model averaging perform best when a large set of predictors is considered. Lasso also shows promising results when time-varying parameters are taken into consideration, see e.g. Kapetanios and Zikes (2018).

We focus on the Phillips curve type model and deal with instability in the predictive content of activity variables using the bagging estimators proposed in Section 2 with a rolling estimation window. We denote the annualized h-quarter inflation by

$$\pi_t^h = \frac{400}{h} \ln(P_t/P_{t-h}), \tag{20}$$

where P_t is the level of quarterly US consumer price index (CPI, All Urban Consumers: All Items). Let π_t denote the annualized quarterly inflation ($\pi_t = \pi_t^1$), then $\pi_t^h = \sum_{i=0}^{h-1} \pi_{t-i}/h$. Multivariate models for forecasting h-quarter inflation generally take the form

$$\pi_{t+h}^h = \mu_h + f_h(\pi_t,\pi_{t-1},...) + g_h(x_t,x_t^*,x_{t-1},x_{t-1}^*,...) + \epsilon_{t+h}, \tag{21}$$

where $f_h(\cdot)$ specifies the dependence of h-quarter inflation on inflations up to time t, and $g_h(\cdot)$ specifies the relationship between inflation and the predictor. Models motivated by the backward-looking Phillips curve specify either inflation or its first difference by an AR process and use the direct (rather than iterated) method for multi-step forecasts. For example, Stock and Watson (2007) set $f_h(\cdot) = \pi_t + \alpha^h(L)\Delta\pi_t$, and Faust and Wright (2013) opt for $f_h(\cdot) = \alpha^h(L)\pi_t$. Neither specification, however, produces forecasts that offer any improvement over the extended random walk model in Atkeson and Ohanian (2001):

$$\pi_{t+h}^h = \pi_t^4 + \epsilon_{t+h}. \tag{22}$$

In a New-Keynesian Phillips curve (NKPC) model, inflation is related to unemployment or output through

$$\pi_t = \alpha \mathbb{E}_t \pi_{t+1} + \beta(x_t - x_t^*) + e_t, \tag{23}$$

where $\mathbb{E}_t \pi_{t+1}$ is the expected future inflation, x_t^* the natural rate of unemployment/output, and $(x_t - x_t^*)$ the unemployment/output gap. Proxies for expected future inflation are often constructed by past inflation, and NKPC differs from backward-looking PC in how to restrict the dependence of inflation on its lags (Gordon, 2013).

Motivated by the NKPC in Equation (23) and the success of the random walk model, we specify $f_h(\cdot) = \alpha\pi_t^4$ for any h. In other words, we use past year's inflation as a proxy for inflation expectation. Our specification has fewer parameters to estimate than the traditional backward-looking PC and hence offers hope for better forecasting performances. Moreover, averaging over 4 quarters of inflation can be viewed as a filtering technique to capture the slowly evolving local mean as Faust and Wright (2013) point out. Setting $g_h(\cdot) = \Delta x_t$, we have the unrestricted model

$$\pi_{t+h}^h = \mu_h + \alpha_h \pi_t^4 + \beta_h \Delta x_t + \epsilon_{t+h}, \tag{24}$$

where $h \in \{1,2,3,4,6\}$. The h-period forecast for inflation in the unrestricted model is given by $\pi_{t+h|t}^h = \hat{\mu}_h + \hat{\alpha}_h \pi_t^4 + \hat{\beta}_h \Delta x_t$, where parameter estimates are obtained from OLS. The restricted model sets $\beta_h = 0$. As in the simulation exercise, we consider three PT-based and three CM-based estimators for β_h as summarized in Table 1. 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). In addition to the models in Table 1, we also consider the random walk (RW) model in Atkeson and Ohanian (2001), where forecasts of π^h_{t+h} are set to π^4_t for any h, and the ridge and Lasso regression specified as follows,

$$(\mu_h,\alpha_h,\beta_h) = \arg\min \left(\sum_{t=1}^{T} \left(\pi^h_{t+h} - \mu_h - \alpha_h \pi^4_t + \beta_h \Delta x_t \right)^2 + \lambda_i ||\beta_h||_i \right), \tag{25}$$

where $i = 2$ for ridge (L_2 norm) and $i = 1$ for Lasso (L_1 norm). For ridge regression, we consider a geometric sequence $\lambda^0_2 = 0.001 \times \{1,2,\ldots,2^{14}\}$, which increases from 0.001 to 16.384 (15 values), then set $\lambda_2 = \lambda^0_2 \times \mathrm{var}(\Delta x)$ for each rolling sample instead of standardizing the data. The best λ_2 in each rolling estimation window is selected using the generalized cross validation formula in Carrasco and Rossi (2016) to produce the forecast. For Lasso regression, we consider $\lambda^0_1 = 0.0001 \times \{1,2,\ldots,2^{14}\}$, which increases from 0.0001 to 1.6384 (15 values), then set $\lambda_1 = \lambda^0_1 \times \mathrm{cov}(\pi^h_{t+h},\Delta x)$ for each rolling sample. The best λ_1 for each sample is chosen by a 10-fold cross validation to produce the forecast.

