Hitotsubashi Journal of Economics 54 (2013), pp.231-250. Ⓒ Hitotsubashi University

ESTIMATION AND INFERENCE IN PREDICTIVE REGRESSIONS*

EIJI KUROZUMI**

Graduate School of Economics, Hitotsubashi University
Kunitachi, Tokyo 186-8601, Japan
kurozumi@stat.hit-u.ac.jp

AND

KOHEI AONO

Faculty of Business Administration, Kyoto Sangyo University
Kyoto, Kyoto 603-8555, Japan
kaono@cc.kyoto-su.ac.jp

Accepted July 2013

Abstract

In this paper, we analyze feasible bias-reduced versions of point estimates for predictive regressions: The plug-in estimates, which are based on the augmented regressions proposed by Amihud and Hurvich (2004) and Amihud, Hurvich and Wang (2010), and the grouped jackknife estimate by Quenouille (1949, 1956). We also derive the correct standard errors associated with these point estimates. The methods thus allow for a unified inferential framework, where point estimates and statistical inference are based on the same methods. Using the new estimates, we investigate U.S. stock returns and find that some variables are able to predict stock returns.

Keywords: near unit root, bias, stock return, jackknife
JEL Classification Codes: C13, C22, C58, G17

I. *Introduction*

Predictive regressions have long been studied in the financial and econometric literature. One of the difficulties in predicting stock returns by financial variables is that the ordinary least squares (OLS) estimate is severely biased, as pointed out by Mankiw and Shapiro (1986) and

* We are grateful to In Choi, Kaddour Hadri, Tokuo Iwaisako and Teruo Nakatsuma for useful comments. Kurozumi's research was supported by the Seimeikai Foundation and the Global COE program, Research Unit for Statistical and Empirical Analysis in Social Sciences, Hitotsubashi University. Aono's research was supported by Fellowship of Japan Center for Economic Research and the Japan Society for Promotion of Science (No. 24730255). All errors are our responsibility.

** Corresponding author.

Elliott and Stock (1994). This bias comes from the fact that typical predictor variables are strongly serially correlated and also contemporarily correlated with prediction errors. In the case of the following model,

$$r_t = \mu_r + \beta x_{t-1} + u_t, \tag{1}$$

$$x_t = \mu_x + \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + e_t, \tag{2}$$

the typical case is that $\rho = \phi_1 + \cdots + \phi_p$ is close to one while $Cov(u_t, e_t)$ is close to -1. Because of the upward bias, the usual t-test suffers from over-size distortions, and hence, we tend to erroneously find evidence of strong predictability.

In order to overcome the problem of the over-size distortions, many efforts have been made and testing procedures have been developed in the literature. Elliott and Stock (1994) considered approximating the distribution of test statistics by the Bayesian mixture procedure, while Cavanagh, Elliott and Stock (1995) proposed to construct test statistics that are free from the true value of ρ by using three intervals—sup-bound, Bonferroni, and Scheffe-type. The latter methods have been applied to stock returns by Torous, Valkanov and Yan (2004), who found evidence of predictability only at short horizons. Stambaugh (1999) tackled this problem from the Bayesian point of view and based on his theory, Lewellen (2004) developed the bias adjustment of the OLS estimate of β. He found stock returns predictability using new tests by combining the t-tests based on the OLS estimate of β and the bias-adjusted estimate. Campbell and Yogo (2006) discussed the optimality of predictive regressions. Theoretically, we can construct conditional optimal tests as developed by Jansson and Moreira (2006), which investigated the optimality in the class of location invariant tests, but such tests require advanced computational methods. Instead, Campbell and Yogo (2006) proposed Bonferroni intervals based on conditionally optimal tests and found stock return predictability. Amihud, Hurvich and Wang (2008) considered the augmented regressions with the bias corrected least squares estimates of the AR coefficients, while Chen and Deo (2009) developed the restricted likelihood-based test using the Bartlett-correction. On the other hand, Lanne (2002) tested predictability not by estimating (1) but by applying stationarity tests, and found it difficult to predict stock returns.

Once the evidence of predictability is observed, it is often the case that we need the point estimate of β. For example, we need the point estimate when actually forecasting future returns by predictive regressions. Another example is the case where a change in predictability is observed, as is often the case with the U.S. stock returns. In this case, we may want to examine the magnitude of change in predictability, which can be measured by using the point estimates of β in two subsamples. Note that although the existing testing methods may give the confidence intervals of β, they do not necessarily give the point estimates of β. Therefore, new estimation methods for the predictive regressions have been developed recently. Amihud and Hurvich (2004) developed the unbiased estimate of β based on the least squares method for possibly multiple AR(1) predictors under the assumption of normality, which is extended to the AR(p) model by Amihud, Hurvich and Wang (2010). Eliasz (2004) proposed the median unbiased estimate for a single AR(1) predictor by inverting the Jansson and Moreira (2006) test statistic, while Chen and Deo (2009) developed the estimate for multiple predictors based on the restricted likelihood, which is free of location parameters and possesses small curvature. Chiquoine and Hjalmarsson (2009) applied the jackknife method, which was originally developed by Quenouille (1949, 1956), to the predictive regressions.

In this paper, we investigate the finite sample properties of the existing estimates and show that the augmented regressions considered by Amihud and Hurvich (2004) and Amihud, Hurvich and Wang (2010) and the jackknife estimates by Chiquoine and Hjalmarsson (2009) are useful and easy to implement. We also derive the correct standard errors accosted with these point estimates. As a result, the estimation and inference can be made in a unified framework where point estimates and statistical inference are based on the same method, which is obviously the usual case but is not necessarily so in the case of predictive regressions; the tests are often based on some conservative method that does not produce a unique point estimates and the reported point estimates is often a simple OLS estimate. The important finite sample properties of the estimates are (i) the grouped jackknife method is quite useful to reduce the bias of the estimate, (ii) one of the plug-in estimates proposed in this paper has correct coverage rates of the confidence intervals. Considering these favorable finite sample properties, our unified framework will complement the existing testing methods to investigate the predictability of stock returns.

In the empirical analysis, we first apply our methods to the U.S. monthly and annual stock returns investigated by Campbell and Yogo (2006). We find the predictability of stock returns using the same predictor variables as observed by Campbell and Yogo (2006). In addition, we also find strong evidence of predictability by the dividend-price ratio, which is not considered as a statistically significant predictor. We then investigate the same stock returns by extending the sample period. Again, we clearly observe the predictability of stock returns by the dividend-price ratio.

