Forecasting Equity Premium: Global Historical Average versus Local Historical Average and Constraints[*]

Tae-Hwy Lee[†]
Department of Economics
University of California, Riverside

Yundong Tu[‡]
Guanghua School of Management and
Center for Statistical Science
Peking University

Aman Ullah[§]
Department of Economics
University of California, Riverside

First version: September 2012
This version: July 2014

Abstract

The equity premium, return on equity minus return on risk-free asset, is expected to be positive. We consider imposing such positivity constraint in local historical average (LHA) in nonparametric kernel regression framework. It is also extended to the semiparametric single index model when multiple predictors are used. We construct the constrained LHA estimator via an indicator function which operates as 'model-selection' between the unconstrained LHA and the bound of the constraint (zero for the positivity constraint). We smooth the indicator function by bagging (Breiman 1996a), which operates as 'model-averaging' and yields a combined forecast of unconstrained LHA forecasts and the bound of the constraint. The local combining weights are determined by the probability that the constraint is binding. Asymptotic properties of the constrained LHA estimators without and with bagging are established, which show how the positive constraint and bagging can help reduce the asymptotic variance and mean squared errors. Monte Carlo simulations are conducted to show the finite sample behavior of the asymptotic properties. In predicting U.S. equity premium, we show that substantial nonlinearity can be captured by LHA and that the local positivity constraint can improve out-of-sample prediction of the equity premium.

Key Words: Equity premium; Nonparametric local historical average model; Positivity constraint; Bagging; Model averaging; Semiparametric single index model.

JEL Classification: C14, C50, C53, G17.

1 Introduction

Goyal and Welch (GW 2008) show that the historical average (HA) forecast of the equity premium (excess return on equity over return on risk-free asset) performs better than forecasts from the predictive regression using covariates (predictors). GW find that numerous economic predictor variables with in-sample significance for the excess stock returns fail to deliver out-of-sample forecasting gains relative to the HA. In GW the benchmark model to beat in out-of-sample forecasting was the 'global historical average' (GHA), which is formed from the sample average of the historical equity premium time series over rolling fixed windows or expanding windows.

While the literature has generally confirmed that it is very hard to beat GHA, there are a few limited demonstrations of some success in beating this simple benchmark GHA. In particular we note the following three approaches here. The first one is Campbell and Thompson (CT 2008), who asked a question in their paper title, "Predicting the equity premium out of sample: Can anything beat the historical average?". They argued that the answer to this question can be "Yes" if theoretically motivated constraints (e.g., monotonicity, positivity) are imposed on the predictive regression function. CT found that the predictive regression models with some sensible constraints can do better than GHA. The second one is Hillebrand, Lee and Medeiros (2009), who use bagging to smooth the CT's constraint and found that bagging can further improve CT's constrained predictive regression forecasts. The third one is Chen and Hong (2009), who show that the nonparametric nonlinear forecasts are better than the parametric linear regression forecasts.

This paper extends the above literature by putting all of these three approaches together. First, following Chen and Hong (2009), we consider nonparametric local models to explore if an LHA model can beat the GHA model.[1] The answer from our empirical analysis (Section 6) is clearly "Yes" using the same data set used in CT (2008). We find that LHA can easily beat GHA for many predictors (especially for the annualized equity premium in monthly frequency).[2] Second, following CT (2008), we consider imposing the *local* positivity constraint on the LHA equity premium forecast to explore if the constraint can improve the LHA. The answer in Section 6 is also

---

[1] Chen and Hong (2009) use the local linear model, while we use the Nadaraya-Watson local constant model.

[2] See Section 6 for the definition of annualized excess return in monthly frequency. $Q_t(k)$ in equation (18) with $k = 12$ is the annualized excess return in month t. $Q_t(k)$ with $k = 1$ is the monthly excess return in month t.

"Yes" for almost all 11 predictors for the annualized excess returns and also for the monthly excess returns. Third, following Hillebrand, Lee and Medeiros (2009), we consider bagging to explore if smoothing of the constraint can further help the positivity-constrained LHA. The answer given in Section 6 to this possibility is again "Yes" for most of the 11 predictors for the annualized equity premium and for the monthly excess returns. In summary, these three considerations give us three models – the LHA forecast, the positivity-constrained LHA forecast (denoted as LHA_p), and the bagged positivity-constrained LHA forecast (denoted as LHA_{pb}). LHA_{pb} is the new equity premium forecast using all three features (local, constrained, and bagged).

The rest of the paper is organized as follows. Section 2 presents four HA models – namely, GHA, LHA, LHA_p, and LHA_{pb}. In Section 3 we derive the asymptotic properties of LHA_p (Theorem 1) and LHA_{pb} (Theorem 2). We also show that LHA_p yields local 'model-selection' between LHA model and the bound of the constraint while LHA_{pb} operates as local 'model-averaging' of LHA and the bound of the constraint with the model-averaging weights determined by the probability that the constraint binds (Theorem 3). Extension to models with multivariate predictors is considered in Section 4. Section 5 examines the finite-sample properties of these models via Monte Carlo simulations. Section 6 evaluates their predictions of equity premium. Section 7 concludes. The Appendix collects all the technical proofs.

2   Historical Average Models

First, we consider the GHA model for the equity premium y as

$$y_{t+1} = \alpha + u_{t+1}, \tag{1}$$

where u_{t+1} is a disturbance term such that $E\left(u_{t+1}\right) = 0, t = 1, ..., n$. The least square estimator of the GHA α,

$$\text{GHA}: \quad \tilde{\alpha} = \frac{1}{n}\sum_{t=1}^{n} y_t, \tag{2}$$

is the unconstrained parametric estimator of α. Note that $\tilde{\alpha}$ is a random variable which is asymptotically normal with mean α.

Now we consider GHA and LHA models for the equity premium. The equity premium is the difference between returns on risky equity and risk-free assets. As the equity premium is the risk premium for the investment on the risky equity, it is expected to be positive, that is $\alpha > 0$. CT (2008) considered imposing such positivity constraint on the linear parametric (global) predictive regression model where the equity premium y is predicted using a predictor x.

We consider such positivity constraint on the LHA model in nonparametric kernel regression framework. Let y be the variable to forecast and x be a predictor. For the ease of exposition, we first consider the LHA model where x contains one regressor. The case in which the predictor x is multivariate is treated in Section 4. Let $I_n = \{x_{t-1}, y_t\}_{t=1}^n$ be an observed training sample (drawn from a stationary process) at time $t = n$ to train the LHA $\alpha(x)$, and x_t is the value of x at time t. The LHA model is

$$y_{t+1} = \alpha(x_t) + u_{t+1}, \tag{3}$$

where $\alpha(x_t) = E(y_{t+1}|x_t)$, u_{t+1} is a disturbance term such that $E(u_{t+1}|x_t) = 0$ by construction, $t = 1, ..., n$. The LHA is the local constant kernel estimator of $\alpha(x)$ trained using I_n

$$\text{LHA}: \quad \tilde{\alpha}(x) = \frac{\sum_{t=1}^n k_h(x_{t-1} - x) y_t}{\sum_{t=1}^n k_h(x_{t-1} - x)}, \tag{4}$$

where h is a bandwidth, $k_h(\cdot) = k(\cdot/h)$, and $k(\cdot)$ is a kernel function. $\tilde{\alpha}(x)$ is shown to be asymptotically normal, c.f. Pagan and Ullah (1999). The LHA equity premium forecast at time n using the predictor value $x = x_n$ is $\tilde{\alpha}(x_n)$.

We construct the constrained estimator via an indicator function. The indicator function selects either the unconstrained LHA or the bound of the constraint (zero for the positivity constraint) as a forecast of the equity premium. We consider the constraint that the LHA of y conditional on x, $\alpha(x) = E(y|x)$, exceeds some known lower bound, $\alpha_1(x)$. That is,

$$\alpha(x) > \alpha_1(x). \tag{5}$$

This information is assumed to be known as a prior to a forecaster. Under this constraint (5), we can easily estimate $\alpha(x)$ with

$$\text{LHA}_p: \quad \bar{\alpha}(x) = \tilde{\alpha}(x) 1_{[\tilde{\alpha}(x) > \alpha_1(x)]} + \alpha_1(x) 1_{[\tilde{\alpha}(x) \leq \alpha_1(x)]}. \tag{6}$$

In the empirical example of this paper, we consider the constraint with the constant bound $\alpha_1(x) = 0$, making $\bar{\alpha}(x)$ the LHA with positivity constraint (denoted as LHA_p). Note that LHA_p operates as local 'model-selection' between LHA $\tilde{\alpha}(x)$ and $\alpha_1(x) = 0$ (the martingale difference, MD, model).