We use quarterly CPI (CPIAUCSL), unemployment (UNEMP), and industrial production (INDPRO) obtained from the latest vintage of FRED-QD database. The data spans 1959Q1 to 2018Q4. Unemployment is inversely related to inflation, and we use the negative of UNEMP as x and impose a positive sign restriction. For all forecast horizons, we set the estimation window T to 40, 60, and 80 quarters, which correspond to 10, 15, and 20 years. The estimation windows are chosen to balance between parameter instability and estimation uncertainty. The out-of-sample period depends on the estimation window length T and the forecast horizon h. For example, for $T = 40$ and $h = 1$ we forecast quarterly inflation over

1969Q3-2018Q4 (198 observations) and for $T = 80$ and $h = 4$ we forecast yearly inflation (overlapping) between1981Q4 and 2018Q4 (149 observations).

Table 6 reports the MSE for the pseudo out-of-sample forecasting exercise. For the three PT-based and three CM-based models, we report results using critical values that correspond to significance level 1% (the results for significance level 5% are qualitatively similar and hence omitted). Results for all MSE are relative to that of the RW model specified in Equation (22). We compute model confidence set (MCS) p-values using Hansen et al. (2011) for each T, h, and x. Models included in the 75% (p-values smaller than 0.25) and 50% MCS (p-values smaller than 0.5) are identified by light and dark shades, respectively.

The first observation from Table 6 is that CM-based bagging methods can indeed improve forecasting accuracy. In 25 out of the 30 cases we considered, CMBG and CMBGA improve upon both the better of the restricted and the unrestricted model, demonstrating the power of shrinkage. CMBG and CMBGA also have lower MSE than their PT-based versions in 27 out of the 30 cases—evidence that shrinking towards a nonzero coefficient can be advantageous when sign restrictions can be imposed.

Second, within the CM-based family, the differences between three methods are small, with MSEs from CMBGA and CMBG particularly similar to each other. Thus, the asymptotic version CMBGA is a computationally attractive alternative to the bootstrap-based predictor CMBG.

Third, between the two predictors, INDPRO has higher predictive power than UNEMP when the estimation window is set to 15 or 20 years but UNEMP fares better with a 10-year window. For $T = 40$, forecasts based on UNEMP improved upon the random walk model for forecast horizons up to one year. For $T = 60$ or 80, forecasts based on INDPRO produce lower MSEs than those based on UNEMP, and they produce a smaller number of final models that are in either the 75% or the 50% MCS, especially with longer forecast horizons ($h \geq 3$).

Fourth, the performance of ridge and Lasso is similar to that of PT-based estimators. Since both PT and ridge/Lasso shrink the coefficients towards zero rather than a positive coefficient, the result is not surprising. Ridge and Lasso, however, are more likely to be included in the model confidence set than those based on PT, possible due to less correlation between the losses produced by ridge and Lasso and the losses produced by other models in the set.

Last but not least, in line with the existing literature on inflation forecasting, it is difficult to beat the RW model statistically. The RW model is included in the 75% MCS in all but two cases ($T = 80$, $h = 1$ or 6, $x =$ INDPRO). Also, the RW forecasts have the lowest MSE when the estimation window is short and the forecast horizon is long. With longer estimation window or short forecast horizon, however, forecasts that uses activity variables fare better than the RW model.

Figures 4 displays the rolling β_h from model UR and CMBGA for forecasting 4-quarter inflation. For $T = 40$, the coefficients from unrestricted estimation are volatile and they change signs for both predictor variables. CMBGA imposes the sign restriction by construction and shrinks the coefficients towards a positive constant, which results in more stable coefficients. For $T = 80$, the coefficients from unrestricted model are less volatile. CMBGA again shrinks the coefficients, resulting in coefficients that are smaller than those from UR in most periods.

Overall the proposed method works well for inflation forecasting, although 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 modest.

