Predictive Regressions: A Reduced-Bias Estimation Method

Yakov Amihud[1] Clifford M. Hurvich[2]

November 26, 2002

[1]Department of Finance, Stern School of Business, New York University, New York NY 10012
[2]Statistics, Stern School of Business, New York University, New York NY 10012

The authors thank Robert Engle, Gary Simon and Jeffrey Simonoff for helpful comments.

Abstract

We propose a direct and convenient reduced-bias estimator of predictive regression coefficients, assuming that the regressors are Gaussian first-order autoregressive with errors that are correlated with the error series of the dependent variable. For the single-regressor model, Stambaugh (1999) shows that the ordinary least squares estimator of the predictive regression coefficient is biased in small samples. Our estimation method employs an augmented regression which uses a proxy for the errors in the autoregressive model. We also develop a heuristic estimator of the standard error of the estimated predictive coefficient which performs well in simulations, and show that the estimated coefficient of the errors and its squared standard error are unbiased. We analyze the case of *multiple* predictors that are first-order autoregressive and derive bias expressions for both the ordinary least squares and our reduced-bias estimated coefficients. The effectiveness of our estimation method is demonstrated by simulations.

Keywords: Stock Returns; Dividend Yields; Autoregressive Models.

1 Introduction

In a recent paper, Stambaugh (1999) shows that there is a bias in the parameter estimation of a standard model that is used in finance and economics. Consider first the following model 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

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

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

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

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

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 ordinary least squares (OLS) estimator of β based on a finite sample will be biased.

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

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

where $\phi = \sigma_{uv}/\sigma_v^2$, and $\hat{\beta}$ and $\hat{\rho}$ are the OLS estimators of β and ρ. 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 $E[\hat{\rho} - \rho] = -(1 + 3\rho)/n + O(n^{-2})$. Applying Stambaugh's result (3), researchers use a bias-corrected estimator of β, which we denote by $\hat{\beta}^s$ as

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

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.[1]

However, there is as yet no theoretical justification for this method of estimation. There is no obvious reason why the sample estimators $\hat{\phi}_s$ and $\hat{\rho}$, 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.[2]

Furthermore, Stambaugh's (1999) analysis is for 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.

In this paper, we propose and derive the properties of reduced-bias estimators, based on augmented regressions, for the vector β in a multiple-predictor generalization of the model (1) and (2). 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

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

[2] Although no one has heretofore explored the theoretical properties of $\hat{\beta}^s$, it turns out that this estimator is closely related to ours and under some specification exactly equal to ours. We are indebted to Gary Simon for producing a proof of this claim. The proof is available on request.

a reduced-bias estimator of the AR parameter. Our method can be used in multiple-predictor models to reduce the bias of the OLS estimator of β. Naturally, our method applies as well in the single-predictor model as a special case. Our proposed estimation method is straightforward and easily implemented. In fact, it is the only direct reduced-bias method available in the literature for the case of multiple predictive variables.

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, among other things, 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 $\hat{\beta}^s$, based on a second-order generalization of Kendall's (1954) formula. We also propose a formula to directly obtain an estimator of the standard error of the bias-corrected estimator of β, which enables us to easily construct confidence intervals and do hypothesis testing. This formula is applicable in the single-predictor case and under one specification of the multi-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.[3]

In addition, our estimation method provides an unbiased estimate of ϕ, which may be useful in the following context. When 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

[3]See Kothari and Shanken (1997).

of x_t on y_t in the model

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

The coefficient β measures[4] 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. We prove that our method provides a reduced-bias estimator of β and an unbiased estimator of ϕ, and that the latter estimator's squared standard error, obtained directly from the regression output, is also unbiased.

The case of multiple predictive variables is presented by a general model in which the predictive variables form a Gaussian multivariate AR(1) series. The analysis is based on a natural generalization of our univariate reduced-bias estimation method, employing a regression which 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 this matrix can be used to produce corresponding improvements in the bias of the coefficients of the predictive variables. We also provide a theoretical expression for the bias in the OLS estimator of β, generalizing Stambaugh's formula (3) to the multiple-predictor case.

The usefulness of our estimation method is demonstrated by simulations for both the single-predictor and the multiple-predictor cases. In implementing our estimators in the

[4]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 $\beta = \delta_1 \rho$.

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 itself follows a univariate AR(1) model. However, the predictive variables can be correlated through the covariance matrix of the errors. In this case, the implementation of our method is simple, and it performs just as well as in the univariate case. In contrast, there is no direct application of Stambaugh's formula in the case of multiple predictors. Finally, we consider the general case where the AR(1) parameter matrix is not constrained to be diagonal. In this case, we construct an estimate of a bias expression for multivariate AR(1) models due to Nicholls and Pope (1988). This indeed reduces the bias, but since the expressions are more complex and more parameters need to be estimated (we use a small sample size), there is some degradation in performance compared to the diagonal case.

Our paper proceeds as follows. In section 2 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 3 describes a heuristic method for estimating the standard error of the estimated predictive regression coefficient. Section 4 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 5, and in section 6 we demonstrate the use of our method in estimating a common predictive model in finance: dividend yield as predictor of expected stock return. Our conclusions are in section 7.

2 Reduced-Bias Estimation of the Regression Coefficient

Stambaugh (1999) shows that given models (1) and (2), the ordinary least squares (OLS) estimator $\hat{\beta}$ has bias $E[\hat{\beta} - \beta] = \phi E[\hat{\rho} - \rho]$, where $\phi = \sigma_{uv}/\sigma_v^2$, and $\hat{\rho}$ is the OLS estimator of ρ based on $x_1, \ldots, x_n$. This expression is exact, for any given sample size n. The expression states that the bias in the OLS estimator of β is proportional to the bias in the OLS estimator of ρ. Thus, if ϕ is large or $\hat{\rho}$ is appreciably biased, $\hat{\beta}$ will be strongly biased as well.

To motivate our proposed reduced-bias estimator of β, we consider first an infeasible estimator, $\tilde{\beta}$, 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 $t = 1, \ldots, n$. It is shown in the appendix that we can write

$$y_t = \alpha + \beta x_{t-1} + \phi v_t + e_t \quad , \tag{6}$$

where $\{e_t\}_{t=1}^n$ are independent and identically distributed 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,*

$$E[\tilde{\beta}] = \beta \quad .$$

Proof: See 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 $\hat{\beta}^c$ 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

$$v_t^c = x_t - (\hat{\theta}^c + \hat{\rho}^c x_{t-1}) \quad , \tag{7}$$

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 will be seen, 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*

$$E[\hat{\beta}^c - \beta] = \phi E[\hat{\rho}^c - \rho] \quad ,$$

where $\phi = \sigma_{uv}/\sigma_v^2$.

