Journal of Empirical Finance 26 (2014) 13–25

Contents lists available at ScienceDirect

Journal of Empirical Finance

journal homepage: www.elsevier.com/locate/jempfin

Predictive regression: An improved augmented regression method☆

Jae H. Kim

Department of Finance, La Trobe Business School, La Trobe University, VIC 3086, Australia

ARTICLE INFO

Article history:
Received 29 August 2012
Received in revised form 8 December 2013
Accepted 10 January 2014
Available online 21 January 2014

JEL classification:
C32
G12
G14

Keywords:
Financial ratios
Forecasting
Return predictability
Stationarity-correction
Statistical power

ABSTRACT

This paper proposes three modifications to the augmented regression method (ARM) for bias-reduced estimation and statistical inference in the predictive regression. They are in relation to improved bias-correction, stationarity-correction, and the use of matrix formulae for bias-correction and covariance matrix estimation. The improved ARM parameter estimators are unbiased to the order of n^{-1}, and always satisfy the condition of stationarity. With the matrix formulae, the improved ARM can easily be implemented for a high order model with multiple predictors. From an extensive Monte Carlo experiment, it is found that the improved ARM delivers substantial gain in parameter estimation, statistical inference, and out-of-sample forecasting in small samples. As an application, the improved ARM is applied to monthly US stock return data to evaluate the predictive power of dividend yield in univariate and bivariate predictive models.

1. Introduction

Stock return predictability is an issue of perennial interest in empirical finance. Central to this issue is predictive regression, where stock return is regressed against the past values of a predictor. The latter often includes a financial ratio such as dividend yield, earnings–price ratio, and book-to-market ratio, which measure stock prices relative to the fundamentals. Predictability of stock return is evaluated by testing for statistical significance of the coefficient associated with the past value of a predictor. Notable research works on predictive regression include Nelson and Kim (1993), Stambaugh (1999), Lewellen (2004), Amihud and Hurvich (2004), Cochrane (2008), Lettau and Van Nieuwerburgh (2008), Welch and Goyal (2007), Ang and Bekaert (2007), Amihud et al. (2008), and Amihud et al. (2010). As Lettau and Ludvigson (2001; p.942) state, it is widely accepted that key financial indicators can predict (excess) stock returns. However, as Welch and Goyal (2007; p.1455) point out, the evidence is rather mixed and often conflicting, depending on the techniques, variables, and time periods employed. This can be attributed to a number of factors, such as estimation biases (Nelson and Kim, 1993; Stambaugh, 1999), persistence of predictors (Lettau and Van Nieuwerburgh, 2008), instability of the relationship (Paye and Timmermann, 2006), and data mining (Malkiel, 2003).

In this paper, I pay attention to small sample biases in parameter estimation, which is an issue of importance especially when the predictor is highly persistent. Nelson and Kim (1993) and Stambaugh (1999) have found that small sample bias in the estimation of predictive coefficient can cause gross over-statement of its statistical significance: see also Cochrane (2008, p.1534).

☆ I would like to thank the referee for constructive comments that substantially improved the paper in many ways. An earlier version of the paper was presented at the 2012 Australian Finance and Banking Conference. The computational resources written in programming language R (R Core Team, 2012) are available on request. I would like to thank Jon Lewellen for sharing the data used in this paper. Helpful comments from Abul Shamsuddin are gratefully acknowledged.

0927-5398/$ – see front matter
http://dx.doi.org/10.1016/j.jempfin.2014.01.002

In particular, Stambaugh (1999) showed, for the case of single predictor with lag order 1, that the least-squares estimator of predictive coefficient is biased in small samples. Lewellen (2004), on the other hand, argues that the Stambaugh's (1999) method can substantially under-state predictive power, if the test is conditional on the most conservative estimate of the autoregressive coefficient of predictor. Amihud and Hurvich (2004) and Amihud et al. (2008) propose a novel bias-corrected estimation and inferential method for predictive regression with the lag order $p = 1$. Amihud et al. (2010) extend this method to the general case of lag order $p > 1$ with multiple predictors. Their method is referred to as the augmented regression method (ARM), since bias-reduced estimation and statistical inference are conducted by augmenting the predictive regression. Although Amihud et al. (2010) finds that their ARM outperforms the least-squares (LS) method in parameter estimation and statistical inference, the ARM-based statistical tests still show substantial size distortion when the sample size is small and the predictor is highly persistent. In addition, Amihud et al. (2010) do not provide general formulae for bias-correction and covariance matrix estimation for $p > 2$, limiting the applicability of their methods in practice. Moreover, their ARM runs into a problem when the bias-correction renders the predictive model non-stationary, as also noted in Stambaugh (1999; p.384).

This paper proposes an improved ARM for predictive regression, with three modifications to the ARM of Amihud et al. (2010). First, an improved bias-correction method is proposed, which provides predictive coefficient estimator with a faster rate of convergence to the true value than the ARM. Second, the stationarity-correction of Kilian (1998) is implemented for bias-correction to ensure that bias-corrected parameter estimates satisfy the condition of stationarity. Third, the improved ARM incorporates matrix formulae for bias-correction and covariance matrix estimation, applicable to multiple predictors of any lag order. With improved bias-correction equipped with stationarity-correction, it is quite possible that the improved ARM shows desirable small sample properties, especially when the predictor is highly persistent.

This main finding of the paper is that the improved ARM delivers substantial gain in small samples. It outperforms the ARM in reducing estimation bias and in controlling the size of the test on predictive coefficients with satisfactory properties in its statistical power. It also performs better than the ARM in the accuracy of out-of-sample and multi-step forecasts. As an application, I apply the improved ARM to the monthly U.S. stock market data set used in Lewellen (2004), in order to evaluate the predictive power of dividend yield for stock return, in both univariate and bivariate settings assuming AR(1) or AR(4) predictors. It is found that the dividend yield shows statistically significant predictive power for the stock return for the period from 1946 to 1994. However, if the data from 1995 are included, the predictive power dividend yield disappears. From 1995 to 2009, the dividend yield shows no predictive power for stock return. The next section presents the predictive regression with a brief review of the key past studies. Section 3 presents the improved ARM for predictive regression. Section 4 presents Monte Carlo results, and Section 5 an empirical application. Section 6 concludes the paper.

2. Predictive regression

The predictive regression with an autoregressive predictor of order 1 (denoted AR(1)) is written as

$$Y_t = \beta_0 + \beta_1 X_{t-1} + u_t; X_t = \alpha_0 + \alpha_1 X_{t-1} + v_t. \tag{1}$$

Y_t is typically (excess) stock return at time t and X_t is a predictor such as dividend yield. It is assumed that X_t is stationary, with the value of α_1 typically less than 1. The error terms $(u_t,v_t)'$ follow independent and identically distributed bivariate normal distribution with covariance matrix $\Sigma \equiv vech(\sigma_u^2,\sigma_{uv},\sigma_v^2)$. Let $\left(\hat{\beta}_1, \hat{\alpha}_1\right)$ denote the ordinary least squares (LS) estimators for (β_1,α_1). The hypothesis of interest is $\beta_1 = 0$, where the predictor X shows no predictive power for Y. Stambaugh (1999) shows that

$$E\left(\hat{\beta}_1 - \beta_1\right) = \phi E(\hat{\alpha}_1 - \alpha_1) = -\frac{\sigma_{uv}}{\sigma_v^2}\left(\frac{1+3\alpha_1}{n}\right) + O\left(n^{-2}\right) \tag{2}$$

where $O()$ denotes the order of magnitude,[1] n sample size, and $\phi \equiv \sigma_{uv}/\sigma_v{}^2$. It means that the LS estimators for (β_1,α_1) are biased to the order of n^{-1}, with their biases decreasing at the rate of n^{-1}. Stambaugh (1999; Section 2) derives the exact moments of $\left(\hat{\beta}_1 - \beta_1\right)$ and expression for the p-value for H_0: $\beta_1 = 0$ against H_1:$\beta_1 > 0$; and shows that $\hat{\beta}_1$ is biased upward when the predictor is highly persistent. In addition, the t-test for H_0: $\beta_1 = 0$ based on the LS method is biased, with its p-value much smaller than the true one in finite samples. Lewellen (2004) proposes a test for β_1 conditional on 0.9999, which is the most conservative value of α_1. That is, the bias-adjusted estimator is obtained as

$$\hat{\beta}_{1,adj} = \hat{\beta}_1 - \phi(\hat{\alpha}_1 - 0.9999), \tag{3}$$

and the test on β_1 is conducted using the sampling distribution of $\hat{\beta}_{1,adj}$. According to Lewellen (2004), this conditional test is useful when $\hat{\alpha}_1 - 1 > -(1 + 3\hat{\alpha}_1)/n$, where the bias-corrected estimate of α_1 is greater than 1 implying non-stationarity. Using this test, he finds that the dividend yield has been a strong predictor for the US stock return, in contrast with Stambaugh's (1999) results.

