JOURNAL OF FINANCIAL AND QUANTITATIVE ANALYSIS VOL. 39, NO. 4, DECEMBER 2004
COPYRIGHT 2004, SCHOOL OF BUSINESS ADMINISTRATION, UNIVERSITY OF WASHINGTON, SEATTLE, WA 98195

https://doi.org/10.1017/S0022109000003227 Published online by Cambridge University Press

Predictive Regressions: A Reduced-Bias Estimation Method

Yakov Amihud and Clifford M. Hurvich*

Abstract

Standard predictive regressions produce biased coefficient estimates in small samples when the regressors are Gaussian first-order autoregressive with errors that are correlated with the error series of the dependent variable. See Stambaugh (1999) for the single regressor model. This paper proposes a direct and convenient method to obtain reduced-bias estimators for single and *multiple* regressor models by employing an augmented regression, adding a proxy for the errors in the autoregressive model. We derive bias expressions for both the ordinary least-squares and our reduced-bias estimated coefficients. For the standard errors of the estimated predictive coefficients, we develop a heuristic estimator that performs well in simulations, for both the single predictor model and an important specification of the multiple predictor model. The effectiveness of our method is demonstrated by simulations and empirical estimates of common predictive models in finance. Our empirical results show that some of the predictive variables that were significant under ordinary least squares become insignificant under our estimation procedure.

I. Introduction

This paper addresses the problem of estimation of predictive regressions, by which the time series of one variable is regressed on the lagged time series of another variable. Standard estimation procedures of these regressions may lead researchers to erroneously conclude that the next period value of the dependent variable can be forecasted by the current known value of the right-hand side variable. Specifically, Stambaugh (1999) shows that in a model where the predictive variable is an AR(1) process and its residuals are correlated with the predictive regressions' residuals, the ordinary least-squares (OLS) estimator of the predictive variable's coefficient, $\hat{\beta}$, will be biased in finite sample.

Examples of such predictive regressions abound. Keim and Stambaugh (1986) propose to predict the excess returns on seven asset classes by three lagged variables: the difference between the yield on long-term BAA underrated corporate bonds and the short-term (one month) Treasury bill rate, the level of the S&P

*Amihud, yamihud@stern.nyu.edu, Department of Finance, and Hurvich, churvich@stern.nyu.edu, Department of Operations and Management Science, New York University, Stern School of Business, 44 W 4th St, New York, NY 10012. We thank Robert Engle, Gary Simon, Jeffrey Simonoff, and Stephen Brown (the editor), for helpful comments and suggestions, as well as Torben Andersen (the referee) for special contributions and suggestions.

500 index relative to its 45-year moving average, and the level of the small firm stock index. Most of the slope coefficients are statistically significant, suggesting that the expected risk premium on many assets changes over time in a predictable manner. Fama and French (1989) apply a similar methodology, using two predictors based on bond yields, the default premium and the term premium. They find that both have predictive power for stock excess return. Fama (1990) shows that stock excess return is predictable by lagged industrial production, using monthly, quarterly, and annual time series.

Other studies use lagged financial ratios as predictors of stock returns. Fama and French (1988) find that the lagged dividend-price ratio has a significant effect on stock returns, and Campbell and Shiller (1988) find that the lagged dividend-price ratio, together with the lagged dividend growth rate, have a significant predictive power on stock returns. The dividend-price ratio is used as predictor in subsequent studies such as Hodrick (1992). Kothari and Shanken (1997) add to the lagged dividend-price ratio the lagged book-to-market ratio as a predictor, and Lewellen (2004) studies, in addition to these two ratios, the predictive power of the earnings-to-price ratio. According to the testing methodology developed in the latter study, these three ratios have stronger forecasting power than previously thought. Also, Pontiff and Schall (1998) find that the book-to-market ratio predicts stock returns.

Another group of studies of predictive regressions uses volatility and liquidity variables. French, Schwert, and Stambaugh (1987) use return variance, obtained from an ARIMA model, and Amihud (2002) and Jones (2002) use various measures of stock market liquidity. The results in these studies generally show that the lagged time series of these variables significantly predict stock excess returns. Baker and Stein (2002) use lagged equity share in new issues in addition to lagged liquidity and the lagged dividend-price ratio to predict stock returns.

The predictive regressions of the type surveyed above raise a problem of estimation. Many of the predictor variables display strong autoregressive structure and tend to have disturbances correlated with the disturbances from the regression. In such situations, the coefficient of the predictive series will be biased in finite samples, which may also lead to incorrect inference on the coefficient. This problem, pointed out by Mankiw and Shapiro (1986), Stambaugh (1986), and Nelson and Kim (1993) is analyzed by Stambaugh (1999) who develops a bias expression for the estimated prediction coefficient. Researchers usually address this problem by correcting the bias in the least-squares estimator using an estimated version of Stambaugh's (1999) bias expression and by doing bootstrap regressions to calculate the distribution of the estimated coefficient (e.g., Kothari and Shanken (1997)). Lewellen (2004) estimates the slope coefficient and its *t*-statistic under a conservative assumption that the true autoregressive coefficient is very close to one, which gives an upper bound for the bias in $\hat{\beta}$.[1]

[1]In predictive regressions, where stock returns are predicted by a lagged variable that is autoregressive, Ferson et al. (2003) show that data mining for predictor variables interacts with spurious regression bias. However, they assume independence between the errors of the predictive regression and the predictor's autoregression (p. 1399). In terms of the model below, they assume that $\phi = 0$. Therefore, their model is different from the model analyzed here.

https://doi.org/10.1017/S0022109000003227 Published online by Cambridge University Press

https://doi.org/10.1017/S0022109000003227 Published online by Cambridge University Press

We propose a new and convenient method to tackle this problem. We develop a procedure to obtain a reduced-bias estimator of the predictor's coefficient and derive its properties, as well as a convenient, easy-to-use method to obtain reduced-bias standard errors. We demonstrate our method's performance by simulations and application to actual data.

Stambaugh (1999) derives the bias expression for $\hat{\beta}$ in terms of the model's parameters. Subsequent research employs a plug-in version of this bias expression, using sample estimates in lieu of the model parameters, to obtain a bias-corrected estimator of β, which we denote by $\hat{\beta}^s$. However, there is not as yet any theoretical justification for this method of estimation. There is no obvious reason why the sample estimators, which are random variables, should be independent of each other, so it is not clear how to obtain the expected value of the bias correction.

Furthermore, Stambaugh (1999) analyzes a *single* predictor model, while the problem of bias in estimating β also arises in the case of *multiple* predictive variables. For the multiple predictor case, there is no available expression for the bias in the OLS estimator of the predictive regression coefficients, nor is there a direct method of estimation to reduce the bias in this case.

Our method of solving the problem of bias in $\hat{\beta}$ is based on augmented regressions that are applicable both in Stambaugh's (1999) single predictor model and in a multiple predictor generalization. The added variables in the regression are proxies for the error series in a Gaussian AR(1) model for the predictors. The proxies are residual series based on a reduced-bias estimator of the AR parameter. Our proposed estimation method is straightforward and easily implemented.

We also propose a formula to directly estimate the standard error of the bias-corrected estimator of β in the single predictor case, which enables us to easily construct confidence intervals and do hypothesis testing. This formula is also found to work well under one specification of the multiple predictor case. No such direct method to estimate the standard error of the bias-corrected estimator of β is available in the literature; instead, it is done by the bootstrapping method. [2]

In the single predictor case, one specification of our approach is equivalent to $\hat{\beta}^s$, although this equivalence is far from obvious. Thus, our theoretical results yield a formula for the bias in $\hat{\beta}^s$. These same theoretical results justify the use of a different version of our approach, which has a smaller bias than $\hat{\beta}^s$.

For the case of multiple predictive variables, we consider a general model in which the predictive variables form a Gaussian multivariate AR(1) series, that is, a Gaussian VAR(1). Our analysis is based on a natural generalization of our univariate reduced-bias estimation method, employing a regression that is augmented by the estimated error series in the multivariate AR(1) model. We derive a general expression for the bias of our proposed reduced-bias estimators of β (in this case, a vector) and show that as in the univariate case, this bias is proportional to the bias in the corresponding estimator of the AR(1) parameter matrix. The importance of this result is in showing that any existing or future methodology that can reduce the bias in estimation of the AR(1) parameter matrix can be used to produce corresponding improvements in the bias of the coefficients of the predic-

[2] See Kothari and Shanken (1997) and Baker and Stein (2002).

tive variables. We also provide a theoretical expression for the bias in the OLS estimator of β, generalizing Stambaugh's bias formula to the multiple predictor case.

We demonstrate the usefulness of our estimation method by presenting simulations for both the single and the multiple predictor cases. In implementing our estimators in the case of multiple predictive variables, we first focus on a special case of our general model in which the AR(1) parameter matrix is known to be diagonal, so that each predictive variable follows a univariate AR(1) model. The predictive variables can still be correlated under this restriction through the covariance matrix of the errors. Our method can be implemented in a direct and simple way, and it performs quite well. Finally, we consider the general case where the AR(1) parameter matrix is not constrained to be diagonal. We construct an estimate of a bias expression for multivariate AR(1) models, due to Nicholls and Pope (1988), which indeed reduces the bias, but since the expressions are more complex and more parameters need to be estimated, there is some degradation in performance compared to the diagonal case.

Finally, we present examples of the application of our estimation method to some common predictive models in finance, by which stock returns are predicted by lagged dividend yield and two lagged financial ratios: book/market and earnings/price. For the monthly predictive model over the period May 1963–December 1994, dividend yield turns from a highly significant predictor—$\beta = 3.05$ and $t = 3.02$—to $\beta = 2.08$ and $t = 2.05$. The other two ratios have positive and highly significant predictive coefficients in OLS, but turn statistically insignificant after applying our estimation method. In an illustration of a multiple predictor regression with both dividend yield and earnings/price ratios as predictors, the former is significant while the latter is not. When extending the estimation period, none of the predictors is significant.

Our paper proceeds as follows. In Section II, we show the basic single predictor model, following Stambaugh (1999), outline our proposal to estimate the predictive regression coefficient, and present the theoretical properties of the reduced-bias estimator. Section III describes a heuristic method for estimating the standard error of the estimated predictive regression coefficient. Section IV presents the multiple predictor model, proposes an augmented regression estimator of the coefficients of the predictive variables, and considers the properties of this estimator. We present simulation results on our method in Section V. Section VI demonstrates the use of our method in empirical estimations of common predictive models in finance. Our conclusions are in Section VII. Proofs of the theoretical results are presented in the Appendix.

II. Reduced-Bias Estimation of the Regression Coefficient

We follow the formal model in Stambaugh (1999), where a scalar time series $\{y_t\}_{t=1}^{n}$ is to be predicted from a scalar first-order autoregressive (AR(1)) time series $\{x_t\}_{t=0}^{n-1}$. The overall model for $t = 1, \ldots, n$ is

$$y_t = \alpha + \beta x_{t-1} + u_t, \tag{1}$$

$$x_t = \theta + \rho x_{t-1} + v_t, \tag{2}$$

https://doi.org/10.1017/S0022109000003227 Published online by Cambridge University Press

where the errors (u_t, v_t) are serially independent and identically distributed as bivariate normal, with contemporaneous correlation, that is,

$$\begin{pmatrix} u_t \\ v_t \end{pmatrix} \sim_{\text{iid}} N(0, \Sigma), \quad \Sigma = \begin{pmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{pmatrix},$$

and the lag-1 autocorrelation ρ of $\{x_t\}$ satisfies the constraint $|\rho| < 1$, thereby ensuring that $\{x_t\}$ is stationary. (The initial value x_0 can be taken to be random or non-random.) Then, Stambaugh (1999) shows that if $\sigma_{uv} \neq 0$, the OLS estimator of β based on a finite sample will be biased.

For the single predictor model given by (1) and (2), Stambaugh (1999) provides an expression for the bias of the OLS estimator of β,

$$E[\hat{\beta} - \beta] = \phi E[\hat{\rho} - \rho], \tag{3}$$

where $\phi = \sigma_{uv}/\sigma_v^2$, and $\hat{\beta}$ and $\hat{\rho}$ are the OLS estimators of β and ρ. This expression is exact, for any given sample size n. Subsequent research employs a plug-in version of this expression by using sample estimators of the two parameters, ϕ and ρ. Specifically, Stambaugh notes, following Kendall (1954), that $E[\hat{\rho} - \rho] = -(1 + 3\rho)/n + O(n^{-2})$. This result is often applied by researchers to obtain

$$\hat{\beta}^s = \hat{\beta} + \hat{\phi}^s(1 + 3\hat{\rho})/n, \tag{4}$$

where $\hat{\phi}^s = \sum \hat{u}_t \hat{v}_t / \sum \hat{v}_t^2$, and $\hat{u}_t$, $\hat{v}_t$ are the residuals from OLS regressions in (1) and (2), respectively.[3]

To motivate our proposed reduced-bias estimator of β, we consider first an infeasible estimator $\tilde{\beta}$,[4] which is the coefficient of x_{t-1} in an OLS regression (with intercept) of y_t on x_{t-1} and v_t, for $t = 1, \ldots, n$, where $v_t = x_t - \rho x_{t-1}$. The Appendix shows that we can write

$$y_t = \alpha + \beta x_{t-1} + \phi v_t + e_t, \tag{5}$$

where $\{e_t\}_{t=1}^n$ are i.i.d. normal random variables with mean zero, and $\{e_t\}$ is independent of both $\{v_t\}$ and $\{x_t\}$. The estimator $\tilde{\beta}$ is exactly unbiased, as stated in the following theorem.

Theorem 1. The infeasible estimator $\tilde{\beta}$, is exactly unbiased, $E[\tilde{\beta}] = \beta$.

Proof. See the Appendix.

In practice, the errors $\{v_t\}$ are unobservable. But the result above suggests that it may be worthwhile to construct a proxy $\{v_t^c\}$ for the errors, on the basis of the available data $\{x_t\}_{t=0}^n$. Define a feasible bias-corrected estimator[5] $\hat{\beta}^c$ to be the coefficient of x_{t-1} in an OLS regression of y_t on x_{t-1} and v_t^c, with intercept.