The rest of the paper is organized as follows. Section 2 reviews the augmented regressions by Amihud and Hurvich (2004) and Amihud, Hurvich and Wang (2010), the jackknife estimate used by Chiquoine and Hjalmarsson (2009) and the restricted maximum likelihood method by Chen and Deo (2009). We also propose versions of the augmented regression estimate and derive the standard errors of the point estimates. We investigate the finite sample properties of the estimates in Section 3. The predictability of the U.S. stock returns is investigated in Section 4, and the concluding remarks are given in Section 5. The technical derivations are relegated to the Appendix.

II. *Estimation Methods for Predictive Regressions*

1. The Feasible Augmented Regression

Let us first briefly review the bias corrected estimates based on the feasible augmented regression. For model (1)–(2), we assume that $u_t \sim i.i.d.(0, \sigma_u^2)$, $e_t \sim i.i.d.(0, \sigma_e^2)$, u_t and e_t are contemporaneously correlated with covariance σ_{ue}, and the AR coefficients $\phi_1, \cdots, \phi_p$ satisfy the stationary condition. In this case, it is well known that the OLS estimate of β,

$$\hat{\beta} = \frac{\sum_{t=1}^{T} \tilde{x}_{t-1} r_t}{\sum_{t=1}^{T} \tilde{x}_{t-1}^2} = \beta + \frac{\sum_{t=1}^{T} \tilde{x}_{t-1} u_t}{\sum_{t=1}^{T} \tilde{x}_{t-1}^2}, \tag{3}$$

where $\tilde{x}_{t-1}=x_{t-1}-\sum_{t=1}^{T} x_{t-1}/T$, is shifted and skewed to the right when σ_{ue} is negative, and hence, $\hat{\beta}$ is positively biased.

In the case of infeasible regressions with known $\phi_1, \cdots, \phi_p$, model (1) can be transformed into

$$r_t=\mu_r^*+\beta x_{t-1}+\frac{\sigma_{ue}}{\sigma_e^2}(x_t-\phi_1 x_{t-1}-\cdots-\phi_p x_{t-p})+\left(u_t-\frac{\sigma_{ue}}{\sigma_e^2}e_t\right), \tag{4}$$

where $\mu_r^*=\mu_r-(\sigma_{ue}/\sigma_e^2)\mu_x$. Then, by regressing r_t on a constant, x_{t-1} and $(x_t-\phi_1 x_{t-1}-\cdots-\phi_p x_{t-p})$, the OLS estimate of β, denoted by $\hat{\beta}^*$, becomes

$$\hat{\beta}^*=\beta+\frac{\sum_{t=1}^{T}\tilde{x}_{t-1}\left(u_t-\frac{\sigma_{ue}}{\sigma_e^2}e_t\right)}{\sum_{t=1}^{T}\tilde{x}_{t-1}^2}+o_p\left(\frac{1}{\sqrt{T}}\right). \tag{5}$$

Since the asymptotic variance of $\hat{\beta}^*$ is $(\sigma_u^2-\sigma_{ue}^2/\sigma_e^2)/Var(x_t)$, which is smaller than the asymptotic variance of $\hat{\beta}$ given by $\sigma_u^2/Var(x_t)$, we can see that $\hat{\beta}^*$ is more efficient than $\hat{\beta}$. Moreover, Amihud and Hurvich (2004) showed that for the AR(1) case, $\hat{\beta}^*$ is unbiased under the assumption of normality [1].

Although augmented regression (4) is desirable, it is infeasible in practice because we do not know the true values of $\phi_1, \cdots, \phi_p$. Instead, Amihud and Hurvich (2004) proposed to plug-in the bias corrected estimate of ϕ_1 in the AR(1) case and proposed the following feasible augmented regression:

$$r_t=\mu_r+\beta x_{t-1}+\gamma(\tilde{x}_t-\hat{\phi}_{1,c}\,\tilde{x}_{t-1})+v_t, \tag{6}$$

where $\hat{\phi}_{1,c}=\hat{\phi}_1+(1+3\hat{\phi}_1)/T+3(1+3\hat{\phi}_1)/T^2$ with $\hat{\phi}_1$ being the OLS estimate for (2) with $p=1$ and v_t represents the regression error. This bias corrected estimate of ϕ_1 is motivated from the well-known fact that $E[\hat{\phi}_1]=\phi_1-(1+3\phi_1)/T+O(1/T^2)$. The expression of $\hat{\phi}_{1,c}$ is obtained by repeatedly replacing ϕ_1 with $\hat{\phi}_1$ in the expression of $E[\hat{\phi}_1]$. Amihud, Hurvich and Wang (2010) extended model (1)–(2) to the AR(p) model for x_t and the multiple predictors given by $x_{t-1}, \cdots, x_{t-p}$ and proposed the similar feasible augmented regression.

2. Other Versions of the Plug-in Estimate

Note that regression (6) proposed by Amihud and Hurvich (2004) is only one feasible version of regression (4) and there might be other estimates whose finite sample property is better than the Amihud-Hurvich estimate. In the following, we consider several estimates of

---

[1] Although we consider only the stationary case, we can show that $\hat{\beta}^*$ is asymptotically unbiased without the assumption of normality even when the AR(1) coefficient is characterized as the local-to-unity given by $\phi_1=1-c/T$ where c is some constant. This is because $u_t-(\sigma_{ue}/\sigma_e^2)e_t$ is an i.i.d. sequence uncorrelated with a sequence of e_t, which drives x_t, so that the limiting distribution of $\hat{\beta}^*$ is mixed normal.

$\phi_1, \cdots, \phi_p$ and propose to plug-in them into infeasible regression (4).

