Indirect Inference on Predictive Regression

Biing-Shen Kuo*
National Chengchi University

Jhih-Gang Chen †
National Chengchi University

Oct. 2011

Abstract

In this paper, an indirect inference method is introduced to address the problem of estimation bias in predictive regression with a general AR(p) regressor. Simulation studies show that the proposed procedure works quite well to reduce bias, at little cost of increase in varaicne. As a result, the asymptotic *t*-test based on the indirect inference estimator is subject to much less size distortion problem than that of the least-squares counterpart.

KEY WORDS: Bias correction; Simulation-based estimation; Autoregression.

*Corresponding author, Dept. of International Business, National Chengchi University, Taipei 116, Taiwan; Tel: (02)29393091 ext. 81029, Fax: (02)29387699, Email: bsku@nccu.edu.tw.

†Dept. of International Business, National Chengchi University, Taipei 116, Taiwan; Tel: (02)29393091 ext. 81262, Fax: (02)29387699, E-mail: jgchen@nccu.edu.tw.

1 INTRODUCTION

This paper is concerned with the estimation of the predictive regression model which has applications that abound in empirical finance and economics. Specifically, the dependent variable in the model usually reflects an asset's price change, while the regressor is some associated lagged variable implied by economic theorems or hypotheses. For example, international finance researchers utilize the predictive regression to investigate the ability of monetary fundamentals to forecast future exchange rate returns, and to examine the long-run relationship between economic fundamentals and exchange rate (Mark, 1995; Kilian, 1999; Mark and Sul, 2001; Groen, 2000, 2005; Engel and West, 2005; Engel, Mark, and West, 2008; Molodtsova and Papell, 2009). In empirical finance, a continuing popular topic is whether the stock returns (either real or nominal) can be predicted by lagged financial ratios such as dividend-price ratio, book-to-market ratio and etc. (Fama and French, 1988; Campbell and Shiller, 1988; Hodrick, 1992; Lewellen, 2004; Welch and Goyal, 2008; Campbell and Thompson, 2008). In macroeconomics, Staiger, Stock, and Watson (1997) and Stock and Watson (1999) examined whether the inflation rate can be predicted by lagged unemployment rate as the well-known Phillips curve postulates.

While the structure of the predictive regression model is simple, the estimation can be troublesome. In usual, the predictor variable in the model displays strong autoregressive behavior, and the innovations in the predictive regression are contemporarily correlated with those from the predictor processes. As pointed out by Mankiw and Sharpio (1986) and Stambaugh (1986), these two common features together lead to biased least-squares estimate of the predictive coefficient in finite sample. Meanwhile, when the predictor variable is nearly or actually integrated, the asymptotic distribution of the least-squares estimator tends to be nonstandard. In consequence, the conventional t-test suffers from the problem of size distortion, and brings about misleading inference on the predictability.

Many efforts have been made in the literature to develop robust test methods. Mark (1995) and Kilian (1999) secure the critical values by bootstrapping the test statistics under the estimated null data generating process. Besides, as the predictor variable is persistent, some studies derive new tests under the framework of local-to-unity . Among these, Cavanagh, Elliott, and Stock (1995) proposed three asymptotically conservative tests based on sup-c, Bonferroni, and Scheffe-type confidence intervals. Torous, Valkanov, and Yan (2004) applied the Scheffe test to stock returns, and found evidence of predictability at shorter rather than at longer horizon.

Lewellen (2004) proposed a test that gains more power by combining the conditional and unconditional tests, that distinguished from the imposition of a unit root on the predictor variable. Campbell and Yogo (2006) developed a new efficient Bonferroni test that is asymptotically valid under fairly general assumptions on the dynamics of the predictor variable.

On the contrary, the literature seems to pay less concern to bias correction of the point estimation. Stambaugh (1986, 1999) derived the $O\left(T^{-1}\right)$ bias formula of the least-squares in the predictive regression model with an AR(1) predictor, and this formula is often utilized by researchers, such as Stambaugh (1999) himself, Kothari and Shanken (1997), and Lewellen (2004), to obtain bias-adjusted estimator. Under the same model, Amihud and Hurvich (2004) obtained bias-reduced estimation via an augmented regression in which the estimated innovations of the predictor variable are included into the predictive regression. However, as shown in this paper, many of the predictor variables used to predict stock returns act more like an AR(p) process, rather than an AR(1). In these cases, the existing AR(1)-assumed methods usually provide poor bias-correction.

So far, none has been done to correct the estimation bias in predictive regression models with a more general AR(p) regressor, except the recent work of Amihud, Hurvich, and Wang (2010). The bias-correction method suggested by Amihud, Hurvich, and Wang (2010) is exactly an extension of Amihud and Hurvich (2004), but the model being considered differs from those usually used in empirical research. Specifically, they allow the predictor variable to be characterized by an AR(p) model; meanwhile, all p lags (rather than single lag) of the predictor variable are also included in the predictive regression. Although they claim that fail to do so could result in extremely low asymptotic power for the predictability test, it remains interesting and important to develop a bias-correction method for the frequently-considered "single-lag" predictive regression model.

The present paper proposes a simulation-based method, named as "indirect inference" in the literature, to address the bias problem in the predictive regression model with an general AR(p) predictor. This methodology is first introduced by Smith (1993) and Gouriéroux, Monfort, and Renault (1993), and exhibits to be applicable to a wide range of structure models. Especially, it has been proven to be useful to reduce finite sample bias in various dynamic models (e.g. MacKinnin and Smith, 1998; Gouriéroux, Renault, and Touzi, 2000; Gouriéroux, Phillips, and Yu, 2007). Yet, to the best of our knowledge, this is the first application to predictive regression models.

Relative to conventional bias-correction methods, the main advantage of indirect inference is its capability to reduce bias without any explicit form of the bias function. Instead, the bias function is implicitly calibrated via simulation with an auxiliary model. This appears to be profoundly desirable under framework we considered because the least-squares bias function becomes much complicated when the AR-order (p) is large, and no exact or approximate bias formula has been provided in the literature.

Our simulation studies indicate that the indirect inference method actually provides a well bias correction to the least-squares, although little increases in variance arise as the cost in some cases. Besides, the indirect inference estimator is proven to be consistent and asymptotically normal, and a new t-test for the predictive coefficient is then derived. As a benefit from the bias reduction, the new test involves little size distortion in most cases considered, and leads to a more reliable inference.

The rest of the paper is organized as follows. Section 2 specifies the predictive regression model with an AR(p) regressor, and discusses the least-squares bias. Section 3 develops the indirect inference method and the corresponding asymptotic theory. Section 4 evaluates the performance of the indirect inference estimator relative to least-squares by means of simulation. Section 5 provides empirical illustrations of the indirect inference method by predicting S&P 500 equity returns with various lagged financial variables, and Section 6 concludes.

2 PREDICTIVE REGRESSION AND LEAST-SQUARES

We consider a predictive model usually seen in the financial literature where the dependent variable y_t is predicted by a lagged autoregressive variable x_{t-1}. The model is specified as

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

$$x_t = \rho_0 + \rho_1 x_{t-1} + \cdots + \rho_p x_{t-p} + v_t, \tag{2}$$

where both y_t and x_t are observed at $t = 0, 1, \cdots, T$, and the innovation vector, $(u_t, v_t)'$, is assumed to be serially independent and bivariate normal, i.e.

$$\begin{bmatrix} u_t \\ v_t \end{bmatrix} \overset{iid}{\sim} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{bmatrix} \right). \tag{3}$$

For stationarity, we further assume that all roots of the polynomial, $\lambda^p - \sum_{i=1}^{p} \rho_i \lambda^{p-i}$, lie within the unit circle. Let $y = (y_1, \cdots, y_T)'$, $X = [\iota_T, x]$, $x = (x_0, \cdots, x_{T-1})'$, and ι_T denote a $T \times 1$ vector of ones. Then the least-squares estimator of (1) is given by

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

and the finite-sample bias of $\hat{\beta}^{LS}$ has the expression in the following proposition.

Proposition 1. *Under the assumption that x_t is covariance-stationary, we have*

$$\mathrm{E}\left(\hat{\beta}^{LS} - \beta\right) = \frac{\sigma_{uv}}{\sigma_v^2}\mathrm{E}\left[\frac{\sum_{t=1}^{T}(x_{t-1} - \bar{x})v_t}{\sum_{t=1}^{T}(x_{t-1} - \bar{x})^2}\right] = \frac{\sigma_{uv}}{\sigma_v^2} f(\rho_1, \cdots, \rho_p, T), \tag{4}$$

where $\bar{x} = \sum_{t=1}^{T} x_{t-1}/T$, and $f(\cdot) = \mathrm{E}\left[\frac{\sum_{t=1}^{T}(x_{t-1}-\bar{x})v_t}{\sum_{t=1}^{T}(x_{t-1}-\bar{x})^2}\right]$ is $O(T^{-1})$.

Remark.

Observing the numerator and the denominator in the expectation term in (4) are correlated, the finite-sample bias of $\hat{\beta}^{LS}$ will generally be nonzero. Besides, the bias could be critical when the sample size is relatively small, and will tend to reduce the prediction accuracy and inference credibility in applications.

Consider the simplest case where x_t is AR(1), we have $v_t = x_t - \rho_0 - \rho_1 x_{t-1}$, and thus the expectation term in (4) equals the bias of the least-square estimator of ρ_1, i.e. $\mathrm{E}(\hat{\rho}_1^{LS} - \rho_1)$, which has an $O(T^{-1})$ approximation as $-(1+3\rho_1)/T$.[1] This leads to

$$\mathrm{E}\left(\hat{\beta}^{LS} - \beta\right) = -\frac{\sigma_{uv}}{\sigma_v^2}\left(\frac{1+3\rho_1}{T}\right) + O(T^{-2}),$$

which is also the result of Stambaugh (1999). Researchers may construct bias-corrected estimators according to the bias formula, and one example is that suggested by Kothari and Shanken (1998) which takes the form:

$$\hat{\beta}^{KS} = \hat{\beta}^{LS} + \frac{\hat{\sigma}_{uv}^{LS}}{\hat{\sigma}_v^{LS^2}}\left(\frac{1+3\hat{\rho}_1^{bc}}{T}\right), \tag{5}$$

where $\hat{\sigma}_{uv}^{LS}$ and $\hat{\sigma}_v^{LS^2}$ are estimated based on the least-squares residuals in (1) and (2), and $\hat{\rho}^{bc} = (T\hat{\rho}_1^{LS} + 1)/(T-3)$ is a bias-adjusted estimator of ρ_1. Although this kind of estimator is supposed to well correct the bias, it has the limitation that the predictor process is AR(1).