The proxy v_t^c takes the form,

$$v_t^c = x_t - \left(\hat{\theta}^c + \hat{\rho}^c x_{t-1}\right), \tag{6}$$

[3] Kothari and Shanken (1997) define $\hat{\beta}^{KS} = \hat{\beta} + \hat{\phi}^s(1 + 3p_A)/n$, where $p_A = (n\hat{\rho} + 1)/(n - 3)$.

[4] An infeasible estimator is not directly computable based on data alone.

[5] A feasible estimator is directly computable from the data alone.

https://doi.org/10.1017/S0022109000003227 Published online by Cambridge University Press

where $\hat{\theta}^c$ and $\hat{\rho}^c$ are any estimators of θ and ρ constructed on the basis of $x_0, x_1, \ldots, x_n$. As we will show, the particular choice of the estimator $\hat{\theta}^c$ has no effect on the bias of $\hat{\beta}^c$. On the other hand, the estimator $\hat{\rho}^c$ should be selected to be as nearly unbiased as possible for ρ, as the bias of $\hat{\beta}^c$ is proportional to the bias of $\hat{\rho}^c$. We have the following theorem, which, like Theorem 1, holds exactly for all values of n.

Theorem 2. The bias of the feasible estimator $\hat{\beta}^c$ is given by $E[\hat{\beta}^c - \beta] = \phi E[\hat{\rho}^c - \rho]$, where $\phi = \sigma_{uv}/\sigma_v^2$.

Proof. See the Appendix.[6]

As a corollary of Theorem 2, setting $\hat{\rho}^c = \hat{\rho}$ (which yields $\hat{\beta}^c = \hat{\beta}$; see the proof of Theorem 3 below), we obtain Stambaugh's (1999) bias expression (3).

Next, we exhibit the relationship between our proposed estimator $\hat{\beta}^c$ and the estimator $\hat{\beta}^s$ motivated by Stambaugh (1999). Suppose the bias-corrected version of $\hat{\rho}$ takes the form $\hat{\rho}^c = \hat{\rho} + \omega$, where ω may depend on the available data $x_0, x_1, \ldots, x_n$. Then both our approach and the Stambaugh approach to correcting the bias in $\hat{\beta}$ yield the same result.

Theorem 3. If $\hat{\rho}^c = \hat{\rho} + \omega$, then $\hat{\beta}^c = \hat{\beta} + \hat{\phi}^s \omega$.

Proof. See the Appendix.

In particular, if we take $\omega = (1+3\hat{\rho})/n$, then Theorem 3 implies that $\hat{\beta}^c = \hat{\beta}^s$, where $\hat{\beta}^s$ is given by (4). Furthermore, it follows from the proof of Theorem 3 that $\hat{\phi}^s = \hat{\phi}^c$, for all values of ω.

There is a large literature on reduced-bias estimation of the lag-1 autocorrelation parameter ρ of AR(1) models, and in view of Theorem 2, this literature is of direct relevance to the construction of reduced-bias estimators of β. Some easily computable and low-bias choices of $\hat{\rho}^c$ include the Burg estimator (see Fuller (1996), p. 418), the weighted symmetric estimator (see Fuller (1996), p. 414), and the tapered Yule-Walker estimator (see Dahlhaus (1988)). Both the Burg estimator and the tapered Yule-Walker estimator have the additional advantage that they are guaranteed to be strictly between -1 and 1.

In this paper, we focus on two estimators based on Kendall's (1954) expression for the bias of the OLS estimator, $\hat{\rho}$, that is, $E[\hat{\rho} - \rho] = -(1+3\rho)/n + O(n^{-2})$. This leads to a first-order bias-corrected estimator $\hat{\rho}^{c,1} = \hat{\rho} + (1 + 3\hat{\rho})/n$ and a second-order bias-corrected estimator,

$$\hat{\rho}^{c,2} = \hat{\rho} + (1 + 3\hat{\rho})/n + 3(1 + 3\hat{\rho})/n^2. \tag{7}$$

The estimator $\hat{\rho}^{c,1}$ was studied by Sawa (1978), and has bias, which is $O(n^{-2})$. The estimator $\hat{\rho}^{c,2}$ is obtained by an iterative correction, $\hat{\rho}^{c,2} = \hat{\rho} + (1 + 3\hat{\rho}^{c,1})/n$. The bias of $\hat{\rho}^{c,2}$ is $O(n^{-2})$ as well, but our simulations indicate that the bias of $\hat{\rho}^{c,2}$ is smaller than that of $\hat{\rho}^{c,1}$. We will therefore restrict attention to $\hat{\rho}^{c,2}$, which we denote by $\hat{\rho}^c$, with the corresponding bias-corrected estimator of β denoted by $\hat{\beta}^c$.

In summary, the procedure we propose for estimating β has two steps:

[6] We thank the referee for suggesting a method that greatly facilitated the proof.

https://doi.org/10.1017/S0022109000003227 Published online by Cambridge University Press

i) Estimate model (2) by OLS and obtain $\hat{\rho}$. Construct the corrected estimator $\hat{\rho}^c = \hat{\rho} + (1 + 3\hat{\rho})/n + 3(1 + 3\hat{\rho})/n^2$ and obtain the corrected residuals v_t^c as in (6) above.

ii) Obtain $\hat{\beta}^c$ as the coefficient of x_{t-1} in an OLS regression of y_t on x_{t-1} and v_t^c, with intercept. This regression also produces $\hat{\phi}^c$ as the estimator of the coefficient of v_t^c.

The coefficient $\hat{\phi}^c$ is an unbiased estimator of ϕ, as stated in the following lemma.

Lemma 1. $E[\hat{\phi}^c] = \phi$.

Proof. See the Appendix.

The unbiasedness of $\hat{\phi}^c$ may be useful in the following context. When a variable x_t is generated by an AR(1) process as in (2), the anticipated component of x_t based on past values of the series is $E(x_t|x_{t-1}) = \theta + \rho x_{t-1}$. Then, the error v_t is the unanticipated component of x_t. A researcher may want to estimate separately the effects of the anticipated and unanticipated components of x_t on y_t in model (5). For example, French, Schwert, and Stambaugh (1987) study the effects of expected and unexpected market volatility on stock market returns, where expected volatility is the fitted values of volatility obtained from an ARIMA model and unexpected volatility is the residual series from this model. Amihud (2002) studies the effect of expected and unexpected stock market illiquidity on stock returns, using lagged illiquidity as expected illiquidity, where illiquidity is assumed to be an AR(1) process. In such a case, the coefficient β measures[7] the effect of the anticipated component of x_t while the coefficient ϕ measures the effect of the unexpected component of x_t on y_t.

III. Estimation of Standard Errors

A. Standard Errors for $\hat{\beta}^c$

For hypothesis tests for β, we need a low-bias finite sample approximation to the standard error of $\hat{\beta}^c$ (Simonoff (1993)). The commonly used estimated standard error $\widehat{SE}(\hat{\beta}^c)$, obtained from the OLS output in a regression of y_t on x_{t-1} and v_t^c, cannot be used for testing hypotheses about β since it is a downward-biased estimator of the standard error of $\hat{\beta}^c$. The reason for this bias is that the OLS standard error fails to take into account the additional variability due to the estimation of ρ. If ρ were a known constant, we would have that the standard error of $\hat{\beta}^c$ equals $\widehat{SE}(\hat{\beta}^c)$. This is a consequence of the following lemma.

Lemma 2.

$$E\left[\hat{\beta}^c - \beta\right]^2 = \phi^2 E\left[\hat{\rho}^c - \rho\right]^2 + E\left[\widehat{SE}\left(\hat{\beta}^c\right)\right]^2, \tag{8}$$

where $\widehat{SE}(\hat{\beta}^c)$ is the estimated standard error for $\hat{\beta}^c$, based on an OLS regression of y_t on x_{t-1} and v_t^c, with intercept (provided by standard regression packages).

[7] Suppose that the model to be estimated is $y_t = \delta_0 + \delta_1 x_t^a + \phi x_t^u + e_t$, where x_t^a and x_t^u are, respectively, the anticipated and unanticipated components of x_t. Then, $\alpha = \delta_0 + \delta_1\theta$ and $\beta = \delta_1\rho$.

https://doi.org/10.1017/S0022109000003227 Published online by Cambridge University Press

Proof. See the Appendix.

Since from Theorem 2, $E[\hat{\beta}^c - \beta] = \phi E[\hat{\rho}^c - \rho] = O(1/n^2)$, we conclude from (8) that

$$\text{Var}\left[\hat{\beta}^c\right] \quad = \quad E\left[\hat{\beta}^c - \beta\right]^2 + O(1/n^4), \tag{9}$$

so a low-bias estimate of the right-hand side of (8) should provide a low-bias estimate of $\text{Var}[\hat{\beta}^c]$. Clearly, $[\widehat{SE}(\hat{\beta}^c)]^2$ provides an unbiased estimator of $E[\widehat{SE}(\hat{\beta}^c)]^2$. We now need to accurately estimate $\phi^2 E[\hat{\rho}^c - \rho]^2$. First, we note that the coefficient $\hat{\phi}^c$ of v_t^c in the OLS regression of y_t on x_{t-1} and v_t^c is unbiased (see Lemma 1).

Next, we need to construct an estimator of $E[\hat{\rho}^c - \rho]^2$ with low bias. Here we use some heuristic approximations, which turn out to work quite well in simulations. Noting that $\hat{\rho}^c$ is a low-bias estimator of ρ, we treat $\hat{\rho}^c$ as if it were unbiased. Then we simply need an expression for $\text{Var}(\hat{\rho}^c)$, where

$$\hat{\rho}^c \quad = \quad \hat{\rho} + \frac{1+3\hat{\rho}}{n} + 3\frac{1+3\hat{\rho}}{n^2} \quad = \quad \frac{1}{n} + \frac{3}{n^2} + (1 + 3/n + 9/n^2)\hat{\rho}.$$

Thus, $\text{Var}(\hat{\rho}^c) = (1+3/n+9/n^2)^2 \text{Var}(\hat{\rho})$. For the OLS estimator $\hat{\rho}$, our simulations indicate that an accurate estimator of $\text{Var}(\hat{\rho})$ is given by $\widehat{\text{Var}}(\hat{\rho})$, the squared standard error of $\hat{\rho}$ (as given by standard regression packages) in an OLS regression of $\{x_t\}_{t=1}^n$ on $\{x_{t-1}\}_{t=1}^n$, with intercept.[8] Thus, a feasible estimator for $\text{Var}(\hat{\rho}^c)$ is given by $\widehat{\text{Var}}(\hat{\rho}^c) = (1+3/n+9/n^2)^2 \widehat{\text{Var}}(\hat{\rho})$. Finally, our estimator for the standard error of $\hat{\beta}^c$ is given by

$$\widehat{SE}^c\left(\hat{\beta}^c\right) \quad = \quad \sqrt{\left\{\hat{\phi}^c\right\}^2 \widehat{\text{Var}}(\hat{\rho}^c) + \left\{\widehat{SE}\left(\hat{\beta}^c\right)\right\}^2}. \tag{10}$$

B. Standard Errors for $\hat{\phi}^c$

Let $\hat{\phi}^c$ be the coefficient of v_t^c in an OLS regression of y_t on x_{t-1} and v_t^c. Lemma 1 shows that $E[\hat{\phi}^c] = \phi$. The following lemma shows that the estimated squared standard error of $\hat{\phi}^c$ is also unbiased.

Lemma 3. $\text{Var}[\hat{\phi}^c] = E[\widehat{SE}(\hat{\phi}^c)]^2$, where $\widehat{SE}(\hat{\phi}^c)$ is the estimated standard error for $\hat{\phi}^c$ as provided by standard regression packages, based on an OLS regression of y_t on x_{t-1} and v_t^c, with intercept.

Proof. See the Appendix.

IV. Predictive Regressions with Multiple Predictors

We consider here a more general model for predictive regressions with several predictor variables, and develop a reduced-bias estimator of the predictive

[8] We find in simulations that this readily available estimator strongly outperforms the asymptotic approximation suggested in Fuller (1996), equation (6.2.9).

https://doi.org/10.1017/S0022109000003227 Published online by Cambridge University Press

regression coefficients in this model. No direct methodology is currently available in this case for either evaluating or reducing the bias in the OLS estimators of the predictive regression coefficients.

We assume that the predictor variables are collected in a p-dimensional vector time series $\{x_t\}$ that evolves according to a stationary Gaussian vector autoregressive VAR(1) model. The overall model is given for $t = 1, \ldots, n$ by

$$y_t = \alpha + \beta' x_{t-1} + u_t, \tag{11}$$

$$x_t = \Theta + \Phi x_{t-1} + v_t, \tag{12}$$

where $\{y_t\}$ is a scalar response variable, α is a scalar intercept, β is a $p \times 1$ vector of regression coefficients, $\{u_t\}$ is a scalar noise term, $\{x_t\}$ is a $p \times 1$ series of predictor variables, Θ is a $p \times 1$ intercept, $\{v_t\}$ is a $p \times 1$ series of shocks such that the vectors $(u_t, v_t')'$ are i.i.d. multivariate normal with mean zero, and Φ is a $p \times p$ matrix such that the absolute values of all its eigenvalues are less than one, to ensure stationarity (see, e.g., Fuller (1996)). It follows from our assumptions that there exists a $p \times 1$ vector ϕ such that

$$u_t = \phi' v_t + e_t, \tag{13}$$

where $\{e_t\}$ are i.i.d. normal random variables with mean zero, and $\{e_t\}$ is independent of both $\{v_t\}$ and $\{x_t\}$. Using (11) and (13) we can write

$$y_t = \alpha + \beta' x_{t-1} + \phi' v_t + e_t. \tag{14}$$

As before, we start by pointing out the unbiasedness of an infeasible estimator $\widetilde{\beta}$, which is the $p \times 1$ vector of coefficients of x_{t-1} in an OLS regression (with intercept) of y_t on x_{t-1} and v_t for $t = 1, \ldots, n$.

Theorem 4.

$$E\left[\tilde{\beta}\right] = \beta. \tag{15}$$

Proof. See the Appendix.

