Multiple-Predictor Regressions: Hypothesis Testing

Yakov Amihud
New York University

Clifford M. Hurvich
New York University

Yi Wang
New York University

We propose a new hypothesis-testing method for multipredictor regressions in small samples, where the dependent variable is regressed on lagged variables that are autoregressive. The new test is based on the augmented regression method (Amihud and Hurvich, 2004), which produces reduced-bias coefficients and is easy to implement. The method's usefulness is demonstrated by simulations and by testing a model where stock returns are predicted by two variables, income-to-consumption and dividend yield. (*JEL* C32, G12)

1. Introduction

In a class of predictive regressions analyzed by Stambaugh (1999), a variable is regressed on the lagged value of a predictor variable, which is autoregressive with errors that are correlated with the errors of the regression model. In small samples, the ordinary least-squares (OLS) estimated predictive slope coefficient is biased, potentially leading to an incorrect conclusion that the lagged variable has predictive power while in fact it does not. Stambaugh (1999) derived the bias expression for the OLS estimate, which was subsequently used in empirical studies to obtain a reduced-bias point estimate of the predictive coefficient. Subsequent papers developed methods for hypothesis testing of single-predictor models.[1]

In this paper, we propose a new method for hypothesis testing in predictive regressions for *multiple*-predictor models. It is based on the multi-predictor augmented regression method (mARM, a multipredictor extension of

---

The authors thank James Stock, the Editor and the Referee for helpful comments and suggestions. Send correspondence to Ira Leon Rennert, Professor of Finance, Stern School of Business, New York University, New York, NY 10012; telephone: 212-998-0720; fax: 212-995-4220; E-mail: yamihud@stern.nyu.edu.

[1] See Kothari and Shanken (1997); Amihud and Hurvich (2004); Lewellen (2004); Campbell and Yogo (2005); Polk et al. (2005).

doi:10.1093/rfs/hhn056 Advance Access publication June 13, 2008

the single-predictor ARM methodology considered in Amihud and Hurvich, 2004). An attractive feature of the ARM-based hypothesis-testing method is that it also generates parameter estimates—the predictor variable's estimated coefficients—thus providing a unified multipredictor regression platform for estimation, testing, and forecasting.

In what follows, we first present the theory underlying mARM-based hypothesis testing, and then perform simulations of two-predictor models with three different autoregressive matrices (one based on estimates from a relevant data set), to compare the actual size obtained under our test with the nominal size. We find that the actual test sizes are quite close to nominal sizes. We compare our mARM-based results with two other methods: The OLS and bootstrapping, the latter being used in some recent papers that use multipredictor models. We find that our mARM-based hypothesis test performs better than either of the alternatives.

Finally, we apply mARM to test the two-predictor model of Santos and Veronesi (2006), where stock returns are predicted by labor income to consumption and by dividend yield. Santos and Veronesi employ multiple-horizon OLS regressions while we employ single-horizon predictive regressions with bias correction, using mARM, following the proposition of Boudoukh et al. (2006) on the relationship between short and long horizon predictions. We find that the evidence for predictability of these economic ratios is weaker than that reported by Santos and Veronesi (2006).

Stock and Watson (1993) propose regressions that are augmented by the errors $\{v_t\}$ of the regressors $\{x_t\}$ in a dynamic OLS method (DOLS) with a single predictor (see also Hamilton, 1994). Their model assumes that the autoregressive process is $I(1)$, meaning—in our context—that the autoregressive parameter of the predictor series is known to be 1. In our model, the autoregressive coefficient matrix of $\{x_t\}$, Φ, is not known and the estimation problem arises from the fact that its OLS estimates are biased in small samples, which in turn creates a bias in the estimated OLS coefficients of the predictive model. This problem does not exist in DOLS.

The mARM methodology does assume that the $\{x_t\}$ process is stationary. This assumption is reasonable for a variety of applications of predictive regressions in Finance. For example, Santos and Veronesi (2006) argued that "the restriction that [the log income/consumption ratio] is stationary rests on solid economic intuition: it is not reasonable to assume that consumption can grow to be infinitely larger than labor income, or, alternatively, that labor income can grow to be several times higher than consumption." For a given sample size, the performance of the mARM method deteriorates as the autoregressive parameter matrix approaches the nonstationarity boundary. However, our simulations indicate that this deterioration is far less for the mARM method than it is for the existing methods for multipredictor hypothesis testing, namely bootstrapping and OLS. Since the mARM test may be somewhat oversized, a predictor found to be statistically significant by mARM-based hypothesis testing should ideally be subjected to further scrutiny before being declared to have predictive power. On the other hand, since an mARM-based p-value may be somewhat understated, a finding based on the mARM test that a predictor is insignificant can be considered trustworthy. (This occurs in several cases in the application we study here.)

Empirical research in finance and economics often deals with *multipredictor*, small-sample models where the predictor variables are autoregressive and may thus be subject to the estimation problem pointed out by Stambaugh (1999). For example, Keim and Stambaugh (1986); Fama and French (1988); Pontiff and Schall (1998); and Keim and Madhavan (2000) estimate models where the excess stock return is regressed on the lagged values variables including dividend yield, book-to-market ratio, default yield premium, term yield premium, and short interest rate, all being autoregressive and some are correlated with others. Also, Keim and Madhavan (2000) estimate the number of transactions in the exchange seat market as a function of the lagged seat prices, lagged change in the seat's quoted bid-ask spread, and lagged absolute seat return, variables that are autocorrelated. Campbell et al. (2001) estimate the prediction of GDP growth by three measures of stock market volatility and find that they are jointly significant. In addition to these predictor variables being autoregressive, they are also contemporaneously negatively correlated with GDP growth, characteristics that are related to those in Stambaugh's (1999) model.

Some multipredictor studies use bootstrapping estimation with the predictors' coefficients set to zero (as under the null). Examples include Lettau and Ludvigson (2001) who predict stock returns by consumption in excess of labor income and asset wealth (CAY), dividend yield, dividend payment, bond yield spreads (term and default), and real interest rate. Baker and Stein (2004) estimate the equity share of firms financing as a function of stock turnover and dividend yield, both being serially correlated, as well as lagged stock returns. Baker and Wurgler (2004) estimate a multipredictor model of dividend initiation as a function of three lagged variables, which are autoregressive: the dividend premium (the difference between the market-to-book ratios for dividend payers and nonpayers), the Citizen Utility dividend premium (the ratio of stock prices of the dividend paying and nonpaying class of this company's stock), and the abnormal stock return at dividend announcements. Another model uses as predictors market-to-book ratio, dividend yield, and tax rates. Other studies run diagnostics to check for potential bias and conclude that it is insignificant. For example, Baker, Greenwood, and Wurgler (2003) estimate a two-predictor model of excess bond return as a function of the lagged values of new issues of short- and long-term debt, both of which are autocorrelated. Guo and Savickas (2006) predict stock returns by both market and idiosyncratic volatility series, which are autoregressive and their errors are correlated with those of the predictive model. Ang and Bekaert (2005) show, using bootstrapping estimation with zero coefficients of their model where stock returns are predicted by lagged dividend yield and the short rate, that their estimated coefficients are indeed biased but they conclude that the bias strengthens their empirical evidence. Goyenko's (2005) prediction of stock and bond returns by illiquidity of both stocks and bonds, which are autocorrelated and related, corrects for the estimation bias using the Amihud-Hurvich (2004) method under the assumption of diagonal covariance matrix between the predictors.

Our paper proceeds as follows. Section 2 describes our proposed method of hypothesis testing in the multipredictor case, based on a newly proposed estimator of the covariance matrix of the estimated slope coefficients. We also present results of simulations that compare the performance of our proposed method with that of methods previously proposed in the literature. In Section 3, we apply mARM to data from Santos and Veronesi (2006),[2] who examine the prediction of stock returns by ratio of labor income to consumption and by dividend yield. Section 4 presents our conclusions. Proofs of the theoretical results and some technical details are presented in the Appendix.

2. Estimation and Testing for Multipredictor Regression

Assume a predictive model where the predictor variables constitute a p-dimensional vector time series $\{x_t\}$, which follows a stationary Gaussian vector autoregressive VAR(1) model. The overall model is given for $t = 1, \ldots, n$ by

$$\begin{aligned} y_t &= \alpha + \beta' x_{t-1} + u_t, \\ x_t &= \Theta + \Phi x_{t-1} + v_t, \end{aligned}$$

where we define the $(p \times 1)$ vectors,

$$x_t = \begin{pmatrix} x_{1t} \\ \vdots \\ x_{pt} \end{pmatrix}, \qquad \Theta = \begin{pmatrix} \Theta_1 \\ \vdots \\ \Theta_p \end{pmatrix}, \qquad \beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}, \qquad v_t = \begin{pmatrix} v_{1t} \\ \vdots \\ v_{pt} \end{pmatrix},$$

and a $(p \times p)$ matrix,[3]

$$\Phi = \begin{pmatrix} \Phi_{11} & \cdots & \Phi_{1p} \\ \vdots & \ddots & \vdots \\ \Phi_{p1} & \cdots & \Phi_{pp} \end{pmatrix}.$$

The quantities y_t, α, and u_t are scalars. The vectors $(u_t, v_t')'$ are i.i.d. multivariate normal with mean zero. We allow u_t and v_t' to be contemporaneously correlated and assume that the absolute values of the eigenvalues of Φ are all less than 1 to ensure stationarity of $\{x_t\}$.