[1]The bias of the least-squares estimator in an AR(1) model has been derived by Marriott and Pope (1954) and Kendall (1954) to the order T^{-1}.

With a general AR(p) predictor, no exact or approximate form of the expectation term, $f(\rho_1, \cdots, \rho_p, T)$, has been derived, and bias-correction based on AR(1) assumption is expected to be less accurate. To see this, we perform some simulations to illustrate the bias performance of the AR(1)-based estimator, $\hat{\beta}^{KS}$, when the true process of the predictor is an AR(2). Table 1 details the settings of the simulations and shows the results. One thing noteworthy is that we fix the value of the largest characteristic root (λ_1) of the AR(2) process, and allow the other root (λ_2) to vary across simulations. Thus, the predictor process is actually AR(1) when $\lambda_2 = 0$, and is less like an AR(1) as λ_2 departs from zero. We might directly jump to the last column which depicts the absolute values of the relative bias ratio between $\hat{\beta}^{KS}$ and $\hat{\beta}^{LS}$. Roughly speaking, the values can be treat as the portion of the least-squares bias that is not corrected by $\hat{\beta}^{KS}$. The pattern is clear and verifies our concern that the far λ_2 is removed from zero, the less portion of the least-squares bias can be taken care by the AR(1)-based estimator, $\hat{\beta}^{KS}$.

3 INDIRECT INFERENCE APPROACH

We attempt to correct the estimation bias by the indirect inference method, which is introduced by Gouriéroux, Monfort, and Renault (1993) and Smith (1993). The major advantage of the procedure is that a given explicit bias function or its expansion is not required but is calibrated via simulation. Under the framework of the predictive regression considered, this feature is profoundly desirable as not much about the least-squares bias function have been known.

To illustrate the basic idea of indirect inference, we shall define the so-called binding function that mapping Θ into $b_T(\Theta)$ given the sample-size T by: [2]

$$b_T(\theta) = \mathrm{E}\left[\hat{\beta}^{LS}(\theta)\right], \tag{6}$$

where $\theta = (\beta, \sigma_{uv}, \sigma_v^2, \rho_1, \cdots, \rho_p) \in \Theta \subset \mathbb{R}^{3+p}$. Under the assumption that the binding function $b_T(\theta)$ is uniformly continuous and one-to-one, the indirect inference estimator can be obtained by:

$$\hat{\theta}^{II} = b_T^{-1}(\hat{\beta}^{LS}).$$

Obviously, the one-to-one assumption is obviously violated here, and thus the binding function b_T is not invertible. In fact, we need $p+2$ more conditions that help define the relationship

[2] By the very definition of (6), the binding function is exactly the mean-function of $\hat{\beta}^{LS}$.

between all the elements of θ.

To get over, we propose a two-stage estimation procedure which is inspired by the fact that the dependent variable $\{y_t\}_{t=0}^T$ never presents in the process of x_t. This infers the estimation of the autoregressive coefficients can be isolated from that of θ, and could be done as the first stage. Then we come back to the estimation of the predictive regression conditional on the estimates from the first stage. The full estimation procedure is elaborated as followed.

1. **Indirect Inference on Autoregression**

   Let $\Gamma^0 = (\rho_1^0, \cdots, \rho_p^0)$ denote the true value of the autoregressive coefficient vector $\Gamma = (\rho_1, \cdots, \rho_p)$ in (2). Given any of $\Gamma \in \mathbb{S}$ with $\mathbb{S}$ being the subset of $\mathbb{R}^p$ that satisfies the stationarity conditions, we can draw the simulated paths according to (2).[3] This is achieved by drawing independent simulated innovation paths $\{v_t^h\}_{t=0}^T$, $h = 1, \cdots, H$, from $N(0, \sigma_v^2)$, and computing the desired simulated paths $\{x_t^h\}_{t=0}^T$. [4] Let $\hat{\Gamma}^{LS}$ and $\hat{\Gamma}^{LS,h}$ denote the least-squares estimates of Γ with the true data $\{x_t\}_{t=0}^T$ and with the h^{th} simulated path $\{x_t^h\}_{t=0}^T$, respectively. Then the indirect inference estimator of Γ is defined by:

   $$\hat{\Gamma}^{II} = \underset{\Gamma \in \mathbb{S}}{\text{argmin}} \left\| \hat{\Gamma}^{LS} - \frac{1}{H} \sum_{h=1}^{H} \hat{\Gamma}^{LS,h}(\Gamma) \right\|, \tag{7}$$

   where $\|\cdot\|$ is some finite-dimensional distance metric. One might note that when the number of the simulated paths tends to infinity (i.e. $H \to \infty$),

   $$\hat{\Gamma}^{II} = \underset{\Gamma \in \mathbb{S}}{\text{argmin}} \left\| \hat{\Gamma}^{LS} - \mathrm{E}\left[\hat{\Gamma}^{LS,h}\left(\Gamma = \hat{\Gamma}^{II}\right)\right] \right\|,$$

   In other words, if we define a vector binding function as $\underline{b}_T(\Gamma) = \mathrm{E}\left[\hat{\Gamma}^{LS}(\Gamma)\right]$,

   $$\hat{\Gamma}^{II} = \underline{b}_T^{-1}(\hat{\Gamma}^{LS}).$$

   Given $\hat{\Gamma}^{II}$, we can further obtain the estimates of ρ_0 and σ_v by

   $$\hat{\rho}_0^{II} = \frac{\sum_{t=p}^T \left(x_t - \sum_{i=1}^p \hat{\rho}_i^{II} x_{t-i}\right)}{T - p + 1},$$

   $$\hat{\sigma}_v^{II2} = \frac{\sum_{t=p}^T \hat{v}_t^{II2}}{T - 2p},$$

[3] As shown by Shaman and Stine (1988), the bias function of the least-squares estimator of Γ does not depend on either the variance of the residual (σ_v^2) or the drift coefficient ρ_0. Thus, the two parameters can be set arbitrarily when drawing the simulated paths.

[4] In general, initial values of $(x_0^h, \cdots, x_{p-1}^h)$ are required when drawing the simulated paths. In practice, we could generate a much longer simulated paths, say $\{x_t^h\}_{t=-m}^T$, with initial condition $(x_{-m+p-1}^h, \cdots, x_{-m}^h) = (x_0, \cdots, x_{p-1})$, and then drop the first m observations. With the stationarity property of the process, the effect of the initial condition diminishes as $m \to \infty$.

where $\hat{v}_t^{II} = x_t - \hat{\rho}_0^{II} - \sum_{i=1}^{p} \hat{\rho}_i^{II} x_{t-i}$.

2. **Indirect Inference on Predictive Regression**

The estimation in this stage is conditional on $\{\hat{\sigma}_v^{II^2}, \hat{\Gamma}^{II}\}$. More specifically, we treat $\{\hat{\sigma}_v^{II^2}, \hat{\Gamma}^{II}\}$ to be the true values of $\{\sigma_v^2, \Gamma\}$, and then the biding function (6) degenerates into

$$\begin{aligned} b_T\left(\theta | \sigma_v^2 = \hat{\sigma}_v^{II^2}, \Gamma = \hat{\Gamma}^{II}\right) =& b_T\left(\beta, \sigma_{uv}; \hat{\sigma}_v^{II^2}, \hat{\Gamma}^{II}\right) \\ =& \mathrm{E}\left[\hat{\beta}^{LS}\left(\beta, \sigma_{uv}; \hat{\sigma}_v^{II^2}, \hat{\Gamma}^{II}\right)\right]. \end{aligned}$$

However, it remains non one-to-one. To make the indirect inference procedure applicable, we allow the covariance coefficient σ_{uv} to be data-based when drawing the simulated paths. That is, given any value of β, we let $\sigma_{uv} = \tilde{\sigma}_{uv}(\beta)$ which is computed as

$$\tilde{\sigma}_{uv}(\beta) = \frac{\sum_{t=p}^{T} \tilde{u}_t \hat{v}_t^{II}}{\sqrt{(T-p-1)(T-2p)}},$$

where $\tilde{u}_t = y_t - \tilde{\alpha} - \beta x_{t-1}$, and $\tilde{\alpha} = \sum_{t=1}^{T} (y_t - \beta x_{t-1})/T$. The binding function further becomes

$$b_T\left(\beta, \tilde{\sigma}_{uv}(\beta); \hat{\sigma}_v^{II^2}, \hat{\Gamma}^{II}\right) = \mathrm{E}\left[\hat{\beta}^{LS}\left(\beta, \tilde{\sigma}_{uv}(\beta); \hat{\sigma}_v^{II^2}, \hat{\Gamma}^{II}\right)\right]. \tag{8}$$

and in which β is the only parameter left to be estimated. Now given some value of β, we are able to draw the simulated paths $\left\{y_t^h, x_t^h \middle| \beta, \tilde{\sigma}_{uv}(\beta), \hat{\sigma}_v^{II}, \hat{\Gamma}^{II}\right\}_{t=0}^{T}$ according to (1) and (2).[5] Letting $\hat{\beta}^{LS}$ and $\hat{\beta}^{LS,h}$ respectively denote the least-squares estimates of β with the true data $\{y_t, x_t\}_{t=0}^{T}$ and with the h^{th} simulated path $\left\{y_t^h, x_t^h \middle| \beta, \tilde{\sigma}_{uv}(\beta), \hat{\sigma}_v^{II}, \hat{\Gamma}^{II}\right\}_{t=0}^{T}$, the "*two-stage indirect inference estimator*" of β is defined as

$$\hat{\beta}^{II} = \underset{\beta \in \mathbb{R}}{\operatorname{argmin}} \left\| \hat{\beta}^{LS} - \frac{1}{H} \sum_{h=1}^{H} \hat{\beta}^{LS,h}\left(\beta, \tilde{\sigma}_{uv}(\beta); \hat{\sigma}_v^{II}, \hat{\Gamma}^{II}\right) \right\|, \tag{9}$$