Next, we develop a class of reduced-bias estimators of β that is based on augmented regressions, where the additional regressors are proxies for the entries of v_t corresponding to an estimate of Φ. Thus, our single predictor methodology generalizes in a very natural way to the setting of multiple predictors. Then, we develop a bias expression for our estimator of β and show that its bias is proportional to the bias of a corresponding estimator of Φ. Thus, bias reduction in estimating β can be achieved through the use of any reduced-bias estimator of Φ, e.g., the one due to Nicholls and Pope (1988), suggested by Stambaugh (1999).

A. The Estimators $\hat{\beta}^c$ and $\hat{\phi}^c$ and Their Properties

Our proposed estimator $\hat{\beta}^c$ of β consists of the estimated coefficients of the vector x_{t-1} in an augmented OLS regression of y_t on all entries of the vectors x_{t-1} and v_t^c, together with a constant. Here, $\{v_t^c\}$ is a proxy for the error series $\{v_t\}$,

$$v_t^c = x_t - \left(\hat{\Theta}^c + \hat{\Phi}^c x_{t-1}\right), \quad t = 1, \ldots, n, \tag{16}$$

https://doi.org/10.1017/S0022109000003227 Published online by Cambridge University Press

where $\hat{\Theta}^c$ and $\hat{\Phi}^c$ are any estimators of Θ and Φ constructed from $\{x_t\}_{t=0}^n$. The following theorem, which is a direct generalization of Theorem 2, shows that the bias in $\hat{\beta}^c$ is proportional to the bias in $\hat{\Phi}^c$, with proportionality constant ϕ.

Theorem 5.

$$E\left[\hat{\beta}^c - \beta\right] = E\left[\hat{\Phi}^c - \Phi\right]' \phi. \tag{17}$$

Proof. See the Appendix.

As a corollary of Theorem 5, setting $\hat{\Phi}^c = \hat{\Phi}$ (which yields $\hat{\beta}^c = \hat{\beta}$), we obtain

$$E\left[\hat{\beta} - \beta\right] = E\left[\hat{\Phi} - \Phi\right]' \phi, \tag{18}$$

where $\hat{\Phi}$ is the OLS estimator of Φ, thereby generalizing Stambaugh's (1999) single predictor bias expression (3) to the multiple predictor case.

The OLS estimators $\hat{\alpha}$ and $\hat{\beta}$ are given by

$$\begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} = (X'X)^{-1}X'y,$$

and the OLS estimators $\hat{\Theta}$ and $\hat{\Phi}$ are given by a $(p+1) \times p$ matrix

$$\begin{pmatrix} \hat{\Theta}' \\ \hat{\Phi}' \end{pmatrix} = (X'X)^{-1}X'x,$$

where $X = [1_n, (x_0, x_1, \ldots, x_{n-1})']$ is an $n \times (p+1)$ matrix of explanatory variables, 1_n is an $n \times 1$ vector of ones, $y = (y_1, \ldots, y_n)'$ is $n \times 1$, and $x = (x_1, \ldots, x_n)'$ is $n \times p$.

The proposed bias-corrected estimator is given by the entries of $\hat{\beta}^c$ in the formula

$$\begin{pmatrix} \hat{\alpha}^c \\ \hat{\beta}^c \\ \hat{\phi}^c \end{pmatrix} = \left(\tilde{X}_c'\tilde{X}_c\right)^{-1}\tilde{X}_c'y, \tag{19}$$

where $\tilde{X}_c = [1_n, (x_0, x_1, \ldots, x_{n-1})', (v_1^c, \ldots, v_n^c)']$ is $1 \times (2p+1)$ and the error series $\{v_t^c\}_{t=1}^n$ are given by (16). The tth row of $\tilde{X}_c$ consists of a 1, the values of all predictive variables at time $t-1$, and the values of all proxies for the autoregressive errors at time t.

Next, we extend the approach motivated by Stambaugh (1999) for bias correction of $\hat{\beta}$ in a single predictor model to the multiple predictor case, and examine its relationship to our proposed estimator, $\hat{\beta}^c$. Denote by $\hat{\phi}^s$ the final p entries in the vector

$$\begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \\ \hat{\phi}^s \end{pmatrix} = \left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'y, \tag{20}$$

where $\tilde{X} = [1_n, (x_0, x_1, \ldots, x_{n-1})', (\hat{v}_1, \ldots, \hat{v}_n)']$ is $1 \times (2p+1)$, and $\hat{v}_t = x_t - (\hat{\Theta} + \hat{\Phi}x_{t-1})$ for $t = 1, \ldots, n$. Suppose now that the bias-corrected version of Φ takes

https://doi.org/10.1017/S0022109000003227 Published online by Cambridge University Press

the form $\hat{\Phi}^c = \hat{\Phi} + \omega$, where the $p \times p$ matrix ω may depend on the available data $x_0, \ldots, x_n$. Consideration of (18) motivates a natural bias-corrected version of $\hat{\beta}$ given by $\hat{\beta} + \omega' \hat{\phi}^s$. This yields the same result as our $\hat{\beta}^c$.

Theorem 6. If $\hat{\Phi}^c = \hat{\Phi} + \omega$, then $\hat{\beta}^c = \hat{\beta} + \omega' \hat{\phi}^s$.

Proof. See the Appendix.

Theorem 6 generalizes Theorem 3 to the multiple predictor case. Again, we also obtain here that $\hat{\phi}^s = \hat{\phi}^c$, for all values of ω. We have the following generalization of Lemma 1, which shows that the $p \times 1$ vector $\hat{\phi}^c$ is unbiased for ϕ.

Lemma 4. If $\{y_t\}$ is given by the multiple predictor model (11) and (12) and $\hat{\phi}^c$ is as defined above, then

$$E\left[\hat{\phi}^c\right] = \phi. \tag{21}$$

Proof. See the Appendix.

B. The Reduced-Bias Estimator $\hat{\Phi}^c$

To give a specific form for our proposed estimator $\hat{\beta}^c$ in the case of multiple predictive variables, we need to construct a reduced-bias estimator $\hat{\Phi}^c$. The theory of this section on the estimator $\hat{\beta}^c$ holds for any estimator $\hat{\Phi}^c$ that is a function of the series of predictor variables $\{x_t\}_{t=0}^n$. Since the bias of $\hat{\beta}^c$ is proportional to the bias of $\hat{\Phi}^c$, as Theorem 5 shows, we now focus on the choice of $\hat{\Phi}^c$. We propose two forms for $\hat{\Phi}^c$. The first applies only in the case where it is known that the true AR(1) parameter matrix Φ is diagonal, while the second is applicable in general. The first performs much better than the second when Φ is in fact diagonal. Although the assumption that Φ is diagonal entails a considerable loss of generality, it should be noted that if the individual entries of $\{x_t\}$ are given by univariate AR(1) models, as would often be assumed in practice, then Φ must be diagonal. Notably, the series $\{x_{1,t}\}, \{x_{2,t}\}, \ldots, \{x_{p,t}\}$ can still be contemporaneously correlated even under the assumption that Φ is diagonal if the error vectors $\{v_{1,t}\}, \{v_{2,t}\}, \ldots, \{v_{p,t}\}$ are contemporaneously correlated, i.e., the covariance matrix $\Sigma_v = \mathrm{Cov}(v_t)$ is non-diagonal.

If Φ is known to be *diagonal*, then each entry of $\{x_t\}$ is a univariate AR(1) process, and the estimation procedure is as follows. We can treat each series $\{x_{i,t}\}$ $(i = 1, 2, \ldots, p)$ separately: estimate its autoregressive coefficient ρ_i by univariate OLS and then correct this estimator as we have proposed for the single predictor case to produce $\hat{\rho}_i^c$. We then construct $\hat{\Phi}^c$ as a diagonal matrix, with diagonal entries being the corrected univariate AR(1) parameter estimates. The error proxies $\{v_{i,t}^c\}$ $(i = 1, \ldots, p)$ are then constructed for each predictor series $\{x_{i,t}\}$ as in the univariate case, $v_{i,t}^c = x_{i,t} - \hat{\theta}_i^c - \hat{\rho}_i^c x_{i,t-1}$, where $\hat{\theta}_i^c$ is the adjusted intercept. Finally, our corrected estimated coefficient vector $\hat{\beta}^c$ is obtained by an OLS regression of y_t on all predictors $\{x_{1,t-1}\}, \{x_{2,t-1}\}, \ldots, \{x_{p,t-1}\}$ and on the corrected error proxies $\{v_{1,t}^c\}, \{v_{2,t}^c\}, \ldots, \{v_{p,t}^c\}$, with intercept. The coefficient of each predictor

https://doi.org/10.1017/S0022109000003227 Published online by Cambridge University Press

series $\{x_{i,t-1}\}$ is $\hat{\beta}_i^c$ and the coefficient of each error proxy vector $\{v_{i,t}^c\}$ is $\hat{\phi}_i^c$. The simulations in the following section indicate that the corresponding reduced-bias estimator $\hat{\beta}^c$ performs quite well compared to the OLS estimator $\hat{\beta}$.

For the general case where Φ may be *non-diagonal*, reduced-bias estimation of Φ is a more difficult problem. We follow the suggestion of Stambaugh (1999) to estimate Φ using the expression of Nicholls and Pope (1988), Theorem 2, for the bias in the OLS estimator $\hat{\Phi}$, $E[\hat{\Phi} - \Phi] = -b/n + O(n^{-3/2})$, where

$$(22) \quad b = \Sigma_v \left[(I - \Phi')^{-1} + \Phi'(I - \Phi'^2)^{-1} + \sum_{\lambda \in \mathrm{Spec}(\Phi')} \lambda (I - \lambda \Phi')^{-1} \right] \Sigma_x^{-1},$$

I is a $p \times p$ identity matrix, $\Sigma_x = \mathrm{Cov}(x_t)$, the symbol λ denotes an eigenvalue of Φ', and the notation $\lambda \in \mathrm{Spec}(\Phi')$ indicates that the sum is to be taken over all p eigenvalues of Φ' with each term repeated as many times as the multiplicity of λ.[9] The $p \times p$ matrix Σ_x may be evaluated using the convenient formula given by Stambaugh ((1999), Eq. (49)), $\mathrm{vec}(\Sigma_x) = [I_{p^2} - (\Phi \otimes \Phi)]^{-1}\mathrm{vec}(\Sigma_v)$, where I_{p^2} is a $p^2 \times p^2$ identity matrix, the vec operator stacks the columns of a matrix into a single column, and $\otimes$ is the Kronecker product.

Expression (22) for the bias in $\hat{\Phi}$ depends on the unknown Φ and Σ_v. We therefore estimate this bias expression iteratively by repeatedly plugging in preliminary estimates of Φ and Σ_v. The bias-corrected estimator $\hat{\Phi}^{c,i}$ at each iteration may then be obtained by subtracting the estimated bias expression from the OLS estimator $\hat{\Phi}$.

The preliminary estimator of Σ_v is obtained first as the sample covariance matrix of the residuals $x_t - \hat{\Theta} - \hat{\Phi}x_{t-1}$, where $\hat{\Theta}$ and $\hat{\Phi}$ are the OLS estimators of Θ and Φ, respectively. It is important that the preliminary estimator of Φ have all eigenvalues less than 1 in absolute value, i.e., that it correspond to a stationary multivariate AR(1) model, since otherwise the bias formula (22) will yield meaningless results. Therefore, we consider the Yule-Walker estimator (see Reinsel (1997), p. 99), which is guaranteed to satisfy the above mentioned condition. The Yule-Walker estimator is given by

$$(23) \quad \hat{\Phi}^{YW} = \left[\sum_{t=1}^{n} (x_t - \overline{x}^*)(x_{t-1} - \overline{x}^*)' \right] \left[\sum_{t=0}^{n} (x_t - \overline{x}^*)(x_t - \overline{x}^*)' \right]^{-1},$$

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

We start our iterative procedure of computing the corrected estimator $\hat{\Phi}^c$ by checking whether the OLS estimator $\hat{\Phi}$ corresponds to a stationary model. If so, it is used as the preliminary estimator and plugged into the bias expression (22) together with the preliminary estimator of Σ_v described above. If the model corresponding to $\hat{\Phi}$ is nonstationary, then $\hat{\Phi}^{YW}$ is used as the preliminary estimator and plugged into (22) together with the preliminary estimator of Σ_v described above. In either case, this yields a first-stage bias-corrected estimator $\hat{\Phi}^{c,1}$ of Φ,

[9] An expression corresponding to (22) can also be found in Stambaugh (1999), Equation (54). There are two typographical errors there: the expression should be multiplied by $-1/n$, and Φ should be replaced by Φ'.

https://doi.org/10.1017/S0022109000003227 Published online by Cambridge University Press

given by the difference between $\hat{\Phi}$ and the estimated bias. If $\hat{\Phi}^{c,1}$ corresponds to a nonstationary model, then we set $\hat{\Phi}^c = \hat{\Phi}^{c,1}$, and the iterative procedure terminates. Otherwise, we proceed to the next stage of the iteration. At the ith stage, for $i > 1$, we reestimate Σ_v as the sample covariance of the residual series $x_t - \hat{\Theta} - \hat{\Phi}^{c,i-1}x_{t-1}$, where $\hat{\Theta}$ is the OLS estimator of Θ. We then plug this estimator of Σ_v, together with $\hat{\Phi}^{c,i-1}$ into the bias expression (22), yielding an estimated bias $-\hat{b}^{i-1}/n$, and then construct $\hat{\Phi}^{c,i} = \hat{\Phi} - (-\hat{b}^{i-1}/n)$. If $\hat{\Phi}^{c,i}$ corresponds to a nonstationary model, then we set $\hat{\Phi}^c = \hat{\Phi}^{c,i}$, and the iterative procedure terminates. Otherwise, we proceed to the next stage of the iteration. In the simulations below, we used a total of 10 iterations.

For the case where Φ may be non-diagonal, our augmented regression procedure works, in summary, as follows. Given the matrix of predictor variables, we construct the bias-corrected AR(1) parameter matrix estimate $\hat{\Phi}^c$ using our iterative procedure based on Nicholls and Pope's (1988) bias expression, as outlined above. Next, we construct the corrected residual series v_t^c using (16). Then, we estimate $\hat{\beta}_i^c$ as the coefficients of $x_{i,t-1}$ $(i = 1, 2, \ldots, p)$ in an OLS regression of y_t on all $x_{i,t-1}$ and $v_{i,t}^c$ with intercept. This regression also produces $\hat{\phi}_i^c$ as the estimators of the coefficients of $v_{i,t}^c$. We present the results of this estimation procedure in the following section.