---

[2] We thank Pietro Veronesi for kindly providing us with the data.

[3] For a diagonal Φ, Amihud and Hurvich (2004) propose a simpler hypothesis testing method, which is similar to the single-predictor testing method.

As shown in Amihud and Hurvich (2004), there exists a $(p \times 1)$ vector ϕ and a sequence $\{e_t\}_{t=1}^{n}$ such that

$$u_t = \phi' v_t + e_t$$

and we can write

$$y_t = \alpha + \beta' x_{t-1} + \phi' v_t + e_t, \tag{1}$$

where $\{e_t\}$ are i.i.d. normal random variables with mean zero, independent of both $\{v_t\}$ and $\{x_t\}$, with $\mathrm{var}(u_t) = \sigma_u^2$, $\mathrm{cov}(v_t) = \Sigma_v$, and $\mathrm{var}(e_t) = \sigma_e^2 = \sigma_u^2 - \phi' \Sigma_v \phi$.

If the true $\{v_t\}$ is used in the above regression (model (1)), the OLS estimate of β is unbiased. Since $\{v_t\}$ is unknown, it is replaced by the proxy $\{v_t^c\}$, the residuals from a fitted VAR(1) model for $\{x_t\}$ whose estimated coefficients have reduced bias. The need for bias correction arises because the OLS estimation of the matrix Φ is biased in small samples (see Nicholls and Pope, 1988). The use of the residuals $\{v_t^c\}$ produces a reduced-bias estimate of β. The proposed reduced-bias estimator of the matrix Φ uses a bias expression due to Nicholls and Pope (1988) and gives the following relationship: $E(\hat{\beta} - \beta) = E[\hat{\Phi} - \Phi]' \phi$. The estimation procedure is summarized as follows:

(1) Use Nicholls and Pope's (1988) expression for the bias of the OLS estimator $\hat{\Phi}$,

$$E[\hat{\Phi} - \Phi] = -b/n + O(n^{-3/2}),$$

where

$$b = \Sigma_v \left[(I - \Phi')^{-1} + \Phi'(I - \Phi'^2)^{-1} + \sum_{\lambda \in \mathrm{Spec}(\Phi')} \lambda (I - \lambda \Phi')^{-1} \right] \Sigma_x^{-1},$$

I is a $p \times p$ identity matrix, $\Sigma_x = \mathrm{cov}(x_t)$, λ denotes an eigenvalue of Φ', and the notation $\lambda \in \mathrm{Spec}(\Phi')$ means that the sum is for all p eigenvalues of Φ' with each term repeated as many times as the multiplicity of λ. We estimate the bias b iteratively by repeatedly plugging in preliminary estimates of Φ and Σ_v.

(2) The preliminary estimator of Σ_v is obtained as the sample covariance matrix of the residuals $(x_t - \hat{\Theta} - \hat{\Phi} x_{t-1})$, if $\hat{\Phi}$ has all its eigenvalues smaller than 1 ($\hat{\Theta}$ and $\hat{\Phi}$ are the OLS estimator of the respective matrices). Otherwise, we use the Yule-Walker estimator, which is guaranteed to satisfy the stationarity condition

$$\hat{\Phi}^{YW} = \left[\sum_{t=1}^{n}(x_t - \overline{x}^*)(x_{t-1} - \overline{x}^*)'\right]\left[\sum_{t=0}^{n}(x_t - \overline{x}^*)(x_t - \overline{x}^*)'\right]^{-1},$$

where $\overline{x}^* = \frac{1}{n+1}\sum_{t=0}^{n} x_t$.

(3) The bias b is estimated iteratively.[4] In the kth iteration, $\hat{\Phi}^{(k-1)}$ and $\hat{\Sigma}_v^{(k-1)}$, which are the results from the previous iteration, are used to construct $\hat{b}^{(k)}$. Then $\hat{b}^{(k)}$ is used to calculate $\hat{\Phi}^{(k+1)} = \hat{\Phi}^k + \hat{b}^{(k)}/n$ and $\hat{\Sigma}_v^{(k+1)}$ from the residuals of x_t using $\hat{\Phi}^{(k+1)}$. This iteration process terminates if either the current $\hat{\Phi}^{(k)}$ corresponds to a nonstationary model or a preset maximum of K (in our case, $K = 10$) iterations is reached. We obtain from the last iteration

$$\hat{\Phi}^c = \hat{\Phi} + \hat{b}/n. \tag{2}$$

(4) The corrected residual series $\{v_t^c\}$ is constructed using

$$v_t^c = x_t - (\hat{\Theta}^c + \hat{\Phi}^c x_{t-1}), \qquad t = 1, \ldots, n, \tag{3}$$

where $\hat{\Theta}^c = \bar{x}_t - \hat{\Phi}^c \bar{x}_{t-1}$, the bar indicating sample mean.[5]

The reduced-bias (mARM) parameter estimates are $\hat{\alpha}^c$, $\hat{\beta}^c$, and $\hat{\phi}^c$, the OLS estimators from the multipredictor augmented regression of y_t on a constant, all $x_{j,t-1}$ and $v_{j,t}^c$ $(j = 1, \ldots, p)$, respectively.

2.1 Estimation of cov[$\hat{\boldsymbol{\beta}}^c$]

For hypothesis testing in the multiple predictor case, we need a low-bias estimator of the covariance matrix, cov[$\hat{\beta}^c$]. We cannot use the estimated OLS covariance matrix from the augmented regression, as there is no theoretical justification for this, and indeed our simulations indicate that it produces extremely poor results. We thus construct an estimator of cov[$\hat{\beta}^c$] using the following result.

Lemma 1.

$$E[(\hat{\beta}^c - \beta)(\hat{\beta}^c - \beta)'] = E[(\hat{\Phi}^c - \Phi)'(\phi\phi')(\hat{\Phi}^c - \Phi)] + E[B] \tag{4}$$

where $\hat{\beta}^c$ *is the reduced-bias estimator of* β *obtained from the mARM,* ϕ *is defined in the multipredictor model* (1), $\hat{\Phi}^c$ *is any estimator of* Φ *based on* $x_0, \ldots, x_n$, *p is the number of predictors, and B is a symmetric* $(p \times p)$ *matrix*

[4] Simulations (not shown here) demonstrate that iteration can reduce both the bias and the variance, hence the mean-squared error, of the slope estimator, especially when the sample size is small. In theory, however, at least for the single predictor case, the bias of the (Kendall) corrected estimator is $O(1/n^2)$, regardless of whether iterations are employed.

[5] The OLS estimator $\hat{\Theta}$ can be used without correction because the bias of $\hat{\beta}^c$ does not depend on the estimated intercept.

given by

$$B = \begin{pmatrix} \frac{(\sum_{t=1}^{n} r_{1t} e_t)^2}{(\sum_{t=1}^{n} r_{1t}^2)^2} & \cdots & \frac{(\sum_{t=1}^{n} r_{1t} e_t)(\sum_{s=1}^{n} r_{ps} e_s)}{(\sum_{t=1}^{n} r_{1t}^2)(\sum_{s=1}^{n} r_{ps}^2)} \\ \vdots & \ddots & \vdots \\ \frac{(\sum_{t=1}^{n} r_{pt} e_t)(\sum_{s=1}^{n} r_{1s} e_s)}{(\sum_{t=1}^{n} r_{pt}^2)(\sum_{s=1}^{n} r_{1s}^2)} & \cdots & \frac{(\sum_{t=1}^{n} r_{pt} e_t)^2}{(\sum_{t=1}^{n} r_{pt}^2)^2} \end{pmatrix}.$$

Here, r_{jt} $(j = 1, \ldots, p)$ are the residuals from an OLS regression of the j'th entry of x_{t-1} on all other $(p-1)$ entries of x_{t-1} as well as all p entries of v_t^c and an intercept, with $v_t^c = x_t - \hat{\Phi}^c x_{t-1} - \hat{\Theta}$.

Proof. See Appendix A. ■

Lemma 2.

$$E[B_{i,j}] = \sigma_e^2 E\left[\frac{\sum_{t=1}^{n} r_{it} r_{jt}}{\sum_{t=1}^{n} r_{it}^2 \sum_{t=1}^{n} r_{jt}^2}\right], \qquad i, j = 1, \ldots, p, \tag{5}$$

$$E[B_{j,j}] = E\big[\widehat{SE}\big(\hat{\beta}_j^c\big)\big], \qquad j = 1, \ldots, p \tag{6}$$

where B, $\hat{\beta}_j^c$, and r_{jt} are defined in Lemma 1, σ_e^2 is the variance of e_t in Equation (1), and $\widehat{SE}(\hat{\beta}_j^c)$ is the estimated standard error of $\hat{\beta}_j^c$ from the augmented OLS regression of y_t on $x_{1,t-1}, \ldots, x_{p,t-1}$ and $v_{1t}^c, \ldots, v_{pt}^c$ with intercept: $\widehat{SE}(\hat{\beta}_j^c) = \frac{\hat{\sigma}_e^2}{\sum_{t=1}^{n} r_{jt}^2}$, where $\hat{\sigma}_e^2$ is the OLS estimate of σ_e^2.