[1] A sequence that is decreasing at the rate of n^{-d} is denoted $O(n^{-d})$. For example, $(1 + 3\alpha_1) / n = O(n^{-1})$.

Amihud et al. (2010) propose the augmented regression method (ARM) for a bias-reduced estimation and inference for predictive regression, extending the earlier works of Amihud and Hurvich (2004) and Amihud et al. (2008) for an AR(1) predictor. Amihud et al. (2010) consider the case of an AR(p) predictor, which can be written as

$$\begin{aligned} Y_t &= \beta_0 + \beta_1 X_{t-1} + \ldots + \beta_p X_{t-p} + u_t \\ X_t &= \alpha_0 + \alpha_1 X_{t-1} + \ldots + \alpha_p X_{t-p} + v_t. \end{aligned} \tag{4}$$

The Eq. (4) represents the case of univariate predictor, but it can be extended to the case of multiple predictors, as we shall see later in this paper. There are two hypotheses of interest in relation to no predictive power of the predictor X for Y: H_{01}: $\beta_1 = \ldots = \beta_p = 0$; and H_{02}: $\beta_1 + \ldots + \beta_p = 0$. Under H_{01}, individual h-period dynamic multipliers relating to X to Y ($h = 1, \ldots, p$) are jointly zero; while under H_{02}, the long-run cumulative multiplier is zero.

The LS estimators for $\alpha \equiv (\alpha_1,\ldots,\alpha_p)$ and for $\beta \equiv (\beta_1,\ldots, \beta_p)$ are denoted as $\hat{\alpha} \equiv (\hat{\alpha}_1, \ldots, \hat{\alpha}_p)$ and $\hat{\beta} \equiv \left(\hat{\beta}_1, \ldots, \hat{\beta}_p\right)$. Shaman and Stine (1988) derive the bias formulae to $O(n^{-1})$ for $\hat{\alpha}$, from which bias-corrected estimators $\hat{\alpha}^c \equiv \left(\hat{\alpha}_1^c, \ldots, \hat{\alpha}_p^c\right)$ for α's are obtained. Amihud et al. (2010) provides an example for $p = 2$. That is,

$$\hat{\alpha}_1^c = \hat{\alpha}_1 + \frac{1 + \hat{\alpha}_1 + \hat{\alpha}_2}{n}; \hat{\alpha}_2^c = \hat{\alpha}_2 + \frac{2 + 4\hat{\alpha}_2}{n}. \tag{5}$$

According to the ARM of Amihud et al. (2010), bias-corrected estimator $\hat{\beta}^c \equiv \left(\hat{\beta}_1^c, \ldots, \hat{\beta}_p^c, \hat{\varphi}^c\right)$ for $\beta \equiv (\beta_1, \ldots, \beta_p, \phi)$ is obtained by regressing Y_t on $(1, X_{t-1}, \ldots, X_{t-p}, v_t^c)$, where $v_t^c = X_t - \hat{\alpha}_0^c - \hat{\alpha}_1^c X_{t-1} - \ldots - \hat{\alpha}_p^c X_{t-p}$. Amihud et al. (2010; Theorem 1) show that

$$E\left(\hat{\beta}_i^c - \beta_i\right) = \phi E\left(\hat{\alpha}_i^c - \alpha_i\right) \tag{6}$$

for $i = 1, \ldots p$. It follows from (5) and (6) that $E\left(\hat{\beta}_i^c - \beta_i\right) = O(n^{-1})$, which means that the bias-corrected estimator of a predictive coefficient is biased to the order of n^{-1}, with its bias diminishing at the rate of n^{-1}.

For statistical inference, the covariance matrix of $\left(\hat{\beta}_1^c, \ldots, \hat{\beta}_p^c\right)$ should be estimated. Amihud et al. (2010) suggest the estimator

$$Cov^c\left(\hat{\beta}_i^c, \hat{\beta}_j^c\right) = \hat{\phi}^{c2} Cov\left(\hat{\alpha}_i^c, \hat{\alpha}_j^c\right) + Cov\left(\hat{\beta}_i^c, \hat{\beta}_j^c\right) \tag{7}$$

where $Cov\left(\hat{\beta}_i^c, \hat{\beta}_j^c\right)$ is obtained from the augmented regression and $Cov\left(\hat{\alpha}_i^c, \hat{\alpha}_j^c\right)$ is obtained from the bias-formulae given by Shaman and Stine (1988). For example, when $p = 2$, the formulae for $Cov\left(\hat{\alpha}_i^c, \hat{\alpha}_j^c\right)$ are derived by evaluating the moments of (5): see Amihud et al. (2010; p.517). For a higher order case, tedious algebra is required to find the expression for $Cov\left(\hat{\alpha}_i^c, \hat{\alpha}_j^c\right)$. Amihud et al. (2010) conduct a Monte Carlo experiment to find that their ARM outperforms the LS method, with substantial bias reduction in parameter estimation. They also find that their bias-reduced t and F-tests provide a better control of the size of the statistical tests on β's than the conventional tests. However, the ARM still shows a degree of size distortion when the sample size is small and the predictor is highly persistent (see Tables 1 and 3 of Amihud et al., 2010).

3. Improved augmented regression method

I propose an improved ARM by providing three modifications to the ARM discussed in Section 2. First, I adopt a matrix formula for bias-corrected parameter estimators for the $AR(p)$ model of any order, which converge to the true parameters at a faster rate than those of Amihud et al. (2010). Second, I adopt a matrix formula for covariance matrix estimation for the bias-corrected estimators which is operational for $AR(p)$ predictor model of any p. Third, when bias-correction pushes the parameter estimates $\left(\hat{\alpha}_1^c, \ldots, \hat{\alpha}_p^c\right)$ to the non-stationary part of the parameter space, the improved ARM makes an adjustment to these estimates in order to maintain the stationarity. This adjustment is crucial because (i) it makes little sense to predict stock return with non-stationary predictors (Lewellen, 2004; p.203); and (ii) the bias formulae for the LS parameter estimators given by Shaman and Stine (1988) are derived under the assumption of a stationary $AR(p)$ model.

3.1. Bias-correction and covariance matrix estimation

According to Patterson (2000, p. 138), $\hat{\alpha}_i^c$ given in (5) is not the best solution because they still tend to under-estimate the true values. Kim (2004) derives a simple matrix formula for bias-corrected estimators for the parameters AR(p) model, which converges to the true values at the rate of n^{-2}. Stine and Shaman (1989) provide analytical expressions for the first order biases (to the order of n^{-1}) of the LS parameter estimators for an AR model of known and finite order in a matrix form, which can be written as

$$E(\hat{\gamma})-\gamma=-\left(A_{p_1}+A_{p_2}\gamma\right)/n+O\left(n^{-2}\right). \tag{8}$$

where $\gamma=(\alpha_1,\ldots,\alpha_p)'$ and $\hat{\gamma}=(\hat{\alpha}_1,\ldots,\hat{\alpha}_p)'$. The rows and columns of A_{p1} and A_{p2} matrices are regular sequences indexed by p (see, for details, Stine and Shaman, 1989). The bias-corrected estimator for γ can obtained by substituting $\hat{\gamma}$ for $E(\hat{\gamma})$ in (8) and solving for γ, which yields

$$\hat{\gamma}^a\equiv\left(\hat{\alpha}_1^a,\ldots,\hat{\alpha}_p^a\right)=\left(nI_p-A_{p2}\right)^{-1}\left(A_{p1}+n\hat{\gamma}\right)+O\left(n^{-2}\right), \tag{9}$$

where I_p is the $p\times p$ identity matrix. The bias-corrected estimator for α_0 is obtained as $\hat{\alpha}_0^a=\left(1-\sum_{i=1}^{p}\hat{\alpha}_i^a\right)\overline{X}$, where $\overline{X}$ is the sample mean of X_t. For the AR(2) model, the matrix formula given in (9) yields

$$\begin{bmatrix}\hat{\alpha}_1^a\\ \hat{\alpha}_2^a\end{bmatrix}=\begin{bmatrix}\dfrac{1+n\hat{\alpha}_1}{n-1}+\dfrac{2+n\hat{\alpha}_2}{(n-1)(n-4)}\\ \dfrac{2+n\hat{\alpha}_2}{n-4}\end{bmatrix}.$$

Kim (2004) shows that $E(\hat{\gamma}^a-\gamma)=O(n^{-2})$, which indicates that $\hat{\gamma}^a$ given in (9) is unbiased to $O(n^{-1})$. That is, the bias-corrected estimator (9) converges to the true parameter value at a faster rate (n^{-2}) than the conventional estimators given in (5).