Again, when $H = \infty$,

$$\hat{\beta}^{II} = \underset{\beta \in \mathbb{R}}{\operatorname{argmin}} \left\| \hat{\beta}^{LS} - \mathrm{E}\left[\hat{\beta}^{LS,h}\left(\beta, \tilde{\sigma}_{uv}(\beta); \hat{\sigma}_v^{II}, \hat{\Gamma}^{II}\right)\right] \right\|.$$

And given (8), we have

$$\hat{\beta}^{II} = b_T^{-1}\left(\hat{\beta}^{LS}, \tilde{\sigma}_{uv}(\hat{\beta}^{II}); \hat{\sigma}_v^{II^2}, \hat{\Gamma}^{II}\right). \tag{10}$$

[5]When drawing simulated paths, we set $\alpha = \tilde{\alpha}$, $\sigma_u^2 = \tilde{\sigma}_u^2 = \sum_{t=1}^{T} \tilde{u}_t^2/(T-2)$, and $\rho_0 = \hat{\rho}_0^{II}$, even though they do not exist in the biding function and thus almostly never affect the indirect inference estimation when H is large enough.

The asymptotic behavior of $\hat{\beta}^{II}$ can be summarized by the following theorem.

Theorem 2. *Under the stationarity assumption of the predictor x_{t-1} and when the number of the simulated paths $H = \infty$, we have*

$$\sqrt{T}\left(\hat{\beta}^{II} - \beta\right) \Rightarrow N\left(0, \sigma_u^2 Q^{-1}\right),$$

where $Q = var(x_t)$ which is a function of the autoregression coefficients. This means the indirect inference estimator is consistent and has the asymptotic distribution identical to that of the least-squares estimator.

4 SIMULATION STUDY

In this section, we investigate the relative finite-sample performance of the indirect inference estimator against the least-squares estimator by means of simulations. For each simulation, the numbers of replications and simulated paths (H) are set to be $5,000$ and $10,000$, respectively.

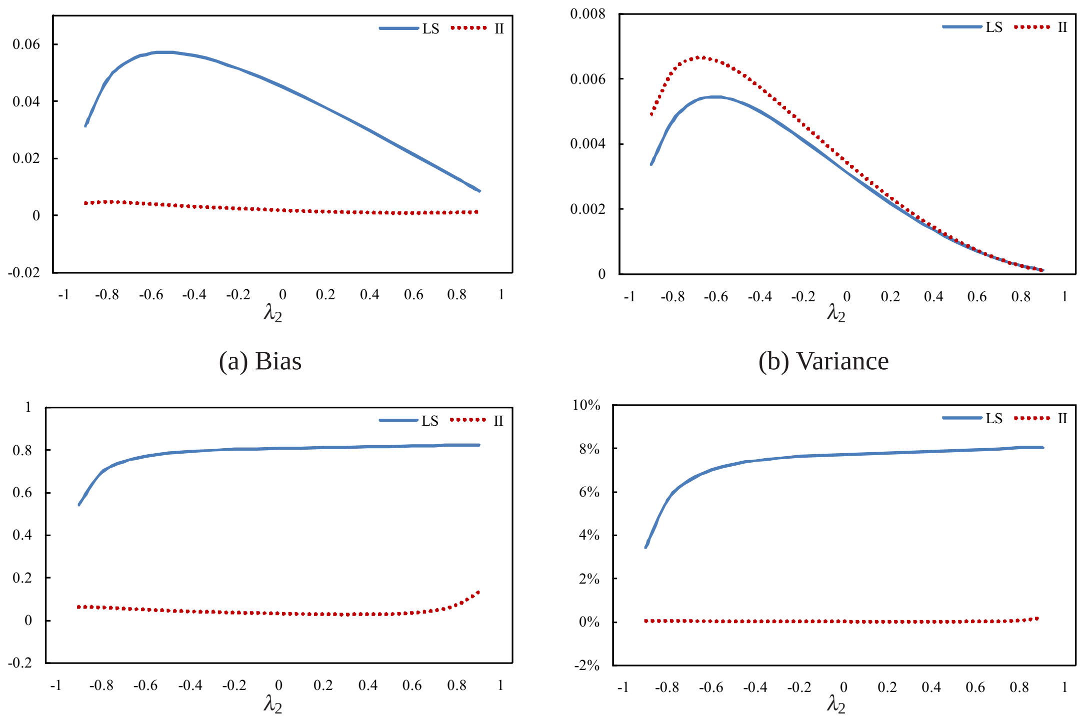
We begin with simulations with arbitrary parameter settings to examine how the estimators perform when the values of some important parameters that determine least-squares bias vary. According to Proposition 1, these important parameters should include the autoregressive coefficients that characterized the predictor variable, the covariance of u_t and v_t, and the variance of v_t relative to that of u_t.

The first simulation design is to illustrate the autoregressive effect of the predictor variable on the least-squares bias. The effect is complicated to analyze when AR order is high because it depends on all of the autoregressive coefficients. To simplify the analysis, the predictor is assumed to follow an AR(2) model with a fixed largest characteristic root equal to 0.95 on purpose to feature its high-persistence usually seen in financial forecasting literature. The full data generating process (DGP) is constituted by (1), (2) and (3) with $T = 80$, $\alpha = \beta = \rho_0 = \sigma_u = \sigma_v = 1$, $\sigma_{uv} = -0.8$, $\rho_1 = \lambda_1 + \lambda_2$, $\rho_2 = -\lambda_1\lambda_2$, $\lambda_1 = 0.95$ and $\lambda_2 = (-0.9, -0.8, \cdots, 0.9)$.

Figure 1 depicts the simulation results by graphing some statistics as functions of λ_2. The least-squares estimator appears to be biased within the whole parameter space considered, and the bias function is nonlinear in the small root, λ_2. It is expected that the bias function would be more complicated as the AR order increases. On the contrary, the indirect inference estimator tends to have very little bias. It should again be emphasized that the indirect inference estimation does not require an explicit form of the bias function of the base estimator (LS) and would remain applicable even when the process of the predictor is involved with a much higher lag-order. As shown by Figure 1(b), the variance of the indirect inference estimator is larger than that of the least-squares although the difference could be very minor when $\lambda_2 > 0$. This is generally true and can be explained by that the indirect inference requires additional estimation of the predictor process while the least-squares does not.

When conducting a conventional t-test for the predictive coefficient in finite-sample, researchers might encounter the problem of size-distortion that could be attributed to two reasons. The first is the estimation bias that leads to a horizontal shift of the null distribution, and the other is the tendency of the explanatory variable(s) to depart from stationarity which results in a change in the shape of the distribution. In general, the only we could take care of in the estimation phase is the bias. Thus, we focus on the size-distortion induced from the bias presuming the shape of the null distribution is not altered. To this end, it would be more meaningful to investigate the standardized bias (SB) calculated as the ratio of the bias to the standard deviation of each estimator, rather than the raw value of the bias which seems to be very small relative to the true $\beta(=1)$. This is because the standardized bias has the sense of the horizontal distance that the null distribution of the t-test statistic is shifted. As shown by Figure 1(c), the standardized bias of the least-squares is much more critical than the counterpart of the indirect inference estimator. Figure 1(d) further illustrates the size distortion of a two-tailed t-test with a nominal size of 5% by assuming that the actual null distribution is $N(0,1)+\text{SB}$. It states that test based on the indirect inference estimator would be preferred because it is involved with very little size distortion induced by the estimation bias.

The next simulation is to investigate how the estimators perform when the covariance of u_t and v_t varies. The parameter settings of the DGP are $T = 80$, $\alpha = \beta = \rho_0 = \sigma_u = \sigma_v = 1$, $\lambda_1 = 0.95$, $\lambda_2 = 0$, and $\sigma_{uv} = (-1, -0.9, \cdots, 1)$. Although $\lambda_2 = 0$ implies that the predictor is AR(1), we still fit an AR(2) model in the first-stage estimation of the indirect inference even it would be less efficient. Simulation results are displayed by Figure 2. Consistent with Proposition 1, the bias of the least-squares is linear in the covariance parameter σ_{uv} and reaches the minimum 0 when $\sigma_{uv} = 0$. For the indirect inference estimator, it still appears to have virtually no bias irrespective of the value of σ_{uv}. One thing interesting is that the change of σ_{uv} brings very little effect to the standard deviation of the least-square and the consequent standardized bias