Proof: See appendix.

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 will 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

$$\hat{\rho}^{c,2} = \hat{\rho} + (1 + 3\hat{\rho})/n + 3(1 + 3\hat{\rho})/n^2. \tag{8}$$

The estimator $\hat{\rho}^{c,1}$ was studied by Sawa (1978), and has bias which is $O(n^{-2})$. We note here the non-obvious fact that if $\hat{\rho}^{c,1}$ is used in constructing the proxy for v_t in the augmented regression, the resulting bias-corrected estimator $\hat{\beta}^c$ is identical to the plug-in estimator $\hat{\beta}^s$ derived from Stambaugh (1999).[5]

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 in fact smaller than that of $\hat{\rho}^{c,1}$. We will therefore restrict attention henceforth to $\hat{\rho}^{c,2}$, which we denote by $\hat{\rho}^c$, and we will henceforth denote the corresponding bias-corrected estimator of β by $\hat{\beta}^c$.

In summary, the procedure we propose for estimating β has two steps:[6]

(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 (7) above.

[5] Indeed, if $(1+3\hat{\rho})/n$ is replaced in (4) by an arbitrary estimator of the negative bias in $\hat{\rho}$, the resulting $\hat{\beta}^s$ can be shown to be equal to the version of our $\hat{\beta}^c$ that uses the corresponding bias-corrected $\hat{\rho}$ in constructing the proxy for v_t in the augmented regression.

[6] See an application of this procedure in Amihud (2002).

(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 appendix.

3 Estimation of Standard Errors

3.1 Standard Errors for $\hat{\beta}^c$

For the construction of valid confidence intervals and hypothesis tests for β, it follows from Simonoff (1993) that a low-bias finite-sample approximation to the standard error of $\hat{\beta}^c$ is needed. While the estimated standard error for $\hat{\beta}^c$ that we develop here is only heuristically motivated, we find that it performs well in simulations. Let $\hat{\sigma}^2$ denote the estimator of the error variance from a regression (with intercept) of y_t on x_{t-1} and v_t^c. Thus, $\hat{\sigma}^2$, which is readily available from standard linear regression programs, is simply the residual sum of squares from this regression, divided by $n-3$. It follows from the proof of Lemma 2 below that $\hat{\sigma}^2$ is a biased estimator of σ_u^2. Therefore, $\widehat{SE}(\hat{\beta}^c)$, the estimated standard error 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. However, feasible and reasonably accurate confidence intervals for β can be constructed, as we show here. The following lemma gives a useful result.

Lemma 2

$$E[\hat{\beta}^c - \beta]^2 = \phi^2 E[\hat{\rho}^c - \rho]^2 + E[\widehat{SE}(\hat{\beta}^c)]^2 \quad , \tag{9}$$

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).*

Proof: See appendix.

Since from Theorem 2

$$E[\hat{\beta}^c - \beta] = \phi E[\hat{\rho}^c - \rho] = O(1/n^2) \quad ,$$

we conclude from (9) that

$$var[\hat{\beta}^c] = E[\hat{\beta}^c - \beta]^2 + O(1/n^4) \quad , \tag{10}$$

so a low-bias estimate of the righthand side of (9) should provide a low-bias estimate of $var[\hat{\beta}^c]$. 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 above).

Next, we need to construct an estimator of $E[\hat{\rho}^c - \rho]^2$ with low bias. We will treat $\hat{\rho}^c$ as if it were unbiased. Then we simply need an expression for $Var(\hat{\rho}^c)$, where

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

Thus,

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

For the OLS estimator $\hat{\rho}$, our simulations indicate that an accurate estimator of $Var(\hat{\rho})$ is given by $\widehat{Var}(\hat{\rho})$, the square of the standard error (as given by standard regression packages), based on an OLS regression of $\{x_t\}_{t=1}^n$ on $\{x_{t-1}\}_{t=1}^n$, with intercept.[7] Thus, a feasible estimator for $Var(\hat{\rho}^c)$ is given by

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

Finally, our estimator for the standard error of $\hat{\beta}^c$ is given by

$$\widehat{SE}^c(\hat{\beta}^c) = \sqrt{\{\hat{\phi}^c\}^2 \widehat{Var}(\hat{\rho}^c) + \{\widehat{SE}(\hat{\beta}^c)\}^2} \quad . \tag{11}$$

3.2 **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. It was shown in Lemma 1 that $E[\hat{\phi}^c] = \phi$. We now consider the problem of estimating the standard error of $\hat{\phi}^c$. The following Lemma shows that the estimated squared standard error is unbiased.

Lemma 3

$$Var[\hat{\phi}^c] = E[\widehat{SE}(\hat{\phi}^c)]^2 \quad ,$$

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 appendix.

[7]Indeed, this readily-available estimator strongly outperforms the asymptotic approximation suggested in Fuller (1966) equation (6.2.9).

4 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 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 $\{x_t\}$ which evolves according to a Gaussian vector autoregressive $VAR(1)$ model. The overall model is given for $t = 1, \ldots, n$ by

$$y_t = \alpha + \beta' x_{t-1} + u_t \quad , \tag{12}$$

$$x_t = \Theta + \Phi x_{t-1} + v_t \quad , \tag{13}$$

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 satisfying the determinantal equation to ensure stationarity (see, e.g., Fuller, 1996). It follows from our assumptions that there exists a $p \times 1$ vector ϕ such that

$$u_t = \phi' v_t + e_t \quad , \tag{14}$$

where $\{e_t\}$ are i.i.d. normal random variables with mean zero, and $\{e_t\}$ is independent of both $\{v_t\}$ and $\{x_t\}$.

We first give the bias of $\hat{\beta}$, the OLS estimator of β in the model given by (12), (13), and (14), thereby generalizing (3) to the multiple-predictor case.

Theorem 3

$$E[\hat{\beta} - \beta] = E[\hat{\Phi} - \Phi]'\phi \quad , \tag{15}$$

where $\hat{\Phi}$ is the OLS estimator of Φ.

Proof: See appendix.

The remainder of our analysis of the multi-predictor case proceeds as follows. First, 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. Second, we develop a bias expression for our estimator of β and show that the bias in it is proportional to the bias in 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).

Specifically, suppose that $\hat{\Theta}^c$ and $\hat{\Phi}^c$ are any estimators of Θ and Φ constructed from $\{x_t\}_{t=0}^n$. Define a proxy $\{v_t^c\}$ for the error series $\{v_t\}$ by

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

To estimate β, we propose to run an OLS regression of y_t on all entries of the vectors x_{t-1} and v_t^c, together with a constant. Our proposed estimator $\hat{\beta}^c$ of β consists of the estimated coefficients of the entries of x_{t-1} in this augmented OLS regression. 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 4

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

Proof: See appendix.