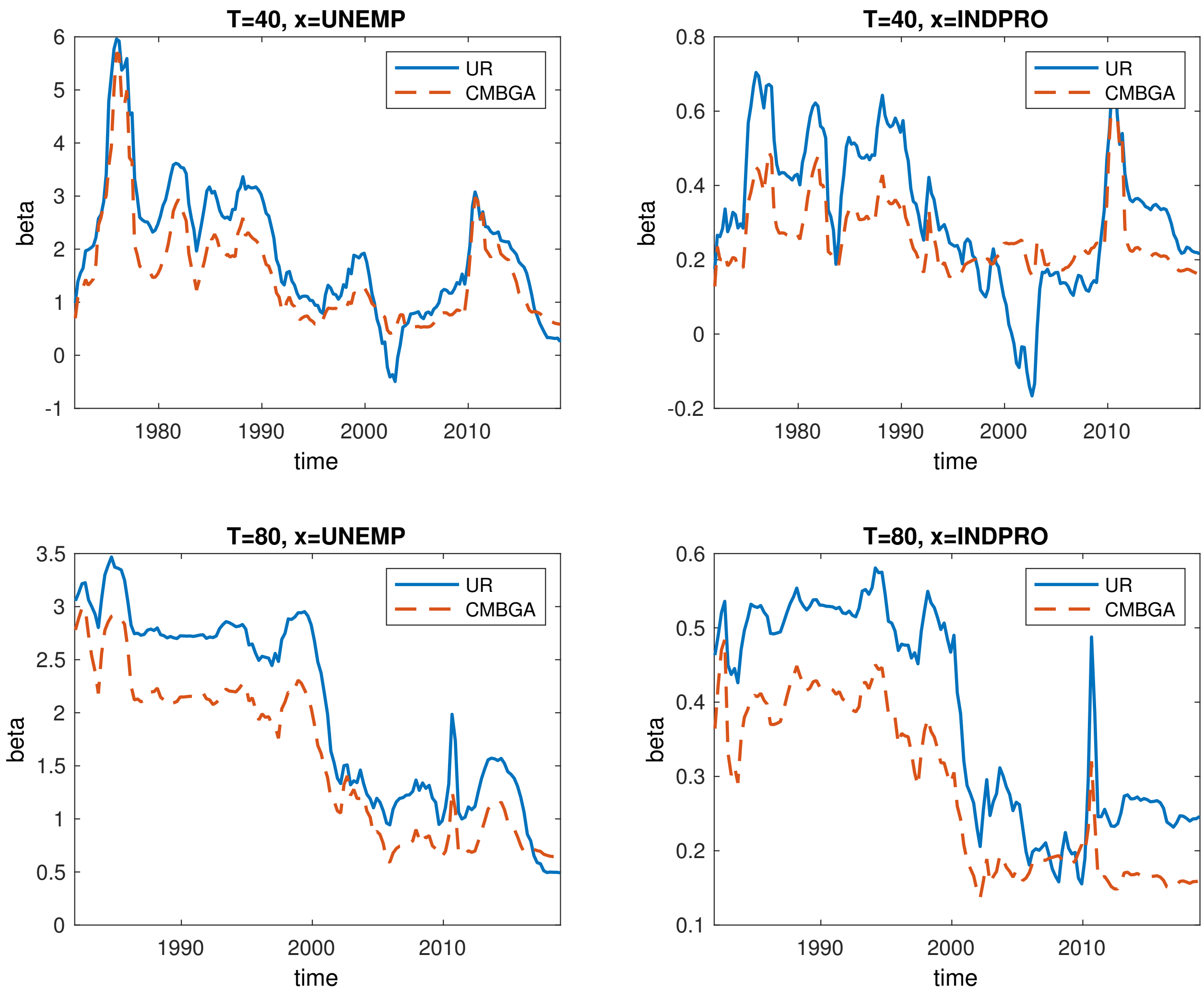

Figure 4: Rolling coefficients for model UR and CMBGA. Forecast horizon $h = 4$ and significance level = 1%.

5 Conclusion

Bootstrap aggregation (bagging) is often applied to t-tests of whether coefficients are significantly different from zero. In finite samples, a significant non-zero coefficient is not sufficient to guarantee that including the predictor improves forecast accuracy. Instead, estimation uncertainty 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 for inflation, as economic theory often predicts relationships between inflation and individual variables. We conduct a pseudo out-of-sample forecasting horserace for inflation, and we find that including either unemployment or industrial production and accounting for in-sample estimation uncertainty can improve upon the univariate random walk model, especially when the estimation window is between 15 to 20 years.

Econometric forecasting models, however, often 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.