(a) Bias

(b) Variance

(c) Standardized bias

(d) Size distortion induced by bias

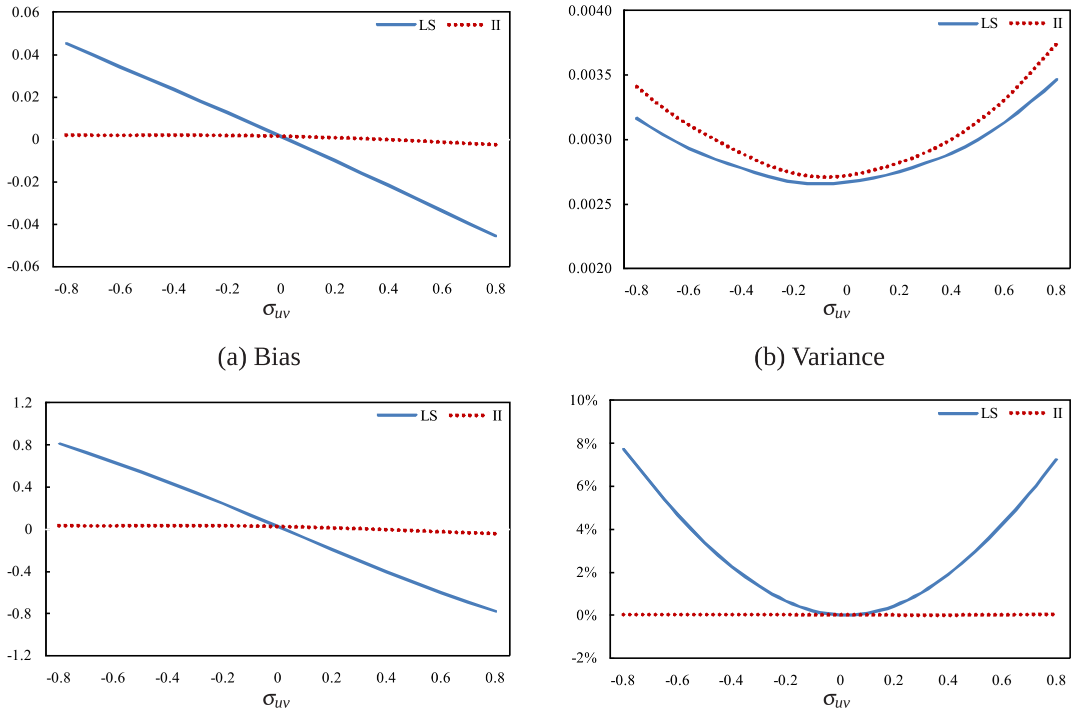
Figure 1: Finite-sample properties of $\hat{\beta}^{LS}$ and $\hat{\beta}^{II}$ as λ_2 varies

seems to also be linear but actually not. Again, we compute the size distortion of the t-test under the assumption that the null distribution is shifted horizontally for the distance equal to standardized bias of each estimator, but not alters in its shape. For the least-squares, the test appears to be well sized when $\sigma_{uv} = 0$, but the extent of distortion is raised as the absolute value of σ_{uv} increases. Benefiting from the bias-reduction, the test related to the indirect inference estimator is involved with little size distortion for all values of σ_{uv}.

Figure 3 shows the simulation results with various σ_u/σ_v ratios holding other parameters unchanged. The parameter settings are $T = 80$, $\alpha = \beta = \rho_0 = 1$, $\lambda_1 = 0.95$, $\lambda_2 = 0$, $\sigma_v = 1$, $\sigma_u = (1, 2, \cdots, 10)$, and $\sigma_{uv} = -0.8\sigma_u$. The setting of σ_{uv} implies a fixed correlation coefficient between u_t and v_t that is equal to -0.8. As shown by Figure 3(a), the absolute value of the least-squares bias is increasing with the σ_u/σ_v ratio abiding by a linear relationship. This can be explained by re-writing (4) in Proposition 1 as

$$\mathrm{E}\left(\hat{\beta}^{LS} - \beta\right) = \frac{\sigma_{uv}}{\sigma_u \sigma_v} \cdot \frac{\sigma_u}{\sigma_v} \mathrm{E}\left[\frac{\sum_{t=1}^{T}(x_{t-1} - \bar{x})v_t}{\sum_{t=1}^{T}(x_{t-1} - \bar{x})^2}\right] = -0.8 \cdot \frac{\sigma_u}{\sigma_v} \mathrm{E}\left[\frac{\sum_{t=1}^{T}(x_{t-1} - \bar{x})v_t}{\sum_{t=1}^{T}(x_{t-1} - \bar{x})^2}\right],$$

and observing that the expectation term is fixed. Besides, the variance of the least-squares is also increasing with the σ_u/σ_v ratio. This is because the unexplained signal from the residual



(a) Bias (b) Variance

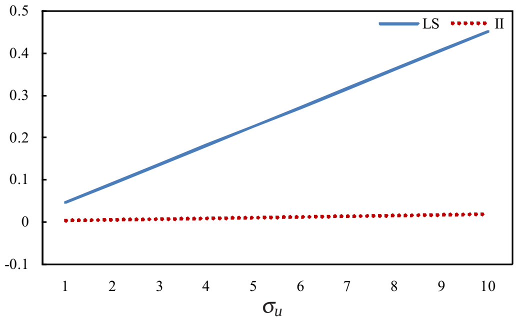
(c) Standardized bias (d) Size distortion induced by bias

Figure 2: Finite-sample properties of $\hat{\beta}^{LS}$ and $\hat{\beta}^{II}$ as σ_{uv} varies

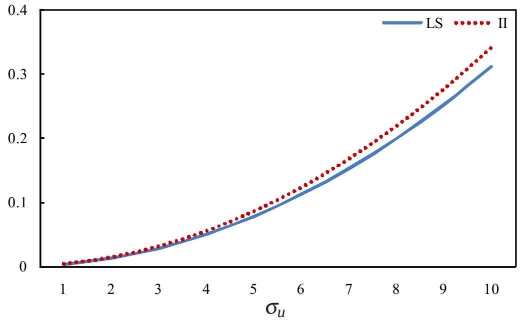
u_t is strengthen as σ_u/σ_v becomes larger, while the signal from the explanatory variable (the predictor) remains unchanged. According to Theorem 2, the standard deviation of the least-squares have a linear relationship with σ_u asymptotically, and it seems to be also true in our finite-sample simulations. As a result, the standardized bias of the least-squares appears to be constant as the σ_u/σ_v ratio varies, and so does the size distortion of the t-test that due to the shift of the null distribution. For the indirect inference estimator, it remains to perform well in both bias, standardized bias, and the size of the t-test, although a little increase of variance is present to be the cost.

We now turn to the simulations based on DGP estimated from real data which might tell us more about the estimators under the situations we actually encounter. There are fourteen designs of simulation. For each simulation, the pseudo data is generated from (1), (2) and (3) with parameter values obtained from the empirical study in the next section. Specifically, we estimate 14 models in which the annual equity premium on S&P500 index is predicted by a different lagged variable, and then treat the indirect inference estimates of the parameters as the true under simulations. The estimated parameter values can be found in Table 5 - Table 7.

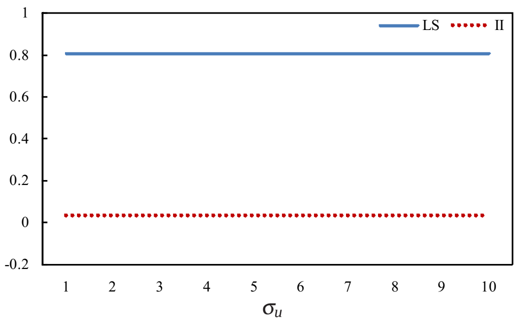
As mentioned earlier, indirect inference on predictive regression model involves a two-stage



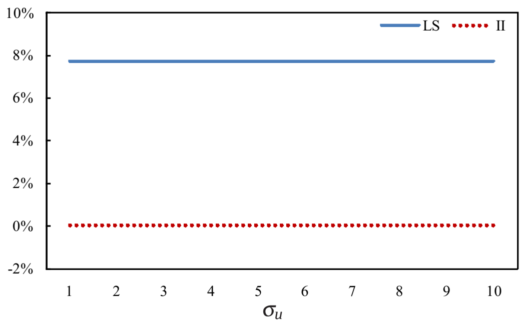
(a) Bias



(b) Variance



(c) Standardized bias



(d) Size distortion induced by bias

Figure 3: Finite-sample properties of $\hat{\beta}^{LS}$ and $\hat{\beta}^{II}$ as σ_u varies

estimation procedure, and the first-stage estimation of the autoregressive model would have a great influence on the second-stage estimation of the desired predictive coefficient. Thus, we would like to see at first how well the widely-known estimation bias problem of the autoregression can be addressed by indirect inference. Table 2 reports the simulation results of the first-stage estimation; meanwhile, the results of the least-squares estimator are also presented for comparison purpose. Overall speaking, the least-squares bias for the autoregressive models is economically significant. On the contrary, the indirect inference estimator has virtually no bias even in the cases where the series of the predictors are featured by a large root and/or a high lag-order. Besides, the indirect inference estimator appears to have smaller root mean square errors (RMSE) than the least-squares when the predictor variable is AR(1). The same seems not to be true when the series is characterized by a higher-order model because the RMSE is almost determined by the variance on which measure the indirect inference estimator tends to be slightly inferior.

We proceed to the simulation results for the predictive regression given in Table 3. It is apparent that the indirect inference estimator has very little bias (in absolute value) compared with that of the least-squares for all cases. Besides, the indirect inference estimator appears to have a slightly larger variance than that of the least-squares, and this could be viewed as the cost of bias reduction. On the measure of RMSE, no estimator could dominate the other in all cases. This is because the two estimators are respectively predominant for the bias and the variance, that together make up the RMSE.

As before, we compute the standardized bias as the ratio of the bias to the standard deviation of each estimator, which would has more sense about the magnitude of the bias. The standardized bias of the indirect inference estimator is far smaller than the counterpart of the least-squares in all cases. According to previous simulation results, it is expected that testing based on the indirect inference estimator could well immune from the bias-induced size distortion that usually accompanies with the least-squares-based test.

Unlike the simulations with arbitrary settings, we actually construct the t-statistic based on each estimators and simulate it in order to account for the "complete" size distortion arising from both the horizontal shift and the change of the shape of the null distribution caused by the autoregressive properties of the predictor. According to the asymptotic results in Theorem 2, the t-statistics are constructed as

$$t^{LS} = \frac{\hat{\beta}^{LS} - \beta}{\left[\sum_{t=1}^{T}(x_{t-1} - \bar{x})^2\right]^{-1} \hat{\sigma}_u^{2LS}} \text{ and } t^{II} = \frac{\hat{\beta}^{II} - \beta}{\left[\sum_{t=1}^{T}(x_{t-1} - \bar{x})^2\right]^{-1} \hat{\sigma}_u^{2II}}, \tag{11}$$

respectively. The last four columns in Table 3 report the realized size when the 5% and 10% nominal size are used for both left-tailed and right-tailed t-tests. Roughly speaking, the degree of size distortion of each test is consistent to the value of the standardized bias of the estimator that the test is based. For cases such as the regressions where the excess stock return is predicted by the book-to-market ratio (**b/m**), default yield spread (**dfy**), or dividend-price ratio (**d/p**), the least-squares estimator tends to be involved with extremely large standardized bias and the t^{LS}-test encounters critical size distortion problem. On the contrary, the t^{II}-test appears to be well sized in all cases but with exceptions for the left-tailed tests under the regression models where either **b/m** or **d/p** serves as the predictor of the excess stock return. Since the indirect inference estimator has very little standardized bias in these exception cases, we can attribute the size distortion to the change in shape of the null distribution induced by the high-persistency of the predictor. One thing interesting is that the left-tailed tests with worse realized size is usually not employed in practice because a positive coefficient is expected when the predictor is **b/m** or **d/p**. So far, we might have the conclusion that testing based on the indirect inference estimator would be more reliable than that based on the least-squares.