If we define $\hat{\phi}^c$ to be the vector of OLS regression coefficients of the entries of v_t^c obtained in the augmented regression described above, then we have the following generalization of Lemma 1, which shows that $\hat{\phi}^c$ is unbiased for ϕ.

Lemma 4 *If $\{y_t\}$ is given by the multiple-predictor model (12) and (13) and $\hat{\phi}^c$ is as defined above, then*

$$E[\hat{\phi}^c] = \phi \quad . \tag{18}$$

Proof: See appendix.

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 $\hat{\beta}^c$ will hold for an estimator $\hat{\Phi}^c$ that is an arbitrary function of the series of predictor variables $\{x_t\}_{t=0}^n$. But as Theorem 4 shows, the bias of $\hat{\beta}^c$ is proportional to the bias of $\hat{\Phi}^c$, so we now focus on the choice of $\hat{\Phi}^c$. We give here two proposals for $\hat{\Phi}^c$, the first of which is applicable only in the case where it is known that the true AR(1) parameter matrix Φ is diagonal, and the second of which 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, entries of $\{x_t\}$ can still be contemporaneously correlated even under the assumption that Φ is diagonal if the covariance matrix $\Sigma_v = Cov(v_t)$ is non-diagonal.

If Φ is known to be diagonal, then each entry of $\{x_t\}$ is a univariate AR(1) process, and therefore we can treat each series separately, estimating its autoregressive coefficient by univariate OLS and then correcting this estimator as we have proposed for the single-predictor case. Then $\hat{\Phi}^c$ is constructed as a diagonal matrix, with diagonal entries given by the corrected univariate AR(1) parameter estimates. 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) for the bias in the OLS estimator $\hat{\Phi}$, that is, $E[\hat{\Phi} - \Phi]$. This expression, which has a remainder term of $O(n^{-3/2})$, can be found in Stambaugh (1999), Equation (54), and in Nicholls and Pope (1988), Theorem 2.[8] The expression for the bias in $\hat{\Phi}$ depends on the unknown Φ and Σ_v. We therefore estimate this bias expression by plugging in preliminary estimates of Φ and Σ_v. The bias-corrected estimator $\hat{\Phi}^c$ is then obtained by subtracting the estimated bias expression from the OLS estimator $\hat{\Phi}$.

[8]Stambaugh's expression contains a typographical error, and should be multiplied by $-1/n$.

The preliminary estimator of Σ_v is obtained as the sample covariance matrix of the residuals $x_t - \hat{\Theta} - \hat{\Phi}x_{t-1}$, where $\hat{\Theta}$ is the OLS estimator of Θ. It is important that the preliminary estimator of Φ have all eigenvalues less than one, i.e., that it correspond to a stationary multivariate AR(1) model. Therefore, for this preliminary estimator we use $\hat{\Phi}$ if it satisfies this condition, and otherwise we use the Yule-Walker estimator of Φ (see Fuller 1996, p. 78), which is guaranteed to satisfy this condition. Iterative versions of our estimation scheme could be tried, but we will not pursue this here.

5 Simulations

5.1 **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. We simulate a total of 1500 replications from the model (1) and (2), with a sample size $n = 30$ and the following parameter values: $\theta = 0.2$, $\rho = 0.8$, $\alpha = 0$, $\beta = 1$, $\phi = -10$. This value for ϕ is achieved by constructing $u_t = \phi v_t + e_t$ with $\phi = -10$, where $\{v_t\}$ and $\{e_t\}$ are mutually independent i.i.d. standard normal random variables. The results are reported in Table 1. Standard errors estimated 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 (11). We now summarize our findings from Table 1.

INSERT TABLE 1 HERE

It can be seen that $\hat{\rho}$ is strongly negatively biased, but that the corrected estimator $\hat{\rho}^c$ is very nearly unbiased, at the cost of a slight inflation in its standard deviation. The estimated standard error $\widehat{SE}(\hat{\rho})$ has an average which is very nearly identical to the true standard deviation of $\hat{\rho}$. We have found, in these simulations and others, that $\widehat{SE}(\hat{\rho})$ is much more accurate in small sample sizes than what would be obtained from using asymptotic expressions for the standard error, such as that given in Fuller (1996, page 318, equation 6.2.9). It follows that the estimated standard error for $\hat{\rho}^c$ (not shown), obtained as $(1+3/n+9/n^2)\widehat{SE}(\hat{\rho})$ is a very nearly unbiased estimate for the true standard deviation of $\hat{\rho}^c$.

Next, we observe that $\hat{\beta}$ is strongly positively biased: the average $\hat{\beta}$ is 2.1646 while $\beta = 1.0$. This is predicted by Stambaugh's (1999) formula $E[\hat{\beta} - \beta] = \phi E[\hat{\rho} - \rho]$, given that both ϕ and the bias in $\hat{\rho}$ are negative.[9] The estimated standard error for $\hat{\beta}$ is within 10% of the true standard deviation.

Our corrected estimator $\hat{\beta}^c$ has a very small bias: the bias is only .046. The actual and theoretical biases match exactly: the bias predicted by our Theorem 2, using simulation means as if they were population means, is $-10.0023(0.7954 - 0.8) = 0.046$. There is no reason in principle for an exact match, however, and indeed the match was not exact in other simulations not shown here.

[9] In fact, using the simulation results we estimate the bias of $\hat{\rho}$ to be $0.68354 - 0.8 = -0.11646$, and plugging this into Stambaugh's formula (3) we should get a corresponding bias in $\hat{\beta}$ of 1.1646. The actual bias in $\hat{\beta}$ from the simulations is $2.1646 - 1 = 1.1646$, an exact match with Stambaugh's equation.

The standard error $\widehat{SE}(\hat{\beta}^c)$ obtained from the regression output greatly underestimates the true standard deviation. This is because $\widehat{SE}(\hat{\beta}^c)$ estimates only the square root of the second term of (9), but ignores the first term, which is much larger than the second term for the parameter configuration and sample size studied here. The corrected estimator $\widehat{SE}^c(\hat{\beta}^c)$ obtained from (11), which takes into account both terms of (9), is much more accurate, having a mean which is within 8% 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 $\hat{\phi}^c$.

5.1 **multiple-predictor model**

Simulations of multiple-predictor models given by (12), (13), and (14) 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 1500 replications with sample size $n = 30$. We use two predictive variables $x_{i,t}$, $i = 1, 2$, with parameter values similar to those used in the simulations for the single-predictor case.

In the simulations, the values of the parameters and the construction of the variables are as follows. $\alpha = 0$, $\beta = (1, 1)'$, $\Theta = (0, 0)'$, $u_t = \phi' v_t + e_t$, the e_t are independent standard normal, $\phi = (\phi_1, \phi_2)' = (-10, -10)'$, 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.