Since the infeasible efficient estimate is expressed as (5), the feasible version can be defined as, by replacing the unknown parameters with some estimates and omitting the $o_p(1/\sqrt{T})$ term,

$$\begin{aligned}\tilde{\beta} &\equiv \beta + \frac{\sum_{t=1}^{T} \tilde{x}_{t-1}\left(u_t - \frac{\hat{\sigma}_{ue}}{\hat{\sigma}_e^2}\tilde{e}_t\right)}{\sum_{t=1}^{T} \tilde{x}_{t-1}^2} \\ &= \hat{\beta} - \frac{\hat{\sigma}_{ue}}{\hat{\sigma}_e^2} \frac{\sum_{t=1}^{T} \tilde{x}_{t-1}\tilde{e}_t}{\sum_{t=1}^{T} \tilde{x}_{t-1}^2}, \end{aligned} \tag{7}$$

where $\hat{\sigma}_{ue}$ and $\hat{\sigma}_e^2$ are consistent estimates of σ_{ue} and σ_e^2, which can be constructed from the regression residuals in (1) and (2), and $\tilde{e}_t = \tilde{x}_t - \tilde{\phi}_1 \tilde{x}_{t-1} - \cdots - \tilde{\phi}_p \tilde{x}_{t-p}$, where $\tilde{x}_{t-j} = x_{t-j} - \sum_{t=j+1}^{T} x_{t-j}/(T-j)$ for $j=1, \cdots, p-1$, is the estimated residual with $\tilde{\phi}_1, \cdots, \tilde{\phi}_p$ being some estimates. From (7), it is observed that the feasible estimate is based on the OLS estimate $\hat{\beta}$ but the bias is adjusted by the second term on the right hand side of (7). We can also see that bias adjustment based on the least squares estimate is meaningless because it is well known that $\hat{e}_t$, the least squares residual, is orthogonal to $\tilde{x}_{t-1}$, that is, $\sum_{t=1}^{T} \tilde{x}_{t-1}\hat{e}_{t-1} = 0$, which implies that $\tilde{\beta} = \hat{\beta}$ in this case. We call $\tilde{\beta}$ the plug-in estimate.

In a special case of $p=1$, the plug-in estimate can be expressed in more compact form. Since

$$\begin{aligned}\frac{\sum_{t=1}^{T} \tilde{x}_{t-1}\tilde{e}_t}{\sum_{t=1}^{T} \tilde{x}_{t-1}^2} &= \frac{\sum_{t=1}^{T} \tilde{x}_{t-1}(\tilde{x}_t - \tilde{\phi}_1 \tilde{x}_{t-1})}{\sum_{t=1}^{T} \tilde{x}_{t-1}^2} \\ &= \frac{\sum_{t=1}^{T} \tilde{x}_{t-1}\{e_t - (\tilde{\phi}_1 - \phi_1)\tilde{x}_{t-1}\}}{\sum_{t=1}^{T} \tilde{x}_{t-1}^2} \\ &= (\hat{\phi}_1 - \phi_1) - (\tilde{\phi}_1 - \phi) \\ &= \hat{\phi}_1 - \tilde{\phi}_1, \end{aligned}$$

where we used the fact that $\hat{\phi}_1 - \phi_1 = \sum_{t=1}^{T} \tilde{x}_{t-1}e_t / \sum_{t=1}^{T} \tilde{x}_{t-1}^2$, we can see that

$$\tilde{\beta} = \hat{\beta} - \frac{\hat{\sigma}_{ue}}{\hat{\sigma}_e^2}(\hat{\phi}_1 - \tilde{\phi}_1),$$

so that the bias adjustment term can be expressed as the difference between the OLS estimate $\hat{\phi}_1$ and the other estimate in the AR(1) case.

(a) Cauchy estimate

To construct the plug-in estimate, we need $\tilde{e}_t$ or $\tilde{\phi}_1, \cdots, \tilde{\phi}_p$. The first estimate we consider is the Cauchy estimate proposed by So and Shin (1999)[2]. This estimate is obtained by the instrumental variable method with a special instrument and by a detrending method other than the usual demeaning. The advantage of this method is that the t-statistic is asymptotically normal irrespective of whether the process is stationary, nearly integrated or integrated, although the estimate is median unbiased but biased in a usual sense in the nearly integrated and integrated cases.

To estimate the model, we first express model (2) as the recursive OLS mean adjustment form given by

$$x_t-\bar{x}_{t-1}=\rho(x_{t-1}-\bar{x}_{t-1})+\psi_1\Delta x_{t-1}+\cdots+\psi_{p-1}\Delta x_{t-p+1}+\varepsilon_t, \tag{8}$$

where $\bar{x}_{t-1}=\sum_{s=1}^{t-1}x_s/(t-1)$, ε_t denotes the error term, $\rho=\phi_1+\cdots+\phi_p$ and $\psi_j=-\sum_{i=j+1}^{p}\phi_i$ for $j=1, \cdots, p-1$. Usually, we subtract the sample mean of x and (8) would be expressed using $\tilde{x}_t$ and $\tilde{x}_{t-1}$. However, So and Shin (1999) recommended subtracting the recursive OLS mean $\bar{x}_{t-1}$ because of the nice finite sample property of the estimate; this method is called the recursive OLS mean adjustment. From (8) we obtain the instrumental variable estimates $\hat{\rho}_{c, rols}$, $\hat{\psi}_{1, c, rols}$, $\cdots$, $\hat{\psi}_{p-1, c, rols}$ with $\mathrm{sign}(x_t-\bar{x}_{t-1})$ and $\Delta x_{t-1}, \cdots, \Delta x_{t-p+1}$ as the instruments and denote the estimate of the $p\times p$ variance-covariance matrix of the estimated coefficients as $\hat{\Sigma}_{c, rols}$. The plug-in estimate of β is obtained by replacing $\tilde{e}_t$ in (7) with $\tilde{e}_{c, rols, t}=\tilde{x}_t-\hat{\rho}_{c, rols}\tilde{x}_{t-1}-\hat{\psi}_{1, c, rols}\Delta x_{t-1}-\cdots-\hat{\psi}_{p-1, c, rols}\Delta x_{t-p+1}$ and denoted by $\tilde{\beta}_{c, rols}$.