Proof. See Appendix B. ■

A feasible approximation for $\text{cov}[\hat{\beta}^c]$ is proposed in the next section.

2.2 Implementation

In the implementation, we make the approximation that $E(\hat{b})$ is b, which is reasonable since $\hat{\Phi}^c$ is a low-bias estimator of Φ. Then, we can approximate $\text{cov}[\hat{\beta}^c]$ by $E[(\hat{\beta}^c - \beta)(\hat{\beta}^c - \beta)']$ and approximate $E[(\hat{\Phi}^c - \Phi)'(\phi\phi')(\hat{\Phi}^c - \Phi)]$ by $\text{cov}[(\hat{\Phi}^c)'\phi]$ (see Equation (4)). Note that the left-hand side of Equation (4) is not $\text{cov}[\hat{\beta}^c] = E[(\hat{\beta}^c - E(\hat{\beta}^c))(\hat{\beta}^c - E(\hat{\beta}^c))']$ because $\hat{\beta}^c$ is a biased estimator of β (although the bias of $\hat{\beta}^c$ is much smaller than that of $\hat{\beta}$). We use the following feasible estimate of $\text{cov}[\hat{\beta}^c]$:

$$\widehat{\text{cov}}^c[\hat{\beta}^c] = \widehat{\text{cov}}[(\hat{\Phi}^c)'\hat{\phi}] + \widehat{E[B]},$$

where $\widehat{\mathrm{cov}}[(\hat{\Phi}^c)'\hat{\phi}]$ and $\widehat{E[B]}$ are estimates of $\mathrm{cov}[(\hat{\Phi}^c)'\hat{\phi}]$ and $E[B]$, corresponding to the formulas

$$
\begin{aligned}
\widehat{\mathrm{var}}^c\left[\hat{\beta}_j^c\right] &= \sum_{i=1}^{p}\left(\hat{\phi}_i^c\right)^2 \widehat{\mathrm{var}}\left[\hat{\Phi}_{ij}^c\right] + \sum_{i=1}^{p}\sum_{k\neq i}^{p} 2\hat{\phi}_i^c\hat{\phi}_k^c \widehat{\mathrm{cov}}\left[\hat{\Phi}_{ij}^c, \hat{\Phi}_{kj}^c\right] \\
&\quad + \left[\widehat{SE}\left(\hat{\beta}_j^c\right)\right]^2, \qquad (j = 1, \ldots, p), \qquad (7)\\
\widehat{\mathrm{cov}}^c\left[\hat{\beta}_i^c, \hat{\beta}_j^c\right] &= \sum_{k=1}^{p}\sum_{l=1}^{p} \hat{\phi}_k^c\hat{\phi}_l^c \widehat{\mathrm{cov}}\left(\hat{\Phi}_{ki}^c, \hat{\Phi}_{lj}^c\right) \\
&\quad + \frac{\hat{\sigma}_e^2\left(\sum_{t=1}^{n} r_{it}r_{jt}\right)}{\left(\sum_{t=1}^{n} r_{it}^2\right)\left(\sum_{t=1}^{n} r_{jt}^2\right)}, \qquad (i, j = 1, \ldots, p) \qquad (8)
\end{aligned}
$$

(Equation (7) is a special case of Equation (8) for $i = j$). Here, $\{\hat{\phi}_k^c\}_{k=1}^p$ are the estimated coefficients of $v_{j,t}^c (j = 1, \ldots, p)$ in the multipredictor augmented regression of y_t on all $x_{j,t-1} (j = 1, \ldots, p)$ and $v_{j,t}^c (j = 1, \ldots, p)$, $v_t^c = x_t - \hat{\Theta}^c - \hat{\Phi}^c x_{t-1}$, $\hat{\Phi}^c$ is the reduced-bias VAR(1) coefficient matrix estimated with iterations based on the bias expression due to Nicholls-Pope (1988) (see Equation (2)), and $\hat{\sigma}_e^2$ is the estimated variance of the error e_t in this regression, $\hat{\sigma}_e^2 = \frac{\mathrm{RSS}}{n-(2p+1)}$, where RSS is the residual sum of squares of the augmented regression, $\{r_{jt}\}_{t=1}^n$ for $(j = 1, \ldots, p)$ are the residuals as defined in Lemma 1. The terms $\widehat{\mathrm{var}}(\hat{\Phi}_{ij}^c)$ and $\widehat{\mathrm{cov}}(\hat{\Phi}_{ki}^c, \hat{\Phi}_{lj}^c)$ are read from the estimated covariance matrix $\hat{\Sigma}_v^c \bigotimes (X'X)^{-1}$ (the covariance matrix of $\hat{\Theta}^c$ and $\hat{\Phi}^c$), where $\hat{\Sigma}_v^c$ is the estimated covariance matrix of v_t, and $X = (\mathbf{1}, X_1, \ldots, X_p)$ is the $(n \times (p + 1))$ matrix of the regressors with a unit vector as the first column.

It is worth noting that for testing the significance of each β_j, $j = 1, \ldots, p$, formula (7) is sufficient. Formula (8) is necessary for joint tests involving all of the β_j's.

The following is a summary of the procedure for an example of a two-predictor model.

(1) Regress x_t on x_{t-1} $(x_t = (x_{1,t}, x_{2,t})')$ to obtain $\hat{\Phi}$, which is used to construct a reduced-bias estimator $\hat{\Phi}^c$ using Nicholls and Pope's (1988) bias expression with iterations.

(2) Construct the bivariate corrected residual series $v_t^c = y_t - \hat{\Theta}^c - \hat{\Phi}^c x_{t-1}$. Denote $v_t^c = (v_{1,t}^c, v_{2,t}^c)'$. Then use v_t^c to estimate $\mathrm{cov}(\hat{\Phi}^c)$, following the procedure described above.

(3) Obtain $\hat{\beta}_1^c$ and $\hat{\beta}_2^c$ as the coefficients of $x_{1,t-1}$ and $x_{2,t-1}$ in a regression of y_t on $x_{1,t-1}$, $x_{2,t-1}$, $v_{1,t}^c$, and $v_{2,t}^c$, with intercept. This regression also produces $\hat{\phi}_1^c$ and $\hat{\phi}_2^c$ as the coefficients of $v_{1,t}^c$ and $v_{2,t}^c$, and it produces the (2×2) estimated covariance matrix $\widehat{\mathrm{cov}}[\hat{\beta}^c]$ whose diagonal elements are used in Equation (7).

(4) Apply formula (7) to obtain $\widehat{\text{var}}^c[\hat{\beta}_1^c]$, $\widehat{\text{var}}^c[\hat{\beta}_2^c]$ whose square roots are used for testing a hypothesis on either β_1 or β_2.

(5) For the joint test on (β_1, β_2), calculate $\{r_{1t}\}$, the residuals from a regression of x_{1t} on x_{2t} and $v_{1,t}^c$, $v_{2,t}^c$, and $\{r_{2t}\}$ accordingly, and then apply formula (8).

2.3 Hypothesis testing

Using mARM, we obtain the reduced-bias estimator $\hat{\beta}^c$ and an estimate of its covariance matrix, which we use in hypothesis testing. Tests of significance of individual coefficients—tests of the null hypothesis $H_0 : \beta_j = 0$—are performed using an ordinary t-statistic, with $t = \hat{\beta}_j^c / \sqrt{\widehat{\text{var}}^c[\hat{\beta}_j^c]}$ as it would be done in OLS, except that the parameters are estimated in a different way.

For the joint test of H_0 : all $\beta_j = 0$, $j = (1, \ldots, p)$ against H_a : at least one $\beta_j \neq 0$, $j = (1, \ldots, p)$, the Wald test is used, employing the matrix $\widehat{\text{cov}}^c[\hat{\beta}^c]$ and the formula $Wald(\hat{\beta}^c) = \hat{\beta}^c(\widehat{\text{cov}}^c[\hat{\beta}^c])^{-1}(\hat{\beta}^c)'$ where $\hat{\beta}^c = (\hat{\beta}_1^c, \ldots, \hat{\beta}_p^c)$.

2.4 Simulations

We use simulations to evaluate the performance of the mARM. We first estimate the predictors' coefficients, showing the bias under the OLS and mARM. We show that the mARM produces a very small bias compared to that under the OLS. Next, we analyze the performance of hypothesis testing under mARM, comparing it with two commonly used methods: OLS and bootstrapping. We show that the size under mARM is very close to the nominal size, compared to the size under the competing test methods.

The analysis employs a two-predictor case with a nondiagonal Φ matrix. $\hat{\beta}^c$ is estimated by mARM as in Equation (1) with v_t replaced by v_t^c, defined in Equation (3). In these simulations, we estimate the covariance matrix of $\hat{\beta}^c$ by Equations (7) and (8) and use it in hypothesis testing.

We use three nondiagonal choices of the AR(1) parameter matrix Φ.