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

$$\Phi = \begin{pmatrix} \rho & 0 \\ 0 & \rho \end{pmatrix} ,$$

with $\rho = .8$. We employ two covariance matrices for the errors of the predictive variables. The first is

$$\Sigma_{1v} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} ,$$

and the second is

$$\Sigma_{2v} = \begin{pmatrix} 10 & 9 \\ 9 & 10 \end{pmatrix} .$$

The estimation procedure for the models in Panel A is as follows:

(I) For each component $x_{1,t}$ and $x_{2,t}$, estimate the univariate AR(1) model (2) by OLS and obtain $\hat{\rho}_1$ and $\hat{\rho}_2$. Construct the corrected estimators $\hat{\rho}_1^c = \hat{\rho}_1 + (1 + 3\hat{\rho}_1)/n + 3(1 + 3\hat{\rho}_1)/n^2$ and $\hat{\rho}_2^c = \hat{\rho}_2 + (1 + 3\hat{\rho}_2)/n + 3(1 + 3\hat{\rho}_2)/n^2$ and obtain the corrected residuals $v_{1,t}^c = x_{1,t} - \hat{\theta}_1^c - \hat{\rho}_1^c x_{1,t-1}$ and $v_{2,t}^c = x_{2,t} - \hat{\theta}_2^c - \hat{\rho}_2^c x_{2,t-1}$, where $\hat{\theta}_1^c$ and $\hat{\theta}_2^c$ are the adjusted intercepts.

(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 $v_{2,t}^c$.

We estimate the corrected standard error for $\hat{\beta}_1^c$ and $\hat{\beta}_2^c$, denoted by $\widehat{SE}^c(\hat{\beta}_i^c)$, using (11) and employing the respective parameter estimates.

INSERT TABLE 2 HERE

The estimation results for the diagonal-Φ two-predictor model are presented in Table 2, Panel A. Consider first the results for Σ_{1v}. Notably, the correlation between the two predictive variables is quite high, $Corr(x_{1,t}, x_{2,t}) = 0.48$. Thus, although we assume a diagonal matrix Φ, our specification generates a high correlation between the two predictors.

We focus on the estimates of the coefficients β_1 and β_2. The OLS estimates are highly biased. Whereas $\beta_1 = \beta_2 = 1$, we find that the average values of $\hat{\beta}_1$ and $\hat{\beta}_2$ are 2.53 and 2.48, respectively. This is quite a large bias, and larger than that in the simulations of the single-predictor model in Table 1. By contrast, the average values of $\hat{\beta}_1^c$ and $\hat{\beta}_2^c$ are 1.07 and 1.07, quite close to the true values. In keeping 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 almost exactly equal to -10.

Not only do the estimates $\hat{\beta}_i^c$ for i= 1 and 2 have very small bias, they also have far smaller standard errors than the OLS estimates $\hat{\beta}_i$. Specifically, the standard errors of $\hat{\beta}_i^c$ are *less than half* the standard error of the OLS estimates. Thus, not only are our estimates almost unbiased compared to the highly biased OLS estimators, they are also far more efficient. Our approximation method for the estimation of the standard errors works quite well. We obtain $\widehat{SE}^c(\hat{\beta}_1^c)$ =1.49 and $\widehat{SE}^c(\hat{\beta}_2^c)$ =1.48 compared to actual standard errors of 1.57 and 1.62, respectively. That is, our estimates are 5% to 9% smaller than the actual standard errors.

Under the covariance matrix Σ_{2v} there is a much greater correlation between the two predictors: $Corr(x_{1,t}, x_{2,t}) = 0.89$. The bias in the OLS predictive coefficients $\hat{\beta}_1$ and $\hat{\beta}_2$ is similar to that under Σ_{1v}, but the increase in the variance and covariance terms in Σ_{2v} greatly increases the standard errors of the OLS estimates of β. However, the standard errors of both our reduced-biased estimates of the entries of β remain similar to those under Σ_{1v}. The notable effect of the change in the covariance matrix is on the efficiency of the OLS estimation versus ours. The standard error of our reduced-bias estimates is one fifth (!) of the standard error of the OLS estimates. This shows that our reduced-bias estimates are quite efficient.

Panel B presents results for a non-diagonal AR(1) parameter matrix

$$\Phi = \begin{pmatrix} .7 & .1 \\ .1 & .7 \end{pmatrix} ,$$

and

$$\Sigma_v = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} .$$

The closer the largest eigenvalue of Φ is to 1, the more nearly nonstationary the multivariate AR(1) model is. The matrix Φ given above has its largest eigenvalue equal to .8, in keeping with all of the other simulations we have done. The structure of Φ accommodates contemporaneous correlation between the predictive variables even when Σ_v is diagonal.

Our estimation procedure for the results in Panel B is as follows:

(I) Construct the bias-corrected AR(1) parameter matrix estimate $\hat{\Phi}^c$ using the method of Nicholls and Pope (1988), suggested in Stambaugh (1999). (See the end of the previous section for more details on the implementation of $\hat{\Phi}^c$.) Next, construct the bivariate corrected residual series $v_t^c = x_t - \hat{\Theta}^c - \hat{\Phi}^c x_{t-1}$ 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 $v_{2,t}^c$.

We obtain that the OLS estimates $\hat{\beta}_1$ and $\hat{\beta}_2$ are strongly biased, in agreement with Theorem 3. The average values for $\hat{\beta}_1$ and $\hat{\beta}_2$ are 2.37 and 2.34, respectively. The corrected estimators $\hat{\beta}_1^c$ and $\hat{\beta}_2^c$ are less biased, averaging to 1.31 and 1.29, respectively. This is in agreement with Theorem 4. It should be noted that the bias reduction here is not as great as in the case where Φ is known to be diagonal. The problem is that the Nicholls-Pope bias-corrected estimator $\hat{\Phi}^c$ still yields an appreciably biased estimator. For example, for the (1,1) entry of Φ, which is 0.7, the OLS estimator $\hat{\Phi}_{11}$ averages to 0.567, while the corrected estimator $\hat{\Phi}_{11}^c$ averages to 0.667, indicating that the bias has not been completely removed. In this regard, it should be noted that the implementation of the Nicholls-Pope bias-corrected estimator of Φ requires the estimation of several additional parameters in comparison to the Kendall method (8). This is a particularly severe problem when the sample size is as small as that considered here ($n = 30$). However, the Kendall method is not applicable in the present case where Φ is not diagonal.