We may be able to consider another version of the Cauchy estimate by changing the demeaning method. It is often the case for the financial time series that the autoregressive root or ρ is close to one and in such a case, the GLS detrending method by Elliott, Rothenberg and Stock (1996) may be useful to obtain the efficient estimates of the coefficients. In the case of the Cauchy estimate, we used the recursive GLS detrending method and estimate the following regression,

$$x_t-\hat{m}_{x, gls, t-1}=\rho(x_{t-1}-\hat{m}_{x, gls, t-1})+\psi_1\Delta x_{t-1}+\cdots+\psi_{p-1}\Delta x_{t-p+1}+\varepsilon_t \tag{9}$$

by the instrumental variable method with the instruments given by $\mathrm{sign}(x_t-\hat{m}_{x, gls, t-1})$ and $\Delta x_{t-1}, \cdots, \Delta x_{t-p+1}$, where

$$\hat{m}_{x, gls, t}=\frac{\sum_{s=1}^{t}1_s^{qd}x_s^{qd}}{\sum_{s=1}^{t}(1_s^{qd})^2}\quad \text{with}\quad x_t^{qd}=\begin{cases}x_1 & : \ t=1\\ x_t-\dfrac{7}{T}x_{t-1} & : \ t\geq 2,\end{cases}\qquad 1_t^{qd}=\begin{cases}1 & : \ t=1\\ \dfrac{7}{T} & : \ t\geq 2.\end{cases}$$

Using the instrumental variable estimates denoted as $\hat{\rho}_{c, rgls}$, $\hat{\psi}_{1, c, rgls}$, $\cdots$, $\hat{\psi}_{p-1, c, rgls}$, we construct $\tilde{e}_{c, rgls, t}=\tilde{x}_t-\hat{\rho}_{c, rgls}\tilde{x}_{t-1}-\hat{\psi}_{1, c, rgls}\Delta x_{t-1}-\cdots-\hat{\psi}_{p-1, c, rgls}\Delta x_{t-p+1}$ and obtain the plug-in estimate $\tilde{\beta}_{c, rgls}$ by replacing $\tilde{e}_t$ with $\tilde{e}_{c, rgls, t}$ in (7). We denote the estimate of the $p\times p$ variance-covariance

[2] In addition to the estimates proposed in this paper, we also considered the weighted symmetric estimate by Park and Fuller (1995), the lag-augmented method by Choi (1993) and Toda and Yamamoto (1995) and its modified version by Kurozumi and Yamamoto (2000), combined with several detrending methods. The mid-point of the confidence interval proposed by Campbell and Yogo (2006) is also investigated as the point estimate. Since their finite sample performance was unsatisfactory, we do not report the result.

matrix of the estimated coefficients as $\hat{\Sigma}_{c,rgls}$, which will be used later to construct the confidence interval of $\tilde{\beta}_{c,rgls}$. We will see in the next section that the recursive GLS detrending is effective to reduce the MSE in finite samples.

(b) Grouped jackknife estimate

The second method we considered is the grouped jackknife estimate by Quenouille (1949, 1956), which is also applied to the option prices by Phillips and Yu (2005) and to the predictive regressions by Chiquoine and Hjalmarsson (2009). Let us express model (2) as

$$x_t=\mu_x+\varphi' z_{t-1}+e_t,$$

where $\varphi=[\phi_1, \cdots, \phi_p]'$ and $z_{t-1}=[x_{t-1}, \cdots, x_{t-p}]'$. Let $\hat{\varphi}=[\hat{\phi}_1, \cdots, \hat{\phi}_p]'$ be the OLS estimate of φ using the whole sample. Next, we split the whole sample into two subsamples ($t=1, \cdots, T/2$ and $t=T/2+1, \cdots, T$) and estimate the model in each subsample. let $\hat{\varphi}_1$ and $\hat{\varphi}_2$ be the OLS estimates from the first and the second subsamples, respectively. Then, the grouped jackknife estimate of φ is defined as

$$\hat{\varphi}_{j,2}=2\hat{\varphi}-\frac{\hat{\varphi}_1+\hat{\varphi}_2}{2}.$$

The jackknife method is motivated from the fact that $\hat{\varphi}_{j,2}$ does not have the first order bias and $E[\hat{\varphi}_{j,2}]=\varphi+O(1/T^2)$ in our case. To see this, recall that, the first order bias is proportional to $1/T$ such that $E[\hat{\varphi}]=\varphi+C_1/T+O(1/T^2)$ where C_1 is a constant independent of the sample size, which is also used in Amihud and Hurvich (2004). Since $\hat{\varphi}_1$ and $\hat{\varphi}_2$ are obtained from the half of the whole sample, we can see that $E[\hat{\varphi}_1]=E[\hat{\varphi}_2]=\varphi+2C_1/T+O(1/T^2)$. From these relations, we can easily see that $E[\hat{\phi}_{j,2}]=\phi+O(1/T^2)$.

In general, we can split the whole sample into m subsamples for $m\geq 2$ and denote the OLS estimates in the subsamples as $\hat{\varphi}_{j,1}, \cdots, \hat{\varphi}_{j,m}$. Then, the grouped jackknife estimate of ϕ is defined as

$$\hat{\varphi}_{j,m}=\frac{m}{m-1}\hat{\varphi}-\frac{1}{m(m-1)}(\hat{\varphi}_1+\cdots+\hat{\varphi}_m).$$

Again, it can be shown that the first order bias disappears from $\hat{\varphi}_{j,m}$ and $E[\hat{\varphi}_{j,m}]=\varphi+O(1/T^2)$. Then, the plug-in estimate is obtained by replacing $\tilde{e}_t$ in (7) with $\tilde{e}_{j,m,t}=\tilde{x}_t-\hat{\varphi}'_{j,m}\tilde{z}_{t-1}$ where $\tilde{z}_{t-1}=[\tilde{x}_{t-1}, \cdots, \tilde{z}_{t-p}]'$. We denote this estimate as $\tilde{\beta}_{j,m}$.

It is known that the jackknife method with larger number of sub-groups may reduce the MSE of the estimate whereas the bias might increase because we cannot ignore the higher order bias if the number of observations in subsamples are too small. We consider $m=2$, 3 and 4 in the following section.