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

	MSE								Rejection %	
	RE	UR	PT	BG	BGA	CM	CMBG	CMBGA	t-test	CM-test
Panel 1 : $b\sigma_\infty^{-1} = 0.0$										
$h = 1$										
$T = 25$	3.98	9.56	6.14	6.47	6.32	10.27	11.15	10.85	7.8	1.6
$T = 50$	2.06	4.36	2.80	2.89	2.86	4.56	4.89	4.83	6.5	1.1
$T = 200$	0.50	1.02	0.63	0.66	0.65	1.05	1.13	1.12	5.4	0.6
$h = 6$										
$T = 25$	10.71	20.00	15.31	15.45	15.29	18.86	20.13	19.59	15.4	6.5
$T = 50$	5.24	9.57	6.79	6.93	6.86	8.97	9.63	9.41	10.6	3.2
$T = 200$	1.36	2.40	1.64	1.70	1.68	2.34	2.50	2.46	7.5	1.2
Panel 2 : $b\sigma_\infty^{-1} = 0.5$										
$h = 1$										
$T = 25$	5.60	9.60	7.93	6.86	6.87	7.11	7.82	7.58	18.5	5.0
$T = 50$	2.63	4.22	3.50	3.01	3.02	3.07	3.35	3.30	15.7	3.2
$T = 200$	0.63	1.01	0.80	0.71	0.70	0.70	0.77	0.77	12.8	2.0
$h = 6$										
$T = 25$	13.29	20.10	17.44	16.05	16.23	16.23	16.91	16.56	29.7	14.1
$T = 50$	6.58	9.89	8.53	7.67	7.76	7.54	7.98	7.83	23.8	8.5
$T = 200$	1.68	2.45	2.05	1.86	1.85	1.78	1.94	1.90	16.9	3.7
Panel 3 : $b\sigma_\infty^{-1} = 1.0$										
$h = 1$										
$T = 25$	10.07	9.75	10.75	8.07	8.20	6.70	6.64	6.56	35.8	13.0
$T = 50$	4.45	4.39	4.83	3.61	3.65	2.90	2.93	2.91	30.3	8.6
$T = 200$	1.03	1.04	1.13	0.86	0.86	0.65	0.67	0.67	27.4	6.0
$h = 6$										
$T = 25$	20.61	19.65	19.95	16.86	17.49	15.88	15.26	15.32	49.2	28.1
$T = 50$	9.96	9.41	9.97	8.17	8.39	7.31	7.13	7.12	40.1	17.5
$T = 200$	2.48	2.40	2.59	2.09	2.10	1.73	1.74	1.71	32.8	9.9
Panel 4 : $b\sigma_\infty^{-1} = 2.0$										
$h = 1$										
$T = 25$	28.19	9.31	13.37	10.10	10.23	10.62	8.11	8.33	71.6	42.1
$T = 50$	11.98	4.38	6.17	4.68	4.72	4.77	3.66	3.72	68.0	35.0
$T = 200$	2.63	1.03	1.51	1.14	1.13	1.14	0.86	0.87	65.2	28.8
$h = 6$										
$T = 25$	51.55	20.11	23.57	20.16	20.90	21.17	18.24	18.90	80.4	61.2
$T = 50$	23.44	9.82	12.21	10.08	10.40	10.52	8.78	9.04	75.5	49.0
$T = 200$	5.57	2.38	3.14	2.51	2.56	2.64	2.11	2.14	69.7	36.9
Panel 5 : $b\sigma_\infty^{-1} = 4.0$										
$h = 1$										
$T = 25$	100.38	9.78	10.58	10.48	10.41	11.87	10.68	10.64	98.4	91.2
$T = 50$	41.50	4.55	4.90	4.91	4.86	5.49	5.02	4.98	98.5	90.4
$T = 200$	9.02	1.03	1.08	1.11	1.10	1.28	1.17	1.16	98.9	90.5
$h = 6$										
$T = 25$	170.43	19.51	20.01	19.59	20.00	21.01	19.80	20.20	99.1	95.5
$T = 50$	79.24	10.00	10.25	10.18	10.35	11.22	10.43	10.64	99.1	94.1
$T = 200$	18.88	2.48	2.57	2.59	2.61	2.90	2.68	2.73	99.2	93.1