The standard errors of the corrected estimators $\hat{\beta}_1^c$ and $\hat{\beta}_2^c$ are approximately 30 percent larger than those of the OLS estimators $\hat{\beta}_1$ and $\hat{\beta}_2$. This may be attributed to the fact that the standard errors of the entries of $\hat{\Phi}^c$ are larger than those of $\hat{\Phi}$. In further simulations not shown here, we tried increasing the off-diagonal entries of Σ_v. The effect of this is to further increase the standard error of both $\hat{\Phi}^c$ and $\hat{\beta}^c$ relative to those of $\hat{\Phi}$ and $\hat{\beta}$.

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.

Overall, in the case of non-diagonal Φ, we find that our method provides bias reduction in estimation of β compared to OLS, but at the cost of a potentially substantial increase in the standard error. Future improvements on our implementation of the Nicholls-Pope bias-correction methodology for estimating Φ could lead to improved performance of the corresponding corrected estimator $\hat{\beta}^c$, in terms of both bias and standard error.

6 Empirical Illustration

In this section, we illustrate our estimation method using a common model of predictive regression that was studied by Stambaugh (1999). Following Kothari and Shanken (1997)[10], we estimate a model where annual stock market return is predicted by the market's dividend yield at the beginning of the year:

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

RM_t is the real (inflation-adjusted) value-weighted annual market return for year t

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

(from April of year $t-1$ to March of year t), and $DIVY_{t-1}$ is the corresponding 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). The dividend yield $DIVY_t$ is assumed to be an AR(1) process

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

Estimates are conducted over three short subperiods of 30 years each to highlight the problem of estimation from a small sample. The series over the period[11] 1934-1991 are split into two equal and (almost) nonoverlapping periods of 30 years each, 1934-1963 and 1962-1991. In addition, we pick a middle period of 30 years, 1953-1982. We follow the procedure in Section 2. The estimation results are presented in Table 3.

INSERT TABLE 3 HERE

(a) 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.

(b) We do a bias-correction of $\hat{\rho}$

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

where n =30 is the sample size. This is reported in Table 3, line 2.

(c) Using these parameters, we calculate the corrected intercept $\hat{\theta}^c$ and corrected residual v_t^c for model (E2):

[11] We start in 1934 because Kothari and Shanken (1997) indicate a problem with extreme observations in 1933.

(E3.2) $\hat{\theta}^c = (1 - \hat{\rho}^c) \sum_{t=1}^{n} DIVY_t / n.$

(E3.3) $v_t^c = DIVY_t - (\hat{\theta}^c + \hat{\rho}^c DIVY_{t-1})$.

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

(e) Using v_t^c from (E3.3), we estimate the augmented model:

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

From this estimation we obtain the parameters $\hat{\beta}^c$ (Table 3, line 4) and $\hat{\phi}^c$ (line 6), their respective standard errors from this regression, $\widehat{SE}(\hat{\beta}^c)$ and $\widehat{SE}(\hat{\phi}^c)$ and t -statistics.

(f) The corrected standard error of $\hat{\beta}^c$, $\widehat{SE}^c(\hat{\beta}^c)$, is calculated according to (11) as follows:

(E6) $\widehat{SE}^c(\hat{\beta}^c) = \sqrt{(\hat{\phi}^c)^2 \{\widehat{SE}(\hat{\rho})\}^2 (1 + 3/n + 9/n^2)^2 + \{\widehat{SE}(\hat{\beta}^c)\}^2}$.

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

The estimation results in Table 3 show that $\hat{\beta}$ is biased upward because $\phi < 0$ (line 5) and $\hat{\rho} < \hat{\rho}^c$ (lines 1 and 2). Indeed, we obtain that $\hat{\beta}^c < \hat{\beta}$ (lines 3 and 4). Next, consider the bias in the standard error of $\hat{\beta}$. Lines 4 and 5 show that $\widehat{SE}^c(\hat{\beta}^c) > \widehat{SE}(\hat{\beta}^c)$. Therefore, in line 5, the null hypothesis $\beta = 0$ is not rejected nearly as strongly as it is in line 4 where the biased standard error is used.

7 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 which have a first order autoregressive structure and whose disturbance terms are contemporaneously correlated with that of the predicted variable. Stambaugh (1999) shows that for the case of a *single* predictor, 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 case of a single regressor, the sample estimated parameters are plugged into Stambaugh's bias expression. In the multi-predictor case, there heretofore exists neither an expression for the bias of the OLS estimators of the coefficients of the predictive variables, nor is there any direct reduced-bias estimation method.

This paper develops an estimation method for both the single-predictor and multi-predictor situations that produces a reduced-bias estimator of the coefficients of the lagged variables. For the single-predictor case, we also develop a straightforward estimation method for a reduced-bias standard error. Our method is particularly useful in the multi-predictor case for which there is no direct reduced-bias estimation method available, even in a "plug in" version.

8 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

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

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

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

Since e has zero mean and is independent of X, we obtain

$$E[\tilde{\beta}] = \beta \quad ,$$

thereby completing the proof □

Proof of Theorem 2: As in the proof of Theorem 1, we use the representation

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

where the error terms e_t are i.i.d. normal with mean zero, and for all t, e_t is independent of $x_0, \ldots, x_n$.

Let $\hat{\alpha}^c$, $\hat{\beta}^c$, $\hat{\phi}^c$ be the coefficients of the constant term, x_{t-1} and v_t^c, respectively, in an OLS regression (with intercept) of y_t on x_{t-1} and v_t^c for $t = 1, \ldots, n$. If $\{r_t\}_{t=1}^n$ is the sequence of residuals obtained in an OLS regression of x_{t-1} on v_t^c (with intercept), then we have

$$\hat{\beta}^c = \frac{\sum_{t=1}^n r_t y_t}{\sum_{t=1}^n r_t^2} \quad . \tag{19}$$

Since the residual vector is orthogonal to the vectors of explanatory variables, we have

$$\sum_{t=1}^n r_t = 0 \quad , \quad \sum_{t=1}^n r_t v_t^c = 0 \quad . \tag{20}$$

Writing $x_{t-1} = a_0 + a_1 v_t^c + r_t$, we obtain from (20) that

$$\sum_{t=1}^n r_t x_{t-1} = \sum_{t=1}^n r_t^2 \quad . \tag{21}$$

Therefore, from (19), we have

$$\hat{\beta}^c = \frac{1}{\sum_{t=1}^n r_t^2} \sum_{t=1}^n r_t(\alpha + \beta x_{t-1} + \phi v_t + e_t)$$

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

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 $r_1, \ldots, r_n$, and therefore

$$E\left[\frac{1}{\sum_{t=1}^n r_t^2} \sum_{t=1}^n r_t e_t\right] = 0 \quad . \tag{22}$$