Case 1.

$$\Phi_1 = \begin{pmatrix} 0.80 & 0.1 \\ 0.1 & 0.85 \end{pmatrix},$$

Case 2.

$$\Phi_2 = \begin{pmatrix} 0.80 & 0.1 \\ 0.1 & 0.94 \end{pmatrix},$$

both with

$$\Sigma_v = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}.$$

Case 3.

$$\Phi_3 = \begin{pmatrix} 0.9928 & -0.0007 \\ 0.0040 & 0.9779 \end{pmatrix},$$

with

$$\Sigma_v = \begin{pmatrix} 5.8252 \times 10^{-5} & 4.3533 \times 10^{-5} \\ 4.3533 \times 10^{-5} & 3.9841 \times 10^{-3} \end{pmatrix}.$$

For all processes, we generated 1500 simulated replications.

In setting the parameter values of Φ, we note that, in general, the closer the largest absolute eigenvalue of Φ to 1, the more nearly nonstationary the multiple VAR(1) model. Here, the largest absolute eigenvalues in Cases 1, 2, and 3 are 0.928, 0.992, and 0.993, respectively. Case 2 is purposely designed to study a case that is close to nonstationarity. In Case 3, the parameter values are estimated from the data in Santos and Veronesi (2006), which we later use in the empirical section. In this model, quarterly stock returns are predicted by the ratio of labor income to disposable income and by the log-dividend yield over the period 1948–1994.

2.4.1 Performance of parameter estimates. Consider first the results for Φ_1 and Φ_2 and recall that the coefficients' values are zero. For Φ_1 and $n = 50$, the OLS regression generates biased average estimates, 6.47 for $\hat{\beta}_1$ and 8.24 for $\hat{\beta}_2$. The mARM greatly reduces the average bias: 0.97 for $\hat{\beta}_1^c$ and 1.88 for $\hat{\beta}_2^c$, respectively. For Φ_2 and $n = 50$, the bias is greater under both the OLS and the mARM, showing the effect of the near unit root. Still, the bias under mARM is much smaller than under OLS. Naturally, when the sample size increases to 200, the average bias is smaller under both OLS and the mARM, with mARM being much better: For Φ_1 and $n = 200$, the averages bias under OLS is 1.24 for $\hat{\beta}_1$ and 2.13 for $\hat{\beta}_2$, while under mARM it is much closer to zero: -0.23 for $\hat{\beta}_1^c$ and 0.41 for $\hat{\beta}_2^c$. As expected, $\hat{\phi}_1^c$ and $\hat{\phi}_2^c$ are unbiased[6] in all cases (true parameter values are $\phi = (-80, -80)$).

We also obtain that the our procedure of estimating the standard errors of the mARM coefficients, detailed in Sections 2.1 and 2.2, produces estimates that are close to the true standard deviations of the estimated coefficients. For example, under Φ_1 and $n = 50$, the average of $\widehat{SE}^c(\hat{\beta}_1^c)$ is 17.49 compared to an estimated standard deviation of 18.22, a relative error of 4%. For Φ_1 and $n = 200$, the mARM-estimated standard error of $\hat{\beta}_1^c$ is 7.83 while its actual standard deviation is 7.86, a difference of only 1%. For Φ_2, where the VAR(1) process is close to nonstationary, the relative errors are 2.2% for $n = 50$ and 1.0% for $n = 200$. For β_2, the respective errors of the average of $\widehat{SE}^c(\hat{\beta}_2^c)$ under Φ_1 are 6.3% and 0.2% for $n = 50$ and $n = 200$, respectively, and under Φ_2, the

[6] This is shown in Amihud and Hurvich (2004).

Table 1
Results for the two-predictor models (9) and (10)

	$(n = 50)$		$(n = 200)$		$(n = 188)$
	Φ_1	Φ_2	Φ_1	Φ_2	Φ_3
OLS mean ($\hat{\beta}_1$)	6.4672	6.9324	1.2433	1.2375	0.0128
mARM mean ($\hat{\beta}_1^c$)	0.9689	2.4161	−0.2314	0.0130	0.0024
simulated standard deviation ($\hat{\beta}_1^c$)	18.2208	17.9288	7.9149	7.9086	0.2169
mARM mean ($\widehat{SE}^c(\hat{\beta}_1^c)$)	17.4850	17.5278	7.8335	7.8333	0.2123
OLS mean ($\hat{\beta}_2$)	8.2419	10.1770	2.1311	2.6111	0.0235
mARM mean ($\hat{\beta}_2^c$)	1.8772	3.0356	0.41452	0.7155	0.0078
simulated standard deviation ($\hat{\beta}_2^c$)	17.0641	14.5353	6.9168	5.0569	0.0324
mARM mean ($\widehat{SE}^c(\hat{\beta}_2^c)$)	15.9910	12.7961	6.9304	4.7442	0.0300
OLS mean ($\hat{\phi}_1^c$)	−79.9978	−79.9984	−79.9978	−79.9977	−0.9321
mARM mean ($\hat{\phi}_2^c$)	−80.0008	−80.0002	−80.0025	−80.0025	−0.6847
simulated cov($\hat{\beta}_1^c, \hat{\beta}_2^c$)	−191.1199	−144.6155	−39.9533	−31.1995	-4.1776×10^{-4}
mARM mean ($\widehat{\text{cov}}^c(\hat{\beta}_1^c, \hat{\beta}_2^c)$)	−199.5683	−158.9612	−42.7049	−32.0820	-3.7616×10^{-4}

Simulation results: parameter estimates for a two-predictor model.
1500 replications are generated from the two-predictor models.

$$y_t = \alpha + \beta' x_{t-1} + u_t, \tag{9}$$

$$x_t = \theta + \Phi x_{t-1} + v_t, \tag{10}$$

where β, x_t, θ and v_t are (2×1) matrices. Φ is a (2×2) matrix, i.e.,

$$y_t = \alpha + \beta_1 x_{1,(t-1)} + \beta_2 x_{2,(t-1)} + u_t, \qquad \begin{pmatrix} x_{1,t} \\ x_{2,t} \end{pmatrix} = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} + \begin{pmatrix} \Phi_{11} & \Phi_{12} \\ \Phi_{21} & \Phi_{22} \end{pmatrix} \begin{pmatrix} x_{1,(t-1)} \\ x_{2,(t-1)} \end{pmatrix} + \begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix}.$$

Table 1 presents estimation results of the two-predictor model by OLS as well as by the multipredictor augmented regression method (mARM). The five-step estimation procedure is described in Section 2.2.

Three cases are considered. In Cases 1 and 2, the parameters are $\alpha = 0$, $\beta = (0, 0)'$, $\Theta = (0, 0)'$, $u_t = \phi' v_t + e_t$; e_t is $N(0, 1)$, $\phi = (\phi_1, \phi_2)' = (-80, -80)'$, v_t is $N(0, \Sigma_v)$. $\{e_t\}$ and $\{v_t\}$ are mutually and serially independent. $n = 50$ or 200. The nondiagonal VAR(1) parameter matrices are

$$\Phi_1 = \begin{pmatrix} 0.80 & 0.1 \\ 0.1 & 0.85 \end{pmatrix},$$

and

$$\Phi_2 = \begin{pmatrix} 0.80 & 0.1 \\ 0.1 & 0.94 \end{pmatrix},$$

all with

$$\Sigma_v = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}.$$

The eigenvalues of Φ_1 are 0.722 and 0.928, and of Φ_2 they are 0.748 and 0.992.

Case 3 uses parameters that are estimated from Santos and Veronesi's (2006) data. $n = 188$, $\alpha = 0$, $\beta = (0, 0)'$, $\Theta = (0, 0)'$, the $\{u_t\}$ are generated as $u_t = \phi' v_t + e_t$, where e_t is normal with mean zero and variance 4.0161×10^{-3}, $\phi = (\phi_1, \phi_2)' = (-0.9307, -0.6824)'$, v_t is $N(0, \Sigma_v)$, and $\{e_t\}$, $\{v_t\}$ are mutually and serially independent. The VAR(1) parameter matrix is

$$\Phi_3 = \begin{pmatrix} 0.9928 & -0.0007 \\ 0.0040 & 0.9779 \end{pmatrix},$$

and

$$\Sigma_v = \begin{pmatrix} 5.8252 \times 10^{-5} & 4.3533 \times 10^{-5} \\ 4.3533 \times 10^{-5} & 3.9841 \times 10^{-3} \end{pmatrix}.$$

The matrix Φ_3 has eigenvalues of 0.978 and 0.993.

Estimates: $\hat{\beta}_j$ and $\hat{\beta}_j^c$ are, respectively, coefficients estimated by OLS and mARM; $\widehat{SE}(\hat{\beta}_j)$ and $\widehat{SE}(\hat{\beta}_j^c)$ are the corresponding estimated standard errors; $\hat{\phi}_j^c$ is the estimated coefficient of $v_{t,j}^c$ from mARM ($v_{t,j}^c$ is the residual of x_t in the VAR(1) model, using the reduced-bias matrix $\hat{\Phi}^c$); $\widehat{SE}^c(\hat{\beta}_j^c)$ and $\widehat{\text{cov}}^c(\hat{\beta}_i^c, \hat{\beta}_j^c)$ are the corrected variance and covariance estimates of $\hat{\beta}^c$ using Equations (7) and (8); the simulated standard deviations and covariance are based on 1500 realizations.

respective errors are 8.8% and 6.2%, respectively. The estimates $\widehat{\text{cov}}^c[\hat{\beta}_1^c, \hat{\beta}_2^c]$, which are used for the joint test, are also quite close to the actual $\text{cov}[\hat{\beta}_1^c, \hat{\beta}_2^c]$. These estimates affect the hypothesis testing, to which we attend in the next subsection.