5 EMPIRICAL ILLUSTRATION

In this section, we illustrate the indirect inference method using the annual equity premium prediction models considered by Goyal and Welch (2008).[6] In each model, the annual S&P500 equity premium is predicted by a single lagged variable listed in Table 4. The data set being used had been updated to 2008 by Amit Goyal and is available on his website.

Assuming the observations of the equity premium and the predictor variable actually come from the model system, (1), (2) and (3), the indirect inference estimate of the predictive coefficient can be obtained following the estimation procedure described in section 2. However, two problems may be encountered when applying the procedure. First, the maximum lag-order (p) of the predictor's process is usually unknown in practice. To get over, we suggest a sequential method to determine it. Specifically, we start by fitting the predictor variable by an "AR(0)" model with a drift, and test the least-squares residuals by the Breusch-Godfrey LM tests for no first- to fourth-order serial correlation. If all of the 4 tests can not be rejected at the 10% level of significance, the lag-order of the autoregressive model is determined to be 0. Else, we increase the lag-order by 1 and repeat the tests until the least-squares residual does not exhibit significant serial-correlation.

Secondly, it is usually impossible to solve the minimization problems, such as (7) and (9), by the widely-used grid search method. For (9), the predictive coefficient β is known to be in the set of real numbers which is unbounded if no other economic constraint is imposed, and the grid search method is not applicable in consequence. For (7) with a lag-order higher than 1, even the value of the autoregressive coefficient vector is bounded by the stationarity conditions, it is necessary but knotty to transform the conditions expressed by the characteristic roots into the one expressed by the autoregressive coefficients. Moreover, it is time-consuming to do multi-parameter grid search, and in this case, the computation time tends to grow geometrically as the lag-order increases.

To overcome such difficulties, we adapt the algorithm that is suggested by Gouriéroux, Renault, and Touzi (2000) for the indirect inference estimation of a AR(2) model. We use the same notation as in section 2. For the general case of the AR(p) model such as (2), suppose

[6] We do not consider the monthly predictive regression models as in Goyal and Welch (2008) because extremely high autoregressive lag-orders are selected for the predictors based on either the criterion we used in this paper, AIC, or SBC. Besides, models with the predictors such as default return spread (**dfr**) and long term return (**ltr**), are also excluded because zero autoregressive lag-order is selected with our criterion.

the real data are generated with the unknown true value of the autoregressive parameter vector $\Gamma = \Gamma_0$. We define a function as:

$$g_\lambda(\Gamma) = \Gamma + \lambda \left[\hat{\Gamma}^{LS} - \frac{1}{H} \sum_{h=1}^{H} \hat{\Gamma}^{LS,h}(\Gamma) \right], \tag{12}$$

where $\hat{\Gamma}^{LS}$ is the least-squares estimate of Γ with the true data, and $\hat{\Gamma}^{LS,h}(\Gamma)$ denotes the least-squares estimate of Γ with the h^{th} simulated path, $\{x_t^h\}_{t=0}^T$, drawn from the AR(p) model with some value of Γ. When $H = \infty$, we hope (12) is a strong contraction with some λ, and $\hat{\Gamma}^{II}$ is the unique fixed point. Thus, for a given $\hat{\Gamma}^{LS}$, we can construct the sequence $\{\hat{\Gamma}^{(n)}\}_{n\geq 0}$:

$$\hat{\Gamma}^{(0)} = \hat{\Gamma}^{LS} \text{ and } \hat{\Gamma}^{(n+1)} = g_\lambda(\hat{\Gamma}^{(n)}),$$

and expect the convergence of the sequence to $\hat{\Gamma}^{II}$. The same algorithm is also used by Tanizaki, Hamori, and Matsubayashi (2005), but with a slightly modification that the contraction parameter λ is allowed to depend on the iteration number, n.[7] In the applications of this paper, the contraction parameter is set as $\lambda = 0.2$ following Gouriéroux, Renault, and Touzi (2000). We obtain convergence in all cases considered with the terminal condition $\sum_{i=1}^{p} |\hat{\rho}_i^{(n+1)} - \hat{\rho}_i^{(n)}| < 10^{-4}$. The same algorithm can also be applied to solve the minimization problem (9), because the predictive coefficient β serves as the only argument.

Now we turn to the empirical results. Table 4 reports the autoregressive order of each predictor process selected with the sequential LM test. Half of the 14 predictor processes appear to have a lag-order higher than 1, which sheds some light to our concern that the bias-correction of the least-squares estimator based on the formula derived under the AR(1)-predictor assumption could usually be inappropriate. Table 5 shows the estimation results of the predictor processes. As shown by previous simulations, the least-squares estimator of the autoregressive coefficients could be much biased. Thus, it is not surprising to see big differences between the the least-squares and the indirect inference estimates, as the latter is able to well take care of the possible bias. The last column reports the largest characteristic root implied by the estimates of the autoregressive coefficients. All of the implied roots are smaller than unity irrespective of the estimation methods. Besides, in 13 of the 14 cases, the largest root implied by the indirect inference estimates is larger than that implied by the least-squares estimates. The only exception is the process of the term spread (**tms**), which is expected to have a much smaller largest-root

[7] In Tanizaki, Hamori, and Matsubayashi (2005), the contraction parameter is set as $\lambda^{(n)} = c^{n-1}$ with $c = 0.9$, and the algorithm is interpreted by the Newton-Raphson method.

compared with that of other processes. It seems to say but requires future verification that the least-squares is inclined to underestimate a large root.

Table 6 reports the estimation results for the predictive regressions. We find a upward bias-correction by comparing the estimates of the least-squares and the indirect inference in four cases in which either the dividend yield (**d/y**), investment capital ratio (**i/k**), T-Bill rate (**tbl**), or net equity expansion (**ntis**), serves as the predictor, and find a downward bias-correction in other cases. To go a step further, the estimates of the covariance between the innovations in the predictive regression and the autoregression are reported in Table 7. According to Proposition 1 and the signs of the ratio $\hat{\sigma}_{uv}/\hat{\sigma}_v^2$, we can expect a negative estimate of the expectation term in the proposition for all cases. As a result, the direction of the bias-correction implied by the indirect inference estimate is always the same as the sign of $\hat{\sigma}_{uv}/\hat{\sigma}_v^2$.

Table 6 also reports the t-statistics based on both estimators computed using formula (11). The values of the two t-statistics could differ a lot, and the indirect inference based t-test which is bias-corrected does not always lower or raise the significance when compared with the least-squares based. A the 10% level, the two tests lead to the same inference about the predictive coefficient in all cases, with the only exception where the equity premium is predicted by dividend-price ratio (**d/p**), and significance is not observed for the indirect inference based test. With the simulation results in the previous section, one might prefer to believe the testing results based on indirect inference estimator. However, as what was shown in the last four columns, the residual in the predictive regression appears to be serially correlated in six of the cases. For these cases, both the two tests might not be reliable even we reconstruct each test with a consistent estimation of the asymptotic covariance matrix. For example, consider the simple model:

$$y_t = \alpha + \beta x_{t-1} + u_t,$$
$$x_t = \theta + \rho x_{t-1} + v_t = \theta/(1-\rho) + v_t + \rho v_{t-1} + \rho^2 v_{t-2} + \cdots,$$
$$u_t = \gamma u_{t-1} + w_t = w_t + \gamma w_{t-1} + \gamma^2 w_{t-2} + \cdots,$$

where $|\rho| < 1$, $|\gamma| < 1$, and the innovation vector $(w_t, v_t)'$ is serially independent and normally distributed with mean zero and covariance matrix $[\sigma_w^2 \ \ \sigma_{wv}; \ \sigma_{wv} \ \ \sigma_v^2]$. Observing that $\mathrm{E}(w_t, v_{t+j}) = 0 \ \forall j \neq 0$, we have

$$\mathrm{E}(x_{t-1}u_t) = \frac{\gamma\sigma_{wv}}{1-\rho\gamma},$$

which is generally not zero regardless of the sample size. As a result, the least-squares estimator of β is expected to be inconsistent, and the conventional t-test would be accompanied with a asymptotic size distortion. Certainly, the indirect inference approach would suffer from the same problem, as it is not designed to take account of the serial correlation. Although there exists a possibility to adapt our approach to this kind of model, it is left to be an issue for future research.

6 CONCLUDING REMARKS

It is well known that the least-squares estimator is biased when the predictive regression contains an autoregressive predictor which has innovations contemporarily correlated with the dependent variable. In consequence, the conventional asymptotic tests, such as t-test, suffer from size-distortion and lead to invalid inference. For the purpose of bias reduction, we have proposed the indirect inference method which is simulation-based and applicable without any explicit form of the bias function or its expansion. Simulation studies show that the bias can be effectively reduced even when the autoregressive order of the predictor is high. Thanks to the little bias, the problem of size distortion is much alleviated when the t-test is built on the indirect inference estimator. The proposed method is also applied to investigate the predictability of stock returns, but less evidence has been found. Although inference based on our approach is supposed to be more reliable, some evidence of serial correlation emerges to challenge the white noise assumption in our model. If this is the case, either the least-squares or the indirect inference could produce misleading inference. Fortunately, with the spirit of simulation, the indirect inference method is expected to deal with serial correlation by some modifications. One drawback of the simulation-based estimation method should be its compute-intensive nature. However, with the rapid advancement of computer technology and the development of numerical methods, the computation cost is becoming a minor problem in practice and we could reasonably expect the widespread use of indirect inference in the near future.