The matrix $\hat{\Phi}$ can be used to test whether Φ is diagonal. A standard way to estimate and test the statistical significance of the coefficients is by applying the SUR method.[10] Although we recognize the bias in the entries of $\hat{\Phi}$, this seems to be a reasonable diagnostic check on whether we can proceed with the restricted version of our estimation method in which the Φ matrix is assumed to be diagonal. If the off-diagonal terms are significantly different from zero, we proceed by constructing $\hat{\Phi}^c$ using the non-diagonal method. The cost of this method is that its estimated corrected autoregressive parameters are more biased than those obtained in the diagonal method. So if the true Φ matrix is diagonal, applying the non-diagonal method would produce inferior results. In the empirical case below, we indeed find that two variables that are commonly used as predictors of stock returns have essentially a diagonal $\hat{\Phi}$ matrix.

V. Simulations

A. Single Predictor Model

We report on the performance of our proposed estimators in a simulation study. First, we study the case of a single predictor model, using 1,500 simulated replications from the model (1) and (2). We perform two simulations, using parameter values that are estimated from a model where stock market returns are predicted by lagged dividend yields. That is, we use the values of the estimated parameters $\hat{\rho}^c$ and $\hat{\beta}^c$ as if they were the true parameter values. The first simulation uses parameter estimates from the annual predictive regression for the

[10] SUR is the seemingly unrelated regressions method. See, e.g., Judge et al. ((1985), Section 11.2). The OLS estimation yields identical results in this case, when the explanatory variables of all equations are identical (Judge et al., p. 448). The SUR method provides, however, a better estimation of the standard errors when the residual vectors of the equations are correlated, as they are in this case.

https://doi.org/10.1017/S0022109000003227 Published online by Cambridge University Press

post-war period 1946–1990 (see Table 3, panel A), and correspondingly we use a sample size of $n = 45$. The parameter values are: $\rho = 0.906$, $\beta = 19.236$, and $\phi = -95.189$. The second simulation uses estimates from the monthly predictive regression in Table 3, panel B, where $n = 379$ with the following parameter values: $\rho = 0.990$, $\beta = 2.080$, and $\phi = -92.196$. We construct $u_t = \phi v_t + e_t$ where $\{v_t\}$ and $\{e_t\}$ are mutually independent i.i.d. normal random variables whose standard deviations are, respectively, 0.137 and 8.621 for the first simulation and 0.041 and 1.8 for the second simulation.

The simulation procedure follows the steps described at the end of Section II. The results are reported in Table 1. Standard errors obtained directly from linear regression output are denoted by $\widehat{SE}$. Thus, for example, $\widehat{SE}(\hat{\rho})$ is the standard error, as given by the OLS regression output, for the estimate of ρ in model (2). Similarly, we obtain $\widehat{SE}(\hat{\beta})$, $\widehat{SE}(\hat{\beta}^c)$, and $\widehat{SE}(\hat{\phi}^c)$. The corrected standard error for $\hat{\beta}^c$ is denoted by $\widehat{SE}^c(\hat{\beta}^c)$, as given by (10). We now summarize our findings from Table 1.

TABLE 1

Simulation Results for Regression Models (1) and (2) with One Predictive Variable

	Case 1 ($n = 45$)		Case 2 ($n = 379$)	
	Mean	Std. Dev.	Mean	Std. Dev.
$\hat{\rho}$	0.81759	0.10154	0.97839	0.012867
$\widehat{SE}(\hat{\rho})$	0.084674	0.01989	0.01024	0.00282
$\hat{\rho}^c$	0.89943	0.10876	0.98886	0.01297
$\hat{\beta}$	27.68732	11.50525	3.14523	1.27993
$\widehat{SE}(\hat{\beta})$	9.76190	2.46097	1.04733	0.29061
$\hat{\beta}^c$	19.84764	12.12282	2.18033	1.28830
$\widehat{SE}(\hat{\beta}^c)$	5.50537	1.58832	0.45299	0.12754
$\widehat{SE}^c(\hat{\beta}^c)$	10.31587	2.57658	1.05468	0.29249
$\hat{\phi}^c$	−95.79690	9.45337	−92.20046	2.27148
$\widehat{SE}(\hat{\phi}^c)$	9.78199	1.52708	2.27326	0.11767

1,500 replications from the single predictor models,

$$(1) \qquad y_t = \alpha + \beta x_{t-1} + u_t,$$

$$(2) \qquad x_t = \theta + \rho x_{t-1} + v_t.$$

Table 1 presents estimation results of the single predictor model by OLS as well as by our estimation procedure, which is as follows: i) Estimate model (2) by OLS and obtain $\hat{\rho}$. Construct the corrected estimator $\hat{\rho}^c = \hat{\rho} + (1+3\hat{\rho})/n + 3(1+3\hat{\rho})/n^2$ and obtain the corrected residuals $v_t^c = x_t - \hat{\theta}^c - \hat{\rho}^c x_{t-1}$, where $\hat{\theta}^c$ is the adjusted intercept. ii) For model (1), obtain $\hat{\beta}^c$ as the coefficient of x_{t-1} in an OLS regression of y_t on x_{t-1} and v_t^c, with intercept. This regression also produces $\hat{\phi}^c$ as the estimator of the coefficient of v_t^c. The parameters $\hat{\beta}$ and $\hat{\rho}$ are obtained from OLS estimation of models (1) and (2), respectively. Standard errors that are estimated directly from linear regression output are denoted by $\widehat{SE}$. The corrected standard error for $\hat{\beta}^c$ is denoted by $\widehat{SE}^c(\hat{\beta}^c)$, as given by (10). Two cases are considered: Case 1 uses parameters from the 45-year predictive regression model in Table 3, panel A: $\rho = 0.906$, $\beta = 19.236$, and $\phi = -95.189$, with $n = 45$. Case 2 uses parameters from the 379-month predictive regression of dividend yield in Table 3, panel B: $\rho = 0.990$, $\beta = 2.080$, and $\phi = -92.196$, with $n = 379$. $u_t = \phi v_t + e_t$ where $\{v_t\}$ and $\{e_t\}$ are mutually independent i.i.d. normal random variables whose standard deviations are, respectively, 0.137 and 8.621 for Case 1 and 0.041 and 1.8 for Case 2.

In both simulations, $\hat{\rho}$ is strongly negatively biased, but the corrected estimator $\hat{\rho}^c$ is very nearly unbiased. Correspondingly, whereas $\hat{\beta}$ is strongly positively biased in both simulations, our proposed $\hat{\beta}^c$ is very nearly unbiased.

Consider first Case 1. The bias in $\hat{\rho}^c$ is only -0.0066, at the cost of a slight inflation in its standard deviation. The estimated standard error $\widehat{SE}(\hat{\rho})$ has an average that is close to the true standard deviation of $\hat{\rho}$. Following the discussion in Section III.A, this supports the use of $\widehat{SE}(\hat{\rho})$ in the calculation of $\widehat{SE}^c(\hat{\beta}^c)$.

https://doi.org/10.1017/S0022109000003227 Published online by Cambridge University Press

Next, we observe that $\hat{\beta}$ is strongly positively biased: the average $\hat{\beta}$ is 27.687 while $\beta = 19.236$, the bias being 8.451, is quite large. However, our corrected estimator $\hat{\beta}^c$ has a very small bias, only 0.612 (=19.848 − 19.236). The actual and theoretical biases match nearly exactly: the bias predicted by our Theorem 2, using the simulation bias for $\hat{\rho}^c$ in place of the actual bias, is $-95.189(0.89943 - 0.906) = 0.625$.

The standard error $\widehat{SE}(\hat{\beta}^c)$ obtained from the regression output greatly underestimates the true standard deviation, consistent with our theoretical discussion in Section III.A. The corrected estimator $\widehat{SE}^c(\hat{\beta}^c)$ obtained from (10) is fairly accurate, having a mean that is within 15% of the true standard deviation.

The estimator $\hat{\phi}^c$ is very nearly unbiased, consistent with Lemma 1, which says that in theory it is exactly unbiased. The standard error of $\widehat{SE}(\hat{\phi}^c)$, obtained directly from the regression output, is very nearly unbiased for the true standard deviation of $\hat{\phi}^c$, consistent with Lemma 3, which says that the square of $\widehat{SE}(\hat{\phi}^c)$ is exactly unbiased for the true variance of $\hat{\phi}^c$.

The simulation results for Case 2 are similar in nature. Here, ρ is closer to 1.0 and the sample size is larger. Still, the mean value of $\hat{\beta}$, 3.145, which is close to the estimated value $\hat{\beta} = 3.046$ obtained from the empirical data on which these simulations are based (Table 3, panel B), exceeds the true $\beta = 2.080$ by 1.065. On the other hand, the mean value of $\hat{\beta}^c$ deviates from the true β by only 0.103. Again, this bias is very close to the bias predicted by Theorem 2 (again using simulation bias for $\hat{\rho}^c$), $-92.196(0.98886 - 0.990) = 0.105$. As before, the mean of $\hat{\phi}^c$ is extremely close to the true value (recall that $\hat{\phi}^c$ is theoretically unbiased).

B. Multiple Predictor Model

Simulations of multiple predictor models given by (11), (12), and (13) are presented in Table 2. We first study the case where the autoregressive matrix Φ is assumed to be diagonal but the errors of the two variables are correlated. We again generate 1,500 replications, this time with sample size $n = 200$. The parameter values here are loosely similar to the actual empirical ones, though they are somewhat varied to illustrate some points.

In all of our multiple predictor simulations, the values of the parameters and the construction of the variables are as follows. $\alpha = 0$, $\beta = (0, 0)'$, $\Theta = (0, 0)'$, $u_t = \phi' v_t + e_t$, the e_t are independent standard normal, $\phi = (\phi_1, \phi_2)' = (-80, -80)'$, the v_t are independent bivariate normal random variables with mean zero and covariance matrix Σ_v, and the sequences $\{e_t\}$ and $\{v_t\}$ are independent of each other.

1. Diagonal Φ Matrix

Panel A in Table 2 presents estimation results for a model with a diagonal AR(1) parameter matrix

$$\Phi = \begin{pmatrix} 0.80 & 0 \\ 0 & 0.95 \end{pmatrix}.$$

https://doi.org/10.1017/S0022109000003227 Published online by Cambridge University Press

We employ two covariance matrices for the errors of the predictive variables. The first is

$$\Sigma_{1v} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}, \quad \text{and the second is} \quad \Sigma_{2v} = \begin{pmatrix} 10 & 9 \\ 9 & 10 \end{pmatrix}.$$

The estimation procedure for the models in panel A follows the estimation procedure of the *univariate* regression described in Section II. We calculate for each $x_{i,t}$ $(i = 1, 2)$ the respective corrected estimators $\hat{\rho}_i^c$ and corrected errors vector $v_{i,t}^c$. (Effectively we set $\hat{\Phi}_{ii}^c = \hat{\rho}_i^c$ with the off-diagonal terms being zero.) Then, we obtain $\hat{\beta}_i^c$ and $\hat{\phi}_i^c$ as the coefficients of $x_{i,t-1}$ and $v_{i,t}^c$, respectively, in an OLS regression of y_t on $x_{1,t-1}$, $x_{2,t-1}$, $v_{1,t}^c$, and $v_{2,t}^c$, with intercept. We also estimate the corrected standard error for $\hat{\beta}_i^c$, denoted by $\widehat{SE}^c(\hat{\beta}_i^c)$, using an analog to (10) and employing the respective parameter estimates,

$$\widehat{SE}^c(\hat{\beta}_i^c) = \sqrt{\{\hat{\phi}_i^c\}^2 \widehat{\mathrm{Var}}(\hat{\rho}_i^c) + \{\widehat{SE}(\hat{\beta}_i^c)\}^2},$$

where $\widehat{\mathrm{Var}}(\hat{\rho}_i^c) = (1 + 3/n + 9/n^2)^2 \mathrm{Var}(\hat{\rho}_i)$ and $\widehat{SE}(\hat{\beta}_i^c)$ is the standard error of $\hat{\beta}_i^c$ obtained directly from the augmented OLS regression.[11]

TABLE 2

Simulation Results for a Model with Multiple Predictive Variables

Panel A. Results with Diagonal AR Matrix Φ

The estimation procedure follows the description in Section V.B.1.

	Results for Σ_{1v}		Results for Σ_{2v}	
	Mean	Std. Dev.	Mean	Std. Dev.
$\mathrm{Corr}(x_1, x_2)$	0.38798	0.14643	0.71154	0.06274
$\hat{\rho}_1$	0.78349	0.04559	0.78309	0.04465
$\widehat{SE}(\hat{\rho}_1)$	0.04385	0.00402	0.04392	0.00397
$\hat{\rho}_1^c$	0.80049	0.04628	0.80008	0.04533
$\hat{\rho}_2$	0.92837	0.03052	0.92953	0.02877
$\widehat{SE}(\hat{\rho}_2)$	0.02581	0.00510	0.02569	0.00495
$\hat{\rho}_2^c$	0.94757	0.03099	0.94876	0.02921
$\hat{\beta}_1$	1.01819	7.82735	−1.07583	10.34441
$\widehat{SE}(\hat{\beta}_1)$	7.25404	0.79978	10.00601	1.12535
$\hat{\beta}_1^c$	−0.07330	4.55905	0.02269	4.00259
$\widehat{SE}(\hat{\beta}_1^c)$	2.76564	0.33749	1.62431	0.20266
$\widehat{SE}^c(\hat{\beta}_1^c)$	4.78765	0.47730	4.22776	0.42323
$\hat{\beta}_2$	2.61877	4.89475	3.85684	6.56538
$\widehat{SE}(\hat{\beta}_2)$	4.27394	0.91415	5.86688	1.26136
$\hat{\beta}_2^c$	0.18096	3.04326	0.10123	2.52673
$\widehat{SE}(\hat{\beta}_2^c)$	1.63067	0.35650	0.95460	0.21242
$\widehat{SE}^c(\hat{\beta}_2^c)$	2.81425	0.57500	2.47409	0.48771
$\hat{\phi}_1^c$	−80.12528	4.71524	−79.95893	4.13121
$\hat{\phi}_2^c$	−79.82175	4.59114	−79.97277	4.17494
$\hat{\Phi}_{11}$	0.78195	0.052150	0.80147	0.06595
$\hat{\Phi}_{12}$	−0.00788	0.03180	−0.02058	0.04130
$\hat{\Phi}_{21}$	0.00492	0.05326	0.01227	0.06491
$\hat{\Phi}_{22}$	0.92495	0.03428	0.92238	0.04199

(continued on next page)

The estimation results for the diagonal-Φ two-predictor model are presented in Table 2, panel A. Although we assume a diagonal matrix Φ, our specification

[11] This is a straightforward extension of the univariate procedure to the diagonal case.

https://doi.org/10.1017/S0022109000003227 Published online by Cambridge University Press

TABLE 2 (continued)

Simulation Results for a Model with Multiple Predictive Variables

Panel B. Results for a Possibly Non-Diagonal AR Matrix Φ

The estimation procedure follows the description in Section V.B.2. The bias-corrected AR(1) parameter matrix estimate $\hat{\Phi}^C$ uses the method of Nicholls and Pope (1988).