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

Table 3: Monte Carlo Results for $\phi = 0.66$ and $c_{0.99}$

	MSE								Rejection %	
	RE	UR	PT	BG	BGA	CM	CMBG	CMBGA	t-test	CM-test
Panel 1 : $b\sigma_\infty^{-1} = 0.0$										
$h = 1$										
$T = 25$	3.98	9.35	5.74	5.76	5.70	9.90	10.33	10.12	2.5	0.6
$T = 50$	2.01	4.44	2.47	2.55	2.52	4.45	4.64	4.57	1.9	0.2
$T = 200$	0.51	1.08	0.57	0.60	0.59	1.04	1.07	1.07	1.2	0.1
$h = 6$										
$T = 25$	10.60	20.16	14.41	14.36	14.25	18.13	19.14	18.50	8.5	3.2
$T = 50$	5.36	9.29	6.37	6.41	6.32	8.82	9.28	9.03	4.7	1.2
$T = 200$	1.35	2.46	1.53	1.57	1.55	2.24	2.33	2.30	2.4	0.3
Panel 2 : $b\sigma_\infty^{-1} = 0.5$										
$h = 1$										
$T = 25$	5.57	9.80	7.49	6.46	6.60	6.59	6.98	6.77	8.0	1.9
$T = 50$	2.60	4.33	3.22	2.83	2.85	2.93	3.12	3.07	5.3	0.9
$T = 200$	0.63	1.05	0.73	0.67	0.66	0.66	0.70	0.70	3.9	0.4
$h = 6$										
$T = 25$	13.37	20.50	17.50	15.49	16.06	15.47	15.76	15.48	17.9	8.6
$T = 50$	6.49	9.78	7.80	6.97	7.12	6.81	7.09	6.91	11.3	3.9
$T = 200$	1.62	2.39	1.86	1.70	1.69	1.63	1.74	1.70	6.6	1.0
Panel 3 : $b\sigma_\infty^{-1} = 1.0$										
$h = 1$										
$T = 25$	9.89	10.10	11.53	8.42	8.76	6.55	6.38	6.30	18.3	5.6
$T = 50$	4.57	4.56	5.17	3.83	3.91	2.72	2.75	2.72	13.7	2.9
$T = 200$	1.01	0.99	1.11	0.84	0.84	0.54	0.57	0.57	10.0	1.1
$h = 6$										
$T = 25$	20.56	19.95	21.70	17.22	18.35	15.77	14.85	15.08	33.3	17.5
$T = 50$	10.06	9.74	10.71	8.44	8.81	7.08	6.90	6.87	23.6	9.6
$T = 200$	2.42	2.45	2.67	2.12	2.15	1.60	1.61	1.59	14.8	3.6
Panel 4 : $b\sigma_\infty^{-1} = 2.0$										
$h = 1$										
$T = 25$	28.39	9.47	17.59	11.58	12.12	11.74	8.42	8.90	51.5	25.6
$T = 50$	11.70	4.35	8.21	5.41	5.55	5.15	3.74	3.86	46.0	18.0
$T = 200$	2.57	1.03	2.03	1.32	1.32	1.13	0.86	0.86	39.2	11.4
$h = 6$										
$T = 25$	50.71	20.28	27.80	21.44	23.05	23.07	18.50	19.84	67.4	46.3
$T = 50$	23.83	9.96	15.07	11.18	11.80	11.71	9.17	9.65	58.5	33.2
$T = 200$	5.65	2.46	4.19	2.92	3.02	2.81	2.18	2.23	46.4	18.1
Panel 5 : $b\sigma_\infty^{-1} = 4.0$										
$h = 1$										
$T = 25$	98.83	9.72	13.36	12.02	11.99	16.94	12.34	12.65	94.2	79.9
$T = 50$	42.03	4.32	5.92	5.42	5.41	7.51	5.76	5.84	94.2	78.2
$T = 200$	8.95	1.03	1.38	1.31	1.30	1.89	1.44	1.44	94.4	75.4
$h = 6$										
$T = 25$	169.83	19.95	22.57	21.10	21.75	25.53	21.81	22.70	97.1	90.2
$T = 50$	78.49	9.93	11.61	10.87	11.20	13.74	11.33	11.85	96.6	86.8
$T = 200$	17.99	2.28	2.65	2.58	2.63	3.43	2.77	2.88	96.4	82.2