APPENDIX: PROOFS

A Proof of Proposition 1

Following Stambaugh (1999), $u = (u_0, \cdots, u_T)'$ can be decomposed as

$$u = \frac{\sigma_{uv}}{\sigma_v^2} v + \varepsilon,$$

where $v = (v_0, \cdots, v_T)'$. With the i.i.d. normality assumption, we have

$$\mathrm{E}(\varepsilon|X) = \mathrm{E}(\varepsilon|x_{-p}, x_{-p+1}, \cdots, x_{-1}, v_0, \cdots, v_T) = 0.$$

Thus,

$$\mathrm{E}\begin{bmatrix} \hat{\alpha}^{LS} - \alpha \\ \hat{\beta}^{LS} - \beta \end{bmatrix} = \mathrm{E}[(X'X)^{-1}X'u] = \frac{\sigma_{uv}}{\sigma_v^2}\mathrm{E}[(X'X)^{-1}X'v].$$

where the second element can be expressed explicitly as

$$\mathrm{E}\left(\hat{\beta}^{LS} - \beta\right) = \frac{\sigma_{uv}}{\sigma_v^2}\mathrm{E}\left[\frac{\sum_{t=1}^{T}(x_{t-1} - \bar{x})v_t}{\sum_{t=1}^{T}(x_{t-1} - \bar{x})^2}\right].$$

The expectation term in the equation will generally be $O(T^{-1})$, because $\hat{\beta}^{LS}$ is $T^{1/2}$ consistent as shown by Theorem 2 (cf. MacKinnon and Smith, 1998). Besides, the expectation term does not depend on σ_v^2, as the result that σ_v serves as the role of scale parameter for both x_t and v_t for all t.

B Proof of Theorem 2

Utilizing (4) and (10), we have

$$\hat{\beta}^{II} = \hat{\beta}^{LS} - \frac{\hat{\sigma}_{uv}^{II}}{\hat{\sigma}_v^{II}}\mathrm{E}\left[\left.\frac{\sum_{t=1}^{T}(x_{t-1}^h - \bar{x}^h)v_t^h}{\sum_{t=1}^{T}(x_{t-1}^h - \bar{x}^h)^2}\right|_{\Gamma=\hat{\Gamma}^{II}}\right] = \hat{\beta}^{LS} - \frac{\hat{\sigma}_{uv}^{II}}{\hat{\sigma}_v^{II}} f\left(\hat{\Gamma}^{II}, T\right), \tag{13}$$

where $\mathrm{E}[\cdot] = f\left(\hat{\Gamma}^{II}, T\right)$, $\{x_{t-1}^h, v_t^h\}_{t=1}^T$ denotes the h^{th} simulated path based on the autoregressive coefficient vector $\Gamma = \hat{\Gamma}^{II}$, and $\bar{x}^h = \sum_{t=1}^{T} x_{t-1}^h / T$. Given that $\hat{\Gamma}^{II}$ is a consistent estimator of the true Γ and x_t is covariance-stationary, x_t^h will meet the stationary condition as well when

$T \to \infty$.[8] By the law of large number (LLN),

$$T^{-1}\sum_{t=1}^{T}\left(x_{t-1}^{h}-\bar{x}^{h}\right)^{2} \xrightarrow{p} Q^{h}, \tag{14}$$

where $Q^h = var(x_t^h)$. Moreover, since $\left(x_{t-1}^{h}-\bar{x}^{h}\right) v_t^h$ is a martingale difference sequence, the central limit theory (CLT; Corollary 5.25 of White, 1984) leads to

$$T^{-1/2}\sum_{t=1}^{T}\left(x_{t-1}^{h}-\bar{x}^{h}\right) v_t^h \Rightarrow N\left(0, \hat{\sigma}_v^{II^2} Q^h\right). \tag{15}$$

Combining (14) and (15) gives that $\sqrt{T} f\left(\hat{\Gamma}^{II}, T\right) \to 0$ as $T \to \infty$. And together with (13), it is revealed that $\sqrt{T}(\hat{\beta}^{II}-\beta)$ and $\sqrt{T}(\hat{\beta}^{LS}-\beta)$ share the same asymptotics.

Now what left to be proven is the $\sqrt{T}$ asymptotics of $\hat{\beta}^{LS}$ or $\hat{\beta}^{II}$. Observing that

$$\sqrt{T}(\hat{\beta}^{LS}-\beta) = \frac{T^{-1/2}\sum_{t=1}^{T}(x_{t-1}-\bar{x})u_t}{T^{-1}\sum_{t=1}^{T}(x_{t-1}-\bar{x})^2},$$

and making use of the similar arguments of (14) and (15) yields the desired results:

$$T^{-1}\sum_{t=1}^{T}(x_{t-1}-\bar{x})^2 \xrightarrow{p} Q,$$

$$T^{-1/2}\sum_{t=1}^{T}(x_{t-1}-\bar{x})u_t \Rightarrow N\left(0, \sigma_u^2 Q\right),$$

where $Q = var(x_t)$.

REFERENCES

Amihud, Y., and Hurvich, C.M. (2004), "Predictive Regression: A Reduced-Bias Estimation Method," *Journal of Financial and Quantitative Analysis*, 39, 813−841.

Amihud, Y., Hurvich, C.M., and Wang, Y. (2010), "Predictive Regression With Order-p Autoregressive Predictors," *Journal of Empirical Finance*, 17, 513−525.

Campbell, J.Y., and Shiller, R.J. (1988), "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors," *The Review of Financial Studies*, 1, 195−228.

Campbell, J.Y., and Thompson, S.B. (2008), "Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average?" *Review of Financial Studies*, 21, 1509−1531.

[8] As shown by Shaman and Stine (1988), the bias function of the least-squares estimator $\hat{\Gamma}^{LS}$ is an $O(T^{-1})$ function of T and Γ itself. Thus, the indirect inference estimator can be expressed as $\hat{\Gamma}^{II} = \hat{\Gamma}^{LS} + O(T^{-1})$. This directly leads to the consistency of $\hat{\Gamma}^{II}$ as long as the well known result that $\hat{\Gamma}^{LS} \xrightarrow{p} \Gamma$ holds.

Campbell, J. Y., and Yogo, M. (2006), "Efficient Tests of Stock Return Predictability," *Journal of Financial Economics*, 81, 27−60.

Cavanagh, C.L., Elliott, G., Stock, J.H. (1995), "Inference in Models With Nearly Integrated Regressors," *Econometric Theory*, 11, 1131−1147.

Engel, C., and West, K.D. (2005), "Exchange Rate and Fundamentals," *Journal of Political Economy*, 113, 485−517.

Engel, C., Mark, N.C., and West, K.D. (2008), "Exchange Rate Models Are Not as Bad as You Think," NBER Chapters, in *NBER Macroeconomics Annual 2007*, Volume 22, National Bureau of Economic Research, Inc., pp. 381−441.

Fama, E. F., and French, K. R. (1988), "Dividend Yields and Expected Stock Returns," *Journal of Financial Economics*, 22, 3−25.

Gouriéroux, C., Monfort, A., Renault, E. (1993), "Indirect Inference," *Journal of Applied Econometrics*, 8, S85−S118.

Gouriéroux, C., Renault, E., and Touzi, N. (2000), "Calibration by Simulation for Small Sample Bias Correction," in *Simulation-Based Inference in Econometrics: Methods and Applications*, eds. Mariano, R.S., Schuermann, T., and Weeks, M., Cambridge University Press, pp. 328−358.

Gouriéroux, C., Phillips, P. C. B., and Yu, J. (2010), "Indirect Inference for Dynamic Panel Models," *Journal of Econometrics*, 157, 68−77.

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

Groen, J.J.J. (2000), "The Monetary Exchange Rate Model as a Long-Run Phenomenon," *Journal of International Economics*, 52, 299−319.

Groen, J.J.J. (2005), "Exchange Rate Predictability and Monetary Fundamentals in a Small Multi-Country Panel," *Journal of Money, Credit, and Banking*, 37, 495−516.

Kendall, M. G. (1954), "Note on the Bias in the Estimation of Autocorrelation," *Biometrika*, 41, 403−404.

Kothari, S., and Shanken, J. (1997), "Book-to-Market, Dividend Yield, and Expected Market Returns: A Time-Series Analysis," *Journal of Financial Economics*, 44, 169−203.

Hodrick, R. J. (1992), “Dividend Yields and Expected Stock Returns: Alternative Procedures for Inference and Measurement,” *The Review of Financial Studies*, 5, 257−86.

Kilian L. (1999), “Exchange Rates and Monetary Fundamentals: What Do We Learn From Long Horizon Regressions?” *Journal of Applied Econometrics*, 14, 491−510.

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

Lewellen, J. (2004), “Predicting Returns With Financial Ratios,” *Journal of Financial Economics*, 74, 209−35.

MacKinnon, J. G., and Smith Jr., A. A. (1998), “Approximate Bias Correction in Econometrics,” *Journal of Econometrics*, 85, 205−230.

Mankiw, N.G., and Shapiro, M.D. (1986), “Do We Reject Too Often? Small Sample Properties of Tests of Rational Expectations Models,” *Economics Letters*, 20, 139−145.

Mark, N.C. (1995), “Exchange Rates and Fundamentals: Evidence on Long-Horizon Predictability,” *American Economic Review*, 85, 201-218.

Mark, N.C., and Sul, D. (2001), “Nominal Exchange Rates and Monetary Fundamentals: Evidence From a Seventeen Country Panel,” *Journal of International Economics*, 53, 29−52.

Marriott, F. H. C., and Pope, J. A. (1954), “Bias in the Estimation of Autocorrelations,” *Biometrika*, 41, 393−402.

Molodtsova, T., and Papell, D. H. (2009), ”Out-of-Sample Exchange Rate Predictability With Taylor Rule Fundamentals,” *Journal of International Economics*, 77, 167−180.