	Results for Φ_D		Results for $\Phi_{ND,1}$		Results for $\Phi_{ND,2}$	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
$\hat{\Phi}_{11}$	0.781944	0.052019	0.779928	0.058582	0.682749	0.066851
$\hat{\Phi}^C_{11}$	0.800771	0.051078	0.800618	0.057919	0.701660	0.066862
$\hat{\Phi}_{12}$	−0.007058	0.030244	0.098094	0.050531	0.196758	0.060437
$\hat{\Phi}^C_{12}$	−0.000932	0.028745	0.098491	0.048575	0.197785	0.059362
$\hat{\Phi}_{21}$	0.006034	0.051610	0.104529	0.059265	0.203699	0.067841
$\hat{\Phi}^C_{21}$	0.002058	0.049432	0.102273	0.057136	0.203001	0.066661
$\hat{\Phi}_{22}$	0.925427	0.032955	0.825254	0.052158	0.726641	0.062244
$\hat{\Phi}^C_{22}$	0.946499	0.032703	0.846313	0.051849	0.746045	0.062403
$\hat{\beta}_1$	0.960970	7.24609	1.24331	8.19269	1.08456	9.30569
$\hat{\beta}^C_1$	−0.227031	6.94037	−0.23136	7.91485	−0.37250	9.15355
$\hat{\beta}_2$	2.52964	4.39790	2.13111	7.13647	2.12654	8.49007
$\hat{\beta}^C_2$	0.353793	4.29205	0.414517	6.91680	0.49123	8.36466
$\hat{\phi}^C_1$	−79.9978	0.058198	−79.99776	0.058296	−79.9978	0.058287
$\hat{\phi}^C_2$	−80.0026	0.058209	−80.0025	0.058143	−80.0026	0.058070

1,500 replications from models,

$$y_t = \alpha + \beta' x_{t-1} + u_t, \tag{11}$$

$$x_t = \Theta + \Phi x_{t-1} + v_t. \tag{12}$$

The values of the parameters and the construction of the variables are: $\alpha = 0$, $\beta = (0,0)'$, $\Theta = (0,0)'$, $u_t = \phi' v_t + e_t$, the e_t are independent standard normal, $\phi = (\phi_1, \phi_2)' = (-80, -80)'$, and the v_t are independent bivariate normal random variables with mean zero and covariance matrix Σ_v. The sequences $\{e_t\}$ and $\{v_t\}$ are independent of each other. $n = 200$.

Panel A presents estimation results of a model with a diagonal AR(1) parameter matrix

$$\Phi = \begin{pmatrix} 0.80 & 0 \\ 0 & 0.95 \end{pmatrix}.$$

Results are presented for two covariance matrices,

$$\Sigma_{1v} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}, \quad \text{and} \quad \Sigma_{2v} = \begin{pmatrix} 10 & 9 \\ 9 & 10 \end{pmatrix}.$$

Panel B presents results for the diagonal AR(1) parameter matrix,

$$\Phi_D = \begin{pmatrix} 0.80 & 0 \\ 0 & 0.95 \end{pmatrix},$$

and non-diagonal AR(1) parameter matrices,

$$\Phi_{ND,1} = \begin{pmatrix} 0.80 & 0.1 \\ 0.1 & 0.85 \end{pmatrix} \quad \text{and} \quad \Phi_{ND,2} = \begin{pmatrix} 0.70 & 0.2 \\ 0.2 & 0.75 \end{pmatrix}, \quad \text{all with} \quad \Sigma_v = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}.$$

generates considerable correlation between the two predictors: $\text{Corr}(x_{1,t}, x_{2,t}) = 0.388$ for Σ_{1v} and $\text{Corr}(x_{1,t}, x_{2,t}) = 0.712$ for Σ_{2v}.

We focus on the estimates of the coefficients β_1 and β_2. Consider first the case of Σ_{1v}. We find that both $\hat{\beta}_1$ and $\hat{\beta}_2$ are biased upward. Whereas in fact $\beta_1 = \beta_2 = 0$, the simulation means of $\hat{\beta}_1$ and $\hat{\beta}_2$ are 1.02 and 2.62, respectively. These biases are not surprising, in light of our theoretical results. Applying (18) and using the simulation means of $\hat{\phi}$ and of the entries of the matrix $\hat{\Phi}$ (at the bottom of the table), we calculate that the bias for $\hat{\beta}_1$ and $\hat{\beta}_2$ should be 1.05 and 2.63, respectively, quite close to the respective simulation means of the biases, 1.02 and 2.62. Intuitively, we would expect a greater upward bias in $\hat{\beta}_2$ since $\Phi_{22} > \Phi_{11}$ and, consequently, the bias in $\hat{\Phi}_{22}$ is greater. Applying our reduced-

https://doi.org/10.1017/S0022109000003227 Published online by Cambridge University Press

https://doi.org/10.1017/S0022109000003227 Published online by Cambridge University Press

bias estimator of β, we obtain a very small bias: the average values for $\hat{\beta}_1^c$ and $\hat{\beta}_2^c$ are -0.07330 and 0.18096, respectively. Consistent with Lemma 4 on the unbiasedness of $\hat{\phi}^c$, we find that the averages of the estimates of ϕ_1 and ϕ_2 are both extremely close to -80.

Not only do the estimates $\hat{\beta}_i^c$ for $i = 1$ and 2 have very small bias, they are also quite efficient, having far smaller standard errors than the OLS estimates $\hat{\beta}_i$. Specifically, the standard errors of $\hat{\beta}_i^c$ are *about 60%* of the standard error of the OLS estimates. Our approximation method for the estimation of the standard errors works quite well and our estimates are within 8% (or less than half the standard deviation of $\widehat{SE}^c(\hat{\beta}_i^c)$ of the actual standard errors. (Here, the bias in $\widehat{SE}^c(\hat{\beta}_1^c)$ is positive, while the bias in $\widehat{SE}^c(\hat{\beta}_2^c)$ is negative.)

At the bottom of Table 2, panel A, we present the estimates of the entries of the matrix $\hat{\Phi}$ obtained by the SUR procedure. These estimates, while biased, give a fairly good approximation of the true entries in Φ. Serving as diagnostics they indicate that the off-diagonal terms are practically zero, suggesting that we are justified in employing here the diagonal estimation method that provides more accurate estimates.

Under the covariance matrix Σ_{2v}, there is a much greater correlation between the two predictors: $\mathrm{Corr}(x_{1,t}, x_{2,t}) = 0.71$. As in the previous case, the upward bias in $\hat{\beta}_2$ is greater. Here, the bias in the OLS predictive coefficients $\hat{\beta}_1$ is *negative*, -1.08, while for $\hat{\beta}_2$ it is positive and quite high, 3.86. Again, the bias values here are quite close to those obtained from the application of (18), using the estimated $\hat{\Phi}$. The increase in the variance and covariance terms in Σ_{2v} greatly increases the standard errors of the OLS estimators of β, but the standard errors of the reduced-bias estimators $\hat{\beta}^c$ are somewhat smaller under Σ_{2v} than under Σ_{1v}. The notable effect of the change in the covariance matrix is on the efficiency of the OLS estimation vs. ours. The standard errors of $\hat{\beta}^c$ are about 40% of the standard errors of $\hat{\beta}$! This shows again the efficiency of our reduced-bias estimators. As before, our corrected estimated standard errors $\widehat{SE}^c(\hat{\beta}_i^c)$ for $i = 1, 2$ are quite close on average to the actual standard errors (here, within 5%).

2. Non-Diagonal Φ Matrix

We summarize our estimation procedure for the bivariate case as follows.

i) Construct the bias-corrected AR(1) parameter matrix estimate $\hat{\Phi}^c$ using our iterative procedure based on Nicholls and Pope's (1988) bias expression, as described earlier. Next, construct the bivariate corrected residual series $v_t^c = y_t - \hat{\Theta}^c - \hat{\Phi}^c x_{t-1}$ where $\hat{\Theta}^c$ is the adjusted intercept. Write $v_t^c = (v_{1,t}^c, v_{2,t}^c)'$ and write $x_t = (x_{1,t}, x_{2,t})'$.

ii) Obtain $\hat{\beta}_1^c$ and $\hat{\beta}_2^c$ as the coefficients of $x_{1,t-1}$ and $x_{2,t-1}$ in an OLS 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 estimators of the coefficients of $v_{1,t}^c$ and $v_{2,t}^c$.

Panel B, Table 2 presents results for a potentially non-diagonal AR(1) parameter matrix Φ. In fact, we considered here both a diagonal AR(1) parameter matrix,

$$\Phi_D = \begin{pmatrix} 0.80 & 0 \\ 0 & 0.95 \end{pmatrix},$$

and non-diagonal AR(1) parameter matrices

$$\Phi_{ND,1} = \begin{pmatrix} 0.80 & 0.1 \\ 0.1 & 0.85 \end{pmatrix}, \quad \text{and} \quad \Phi_{ND,2} = \begin{pmatrix} 0.70 & 0.2 \\ 0.2 & 0.75 \end{pmatrix},$$

$$\text{all with} \quad \Sigma_v = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}.$$

For all processes, we generated 1,500 simulated replications, with sample size $n = 200$.

In setting the parameter values of Φ, we note that, in general, the closer the largest eigenvalue of Φ is to 1, the more nearly nonstationary the multivariate AR(1) model is. In the case here, the largest absolute eigenvalues of Φ_D, $\Phi_{ND,1}$, and $\Phi_{ND,2}$ are similar, at 0.95, 0.93, and 0.93, respectively. The structure of Φ_{ND} would accommodate contemporaneous correlation between the predictive variables even if Σ_v were diagonal.

We focus first on the case where the AR(1) parameter matrix is Φ_D, for which the numerical results are given in the left-hand side of panel B, Table 2. We find that the OLS estimates $\hat{\beta}_1$ and $\hat{\beta}_2$ are strongly biased, in agreement with (18). The average values for $\hat{\beta}_1$ and $\hat{\beta}_2$ are 0.96 and 2.53, respectively. The corrected estimators $\hat{\beta}_1^c$ and $\hat{\beta}_2^c$ are less biased, averaging to -0.22 and 0.35, respectively. This is in agreement with Theorem 5. It is instructive to compare the results here with those given in the left-hand side of panel A, Table 2, which correspond to the same model as used here but use an estimation method that requires the knowledge that Φ is diagonal. The bias reduction in $\hat{\beta}_1^c$ and $\hat{\beta}_2^c$ here is somewhat weaker than for the diagonal method, and the variance reduction in $\hat{\beta}_1^c$ and $\hat{\beta}_2^c$ compared to the OLS estimators is no longer found here. Correspondingly, the entries of the corrected estimator $\hat{\Phi}^c$ are less biased than the corresponding entries of $\hat{\Phi}$, though the bias reduction here is not quite as impressive as in the diagonal method. The standard errors of the entries of $\hat{\Phi}^c$ and $\hat{\Phi}$ are comparable. The estimators $\hat{\phi}_1^c$ and $\hat{\phi}_2^c$ average to values very close to the true value of -10, in agreement with Lemma 4.

For the cases where the AR(1) parameter matrices are actually non-diagonal, given by $\Phi_{ND,1}$ and $\Phi_{ND,2}$, the results, reported in the four right-hand side columns of panel B, Table 2 are similar to those reported in the left-hand side. The remaining biases in the coefficients $\hat{\beta}_1^c$ and $\hat{\beta}_2^c$ seem to slightly increase as the magnitude of the off-diagonal terms increases. Altogether, our estimated coefficients show a substantial reduction in the bias compared to the OLS coefficients. The results thus support the applicability of our methodology without the need for either the assumption or the fact that Φ is diagonal.

VI. Empirical Illustration

In this section, we illustrate our estimation method using a common model of predictive regression that is studied by Stambaugh (1999). We employ two sets of data employed in earlier predictive regressions models. The first one is the annual data of Kothari and Shanken (1997) and the second is the monthly data

https://doi.org/10.1017/S0022109000003227 Published online by Cambridge University Press

https://doi.org/10.1017/S0022109000003227 Published online by Cambridge University Press

of Lewellen (2004).[12] We estimate models where stock returns are predicted by lagged financial ratios, most notably the dividend yield, also analyzed by Stambaugh (1999). Our objective here is not to establish whether stock returns are predictable by any financial ratio, but rather to illustrate the usefulness of our estimation method.

A. Univariate Predictive Models

We follow the estimation procedure described in Section II. The predictive model is

$$RM_t = \alpha + \beta X_{t-1} + u_t, \tag{E1}$$

where RM_t is the market return in period t and X_{t-1} is the lagged value of a financial ratio from data in period $t-1$ that is known at the beginning of period t. The predictive variable is assumed to be an AR(1) process,

$$X_t = \theta + \rho X_{t-1} + v_t. \tag{E2}$$

Consider first the annual model of Kothari and Shanken (1997). RM_t is the real (inflation-adjusted) value-weighted annual market return for year t (from April of year $t-1$ to March of year t), and the predictor X_{t-1} is $DIVY_{t-1}$, the value-weighted dividend yield for the preceding year (the dividend paid over year $t-1$ divided by the price level at the end of that year). Throughout, we use a logarithmic transformation of the dividend yield series to reduce its positive skewness.[13] We estimate the models over the entire 65-year sample period 1926–1990 and over the 45-year post-war period 1946–1990, since the autoregressive process of $DIVY_t$ seems to have changed after 1946 (see discussion below). The estimation results are presented in Table 3, panel A.