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

Table 4: Monte Carlo Results for $\phi = 0.9$ and $c_{0.95}$

	MSE								Rejection %	
	RE	UR	PT	BG	BGA	CM	CMBG	CMBGA	t-test	CM-test
Panel 1 : $b\sigma_\infty^{-1} = 0.0$										
$h = 1$										
$T = 25$	3.97	11.10	10.20	9.72	9.58	18.08	18.97	18.61	7.8	1.7
$T = 50$	2.00	5.10	3.68	3.69	3.65	6.37	6.77	6.70	6.5	0.9
$T = 200$	0.50	1.06	0.68	0.70	0.70	1.14	1.22	1.21	5.7	0.7
$h = 6$										
$T = 25$	10.74	23.32	22.18	21.74	21.25	30.37	32.08	30.96	16.8	7.4
$T = 50$	5.37	12.25	9.42	9.47	9.30	13.26	14.19	13.78	12.7	4.0
$T = 200$	1.36	2.74	1.91	1.96	1.93	2.76	3.00	2.94	8.3	1.7
Panel 2 : $b\sigma_\infty^{-1} = 0.5$										
$h = 1$										
$T = 25$	6.42	10.92	11.55	9.14	9.17	12.41	12.90	12.63	19.4	5.6
$T = 50$	2.85	5.06	4.49	3.75	3.74	4.26	4.56	4.49	16.6	3.5
$T = 200$	0.65	1.09	0.88	0.76	0.75	0.76	0.84	0.83	13.2	2.1
$h = 6$										
$T = 25$	15.39	25.18	24.25	21.43	21.71	24.74	25.01	24.42	33.2	17.4
$T = 50$	7.38	11.92	10.68	9.34	9.40	9.90	10.38	10.12	25.9	10.2
$T = 200$	1.73	2.68	2.26	1.99	1.98	1.90	2.08	2.03	19.2	4.8
Panel 3 : $b\sigma_\infty^{-1} = 1.0$										
$h = 1$										
$T = 25$	13.47	11.38	14.79	10.62	10.74	11.82	10.97	10.83	38.6	15.0
$T = 50$	5.49	4.83	5.77	4.24	4.26	3.91	3.76	3.72	32.4	10.1
$T = 200$	1.10	1.08	1.22	0.91	0.91	0.70	0.70	0.70	27.8	6.4
$h = 6$										
$T = 25$	28.62	24.23	26.93	22.02	22.92	23.70	21.78	21.88	52.0	31.9
$T = 50$	13.31	12.26	13.39	10.78	11.06	10.24	9.75	9.67	45.8	23.4
$T = 200$	2.90	2.75	2.97	2.35	2.36	1.95	1.93	1.89	34.9	11.4
Panel 4 : $b\sigma_\infty^{-1} = 2.0$										
$h = 1$										
$T = 25$	41.06	10.87	15.44	11.75	11.82	13.82	10.42	10.50	73.9	46.3
$T = 50$	16.70	5.04	7.36	5.55	5.56	6.04	4.54	4.58	70.0	38.7
$T = 200$	2.85	1.06	1.52	1.15	1.15	1.18	0.89	0.89	66.3	29.8
$h = 6$										
$T = 25$	83.50	24.91	28.85	25.15	25.85	27.78	23.82	24.48	83.2	66.4
$T = 50$	37.29	12.32	15.14	12.72	13.03	13.54	11.32	11.59	78.5	56.5
$T = 200$	7.32	2.80	3.66	2.95	3.00	3.15	2.47	2.50	72.2	41.0
Panel 5 : $b\sigma_\infty^{-1} = 4.0$										
$h = 1$										
$T = 25$	158.72	11.25	12.73	12.18	12.04	14.17	12.37	12.24	97.5	90.1
$T = 50$	58.95	5.04	5.44	5.46	5.39	6.37	5.60	5.54	98.0	89.4
$T = 200$	9.94	1.07	1.13	1.14	1.14	1.30	1.20	1.19	98.4	90.2
$h = 6$										
$T = 25$	305.36	24.91	25.95	25.26	25.62	27.46	25.61	25.94	98.9	95.3
$T = 50$	132.11	11.77	12.17	12.03	12.13	12.98	12.26	12.36	98.7	94.2
$T = 200$	24.95	2.77	2.86	2.89	2.90	3.22	2.98	3.02	99.0	93.3

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	BG	BGA	CM	CMBG	CMBGA	t-test	CM-test
Panel 1 : $b\sigma_{\infty}^{-1} = 0.0$										
$h = 1$										
$T = 25$	4.02	11.20	9.65	9.03	8.97	17.96	18.48	18.11	2.8	0.6
$T = 50$	1.98	4.91	3.34	3.25	3.22	6.11	6.29	6.22	1.8	0.2
$T = 200$	0.51	1.10	0.59	0.62	0.61	1.14	1.17	1.17	1.0	0.1
$h = 6$										
$T = 25$	10.88	24.68	22.72	21.57	21.34	30.87	32.14	31.01	9.3	4.0
$T = 50$	5.40	12.11	8.90	8.77	8.68	12.86	13.45	13.06	6.2	1.8
$T = 200$	1.38	2.79	1.71	1.76	1.72	2.65	2.79	2.73	2.8	0.4
Panel 2 : $b\sigma_{\infty}^{-1} = 0.5$										
$h = 1$										
$T = 25$	6.43	10.86	11.62	8.89	9.13	11.43	11.53	11.28	8.1	2.0
$T = 50$	2.98	5.26	4.51	3.67	3.71	4.16	4.31	4.25	5.9	0.9
$T = 200$	0.63	1.07	0.78	0.68	0.68	0.70	0.75	0.74	4.2	0.3
$h = 6$										
$T = 25$	15.09	24.59	25.00	20.95	21.72	23.83	23.50	23.04	21.9	10.3
$T = 50$	7.34	11.80	10.53	8.73	9.06	9.24	9.47	9.22	13.7	5.3
$T = 200$	1.75	2.80	2.21	1.96	1.96	1.90	2.02	1.96	7.4	1.5
Panel 3 : $b\sigma_{\infty}^{-1} = 1.0$										
$h = 1$										
$T = 25$	13.82	11.40	17.64	11.61	12.19	11.47	10.22	10.27	20.2	6.6
$T = 50$	5.60	5.14	6.74	4.64	4.73	3.71	3.52	3.49	16.3	3.6
$T = 200$	1.11	1.09	1.28	0.94	0.94	0.63	0.64	0.64	10.6	1.7
$h = 6$										
$T = 25$	28.89	24.23	30.38	22.74	24.53	24.52	21.32	21.80	38.9	22.4
$T = 50$	13.06	12.13	14.64	10.86	11.55	9.86	9.11	9.14	27.7	12.3
$T = 200$	2.84	2.76	3.17	2.42	2.47	1.80	1.79	1.75	17.6	4.8
Panel 4 : $b\sigma_{\infty}^{-1} = 2.0$										
$h = 1$										
$T = 25$	42.38	11.28	23.68	15.01	15.69	17.56	11.80	12.50	54.9	30.2
$T = 50$	16.09	4.89	9.94	6.40	6.53	6.74	4.74	4.88	49.2	22.1
$T = 200$	2.88	1.06	2.13	1.39	1.38	1.26	0.93	0.93	40.8	12.6
$h = 6$										
$T = 25$	86.55	25.04	37.00	27.62	29.68	32.77	25.27	27.10	72.1	54.5
$T = 50$	36.41	12.30	19.10	14.07	14.96	15.84	11.83	12.62	63.3	40.5
$T = 200$	7.19	2.83	4.85	3.36	3.50	3.43	2.57	2.65	51.4	23.4
Panel 5 : $b\sigma_{\infty}^{-1} = 4.0$										
$h = 1$										
$T = 25$	159.41	11.26	16.10	13.93	13.77	19.53	14.49	14.71	93.2	81.2
$T = 50$	58.95	4.87	6.72	6.02	6.02	8.60	6.37	6.47	92.9	77.7
$T = 200$	9.84	1.08	1.43	1.35	1.35	1.94	1.46	1.46	93.7	75.1
$h = 6$										
$T = 25$	307.42	25.20	29.95	26.94	27.52	32.31	27.82	28.55	97.0	91.6
$T = 50$	129.44	11.96	13.82	13.08	13.26	16.24	13.50	13.95	96.2	87.8
$T = 200$	25.26	2.83	3.41	3.23	3.29	4.31	3.43	3.58	96.0	83.4