The LHA_p estimator $\bar{\alpha}(x)$ involves an indicator and is not stable in the sense of Breiman (1996b) and Bühlmann and Yu (2002). Following Bühlmann and Yu (2002), we smooth the indicator function by bagging (Breiman 1996a). To define the "bagging positivity-constrained L-HA' of $\alpha(x)$, we construct a bootstrap sample $\{x_{t-1}^*, y_t^*\}_{t=1}^n$ which is used to derive a bootstrap constrained estimator via (6) using the plug-in principle. The bagging predictor is an expectation of this estimator over the bootstrapped samples. To be precise, denote $\tilde{\alpha}^{*(j)}(x)$ as the unconstrained estimator of $\alpha(x)$ computed from the j-th bootstrapped sample $\left\{x_{t-1}^{*(j)}, y_t^{*(j)}\right\}_{t=1}^n$, $j = 1, \ldots, J$. Then the plug-in constrained estimator in the j-th bootstrap sample $\bar{\alpha}^{*(j)}(x) = \tilde{\alpha}^{*(j)}(x) 1_{\left[\tilde{\alpha}^{*(j)}(x) > \alpha_1(x)\right]} + \alpha_1(x) 1_{\left[\tilde{\alpha}^{*(j)}(x) \leq \alpha_1(x)\right]}$. We define the bagging positivity-constrained LHA estimator (denoted as LHA_{pb}) as

$$\text{LHA}_{pb}: \quad \hat{\alpha}(x) = \frac{1}{J} \sum_{j=1}^J \bar{\alpha}^{*(j)}(x) := E^* \bar{\alpha}^*(x). \tag{7}$$

in line with that of Breiman (1996a).

In the next section, we will show that

$$\hat{\alpha}(x) \approx w(x)\tilde{\alpha}(x) + (1 - w(x))\alpha_1(x). \tag{8}$$

In Theorem 3, we show that the combining weight $w(x)$ is the limiting probability that the local positivity-constraint is non-binding. If $w(x) > 0$, bagging operates as convex 'model-averaging' locally instead of as 'model-selection' and yields a combined forecast of unconstrained LHA $\tilde{\alpha}(x)$ and the bound of the constraint $\alpha_1(x)$. The underlying reason for the benefit of imposing the (correct) constraint is the 'shrinkage' principle with $(1 - w(x))$ being the extent of the shrinkage towards the bound. Breiman (1996a) shows that bagging estimator enjoys a smaller mean squared error loss. Bühlmann and Yu (2002) establish the asymptotic properties of bagging estimators in variable selection scenario and show that bagging estimator has a much smaller variance, albeit introducing an additional bias. In the next section, we will study sampling properties of LHA_p and LHA_{pb}.

3 Asymptotic Properties of Local Historical Average

Denote Z as a standard normal random variable with CDF $\Phi\left(\cdot\right)$ and PDF $\varphi\left(\cdot\right)$. Furthermore, define $Z_{b(x)} = Z + b(x)$. The following assumptions will be used to establish the asymptotic properties of the constrained estimator and its bagging version.

Assumption A

(A.1) As $n \to \infty$, (i) $\gamma\left(n,h\right) \to \infty$, (ii) $h \to 0$, (iii) $\gamma\left(n,h\right)h^2 \to 0$.

(A.2) $\gamma\left(n,h\right)\sigma_\alpha^{-1}\left(x\right)\left(\tilde{\alpha}\left(x\right) - \alpha\left(x\right)\right) \overset{d}{\to} Z$, where $\sigma_\alpha\left(x\right) > 0$.

(A.3) $\alpha\left(x\right) = \alpha_1\left(x\right) + \gamma^{-1}\left(n,h\right)\sigma_\alpha\left(x\right)b\left(x\right)$ for some real function $b\left(\cdot\right)$.

Assumption (A.1) places conditions on the bandwidth parameter. Assumption (A.2) states that the unconstrained estimator $\tilde{\alpha}\left(x\right)$ is asymptotically normal. We note that (A.2) is a high-level assumption whose lower-level assumptions would depend on the persistence in the predictor x: (a) when x is strongly mean-reverting (stationary), $\gamma\left(n,h\right)$ is a function of n and h, usually taking the form of $\sqrt{nh}$. In this case, lower-level assumptions that leads to (A.2) can be found in Li and Racine (2007), for example; (b) when x is highly persistent or unit root, $\gamma\left(n,h\right) = \sqrt{\sum_{t=1}^n K\left(\frac{x - x_{t-1}}{h}\right)}$, with convergence in (i) and (iii) of Assumption (A.1) adjusted to convergence in probability, and lower-level assumptions for (A.2) have been studied in Bandi (2004), Wang and Phillips (2009a,b) among others. These lower level assumptions are not repeated here. It can be seen that $\sigma_\alpha\left(x\right)$ represents the asymptotic standard deviation of $\tilde{\alpha}\left(x\right)$, whose expression can be found in earlier references. We emphasize that if $\gamma\left(n,h\right)h^2 \to 0$, the asymptotic bias term of $\tilde{\alpha}\left(x\right)$ vanishes to zero. Assumption (A.3) describes that the distance between $\alpha\left(x\right)$ and the lower bound $\alpha_1\left(x\right)$ is controlled by the drift function $b\left(\cdot\right)$. This assumption is only relevant when local asymptotics are considered. It will be made explicit that asymptotic distributions of constrained estimators will depend on $b\left(\cdot\right)$.

We first establish the following theorem for the constrained estimator $\bar{\alpha}\left(x\right)$.

Theorem 1. (i) *Under A.1-A.2, we have,*

(a) *when $\alpha(x) > \alpha_1(x)$, $\gamma(n,h)\sigma_\alpha^{-1}(x)(\bar{\alpha}(x) - \alpha(x)) \xrightarrow{d} Z$.*

(b) *when $\alpha(x) = \alpha_1(x)$, $\Pr\left[\gamma(n,h)\sigma_\alpha^{-1}(x)(\bar{\alpha}(x) - \alpha(x)) < z\right] \xrightarrow{d} \Phi(z) \cdot 1_{\{z \geq 0\}}.$*

(ii) *If we further assume A.3, then*

$$\gamma(n,h)\sigma_\alpha^{-1}(x)(\bar{\alpha}(x) - \alpha(x)) \xrightarrow{d} Z_{b(x)}1_{\left[Z_{b(x)}>0\right]} - b(x).$$

Remark 1. The proofs are collected in the Appendix. Theorem 1 states the limiting distribution of $\bar{\alpha}(x)$. Part **(i)** presents the usual asymptotic distribution when the constraint is strict and when the $\alpha(x)$ is on the boundary. The result confirms the intuition that, as long as the constraint is strict, it will be met by the unconstrained estimator $\tilde{\alpha}(x)$ when the sample size is large enough. This leads to the conclusion that $\bar{\alpha}(x)$ would be asymptotically equivalent to $\tilde{\alpha}(x)$. When $\alpha(x)$ is on the boundary, the limiting CDF compresses all the mass of negative values at 0. Part **(ii)** establishes the local asymptotic distribution of $\bar{\alpha}(x)$ that depends on the drift parameter $b(x)$. It is easy to see that, if $b(x)$ is allowed to grow as n, $Z_{b(x)}1_{\left[Z_{b(x)}>0\right]} - b(x)$ will collapse to Z, and result in **(ii)** becomes that in **(i.a).** Similarly, **(ii)** reproduces the result of **(i.b)** when $b(x) = 0$.□

The following corollary presents the asymptotic bias and variance of the constrained estimator.

Corollary 1. *Under A.1-A.3, it follows that*

(a) $\lim_{n \to \infty} \gamma(n,h)\sigma_\alpha^{-1}(x)E[\bar{\alpha}(x) - \alpha(x)] = \varphi(b(x)) + b(x)\Phi(b(x)) - b(x).$

(b) $\lim_{n \to \infty} Var\left[\gamma(n,h)\sigma_\alpha^{-1}(x)\bar{\alpha}(x)\right] = \Phi(b(x)) + b(x)\varphi(b(x)) - \varphi^2(b(x)) - 2b(x)\varphi(b(x))\Phi(b(x)) + b^2(x)\Phi(b(x))[1 - \Phi(b(x))].$

Now we consider the LHA_{pb} with the constraint and bagging. To apply bagging, we need an additional assumption:

Assumption A (continued)

(A.4) $\gamma(n,h)\sigma_\alpha^{-1}(x)(\tilde{\alpha}^*(x) - \tilde{\alpha}(x)) \xrightarrow{d} Z.$

Assumption (A.4) requires the bootstrap consistency for the unconstrained LHA estimator $\tilde{\alpha}(x)$. Validity for bootstrap for local nonparametric estimators can be found in Hall (1992) or Horowitz (2001).