We do the following estimation procedure.

i) We estimate model (E2) by OLS and obtain $\hat{\rho}$, its standard error $\widehat{SE}(\hat{\rho})$ and t-statistic. These are presented in Table 3, line 1.

ii) The bias correction of $\hat{\rho}$, reported in line 2, is

$$\hat{\rho}^c = \hat{\rho} + (1 + 3\hat{\rho})/n + 3(1 + 3\hat{\rho})/n^2, \tag{E3.1}$$

iii) Using these parameters, the corrected residual v_t^c is calculated as

$$v_t^c = DIVY_t - \left(\hat{\theta}^c + \hat{\rho}^c DIVY_{t-1}\right), \tag{E3.2}$$

where the corrected intercept is $\hat{\theta}^c = (1 - \hat{\rho}^c) \sum_{t=1}^{n} DIVY_t/n$.

iv) Model (E1) is estimated to obtain the coefficient $\hat{\beta}$ and its standard error $\widehat{SE}(\hat{\beta})$, reported in line 3.

v) Using v_t^c from (E3.2), we estimate the augmented model by OLS:

$$RM_t = \alpha + \beta DIVY_{t-1} + \phi v_t^c + e_t. \tag{E4}$$

[12] We thank these authors for kindly providing us their data.

[13] See Lewellen (2004).

TABLE 3

Regression Estimates of Annual Stock Return on Lagged Financial Ratios

Panel A. Annual Predictive Model

RM_t is the value-weighted market real return for year t and $X_t = \log(DIVY_t)$ is the logarithm of the value-weighted annual dividend yield.

	Coefficient	From Model	1926–1990 ($n = 65$)	1946–1990 ($n = 45$)
1	$\hat{\rho}$	(E2)	0.627	0.823
	$(\widehat{SE}\hat{\rho})$ [t]	(OLS)	(0.098) [6.41]	(0.087) [9.52]
2	$\hat{\rho}^C$	(E3.1)	0.673	0.906
3	$\hat{\beta}$	(E1)	25.250	27.066
	$(\widehat{SE}\hat{\beta})$ [t]	(OLS)	(9.847) [2.56]	(8.872) [3.05]
4	$\hat{\beta}^C$	(E4)	21.343	19.236
	$(\widehat{SE}\hat{\beta}^C)$ [t]		(5.440) [3.92]	(3.370) [5.71]
5	$(\widehat{SE}^C\hat{\beta}^C)$ [t]	(E5)	(10.212) [2.09]	(9.448) [2.04]
6	$\hat{\phi}^C$		−84.283	−95.189
	$(\widehat{SE}\hat{\phi}^C)$ [t]	(E4)	(6.994) [12.05]	(5.878) [16.20]

Panel B. Monthly Predictive Model

The model uses for X_t the logarithm of three financial ratios: $DIVY_t$, the value-weighted dividend yield, which is the annual dividend divided by the current index level; B/M_t, the book-to-market ratio; and E/P_t, the earnings-to-price ratio. RM_t is the value-weighted market return for month t. Data are for the NYSE, 5/1963–12/1994 (379 months).

	Coefficient	From Model	Dividend Yield	Book/Market	Earnings/Price
1	$\hat{\rho}$	(E2)	0.979	0.985	0.985
	$(\widehat{SE}\hat{\rho})$ [t]	(OLS)	(0.011) [92.51]	(0.009) [106.90]	(0.008) [123.01]
2	$\hat{\rho}^C$	(E3.1)	0.990	0.995	0.996
3	$\hat{\beta}$	(E1)	3.046	1.935	1.638
	$(\widehat{SE}\hat{\beta})$ [t]	(OLS)	(1.009) [3.02]	(0.805) [2.40]	(0.673) [2.43]
4	$\hat{\beta}^C$	(E4)	2.080	1.101	0.879
	$(\widehat{SE}\hat{\beta}^C)$ [t]		(0.258) [8.07]	(0.340) [3.24]	(0.347) [2.53]
5	$(\widehat{SE}^C\hat{\beta}^C)$ [t]	(E5)	(1.017) [2.05]	(0.811) [1.36]	(0.678) [1.30]
6	$\hat{\phi}^C$		−92.196	−79.221	−72.138
	$(\widehat{SE}\hat{\phi}^C)$ [t]	(E4)	(1.250) [73.61]	(1.898) [41.73]	(2.227) [32.39]

Table 3 presents results of the following models:

(E1) $$RM_t = \alpha + \beta X_{t-1} + u_t.$$

(E2) $$X_t = \theta + \rho X_{t-1} + v_t.$$

(E3.1) $$\hat{\rho}^C = \hat{\rho} + (1 + 3\hat{\rho})/n + 3(1 + 3\hat{\rho})/n^2.$$

(E3.2) $$v_t^C = X_t - (\hat{\theta}^C + \hat{\rho}^C X_{t-1}). \quad (\hat{\theta}^C = (1 - \hat{\rho}^C)\sum_{t=1}^{n} X_t/n).$$

(E4) $$RM_t = \alpha + \beta X_{t-1} + \phi v_t^C + e_t.$$

(E5) $$\widehat{SE}^C(\hat{\beta}^C) = \sqrt{(\hat{\phi}^C)^2\{\widehat{SE}(\hat{\rho})\}^2(1 + 3/n + 9/n^2)^2 + \{\widehat{SE}(\hat{\beta}^C)\}^2}.$$ Also, $t = \hat{\beta}^C/\widehat{SE}^C(\hat{\beta}^C)$.

Estimators $\hat{\theta}$, $\hat{\rho}$, $\hat{\alpha}$, and $\hat{\beta}$ are obtained from OLS regressions. Estimators $\hat{\theta}^C$, $\hat{\rho}^C$, $\hat{\alpha}^C$, and $\hat{\beta}^C$ are obtained under our estimation procedure described in the text. In parentheses there are the standard errors of the estimated coefficients and [t] is the corresponding t-statistic. All standard errors and t-statistics are obtained directly from OLS, except for those in (E5).

We obtain the parameters $\hat{\beta}^C$ (line 4) and $\hat{\phi}^C$ (line 6), their respective standard errors $\widehat{SE}(\hat{\beta}^C)$ and $\widehat{SE}(\hat{\phi}^C)$, and their t-statistics.

vi) The corrected standard error of $\hat{\beta}^C$, $\widehat{SE}^C(\hat{\beta}^C)$, is calculated according to (10) as follows

$$\text{(E5)} \quad \widehat{SE}^C\left(\hat{\beta}^C\right) = \sqrt{\left(\hat{\phi}^C\right)^2\left\{\widehat{SE}\left(\hat{\rho}\right)\right\}^2(1 + 3/n + 9/n^2)^2 + \left\{\widehat{SE}\left(\hat{\beta}^C\right)\right\}^2}.$$

This is reported in line 5. The corresponding t-statistic is calculated as $\hat{\beta}^C/\widehat{SE}^C(\hat{\beta}^C)$.

https://doi.org/10.1017/S0022109000003227 Published online by Cambridge University Press

The estimation results in Table 3 indicate that $\hat{\beta}$ may be biased upward because $\hat{\phi}^c < 0$ (line 6) and $\hat{\rho} < \hat{\rho}^c$ (lines 1 and 2). Indeed, we obtain that $\hat{\beta}^c < \hat{\beta}$ (lines 3 and 4). Therefore, in line 5, $\hat{\beta}^c / \widehat{SE}^c(\hat{\beta}^c) = 2.13$ and the null hypothesis $\beta = 0$ is not rejected as strongly as it is based on the *t*-statistic in line 3. The autoregressive process of the dividend yield series may well have changed over time. In the 20 years 1926–1945, $\hat{\rho} = 0.218$ ($t = 0.96$), which is much smaller than the estimate of $\hat{\rho}$ for the following period 1946–1990. If $\rho = 0$ in the earlier period, we would not expect $\hat{\beta}$ to be biased. Hence we focus on the 45-year period 1946–1990, where $\hat{\rho}$ is large and statistically significant. For this period, the OLS regression produces $\hat{\beta} = 27.066$ and $t = 3.05$, whereas by our procedure the slope coefficient is smaller and has lower statistical significance: $\hat{\beta}^c = 19.236$ with $t = 2.04$. The bias in $\hat{\beta}$ may be greater for the 45-year period than for the entire 65-year period because in the shorter period, the bias in $\hat{\rho}$ is apparently greater and also $\hat{\phi}^c$ is more negative.

Next, we apply our method to the predictive models analyzed by Lewellen (2004), where monthly stock returns *RM* (value-weighted index) are predicted by lagged monthly values (in logarithm) of three financial ratios. Each model employs one predictor only and consequently we employ here our univariate method. The predictive variables are: i) the dividend yield for the value-weighted index, *DIVY*, calculated as the annual dividend divided by the current index level; ii) the book-to-market ratio B/M, the book value of equity in the previous year divided by the market equity in the previous month; and iii) the earnings-to-price ratio E/P, the operating earnings divided by market equity. All variables and returns are for the NYSE. We start with 1963, when Compustat data became available, and end in 1994 since Lewellen (2004) points out that the predictive power of these three ratios declined considerably starting in 1995. The sample period is thus 379 months. The estimation process is similar to that outlined in steps i)–vi) above, and the results are presented in Table 3, panel B.

In the monthly models, the estimates of the autoregressive coefficients ρ for all three predictive variables are quite high. Our estimated $\hat{\rho}^c$ is 0.990 for *DIVY*, 0.995 for B/M, and 0.996 for E/P. Given the large sample size, the bias in $\hat{\rho}$ should be relatively small. However, the relatively large negative value of $\hat{\phi}^c$ should produce a strong bias in $\hat{\beta}$ for all three predictors.

By the OLS estimator $\hat{\beta}$, we reject the null hypothesis that $\beta = 0$ at well below the standard 5% level for all three predictors (two-tail tests). However, using $\hat{\beta}^c$ and $\widehat{SE}^c(\hat{\beta}^c)$, we cannot reject the null for the predictors B/M and E/P. For *DIVY*, the significance is lower. The conclusion according to these results is that the series B/M and E/P do not have predictive power while the series *DIVY* does.

Our results cast doubt on the conclusions drawn in earlier studies on the significance of the three predictors. The results of Fama and French (1984) and Campbell and Shiller (1988) on the predictive power of the dividend-price ratio on future stock returns are called into question, given the marginal significance of its coefficient. When we extend the study period to be 1927–2001 (900 months compared with 379 months in the sample studied in Table 3, panel B), we obtain that the coefficient $\hat{\beta}^c$ of $DIVY_{t-1}$ is 0.366 with $t = 0.70$, insignificant. This suggests that the lagged dividend-price ratio is not a significant predictor of stock

https://doi.org/10.1017/S0022109000003227 Published online by Cambridge University Press

returns. Our results also question the conclusions of Pontiff and Schall (1998) that the aggregate book-to-market ratio forecasts future market returns. We find that while the OLS slope coefficient of the book-to-market ratio is statistically significant, the adjusted coefficient and its adjusted standard error show that the effect of the book-to-market ratio on future stock returns is statistically insignificant.

B. A Bivariate Predictive Model

We apply our multivariate estimation procedure, where the matrix Φ is diagonal, to a bivariate model where the predictive variables are the dividend yield *DIVY* and the earnings-to-price ratio E/P both in logarithms. We first examine whether the nature of the covariance matrix Φ by estimating the system of equations, $DIVY_t = \Phi_{10} + \Phi_{11} DIVY_{t-1} + \Phi_{12} E/P_{t-1} + v_{1,t}$ and $E/P_t = \Phi_{20} + \Phi_{21} DIVY_{t-1} + \Phi_{22} E/P_{t-1} + v_{2,t}$, where $v_{1,t}$ and $v_{2,t}$ are the error terms, which are serially independent but may be mutually correlated. The estimation results, presented in Table 4, panel A, indicate that the off-diagonal terms of Φ are essentially zero. However, the correlation of their error terms is quite high: $\mathrm{Corr}(\hat{v}_{DIVY,t}, \hat{v}_{E/P,t}) = 0.858$. This means that while Φ is apparently diagonal, Σ_v is not. Thus, the scenario here resembles the one in our multivariate diagonal model and in the corresponding simulations.

We can therefore proceed with our bivariate estimation method where Φ is assumed to be diagonal. That is, we estimate the augmented regression, $RM_t = \beta_0 + \beta_1 DIVY_{t-1} + \beta_2 E/P_{t-1} + \phi_1 v^c_{DIVY,t} + \phi_2 v^c_{E/P,t}$. where $v^c_{DIVY,t}$ and $v^c_{E/P,t}$ are the error terms obtained from the univariate estimation procedure, steps ii) and iii), described in Section VI.A. By this augmented regression we obtain the reduced-bias coefficients $\hat{\beta}^c$ and the coefficients $\hat{\phi}^c$, which are unbiased. In addition, we calculate the corrected standard errors of $\hat{\beta}^c$, denoted by $\widehat{SE}^c(\hat{\beta}^c)$, as we do in the univariate case, step vi). The estimation results are presented in Table 4, panel B.