Next, we consider Case 3, in which the parameter values are obtained from estimates based on the data of Santos and Veronesi (2006). Here, the bias under OLS is small. Still, mARM produces strong bias reduction, similar to that obtained in Cases 1 and 2.

2.4.2 Performance of hypothesis tests. We compare the performance of mARM in hypothesis testing the OLS and the widely used bootstrapping procedure. The bootstrapping tests are done in two different ways. In one, we bootstrap on the estimated coefficients, β_1 and β_2, and in the other, we simulate on their t-statistics t_1 and t_2. We denote these two methods by BS_{t_1} and BS_{t_2}. It is customary in the finance literature to bootstrap on the estimated predictive coefficient, e.g. Kothari and Shanken (1997) and Baker and Stein (2004). However, we show that the performance of bootstrapping test is better when doing it on the coefficient's t-statistic. Indeed, it is known that bootstrapping on the corresponding t-statistic produces better sizes than bootstrapping on $\hat{\beta}$.[7] (Our bootstrap procedure used here is detailed in the Appendix C.) We also present simulations of the joint Wald test. The null hypothesis in all tests is that the values of the coefficients is zero.

The analysis examines the actual sizes that are obtained when the nominal size is 5%, using the same model parameters in all methods.[8] We present results for individual coefficient tests—one (right)-tail test and two-tail tests—and for the joint Wald test.

The results are presented in Table 2. For individual tests of each coefficient β_1 and β_2, the sizes under mARM tests are uniformly better—closer to nominal—than under OLS and both bootstrap-based tests. Consider, for example, the right-tailed test (panel A(1)) for β_1. For Φ_1 and $n = 50$, the sizes under mARM is 6.4% (nominal size is 5%). For OLS, the respective size is 11.9% and, under bootstrapping using OLS, it is 11.8% when doing the bootstrapping using the estimate of $beta_1$ and 7.6% when using its t-value. The size under bootstrapping when using the reduced-bias $\hat{\beta}_1^c$ is 6.2% and when using its t-value, the size is 6.3%, similar to that obtained under mARM with a standard t-test. It follows that the mARM-based simple t-test performs as well as the more cumbersome bootstrapping when using the estimated parameters from mARM, albeit with much less effort. It surely outperforms bootstrapping using OLS estimates, which produce inflated size. When the sample size increases, the sizes for mARM are quite close to the nominal sizes. The sizes of the mARM tests are

---

7 See Hall and LePage (1996).

8 We also performed the hypothesis test for the nominal sizes 1% and 10%. The results are qualitatively similar.

Table 2
Simulation results: hypothesis testing size and power for two-predictor model

Panel A: Size for nominal 5% test

(1) Right-tailed test and joint test

OLS		$t(\hat{\beta}_1)$ (%)		$t(\hat{\beta}_2)$ (%)		Wald (%)
$n=50$	Φ_1	11.9		15.3		12.3
	Φ_2	12.9		23.3		21.8
$n=200$	Φ_1	7.3		8.9		7.9
	Φ_2	7.4		14.6		15.2
$n=188$	Φ_3	8.3		18.1		8.9
Bootstrap using OLS estimates		BS_{β_1} (%)	BS_{t_1} (%)	BS_{β_2} (%)	BS_{t_2} (%)	Wald (%)
$n=50$	Φ_1	11.8	7.6	21.5	14.9	9.5
	Φ_2	17.8	7.9	30.5	21.5	17.7
$n=200$	Φ_1	5.3	4.5	15.9	10.2	7.3
	Φ_2	9.5	4.8	26.3	18.8	13.5
$n=188$	Φ_3	42.2	7.9	14.7	16.8	8.8
Bootstrap using mARM estimates		BS_{β_1} (%)	BS_{t_1} (%)	BS_{β_2} (%)	BS_{t_2} (%)	Wald (%)
$n=50$	Φ_1	6.2	6.3	24.2	12.0	8.0
	Φ_2	6.1	6.0	31.8	20.2	13.0
$n=200$	Φ_1	5.4	4.6	18.7	8.9	6.3
	Φ_2	6.0	4.5	28.5	18.7	12.8
$n=188$	Φ_3	15.5	5.1	37.3	15.4	7.6
mARM		$t(\hat{\beta}_1^c)$ (%)		$t(\hat{\beta}_2^c)$ (%)		Wald (%)
$n=50$	Φ_1	6.4		6.8		9.1
	Φ_2	6.5		9.8		9.6
$n=200$	Φ_1	4.8		5.5		6.9
	Φ_2	5.1		7.7		10.5
$n=188$	Φ_3	5.1		7.4		6.0

(2) Two-tailed test

OLS		$t(\hat{\beta}_1)$ (%)		$t(\hat{\beta}_2)$ (%)	
$n=50$	Φ_1	9.2		10.6	
	Φ_2	9.2		16.1	
$n=200$	Φ_1	6.3		6.1	
	Φ_2	6.7		9.7	
$n=188$	Φ_3	7.5		11.1	
Bootstrap using OLS estimates		BS_{β_1} (%)	BS_{t_1} (%)	BS_{β_2} (%)	BS_{t_2} (%)
$n=50$	Φ_1	25.6	9.3	25.8	10.4
	Φ_2	36.5	7.4	31.9	14.3
$n=200$	Φ_1	16.6	6.7	16.7	7.3
	Φ_2	24.1	7.0	26.2	12.4
$n=188$	Φ_3	79.8	7.4	9.3	9.8
Bootstrap using mARM estimates		BS_{β_1} (%)	BS_{t_1} (%)	BS_{β_2} (%)	BS_{t_2} (%)
$n=50$	Φ_1	19.4	12.2	17.8	7.5
	Φ_2	21.5	13.9	21.6	12.3
$n=200$	Φ_1	16.2	8.1	13.9	5.5
	Φ_2	17.5	9.3	21.9	11.4
$n=188$	Φ_3	29.9	7.9	34.9	9.2
mARM		$t(\hat{\beta}_1^c)$ (%)		$t(\hat{\beta}_2^c)$ (%)	
$n=50$	Φ_1	6.3		6.3	
	Φ_2	5.7		8.3	
$n=200$	Φ_1	6.2		5.4	
	Φ_2	6.0		6.4	
$n=188$	Φ_3	5.5		6.4	

Table 2
(Continued)

Panel B: Size-controlled power comparison between mARM-based and bootstrap-based hypothesis tests

(1) Right-tailed tests and joint tests

		Bootstrap			mARM		
VAR(1) matrix	True β	BS_{t_1} (%)	BS_{t_2} (%)	Wald (%)	$t(\hat{\beta}_1^c)$ (%)	$t(\hat{\beta}_2^c)$ (%)	Wald (%)
Φ_1	(0,0)	5.0	5.0	5.0	5.0	5.0	5.0
	(5,5)	13.8	16.4	60.4	14.2	17.5	29.0
	(10,10)	33.5	40.7	100.0	36.0	42.1	100.0
	(20,20)	79.7	87.5	100.0	82.4	90.3	100.0
Φ_2	(0,0)	5.0	5.0	5.0	5.0	5.0	5.0
	(5,5)	13.1	28.4	100.0	13.7	26.6	96.1
	(10,10)	30.0	66.7	100.0	33.6	66.4	100.0
	(20,20)	79.5	97.7	100.0	82.9	98.8	100.0

(2) Two-tailed test

		Bootstrap		mARM	
VAR(1) matrix	True β	BS_{t_1} (%)	BS_{t_2} (%)	$t(\hat{\beta}_1^c)$ (%)	$t(\hat{\beta}_2^c)$ (%)
Φ_1	(0,0)	5.0	5.0	5.0	5.0
	(5,5)	6.9	14.5	8.0	10.7
	(10,10)	16.8	37.2	21.3	30.7
	(20,20)	62.5	85.8	70.1	83.8
Φ_2	(0,0)	5.0	5.0	5.0	5.0
	(5,5)	6.1	27.1	7.4	19.4
	(10,10)	15.9	65.6	20.1	56.5
	(20,20)	60.3	97.7	70.4	97.7

The parameters are the same as in Table 1. Results are based on 1500 replications from the two-predictor models. The results pertain to nominal size of 5%. The sizes obtained from the multipredictor augmented regression method (mARM) are compared with OLS and bootstrapping.

BS denotes the bootstrap method and the subscripts "β" or "t" imply that the test is based on the empirical distribution of the slope coefficient or its t-statistic, respectively. Subscripts "1" or "2" are for the first or the second slope coefficient. t_1 and t_2 are the t-statistics for coefficients β_1 and β_2 from the mARM. The Wald test in the bootstrap method is based on the empirical distribution of the Wald-statistic and the Wald test in mARM is calculated based on $\hat{\beta}_1^c$, $\hat{\beta}_2^c$ and the corresponding estimated covariance matrix from the augmented regression.