Theorem 2. *Under A.1-A.4, we have,*

$$\gamma(n,h)\,\sigma_\alpha(x)^{-1}\left(\hat{\alpha}(x) - \alpha(x)\right) \xrightarrow{d} Z - Z_{b(x)}\Phi\left(-Z_{b(x)}\right) + \varphi\left(-Z_{b(x)}\right).$$

Corollary 2. *If A.1-A.4 hold, then*

(a) $\lim_{n\to\infty}\gamma(n,h)\,\sigma_\alpha^{-1}(x)E\left[\hat{\alpha}(x) - \alpha(x)\right] = 2\varphi * \varphi\left(-b(x)\right) - b(x)\,\Phi * \varphi\left(-b(x)\right).$

(b) $\lim_{n\to\infty}Var\left[\gamma(n,h)\,\sigma_\alpha^{-1}(x)\,\hat{\alpha}(x)\right] = 1+\Phi^2*\varphi''\left(-b(x)\right)+\Phi^2*\varphi\left(-b(x)\right)-2b\Phi^2*\varphi'\left(-b(x)\right)+$

$b^2(x)\,\Phi^2*\varphi\left(-b(x)\right)+\varphi^2*\varphi\left(-b(x)\right)-2\Phi*\varphi''\left(-b(x)\right)-2\Phi*\varphi\left(-b(x)\right)+2b(x)\,\Phi*\varphi'\left(-b(x)\right)-$

$2\varphi*\varphi'\left(-b(x)\right)+2\left(\Phi\cdot\varphi\right)*\varphi'\left(-b(x)\right)-2b(x)\left(\Phi\cdot\varphi\right)*\varphi\left(-b(x)\right)-\left[2\varphi*\varphi\left(-b(x)\right) - b(x)\,\Phi*\varphi\left(-b(x)\right)\right]^2.$

Remark 2. We adopt the notation $f*g$ to denote the convolution of two functions f and g, defined as $f*g(s) = \int f(t) \times g(s-t)\,dt$. Theorem 2 states the limiting distribution of $\hat{\alpha}(x)$ and Corollary 2 shows the explicit expression for its asymptotic bias and variance. The dependence of the limiting distribution on the drift parameter $b(x)$ is explicit through $Z_{b(x)}$. In order to compare the performance of bagging constrained estimator $\hat{\alpha}(x)$ and constrained estimator $\bar{\alpha}(x)$ without bagging, we plot asymptotic variance, squared bias and MSE against the drift function $b(\cdot) = b$ in Figure 1. Figure 1 should be understood for GHA or for LHA for a fixed value of x. We notice from the figure that there is a trade off using bagging, which reduce asymptotic variance while incurring some additional bias. Overall, it is clear that for a large range of values of $b(\cdot)$ (≥ 0.391), bagging estimator enjoys a reduction in asymptotic MSE (AMSE).

Figure 1 About Here

Theorem 3. *Under A.1-A.4, we have*

$$\hat{\alpha}(x) = \tilde{\alpha}(x)\,\Phi\left(Z_{b(x)}\right) + \alpha_1(x)\,\Phi\left(-Z_{b(x)}\right) + O_p\left(\frac{1}{\gamma(n,h)}\right).$$

Remark 3. Theorem 3 establishes that the LHA_{pb} estimator $\hat{\alpha}(x)$ is a model averaging type estimator with a weight $\Phi\left(Z_{b(x)}\right)$ assigned to the unconstrained estimator LHA $\tilde{\alpha}(x)$ and a weight $\Phi\left(-Z_{b(x)}\right)$ to the lower bound $\alpha_1(x)$, up to order $O_p\left(\frac{1}{\gamma(n,h)}\right)$. Note that as $b(x)$ increases to infinity, i.e., when the constraint becomes less binding, $\tilde{\alpha}(x)$ will receive probability weight that goes to 1 (since $\Phi\left(Z_{b(x)}\right)$ approaches 1). On the other hand, as $b(x)$ decreases to zero, i.e., when the constraint becomes more binding, $\Phi\left(Z_{b(x)}\right)$ will become closer to a uniform random variable (since $\Phi(Z)$ is uniform by probability integral transformation). Overall, the performance of the bagging constrained estimator, compared to constrained estimator and unconstrained estimator, can be sensitive to the distance of the lower bound to the true function value, as depicted in Figure 1.

4 Semiparametric Extensions

In this section, we extend the results developed in the previous section to models with multivariate predictors. It has been long recognized that kernel regressions with multivariate regressors suffer from the "curse of dimensionality", i.e., the convergence rate of the kernel estimators will deteriorate as the dimension of the regressors increases. To circumvent this challenge, semiparametric models have become popular. Many recent work has focused on the single index model that enjoys easy implementation. For more details, see Gao (2007) and references therein. This section will illustrate the extension on single index model. We note that the results would be similarly extended to other semiparametric models.

Consider the single index model of the form

$$y_{t+1} = \alpha\left(X_t'\beta_0\right) + u_{t+1}. \tag{9}$$

The model for i.i.d. data has been extensively studied by many authors, to cite a few, Ichimura (1993) and Härdle, Hall and Ichimura (1993). In time series setting, (9) is a special case of the model studied by Xia, Tong and Li (1999). The estimation procedure follows from Ichimura (1993). Let $\alpha(z) = E\left(y_{t+1}|X_t'\beta = z\right)$. Denote its (leave-one-out) Nadaraya-Watson kernel estimator (with

the s-th observation omitted) as

$$\tilde{\alpha}_{-s}(z) = \frac{\sum_{t=1, t \neq s}^{n} k_h \left(X'_{t-1} \beta - z \right) y_t}{\sum_{t=1, t \neq s}^{n} k_h \left(X'_{t-1} \beta - z \right)}, \tag{10}$$

where h is a bandwidth, and $k_h(\cdot) = k(\cdot/h)$, $k(\cdot)$ is a kernel function. The estimation of β_0 and the choice of h can be performed by selecting the orientation β and h that minimize a measure of the distance. That is,

$$\left(\hat{\beta}, \hat{h} \right) = \arg \min_{\beta, h} S_n(\beta, h), \tag{11}$$

where $S_n(\beta, h) = \sum_{s=1}^{n} \left[y_s - \tilde{\alpha}_{-s} \left(X'_{s-1} \beta \right) \right]^2$. With $\hat{\beta}$ and $\hat{h}$, the semiparametric single index local historical average forecast at time n using $z = X'_n \hat{\beta}$ is obtained from

$$\tilde{\alpha} \left(X'_n \hat{\beta} \right) = \frac{\sum_{t=1}^{n} k_{\hat{h}} \left(X'_{t-1} \hat{\beta} - X'_n \hat{\beta} \right) y_t}{\sum_{t=1}^{n} k_{\hat{h}} \left(X'_{t-1} \hat{\beta} - X'_n \hat{\beta} \right)}. \tag{12}$$

Conditions for $\tilde{\alpha} \left(X'_n \hat{\beta} \right)$ to satisfy Assumption (A.1)-(A.2) are given in Xia, Tong and Li (1999). Under the constraint of (5), $\bar{\alpha} \left(X'_n \hat{\beta} \right)$ and $\hat{\alpha} \left(X'_n \hat{\beta} \right)$ can be defined analogous to (6) and (7), respectively. It follows that Theorem 1-3 also hold in the semiparametric single index model.

5   Finite Sample Properties of Local Historical Average

In this section, we study the finite sample performance of the constrained estimator LHA_p $\bar{\alpha}(x)$ and its bagging version LHA_{pb} $\hat{\alpha}(x)$. We first consider the following data generating process (DGP)

$$\text{DGP 1}: \quad y_{t+1} = a \left(4x_t - 2 \right)^3 + e_{t+1}, \tag{13}$$

where

$$x_t - \mu = \rho \left(x_{t-1} - \mu \right) + u_t, \tag{14}$$

with $\mu = 1$, e_t and u_t are i.i.d. normal r.v. with mean 0 and $\sigma_e = 1$ and $\sigma_u = 0.5$, and $a \in \{0.001, 0.004, 0.007, 0.010, 0.020\}$ that controls the distance between $\alpha(x)$ and the bound $\alpha_1(x) = 0$. Hence, from Assumption (A.3), note that the value of a also controls $b(x)$ for given $\gamma(n, h)$ and $\sigma_\alpha(x)$. We follow the design of Chen and Hong (2009) to allow time series dependence in the predictor and consider different values of ρ chosen from $\{0, 0.1, 0.9, 1\}$. We evaluate the estimators of

9

$\alpha(x)$ at $x = 1$ and 1.5. We compute the mean of squared errors out of 200 Monte Carlo replications. In each replication, we experiment with sample size $n = 50, 100, 200$, and the bootstrap sample size $J = 100$ for bagging in each replication. The relative mean squared errors are reported in Table 1. We use cross-validation to select a bandwidth h that minimizes the integrated mean squared error and use this same bandwidth for the $J = 100$ bootstrap samples generated within each replication. The block bootstrap method is used to generate bootstrap samples. We consider the block length to be 1, 4 and 12 but the main results do not change much. Therefore, the result for block length equal to 4 will be reported. See Härdle, Horowitz and Kreiss (2003) and references therein for details of block bootstrap method for time series.

Consider a forecasting model

$$\text{Model}: \quad y_{t+1} = \alpha(x) + u_{t+1}. \tag{15}$$

For a given evaluation predictor value x, we are interested in forming a forecast $\hat{y}_{n+1} = \alpha(x|I_n)$, where $I_n = \{x_{n_0}, ..., x_n, y_{n_0}, ..., y_n\}$ is used to estimate a model. In this section for simulation we fix both $n_0 = 1$ and $n = 50, 100, 200$, and estimate various models using the $R \equiv n - n_0 + 1 = n$ observations. In each Monte Carlo replication i ($i = 1, \ldots, 200$), 200 values of $\{\hat{m}^{(i)}(x)\}$ are computed at various fixed x values, and also 200 values of $\{\hat{u}^{(i)}(x) \equiv \alpha(x) - \hat{m}^{(i)}(x)\}_{i=1}^{200}$ are computed. Here, $\hat{m}(x) = \tilde{\alpha}(x), \bar{\alpha}(x)$, or $\hat{\alpha}(x)$. We compute the Monte Carlo average of the squared $\hat{u}^{(i)}(x)$ over i for each evaluation point x, $MSE \equiv \frac{1}{200} \sum_{i=1}^{200} \hat{u}^{(i)2}(x)$. We compare the three models pointwise for different values of x, the results are reported in Table 1 at $x = 1$ and in Table 2 at $x = 1.5$. We report the relative MSE of LHA_p $\bar{\alpha}(x)$ and LHA_{pb} $\hat{\alpha}(x)$ w.r.t. that of LHA $\tilde{\alpha}(x)$.