From (20) and (21), we have

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

Thus,

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

Since $v_t - v_t^c = (\hat{\theta}^c - \theta) + (\hat{\rho}^c - \rho)x_{t-1}$, we find from (20) and (21) that

$$E[\hat{\beta}^c - \beta] = \phi E[\hat{\rho}^c - \rho] \quad ,$$

thereby completing the proof □

Proof of Lemma 1: 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, $e = u - \phi v$. Using the representation

$$y_t = \alpha + \phi(\hat{\theta}^c - \theta) + \beta x_{t-1} + \phi v_t^c + \phi(\hat{\rho}^c - \rho)x_{t-1} + e_t \quad ,$$

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

$$\hat{\phi}^c = \frac{\sum_{t=1}^n q_t y_t}{\sum_{t=1}^n q_t^2} = \phi + \frac{\sum_{t=1}^n q_t e_t}{\sum_{t=1}^n q_t^2} \quad . \tag{23}$$

Since $\{e_t\}$ is independent of $\{r_t\}$ and $E[e_t] = 0$, the expectation of the second term on the righthand side of the above equation is zero, so we obtain $E[\hat{\phi}^c] = \phi$ □

Proof of Lemma 2: Note first that

$$[\widehat{SE}(\hat{\beta}^c)]^2 = \frac{\hat{\sigma}^2}{\sum_{t=1}^n r_t^2} \quad ,$$

where $\{r_t\}_{t=1}^n$ is the sequence of residuals obtained in a simple OLS regression of x_{t-1} on v_t^c (with intercept). We use the error $e_t = u_t - \phi v_t$ as in the previous proofs. Note that the variance of e_t is $\sigma_e^2 = Var(e_t) = \sigma_u^2 - \sigma_{uv}^2/\sigma_v^2$. From the proof of Theorem 2, it can be seen that

$$\hat{\beta}^c - \beta = \phi(\hat{\rho}^c - \rho) + \frac{\sum_{t=1}^n r_t e_t}{\sum_{t=1}^n r_t^2} \quad . \tag{24}$$

The two terms on the righthand side of (24) are uncorrelated, and the second term has mean zero. It follows that

$$E[\hat{\beta}^c - \beta]^2 = \phi^2 E[\hat{\rho}^c - \rho]^2 + \sigma_e^2 E\left[\frac{1}{\sum_{t=1}^n r_t^2}\right] \quad .$$

It remains to be shown that

$$\sigma_e^2 E\left[\frac{1}{\sum_{t=1}^n r_t^2}\right] = E\left[\frac{\hat{\sigma}^2}{\sum_{t=1}^n r_t^2}\right] \quad . \tag{25}$$

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

$$\sum_{t=1}^n r_{0t}^2 = e'(I - H)e \sim \sigma_e^2 \chi_{n-3}^2 \quad ,$$

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

$$\hat{\sigma}^2 = \frac{1}{n-3} \sum_{t=1}^n r_{0t}^2$$

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

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

Taking expectations of both sides and using the double expectation theorem yields (25) □

Proof of Lemma 3: Note first that

$$[\widehat{SE}(\hat{\phi}^c)]^2 = \frac{\hat{\sigma}^2}{\sum_{t=1}^n q_t^2} \quad .$$

From (23), we obtain

$$Var[\hat{\phi}^c] = \sigma_e^2 E\left[\frac{1}{\sum_{t=1}^n q_t^2}\right] \quad . \tag{26}$$

Proceeding as in the proof of Lemma 2, we have

$$E\left[\frac{\hat{\sigma}^2}{\sum_{t=1}^n q_t^2} \mid X\right] = E\left[\frac{1}{n-3}\frac{e'(I-H)e}{\sum_{t=1}^n q_t^2} \mid X\right]$$
$$= \frac{1}{\sum_{t=1}^n q_t^2}\,\frac{1}{n-3}\,E[\sigma_e^2\chi_{n-3}^2] = \sigma_e^2\frac{1}{\sum_{t=1}^n q_t^2} \quad .$$

Taking expectations of both sides and using the double expectation theorem yields

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

The Lemma now follows from (26) □

Proof of Theorem 3: Using (12) and (14) we can write

$$y_t = \alpha + \beta' x_{t-1} + \phi' v_t + e_t \quad , \tag{27}$$

where $\{e_t\}$ has zero mean and is independent of both $\{v_t\}$ and $\{x_t\}$. The OLS estimators of α and β are given by

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

where

$$X = [1_n, (x_0, x_1, \dots\, x_{n-1})']$$

is an $n \times (p+1)$ matrix of predictor variables, and $y = (y_1, \dots, y_n)'$. The OLS estimators of Θ and Φ are given by

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

a $(p+1) \times p$ matrix, where $x = (x_1, \ldots, x_n)'$ is $n \times p$. In vector form, we can write (27) as

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

where $v = (v_1, \ldots, v_n)'$ is $n \times p$, and $e = (e_1, \ldots, e_n)'$ is $n \times 1$. Thus,

$$\begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} - \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = (X'X)^{-1}X'v\phi + (X'X)^{-1}X'e \quad . \tag{28}$$

Similarly, since

$$x = X \begin{pmatrix} \Theta' \\ \Phi' \end{pmatrix} + v \quad ,$$

we have

$$\begin{pmatrix} \hat{\Theta}' \\ \hat{\Phi}' \end{pmatrix} - \begin{pmatrix} \Theta' \\ \Phi' \end{pmatrix} = (X'X)^{-1}X'v \quad . \tag{29}$$

Taking the expectation of (28) gives

$$E\left[\begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} - \begin{pmatrix} \alpha \\ \beta \end{pmatrix}\right] = E[(X'X)^{-1}X'v]\phi \quad .$$

Taking the expectation of (29) gives

$$E\left[\begin{pmatrix} \hat{\Theta}' \\ \hat{\Phi}' \end{pmatrix} - \begin{pmatrix} \Theta' \\ \Phi' \end{pmatrix}\right] = E[(X'X)^{-1}X'v] \quad .$$

Thus,

$$E\left[\begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} - \begin{pmatrix} \alpha \\ \beta \end{pmatrix}\right] = E\left[\begin{pmatrix} \hat{\Theta}' \\ \hat{\Phi}' \end{pmatrix} - \begin{pmatrix} \Theta' \\ \Phi' \end{pmatrix}\right]\phi \quad .$$

In particular, considering the final p entries of this equation, we obtain

$$E[\hat{\beta} - \beta] = E[\hat{\Phi} - \Phi]'\phi \quad \square$$

Proof of Theorem 4: Using (12), (13), (14) and (16), we can write

$$y_t = \tilde{\alpha} + \{\beta' + \phi'(\hat{\Phi}^c - \Phi)\}x_{t-1} + \phi' v_t^c + e_t \quad , \tag{30}$$