3. Direct Application of the Jackknife Method

We have so far considered to reduce the bias of the least squares estimate $\hat{\beta}$ through the estimation of the AR parameter in (2). However, the formula of the plug-in estimate is available only when model (1)–(2) is assumed and we need to change this formula if the underlying model has been changed. For example, as discussed in Amihud and Hurvich (2004), Amihud, Hurvich and Wang (2008) and Chen and Deo (2009), if we consider predictive regressions with multiple predictors, the modification may be straightforward if the coefficient matrices associated with the vector autoregressive representation of the predictors are assumed to be diagonal. However, this is not the case in general and then the modification would be more complicated if the off-diagonal elements are taken into account. On the other hand, if a single predictor variable is used but several lagged variables are included as predictors as considered in Amihud, Hurvich and Wang (2010), the modification may be different from those considered in the standard multiple predictors case. Moreover, there is no guarantee that the predictor variables are well characterized as AR(p) models but we may have to consider more general models. In such cases, we have to investigate how to reduce the bias of the estimate and have to find the bias correction formula depending on the models, which is tedious in practice. In this sense, the plug-in estimate considered in the previous subsections is limited, although it is useful in some cases because some financial time series is known to be well approximated by autoregressive processes.

On the other hand, we can directly apply the jackknife method for the estimation of β without assuming specific models for the predictor variables as is considered in Chiquoine and Hjalmarsson (2009). To see this, let us consider general multiple predictors model given by

$$r_t = \mu_r + \boldsymbol{\beta}' \boldsymbol{x}_{t-1} + e_t, \tag{10}$$

where $\boldsymbol{x}_{t-1}$ is a q-dimensional predictor variable uncorrelated with e_t and $\boldsymbol{\beta}$ is an associated coefficient. As explained before, let $\hat{\boldsymbol{\beta}}$ be the least squares estimate using the whole sample while $\hat{\boldsymbol{\beta}}_1, \cdots, \hat{\boldsymbol{\beta}}_m$ be the subsample estimates when the whole sample is split into m-subsamples. Then, the grouped jackknife estimate is defined by

$$\hat{\boldsymbol{\beta}}_{j,m} = \frac{m}{m-1}\hat{\boldsymbol{\beta}} - \frac{1}{m(m-1)}(\hat{\boldsymbol{\beta}}_1 + \cdots + \hat{\boldsymbol{\beta}}_m). \tag{11}$$

As long as $\boldsymbol{x}_{t-1}$ is stationary, which is often assumed in the literature, it is known that the first order bias of the least squares estimate is proportional to $1/T$ and hence we can see that $\hat{\boldsymbol{\beta}}_{j,m}$ has only the second order bias. This is a great advantage over the other estimates because we do not have to assume any specific structure for $\boldsymbol{x}_{t-1}$, so that we can apply the jackknife method in various situations. We will investigate the difference in finite samples between the plug-in estimate with the jackknife method and the direct jackknife estimate in the next section.

4. Restricted Likelihood Estimate

Instead of relying on the least squares principle, Chen and Deo (2009) proposed to estimate the predictive regression using the likelihood method. They first pointed out that the maximum likelihood estimate behaves very well when there is no constant whereas it behaves badly with a constant. They then proposed to estimate the model by maximizing the restricted likelihood function, which is free of the nuisance intercept parameter. They also found the bias expression for the restricted maximum likelihood (REML) estimate and proposed the bias corrected version of the REML estimate. We omit the exact expression of the estimate because it is complicated. See Chen and Deo (2009) for details. Chen and Deo (2009) showed that their estimate is less biased than the Amihud and Hurvich (2004) estimate. Further, their test based on the restricted likelihood function is more powerful than the test proposed by Campbell and Yogo (2006) and Jansson and Moreira (2006).

Although the REML method is theoretically interesting and the bias corrected REML estimate is less biased according to Chen and Deo (2009), its finite sample property is much affected by the true distribution of the error term and the initial value condition on the predictor variable. In fact, our preliminary simulation (which is not reported to save space) shows that the bias corrected REML estimate is less biased when the error term is normal with stationary initial values whereas it is more biased either when the distribution of the error term is log-normal or when the initial value of x_t is fixed at $x_0=0$. Our simulation shows that neither the REML estimate nor the Amihud-Hurvich estimate dominate the other estimate; their finite sample performance depends on the distributions of the error terms and the initial value condition on x_t.

5. Confidence Intervals

The estimates considered in this section, except for the REML estimate [3], are shown to be asymptotically normal as long as the predictor variable in (2) is stationary and we can construct the confidence intervals based on those estimates. Since the plug-in estimate is given by (7), what we need is the variance estimate of $\tilde{\beta}$.

As explained in the Appendix, the variance estimate of $\tilde{\beta}_{c,\,rols}$ is given by

$$\widehat{Var(\tilde{\beta}_{c,\,rols})}=\widehat{Var(\hat{\beta}^*)}+\left(\frac{\hat{\sigma}_{ue}}{\hat{\sigma}_e^2}\right)^2 J'\,\hat{\Sigma}_{c,\,rols}\,J, \tag{12}$$

where $Var(\hat{\beta}^*)=\dfrac{\hat{\sigma}_e^2\hat{\sigma}_u^2-\hat{\sigma}_{ue}^2}{\hat{\sigma}_e^2\sum_{t-1}^{T}\tilde{x}_{t-1}^2}$ and $J=\left[1,\ \dfrac{\sum_{t=1}^{T}\tilde{x}_{t-1}\Delta x_{t-1}}{\sum_{t=1}^{T}\tilde{x}_{t-1}^2},\ \cdots,\ \dfrac{\sum_{t=1}^{T}\tilde{x}_{t-1}\Delta x_{t-p+1}}{\sum_{t=1}^{T}\tilde{x}_{t-1}^2}\right]'$.

Similarly, when we plug-in the Cauchy estimate with the recursive GLS detrending method, the variance estimate is given by (12) with $\hat{\Sigma}_{c,\,rols}$ replaced by $\hat{\Sigma}_{c,\,rgls}$.

Expression for the variance estimate of the plug-in estimate using the jackknife method is expressed as

$$\widehat{Var(\tilde{\beta}_{j,\,m})}=\widehat{Var(\hat{\beta}^*)}+\left(\frac{\hat{\sigma}_{ue}}{\hat{\sigma}_e^2}\right)^2 K'\hat{\Sigma}_{j,\,m}K \quad , \tag{13}$$

where $\hat{\Sigma}_{j,\,m}$ is the variance estimate of $\hat{\varphi}_{j,\,m}$ and

$$K=\left[1,\ \frac{\sum_{t=1}^{T}\tilde{x}_{t-1}\tilde{x}_{t-2}}{\sum_{t=1}^{T}\tilde{x}_{t-1}^2},\ \cdots,\ \frac{\sum_{t=1}^{T}\tilde{x}_{t-1}\tilde{x}_{t-p}}{\sum_{t=1}^{T}\tilde{x}_{t-1}^2}\right]'.$$

[3] Although the REML estimate is expected to be asymptotically normal, its limiting distribution is not derived by Chen and Deo (2009).