The OLS estimation yields $\hat{\beta}_1 = 5.269$ and $\hat{\beta}_2 = -1.599$. The first coefficient is marginally significant while the second is not, using the OLS t-statistics. The two variables in the regression are highly correlated, $\mathrm{Corr}(DIVY_t, E/P_t) = 0.924$, which presents a collinearity problem. Taken together, the two variables have a significant effect on RM_t, as measured by the F-value of the regression: $F = 4.97$ ($p = 0.007$). However, as we know from our theory and from the simulation results for the bivariate case, the OLS coefficients $\hat{\beta}$ (and, correspondingly, the test statistics) can be biased in this case since we find that the coefficients ϕ are negative. Indeed, we obtain that $\hat{\beta}^c_1 = 2.509$, which is much smaller than $\hat{\beta}_1$. Bias in $\hat{\beta}_1$ may be caused by the large negative value -82.582 of $\hat{\phi}^c_1$. After correcting the standard error, $\hat{\beta}^c_1$ is still significant. The coefficient $\hat{\beta}^c_2$ is smaller (in absolute value) than $\hat{\beta}_2$ and insignificant, while $\hat{\phi}_2$ is highly significant. By these results, only *DIVY* has significant predictive power. However, when extending the sample by 84 months to 2001, none of the predictive variables is significant, which again casts doubt on the conclusions in earlier studies that suggest that these variables are significant predictors of stock returns.

https://doi.org/10.1017/S0022109000003227 Published online by Cambridge University Press

TABLE 4

Regression Estimates of a Bivariate Model with a Diagonal Covariance Matrix

Panel A. The Covariance Matrix $\hat{\Phi}$ for the Bivariate Model

Table 4 presents a system estimate (SUR) of the following system of equations: $DIVY_t = \Phi_{10} + \Phi_{11} DIVY_{t-1} + \Phi_{12} E/P_{t-1} + v_{1,t}$, $E/P_t = \Phi_{20} + \Phi_{21} DIVY_{t-1} + \Phi_{22} E/P_{t-1} + v_{2,t}$, where $v_{1,t}$ and $v_{2,t}$ are the error terms, which are serially independent but may be mutually correlated. Table 4 presents the parameter estimates and the respective t-statistics.

	$DIVY_{t-1}$	E/P_{t-1}
$DIVY_t$ [t]	0.957 [34.65]	0.016 [0.87]
E/P_t [t]	−0.006 [0.20]	0.989 [47.29]

Panel B. Bivariate Predictive Model

$RM_t = \beta_0 + \beta_1 DIVY_{t-1} + \beta_2 E/P_{t-1} + \phi_1 v^c_{DIVY,t} + \phi_2 v^c_{E/P,t}$. The subscripts of the error terms v indicate that they pertain to $DIVY$ and E/P from the univariate regression (F2) and the estimation procedure (F3). $\text{Corr}(\hat{v}^c_{DIVY,t}, \hat{v}^c_{E/P,t}) = 0.858$.

	Coefficient	From Model	$DIVY_{t-1}$	E/P_{t-1}
1	$\hat{\beta}$	(F1)	5.269	−1.599
	$(\widehat{SE}\hat{\beta})$ [t]	(OLS)	(2.643) [1.99]	(1.757) [0.91]
2	$\hat{\beta}^c$	(F4)	2.509	−0.345
	$(\widehat{SE}\hat{\beta}^c)$ [t]		(0.660) [3.80]	(0.439) [0.79]
3	$(\widehat{SE}^c\hat{\beta}^c)$ [t]	(F5)	(0.938) [2.67]	(0.447) [0.77]
4	$\hat{\phi}^c$	(F4)	−82.582	−9.800
	$(\widehat{SE}\hat{\phi}^c)$ [t]		(2.389) [34.57]	(2.096) [4.68]

Table 4 presents results of the following models:

(F1) $RM_t = \alpha + \beta_1 DIVY_{t-1} + \beta_2 E/P_{t-1} + u_t$.

(F2) $X_t = \theta + \rho X_{t-1} + v_t$, for $X = DIVY$ and E/P.

(F3) The calculation of $\hat{v}^c_{DIVY,t}$ and $v^c_{E/P,t}$ follows the procedure outlined in Table 3, procedure (E3), for the univariate regressions for each of the predictors $DIVY$ and E/P.

(F4) $RM_t = \alpha + \beta_1 DIVY_{t-1} + \beta_2 E/P_{t-1} + \phi_1 \hat{v}^c_{DIVY,t} + \phi_2 \hat{v}^c_{E/P,t} + u_t$.

(F5) $\widehat{SE}^c(\hat{\beta}^c)$ is calculated, as in (E5) in Table 3, separately for $DIVY$ and E/P using the respective estimates.

Estimators $\hat{\theta}$, $\hat{\rho}$, $\hat{\alpha}$, and $\hat{\beta}$ are obtained from OLS regressions of model (F1). Estimators of $\hat{\beta}^c$ are obtained from an OLS regression of model (F4). In parentheses there are the standard errors of the estimated coefficients and [t] is the corresponding t-statistic.

VII. Concluding Remarks

This paper provides a convenient way to estimate a predictive regression model where a time series of one variable is regressed on lagged variables that are assumed to predict it. The predictive variables have a first-order autoregressive structure and their disturbance terms are contemporaneously correlated with that of the predicted variable. For the *single* predictor case, Stambaugh (1999) shows that the OLS-estimated coefficient of the lagged variable is biased when computed from a small sample. There is no available estimation method for this model, except for a plug-in version where, in the single predictor case the sample estimated parameters are plugged into Stambaugh's bias expression. In the multiple predictor case, there heretofore existed neither an expression for the bias of the OLS estimators of the coefficients of the predictive variables, nor any reduced-bias estimation method.

We develop an estimation method for both the single and multiple predictor models that produces a reduced-bias estimator of the coefficients of the lagged variables, which turns out to be identical in the single predictor case to the estimator motivated by Stambaugh's (1999) bias expression. We also develop a straightforward estimation method for a reduced-bias standard error, which we

https://doi.org/10.1017/S0022109000003227 Published online by Cambridge University Press

find works well in some versions of the multiple predictor models. With these standard errors, it is easy to perform tests of statistical significance. Our method is particularly useful in the multiple predictor case for which there is no direct reduced-bias estimation method available. The performance and usefulness of our method is demonstrated in simulations, and we illustrate its use in some applications with actual data employed in other studies.

Appendix

Proof of Theorem 1. As in Stambaugh (1999), we define the error process $\{e_t\}$ by $u_t = \phi v_t + e_t = E[u_t|v_t] + e_t$. Since $(e_t, v_t)'$ is bivariate normal and $E[e_t|v_t] = 0$, e_t and v_t must be independent for all t. Since the vectors $(u_t, v_t)'$ are independent, e_t must be independent of $v_1, \ldots, v_n$, and x_0. Thus, for all t, e_t is independent of $x_0, \ldots, x_n$.

Let 1_n be an $n \times 1$ vector of ones, and define the matrix $\tilde{X} = [1_n, \{x_{t-1}\}_{t=1}^n, \{v_t\}_{t=1}^n]$. Let $y = (y_1, \ldots, y_n)'$. Since $y_t = \alpha + \beta x_{t-1} + \phi v_t + e_t$, we have

$$y = \tilde{X} \begin{pmatrix} \alpha \\ \beta \\ \phi \end{pmatrix} + e,$$

where $e = (e_1, \ldots, e_n)'$, and the vector $(\tilde{\alpha}, \tilde{\beta}, \tilde{\phi})$ of least-squares estimators is given by

$$\begin{pmatrix} \tilde{\alpha} \\ \tilde{\beta} \\ \tilde{\phi} \end{pmatrix} = \left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'y = \begin{pmatrix} \alpha \\ \beta \\ \phi \end{pmatrix} + \left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'e,$$

Since e has zero mean and is independent of $\tilde{X}$, we obtain $E[\tilde{\beta}] = \beta$, thereby completing the proof. □

Proof of Theorem 2. As in the proof of Theorem 1, we use the representation $y_t = \alpha + \beta x_{t-1} + \phi v_t + e_t$, where the error terms e_t are i.i.d. normal with mean zero, and for all t, e_t is independent of $x_0, \ldots, x_n$. Since $v_t - v_t^c = (\hat{\theta}^c - \theta) + (\hat{\rho}^c - \rho)x_{t-1}$, we can write

$$y_t = \left[\alpha + \phi\left(\hat{\theta}^c - \theta\right)\right] + \left[\beta + \phi\left(\hat{\rho}^c - \rho\right)\right] x_{t-1} + \phi v_t^c + e_t. \tag{24}$$

Since $\hat{\theta}^c$, $\hat{\rho}^c$, and $\{v_t^c\}$ are all functions of $x_0, \ldots, x_n$, it follows that, conditionally on $x_0, \ldots, x_n$, equation (24) satisfies all the regularity conditions needed for a linear regression model, and therefore $E[\hat{\beta}^c|x_0, \ldots, x_n] = \beta + \phi(\hat{\rho}^c - \rho)$. Taking the expectation of the formula above and applying the double expectation theorem yields $E[\hat{\beta}^c - \beta] = \phi E\left[\hat{\rho}^c - \rho\right]$, thereby completing the proof. □

Proof of Theorem 3. First consider the case $\omega = 0$. Let $\hat{v}_t = x_t - (\hat{\theta} + \hat{\rho}x_{t-1})$ where $\hat{\theta}$ and $\hat{\rho}$ are the usual OLS estimates of θ and ρ. Then (24) becomes

$$y_t = \left[\alpha + \phi\left(\hat{\theta} - \theta\right)\right] + \left[\beta + \phi\left(\hat{\rho} - \rho\right)\right] x_{t-1} + \phi\hat{v}_t + e_t. \tag{25}$$

Now, $\hat{v}_t$ is orthogonal to x_{t-1} in the sense that $\sum_{t=1}^n \hat{v}_t x_{t-1} = 0$, so that if $\omega = 0$, $\hat{\beta}^c$ is equal to the OLS slope estimate in a simple regression of y_t on x_{t-1}. Hence, $\hat{\beta}^c = \hat{\beta}$ if $\omega = 0$. Due to the orthogonality just mentioned, the OLS coefficient of x_{t-1} in the multiple regression of y_t on x_{t-1} and $\hat{v}_t$ remains unchanged from the value it took in the simple regression of y_t on x_{t-1}. This same orthogonality also implies that the coefficient of $\hat{v}_t$ in the regression of y_t on x_{t-1} and $\hat{v}_t$ in (25) is $\hat{\phi}^s$.

Next, suppose that $\hat{\rho}^c = \hat{\rho} + \omega$, where ω is any nonzero real number. Since $v_t^c = x_t - (\hat{\theta}^c + \hat{\rho}^c x_{t-1})$, we obtain

$$v_t^c = \hat{v}_t + \left(\hat{\theta} - \hat{\theta}^c\right) - \omega x_{t-1}, \tag{26}$$

https://doi.org/10.1017/S0022109000003227 Published online by Cambridge University Press

so that the vector v_t^c is the sum of the vector $\hat{v}_t$ with a constant vector and a vector that is collinear with x_{t-1}. Thus the linear space spanned by a constant vector together with x_{t-1} and $\hat{v}_t$ is the same as the linear space spanned by a constant vector together with x_{t-1} and v_t^c. It follows that the fitted values from the regressions of y_t on these two spaces are the same, that is,

$$\hat{\alpha} + \hat{\beta} x_{t-1} + \hat{\phi}^s \hat{v}_t \quad = \quad \hat{\alpha}^c + \hat{\beta}^c x_{t-1} + \hat{\phi}^c v_t^c, \tag{27}$$

for $t = 1, \ldots, n$. The subtraction of ωx_{t-1} in (26) has no effect on the OLS coefficient of v_t^c, and hence $\hat{\phi}^c = \hat{\phi}^s$ for all values of ω. Now, using (26) to substitute for v_t^c in the right-hand side of (27), we obtain $\hat{\alpha} + \hat{\beta} x_{t-1} + \hat{\phi}^s \hat{v}_t = [\hat{\alpha}^c + \hat{\phi}^s(\hat{\theta} - \hat{\theta}^c)] + (\hat{\beta}^c - \hat{\phi}^s \omega) x_{t-1} + \hat{\phi}^s \hat{v}_t$. Since the least-squares coefficients are uniquely determined by the least-squares algorithm, the coefficients of x_{t-1} on the left-hand and right-hand sides of the above equation must be the same, so that $\hat{\beta}^c = \hat{\beta} + \hat{\phi}^s \omega$. □

Proof of Lemma 1. Using equation (24) and arguing as in the proof of Theorem 2, we obtain $E[\hat{\phi}^c | x_0, \ldots, x_n] = \phi$. Now, taking the expectation and applying the double expectation theorem yields $E[\hat{\phi}^c] = \phi$. □

Proof of Lemma 2. Let $\{r_t\}_{t=1}^n$ be the sequence of residuals obtained in an OLS regression of x_{t-1} on v_t^c (with intercept). Then we have

$$\hat{\beta}^c \quad = \quad \frac{\sum_{t=1}^n r_t y_t}{\sum_{t=1}^n r_t^2}, \qquad \text{and} \qquad \left[\widehat{SE}\left(\hat{\beta}^c\right)\right]^2 \quad = \quad \frac{\hat{\sigma}^2}{\sum_{t=1}^n r_t^2}, \tag{28}$$

where $\hat{\sigma}^2$ is the estimator of the error variance from a regression (with intercept) of y_t on x_{t-1} and v_t^c. Note that $\hat{\sigma}^2$ is simply the residual sum of squares from this regression divided by $n - 3$. We use the error $e_t = u_t - \phi v_t$ as in the previous proofs. Note that the variance of e_t is $\sigma_e^2 = \mathrm{Var}(e_t) = \sigma_u^2 - \sigma_{uv}^2/\sigma_v^2$. We first obtain a convenient expression for $\hat{\beta}^c - \beta$. Since the residual vector is orthogonal to the vectors of explanatory variables, we have

$$\sum_{t=1}^n r_t \quad = \quad 0, \quad \sum_{t=1}^n r_t v_t^c \quad = \quad 0. \tag{29}$$

Writing $x_{t-1} = a_0 + a_1 v_t^c + r_t$, we obtain from (29) that

$$\sum_{t=1}^n r_t x_{t-1} \quad = \quad \sum_{t=1}^n r_t^2. \tag{30}$$

Therefore, from (28), we have

$$\begin{aligned} \hat{\beta}^c \quad &= \quad \frac{1}{\sum_{t=1}^n r_t^2} \sum_{t=1}^n r_t (\alpha + \beta x_{t-1} + \phi v_t + e_t) \\ &= \quad \frac{1}{\sum_{t=1}^n r_t^2} \sum_{t=1}^n r_t \left[\beta x_{t-1} + \phi v_t^c + \phi (v_t - v_t^c) + e_t\right]. \end{aligned}$$

From (29) and (30), we have

$$\frac{1}{\sum_{t=1}^n r_t^2} \sum_{t=1}^n r_t \left(\beta x_{t-1} + \phi v_t^c\right) \quad = \quad \beta.$$