For conciseness, in panel B, we present the power comparison between methods only for $n = 200$ and VAR(1) coefficient matrices Φ_1 and Φ_2. (The results are qualitatively similar for $n = 50$ or $n = 188$.)

more inflated—though still closer to the nominal values than the bootstrapping-based tests—when the absolute eigenvalue of Φ is very close to 1 and the sample size is small. The advantage of mARM over OLS and bootstrapping also holds under two-tailed tests (panel A(2)). The sizes under mARM are even closer to nominal than under the right-tailed tests, and they are closer to nominal value than under bootstrapping.

For the joint test of $(\beta_1, \beta_2) = (0, 0)$, the mARM-based test again improves on the bootstrap-based test. For Φ_1 and $n = 50$, the size of the joint test under mARM is 9.1% while it is 12.3% under OLS and 9.5% under bootstrapping. Again, the sizes are closer to nominal when the largest absolute eigenvalue of matrix Φ is not too close to 1 and when n is larger. Furthermore, the improvement of mARM over OLS and bootstrapping is greater under Φ_2 and $n = 50$, i.e., when the potential bias problem is greater.

In Table 2, panel B, we present results on the power of the tests, comparing the *size-adjusted* power of mARM with that under bootstrapping. By construction, the power equals the nominal size at $(\beta_1, \beta_2) = (0, 0)$. The power for individual coefficients tests are roughly comparable for the two methods while the power is better under bootstrapping for the joint test. Of course, the size-adjusted test is infeasible for practical implementation.

For Case 3, as in Cases 1 and 2, the mARM-based hypothesis test provides far more accurate size than the OLS-based and both bootstrap-based tests. The mARM-based bootstrap test generally performs better than the OLS-based bootstrap test, but both are strongly outperformed by our proposed mARM test based on straightforward t-values.

3. Case Study: Santos and Veronesi Revisited

We apply our mARM to test the two-predictor model proposed by Santos and Veronesi (2006). In their model, stock returns are predicted by two variables: dividend yield and one of the following three ratios: (1) labor income to consumption, $s^w_{t,1} = w_t/C_t$, where w_t is the labor income and C_t is the consumption of nondurables plus services; (2) employee compensation to consumption, $s^w_{t,2} = w^{ce}_t/C_t$; and (3) labor income to disposable income, $s^w_{t,3} = w_t/Y_t$. Using quarterly data, Santos and Veronesi (2006) estimate long-horizon four to predictability by regressing the excess stock returns on lagged values of $\log(D/P)$ and one of the predictors $s^w_{t,j}$, $(j = 1, 2, 3)$, at lags of 4, 8, 12, and 16 quarters. They do the estimation by OLS, which does not attend to the potential bias that may arise in predictive regressions with persistent predictors and small samples (Stambaugh, 1999). Indeed, in this case, all predictors are quite persistent: the sample autocorrelations of $s^w_{t,j}$, $(j = 1, 2, 3)$ and of $\log(D/P)$ are all greater than 0.975 (see their Table 1).

Boudoukh et al. (2006) show that statistical significance in a long-horizon predictive model can be obtained even under the null of no predictability. This may arise from the combined effects of overlapping returns and the persistence of the predictors. Then, the estimated OLS coefficients and R^2 are approximately proportional to the horizon. It may therefore be that a predictor's coefficient that is insignificantly different from zero at a short horizon becomes significant in a longer-horizon estimation. This holds even after the standard errors are corrected for the overlapping returns.[9] Indeed, in Santos and Veronesi's (2006) Tables 2 and 3, the OLS estimates as well as the adjusted R^2s increase monotonically (and roughly proportionally) as the horizon increases. Therefore, we estimate here one-quarter-ahead predictive regressions. This may be regarded as a robustness check on the conclusions of Santos and Veronesi (2006). In particular, our estimation attends to the potential bias that may arise in small-sample multipredictor regressions.

The results are summarized in Table 3. We consider two-predictor regressions, consisting of one of the ratios $s^w_{t,j}$, $j = 1, 2, 3$ together with the

[9] Boudoukh et al. (2006) show their results for a single-predictor model. A similar problem presumably holds in long-horizon regressions with multiple predictors.

Table 3
Empirical tests: predicting excess stock returns by ratio of labor income to consumption and log-dividend yield

		1948–1994			1948–2001		
Variable	Method	Estimate	t-stat (NW)	p-value	Estimate	t-stat (NW)	p-value
$s^w_{t,1}$	OLS	−0.2517	−1.37	0.0853	−0.4035	−2.35	0.0094**
	mARM	−0.2592	−1.40	0.0808	−0.3918	−2.26	0.0119*
$\log(DP)$	OLS	0.0657	3.42	0.0003**	0.0444	2.86	0.0021**
	mARM	0.0467	2.41	0.0080**	0.0288	1.84	0.0329*
$s^w_{t,2}$	OLS	−0.2504	−1.73	0.0418*	−0.3879	−2.82	0.0024**
	mARM	−0.1353	−0.91	0.1814	−0.3715	−2.66	0.0039**
$\log(DP)$	OLS	0.0628	3.40	0.0003**	0.0428	2.79	0.0026**
	mARM	0.0590	3.17	0.0008**	0.0270	1.74	0.0409*
$s^w_{t,3}$	OLS	−0.2803	−0.86	0.1949	−0.5337	−1.68	0.0465*
	mARM	−0.0534	−0.16	0.4365	−0.4136	−1.29	0.0986
$\log(DP)$	OLS	0.0665	3.27	0.0005**	0.0315	2.14	0.0162*
	mARM	0.0728	3.54	0.0002**	0.0216	1.45	0.0735

The table presents results for mARM regression using the following model:

$$R_t = \beta_0 + \beta_1 s^w_{t-1,j} + \beta_2 \log(DP) + \phi_1 v^c_{s,t} + \phi_2 v^c_{DP,t}, \qquad (j = 1, 2, 3)$$

where R_t is the value-weighted market excess return, $s^w_{t,j}$ is the ratio of labor income to consumption, and $\log(DP)$ is the log-dividend yield. $\{v^c_{s,t}\}$ and $\{v^c_{DP,t}\}$ are the residuals obtained from the multivariate estimation procedure of the autoregressive matrix Φ of the two-predictor variables after using Nicholls and Pope's (1988) correction (see Equation (3)). The corresponding OLS model does not include $v^c_{s,t}$ and $v^c_{DP,t}$.

Following Santos and Veronesi (2006), there are three labor income to consumption ratios, $(j = 1, 2, 3)$: (1) labor income to consumption, $s^w_{t,1} = w_t/C_t$, where w_t is the labor income and C_t is the consumption of nondurables plus services; (2) employee compensation to consumption, $s^w_{t,2} = w^{ce}_t/C_t$; and (3) labor income to disposable income, $s^w_{t,3} = w_t/Y_t$.

In the table, the p-values for $s^w_t = w_t/C_t$, w^{ce}_t/C_t, or w_t/Y_t are left-tailed and the p-values for $\log(D/P)$ are right-tailed (reflecting the alternative hypotheses to the null). * and ** indicate significance at 5% and 1%, respectively.

log-dividend yield, and do the estimations for two sample periods, 1948–2001 and 1948–1994, as do Santos and Veronesi. In the OLS regressions, we report both the ordinary t-statistic and the Newey-West adjusted t-statistic, as did Santos and Veronesi (2006).[10] For mARM, we report test statistics using the adjusted standard error calculated by Lemmas 1 and 2. To account for potential error autocorrelation, we performed a simple further modification of the mARM standard errors, multiplying them by the ratio of the (OLS-based) Newey-West standard error to the OLS standard error. We refer to the resulting modified standard error as the Newey-West standard error for mARM.[11]

The p-values that we present for the tests of significance of the coefficients are left-tailed for $s^w_{t,j}$, $(j = 1, 2, 3)$, since the alternative hypothesis to the null is that the coefficient of $s^w_{t,j}$ is negative. This follows from Santos and Veronesi (2006, p. 19) statement: "The main prediction of our model is that a high share of labor income to consumption ... predicts low future returns." The predictive

[10] We do not use the Hodrick (1992) standard errors, which account for the overlapping-errors structure induced by long-horizon regression (see Ang and Bekeart, 2005), since we perform only nonoverlapping one-quarter-ahead predictive regressions.

[11] For the data set considered here, the Newey-West correction has little effect, and and it does not change any conclusion.

coefficient of $\log(D/P)$ is well known to be positive, so for this predictor the p-value corresponds to a right-tailed test.