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

Table 6: Out-of-sample inflation forecasting

	$T = 40$									
	UNEMP					INDPRO				
$h =$	1	2	3	4	6	1	2	3	4	6
RW	1.000	1.000	1.000	1.000	**1.000**	1.000	1.000	**1.000**	**1.000**	**1.000**
RE	0.978	1.012	1.085	1.132	1.122	0.978	1.012	1.085	1.132	1.122
UR	0.893	0.936	0.982	0.990	1.041	0.927	1.007	1.025	1.011	1.001
PT	0.892	0.944	1.034	1.025	1.038	0.981	1.156	1.163	1.109	1.010
PTBG	0.893	0.940	1.002	1.011	1.033	0.916	1.032	1.074	1.052	1.011
PTBGA	0.892	0.931	0.989	0.997	1.038	0.940	1.031	1.059	1.046	1.014
CMPT	0.906	0.949	0.979	1.013	1.111	0.924	**0.973**	1.036	1.056	1.038
CMBG	0.884	0.937	0.991	1.000	1.029	**0.879**	0.989	1.026	1.011	1.000
CMBGA	0.883	**0.930**	**0.975**	0.986	1.035	0.891	0.987	1.022	1.021	1.007
Ridge	0.893	0.933	0.982	0.991	1.040	0.929	1.006	1.033	1.017	1.006
Lasso	**0.882**	0.958	0.988	**0.979**	1.054	0.908	0.998	1.042	1.034	1.021

	$T = 60$									
	UNEMP					INDPRO				
$h =$	1	2	3	4	6	1	2	3	4	6
RW	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
RE	0.946	**0.910**	0.926	0.952	0.984	0.946	0.910	0.926	0.952	0.984
UR	0.913	0.947	0.932	0.886	0.914	0.925	0.940	0.895	0.874	0.872
PT	0.933	0.948	0.931	0.910	0.929	0.891	0.946	0.930	0.879	0.863
PTBG	0.897	0.931	0.930	0.890	0.909	0.897	0.924	0.901	0.865	0.846
PTBGA	0.902	0.935	0.931	0.894	0.910	0.901	0.929	0.898	0.874	0.858
CMPT	0.893	0.972	0.969	0.915	**0.859**	0.881	**0.857**	**0.851**	0.866	0.839
CMBG	**0.891**	0.927	0.920	0.887	0.888	0.867	0.888	0.865	0.843	**0.829**
CMBGA	0.895	0.927	**0.912**	**0.879**	0.895	**0.865**	0.881	0.853	**0.837**	0.836
Ridge	0.911	0.945	0.935	0.887	0.915	0.913	0.952	0.900	0.881	0.876
Lasso	0.914	0.951	0.939	0.900	0.922	0.913	0.963	0.907	0.888	0.874