Since $v_t - v_t^c = (\hat{\theta}^c - \theta) + (\hat{\rho}^c - \rho) x_{t-1}$, we conclude that

$$\hat{\beta}^c - \beta \quad = \quad \phi (\hat{\rho}^c - \rho) + \frac{\sum_{t=1}^n r_t e_t}{\sum_{t=1}^n r_t^2}. \tag{31}$$

https://doi.org/10.1017/S0022109000003227 Published online by Cambridge University Press

Since the $\{r_t\}$ are functions of $\{x_t\}$, and since for all t, e_t is independent of $\{x_t\}_{t=0}^n$, it follows that for all t, e_t is independent of $r_1, \ldots, r_n$. Therefore, the two terms on the right-hand side of (31) are uncorrelated, and the second term has mean zero. It follows that $E[\hat{\beta}^c - \beta]^2 = \phi^2 E[\hat{\rho}^c - \rho]^2 + \sigma_e^2 E[1/(\sum_{t=1}^n r_t^2)]$. It remains to be shown that

$$\sigma_e^2 E\left[\frac{1}{\sum_{t=1}^n r_t^2}\right] = E\left[\frac{\hat{\sigma}^2}{\sum_{t=1}^n r_t^2}\right]. \tag{32}$$

Let H denote the hat matrix corresponding to $X = [1_n, x_{t-1}, v_t^c]$ for the regression of y_t on x_{t-1}, v_t^c. That is, $H = X(X'X)^{-1}X'$. Let r_0 denote the residual vector from this regression, so that $r_0 = (I - H)y = (I - H)e$, where I denotes an $n \times n$ identity matrix. Conditionally on X, we have $\sum_{t=1}^n r_{0t}^2 = e'(I - H)e \sim \sigma_e^2 \chi_{n-3}^2$, and since the random variable on the right-hand side does not depend on X, the result is true unconditionally as well. Thus, $\hat{\sigma}^2 = 1/(n-3)\sum_{t=1}^n r_{0t}^2$ is an unbiased estimator of σ_e^2, that is, $E[\hat{\sigma}^2] = \sigma_e^2$. Now, we have

$$\begin{aligned} E\left[\frac{\hat{\sigma}^2}{\sum_{t=1}^n r_t^2}\middle| X\right] &= E\left[\frac{1}{n-3}\frac{e'(I-H)e}{\sum_{t=1}^n r_t^2}\middle| X\right] \\ &= \frac{1}{\sum_{t=1}^n r_t^2}\frac{1}{n-3}E\left[\sigma_e^2\chi_{n-3}^2\right] = \sigma_e^2\frac{1}{\sum_{t=1}^n r_t^2}. \end{aligned}$$

Taking expectations of both sides and using the double expectation theorem yields (32). □

Proof of Lemma 3. Let q be the residual vector in an OLS regression of v_t^c on x_{t-1}. Note that q is independent of the error vector, $e = u - \phi v$. Then $[\widehat{SE}(\hat{\phi}^c)]^2 = \hat{\sigma}^2/(\sum_{t=1}^n q_t^2)$. Using the representation, $y_t = \alpha + \phi(\hat{\theta}^c - \theta) + \beta x_{t-1} + \phi v_t^c + \phi(\hat{\rho}^c - \rho)x_{t-1} + e_t$, together with the properties $\sum q_t v_t^c = \sum q_t^2$ and $\sum q_t x_{t-1} = \sum q_t = 0$, we obtain

$$\hat{\phi}^c = \frac{\sum_{t=1}^n q_t y_t}{\sum_{t=1}^n q_t^2} = \phi + \frac{\sum_{t=1}^n q_t e_t}{\sum_{t=1}^n q_t^2}. \tag{33}$$

Since $\{e_t\}$ is independent of $\{q_t\}$ and $E[e_t] = 0$, the expectation of the second term on the right-hand side of the above equation is zero, and we obtain

$$\mathrm{Var}\left[\hat{\phi}^c\right] = \sigma_e^2 E\left[\frac{1}{\sum_{t=1}^n q_t^2}\right]. \tag{34}$$

Proceeding as in the proof of Lemma 2, we have

$$\begin{aligned} E\left[\frac{\hat{\sigma}^2}{\sum_{t=1}^n q_t^2}\middle| X\right] &= E\left[\frac{1}{n-3}\frac{e'(I-H)e}{\sum_{t=1}^n q_t^2}\middle| X\right] \\ &= \frac{1}{\sum_{t=1}^n q_t^2}\frac{1}{n-3}E\left[\sigma_e^2\chi_{n-3}^2\right] = \sigma_e^2\frac{1}{\sum_{t=1}^n q_t^2}. \end{aligned}$$

Taking expectations of both sides and using the double expectation theorem yields $E[\hat{\sigma}^2/(\sum_{t=1}^n q_t^2)] = \sigma_e^2 E[1/(\sum_{t=1}^n q_t^2)]$. Lemma 3 now follows from (34). □

Proof of Theorem 4. As pointed out in (14), we have $y_t = \alpha + \beta' x_{t-1} + \phi' v_t + e_t$, where $\{e_t\}$ has zero mean and is independent of both $\{v_t\}$ and $\{x_t\}$. Let 1_n be an $n \times 1$ vector of ones, and define the $n \times (2p+1)$ matrix $\tilde{X} = [1_n, (x_0, x_1, \ldots x_{n-1})', (v_1, \ldots, v_n)']$. Let $y = (y_1, \ldots, y_n)'$. We have

$$y = \tilde{X}\begin{pmatrix} \alpha \\ \beta \\ \phi \end{pmatrix} + e,$$

https://doi.org/10.1017/S0021090000003227 Published online by Cambridge University Press

where $e = (e_1, \ldots, e_n)'$, and the vector $(\tilde{\alpha}, \tilde{\beta}, \tilde{\phi})$ of least-squares estimators is given by

$$\begin{pmatrix} \tilde{\alpha} \\ \tilde{\beta} \\ \tilde{\phi} \end{pmatrix} = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'y = \begin{pmatrix} \alpha \\ \beta \\ \phi \end{pmatrix} + (\tilde{X}'\tilde{X})^{-1}\tilde{X}'e.$$

Since e has zero mean and is independent of $\tilde{X}$, we obtain $E[\tilde{\beta}] = \beta$, thereby completing the proof. □

Proof of Theorem 5. Using (11), (12), (13), and (16), we can write

$$y_t = \left[\alpha + \phi'\left(\hat{\Theta}^c - \Theta\right)\right] + \left\{\beta' + \phi'\left(\hat{\Phi}^c - \Phi\right)\right\} x_{t-1} + \phi' v_t^c + e_t. \tag{35}$$

Since $\hat{\Theta}^c$, $\hat{\Phi}^c$, and $\{v_t^c\}$ are all functions of $x_0, \ldots, x_n$, it follows that, conditionally on $x_0, \ldots, x_n$, equation (35) satisfies all the regularity conditions needed for a liner regression model, and therefore $E[\hat{\beta}^c | x_0, \ldots, x_n] = \beta + [\hat{\Phi}^c - \Phi]'\phi$. Taking the expectation of the formula above and applying the double expectation theorem yields $E[\hat{\beta}^c - \beta] = E[\hat{\Phi}^c - \Phi]'\phi$. □

Proof of Lemma 4. Using equation (35) and arguing as in the proof of Theorem 5, we obtain $E[\hat{\phi}^c | x_0, \ldots, x_n] = \phi$. Now, taking the expectation and applying the double expectation theorem yields $E[\hat{\phi}^c] = \phi$. □

Proof of Theorem 6. Our proof is a direct generalization of the proof of Theorem 3. We first consider the case $\omega = 0$. Then (35) becomes

$$y_t = [\alpha + \phi'(\hat{\Theta} - \Theta)] + \{\beta' + \phi'(\hat{\Phi} - \Phi)\}x_{t-1} + \phi'\hat{v}_t + e_t. \tag{36}$$

Since $\hat{\Theta}$ and $\hat{\Phi}$ are the OLS estimates, each of the first $p + 1$ columns of $\tilde{X}$ is orthogonal to each of the final p columns of $\tilde{X}$. Hence $\hat{\beta}^c = \hat{\beta}$ if $\omega = 0$. Next, suppose that $\hat{\Phi}^c = \hat{\Phi} + \omega$, where the $p \times p$ matrix ω is not identically zero. Note that

$$v_t^c = \hat{v}_t + \left(\hat{\Theta} - \hat{\Theta}^c\right) - \omega x_{t-1}, \tag{37}$$

so that the linear space of n-dimensional vectors spanned by a constant vector together with the rows of $[x_0, \ldots, x_{n-1}]$ and the rows of $[\hat{v}_1, \ldots, \hat{v}_n]$ is the same as the linear space spanned by a constant vector together with the rows of $[x_0, \ldots, x_{n-1}]$ and the rows of $[v_1^c, \ldots, v_n^c]$. In other words, the columns of $\tilde{X}$ span the same space as the columns of $\tilde{X}_c$. Thus, the fitted values from the regressions of y_t on these two spaces are the same, that is,

$$\hat{\alpha} + \hat{\beta}' x_{t-1} + \left(\hat{\phi}^s\right)' \hat{v}_t = \hat{\alpha}^c + \left(\hat{\beta}^c\right)' x_{t-1} + \left(\hat{\phi}^c\right)' v_t^c, \tag{38}$$

for $t = 1, \ldots, n$, and $\hat{\phi}^c = \hat{\phi}^s$ for all ω. Now, using (37) to substitute for v_t^c on the right-hand side of (38), we obtain

$$\hat{\alpha} + \hat{\beta}' x_{t-1} + \left(\hat{\phi}^s\right)' \hat{v}_t = \left[\hat{\alpha}^c + \left(\hat{\phi}^s\right)'\left(\hat{\Theta} - \hat{\Theta}^c\right)\right] + \left[\left(\hat{\beta}^c\right)' - \left(\hat{\phi}^s\right)'\omega\right] x_{t-1} + \left(\hat{\phi}^s\right)' \hat{v}_t. \tag{39}$$

Since the least-squares coefficients are uniquely determined by the least-squares algorithm, the vectors of coefficients of x_{t-1} on the left- and right-hand sides of (39) must be the same, so that $\hat{\beta}' = (\hat{\beta}^c)' - (\hat{\phi}^s)'\omega$, and, therefore, $\hat{\beta}^c = \hat{\beta} + \omega'\hat{\phi}^s$. □

References

Amihud, Y. "Illiquidity and Stock Returns: Cross-Sectional and Time-Series Effects." *Journal of Financial Economics*, 5 (2002), 31–56.

Baker, M., and J. C. Stein. "Market Liquidity as a Sentiment Indicator." Working Paper, Harvard Business School (2002).

https://doi.org/10.1017/S0022109000003227 Published online by Cambridge University Press

Campbell, J. Y., and R. J. Shiller. "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors." *Review of Financial Studies*, 1 (1988), 195–228.
Dahlhaus, R. "Small Sample Effects in Time Series Analysis: A New Asymptotic Theory and a New Estimate." *Annals of Statistics*, 16 (1988), 808–841.
Fama, E. F. "Stock Returns, Expected Returns, and Real Activity." *Journal of Finance*, 45 (1990), 1089–1108.
Fama, E. F., and K. R. French. "Dividend Yields and Expected Stock Returns." *Journal of Financial Economics*, 22 (1988), 3–25.
_____. "Business Conditions and Expected Returns on Stocks and Bonds." *Journal of Financial Economics*, 25 (1989), 23–49.
Ferson, W. E.; S. Sarkissian; and T. T. Simin. "Spurious Regressions in Financial Economics?" *Journal of Finance*, 58 (2003), 1393–1413.
French, K. R.; G. W. Schwert; and R. F. Stambaugh. "Expected Stock Returns and Volatility." *Journal of Financial Economics*, 19 (1987), 3–29.
Fuller, W. A. *Introduction to Statistical Time Series*. New York, NY: John Wiley and Sons (1996).
Hodrick, R. "Dividend Yields and Expected Stock Returns: Alternative Procedures for Inference and Measurement." *Review of Financial Studies*, 5 (1992), 357–386.
Jones, C. M. "A Century of Stock Market Liquidity and Trading Costs." Working Paper, Columbia Business School (2002).
Judge, G. G.; R. C. Hill; W. E. Griffiths; H. Lutkepohl; and T. Lee. *Introduction to the Theory and Practice of Econometrics*. New York, NY: John Wiley and Sons (1985).
Keim, D. B., and R. F. Stambaugh. "Predicting Returns in the Stock and Bond Market." *Journal of Financial Economics*, 17 (1986), 357–396.
Kendall, M. G. "Note on Bias in the Estimation of Autocorrelation." *Biometrika*, 41 (1954), 403–404.
Kothari, S. P., and J. Shanken. "Book-to-Market, Dividend Yield, and Expected Market Returns: A Time-Series Analysis." *Journal of Financial Economics*, 18 (1997), 169–203.
Lewellen, J. "Predicting Returns with Financial Ratios." *Journal of Financial Economics* (forthcoming 2004).
Mankiw, N. G., and M. Shapiro. "Do We Reject Too Often? Small Sample Properties of Tests of Rational Expectations Models." *Economic Letters*, 20 (1986), 139–145.
Nelson, C. R., and M. J. Kim. "Predictable Stock Returns: The Role of Small Sample Bias." *Journal of Finance*, 48 (1993), 641–661.
Nicholls, D. F., and A. L. Pope. "Bias in the Estimation of Multivariate Autoregressions." *Australian Journal of Statistics*, 30A (1988), 296–309.
Pontiff, J., and L. D. Schall. "Book-to-Market Ratios as Predictors of Market Returns." *Journal of Financial Economics*, 49 (1998), 141–160.
Reinsel, G. C. *Elements of Multivariate Time Series Analysis*. New York, NY: Springer-Verlag (1997).
Sawa, T. "The Exact Moments of the Least Squares Estimator for the Autoregressive Model." *Journal of Econometrics*, 8 (1978), 159–172.
Simonoff, J. S. "The Relative Importance of Bias and Variability in the Estimation of the Variance of a Statistic." *The Statistician*, 42 (1993), 3–7.
Stambaugh, R. F. "Bias in Regressions with Lagged Stochastic Regressors." Working Paper, Univ. of Chicago (1986).
_____. "Predictive Regressions." *Journal of Financial Economics*, 54 (1999), 375–421.

https://doi.org/10.1017/S0022109000003227 Published online by Cambridge University Press