We summarize the main findings as follows. At $x = 1$, the constrained estimator works better than unconstrained estimator for small values of a in all sample sizes. The gain in relative mean squared errors $\left(MSE_{\text{LHA}_p}/MSE_{\text{LHA}}\right)$ can be as big as 50%. When a gets larger, the gain of constrained estimator starts to decrease, as noted by the increase of relative MSE. The constraint will become non-binding eventually and thus constrained estimator performs the same as the unconstrained. Bagging does not tend to work for sample size $n = 50$ for small values of a considered here. When a and n get larger ($a = 0.02, n = 100, 200$), bagging improves upon the constrained

10

estimator for all values of ρ, with the gain in relative mean squared error $\left(MSE_{\text{LHA}_{pb}}/MSE_{\text{LHA}}\right)$ as large as 5%. This is consistent with the theory that bagging estimator works better than the constrained estimator when the sample size n and the level of the function determined by a are of suitable proportion for $b(x)$. For large values of a, the relative mean squared errors that are larger than 1 are due to sampling errors incurred in the bootstrap procedure.

As shown from Table 2, the results become more apparent when the estimators are evaluated at $x = 1.5$. Again, the role of the constraint becomes less important as a gets larger. Bagging's role become more salient in this case, with gain in MSE more than 10% when $a = 0.007$ and $n = 50$. As Figure 1 shows, the AMSE of bagging estimator can be over 10% smaller than constrained estimator. So the result we find is congruent with the asymptotic theory. Bagging achieved the maximal amount of gain in relative MSE (16%) when $a = 0.02$, $\rho = 1$ and $n = 100$.

We next consider the following DGP

$$\text{DGP 2}: \quad y_{t+1} = a \exp\left(x_t'\beta\right) + e_{t+1}, \tag{16}$$

where $\beta = (1, 0.5, 0.5)'$, $x_t = (x_{1,t}, x_{2,t}, x_{3,t})'$, $x_{k,t}$, for $k = 1, 2, 3$, is generated independently from an AR(1) process as in (14) with $\mu_k = 1$, e_t and $u_{k,t}$ are i.i.d. normal r.v. with mean 0 and $\sigma_e = 1$ and $\sigma_{u,k} = 1$, $a \in \{0.001, 0.004, 0.007, 0.010, 0.020\}$. ρ is set to be 0 or 0.9. Other specifications on the simulation are the same as those in DGP 1. We consider two forecasting models. One is the multivariate local constant least square estimator or the Nadaraya-Watson (NW) estimator, and the other is the estimator derived from the single index model (SIM). The models are compared at evaluation point $x_t = (1, 1, 1)'$. The relative forecasting MSEs are reported in Table 3. It can be seen from Table 3 that the constrained estimator and the bagging constrained estimator achieve the uniform reduction in MSE for the multivariate NW estimator and the SIM estimator. Furthermore, as a gets larger, the bagging constrained estimator tends to outperform the constrained estimator without bagging. The role of imposing constraint becomes less important for larger a. These findings are very much similar to those from DGP 1.

Table 1 About Here

11

6   Empirical Properties of Local Historical Average

To put our proposed constrained local historical average estimators LHA_p and LHA_{pb} in practice, we consider forecasting U.S. equity premium. Equity premium or excess return is defined as return of the S&P500 Index over the risk-free short-term interest rate. Denote by P_t the S&P500 Index at month t. The monthly simple one-month return from month t to month $t+1$ is defined as $R_t(1) \equiv P_{t+1}/P_t - 1$, and one-month excess return is $Q_t(1) \equiv R_t(1) - r_t$ with r_t being the risk-free interest rate.

Following Campbell, Lo and MacKinlay (1997, p. 10), we define the k-period return from month t to month $t+k$ as

$$
\begin{aligned}
R_t(k) &\equiv \frac{P_{t+k}}{P_t} - 1 \\
&= \left(\frac{P_{t+k}}{P_{t+k-1}} \right) \times \cdots \times \left(\frac{P_{t+1}}{P_t} \right) - 1 \\
&= (1 + R_{t+k-1}(1)) \times \cdots \times (1 + R_t(1)) - 1
\end{aligned}
\tag{17}
$$

and following CT (2008) we define the k-period excess return as

$$
\begin{aligned}
Q_t(k) &\equiv (1 + R_{t+k-1}(1) - r_{t+k-1}) \times \cdots \times (1 + R_t(1) - r_t) - 1 \\
&= Q_{t+k-1}(1) \times \cdots \times Q_t(1) - 1 \\
&= \left[\prod_{j=1}^{k} Q_{t+k-j}(1) \right] - 1.
\end{aligned}
\tag{18}
$$

We let $y_{t+1} = Q_t(k)$, and consider $k = 1, 12$ as reported in CT (2008). The results presented in Table 4 are with this definition of the equity premium in (18). We have conducted the same analysis with $k = 3, 6$ but their results turn out to be what may be easily expected from $k = 1, 12$, and thus we do not report them for space.

We use 11 predictors including dividend price ratio (d/p), earning price ratio (e/p), smooth earning price ratio (se/p), book to market ratio (b/m), return on equity (roe), treasure bill (t-$bill$),

long term yield (lty), term spread (ts), default spread (ds), inflation (inf) and net equity issuance (nei). We thank John Campbell and Sam Thompson for sharing their data used in CT (2008). It is found that d/p, e/p, se/p, b/m, roe, $t\text{-}bill$, and lty are unit root processes, while the equity premium $(Q_t(1), Q_t(12))$, ts, ds, inf and nei are not. To save space, we use the first-order difference of the unit root variables when they are used as predictor. The results using the unit root variables (not their differences) as predictors are quite similar and are available from the authors upon request.

We follow CT (2008) to impose a constraint that the equity premium should be positive. We consider the annualized monthly equity premium $Q_t(12)$ and monthly equity premium $Q_t(1)$, with forecasts starting from 1960:01 and 1980:01 and rolling till 2005:12. The in-sample size for model estimation is kept fixed as $R = 120$. We report the results for mean squared forecast errors (MSFE) relative to the global historical average (GHA) forecast in Table 4.

In Table 4A with $k = 12$, we are forecasting the annualized equity premium $Q_t(12)$ at month t. We first see that nonparametric LHA forecasts $\tilde{\alpha}(x)$ outperform the global historical average GHA $\tilde{\alpha}$, for the predictor d/p, e/p, ds, and nei in both forecasting periods. Second, for these predictors, we observe that imposing the positivity constraint generally reduces the MSFE, which may be further reduced after the bagging procedure. The largest reduction for imposing the constrain occurs for ds when forecasts begins at 1960:01, and it achieves more than 5%. Third, bagging works for annualized equity premium forecasts for almost all predictors in both forecasting periods, though the improvement is often small. However, this is consistent with the theory in Section 3 as summarized in Figure 1. Compared to local GHA, the bagging constrained forecasts are better, except for one case. The maximum gain in MSFE is over 16%, for the predictor ds, in the forecasting sample 1960:01. Fourth, for the semiparametric single index model that uses all the 11 predictors, the positivity constraint improves the MSFE from 1.006 to 0.998, which is further improved by the bagging procedure that achieves a relative MSFE 0.985 when forecasts begins at 1960:01. The similar result in improvement direction is also seen when forecasts begins at 1980:01.

In Table 4B with $k = 1$, we are forecasting the monthly equity premium $Q_t(1)$ at month t. We hardly see much gain using unconstrained nonparametric methods over the GHA. The best that nonparametric MSFE gains, with 0.7% reduction, is for the predictor e/p when forecasts start from

13

1980:01. However, imposing the positivity constraint LHA$_p$ almost always improves MSFE. We observe that bagging works most of the time. Especially for the predictors d/p, e/p, b/m, bagging even help the nonparametric LHA$_{pb}$ forecast to beat the "unbeatable" global historical average GHA in both forecast samples. This gain is as large as 1.1% for e/p when forecasts start from 1980:01, which is economically significant according to Campbell and Thompson (2008).

Table 4 About Here

7 Conclusions

In this paper, we investigate the use of nonparametric local historical average and semiparametric single index local historical average in forecasting of equity premium, compared to the global historical average which is traditionally used. In addition, we consider imposing a local constraint that the equity premium is expected to be positive. We define the constrained local historical average forecast and its bagging version. Asymptotic properties of these constrained/bagged forecasts are established. We show that the constrained local historical average forecast operates as model-selection between the local historical average and zero equity premium, and that the bagged constrained local historical average forecast yields a locally shrunken combined forecast of the local historical average forecast and the zero equity premium forecast. The local combining weights are determined by the probability that the local constraint is binding. Significant gains in MSE can be achieved by using a local model, a local constraint, and bagging as shown in our simulation. In predicting U.S. equity premium, we show that substantial nonlinearity is present which can be captured by the nonparametric local historical average and that the local positivity constraint of the equity premium provides valuable prior information in improving its out-of-sample prediction.