Following the ARM of Amihud et al. (2010), the improved bias-corrected estimator $\hat{\beta}^a\equiv\left(\hat{\beta}_1^a,\ldots,\hat{\beta}_p^a,\hat{\phi}^a\right)$ for $\beta\equiv(\beta_1,\ldots,\beta_p,\phi)$ is obtained by regressing Y_t on $(1, X_{t-1},\ldots,X_{t-p}, v_t^a)$,where $v_t^a=X_t-\hat{\alpha}_0^a-\hat{\alpha}_1^aX_{t-1}-\ldots-\hat{\alpha}_p^aX_{t-p}$. It follows from (6) that

$$E\left(\hat{\beta}_i^a-\beta_i\right)=\phi E(\hat{\alpha}_i^a-\alpha_i)=O\left(n^{-2}\right),$$

while $E\left(\hat{\beta}_i^c-\beta_i\right)=\phi E(\hat{\alpha}_i^c-\alpha_i)=O(n^{-1})$ as we have seen earlier. That is, the improved ARM provides the bias-corrected estimators for the predictive coefficients unbiased to $O(n^{-1})$, with their biases diminishing at the rate of n^{-2}. It is worth noting that Chiquoine and Hjalmarsson (2009) also proposed the bias-corrected estimator of the predictive coefficient unbiased to $O(n^{-1})$, based on the jackknife method when $p=1$. From (9), the matrix formula for the covariance matrix for $\hat{\gamma}^a$ can be obtained:

$$Cov(\hat{\gamma}^a)=\left(I_p-A_{p2}/n\right)^{-1}Cov(\hat{\gamma})\left[\left(I_p-A_{p2}/n\right)^{-1}\right]' \tag{10}$$

where $Cov(\hat{\gamma})$ is the covariance matrix of $\hat{\gamma}=\left(\hat{\alpha}_1,\ldots,\hat{\alpha}_p\right)$. Taking this into consideration, the improved covariance matrix estimator for $\left(\hat{\beta}_1^c,\ldots,\hat{\beta}_p^c\right)$ is given by

$$Cov^a\left(\hat{\beta}_i^c,\hat{\beta}_j^c\right)=\hat{\phi}^{c2}Cov\left(\hat{\alpha}_i^a,\hat{\alpha}_j^a\right)+Cov\left(\hat{\beta}_i^c,\hat{\beta}_j^c\right).$$

3.2. Stationarity-correction

It is possible that bias-corrected estimate is pushed to non-stationarity, while the underlying model is stationary. An example can be found in Table 2 of Lewellen (2004), where it appears that $\hat{\alpha}_1=0.997$, $Bias(\hat{\alpha}_1)=-(1+3\hat{\alpha}_1)/660=-0.006$, which yields $\hat{\alpha}_1^c=1.003$ and $\hat{\alpha}_1^a=1.003$.[2] To adjust for this, I employ Kilian's (1998) stationarity-correction, which can be described as follows:

If $\hat{\gamma}^a$ implies non-stationarity, then let $\delta_1=1$, $\Delta_1=Bias(\hat{\gamma})$ and $\hat{\gamma}^a=\hat{\gamma}-\Delta_1$. Set $\Delta_{i+1}=\delta_i\Delta_i$, $\delta_{i+1}=\delta_i-0.01$ for $i=1,2,3,\ldots$. Iterate until $\hat{\gamma}_i^a$ satisfies the condition of stationarity and set $\hat{\gamma}^a=\hat{\gamma}_i^a$. If $\hat{\gamma}^a$ implies stationarity, no correction is required.

The correction shrinks the bias until the bias-corrected estimate satisfies the condition of stationarity. Kilian (1998: Section IIIC) provides the theoretical justification for this correction. In Lewellen's (2004) example given above, $\hat{\alpha}_1^a$ is adjusted to 0.9998 (similarly for $\hat{\alpha}_1^c$). This adjustment provides an adjusted estimate of α_1, fairly close to its most conservative value of 0.9999 used in Lewellen (2004). However, Kilian's (1998) correction is implemented only when the stationarity condition is violated.

[2] In this case, the two bias-corrected estimators are identical to the third decimal points (rounded off figures) because the sample size is large ($n=660$). But, in general, they give numerically different estimates.

3.3. Improved ARM for multiple predictors

Similarly to Amihud et al. (2010), the improved ARM is also applicable to the case of multiple predictors where the vector of predictors follows a vector AR model of order p with diagonal coefficient matrices. For the bivariate predictor case order p, the model is written as

$$\begin{aligned} Y_t &= \beta_0 + \beta_{11}X_{1t-1} + \ldots + \beta_{1p}X_{1t-p} + \beta_{21}X_{2t-1} + \ldots + \beta_{2p}X_{2t-p} + u_t; \\ X_{1t} &= \alpha_1 + \alpha_{11}X_{1t-1} + \ldots + \alpha_{1p}X_{1t-p} + v_{1t} \\ X_{2t} &= \alpha_2 + \alpha_{21}X_{2t-1} + \ldots + \alpha_{2p}X_{2t-p} + v_{2t}. \end{aligned} \tag{11}$$

Extension to the case of higher number of predictors ($k \geq 3$) is straightforward. The model follows all the assumptions given in Amihud et al. (2010; Section 4). Following the ARM of Amihud et al. (2010), the improved ARM is conducted based on the augmented regression of the following form:

$$Y_t on\, 1, X_{1t-1}, \ldots, X_{1t-p}, X_{2t-1}, \ldots, X_{2t-p}, \hat{v}_1^a\, and\, \hat{v}_2^a,$$

where $\hat{v}_{it}^a = X_t - \hat{\alpha}_i^a - \hat{\alpha}_{i1}^a X_{it-1} - \ldots - \hat{\alpha}_{ip}^a X_{it-p}$ and $\hat{\gamma}_i^a \equiv \left(\hat{\alpha}_i^a, \hat{\alpha}_{i1}^a, \ldots, \hat{\alpha}_{ip}^a\right)$ $(i = 1, 2)$ are the bias-corrected estimators for $(\alpha_i, \alpha_{i1}, \ldots \alpha_{ip})$ using the formula (9) implementing Kilian's (1998) stationarity-correction when required. The bias-corrected covariance matrix estimation is conducted in a similar manner as in the univariate predictor case. That is,

$$\begin{aligned} Cov(\hat{\gamma}_1^a) &= \left(I_p - A_{p2}/n\right)^{-1} Cov(\hat{\gamma}_1)[\left(I_p - A_{p2}/n\right)^{-1}]'; \\ Cov(\hat{\gamma}_2^a) &= \left(I_p - A_{p2}/n\right)^{-1} Cov(\hat{\gamma}_2)[\left(I_p - A_{p2}/n\right)^{-1}]'; and \\ Cov(\hat{\gamma}_1^a, \hat{\gamma}_2^a) &= \left(I_p - A_{p2}/n\right)^{-1} Cov(\hat{\gamma}_1, \hat{\gamma}_2)[\left(I_p - A_{p2}/n\right)^{-1}]'; \end{aligned}$$

where $\hat{\gamma}_1 = \left(\hat{\alpha}_{11}, \ldots, \hat{\alpha}_{1p}\right)'; \hat{\gamma}_2 = \left(\hat{\alpha}_{21}, \ldots, \hat{\alpha}_{2p}\right)'; Cov(\hat{\gamma}_1^a, \hat{\gamma}_2^a) = \left(X_1' X_1\right)^{-1} X_1' \hat{v}_1 \hat{v}_2' X_2 \left(X_1' X_1\right)^{-1}$, while X_1 and X_2 are the data matrices of $[X_{1t-1}; \ldots, X_{1t-p}]$ and $[X_{2t-1}, \ldots, X_{2t-p}]$ and $\hat{v}_i$'s are vector of residuals.