	$T = 80$									
	UNEMP					INDPRO				
$h =$	1	2	3	4	6	1	2	3	4	6
RW	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
RE	0.930	0.893	0.882	**0.894**	0.942	0.930	0.893	0.882	0.894	0.942
UR	0.920	0.939	0.926	0.934	0.966	0.915	0.926	0.869	0.848	0.820
PT	**0.867**	**0.876**	**0.875**	0.936	0.953	**0.864**	0.892	0.842	0.868	0.858
PTBG	0.895	0.900	0.904	0.922	0.947	0.891	0.896	0.854	0.846	0.804
PTBGA	0.896	0.892	0.898	0.921	0.942	0.897	0.897	0.852	0.846	0.809
CMPT	0.902	0.926	0.920	0.906	0.939	0.882	**0.870**	**0.817**	0.846	**0.738**
CMBG	0.899	0.911	0.910	0.918	0.925	0.873	0.889	0.842	**0.818**	0.766
CMBGA	0.901	0.906	0.898	0.908	**0.917**	0.873	0.880	0.834	0.819	0.773
Ridge	0.917	0.933	0.921	0.928	0.962	0.922	0.927	0.868	0.849	0.822
Lasso	0.893	0.926	0.947	0.945	0.925	0.911	0.957	0.910	0.858	0.839

Note: this table reports the MSE of the Phillips curve model in (24) relative to the MSE of the random walk (RW) model in (22). Estimators for model (24) include restricted (RE), unrestricted (UR), pre-test using a t-test (PT), bagging t-test (PTBG), asymptotic bagging t-test (PTBGA), pre-test using a CM-test (CMPT), Bagging CM-test (CMBG) and asymptotic bagging CM-test (CMBGA) , ridge regression and Lasso regression. See Table 1 and Equation 25 for specifications for these estimators. Models included in the 75% (p-values smaller than 0.25) and 50% MCS (p-values smaller than 0.5) are identified by light and dark shades, respectively.

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.

Ball, L. and S. Mazumder (2019): "A Phillips Curve with Anchored Expectations and Short-Term Unemployment," *Journal of Money, Credit and Banking*, 51, 111–137.

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.

Carrasco, M. and B. Rossi (2016): "In-Sample Inference and Forecasting in Misspecified Factor Models," *Journal of Business & Economic Statistics*, 34, 313–338.

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.

Faust, J. and J. H. Wright (2013): "Chapter 1 - Forecasting Inflation," in *Handbook of Economic Forecasting*, ed. by G. Elliott and A. Timmermann, Elsevier, vol. 2, Part A of *Handbook of Economic Forecasting*, 2 – 56.

Galí, J. (2015): *Monetary policy, inflation, and the business cycle: an introduction to the new Keynesian framework and its applications*, Princeton University Press.

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.

Gordon, R. J. (2013): "The Phillips Curve is Alive and Well: Inflation and the NAIRU During the Slow Recovery," Working Paper 19390, National Bureau of Economic Research.

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.

Jin, S., L. Su, and A. Ullah (2014): “Robustify Financial Time Series Forecasting with Bagging,” *Econometric Reviews*, 33, 575–605.

Jordan, S. J., A. Vivian, and M. E. Wohar (2017): “Forecasting market returns: bagging or combining?” *International Journal of Forecasting*, 33, 102 – 120.

Kapetanios, G. and F. Zikes (2018): “Time-varying Lasso,” *Economics Letters*, 169, 1 – 6.

Kim, H. H. and N. R. Swanson (2014): “Forecasting financial and macroeconomic variables using data reduction methods: New empirical evidence,” *Journal of Econometrics*, 178, 352 – 367, recent Advances in Time Series Econometrics.

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.

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. and A. Timmermann (2017): “Forecasting Macroeconomic Variables Under Model Instability,” *Journal of Business & Economic Statistics*, 35, 183–201.

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.

Rossi, B. (2013): "Chapter 21 - Advances in Forecasting under Instability," in *Handbook of Economic Forecasting*, ed. by G. Elliott and A. Timmermann, Elsevier, vol. 2 of *Handbook of Economic Forecasting*, 1203 – 1324.

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.

Watson, M. W. (2014): "Inflation Persistence, the NAIRU, and the Great Recession," *American Economic Review*, 104, 31–36.

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

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

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

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

Next consider the bagged version

$$T^{1/2}\hat{\sigma}_{\infty}^{-1}\hat{\beta}^{CMBG} = \frac{1}{B}\sum_{b=1}^{B}\left[T^{1/2}\hat{\sigma}_{\infty}^{-1}\hat{\beta}_b^{*}\mathbf{I}(\hat{\tau}_b^{*} > c) + \mathbf{I}(\hat{\tau}_b^{*} \le c)\right]. \tag{26}$$

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,

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

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:

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