The paper studies the role of constrained estimation and that of bagging under the condition that the nonparametric estimator is asymptotically normal. In the case of structural change, this condition is often violated. Thus the comparison becomes quite challenging and needs further detailed investigation, which is beyond the scope of this paper. Such a potential research topic is left for future study.

Appendix

Proof of Theorem 1. We first prove (**i**). For any $z \in \mathbb{R}$,

$$\Pr\left(\gamma\left(n,h\right)\sigma_\alpha^{-1}\left(x\right)\left(\bar{\alpha}\left(x\right)-\alpha\left(x\right)\right)<z\right)$$

$$=\Pr\left(\gamma\left(n,h\right)\sigma_\alpha^{-1}\left(x\right)\left(\max\{\tilde{\alpha}\left(x\right),\alpha_1\left(x\right)\}-\alpha\left(x\right)\right)<z|\tilde{\alpha}\left(x\right)<\alpha_1\left(x\right)\right)\times\Pr\left(\tilde{\alpha}\left(x\right)<\alpha_1\left(x\right)\right)$$

$$+\Pr\left(\gamma\left(n,h\right)\sigma_\alpha^{-1}\left(x\right)\left(\max\{\tilde{\alpha}\left(x\right),\alpha_1\left(x\right)\}-\alpha\left(x\right)\right)<z|\tilde{\alpha}\left(x\right)\geq\alpha_1\left(x\right)\right)\times\Pr\left(\tilde{\alpha}\left(x\right)\geq\alpha_1\left(x\right)\right)$$

$$=\Pr\left(\gamma\left(n,h\right)\sigma_\alpha^{-1}\left(x\right)\left(\alpha_1\left(x\right)-\alpha\left(x\right)\right)<z\right)\times\Pr\left(\tilde{\alpha}\left(x\right)<\alpha_1\left(x\right)\right)+ \tag{19}$$

$$\Pr\left(\gamma\left(n\right)\sigma^{-1}\left(\tilde{\alpha}\left(x\right)-\alpha\left(x\right)\right)<z|\tilde{\alpha}\left(x\right)\geq\alpha_1\left(x\right)\right)\times\Pr\left(\tilde{\alpha}\left(x\right)\geq\alpha_1\left(x\right)\right).$$

First term in (19): When $\alpha\left(x\right)>\alpha_1\left(x\right)$,

$$\Pr\left(\gamma\left(n,h\right)\sigma_\alpha^{-1}\left(x\right)\left(\alpha_1\left(x\right)-\alpha\left(x\right)\right)<z\right)\to\Pr\left(-\infty<z\right)=1.$$

When $\alpha\left(x\right)=\alpha_1\left(x\right)$,

$$\Pr\left(\gamma\left(n,h\right)\sigma_\alpha^{-1}\left(x\right)\left(\alpha_1\left(x\right)-\alpha\left(x\right)\right)<z\right)\to\left\{\begin{array}{ll}1,&\text{if }z>0\\0,&\text{if }z\leq0\end{array}\right..$$

Second term in (19):

$$\Pr\left(\tilde{\alpha}\left(x\right)<\alpha_1\left(x\right)\right)$$

$$=\Pr\left(\gamma\left(n,h\right)\sigma_\alpha^{-1}\left(x\right)\sigma^{-1}\left(\tilde{\alpha}\left(x\right)-\alpha\left(x\right)\right)<\gamma\left(n,h\right)\sigma_\alpha^{-1}\left(x\right)\sigma^{-1}\left(\alpha_1\left(x\right)-\alpha\left(x\right)\right)\right)$$

$$\to\left\{\begin{array}{ll}\Pr\left(Z<-\infty\right)=0,&\text{if }\alpha\left(x\right)>\alpha_1\left(x\right)\\\Pr\left(Z<0\right)=\Phi\left(0\right),&\text{if }\alpha\left(x\right)=\alpha_1\left(x\right)\end{array}\right..$$

Third term in (19): When $\alpha\left(x\right)=\alpha_1\left(x\right)$,

$$\Pr\left(\gamma\left(n,h\right)\sigma_\alpha^{-1}\left(x\right)\left(\tilde{\alpha}\left(x\right)-\alpha\left(x\right)\right)<z|\tilde{\alpha}\left(x\right)\geq\alpha_1\left(x\right)\right)$$

$$=\frac{\Pr\left(\gamma\left(n,h\right)\sigma_\alpha^{-1}\left(x\right)\left(\tilde{\alpha}\left(x\right)-\alpha\left(x\right)\right)<z,\gamma\left(n,h\right)\sigma_\alpha^{-1}\left(x\right)\left(\tilde{\alpha}\left(x\right)-\alpha_1\left(x\right)\right)\geq0\right)}{\Pr\left(\gamma\left(n,h\right)\sigma_\alpha^{-1}\left(x\right)\left(\tilde{\alpha}\left(x\right)-\alpha_1\left(x\right)\right)\geq0\right)}$$

$$=\frac{\Pr\left(\gamma\left(n,h\right)\sigma_\alpha^{-1}\left(x\right)\left(\tilde{\alpha}\left(x\right)-\alpha\left(x\right)\right)<z,\gamma\left(n,h\right)\sigma_\alpha^{-1}\left(x\right)\left(\tilde{\alpha}\left(x\right)-\alpha\left(x\right)\right)\geq\gamma\left(n,h\right)\sigma_\alpha^{-1}\left(x\right)\left(\alpha_1\left(x\right)-\alpha\left(x\right)\right)\right)}{\Pr\left(\gamma\left(n,h\right)\sigma_\alpha^{-1}\left(x\right)\left(\tilde{\alpha}\left(x\right)-\alpha\left(x\right)\right)\geq\gamma\left(n\right)\left(\alpha_1\left(x\right)-\alpha\left(x\right)\right)\right)}$$

$$\to\left\{\begin{array}{ll}\frac{\Phi\left(z\right)-\Phi\left(0\right)}{1-\Phi\left(0\right)},&\text{if }z>0\\0,&\text{otherwise}\end{array}\right..$$

When $\alpha\left(x\right)>\alpha_1\left(x\right)$,

$$\Pr\left(\gamma\left(n,h\right)\sigma_\alpha^{-1}\left(x\right)\left(\tilde{\alpha}\left(x\right)-\alpha\left(x\right)\right)<z|\tilde{\alpha}\left(x\right)\geq\alpha_1\left(x\right)\right)=\Phi\left(z\right).$$

15

Fourth term in (19):

$$\Pr\left(\tilde{\alpha}\left(x\right) \geq \alpha_1\left(x\right)\right)$$

$$= 1 - \Pr\left(\gamma\left(n,h\right)\sigma_\alpha^{-1}\left(x\right)\sigma^{-1}\left(\tilde{\alpha}\left(x\right) - \alpha\left(x\right)\right) < \gamma\left(n,h\right)\sigma_\alpha^{-1}\left(x\right)\left(\alpha_1\left(x\right) - \alpha\left(x\right)\right)\right)$$

$$\rightarrow \begin{cases} 1 - \Pr\left(Z < -\infty\right) = 1, & \text{if } \alpha\left(x\right) > \alpha_1\left(x\right) \\ 1 - \Pr\left(Z < 0\right) = 1 - \Phi\left(0\right), & \text{if } \alpha\left(x\right) = \alpha_1\left(x\right) \end{cases}.$$

Therefore, combining the four terms leads to, (**i.a**) when $\alpha\left(x\right) > \alpha_1\left(x\right)$, $\Pr\left(\gamma\left(n,h\right)\sigma_\alpha^{-1}\left(x\right)\left(\bar{\alpha}\left(x\right) - \alpha\left(x\right)\right) < z\right)$

$\Phi\left(z\right)$ and (**i.b**) when $\alpha\left(x\right) = \alpha_1\left(x\right)$, for $z > 0$, $\Pr\left(\gamma\left(n,h\right)\sigma_\alpha^{-1}\left(x\right)\left(\bar{\alpha}\left(x\right) - \alpha\left(x\right) - B_m\left(x\right)\right) < z\right) \rightarrow$

$\Phi\left(z\right)$; for $z = 0$, $\Pr\left(\gamma\left(n,h\right)\sigma_\alpha^{-1}\left(x\right)\left(\bar{\alpha}\left(x\right) - \alpha\left(x\right)\right) < z\right) \rightarrow \Phi\left(0\right)$; for $z < 0$, $\Pr\left(\gamma\left(n,h\right)\sigma_\alpha^{-1}\left(x\right)\left(\bar{\alpha}\left(x\right) - \alpha\left(x\right)\right) < z\right)$

0. Written compactly, we have $\Pr\left(\gamma\left(n,h\right)\sigma_\alpha^{-1}\left(x\right)\left(\bar{\alpha}\left(x\right) - \alpha\left(x\right)\right) < z\right) = \Phi\left(z\right)1_{\{z>0\}}$.