Note that the improved ARM for multiple predictors presented here is applicable when the predictors follow a vector AR model of order p with diagonal coefficient matrices. This means that X_1 and X_2 in (11) are not allowed of being inter-related dynamically. Amihud and Hurvich (2004) and Amihud et al. (2008) consider a case where the predictors follow a full vector AR(1) model, in which X_1 and X_2 can be inter-related dynamically. They adopt the bias formula derived by Nicholls and Pope (1988) for vector AR parameter estimators for bias-correction. This bias formula is of $O(n^{-1})$, as in the univariate AR case of Shaman and Stine (1988) and Stine and Shaman (1989). To the best of my knowledge, a bias-correction method which delivers bias-corrected estimators unbiased to $O(n^{-1})$ for a vector AR model is not available. Hence, the improved ARM as proposed in this paper is not applicable to the vector AR case, although Kilian's (1998) stationarity-correction is still applicable. Given this technical difficulty, this line of research is left as a subject of future research.

4. Monte Carlo experiment

A Monte Carlo experiment is conducted under the same simulation setting as in Amihud et al. (2010), in order to evaluate the properties of the improved ARM. First, I evaluate the bias and standard deviation of alternative predictive coefficient estimators. Second, I evaluate the size and power properties of the t-test and F-test for predictive coefficients. The size refers to the probability of rejecting the true null hypothesis; while the power is the probability of rejecting a false null hypothesis. For sound statistical inference, the size value should be close to the level of significance, with the power being reasonably high when the null hypothesis is violated to an economically meaningful extent. Third, I evaluate out-of-sample and multi-step ahead predictive performance of the improved ARM.

4.1. Experimental design

For the case of univariate predictor, Amihud et al. (2010) simulate a predictive model with an AR(2) predictor of the form

$$\begin{aligned} Y_t &= \beta_0 + \beta_1 X_{t-1} + \beta_2 X_{t-2} + u_t \\ X_t &= \alpha_0 + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + v_t. \end{aligned}$$

Amihud et al. (2010) consider two types of predictor X: Case I where $\alpha_1 = 1.1053$ and $\alpha_2 = -0.1430$; and Case II where $\alpha_1 = 1.0553$ and $\alpha_2 = -0.1430$. Case I represents a highly persistent predictor with the dominant root of the model being 0.9557; while Case II a less persistent one with the dominant root 0.8956. The error terms u and v are highly correlated with the correlation coefficient -0.9685, with $u_t = -92.17v_t + e_t$ where e_t and v_t are generated from zero-mean normal distributions

Table 1
Bias and standard deviation in parameter estimation of predictive coefficients: bivariate predictor, Case I.

	LS		ARM		Improved ARM	
	Bias	Stand dev	Bias	Stand dev	Bias	Stand dev
Error terms: normal distributions						
n = 50						
β_{11}	7.37	24.72	1.95	14.97	1.18	14.70
β_{12}	1.80	24.12	−0.42	14.95	0.08	14.51
β_{21}	8.02	31.44	2.04	14.49	1.18	15.07
β_{22}	1.65	30.94	−0.69	14.43	0.08	14.84
n = 200						
β_{11}	1.49	11.34	0.21	6.86	0.03	6.79
β_{12}	0.43	11.21	−0.005	6.83	−0.008	6.75
β_{21}	1.65	14.50	0.004	6.92	0.26	6.92
β_{22}	0.61	14.33	0.08	6.94	−0.10	6.84
Error terms: fat-tailed distributions						
n = 50						
β_{11}	9.81	25.49			2.53	14.43
β_{12}	1.01	24.57			−0.13	14.07
β_{21}	9.18	31.83			2.61	14.68
β_{22}	0.75	31.54			−0.13	14.46
n = 200						
β_{11}	2.03	11.23			0.34	6.71
β_{12}	0.15	11.15			−0.19	6.78
β_{21}	1.64	14.28			0.12	6.86
β_{22}	0.63	14.34			0.05	6.84

The columns labelled ARM contain the results based on augmented regression method proposed by Amihud et al (2010). These figures are reproduced from their Table 3 of Amihud et al (2010).
The columns labelled "Improved ARM" contain the results based on the improved version of the augmented regression method proposed in this paper.
The column labelled "LS" contain the results based on the LS estimation.

with standard deviations 0.01844 and 0.0007746 respectively. The hypotheses of interest are H_0: $\beta_i = 0$ against H_1: $\beta_i > 0$[3] for all i; or H_{01}: $\beta_1 = \beta_2 = 0$ (no predictive power of X). To evaluate the size properties, the probability of rejecting the null hypothesis is calculated setting $\beta_1 = \beta_2 = 0$.

For the case of bivariate predictors, Amihud et al. (2010) simulate a model of the form

$$\begin{aligned} Y_t &= \beta_0 + \beta_{11}X_{1t-1} + \beta_{12}X_{1t-2} + \beta_{21}X_{2t-1} + \beta_{22}X_{2t-2} + u_t; \\ X_{1t} &= \alpha_1 + \alpha_{11}X_{1t-1} + \alpha_{12}X_{1t-2} + v_{1t} \\ X_{2t} &= \alpha_2 + \alpha_{21}X_{2t-1} + \alpha_{22}X_{2t-2} + v_{2t} \end{aligned}$$

considering two different sets of α values (Cases I and II as in the univariate case) for the two predictors X_1 and X_2. The error terms (u and v's) are generated in a similar way to the univariate case above, with the correlation between v_1 and v_2 set to 0.5. The hypotheses of interest are H_0: $\beta_i = 0$ against H_1: $\beta_i > 0$; or H_{01}: $\beta_{11} = \beta_{12} = \beta_{21} = \beta_{22} = 0$ (no predictive power of X_1 and X_2). Again, the size values are calculated as the probability of rejecting the null hypothesis, setting $\beta_{11} = \beta_{12} = \beta_{21} = \beta_{22} = 0$. The level of significance for the test is set to 5%. Following Amihud et al. (2010), the numbers of Monte Carlo trials are set to 10,000 and 2000, respectively for the univariate and bivariate cases.

One of the key assumptions of the (improved) ARM is the normality of error terms. To evaluate its robustness to the departure from normality, I consider e_t generated from a GARCH(1,1) model and v_t from the Student-t distribution with 5 degrees of freedom. For the GARCH(1,1), the AR parameter is set to 0.8 and the MA parameter to 0.1. Both distributions are adjusted so that their unconditional standard deviations are identical to those under the normal error terms specified above. These error distributions simulate the shocks to the predictive model with conditional heteroskedasticity and fat-tailed behaviours.

4.2. Parameter estimation

Table 1 reports the Monte Carlo results for the estimation of predictive coefficients when the bivariate predictors are considered (Case I). Although not reported in detail, the results from the univariate case are found to be similar. The bias and standard deviation values of the predictive coefficient estimators based on the LS method replicate those reported in Amihud et al. (2010: Table 3) with reasonable accuracy. The values associated with the ARM are reproduced from Amihud et al. (2010). The first panel of Table 1 shows the bias and standard deviations under normal error terms, while the second reports those under

[3] The two-tailed alternative hypothesis is also considered. But the results are not reported for simplicity as they are qualitatively similar to the case of one-tailed alternative.

fat-tailed distributions (combination of GARCH(1,1) and Student-t). As expected, the bias and standard deviation values from LS estimation are substantially larger than those of the ARM and improved ARM. Comparing the ARM and improved ARM, the biases of the latter are substantially smaller than those of the former. For example, when the sample size is 50, the mean absolute bias of the ARM is 1.28, while that of the improved ARM is 0.63. This indicates that the improved ARM can deliver substantial bias reduction in small samples, although the reduction becomes negligible as the sample size increases. The second panel of Table 1 presents the case under fat-tailed error distributions: the results are qualitatively similar with those under normal errors: the improved ARM shows substantial efficiency gain in parameter estimation in comparison with the LS method.

4.3. Size and power properties

Table 2 reports the size values of the *t*- and *F*-tests when the error terms are generated from normal distributions. For nearly all cases, the *t*-test and *F*-test from the LS method show size values higher than 5%, over-rejecting the true null hypothesis in small samples. These size values replicate those of LS estimation reported in Amihud et al. (2010; Tables 1 and 3) with reasonable accuracy. The size values from the ARM are reproduced from Amihud et al. (2010). It is found that the improved ARM provides size values closer to 5% than the ARM does when the sample size is small, especially for the *F*-test. When the sample size is 50, for the univariate predictor under Case I, the *F*-test based on ARM provides the size value of 9.9%, while the improved ARM provides the size value of 5.0%. For the bivariate predictors under Case I when the sample size is 50, the *F*-test based on the improved ARM has size value of 5.3%, while that of the ARM gives 11.5%. These results show that the improved ARM controls the size of the test substantially better than the ARM of Amihud et al. (2010) in small samples, especially when the predictor is highly persistent. Although not reported in detail, the size values under the fat-tailed error terms are found to be similar to those under normal distributions.