Shaman, P., and Stine, R. A. (1988), “The Bias of Autoregressive Coefficient Estimators,” *Journal of the American Statistical Association*, 83, 842−848.

Staiger, D., Stock, J.H., and Watson, M.W. (1997), “The NAIRU, Unemployment and Monetary Policy,” *Journal of Economic Perspectives*, 11, 33−49.

Stambaugh, R. (1986), “Bias in Regressions With Lagged Stochastic Regressors,” Unpublished Manuscript, University of Chicago, Chicago, IL.

Stambaugh, R. F. (1999), “Predictive Regressions,” *Journal of Financial Economics*, 54, 375−421.

Stock, J.H., and Watson, M.W. (1999), “Forecasting Inflation,” *Journal of Monetary Economics*, 44, 293−335.

Tanizaki, H., Hamori, S., and Matsubayashi, Y. (2005), "On Least-Squares Bias in the AR(p) Models: Bias Correction Using the Bootstrap Methods," *Statistical Papers*, 47, 109−124.

Torous, W., Valkanov, R. and Yan, S. (2004), "On Predicting Stock Returns with Nearly Integrated Explanatory Variables," *Journal of Business*, 77, 937−66.

White, H. (1984), *Asymptotic Theory for Econometricians*, Academic Press.

Table 1: Bias of AR(1)-Based Estimator When Predictor is AR(2)

λ_2	$bias\left(\hat{\beta}^{LS}\right)$	$bias\left(\hat{\beta}^{KS}\right)$	$\frac{bias\left(\hat{\beta}^{LS}\right)}{bias\left(\hat{\beta}^{KS}\right)}$
−0.8	0.0219	0.0147	0.671
−0.7	0.0322	0.0194	0.602
−0.6	0.0361	0.0183	0.506
−0.5	0.0396	0.0172	0.434
−0.4	0.0398	0.0134	0.338
−0.3	0.0399	0.0103	0.258
−0.2	0.0400	0.0078	0.196
−0.1	0.0384	0.0042	0.111
0.0	0.0363	0.0008	0.022
0.1	0.0350	-0.0011	0.032
0.2	0.0332	-0.0028	0.085
0.3	0.0303	-0.0051	0.168
0.4	0.0275	-0.0065	0.236
0.5	0.0242	-0.0079	0.326
0.6	0.0213	-0.0082	0.386
0.7	0.0182	-0.0079	0.436
0.8	0.0151	-0.0071	0.469

1 The DGP is (1), (2), and (3), with $T = 80$, $\alpha = \rho_0 = \sigma_u = \sigma_v = 1$, $\beta = 0$, $\sigma_{uv} = -0.8$, $\rho_1 = \lambda_1 + \lambda_2$, and $\rho_2 = -\lambda_1\lambda_2$. The two characteristic roots are set as $\lambda_1 = 0.9$, and $\lambda_2 = (-0.8, -0.7, \cdots, 0.8)$.

2 The formula of $\hat{\beta}^{KS}$ is given in (5) in the text with $\hat{\rho}_1^{LS}$ obtained from an AR(1) regression of the predictor.

3 $10,000$ replications are conducted.

Table 2: Simulation Results: Autoregression of Predictor

Predictor	T	p	Largest-Root		$\lvert\text{Bias}\rvert$	Variance	RMSE
b/m	88	1	0.886	LS	0.044	0.004	0.078
				II	0.001	0.005	0.069
dfy	90	1	0.828	LS	0.041	0.005	0.081
				II	0.000	0.006	0.074
d/y	137	1	0.980	LS	0.034	0.001	0.049
				II	0.008	0.001	0.032
e/p	137	1	0.785	LS	0.026	0.003	0.062
				II	0.000	0.003	0.058
i/k	62	1	0.773	LS	0.058	0.009	0.112
				II	0.001	0.010	0.102
lty	90	1	0.977	LS	0.052	0.003	0.073
				II	0.016	0.002	0.046
svar	124	1	0.711	LS	0.025	0.005	0.072
				II	0.001	0.005	0.069
tms	89	2	0.407	LS	0.036	0.023	0.153
				II	0.002	0.024	0.155
d/p	137	3	0.966	LS	0.049	0.026	0.163
				II	0.006	0.027	0.164
infl	90	4	0.920	LS	0.098	0.054	0.241
				II	0.006	0.059	0.243
tbl	89	4	0.997	LS	0.084	0.075	0.280
				II	0.041	0.081	0.286
ntis	82	5	0.794	LS	0.078	0.082	0.289
				II	0.006	0.091	0.302
d/e	137	6	0.927	LS	0.094	0.060	0.250
				II	0.004	0.064	0.254
eqis	82	7	0.831	LS	0.153	0.110	0.338
				II	0.011	0.125	0.354

[1] Estimated model: $x_t = \rho_0 + \rho_1 x_{t-1} + \cdots + \rho_p x_{t-p} + v_t$ where x_t is the predictor in the predictive regression model.

[2] $\lvert\text{Bias}\rvert$, Variance, and RMSE are defined as $\sum_{i=1}^{p} \lvert \mathrm{E}(\hat{\rho}_i - \rho_i) \rvert$, $\sum_{i=1}^{p} \mathrm{Var}(\hat{\rho}_i)$ and $\left[\sum_{i=1}^{p} \mathrm{E}(\hat{\rho}_i - \rho_i)^2\right]^{1/2}$, respectively.

Table 3: Simulation Results: Predictive Regression

Predictor	T	β		Bias($\hat{\beta}$)	Var($\hat{\beta}$)	RMSE($\hat{\beta}$)	$\frac{\lvert\text{Bias}(\hat{\beta})\rvert}{\text{std}(\hat{\beta})}$	$\text{Size}^L_{5\%}$	$\text{Size}^R_{5\%}$	$\text{Size}^L_{10\%}$	$\text{Size}^R_{10\%}$
b/m	88	0.131	LS	0.049	0.007	0.097	**0.575**	1.9%	12.5%	3.3%	21.3%
			II	-0.002	0.008	0.088	0.022	10.5%	5.2%	16.8%	10.0%
dfy	90	-0.642	LS	1.005	6.454	2.732	**0.396**	2.6%	9.6%	5.75%	17.4%
			II	0.015	6.781	2.604	0.006	6.1%	4.8%	12.5%	10.0%
d/y	137	0.077	LS	-0.002	0.001	0.037	0.062	6.0%	4.7%	11.2%	9.2%
			II	0.000	0.001	0.037	0.008	5.2%	5.5%	10.2%	10.7%
e/p	137	0.074	LS	0.005	0.001	0.039	0.139	4.2%	6.7%	8.1%	12.4%
			II	0.000	0.001	0.039	0.006	5.5%	5.2%	10.6%	10.0%
i/k	62	-13.2	LS	-0.149	37.710	6.143	0.024	5.5%	4.9%	11.0%	9.9%
			II	0.038	37.974	6.162	0.006	5.3%	5.4%	10.4%	10.6%
lty	90	-0.584	LS	0.138	1.124	1.069	0.130	4.1%	6.5%	8.2%	11.9%
			II	0.018	1.140	1.068	0.017	5.5%	5.3%	11.0%	10.0%
svar	124	0.085	LS	0.062	0.162	0.407	0.155	3.5%	6.4%	7.5%	11.8%
			II	-0.005	0.164	0.405	0.011	5.3%	4.8%	10.7%	9.4%
tms	89	1.497	LS	0.072	2.308	1.521	0.047	4.7%	5.6%	9.3%	11.1%
			II	0.002	2.315	1.522	0.001	5.2%	5.0%	10.3%	10.3%
d/p	137	0.031	LS	0.030	0.002	0.050	**0.761**	0.7%	15.3%	2.1%	26.0%
			II	0.001	0.002	0.042	0.014	10.9%	5.5%	18.0%	10.3%
infl	90	-0.218	LS	0.016	0.236	0.487	0.032	4.8%	5.4%	9.6%	10.4%
			II	0.001	0.238	0.488	0.002	5.3%	5.1%	10.2%	10.1%
tbl	89	-0.592	LS	-0.126	0.395	0.641	**0.200**	7.3%	3.4%	13.8%	7.0%
			II	-0.036	0.403	0.636	0.057	5.4%	5.1%	10.6%	9.9%
ntis	82	-1.450	LS	-0.024	1.485	1.219	0.020	5.3%	5.0%	10.4%	9.8%
			II	-0.011	1.491	1.221	0.009	5.2%	5.1%	10.3%	10.1%
d/e	137	-0.001	LS	0.007	0.002	0.047	0.157	3.9%	6.8%	7.8%	12.7%
			II	0.000	0.002	0.048	0.001	5.7%	5.3%	11.1%	10.0%
eqis	82	-0.470	LS	0.006	0.062	0.249	0.025	5.2%	5.5%	10.1%	10.5%
			II	0.001	0.062	0.249	0.003	5.4%	5.3%	10.5%	10.1%

[1] **Boldface** number denotes the value of $\frac{\lvert\text{Bias}(\hat{\beta})\rvert}{\text{std}(\hat{\beta})}$ is larger than 0.2.

[2] Size^L_α and Size^R_α are respectively the realized sizes of the left-tailed test and right-tailed test with a nominal size of α.