The estimate $\hat{\Sigma}_{j,m}$ is obtained as follows. Note that

$$E[(\hat{\varphi}_{j,m}-\varphi)(\hat{\varphi}_{j,m}-\varphi)']=E[\{(\hat{\varphi}-\varphi)+(\hat{\varphi}_{j,m}-\hat{\varphi})\}\{(\hat{\varphi}-\varphi)+(\hat{\varphi}_{j,m}-\hat{\varphi})\}']$$
$$=Var(\hat{\varphi})+E[(\hat{\varphi}_{j,m}-\hat{\varphi})(\hat{\varphi}_{j,m}-\hat{\varphi})']$$
$$+E[(\hat{\varphi}-\varphi)(\hat{\varphi}_{j,m}-\hat{\varphi})]+E[(\hat{\varphi}_{j,m}-\hat{\varphi})(\hat{\varphi}-\varphi)'].$$

Since $\hat{\varphi}$ is efficient while $\hat{\varphi}_{j,m}$ is inefficient, we can see from Lemma 2.1 in Hausman (1978) that the last two terms asymptotically equal zero. Then, the variance estimate of $\hat{\varphi}_{j,m}$ is given by

$$\hat{\Sigma}_{j,m}=\widehat{Var(\hat{\varphi})}+(\hat{\varphi}_{j,m}-\hat{\varphi})(\hat{\varphi}_{j,m}-\hat{\varphi})', \tag{14}$$

where $\widehat{Var(\hat{\varphi})}$ is the variance estimate of the OLS estimate using the whole sample. The same expression is obtained by Yamamoto and Kurozumi (2005) in a different manner in a different situation.

Note that both (12) and (13) are expressed as the sum of the variance estimate of the infeasible efficient estimate $\hat{\beta}^*$ and the additional term, which is nonnegative. In other words, the plug-in estimate cannot be as efficient as the infeasible estimate $\hat{\beta}^*$ because of the uncertain variation of the estimated AR coefficients unless $\hat{\sigma}_{ue}=0$. We can also show that when $p=1$ and $\tilde{\phi}_1$ is plugged-in, the variance estimate becomes in simpler form as

$$\widehat{Var(\tilde{\beta})}=\widehat{Var(\hat{\beta}^*)}+\left(\frac{\hat{\sigma}_{ue}}{\hat{\sigma}_e^2}\right)^2\widehat{Var(\tilde{\phi}_1)}.$$

In exactly the same way as (14), the variance estimates for (11) is given by

$$\widehat{Var(\hat{\boldsymbol{\beta}}_{j,m})}=\widehat{Var(\hat{\boldsymbol{\beta}})}+(\hat{\boldsymbol{\beta}}_{j,m}-\hat{\boldsymbol{\beta}})(\hat{\boldsymbol{\beta}}_{j,m}-\hat{\boldsymbol{\beta}})', \tag{15}$$

where $\widehat{Var(\hat{\boldsymbol{\beta}})}$ is the OLS estimate of the variance of $\hat{\boldsymbol{\beta}}$ using the whole sample.

In any case, because the limiting distribution of the point estimates considered in this paper is normal, we can easily construct the confidence intervals in a standard way using the variance estimates.

III. *Finite Sample Properties*

In this section, we investigate the finite sample properties of the estimates considered in the previous section by Monte Carlo simulations. The data generating process is the same as (1)–(2) with $p=1$, $u_t \sim i.i.d.N(0, 1)$, $e_t \sim i.i.d.N(0, 1)$, $Cov(u_t, e_t)=\sigma_{ue}$, and $x_0=0$. We set $\mu_r=\mu_x=0$, $\beta=0$, $\rho=0.7$, 0.8, 0.9, 0.95, and 0.99; $\sigma_{ue}=-0.95$ and -0.55; and $T=50$, 100, 250, and 500. Although the true values of μ_r and μ_x equal zero, the models are estimated with a constant term. For the jackknife method, we consider $m=2$, 3 and 4 in this simulation. The number of replications is 10, 000 and all computations are conducted by using the GAUSS matrix language.

Table 1 summarizes the biases of the various estimates when $\sigma_{ue}=-0.95$. In the table, the

Table 1. Bias of Various Estimates

ϕ_1	T	$\tilde{\beta}_{ah}$	$\tilde{\beta}_{c,ols}$	$\tilde{\beta}_{c,gls}$	$\tilde{\beta}_{j,2}$	$\tilde{\beta}_{j,3}$	$\tilde{\beta}_{j,4}$	$\hat{\beta}_{j,2}$	$\hat{\beta}_{j,3}$	$\hat{\beta}_{j,4}$	$\hat{\beta}_{reml}$
	50	0.0046	0.0121	0.0492	0.0019	−0.0004	0.0021	0.0019	0.0049	0.0079	0.0165
	100	0.0015	0.0034	0.0269	−0.0002	−0.0007	0.0003	−0.0002	0.0008	0.0004	0.0080
0.7	250	−0.0003	−0.0020	0.0105	−0.0006	−0.0008	−0.0008	−0.0006	−0.0006	−0.0006	0.0023
	500	−0.0003	−0.0022	0.0054	−0.0004	−0.0004	−0.0004	−0.0003	−0.0003	−0.0004	0.0010
	50	0.0073	0.0152	0.0426	0.0017	−0.0001	0.0032	0.0016	0.0058	0.0096	0.0195
	100	0.0023	0.0045	0.0228	−0.0008	−0.0013	−0.0002	−0.0008	0.0004	−0.0001	0.0094
0.8	250	−0.0002	−0.0021	0.0084	−0.0007	−0.0011	−0.0011	−0.0008	−0.0008	−0.0008	0.0027
	500	−0.0003	−0.0021	0.0039	−0.0004	−0.0004	−0.0005	−0.0003	−0.0004	−0.0005	0.0012
	50	0.0138	0.0229	0.0351	0.0046	0.0039	0.0080	0.0043	0.0102	0.0147	0.0220
	100	0.0046	0.0078	0.0179	−0.0017	−0.0016	−0.0001	−0.0016	0.0003	0.0000	0.0112
0.9	250	0.0003	0.0002	0.0067	−0.0010	−0.0014	−0.0015	−0.0011	−0.0012	−0.0012	0.0035
	500	0.0000	−0.0013	0.0034	−0.0004	−0.0005	−0.0006	−0.0003	−0.0004	−0.0006	0.0016
	50	0.0204	0.0265	0.0314	0.0102	0.0099	0.0139	0.0098	0.0164	0.0208	0.0208
	100	0.0079	0.0119	0.0156	−0.0007	0.0001	0.0019	−0.0007	0.0020	0.0020	0.0120
0.95	250	0.0013	0.0019	0.0060	−0.0013	−0.0016	−0.0017	−0.0014	−0.0013	−0.0013	0.0042
	500	0.0004	0.0000	0.0029	−0.0004	−0.0006	−0.0007	−0.0004	−0.0005	−0.0007	0.0020
	50	0.0235	0.0251	0.0252	0.0119	0.0123	0.0162	0.0114	0.0190	0.0232	0.0144
	100	0.0131	0.0138	0.0128	0.0037	0.0044	0.0064	0.0035	0.0064	0.0064	0.0088
0.99	250	0.0045	0.0053	0.0053	0.0003	0.0005	0.0005	0.0003	0.0008	0.0009	0.0045
	500	0.0018	0.0023	0.0026	−0.0003	−0.0003	−0.0003	−0.0003	−0.0002	−0.0003	0.0026