To prove (**ii**), note that

$$\gamma\left(n,h\right)\sigma_\alpha^{-1}\left(x\right)\left(\bar{\alpha}\left(x\right) - \alpha\left(x\right)\right)$$

$$= \gamma\left(n,h\right)\sigma_\alpha^{-1}\left(x\right)\left(\alpha_1\left(x\right) - \alpha\left(x\right)\right) + \gamma\left(n,h\right)\sigma_\alpha^{-1}\left(x\right)\left(\tilde{\alpha}\left(x\right) - \alpha_1\left(x\right)\right)1_{\left[\gamma\left(n,h\right)\sigma_\alpha^{-1}\left(x\right)\left(\tilde{\alpha}\left(x\right) - \alpha_1\left(x\right)\right)>0\right]}$$

$$= \gamma\left(n,h\right)\sigma_\alpha^{-1}\left(x\right)\left(\alpha_1\left(x\right) - \alpha\left(x\right)\right) +$$

$$\gamma\left(n,h\right)\sigma_\alpha^{-1}\left(x\right)\left(\tilde{\alpha}\left(x\right) - \alpha\left(x\right) + \alpha\left(x\right) - \alpha_1\left(x\right)\right)1_{\left[\gamma\left(n,h\right)\sigma_\alpha^{-1}\left(x\right)\left(\tilde{\alpha}\left(x\right) - \alpha_1\left(x\right)\right)>0\right]}$$

$$\xrightarrow{d} Z_{b(x)}1_{\left[Z_{b(x)}>0\right]} - b\left(x\right).$$

by Assumption (A.1) and (A.2).

Proof of Corollary 1. For a standard normal random variable Z and a constant b, we can easily show that $E1_{[Z_b>0]} = \Phi\left(b\right)$, $E\left[Z1_{[Z_b>0]}\right] = \varphi\left(b\right)$, $E\left[Z^21_{[Z_b>0]}\right] = -b\varphi\left(b\right) + \Phi\left(b\right)$, $E\left[Z_b1_{[Z_b>0]}\right] = \varphi\left(b\right) + b\Phi\left(b\right)$ and $E\left[Z_b^21_{[Z_b>0]}\right] = \Phi\left(b\right) + b\varphi\left(b\right) + b^2\Phi\left(b\right)$. Therefore, we have

$$E\left[Z_{b(x)}1_{\left[Z_{b(x)}>0\right]} - b\left(x\right)\right] = EZ1_{\left[Z_{b(x)}>0\right]} + b\left(x\right)E1_{\left[Z_{b(x)}>0\right]} - b\left(x\right)$$

$$= \varphi\left(b\left(x\right)\right) + b\left(x\right)\Phi\left(b\left(x\right)\right) - b\left(x\right),$$

and

$$Var\left[Z_{b(x)}1_{\left[Z_{b(x)}>0\right]} - b\left(x\right)\right] = Var\left[Z_{b(x)}1_{\left[Z_{b(x)}>0\right]}\right]$$

$$= E\left\{\left[Z_{b(x)}1_{\left[Z_{b(x)}>0\right]}\right]^2\right\} - \left\{E\left[Z_{b(x)}1_{\left[Z_{b(x)}>0\right]}\right]\right\}^2$$

$$= \Phi\left(b\left(x\right)\right) + b\left(x\right)\varphi\left(b\left(x\right)\right) + b^2\left(x\right)\Phi\left(b\left(x\right)\right) - \left[\varphi\left(b\left(x\right)\right) + b\Phi\left(b\left(x\right)\right)\right]^2.$$

Proof of Theorem 2. Write

$$\gamma\left(n,h\right)\sigma_\alpha^{-1}\left(x\right)\left(\hat\alpha\left(x\right)-\alpha\left(x\right)\right)$$

$$= \gamma\left(n,h\right)\sigma_\alpha^{-1}\left(x\right)\left(E^*\left[\tilde\alpha^*\left(x\right)1_{\left[\tilde\alpha^*\left(x\right)\geq\alpha_1\left(x\right)\right]}\right]+\alpha_1\left(x\right)E^*\left[1_{\left[\tilde\alpha^*\left(x\right)<\alpha_1\left(x\right)\right]}\right]-\alpha\left(x\right)\right)$$

$$= \gamma\left(n,h\right)\sigma_\alpha^{-1}\left(x\right)\left(E^*\left[\left(\tilde\alpha^*\left(x\right)-\alpha\left(x\right)\right)1_{\left[\tilde\alpha^*\left(x\right)\geq\alpha_1\left(x\right)\right]}\right]+\left(\alpha_1\left(x\right)-\alpha\left(x\right)\right)E^*\left[1_{\left[\tilde\alpha^*\left(x\right)<\alpha_1\left(x\right)\right]}\right]\right)\tag{20}$$

For the first term in (20) we have

$$\gamma\left(n,h\right)\sigma_\alpha^{-1}\left(x\right)\left(E^*\left[\left(\tilde\alpha^*\left(x\right)-\alpha\left(x\right)\right)1_{\left[\tilde\alpha^*\left(x\right)\geq\alpha_1\left(x\right)\right]}\right]\right)$$

$$= E^*\left[\gamma\left(n,h\right)\sigma_\alpha^{-1}\left(x\right)\left(\tilde\alpha^*\left(x\right)-\alpha\left(x\right)\right)1_{\left[\tilde\alpha^*\left(x\right)\geq\alpha_1\left(x\right)\right]}\right]$$

$$= E^*\left[\gamma\left(n,h\right)\sigma_\alpha^{-1}\left(x\right)\left(\tilde\alpha^*\left(x\right)-\alpha\left(x\right)\right)\times 1_{\left[\gamma\left(n,h\right)\sigma_\alpha^{-1}\left(x\right)\left(\tilde\alpha^*\left(x\right)-\alpha\left(x\right)\right)\geq\gamma\left(n,h\right)\sigma_\alpha^{-1}\left(x\right)\left(\alpha_1\left(x\right)-\alpha\left(x\right)\right)\right]}\right]$$

$$\overset{d}{\to} E_W\left[W1_{\left[W\geq-b\left(x\right)\right]}|Z\right],\tag{21}$$

where $W|Z\sim N\left(Z,1\right)$. Note that

$$E_W\left[W1_{\left[W\geq-b\left(x\right)\right]}|Z\right]$$

$$= E_W\left[W|Z\right]-E_W\left[W1_{\left[W<-b\left(x\right)\right]}|Z\right]$$

$$= Z-\int_{-\infty}^{-b\left(x\right)}w\varphi\left(w-Z\right)dw$$

$$= Z-\int_{-\infty}^{-b\left(x\right)-Z}\left(s+Z\right)\varphi\left(s\right)ds$$

$$= Z-Z\Phi\left(-b\left(x\right)-Z\right)-\int_{-\infty}^{-b\left(x\right)-Z}s\varphi\left(s\right)ds$$

$$= Z-Z\Phi\left(-Z_{b\left(x\right)}\right)+\varphi\left(-Z_{b\left(x\right)}\right).$$

Similarly, for the second term in (20) we get

$$\gamma\left(n,h\right)\sigma_\alpha^{-1}\left(x\right)\left(\alpha_1\left(x\right)-\alpha\left(x\right)\right)E^*\left[1_{\left[\tilde\alpha^*\left(x\right)<\alpha_1\left(x\right)\right]}\right]\overset{p}{\to}-b\left(x\right)\Phi_Z\left(-Z_{b\left(x\right)}\right),\tag{22}$$

by Slutsky's theorem.

Putting (21) and (22) together into (20) gives the desired result.

Proof of Corollary 2. We can first show that $E\varphi\left(-Z_b\right)=\varphi*\varphi\left(-b\right)$, $E\varphi^2\left(-Z_b\right)=\varphi^2*\varphi\left(-b\right)$, $E\left[Z\varphi\left(-Z_b\right)\right]=-\varphi*\varphi'\left(-b\right)$, $E\left[Z\Phi\left(-Z_b\right)\right]=-\varphi*\varphi\left(b\right)$, $E\left[Z^2\Phi\left(-Z_b\right)\right]=\Phi*\varphi''\left(-b\right)+\Phi*\varphi\left(-b\right)$,