For the 5% benchmark of significance, there are two cases where the income-to-consumption ratios are significant under OLS but insignificant under mARM. In the 1948–1994 regression using the ratio $s^w_{t,2}$ (employee compensation to consumption), the OLS p-value is 0.0418 while under mARM, the p-value is 0.1814, indicating insignificance. Notably, the estimated OLS coefficient of $s^w_{t,2}$ is twice as large as the mARM reduced-bias coefficient. In the 1948–2001 regression, the coefficient of $s^w_{t,3}$ (ratio of labor income to disposable income) has p-value of 0.0465 under OLS but 0.0986 under mARM, insignificant. If the benchmark of significance is 1%, the coefficient of $s^w_{t,1}$ in the 1948–2001 regression is significant under OLS but insignificant under mARM. In five of the six cases considered, the mARM-based p-values of the $s^w_{t,j}$ variables are larger (indicating weaker significance) than the OLS-based p-value. The log-dividend yield too appears to have weaker significance—lower p-values—under mARM compared with OLS in five out of six cases. The coefficient of the log-dividend yield fails to pass the 5% hurdle in the third regression for 1948–2001 using mARM while it appears significant under OLS. Under a 1% benchmark of significance, it is significant under OLS in the second regression for the 1948–2001 period but insignificant under mARM. In five out of the six cases, the p-values under mARM are larger (indicating weaker significance) than under OLS. In summary, the case for predictability is weaker in estimations that employ the mARM compared to OLS, and in some cases significant coefficients become insignificant when employing mARM.

4. Concluding Remarks

In this paper, we propose a hypothesis-testing method for multipredictor regression models and have compared it to results obtained under the standard OLS and bootstrap methods. Our method, denoted by mARM, provides for both parameter estimation and hypothesis testing in a unified and convenient way. We evaluate the performance of our method in a simulation study. The individual t-tests, especially the right-tailed case, perform well in terms of size, and have size closer to nominal than tests based on OLS or on the bootstrap method. Our method also enables *joint* testing of all predictive coefficients using the Wald test, which also produces size closer to nominal than OLS and bootstrapping. The advantage of the mARM test is greater when the sample size is small or when the matrix of the regressors' vector autoregressive coefficients has eigenvalues close to 1.

Applying our methodology to estimate multipredictor models where stock returns are predicted by variables employed in a previous study, log-dividend yield together with an income/consumption ratio, we find that in some cases, the null hypothesis of no predictive power is rejected based on OLS and bootstrapping while it is not rejected when using mARM. This suggests that, in general, some previously reported conclusions about the predictive power of certain economic variables may need to be revisited.

Appendix

A. Proof of Lemma 1

We can write

$$y_t = \tilde{\alpha} + \{\beta' + \phi'(\hat{\Phi}^c - \Phi)\}x_{t-1} + \phi' v_t^c + e_t, \tag{A1}$$

where $\tilde{\alpha} = \alpha + \phi'(\hat{\Theta}^c - \Theta)$ is a constant with respect to t. Next, define the $p \times 1$ vectors $\{r_t\}_{t=1}^n$ by $r_t = (r_{1t}, \ldots, r_{pt})'$ where for $j = 1, \ldots, n$, $\{r_{jt}\}_{t=1}^n$ is the (row) vector of residuals from a $2p-1$-variable OLS regression of the j'th entry of x_{t-1} on all other entries of x_{t-1} as well as all p entries of v_t^c and an intercept. Correspondingly, define $\{\tilde{r}_t\}_{t=1}^n$ by $\tilde{r}_t = (r_{1t}/\Sigma r_{1t}^2, \ldots, r_{pt}/\Sigma r_{pt}^2)'$ and write $x_t = (x_{1t}, \ldots, x_{pt})'$ and $v_t^c = (v_{1t}^c, \ldots, v_{pt}^c)'$. It follows that

$$\hat{\beta}^c = \sum_{t=1}^{n} \tilde{r}_t y_t, \tag{A2}$$

and for all $j, k \in \{1, \ldots, p\}$ with $j \neq k$,

$$\sum_{t=1}^{n} \tilde{r}_{jt} = \sum_{t=1}^{n} \tilde{r}_{jt}\, x_{k,t-1} = \sum_{t=1}^{n} \tilde{r}_{jt}\, v_{jt}^c = \sum_{t=1}^{n} \tilde{r}_{jt}\, v_{kt}^c = 0, \tag{A3}$$

and

$$\sum_{t=1}^{n} \tilde{r}_{jt}\, x_{j,t-1} = \sum_{t=1}^{n} \tilde{r}_{jt}\, r_{jt} = 1. \tag{A4}$$

Substituting y_t from Equation (A1) in Equation (A2) and using Equations (A3) and (A4) yields

$$\hat{\beta}^c = \beta + (\hat{\Phi}^c - \Phi)'\phi + \sum_{t=1}^{n} \tilde{r}_t e_t. \tag{A5}$$

The lemma now follows, since e_t has mean 0 and is independent of $\tilde{r}_t$ and $\hat{\Phi}^c$. ■

B. Proof of Lemma 2

The matrix B is defined as in Lemma 1:

$$B = \begin{pmatrix} \frac{(\sum_{t=1}^n r_{1t}e_t)^2}{(\sum_{t=1}^n r_{1t}^2)^2} & \cdots & \frac{(\sum_{t=1}^n r_{1t}e_t)(\sum_{s=1}^n r_{ps}e_s)}{(\sum_{t=1}^n r_{1t}^2)(\sum_{s=1}^n r_{ps}^2)} \\ \vdots & \ddots & \vdots \\ \frac{(\sum_{t=1}^n r_{pt}e_t)(\sum_{s=1}^n r_{1s}e_s)}{(\sum_{t=1}^n r_{pt}^2)(\sum_{s=1}^n r_{1s}^2)} & \cdots & \frac{(\sum_{t=1}^n r_{pt}e_t)^2}{(\sum_{t=1}^n r_{pt}^2)^2} \end{pmatrix}.$$

First, we show the independence between $\{e_t\}$ and $\{r_{jt}\}$.

Because $v_t^c = x_t - \hat{\Theta} - \hat{\Phi}^c$ and $\hat{\Theta}$, $\hat{\Phi}^c$ are all functions of x_t, all entries of v_t^c must be functions of x_t. But e_t is independent of x_t and v_t, so e_t must be independent of v_t^c and therefore independent of r_{it} or r_{jt} for any $i, j = 1, \ldots, p$. Recall that e_t has mean zero.

Therefore,

$$
\begin{aligned}
E[B_{i,j}] &= E\left[\left(\frac{\sum_{t=1}^{n} r_{it}e_t}{\sum_{t=1}^{n} r_{it}^2}\right)\left(\frac{\sum_{s=1}^{n} r_{js}e_s}{\sum_{s=1}^{n} r_{js}^2}\right)\right] \\
&= E\left[\frac{\sum_{t=1}^{n}\sum_{s=1}^{n} r_{it}r_{js}e_t e_s}{\sum_{t=1}^{n} r_{it}^2 \sum_{s=1}^{n} r_{js}^2}\right] = E\left[\frac{\sum_{t=1}^{n} r_{it}r_{jt}e_t^2}{\sum_{t=1}^{n} r_{it}^2 \sum_{t=1}^{n} r_{jt}^2}\right] \\
&= E\left[\frac{r_{i1}r_{j1}e_1^2}{\sum_{t=1}^{n} r_{it}^2 \sum_{t=1}^{n} r_{jt}^2}\right] + \cdots + E\left[\frac{r_{in}r_{jn}e_n^2}{\sum_{t=1}^{n} r_{it}^2 \sum_{t=1}^{n} r_{jt}^2}\right] \\
&= E\left[\frac{r_{i1}r_{j1}}{\sum_{t=1}^{n} r_{it}^2 \sum_{t=1}^{n} r_{jt}^2}\right] E\left[e_1^2\right] + \cdots + E\left[\frac{r_{in}r_{jn}}{\sum_{t=1}^{n} r_{it}^2 \sum_{t=1}^{n} r_{jt}^2}\right] E\left[e_n^2\right] \\
&= E\left[\frac{r_{i1}r_{j1}}{\sum_{t=1}^{n} r_{it}^2 \sum_{t=1}^{n} r_{jt}^2}\right] \sigma_e^2 + \cdots + E\left[\frac{r_{in}r_{jn}}{\sum_{t=1}^{n} r_{it}^2 \sum_{t=1}^{n} r_{jt}^2}\right] \sigma_e^2 \\
&= \sigma_e^2 E\left[\frac{\sum_{t=1}^{n} r_{it}r_{jt}}{\sum_{t=1}^{n} r_{it}^2 \sum_{t=1}^{n} r_{jt}^2}\right],
\end{aligned}
$$

which is formula (5).