In the bivariate model, it is also of interest to test for $H_{01,1}$: $\beta_{11} = \beta_{12} = 0$ and/or $H_{02,1}$: $\beta_{11} + \beta_{12} = 0$. Both hypotheses imply no predictive power of X_1 for Y in the bivariate predictive regression, when the partial effects of X_2 are netted out. However, they are different in that H_{01} indicates that individual lagged multipliers are jointly equal to zero, while H_{02} indicates that the total effects of X_1 on Y are zero. These tests are useful when one is primarily interested in the marginal predictive power of X_1 for Y, but

Table 2
Probability of rejection of the true null hypothesis (level of significance = 5%).

	LS	ARM	Improved ARM	LS	ARM	Improved ARM
	Case I			Case II		
Univariate predictor						
n = 50						
Right-tailed						
$\beta_1 = 0$	11.1	8.2	8.1	9.9	6.8	6.8
$\beta_2 = 0$	5.9	4.9	4.4	6.2	5.1	4.5
Joint						
$\beta_1 = \beta_2 = 0$	11.0	9.9	5.0	8.0	8.1	5.3
n = 200						
Right-tailed						
$\beta_1 = 0$	7.1	5.2	5.5	6.7	5.1	5.2
$\beta_2 = 0$	5.6	4.5	4.6	5.6	4.5	4.7
Joint						
$\beta_1 = \beta_2 = 0$	7.4	7.0	6.8	5.8	5.8	5.8
Bivariate predictor						
n = 50						
Right-tailed						
$\beta_{11} = 0$	10.9	8.8	8.3	9.6		7.1
$\beta_{12} = 0$	6.0	5.3	5.0	6.3		4.6
$\beta_{21} = 0$	9.6	7.0	7.8	8.3		6.5
$\beta_{22} = 0$	5.6	4.2	5.2	5.5		5.5
Joint						
$\beta_{11} = \beta_{12} = \beta_{21} = \beta_{22} = 0$	13.9	11.5	5.3	9.9		5.7
n = 200						
Right-tailed						
$\beta_{11} = 0$	6.6	5.6	4.7	6.4		4.3
$\beta_{12} = 0$	5.5	4.8	5.0	5.8		5.0
$\beta_{21} = 0$	6.7	5.6	5.7	6.5		5.5
$\beta_{22} = 0$	5.4	5.0	4.6	5.3		4.6
Joint						
$\beta_{11} = \beta_{12} = \beta_{21} = \beta_{22} = 0$	8.6	8.1	8.0	6.6		6.8

The entries are in percentages.
The columns labelled "ARM" contain the results based on the augmented regression method proposed by Amihud et al (2010). The figures for the ARM column are reproduced from Tables 1 and 3 of Amihud et al (2010). Amihud et al. (2010) did not report the size values for Case II of bivariate case, and this column is left blank.
The columns labelled "Improved ARM" contain the results based on the improved version of the augmented regression method proposed in this paper.
The column labelled "LS" contain the results based on the LS estimation.

a bivariate model is more appropriate for Y than a univariate one (see, for example, Ang and Bekaert, 2007). A Monte Carlo experiment is conducted to evaluate the size properties of the F-test for $H_{01,1}$ and $H_{02,1}$. Again, the improved ARM provides desirable size properties, while the LS method grossly over-rejects the true null hypotheses in small samples. For example, when the sample size is 50 under Case I, the size values for $H_{01,1}$ are found to be 13.8% and 5.1%, respectively from the LS and improved ARM; while those for $H_{02,1}$ are 18.0% and 4.0%. The full results are not reported in this paper for brevity.

Table 3 reports the (size-adjusted) power of the F-test based on the ARM and improved ARM for the case of univariate predictor, when the error terms are generated from normal distributions. The probability of rejecting H_{01}: $\beta_1 = \beta_2 = 0$ is calculated, setting $\beta_1 = \beta_2 \in \{0.1, 0.5, 1, 2, 4, 6, 8\}$. The power values of the t-tests are also calculated in a similar way, but not reported since the results were found to be similar. Both methods provide statistical tests whose power increases with the sample size. It is evident that the improved ARM test is more powerful than the ARM. Its power gain is higher when the sample size is smaller and when the predictor is more persistent. Similar power properties are evident when the model is subject to the error terms with fat-tailed distributions, although their details are not reported for brevity. An interesting feature from Table 3 is that, as the values of $\beta_1 = \beta_2$ approaches 0 from 8 to the direction of the null hypothesis, the power converges to the size value of 5%. It appears that the speed of convergence depends on the sample size and the degree of persistence of the model. This property is as expected from the underlying asymptotic theory, and is a reflection that the asymptotic theory provides a reliable guidance in small samples.

4.4. Forecasting performance

For investors and market participants seeking to exploit profit opportunities, out-of-sample forecast accuracy of predictive regression is an important input for their decision-making. To this end, I compare accuracy of out-of-sample forecasts generated from the ARM and improved ARM, relative to those of the LS method. For each Monte Carlo trial, I generate the data $\{(Y_t, X_t\}_{t=1}^{n+h}$, and use $\{(Y_t, X_t\}_{t=1}^{n}$ for model estimation and $\{(Y_t, X_t\}_{t=n+1}^{n+h}$ for evaluation of forecast accuracy. Using the simple model (1) as an example, the forecast for the predictor X_{n+h} is recursively generated as

$$X_n(h) = \tilde{\alpha}_0 + \tilde{\alpha}_1 X_n(h-1),$$

where $X_n(k) = X_{n+k}$ for $k \le 0$ and $\tilde{\alpha}$'s are the estimators for α's (based on the LS, ARM, or improved ARM). Then, the forecast for the future asset return Y_{n+h} is generated as

$$Y_n(h) = \tilde{\beta}_0 + \tilde{\beta}_1 X_n(h-1),$$

where $Y_n(k) = Y_{n+k}$ for $k \le 0$ and $\tilde{\beta}$'s are the estimators for β's (based on the LS, ARM, or improved ARM). As a measure of forecast accuracy, the root mean squared forecast error is used:

$$RMSE(h) = \sqrt{\sum_{k=1}^{h} \left(Y_{n+k} - Y_n(k)\right)^2 / h}.$$

Table 3
Size-adjusted power values of the F-test (level of significance = 5%).

	ARM	Improved ARM	ARM	Improved ARM
	n = 50		n = 200	
Case 1				
$\beta_1 = \beta_2$				
0.1	4.88	5.09	4.54	6.24
0.5	4.46	5.38	4.33	5.73
1	4.27	5.90	6.99	9.25
2	5.29	8.05	25.27	31.24
4	10.06	16.03	99.82	99.93
6	26.26	41.10	100.00	100.00
8	60.36	81.05	100.00	100.00
Case II				
$\beta_1 = \beta_2$				
0.1	4.84	5.24	4.73	5.51
0.5	4.46	5.15	4.67	5.33
1	4.28	5.24	5.82	6.85
2	4.58	6.18	15.03	16.72
4	7.83	11.00	69.85	72.71
6	16.19	21.24	99.99	99.99
8	32.97	42.01	100.00	100.00

The entries are in percentages and represent the probability of rejecting H_0: $\beta_1 = \beta_2 = 0$ when their true values are $\beta_1 = \beta_2 \in \{0.1, 0.5, 1, 2, 4, 6, 8\}$ in the predictive model with a univariate predictor. The above figures are from the model with normal error terms.

I calculate the ratio of $RMSE(h)$ of the ARM forecasts to that of the LS forecasts; and the ratio of $RMSE(h)$ of the improved ARM forecasts to that of the LS forecasts for $h = 1, \ldots, 6$.

Fig. 1 plots the mean of the *RMSE* ratios for the predictive model with univariate predictor when the sample size is 50. In Case I, the improved ARM forecasts are clearly more accurate than the LS forecasts, while the ARM forecasts are inferior to LS forecasts. For Case II, both ARM-based forecasts outperform the LS forecasts, but the improved ARM forecasts are more accurate than the ARM. Although not reported in detail, the improved ARM forecasts are still more accurate than the ARM forecasts when the sample is 200, although the difference is not substantial. In general, the gain of forecast accuracy diminishes as the forecast horizon increases. The poor performance of the ARM forecasts, observed in Case I when the sample size is 50, is likely to be caused by the bias-corrected parameter estimates which imply non-stationary.