$E\left[Z^2\Phi^2\left(-Z_b\right)\right] = \Phi^2 * \varphi''\left(-b\right) + \Phi^2 * \varphi\left(-b\right)$, and $E\left[Z\Phi\left(-Z_b\right)\varphi\left(-Z_b\right)\right] = -\left(\Phi\cdot\varphi\right)*\varphi'\left(-b\right)$. To complete the proof, we only need to show that,

$$E\left[Z - Z_{b(x)}\Phi\left(-Z_{b(x)}\right) + \varphi\left(-Z_{b(x)}\right)\right]$$
$$= 0 + \varphi * \varphi\left(-b\left(x\right)\right) - b\left(x\right)\Phi * \varphi\left(-b\left(x\right)\right) + \varphi * \varphi\left(-b\left(x\right)\right)$$
$$= 2\varphi * \varphi\left(-b\left(x\right)\right) - b\left(x\right)\Phi * \varphi\left(-b\left(x\right)\right).$$

and

$$E\left[Z - Z_{b(x)}\Phi\left(-Z_{b(x)}\right) + \varphi\left(-Z_{b(x)}\right)\right]^2$$
$$= EZ^2 + E\left[Z_{b(x)}\Phi\left(-Z_{b(x)}\right)\right] + E\left[\varphi\left(-Z_{b(x)}\right)\right]^2 - 2E\left[ZZ_{b(x)}\Phi\left(-Z_{b(x)}\right)\right]$$
$$+2E\left[Z\varphi\left(-Z_{b(x)}\right)\right] - 2E\left[Z_{b(x)}\Phi\left(-Z_{b(x)}\right)\varphi\left(-Z_{b(x)}\right)\right]$$
$$= 1 + \Phi^2 * \varphi''\left(-b\left(x\right)\right) + \Phi^2 * \varphi\left(-b\left(x\right)\right)$$
$$-2b\Phi^2 * \varphi'\left(-b\left(x\right)\right) + b^2\left(x\right)\Phi^2 * \varphi\left(-b\left(x\right)\right) + \varphi^2 * \varphi\left(-b\left(x\right)\right)$$
$$-2\Phi * \varphi''\left(-b\left(x\right)\right) - 2\Phi * \varphi\left(-b\left(x\right)\right) + 2b\left(x\right)\Phi * \varphi'\left(-b\left(x\right)\right)$$
$$-2\varphi * \varphi'\left(-b\left(x\right)\right) + 2\left(\Phi\cdot\varphi\right)*\varphi'\left(-b\left(x\right)\right) - 2b\left(x\right)\left(\Phi\cdot\varphi\right)*\varphi\left(-b\left(x\right)\right).$$

Proof of Theorem 3. By definition,

$$\hat{\alpha}\left(x\right) = E^*\bar{\alpha}^*\left(x\right)$$
$$= E^*\left[\tilde{\alpha}^*\left(x\right)1_{\left[\tilde{\alpha}^*\left(x\right)\geq\alpha_1\left(x\right)\right]}\right] + E^*\alpha_1\left(x\right)\left[1_{\left[\tilde{\alpha}^*\left(x\right)<\alpha_1\left(x\right)\right]}\right]$$
$$\equiv A_1 + A_2,$$

where

$$A_2 = E^*\left[\alpha_1\left(x\right)1_{\left[\tilde{\alpha}^*\left(x\right)<\alpha_1\left(x\right)\right]}\right]$$
$$= \alpha_1\left(x\right)E^*\left[1_{\left[\tilde{\alpha}^*\left(x\right)-\tilde{\alpha}\left(x\right)<\alpha_1\left(x\right)-\tilde{\alpha}\left(x\right)\right]}\right]$$
$$= \alpha_1\left(x\right)E_W\left[1_{\left[W<-b\left(x\right)\right]}|Z\right] + O_p\left(\frac{1}{\gamma\left(n,h\right)}\right)$$
$$= \alpha_1\left(x\right)\Phi\left(-Z_{b(x)}\right) + O_p\left(\frac{1}{\gamma\left(n,h\right)}\right),$$

and

$$
\begin{aligned}
A_1 &= E^* \left[\tilde{\alpha}^* (x) \, 1_{[\tilde{\alpha}^*(x) \geq \alpha_1(x)]} \right] \\
&= E^* \left\{ [\tilde{\alpha}^* (x) - \tilde{\alpha} (x)] \, 1_{[\tilde{\alpha}^*(x) \geq \alpha_1(x)]} \right\} + E^* \tilde{\alpha} (x) \, 1_{[\tilde{\alpha}^*(x) \geq \alpha_1(x)]} \\
&\equiv A_{11} + A_{12},
\end{aligned}
$$

with

$$
\begin{aligned}
A_{11} &= E^* \left\{ [\tilde{\alpha}^* (x) - \tilde{\alpha} (x)] \, 1_{[\tilde{\alpha}^*(x) \geq \alpha_1(x)]} \right\} \\
&= \frac{1}{\gamma(n,h)} E^* \left\{ \gamma(n,h) [\tilde{\alpha}^* (x) - \tilde{\alpha} (x)] \, 1_{[\gamma(n,h)(\tilde{\alpha}^*(x) - \alpha(x)) \geq \gamma(n,h)(\alpha_1(x) - \alpha(x))]} \right\} \\
&= \frac{1}{\gamma(n,h)} E_W \left[W 1_{[W \geq -b(x)]} | Z \right] + o_p \left(\frac{1}{\gamma(n,h)} \right) \\
&= o_p \left(\frac{1}{\gamma(n,h)} \right),
\end{aligned}
$$

and

$$
\begin{aligned}
A_{12} &= \tilde{\alpha} (x) \, E^* \left[1_{[\tilde{\alpha}^*(x) \geq \alpha_1(x)]} \right] \\
&= \tilde{\alpha} (x) \, E_W \left[1_{[W \geq -b(x)]} | Z \right] \\
&= \tilde{\alpha} (x) \, \Phi \left(Z_{b(x)} \right) + O_p \left(\frac{1}{\gamma(n,h)} \right).
\end{aligned}
$$

Combining the results completes the proof.

References

Bandi, F.M. (2004), "On Persistence and Nonparametric Estimation," Unpublished manuscript, University of Chicago.

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

Breiman, L. (1996b), "Heuristics of Instability and Stabilization in Model Selection", *Annals of Statistics* 24, 2350–2383.

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

Campbell, J.Y., Lo A.W., and MacKinlay, A.C. (1997), *The Econometrics of Financial Markets*, Princeton University Press.

Campbell, J.Y. and Thompson, S. (2008), "Predicting the Equity Premium Out of Sample: Can Anything Beat the Historical Average?," *Review of Financial Studies* 21(4), 1511-1531.

Chen, Q. and Hong, Y. (2009), "Predictability of Equity Returns Over Different Time Horizons: A Nonparametric Approach," Working paper, Cornell University.

Gao, J. (2007), *Nonlinear Time Series: Nonparametric and Semiparametric Methods*, Chapman&Hall/CRC.

Goyal, A. and Welch, I. (2008), "A Comprehensive Look at the Empirical Performance of Equity Premium Prediction," *Review of Financial Studies* 21, 1455-1508.

Hall, P. (1992), *The Bootstrap and Edgeworth Expansion.* New York: Springer-Verlag.

Härdle, W., Hall, P. and Ichimura, H. (1993), "Optimal Smoothing in Single Index Model," *Annals of Statistics* 21, 157-178.

Härdle, W., Horowitz, J. and Kreiss, J.P. (2003), "Bootstrap Methods for Time Series," *International Statistical Review* 71, 435-459.

Hillebrand, E., Lee, T.-H., and Medeiros, M. (2009), "Bagging Constrained Forecasts with Application to Forecasting Equity Premium," *JSM Proceedings for Business and Economic Statistics.*

Horowitz, J.L. (2001), "The Bootstrap," *Handbook of Econometrics*, Volume 5, Edited by J.J. Heckman and E. Leamer.

Ichimura, H. (1993), "Semiparametric Least Squares (SLS) and Weighted SLS Estimation of Single-index Models," *Journal of Econometrics* 58, 71–120.

Li, Q. and Racine, J.S. (2007), *Nonparametric Econometrics*, Princeton University Press.

Pagan, A. and Ullah, A. (1999), *Nonparametric Econometrics*, Cambridge University Press.

Wang, Q. and Phillips, P.C.B. (2009a), "Asymptotic Theory for Local Time Density Estimation and Nonparametric Cointegrating Regression," *Econometric Theory* 25, 710-738.

Wang, Q. and Phillips, P.C.B. (2009b), "Structural Nonparametric Cointegrating Regression," *Econometrica* 77, 1901-1948.

Xia, Y., Tong, H. and Li, W.K. (1999), "On Extended Partially Linear Single-index Models," *Biometrika* 86, 831–842.

Figure 1: Asymptotic properties of the local historical average with positivity constraint (LHA$_p$, solid line) and the local historical average with positivity constraint and bagging (LHA$_{pb}$, dashed line)
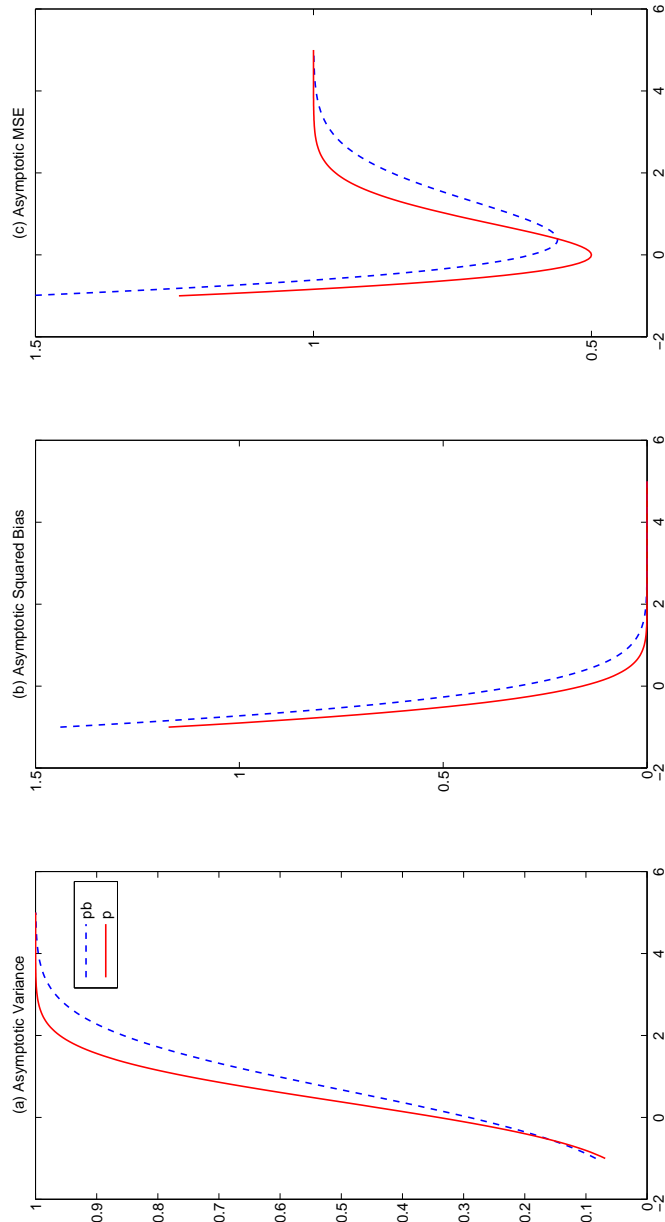
Table 1: Simulation Results for DGP 1: evaluation point, $x = 1$