As a special case, when $i = j$, formula (5) simplifies to

$$E[B_{j,j}] = E\left[\left(\frac{\sum_{t=1}^{n} r_{jt}e_t}{\sum_{t=1}^{n} r_{jt}^2}\right)^2\right] = \sigma_e^2 E\left[\frac{1}{\sum_{t=1}^{n} r_{jt}^2}\right].$$

It remains to be shown that,

$$\sigma_e^2 E\left[\frac{1}{\sum_{t=1}^{n} r_{jt}^2}\right] = E\left[\frac{\hat{\sigma}_e^2}{\sum_{t=1}^{n} r_{jt}^2}\right]. \tag{B1}$$

Let $H = X(X'X)^{-1}X'$ denote the hat matrix for the regression of y_t on all $x_{j,t-1}$ and v_{jt}^c, $(j = 1, \ldots, p)$, where $X = [1_n, x_{1,t-1}, \ldots, x_{p,t-1}, v_{1t}^c, \ldots, v_{pt}^c]$. Let ϵ denote the residual vector and e the error vector from this regression, so that $\epsilon = (I - H)y = (I - H)e$, where I denotes an $n \times n$ identity matrix. Conditionally on X, we have

$$\sum_{t=1}^{n} \epsilon_t^2 = e'(I - H)e \sim \sigma_e^2 \chi_{n-(2p+1)}^2,$$

and since the random variable on the right-hand side does not depend on X, the result is true unconditionally as well. Thus,

$$\hat{\sigma}_e^2 = \frac{1}{n - (2p + 1)} \sum_{t=1}^{n} \epsilon_t^2$$

is an unbiased estimator of σ_e^2, that is, $E[\hat{\sigma}_e^2] = \sigma_e^2$. Now, we have

$$E\left[\frac{\hat{\sigma}_e^2}{\sum_{t=1}^n r_{jt}^2} \mid X\right] = E\left[\frac{1}{n-(2p+1)} \frac{e'(I-H)e}{\sum_{t=1}^n r_{jt}^2} \mid X\right]$$

$$= \left(\frac{1}{\sum_{t=1}^n r_{jt}^2}\right)\left(\frac{1}{n-(2p+1)}\right) E\left[\sigma_e^2 \chi_{n-(2p+1)}^2\right]$$

$$= \sigma_e^2 \frac{1}{\sum_{t=1}^n r_{jt}^2}.$$

Taking expectations of both sides and using the double expectation theorem yields formula (B1). Thus formula (6) is proved. ■

C. Multipredictor bootstrap procedure

The following describes the procedure employed for the bootstrap method. We illustrate here the bootstrap test of the null hypothesis $\beta_1 = 0$; the tests of $\beta_2 = 0$ and of the joint hypothesis $\beta_1 = \beta_2 = 0$ are analogous.[12]

(1) Given the actual time series $\{y_t, x_{1,t}, x_{2,t}\}$, obtain OLS estimates of $\hat{\alpha}$, $\hat{\beta}$, $\hat{\Theta}$, $\hat{\Phi}$, and t-statistic, $t(\hat{\beta}_1)$, as well as the residuals $\{\hat{u}_t\}$, $\{\hat{v}_{1,t}\}$, $\{\hat{v}_{2,t}\}$.

(2) In each replication, we fix the first observations $x_{1,0}$ and $x_{2,0}$ from the data and construct $\{y_t^\star, x_{1,t}^\star, x_{2,t}^\star\}$ iteratively using the predictive regression model as well as the OLS estimates $\hat{\alpha}$, $\hat{\beta}_2$, $\hat{\Theta}$, and $\hat{\Phi}$, fixing $\beta_1 = 0$. Simulated residuals $(u_t^\star, v_{1,t}^\star, v_{2,t}^\star)$ are drawn from a multivariate normal distribution whose covariance matrix is estimated from the actual residuals $\{\hat{u}_t, \hat{v}_{1,t}, \hat{v}_{2,t}\}$. The bootstrap replication is denoted by $\{y_t^\star, x_{1,t}^\star, x_{2,t}^\star\}_{t=0}^n$.

(3) Using the bootstrap replication $\{y_t^\star, x_{1,t}^\star, x_{2,t}^\star\}$, repeat Step (1) to obtain $t(\hat{\beta}_1^\star)$. Repeat Steps (1) and (2) under the null M times under the null (we use $M = 2500$). From the M values of $t(\hat{\beta}_1^\star)$ thus obtained, we generate the empirical bootstrap distribution of $t(\hat{\beta}_1^\star)$, which are then ranked. The null hypothesis is tested by comparing $t(\hat{\beta}_1)$ to the bootstrap distribution of $t(\hat{\beta}_1^\star)$. For example, in a right-tailed nominal size α test of $H_0 : \beta_1 = 0$ against $H_a : \beta_1 > 0$, the null hypothesis is rejected if $t(\hat{\beta}_1)$ is larger than the $1 - \alpha$ quantile of $t(\hat{\beta}_1^\star)$. We then calculate the size (the proportion of rejecting the null) by replicating the procedure 1500 times.

We also implemented the bootstrapping method using mARM-estimated parameter values, e.g., using $\hat{\beta}^c$ instead of $\hat{\beta}$. In our simulations, we study the performance of both OLS- and mARM-based bootstrap methods.

References

Amihud, Y., and C. M. Hurvich. 2004. Predictive Regression: A Reduced-Bias Estimation Method. *Journal of Financial and Quantitative Analysis* 39:813–41.

Ang, A., and G. Bekaert. 2007. Stock Return Predictability: Is It There? *Review of Financial Studies* 20:651–707.

Baker, M. P., R. Greenwood, and J. Wurgler. 2003. The Maturity of Debt Issues and Predictable Variation in Bond Returns. *Journal of Financial Economics* 70:261–91.

Baker, M. P., and J. Stein. 2004. Market Liquidity as a Sentiment Indicator. *Journal of Financial Markets* 7:1125–65.

Baker, M. P., and J. Wurgler. 2004. A Catering Theory of Dividends. *Journal of Finance* 59:1125–65.

[12] A different bootstrapping method is in Polk, Thompson, and Vuolteenaho (2005).

Boudoukh, J., M. Richardson, and R. Whitelaw. 2006. The Myth of Long-horizon Predictability. *Review of Financial Studies*, Advance Access published on October 25, 2006, 10.1093/rfs/hhn1042.

Campbell, J. Y., M. Lettau, B. G. Malkiel, and Y. Xu. 2001. Have Individual Stocks Become More Volatile? An Empirical Exploration of Idiosyncratic Risk. *Journal of Finance* 56:1–43.

Campbell, J. Y., and M. Yogo. 2005. Efficient Tests of Stock Return Predictability. *Journal of Financial Economics* 81:27–60.

Cavanagh, C. L., G. Elliott, and J. Stock. 1995. Inference in Models with Nearly Integrated Regressors. *Econometric Theory* 11:1131–47.

Fama, E., and K. R. French. 1988. Dividend Yield and Expected Stock Returns. *Journal of Financial Economics* 22:3–25.

Fuller, W. A. 1996. *Introduction to Statistical Time Series*. New York: Wiley.

Goyenko, R. 2005. Stock and Bond Pricing with Liquidity Risk. Working Paper, Indiana University.

Greene, W. H. 2002. *Econometric Analysis*. Upper Saddle River, NJ: Prentice-Hall.

Guo, H., and R. Savickas. 2008. Aggregate Idiosyncratic Volatility in G7 Countries. *Review of Financial Studies* 21:1259–96 .

Hall, P., and R. LePage. 1996. On Bootstrap Estimation of the Studentized Mean. *Annals of the Institute of Statistical Mathematics* 48:403–21.

Hamilton, J. D. 1994. *Time Series Analysis*. Princeton, NJ: Princeton University Press.

Hayashi, F. 2000. *Econometrics*. Princeton, NJ: Princeton University Press.

Keim, D. B., and A. Madhavan. 2000. The Relation Between Stock Market Movement and NYSE Seat Price. *Journal of Finance* 55:2817–40.

Keim, D. B., and R. F. Stambaugh. 1986. Predicting Returns in the Stock and Bond Market. *Journal of Financial Economics* 17:357–96.

Kothari, S. P., and J. Shanken. 1997. Book-to-Market, Dividend Yield, and Expected Market Returns: A Time-Series Analysis. *Journal of Financial Economics* 2:169–203.

Lettau, M., and S. Ludvigson. 2001. Consumption, Aggregate Wealth, and Expected Stock Returns. *Journal of Finance* 3:815–49.

Lewellen, J. 2004. Predicting Returns with Financial Ratios. *Journal of Financial Economics* 74:209–35.

MacKinnon, J., and A. Smith Jr. 1998. Approximate Bias Correction in Econometrics. *Journal of Econometrics* 85:205–30.

Nelson, C. R., and M. J. Kim. 1993. Predictable Stock Returns: The Role of Small Sample Bias. *Journal of Finance* 48:641–61.

Nicholls, D. F., and A. L. Pope. 1988. Bias in the Estimation of Multiple Autoregressions. *Australian Journal of Statistics* 30A:296–309.

Orcutt, G. H., and H. S. Winocur Jr. 1969. First Order Regression: Inference, Estimation and Prediction. *Econometrica* 37:1–14.

Phillips, P. C. B., and M. Loretan. 1991. Estimating Long-Run Economic Equilibria. *Review of Economic Studies* 58:407–36.

Polk, C., S. Thompson, and T. Vuolteenaho. 2006. Cross-Sectional Forecasts of the Equity Premium. *Journal of Financial Economics* 81:101–41.

Pontiff, J., and L. D. Schall. 1998. Book-to-Market Ratios as Predictors of Market Returns. *Journal of Financial Economics* 49:141–60.

Saikkonen, P. 1991. Asymptotically Efficient Estimation of Cointegration Regressions. *Econometric Theory* 7:1–21.

Santos, T., and P. Veronesi. 2006. Labor Income and Predictable Stock Returns. *Review of Financial Studies* 19:1–44.

Stambaugh, R. F. 1999. Predictive Regressions. *Journal of Financial Economics* 54:375–421.

Stock, J. H., and M. W. Watson. 1988. Testing for Common Trends. *Journal of American Statistical Association* 83:1097–1107.

Wooldridge, J. M. 1991. Notes on Regression with Difference-Stationary Data. Working Paper, Michigan State University.