The above result has strong implications to investors and market participants. It indicates that, given the signal from a predictor, the improved ARM has the best ability to predict the future changes of an asset price. This means that, given the level of risk that they are facing, investors are best placed to formulate profit-maximising trading rules when they use the improved ARM for forecasting. The question as to how large the potential profit gains would be is an important one, which could be addressed by evaluating profitability of trading rules based on alternative forecasting methods. However, I leave this line of investigation as a possible future research.

We have found the evidence that the improved ARM performs better than the ARM in parameter estimation, statistical inference, and forecasting, especially when the sample size is small and the predictors are highly persistent. The estimation bias is particularly large in the latter case. The improved ARM, equipped with the bias-correction with higher precision and stationarity-correction, is more effective for bias reduction, which in turn improves the properties of statistical test and multi-step forecasting. For example, the improved ARM forecasts are more accurate since they are generated with parameter estimators converging to the true value at a faster rate. Intuitively, forecasts generated from more precise parameter estimators should be more accurate, since they are generated as a linear combination of the parameter estimators and the last p observations of the predictor. This is consistent with the findings that the use of bias-corrected estimator generates more accurate (multi-step) out-of-sample forecasts for a time series model (see, for example, Kim, 2003). In addition, the ARM often generates the forecasts which imply non-stationarity of asset return, which is another intuitive reason as to why the improved ARM forecasts are superior to the ARM forecasts.

5. Empirical application

Whether stock return can be predicted by dividend yield has been a subject of much interest in empirical finance since the seminal work of Fama and French (1988). In this section, I employ the improved ARM to evaluate whether dividend yield shows predictive ability for stock return, using the US data analysed in Lewellen (2004). The data includes the stock return (value-weighted) and dividend yield from the CRSP database, monthly from 1946 to 2009. Further data details are available in Lewellen (2004). I also consider the excess return and (excess) returns from the equal-weighted index, but the results are found to be qualitatively similar. Guided by Lewellen (2004; Section 4.2) who has observed unusual decline of the dividend yield from 1995, I conduct sub-sample analyses with the breaking point of 1995. Based on his conditional test on predictive coefficient, Lewellen (2004) finds highly significant predictive power of dividend yield during the period of 1946–2000, despite unusual price run-up after 1995. In contrast, applying their ARM to quarterly data from 1946 to 1994, Amihud et al. (2010) find no significant predictive power of dividend yield under the predictive model of lag order $p = 2$.

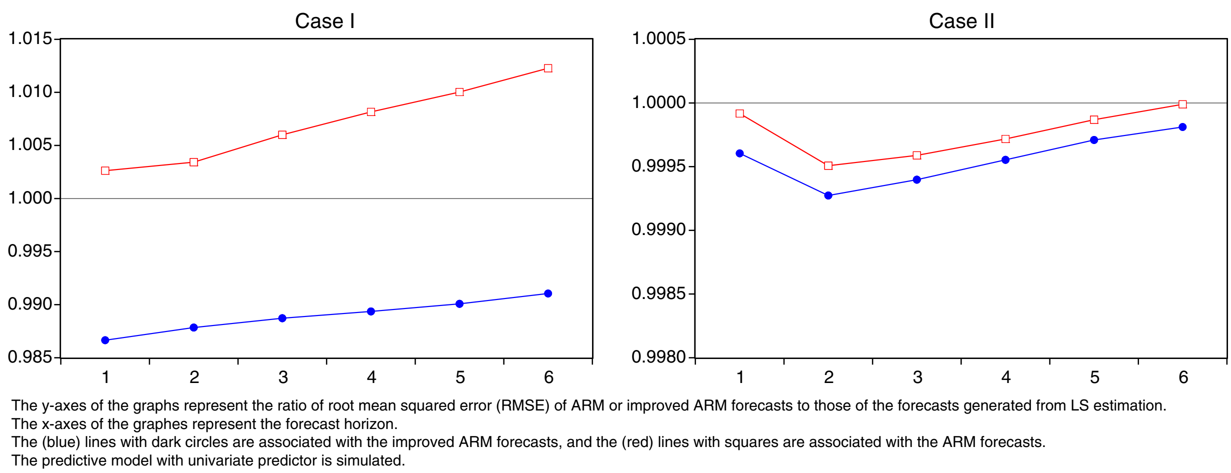

Fig. 1. Forecasting Performance of the ARM and Improved ARM (n = 50). The y-axes of the graphs represent the ratio of root mean squared error (RMSE) of ARM or improved ARM forecasts to those of the forecasts generated from LS estimation. The x-axes of the graphs represent the forecast horizon. The (blue) lines with dark circles are associated with the improved ARM forecasts, and the (red) lines with squares are associated with the ARM forecasts. The predictive model with univariate predictor is simulated.

5.1. Predictive regression with univariate predictor

Let Y be the value-weighted return and X is the natural log of dividend yield. Following Lewellen (2004), I first consider the AR(1) predictor case. The estimation and test results are reported in the first panel of Table 4. For the period from 1946 to 2000, the LS method provides β_1 estimate of 0.92 which is significant at 5% level of significance against one-tailed alternative; i.e. H_0: $\beta_1 = 0$ is rejected in favour of H_1: $\beta_1 > 0$. As mentioned earlier, $\hat{\alpha}_1^c = 1.003$ and $\hat{\alpha}_1^a = 0.9997$, with the corresponding β_1 estimate 0.68 which is statistically no different from 0 as H_0: $\beta_1 = 0$ is not rejected. The Lewellen's (2004) method, conditional on $\alpha_1 =$ 0.9999, provides β_1 estimate of 0.66 which is highly statistically significance; whereas the Stambaugh's (1999) method provides β_1 estimate of 0.20 with no statistical significance. In this case, the improved ARM provides the predictive coefficient estimate similar to that of the Lewellen's (2004) method, while the outcome of the t-test on β_1 is consistent with Stambaugh's (1999). The value of ϕ estimate is negative and significant statistically, indicating strong negative correlation between the error terms of stock return and dividend yield equations.

Following Lewellen (2004), the ARM is also applied to the data from 1946 to 1994. The LS provides β_1 estimate of 2.23 statistically significant at 1% level of significance, and the improved ARM gives β_1 estimate of 1.62, which is also statistically significant. The Lewellen's (2004) method provides β_1 estimate of 0.98, which is statistically significant; while Stambaugh's (1999) method yields β_1 estimate of 1.53 but it is statistically insignificant. In this case, the improved ARM provides predictive coefficient estimate similar to Stambaugh's (1999), while its t-test outcome is in agreement with Lewellen's (2004). For the entire

Table 4
Predictive regression for the US stock return with dividend yield as predictor.

AR(1) predictor

	β_1	α_1	ϕ
1946–2000			
LS	0.92*	0.9971	
Improved ARM	0.68	0.9998	−90.44**
Stambaugh	0.20		
Lewellen	0.66**	0.9999	
1946–1994			
LS	2.23**	0.9860	
Improved ARM	1.62**	0.9928	−90.05**
Stambaugh	1.53		
Lewellen	0.98**	0.9999	
1946–2009			
LS	1.07**	0.9936	
Improved ARM	0.60	0.9988	−90.42**
1995–2009			
LS	1.90	0.9714	−91.60**
Improved ARM	−0.13	0.9936	

The results for Stambaugh and Lewellen are reproduced from Lewellen (2004; Table 4).
The signs * and ** indicate 5% and 1% level of significance, against one-tailed alternative

AR(4) predictor

	β	α	ϕ	H_{01}	H_{02}
1946–2000					
LS	0.92	0.9967		1.25	3.62
Improved ARM	0.65	0.9997	−90.55**	0.79	1.78
1946–1994					
LS	2.28	0.9850		3.24**	11.32**
Improved ARM	1.70	0.9914	−90.18**	1.96	6.23**
1946–2009					
LS	1.12	0.9929		3.23**	7.77**
Improved ARM	0.70	0.9975	−90.57**	0.97	2.98
1995–2009					
LS	2.18	0.9678		2.38	2.01
Improved ARM	0.66	0.9839	−91.89**	1.87	0.18

The symbols * and ** indicate 5% and 1% level of significance.
$\beta = \beta_1 + \beta_2 + \beta_3 + \beta_4$ is the sum of predictive coefficients.
$\alpha = \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4$ is the sum of AR coefficients for the predictor.
H_{01}: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$; F-statistic is reported;
H_{02}: $\beta_1 + \beta_2 + \beta_3 + \beta_4 = 0$; F-statistic is reported.

period of 1994 to 2009, the LS estimate of predictive coefficient is 1.07 and is statistically significant, while the improved ARM provides the estimate of 0.60 that is statistically insignificant. For the data from 1995 to 2009, the predictive coefficient is found to be statistically insignificant, regardless of the estimation methods.