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

$$\hat{\beta}^c = \sum_{t=1}^n \tilde{r}_t y_t \quad , \tag{31}$$

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

$$\sum_{t=1}^n \tilde{r}_{jt} = \sum_{t=1}^n \tilde{r}_{jt}\, x_{k,t-1} = \sum_{t=1}^n \tilde{r}_{jt}\, v_{jt}^c = \sum_{t=1}^n \tilde{r}_{jt}\, v_{kt}^c = 0 \quad , \tag{32}$$

and

$$\sum_{t=1}^n \tilde{r}_{jt}\, x_{j,t-1} = \sum_{t=1}^n \tilde{r}_{jt}\, r_{jt} = 1 \quad . \tag{33}$$

Substituting y_t from (30) in (31) and using (32) and (33) yields

$$\hat{\beta}^c = \beta + (\hat{\Phi}^c - \Phi)'\phi + \sum_{t=1}^n \tilde{r}_t e_t \quad . \tag{34}$$

Since e_t has mean 0 and is independent of $\tilde{r}_t$, the expectation of the final term in (34) is zero, and after taking expectations of both sides of (34), we obtain

$$E[\hat{\beta}^c - \beta] = E[\hat{\Phi}^c - \Phi]'\phi \quad \square \tag{35}$$

Proof of Lemma 4: Define the $p \times 1$ vectors $\{q_t\}_{t=1}^n$ by $q_t = (q_{1t}, \ldots, q_{pt})'$ where for $j = 1, \ldots n$, $\{q_{jt}\}_{t=1}^n$ is the (row) vector of residuals from a $2p-1$-variable OLS regression of the j'th entry of v_t^c on all other entries of v_t^c as well as all p entries of x_{t-1} and an intercept. Correspondingly, define $\{\tilde{q}_t\}_{t=1}^n$ by $\tilde{q}_t = (q_{1t}/\Sigma\, q_{1t}^2, \ldots, q_{pt}/\Sigma\, q_{pt}^2\ )'$ and write $x_t = (x_{1t}, \ldots, x_{pt})'$, and $v_t^c = (v_{1t}^c, \ldots, v_{pt}^c)'$. It follows that

$$\hat{\phi}^c = \sum_{t=1}^{n} \tilde{q}_t y_t \quad , \tag{36}$$

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

$$\sum_{t=1}^{n} \tilde{q}_{jt} = \sum_{t=1}^{n} \tilde{q}_{jt}\, v_{kt}^c = \sum_{t=1}^{n} \tilde{q}_{jt}\, x_{j,t-1} = \sum_{t=1}^{n} \tilde{q}_{jt}\, x_{k,t-1} = 0 \quad , \tag{37}$$

and

$$\sum_{t=1}^{n} \tilde{q}_{jt}\, v_{jt}^c = \sum_{t=1}^{n} \tilde{q}_{jt}\, q_{jt} = 1 \quad . \tag{38}$$

Substituting y_t from (30) in (36) and using (37) and (38) yields

$$\hat{\phi}^c = \phi + \sum_{t=1}^{n} \tilde{q}_t e_t \quad . \tag{39}$$

Since e_t has mean 0 and is independent of $\tilde{q}_t$, the expectation of the final term in (39) is zero, and after taking expectations of both sides of (39), we obtain

$$E[\hat{\phi}^c] = \phi \quad \Box \tag{40}$$

References

[1] Amihud, Y., 2002, Illiquidity and stock returns: cross-sectional and time-series effects. *Journal of Financial Economics* **5**, 31-56.

[2] Dahlhaus, R., 1988, Small sample effects in time series analysis: a new asymptotic theory and a new estimate. *Annals of Statistics* **16**, 808-841.

[3] Fuller, W.A., 1996, *Introduction to Statistical Time Series, Second Edition.* New York: Wiley.

[4] Kendall, M.G., 1954, Note on bias in the estimation of autocorrelation. *Biometrika* **41**, 403-404.

[5] Kothari, S.P. and J. Shanken, 1997, Book-to-market, dividend yield, and expected market returns: A time-series analysis. *Journal of Financial Economics* **44**, 169-203.

[6] Nelson, C.R. and M.J. Kim, 1993, Predictable stock returns: the role of small sample bias. *Journal of Finance* **48**, 641-661.

[7] Nicholls, D.F. and A.L. Pope, 1988, Bias in the estimation of multivariate autoregressions. *Australian Journal of Statistics* **30A**, 296-309.

[8] Sawa, T., 1978, The exact moments of the least squares estimator for the autoregressive model. *Journal of Econometrics* **8**, 159-172.

[9] Simonoff, J.S., 1993, The relative importance of bias and variability in the estimation of the variance of a statistic. *The Statistician* **42**, 3-7.

[10] Stambaugh, R.F., 1999, Predictive Regressions. *Journal of Financial Economics* **54**, 375-421.

Table 1: Simulation results for regression model (1) and (2) with one predictive variable

1500 replications from the single-predictor models

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

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

The sample size is $n = 30$. The values of the parameters and the construction of the variables are as follows: $\theta = 0.2$, $\rho = 0.8$, $\alpha = 0$, $\beta = 1$, $u_t = \phi v_t + e_t$ with $\phi = -10$ and $\{v_t\}$ and $\{e_t\}$ are mutually independent i.i.d. standard normal random variables. The table presents estimation results of the single-predictor model by OLS as well as by our estimation procedure.

Our estimation procedure 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 (11).

Table 1: Results for the single-predictor model (1) and (2)

	Mean	Std Dev
$\hat{\rho}$	0.68354	0.144900
$\widehat{SE}(\hat{\rho})$	0.14938	0.027022
$\hat{\rho}^c$	0.79539	0.160840
$\hat{\beta}$	2.16466	1.457300
$\widehat{SE}(\hat{\beta})$	1.35350	0.247760
$\hat{\beta}^c$	1.04597	1.615370
$\widehat{SE}(\hat{\beta}^c)$	0.14091	0.037945
$\widehat{SE}^c(\hat{\beta}^c)$	1.50131	0.274680
$\hat{\phi}^c$	-10.00231	0.198450
$\widehat{SE}(\hat{\phi}^c)$	0.19491	0.037140

Table 2: Simulation results for a model with multiple predictive variables

1500 replications from the models

$$y_t = \alpha + \beta' x_{t-1} + u_t \quad , \tag{12}$$

$$x_t = \Theta + \Phi x_{t-1} + v_t \quad . \tag{13}$$

The sample size is $n = 30$. There are two predictors. The values of the parameters and the construction of the variables are as follows: $\alpha = 0$, $\beta = (1, 1)'$, $\Theta = (0, 0)'$, $u_t = \phi' v_t + e_t$, the e_t are independent standard normal, $\phi = (\phi_1, \phi_2)' = (-10, -10)'$, 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.