third to the eighth columns correspond to the plug-in estimates, including the Amihud-Hurvich estimate, while the biases of the direct jackknife estimates are given in the ninth to the eleventh columns. The bias of the bias corrected REML estimate is given in the last column. As a whole, the biases of all the estimates tend to become smaller as T gets larger or as ϕ_1 goes away from 1, except for $\tilde{\beta}_{c,\,rols}$, $\tilde{\beta}_{c,\,rgls}$ and $\hat{\beta}_{reml}$, those of which have smaller biases when ϕ_1 is very close to 1. Among the estimates, the two plug-in estimates using the Cauchy method have larger biases compared with the other estimates except for the case where $\phi_1=0.99$; in this case, the biases of these two estimates are almost the same as the bias of the Amihud-Hurvich estimate. On the other hand, we can see that the jackknife method is very effective to reduce the bias. It seems that the larger number of groups for the jackknife method tends to result in slightly larger biases, although this is not always the case. Comparing the plug-in estimate with the jackknife method with the direct jackknife estimate with the same number of groups, the performance of $\tilde{\beta}_{j,\,m}$ and $\hat{\beta}_{j,\,m}$ is very similar. This implies that as long as the bias is concerned, the information about the AR structure is not so important for the jackknife method to reduce the bias.

Table 2 shows the MSE of the estimates. In this case, $\tilde{\beta}_{ah}$ and $\tilde{\beta}_{reml}$ perform better than the other estimates in many cases. As expected in the previous section, $\tilde{\beta}_{c,\,rgls}$ has small MSE when ϕ_1 is close to one. In fact, its MSE is smaller than that of the Amihud-Hurvich estimate for $\phi_1=0.99$. On the other hand, the jackknife method tends to result in the larger MSE. However, as the number of groups (m) increases, the MSE of the jackknife estimate decreases. For example, when $\phi_1=0.95$ and $T=500$, the MSEs of $\tilde{\beta}_{j,\,2}$, $\tilde{\beta}_{j,\,3}$ and $\tilde{\beta}_{j,\,4}$ are 0.0327, 0.0296, 0.0287, respectively. As in the case of the bias, we can observe that the performance of $\tilde{\beta}_{j,\,m}$ and $\hat{\beta}_{j,\,m}$ is

TABLE 2. MSE($\times$100) OF VARIOUS ESTIMATES

ϕ_1	T	$\tilde{\beta}_{ah}$	$\tilde{\beta}_{c,ols}$	$\tilde{\beta}_{c,gls}$	$\tilde{\beta}_{j,2}$	$\tilde{\beta}_{j,3}$	$\tilde{\beta}_{j,4}$	$\hat{\beta}_{j,2}$	$\hat{\beta}_{j,3}$	$\hat{\beta}_{j,4}$	$\hat{\beta}_{reml}$
	50	1.5206	2.1768	2.4361	1.9858	1.7935	1.6820	2.0114	1.7818	1.6670	1.5127
	100	0.6369	0.9660	1.0225	0.7510	0.7031	0.6849	0.7558	0.7032	0.6840	0.6391
0.7	250	0.2239	0.3386	0.3475	0.2410	0.2335	0.2306	0.2415	0.2338	0.2309	0.2237
	500	0.1070	0.1615	0.1635	0.1107	0.1088	0.1080	0.1110	0.1086	0.1081	0.1070
	50	1.3041	1.7892	1.8918	1.8780	1.6241	1.4860	1.9104	1.6127	1.4734	1.3089
	100	0.5066	0.7491	0.7693	0.6465	0.5869	0.5622	0.6525	0.5864	0.5625	0.5113
0.8	250	0.1666	0.2469	0.2478	0.1876	0.1785	0.1751	0.1881	0.1786	0.1754	0.1667
	500	0.0774	0.1183	0.1168	0.0819	0.0796	0.0789	0.0822	0.0795	0.0790	0.0774
	50	1.0729	1.3637	1.2896	1.7536	1.4099	1.2542	1.7932	1.3963	1.2515	1.0732
	100	0.3610	0.4878	0.4655	0.5428	0.4564	0.4209	0.5504	0.4553	0.4230	0.3675
0.9	250	0.1012	0.1440	0.1401	0.1286	0.1171	0.1126	0.1292	0.1172	0.1127	0.1019
	500	0.0436	0.0661	0.0633	0.0493	0.0467	0.0460	0.0497	0.0468	0.0460	0.0438
	50	0.9475	1.1051	0.9858	1.6743	1.2786	1.1239	1.7169	1.2663	1.1293	0.9044
	100	0.2875	0.3599	0.3024	0.4966	0.3866	0.3445	0.5049	0.3857	0.3472	0.2869
0.95	250	0.0661	0.0892	0.0820	0.0995	0.0850	0.0791	0.1003	0.0853	0.0790	0.0670
	500	0.0256	0.0368	0.0343	0.0327	0.0296	0.0287	0.0330	0.0297	0.0288	0.0259
	50	0.8189	0.8675	0.7099	1.5889	1.1530	1.0043	1.6337	1.1419	1.0087	0.6937
	100	0.2326	0.2444	0.1799	0.4589	0.3332	0.2853	0.4697	0.3322	0.2882	0.1978
0.99	250	0.0404	0.0456	0.0350	0.0795	0.0591	0.0512	0.0809	0.0592	0.0513	0.0375
	500	0.0116	0.0141	0.0112	0.0214	0.0163	0.0147	0.0217	0.0165	0.0149	0.0116

very similar for the same number of m.