The second panel of Table 4 reports the case where the lag order of 4 is assumed for the predictive regression and dividend yield equation. Statistically, this choice is justified by AIC which selects the order of 4. Economically, stock returns may depend on dividend yield with a lag of up to 4 months, since dividends are usually paid out quarterly. This case also serves as a robustness check, since it is possible that AR(1) predictor model may be subjected to model specification bias, if the true lag order was higher. Stambaugh's (1999) and Lewellen's (2004) methods are not considered for this case, since they are proposed for the AR(1) predictor. Amihud's (2010) ARM is in principle applicable to the AR(4) structure, but derivation of the covariance matrix of bias-corrected predictive coefficient estimators is highly demanding, as mentioned earlier. However, it is well expected that the original ARM provides similar estimation and inferential results to those of the improved ARM, as long as it is operational with the bias-corrected coefficient estimates satisfying the assumption of stationarity.

The model with lag order $p = 4$ can be written as

$$\begin{aligned} Y_t &= \beta_0 + \beta_1 X_{1t-1} + \ldots + \beta_4 X_{1t-4} + u_t; \\ X_{1t} &= \alpha_1 + \alpha_{11} X_{1t-1} + \ldots + \alpha_{14} X_{1t-4} + v_t. \end{aligned}$$

For simplicity, the sum of predictive coefficients ($\beta = \beta_1 + \beta_2 + \beta_3 + \beta_4$) is reported, which represents the total effect of dividend yield on stock return. The sum of AR coefficients (α) is also reported, which represents the degree of persistence of dividend yield. The estimates for β and α are qualitatively no different from those obtained from the AR(1) predictor case, for all periods. From 1946 to 2000, both LS and improved ARM provide evidence of no return predictability from the dividend yield. The F-tests for both H_{01}: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ and H_{02}: $\beta = 0$ are not rejected based on both LS and the improved ARM, at 5% level of significance. From 1946 to 1994, the *F*-test based on the LS rejects both H_{01} and H_{02}, while the improved ARM rejects only H_{02} in support of predictability. For period of 1946–2009, the LS and improved ARM show different results: the *F*-test based on the LS rejects both H_{01} and H_{02}, but the improved ARM rejects neither. For the data from 1995 to 2009, both H_{01} and H_{02} are not rejected, regardless of the estimation methods.

Keeping in mind the Monte Carlo results that the *F*-test based on the LS seriously over-rejects the true null hypothesis while the improved ARM performs desirably, my henceforth assessment on the empirical result is based on the inferential outcomes of the latter. The above results (based on the improved ARM) suggest that the dividend yield shows predictive power for stock return only for the pre-1994 period. Its predictive ability for stock return becomes statistically insignificant if the data after 1995 are included. It is also noteworthy that the LS estimates of predictive coefficients are substantially higher than the improved ARM estimates, consistent with the well-known property that the LS estimators are biased upward.

5.2. Predictive regression with bivariate predictors

Ang and Bekaert (2007) find that the predictive ability of the dividend yield for excess stock return is considerably enhanced in a bivariate regression with short-term interest rate. This finding provides a good case for a bivariate predictive model for excess stock return using the dividend yield and short-term interest rate. In keeping with Ang and Bekaert (2007), let *Y* be the excess stock return calculated as value-weighted return minus short-term interest rate (T-bill rate); and let X_1 and X_2 respectively be the natural log of dividend yield and the short-term interest rate. I first consider AR(1) predictor case, which can be written as

$$\begin{aligned} Y_t &= \beta_0 + \beta_{11} X_{1t-1} + \beta_{21} X_{2t-1} + u_t; \\ X_{1t} &= \alpha_1 + \alpha_{11} X_{1t-1} + v_{1t}; X_{2t} = \alpha_2 + \alpha_{21} X_{2t-1} + v_{2t}. \end{aligned}$$

The estimation and test results are reported in the first panel of Table 5. Overall, the predictive coefficient estimates of dividend yield (β_{11}) and their statistical significance are qualitatively no different from the univariate case reported in Table 4. The predictive coefficient estimates attached to short-term interest rates (β_{21}) are negative and statistically significant, except for the period of 1995–2009. The AR(1) coefficient estimates indicate that both dividend yield and short-term interest rates are highly persistent. As expected, the error term of the dividend yield equation is found to be highly (negatively) correlated to that of the excess stock return with highly significant ϕ_1 value; but the error term of the interest rate equation does not show such strong correlation as the value of ϕ_2 is not statistically different from zero. According to the improved ARM, the dividend yield has predictive power only for the period of 1946–1994, as it is the only period in which β_{11} is statically significant. The short-term interest rate is found to show significant predictive power for excess stock return, except for the period of 1995–2009, regardless of the estimation methods.

As a robustness check, I again consider the AR(4) predictor case, which is written as

$$\begin{aligned} Y_t &= \beta_0 + \beta_{11} X_{1t-1} + \ldots + \beta_{14} X_{1t-4} + \beta_{21} X_{2t-1} + \ldots + \beta_{24} X_{2t-4} + u_t; \\ X_{1t} &= \alpha_1 + \alpha_{11} X_{1t-1} + \ldots + \alpha_{14} X_{1t-4} + v_{1t}; \\ X_{2t} &= \alpha_2 + \alpha_{21} X_{2t-1} + \ldots + \alpha_{24} X_{2t-4} + v_{2t}. \end{aligned}$$

Table 5
Predictive regression for the US stock return with dividend yield and short-term interest rate as bivariate predictors.

AR(1) predictors

	β_{11}	β_{21}	α_{11}	α_{21}	ϕ_1	ϕ_2
1946–2000						
LS	0.93*	−1.87**	0.9971	0.9769		
Improved ARM	0.68	−0.97**	0.9998	0.9829	−90.45**	−0.75
1946–1994						
LS	2.28**	−1.98**	0.9860	0.9774		
Improved ARM	1.62**	−1.00**	0.9928	0.9842	−90.04**	−0.81
1946–2009						
LS	1.13**	−1.70**	0.9936	0.9804		
Improved ARM	0.59	−0.85**	0.9988	0.9855	−90.46**	−0.71
1995–2009						
LS	3.09*	3.21	0.9714	0.9904		
Improved ARM	0.19	0.07	0.9936	0.9991	−91.03**	1.34

The symbols * and ** indicate 5% and 1% level of significance, against one-tailed alternative.

AR(4) predictors

	β_1	β_2	$H_{01,1}$	$H_{01,2}$	$H_{02,1}$	$H_{02,2}$
1946–2000						
LS	0.85	−1.52	0.91	4.94**	3.10	5.18**
Improved ARM	0.66	−1.00	0.83	6.53**	1.81	25.56**
1946–1994						
LS	2.18	−1.63	2.79*	4.72**	10.40**	5.94**
Improved ARM	1.72	−1.03	2.02	6.60**	6.32**	25.83**
1946–2009						
LS	1.15	−1.40	3.24**	4.16**	8.09**	4.76**
Improved ARM	0.69	−0.88	2.00	5.78**	2.89	21.85**
1995–2009						
LS	3.23	2.25	2.56*	0.93	3.65	0.86
Improved ARM	0.98	−0.12	1.74	1.94	0.87	0.04

$\beta_1 = \beta_{11} + \beta_{12} + \beta_{13} + \beta_{14}$ is the sum of predictive coefficients associated with dividend yield.
$\beta_2 = \beta_{21} + \beta_{22} + \beta_{23} + \beta_{24}$ is the sum of predictive coefficients associated with short-term interest rate.
$H_{01,1}$: $\beta_{11} = \beta_{12} = \beta_{13} = \beta_{14} = 0$; F-statistic is reported.
$H_{01,2}$: $\beta_{21} = \beta_{22} = \beta_{23} = \beta_{24} = 0$; F-statistic is reported.
$H_{02,1}$: $\beta_{11} + \beta_{12} + \beta_{13} + \beta_{14} = 0$; F-statistic is reported.
$H_{02,2}$: $\beta_{21} + \beta_{22} + \beta_{23} + \beta_{124} = 0$; F-statistic is reported.
The symbols * and ** indicate 5% and 1% level of significance.