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

$$\Phi = \begin{pmatrix} \rho & 0 \\ 0 & \rho \end{pmatrix} \quad ,$$

with $\rho = .8$. Results are presented for two covariance matrices:

$$\Sigma_{1v} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \quad .$$

and

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

Panel B presents results for a non-diagonal AR(1) parameter matrix

$$\Phi = \begin{pmatrix} .7 & .1 \\ .1 & .7 \end{pmatrix} \quad ,$$

and

$$\Sigma_v = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \quad .$$

Our estimation procedure for the results in Panel A is as follows:

(I) For each component $x_{1,t}$ and $x_{2,t}$, estimate the univariate AR(1) model (2) by OLS and obtain $\hat{\rho}_1$ and $\hat{\rho}_2$. Construct the corrected estimators $\hat{\rho}_i^c = \hat{\rho}_i + (1 + 3\hat{\rho}_i)/n + 3(1 + 3\hat{\rho}_i)/n^2$, i =1 and 2, and obtain the corrected residuals $v_{i,t}^c = x_{i,t} - \hat{\theta}_i^c - \hat{\rho}_i^c x_{i,t-1}$, where $\hat{\theta}_i^c$ is the adjusted intercept.

(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 $v_{2,t}^c$.

Standard errors are shown for the two autoregressive coefficients and for the two β coefficients. Standard errors that are estimated directly from OLS regression output are denoted by $\widehat{SE}$. The corrected standard errors for $\hat{\beta}_i^c$ are denoted by $\widehat{SE}^c(\hat{\beta}_i^c)$, as given by (11).

Panel A: Results with diagonal autoregressive matrix Φ

	Results for Σ_{1v}		Results for Σ_{2v}	
	Mean	Std Dev	Mean	Std Dev
$Corr(x_1, x_2)$	0.47812	0.26081	0.88823	0.080110
$\hat{\rho}_1$	0.68093	0.14159	0.68131	0.15099
$\widehat{SE}(\hat{\rho}_1)$	0.13380	0.024649	0.13336	0.025155
$\hat{\rho}_1^c$	0.79249	0.15716	0.79292	0.16759
$\hat{\rho}_2$	0.68137	0.14550	0.68584	0.14463
$\widehat{SE}(\hat{\rho}_2)$	0.13324	0.024493	0.13285	0.024479
$\hat{\rho}_2^c$	0.79299	0.16150	0.79794	0.16054
$\hat{\beta}_1$	2.53264	3.45941	2.35615	7.90746
$\widehat{SE}(\hat{\beta}_1)$	2.87538	0.76788	6.43206	1.81446
$\hat{\beta}_1^c$	1.07288	1.57362	1.07419	1.68432
$\widehat{SE}(\hat{\beta}_1^c)$	0.12675	0.037893	0.11451	0.035915
$\widehat{SE}^c(\hat{\beta}_1^c)$	1.49123	0.27377	1.48545	0.28011
$\hat{\beta}_2$	2.48274	3.50385	2.60638	7.82477
$\widehat{SE}(\hat{\beta}_2)$	2.85577	0.73630	6.39806	1.83349
$\hat{\beta}_2^c$	1.06690	1.62275	1.01614	1.60598
$\widehat{SE}(\hat{\beta}_2^c)$	0.12596	0.037133	0.11391	0.036226
$\widehat{SE}^c(\hat{\beta}_2^c)$	1.48368	0.27232	1.47940	0.27149
$\hat{\phi}_1^c$	-10.00283	0.16404	-9.99990	0.15134
$\hat{\phi}_2^c$	-9.99360	0.16182	-9.99833	0.14949

Our estimation procedure for the results in Panel B is as follows:

(I) Construct the bias-corrected AR(1) parameter matrix estimate $\hat{\Phi}^c$ using the method of Nicholls and Pope (1988) as described in the text. 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: Results with non-diagonal autoregressive matrix Φ

	Mean	Std Dev
$\hat{\Phi}_{11}$	0.567259	0.168382
$\hat{\Phi}^c_{11}$	0.666770	0.195825
$\hat{\Phi}_{12}$	0.097958	0.175846
$\hat{\Phi}^c_{12}$	0.104370	0.227690
$\hat{\Phi}_{21}$	0.095680	0.177392
$\hat{\Phi}^c_{21}$	0.102141	0.227899
$\hat{\Phi}_{22}$	0.567612	0.166191
$\hat{\Phi}^c_{22}$	0.666601	0.194369
$\hat{\beta}_1$	2.369189	2.471143
$\hat{\beta}^c_1$	1.308748	3.201779
$\hat{\beta}_2$	2.340689	2.489899
$\hat{\beta}^c_2$	1.287487	3.239454
$\hat{\phi}^c_1$	-10.00009	0.147507
$\hat{\phi}^c_2$	-9.997331	0.145990

Table 3 Small-sample estimates of a regression of stock return on lagged dividend yield

The table presents results of the following models:

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

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

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

(E3.2) $\hat{\theta}^c = (1 - \hat{\rho}^c) \sum_{t=1}^{n} DIVY_t/n$.

(E3.3) $v_t^c = DIVY_t - (\hat{\theta}^c + \hat{\rho}^c DIVY_{t-1})$.

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

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

RM_t is the value weighted market real return for year t and $DIVY_{t-1}$ is the value weighted dividend yield for the preceding year. 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. [t] is the t-statistic.

	Coefficient	From model	Case 1: 1934-1963	Case 2: 1962-1991	Case 3: 1953-1982
1	$\hat{\rho}$	(E2)	0.448	0.7845	0.7184
	$(\widehat{SE}\hat{\rho})$ [t]	(OLS)	(0.1711) [2.62]	(0.1904) [4.12]	(0.1207) [5.95]
2	$\hat{\rho}^c$	(E3.1)	0.534	0.9075	0.8341
3	$\hat{\beta}$	(E1)	5.4062	7.7607	8.7435
	$(\widehat{SE}\hat{\beta})$ [t]	(OLS)	(2.99) [1.81]	(3.0428) [2.55]	(3.0916) [2.83]
4	$\hat{\beta}^c$	(E4)	4.1705	6.4835	5.9497
	$(\widehat{SE}\hat{\beta}^c)$ [t]		(1.7463) [2.39]	(2.3727) [2.73]	(1.0684) [5.57]
5	$(\widehat{SE}^c\hat{\beta}^c)$ [t]	(E5)	(3.1256) [1.33]	(3.1675) [2.05]	(3.2508) [1.83]
6	$\hat{\phi}^c$		-14.3771	-10.3867	-24.1489
	$(\widehat{SE}\hat{\phi}^c)$ [t]	(E4)	(1.9197) [2.39]	(2.3380) [4.44]	(1.6465) [14.67]