Table 4: Predictors for S&P500 Equity Premium and the AR-Order Selections

Predictor	Definition	Time Span	AR-Order	$\chi^2_{[1]}$	$\chi^2_{[2]}$	$\chi^2_{[3]}$	$\chi^2_{[4]}$
b/m	Book to Market	1921-2008	1	1.511	3.447	3.920	5.890
dfy	Default Yield Spread	1919-2008	1	2.488	2.472	2.837	3.612
d/y	Dividend Yield	1872-2008	1	0.016	2.983	3.219	3.613
e/p	Earning Price Ratio	1872-2008	1	0.694	1.688	1.723	3.431
i/k	Investment Capital Ratio	1947-2008	1	1.592	2.108	3.596	5.545
lty	Long Term Yield	1919-2008	1	1.466	1.416	2.434	3.575
svar	Stock Variance	1885-2008	1	2.569	4.469	4.729	5.945
tms	Term Spread	1920-2008	2	0.607	0.782	5.727	5.219
d/p	Dividend Price Ratio	1872-2008	3	0.609	1.081	2.723	2.756
infl	Inflation	1919-2008	4	0.150	1.413	2.856	3.317
tbl	T-Bill Rate	1920-2008	4	2.538	2.580	2.906	3.617
ntis	Net Equity Expansion	1927-2008	5	0.055	1.525	2.363	3.027
d/e	Dividend Payout Ratio	1872-2008	6	0.536	2.185	5.484	7.250
eqis	Pct Equity Issuing	1927-2008	7	2.193	2.412	2.343	4.192

[1] See Goyal and Welch (2008) and Amit Goyal's website for detailed variable description.

[2] $\chi^2_{[q]}$ is the Breusch-Godfrey LM test statistic, with a null hypothesis that there is no serial correlation up to order q.

[3] **Boldface** number denotes significance at 10% level.

[4] The lag-order is selected by continuously increasing the lag-order of the AR model from a initialization "0", until all the four $\chi^2_{[q]}$ statistics are not significant.

Table 5: Estimation of the Autoregression

Predictor	Method	$\hat{\rho}_0$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3$	$\hat{\rho}_4$	$\hat{\rho}_5$	$\hat{\rho}_6$	$\hat{\rho}_7$	Largest-Root
b/m	LS	0.089	0.842	–	–	–	–	–	–	0.842
	II	0.063	0.886	–	–	–	–	–	–	0.886
dfy	LS	0.003	0.787	–	–	–	–	–	–	0.787
	II	0.002	0.828	–	–	–	–	–	–	0.828
d/y	LS	−0.179	0.946	–	–	–	–	–	–	0.946
	II	−0.071	0.980	–	–	–	–	–	–	0.980
e/p	LS	−0.647	0.760	–	–	–	–	–	–	0.760
	II	−0.581	0.785	–	–	–	–	–	–	0.785
i/k	LS	0.010	0.717	–	–	–	–	–	–	0.717
	II	0.008	0.773	–	–	–	–	–	–	0.773
lty	LS	0.002	0.962	–	–	–	–	–	–	0.962
	II	0.001	0.977	–	–	–	–	–	–	0.977
svar	LS	0.010	0.685	–	–	–	–	–	–	0.685
	II	0.009	0.711	–	–	–	–	–	–	0.711
tms	LS	0.007	0.691	−0.181	–	–	–	–	–	0.426
	II	0.007	0.709	−0.166	–	–	–	–	–	0.407
d/p	LS	−0.314	0.807	−0.164	0.261	–	–	–	–	0.932
	II	−0.163	0.829	−0.158	0.281	–	–	–	–	0.966
infl	LS	0.007	0.671	−0.097	−0.120	0.303	–	–	–	0.866
	II	0.005	0.700	−0.076	−0.125	0.351	–	–	–	0.920
tbl	LS	0.003	1.062	−0.351	0.047	0.158	–	–	–	0.933
	II	−0.000	1.111	−0.349	0.045	0.190	–	–	–	0.997
ntis	LS	0.005	0.626	−0.006	0.019	0.041	−0.025	–	–	0.671
	II	0.004	0.650	0.012	0.020	0.063	−0.012	–	–	0.794
d/e	LS	−0.126	0.799	−0.250	0.115	−0.053	−0.215	0.362	–	0.892
	II	−0.088	0.818	−0.238	0.114	−0.036	−0.229	0.402	–	0.927
eqis	LS	0.075	0.504	0.142	0.023	0.268	−0.280	−0.103	0.044	0.819
	II	0.060	0.529	0.165	0.014	0.302	−0.309	−0.090	0.065	0.831

Table 6: Estimation of the Predictive Regression

Predictor	T		$\hat{\alpha}$	$\hat{\beta}$	$t_{\hat{\beta}}$	$\chi^2_{[1]}$	$\chi^2_{[2]}$	$\chi^2_{[3]}$	$\chi^2_{[4]}$
b/m	88	LS	−0.046	0.180	**2.259**	2.459	2.607	3.432	3.658
		II	−0.018	0.131	**1.645**	2.691	2.846	3.683	4.030
dfy	90	LS	0.052	0.327	0.122	0.751	2.314	2.085	4.328
		II	0.064	−0.642	−0.240	1.067	2.531	2.283	4.379
d/y	137	LS	0.280	0.075	**1.877**	0.088	**4.683**	5.968	6.914
		II	0.286	0.077	**1.930**	0.081	**4.680**	5.952	6.882
e/p	137	LS	0.255	0.079	**1.843**	0.533	**7.102**	**8.830**	**10.148**
		II	0.241	0.074	**1.725**	0.569	**7.149**	**8.863**	**10.171**
i/k	62	LS	0.534	−13.343	**−2.156**	0.119	1.943	2.207	3.344
		II	0.529	−13.225	**−2.137**	0.120	1.936	2.195	3.339
lty	90	LS	0.080	−0.460	−0.591	0.415	2.428	2.158	4.482
		II	0.086	−0.584	−0.750	0.442	2.450	2.19	4.489
svar	124	LS	0.043	0.150	0.337	0.096	**5.011**	6.062	7.177
		II	0.045	0.084	0.190	0.118	**4.967**	6.039	7.141
tms	89	LS	0.038	1.559	1.025	0.498	1.739	1.810	3.580
		II	0.038	1.497	0.984	0.506	1.744	1.816	3.588
d/p	137	LS	0.239	0.061	**1.655**	0.939	**5.317**	**6.773**	**8.181**
		II	0.143	0.031	0.837	1.819	**6.166**	**7.745**	**9.180**
infl	90	LS	0.062	−0.210	−0.458	0.671	2.637	2.729	4.701
		II	0.062	−0.218	−0.475	0.686	2.651	2.751	4.718
tbl	89	LS	0.087	−0.714	−1.025	0.427	1.975	2.016	4.045
		II	0.082	−0.592	−0.852	0.458	1.998	2.041	4.084
ntis	82	LS	0.080	−1.464	**−1.764**	0.151	1.110	0.963	**7.879**
		II	0.080	−1.450	**−1.748**	0.164	1.123	0.961	**7.842**
d/e	137	LS	0.048	0.006	0.115	0.247	**5.955**	**6.974**	**8.266**
		II	0.044	−0.001	−0.009	0.281	**5.946**	**7.010**	**8.298**
eqis	82	LS	0.141	−0.463	**−2.408**	0.121	0.692	0.738	2.406
		II	0.143	−0.470	**−2.441**	0.114	0.687	0.752	2.435

[1] **Boldface** number denotes the significance at 10% level when a two-tailed test is conducted.

[2] The same as note 2 in Table 4.

Table 7: Estimation of the Residual Covariance Matrix

Predictor		$\hat{\sigma}_u^2$	$\hat{\sigma}_v^2$	$\hat{\sigma}_{uv}$	$\hat{\sigma}_{uv}/\hat{\sigma}_v^2$
b/m	LS	3.574×10^{-2}	1.992×10^{-2}	-2.196×10^{-2}	-1.103
	II	3.590×10^{-2}	2.005×10^{-2}	-2.210×10^{-2}	-1.103
dfy	LS	3.811×10^{-2}	2.784×10^{-5}	-6.717×10^{-4}	-24.131
	II	3.812×10^{-2}	2.794×10^{-5}	-6.741×10^{-4}	-24.131
d/y	LS	3.257×10^{-2}	1.984×10^{-2}	1.453×10^{-3}	0.073
	II	3.257×10^{-2}	2.002×10^{-2}	1.464×10^{-3}	0.073
e/p	LS	3.260×10^{-2}	7.081×10^{-2}	-1.439×10^{-2}	-0.203
	II	3.260×10^{-2}	7.090×10^{-2}	-1.441×10^{-2}	-0.203
i/k	LS	2.697×10^{-2}	5.917×10^{-6}	2.033×10^{-5}	3.436
	II	2.697×10^{-2}	5.956×10^{-6}	2.041×10^{-5}	3.427
lty	LS	3.780×10^{-2}	5.945×10^{-5}	-1.425×10^{-4}	-2.396
	II	3.780×10^{-2}	5.962×10^{-5}	-1.438×10^{-4}	-2.412
svar	LS	3.447×10^{-2}	9.273×10^{-4}	-2.401×10^{-3}	-2.590
	II	3.447×10^{-2}	9.282×10^{-4}	-2.404×10^{-3}	-2.590
tms	LS	3.700×10^{-2}	1.208×10^{-4}	-3.096×10^{-4}	-2.562
	II	3.700×10^{-2}	1.208×10^{-4}	-3.096×10^{-4}	-2.582
d/p	LS	3.275×10^{-2}	4.139×10^{-2}	-3.015×10^{-2}	-0.728
	II	3.291×10^{-2}	4.175×10^{-2}	-3.042×10^{-2}	-0.729
infl	LS	3.802×10^{-2}	9.331×10^{-4}	-2.603×10^{-4}	-0.279
	II	3.802×10^{-2}	9.430×10^{-4}	-2.755×10^{-4}	-0.292
tbl	LS	3.700×10^{-2}	1.904×10^{-4}	3.344×10^{-4}	1.756
	II	3.701×10^{-2}	1.957×10^{-4}	3.439×10^{-4}	1.757
ntis	LS	3.767×10^{-2}	1.919×10^{-4}	9.756×10^{-5}	0.509
	II	3.767×10^{-2}	1.930×10^{-4}	5.340×10^{-5}	0.277
d/e	LS	3.342×10^{-2}	5.095×10^{-2}	-1.352×10^{-2}	-0.265
	II	3.342×10^{-2}	5.120×10^{-2}	-1.369×10^{-2}	-0.267
eqis	LS	3.647×10^{-2}	4.165×10^{-3}	-5.725×10^{-4}	-0.137
	II	3.647×10^{-2}	4.195×10^{-3}	-6.826×10^{-4}	-0.163