For all periods considered, the sum of predictive coefficients attached to dividend yield ($\beta_1 = \beta_{11} + \beta_{12} + \beta_{13} + \beta_{14}$) is found to be close to the β_{11} estimate in the AR(1) case. Similarly, the sum of predictive coefficients attached to interest rate ($\beta_2 = \beta_{21} + \beta_{22} + \beta_{23} + \beta_{24}$) is found to be close to β_{21} estimate in the AR(1) case, for all periods. When the improved ARM is used, the *F*-test for $H_{01,1}$: $\beta_{11} = \beta_{12} = \beta_{13} = \beta_{14} = 0$ is not rejected for all periods; while $H_{02,1}$: $\beta_{11} + \beta_{12} + \beta_{13} + \beta_{14} = 0$ is rejected only for 1946–1994, at 5% level of significance. This suggests that the dividend yield has predictive power before 1994 only. The *F*-test based on the LS method rejects either $H_{01,1}$ or $H_{02,1}$ more frequently, possible reflection of its property of over-rejecting the true null hypothesis. For short-term interest rate, both $H_{01,2}$: $\beta_{21} = \beta_{22} = \beta_{23} = \beta_{24} = 0$ and $H_{02,2}$: $\beta_{21} + \beta_{22} + \beta_{23} + \beta_{24} = 0$ are soundly rejected, except for the period of 1995–2009, regardless of the estimation methods. This indicates strong predictive power of short-term interest rate for stock return, but this result does not hold for the period from 1995 to 2009.

From the univariate and bivariate predictive regressions, the improved ARM results indicate that the dividend yield has statistically significant predictive power for stock return when the data from 1946 to 1994 is used. However, its predictive power is no longer statistically significant if the data from 1995 is included. The analysis based on the data from 1995 to 2009 shows no predictive power of dividend yield. This finding is different from the results of the conditional test of Lewellen (2004), who find statistically significant predictive power of dividend yield from 1994 to 2000. Applying their ARM to quarterly data from 1946 to 1994, Amihud et al. (2010) report significant predictive power of dividend yield when $p = 1$, but otherwise when $p = 2$. Based on monthly data from 1946 to 1994, the improved ARM finds significant predictive power of the dividend yield when $p = 1$. When $p = 4$, similarly to Amihud et al. (2010), $H_{01,1}$: $\beta_{11} = \beta_{12} = \beta_{13} = \beta_{14} = 0$ is not rejected; however, the improved ARM finds the total effect of dividend yield to be statistically significant, since $H_{02,1}$: $\beta_{11} + \beta_{12} + \beta_{13} + \beta_{14} = 0$ is soundly rejected.[4] As for the short-term interest, while it has been a strong predictor for stock return from 1946, it shows little predictive power from 1995 to 2009.

[4] Amihud et al. (2010) did not test for this hypothesis on the total effect.

6. Concluding remarks

Predictive regression is widely used in empirical finance to test for the predictability of asset return by a range of predictors which represent the fundamentals. It is well-known that statistical inference on predictive coefficient is fraught with small sample deficiencies, which are in large part caused by the biases of parameter estimation (see, for example, Stambaugh, 1999). In response to this, Amihud et al. (2010), extending the earlier works by Amihud and Hurvich (2004) and Amihud et al. (2008), propose the augmented regression method for bias-corrected estimation and statistical inference when the predictors follow an AR(p) structure. Although their methods are found to be effective in reducing small sample biases in parameter estimation, the statistical test on predictive coefficients still shows a degree of size distortion in small samples when the predictor is highly persistent. In addition, the usefulness of their method for a higher order case ($p \geq 3$) is somewhat limited, because analytical formulae for bias and covariance matrix estimators are not clearly given, although they can be in principle obtained through tedious algebraic derivations. Moreover, their method runs into a problem when bias-correction renders the predictive model non-stationary.

This paper proposes three modifications to the augmented regression method of Amihud et al. (2010). They include (i) an improved bias-correction method which provides predictive coefficient estimators unbiased to the order of n^{-1}; (ii) stationarity-correction for bias-correction to ensure the stationarity of predictive model; and (iii) the use of matrix formulae for bias-correction and covariance matrix estimation. From an extensive Monte Carlo experiment, it is found that the improved augmented regression method proposed delivers substantial gain in small samples, especially when the predictor is highly persistent. In comparison with the original method of Amihud et al. (2010), the improved method delivers additional reduction in estimation bias. It also provides the statistical test on predictive coefficients with better size and power properties. Moreover, the improved method generates out-of-sample (multi-step) forecasts more accurate than the original method. The improved augmented regression method proposed in this paper is applied to a set of monthly US stock market data examined in Lewellen (2004) to assess the predictive power of dividend yield for stock return.

The predictive model considered in this study assumes that the predictability is constant over time. However, in financial markets, the degree of predictability may change due to time-varying predictability or structural breaks. In these circumstances, the model examined in this study is misspecified and may perform poorly. The bias-reduced estimation and statistical inference under the departure from the constant predictability are an issue of importance, given the dynamic nature of financial markets. However, since this line of research is beyond the scope of the current study, it is left as a future research subject.

References

Amihud, Y., Hurvich, C.M., 2004. Predictive regression: a reduced-bias estimation method. J. Financ. Quant. Anal. 39, 813–841.
Amihud, Y., Hurvich, C.M., Wang, Y., 2008. Multiple-predictor regressions: hypothesis testing. Rev. Financ. Stud. 22, 413–434.
Amihud, Y., Hurvich, C.M., Wang, Y., 2010. Predictive regression with order-p autoregressive predictors. J. Empir. Financ. 17, 513–525.
Ang, A., Bekaert, G., 2007. Stock return predictability: is it there? Rev. Financ. Stud. 20, 651–707.
Chiquoine, B., Hjalmarsson, E., 2009. Jackknifing stock return predictions. J. Empir. Financ. 16, 793–803.
Cochrane, J.H., 2008. The dog that did not bark: a defence of return predictability. Rev. Financ. Stud. 21, 1533–1575.
Fama, E.F., French, K.R., 1988. Dividend yields and expected stock returns. J. Financ. Econ. 22 (1), 3–25.
Kilian, L., 1998. Small sample confidence intervals for impulse response functions. Rev. Econ. Stat. 80, 218–230.
Kim, J.H., 2003. Forecasting autoregressive time series with biascorrected parameter estimators. Int. J. Forecast. 19, 493–502.
Kim, J.H., 2004. Bootstrap prediction intervals for autoregression using asymptotically mean-unbiased parameter estimators. Int. J. Forecast. 20, 85–97.
Lettau, M., Ludvigson, S., 2001. Consumption, aggregate wealth, and expected stock returns. J. Financ. 56 (3), 815–849.
Lettau, M., Van Nieuwerburgh, S., 2008. Reconciling the return predictability evidence. Rev. Financ. Stud. 21 (4), 1607–1652.
Lewellen, J., 2004. Predicting returns with financial ratios. J. Financ. Econ. 74, 209–235.
Malkiel, B.G., 2003. The efficient market hypothesis and its critics. J. Econ. Perspect. 17 (1), 59–82.
Nelson, C.R., Kim, M.J., 1993. Predictable stock returns: the role of small sample bias. J. Financ. 48, 641–661.
Nicholls, D.F., Pope, A.L., 1988. Bias in the estimation of multivariate autoregressions. Aust. J. Stat. 30A, 296–309.
Patterson, K.D., 2000. Bias reduction in autoregressive models. Econ. Lett. 68, 15–141.
Paye, B., Timmermann, A., 2006. Instability of return prediction models. J. Empir. Financ. 13 (3), 274–315.
R Core Team, 2012. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria3-900051-07-0 (http://www.R-project.org/).
Shaman, P., Stine, R.A., 1988. The bias of autoregressive coefficient estimators. J. Am. Stat. Assoc. 83, 842–848.
Stambaugh, R.F., 1999. Predictive regressions. J. Financ. Econ. 54, 375–421.
Stine, R.A., Shaman, P., 1989. A fixed point characterization for bias of autoregressive estimators. Ann. Stat. 17, 1275–1284.
Welch, I., Goyal, A., 2007. A comprehensive look at the empirical performance of equity premium prediction. Rev. Financ. Stud. 21 (4), 1455–1508.