	$n = 50$		$n = 100$		$n = 200$		$n = 50$		$n = 100$		$n = 200$	
a	p	pb[a]	p	pb	p	pb	p	pb	p	pb	p	pb
	$\rho = 0$						$\rho = 0.1$					
0.001	0.522	0.561	0.556	0.589	0.619	0.644	0.506	0.525	0.624	0.657	0.461	0.483
0.004	0.516	0.539	0.586	0.595	0.572	0.579	0.606	0.633	0.573	0.583	0.534	0.548
0.007	0.644	0.661	0.673	0.657	0.545	0.525	0.670	0.679	0.753	0.766	0.555	0.549
0.010	0.619	0.638	0.707	0.676	0.765	0.725	0.784	0.779	0.720	0.720	0.614	0.598
0.020	0.689	0.681	0.813	0.790	0.818	0.750	0.742	0.715	0.939	0.919	0.745	0.695
	$\rho = 0.9$						$\rho = 1$					
0.001	0.605	0.634	0.475	0.521	0.538	0.582	0.412	0.356	0.518	0.551	0.587	0.616
0.004	0.550	0.580	0.638	0.663	0.785	0.804	0.481	0.442	0.672	0.712	0.649	0.687
0.007	0.757	0.771	0.610	0.623	0.809	0.813	0.474	0.397	0.811	0.844	0.862	0.901
0.010	0.615	0.623	0.483	0.494	0.946	0.942	0.583	0.447	0.825	0.821	0.895	0.893
0.020	0.707	0.700	0.707	0.697	1.000	0.986	0.743	0.597	0.981	0.897	0.998	0.993

[a]The column "p" denotes Relative MSE of LHA$_p$ to that of LHA and "pb" denotes that of LHA$_{pb}$ to LHA.

Table 2: Simulation Results for DGP 1: evaluation point, $x = 1.5$

	$n=50$		$n=100$		$n=200$		$n=50$		$n=100$		$n=200$	
a	p	pb[a]	p	pb	p	pb	p	pb	p	pb	p	pb
	$\rho=0$						$\rho=0.1$					
0.001	0.555	0.552	0.398	0.389	0.723	0.692	0.534	0.524	0.510	0.489	0.674	0.641
0.004	0.682	0.618	0.759	0.708	0.824	0.715	0.613	0.539	0.607	0.541	0.843	0.763
0.007	0.900	0.802	0.796	0.732	0.957	0.905	0.834	0.754	0.976	0.917	0.990	0.936
0.010	0.997	0.965	0.983	0.920	0.964	0.956	0.8589	0.803	0.982	0.921	1.000	0.999
0.020	1.000	0.995	1.000	0.992	1.000	0.995	1.000	0.972	1.000	0.982	1.000	1.000
	$\rho=0.9$						$\rho=1$					
0.001	0.389	0.404	0.619	0.631	0.717	0.698	0.403	0.392	0.518	0.511	0.452	0.474
0.004	0.681	0.585	0.837	0.796	0.814	0.754	0.407	0.307	0.465	0.342	0.632	0.607
0.007	0.752	0.643	0.948	0.905	0.979	0.912	0.436	0.331	0.563	0.445	0.745	0.650
0.010	0.782	0.702	0.991	0.960	0.968	0.934	0.415	0.295	0.616	0.498	0.877	0.790
0.020	0.987	0.941	1.000	1.010	1.000	0.983	0.660	0.602	0.926	0.759	1.000	0.958

[a]The column "p" denotes Relative MSE of LHA$_p$ to that of LHA and "pb" denotes that of LHA$_{pb}$ to LHA.

Table 3: Simulation Results for DGP 2: evaluation point, $x = (1,1,1)'$

	$n=50$		$n=100$		$n=200$		$n=50$		$n=100$		$n=200$	
a	p	pb[a]	p	pb	p	pb	p	pb	p	pb	p	pb
	$\rho=0$, NW						$\rho=0$, SIM					
0.001	0.558	0.595	0.581	0.620	0.512	0.539	0.568	0.563	0.549	0.514	0.558	0.516
0.004	0.451	0.487	0.571	0.578	0.618	0.624	0.516	0.509	0.585	0.471	0.659	0.541
0.007	0.594	0.602	0.708	0.695	0.607	0.574	0.597	0.545	0.664	0.585	0.766	0.586
0.010	0.595	0.611	0.780	0.753	0.705	0.681	0.594	0.535	0.783	0.567	0.771	0.612
0.020	0.699	0.648	0.954	0.926	0.824	0.772	0.809	0.594	0.964	0.716	0.962	0.693
	$\rho=0.9$, NW						$\rho=0.9$, SIM					
0.001	0.600	0.623	0.469	0.501	0.528	0.575	0.610	0.589	0.552	0.532	0.531	0.451
0.004	0.554	0.569	0.470	0.486	0.701	0.707	0.585	0.530	0.732	0.550	0.782	0.621
0.007	0.599	0.587	0.518	0.523	0.581	0.587	0.688	0.588	0.875	0.562	0.699	0.545
0.010	0.519	0.524	0.602	0.592	0.661	0.661	0.681	0.590	0.946	0.533	0.862	0.635
0.020	0.673	0.656	0.352	0.336	0.809	0.784	0.853	0.741	0.984	0.537	0.941	0.706

[a]The column "p" denotes Relative MSE of LHA_p to that of LHA and "pb" denotes that of LHA_{pb} to LHA. NW denotes the Nadaraya-Watson estimator, while SIM denotes the estimator obtained from Single Index Model.

Table 4: Empirical Results

Panel A. Forecasting Annualized Equity Premium $Q_t(k = 12)$ at Month t

	Forecast begins from 1960:01			Forecast begins from 1980:01		
	LHA/GHA	LHA$_p$/GHA	LHA$_{pb}$/GHA	LHA/GHA	LHA$_p$/GHA	LHA$_{pb}$/GHA
d/p	0.974	0.963	0.963	0.984	0.975	0.978
e/p	0.979	0.966	0.966	0.990	0.981	0.981
se/p	1.045	1.016	1.009	0.986	0.978	0.976
b/m	1.012	0.999	0.993	1.007	0.997	0.993
roe	1.010	1.000	0.999	1.003	0.999	0.998
$t\text{-}bill$	1.013	0.986	0.983	1.002	0.996	0.993
lty	1.017	0.997	0.989	0.977	0.973	0.971
ts	1.071	1.031	1.017	0.923	0.929	0.932
ds	0.965	0.907	0.899	0.907	0.879	0.871
inf	0.990	0.975	0.979	1.044	1.013	1.008
nei	0.963	0.962	0.962	0.850	0.851	0.848
$index$	1.006	0.998	0.985	1.033	1.028	1.003

Panel B. Forecasting Monthly Equity Premium $Q_t(k = 1)$ at Month t

	Forecast begins from 1960:01			Forecast begins from 1980:01		
	LHA/GHA	LHA$_p$/GHA	LHA$_{pb}$/GHA	LHA/GHA	LHA$_p$/GHA	LHA$_{pb}$/GHA
d/p	1.015	0.993	0.991	0.996	0.994	0.990
e/p	1.028	0.992	0.991	0.993	0.992	0.989
se/p	1.035	1.007	1.003	1.002	1.000	0.996
b/m	1.008	1.004	0.998	0.999	0.999	0.995
roe	1.043	1.021	1.025	1.002	1.000	0.995
$t\text{-}bill$	1.047	1.026	1.015	1.017	1.024	1.002
lty	1.029	1.022	1.008	1.011	1.008	0.998
ts	1.025	1.011	1.024	1.090	1.053	1.046
ds	1.012	1.005	1.009	1.034	1.021	1.022
inf	1.011	0.999	0.997	1.034	1.025	1.023
nei	1.030	1.012	1.007	1.048	1.023	1.008
$index$	1.023	1.011	1.002	1.024	1.012	1.003