Figure 1 shows the probability density functions[4] of the estimates in the representative case where $\rho=0.95$, $\sigma_{ue}=-0.95$ and $T=250$. As in Figure 1(a), the plug-in estimates using the Amihud-Hurvich method and the Cauchy method and the REML estimate are skewed to the right, whereas the plug-in estimates using the jackknife estimates and the direct estimates based on the jackknife method are almost symmetric around the origin, as is observed in Figure 1(b) and (c).

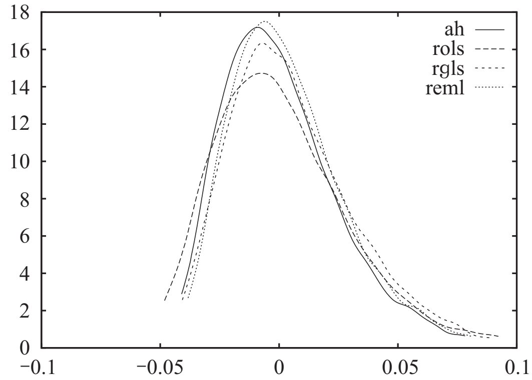
Table 3 reports the coverage rates of the 90% confidence intervals.[5] When the predictor variable is moderately persistent with $\phi_1=0.7$, all the confidence intervals tend to have correct coverage rates as the sample size increases. As ϕ_1 approaches 1, the confidence interval based on the Amihud-Hurvich estimate tends to be liberal whereas the jackknife method leads to a slightly larger coverage rate. Of interest is that $\tilde{\beta}_{c,\,rols}$ and $\tilde{\beta}_{c,\,rgls}$ have correct coverage rates irrespective of the true values of ϕ_1 and T. In fact, the q-q plots of the t-statistics based on these two plug-in estimates against $N(0,\,1)$ are shown to be straight lines, which implies that their finite sample distributions are well approximated by a standard normal distribution, although the distributions of the estimates themselves are skewed to the right as is seen in Figure 1(a) (we omit the figure to save space).

To summarize, as long as the bias is concerned, the jackknife method works quite well

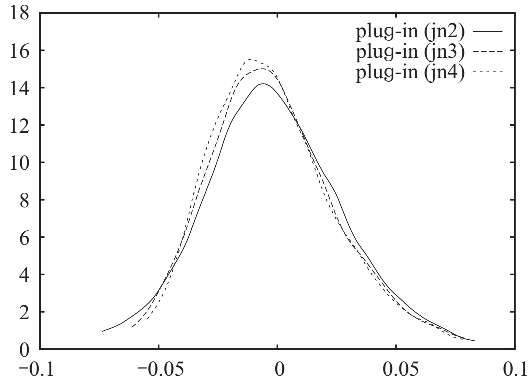
[4] These densities are drawn for the range of 1% to 99% points by the kernel method with a Gaussian kernel. The smoothing parameter, h, is decided by equation (3.31) in Silverman (1986): $h=0.9AT^{-1/5}$ where $A=\min$(standard deviation, interquartile range/1.34).

[5] We do not report the confidence intervals based on the bias corrected REML estimate because Chen and Deo (2009) do not show how to construct the standard error of the estimate.

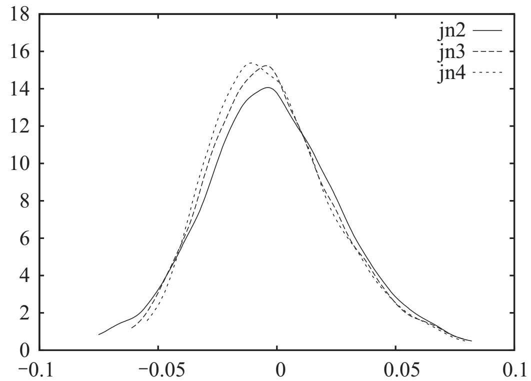
FIG. 1. THE PROBABILITY DENSITY FUNCTIONS OF THE ESTIMATES



(a) pdfs of $\tilde{\beta}_{ah}$, $\tilde{\beta}_{rols}$, $\tilde{\beta}_{rgls}$ and $\tilde{\beta}_{reml}$



(b) pdfs of $\tilde{\beta}_{j,2}$, $\tilde{\beta}_{j,3}$ and $\tilde{\beta}_{j,4}$



(c) pdfs of $\hat{\beta}_{j,2}$, $\hat{\beta}_{j,3}$ and $\hat{\beta}_{j,4}$

TABLE 3. COVERAGE RATE OF CONFIDENCE INTERVAL

ϕ_1	T	$\tilde{\beta}_{ah}$	$\tilde{\beta}_{c,ols}$	$\tilde{\beta}_{c,gls}$	$\tilde{\beta}_{j,2}$	$\tilde{\beta}_{j,3}$	$\tilde{\beta}_{j,4}$	$\hat{\beta}_{j,2}$	$\hat{\beta}_{j,3}$	$\hat{\beta}_{j,4}$
	50	0.8744	0.8911	0.8955	0.9049	0.9094	0.9108	0.9044	0.9056	0.9056
	100	0.8851	0.8898	0.8930	0.9041	0.9050	0.9034	0.9031	0.9034	0.9013
0.7	250	0.8911	0.8968	0.8971	0.9001	0.8987	0.8986	0.8989	0.8998	0.8990
	500	0.8950	0.8965	0.8998	0.8990	0.9007	0.9000	0.8994	0.9013	0.9000
	50	0.8655	0.8906	0.8998	0.9097	0.9167	0.9205	0.9063	0.9115	0.9132
	100	0.8766	0.8860	0.8948	0.9113	0.9142	0.9114	0.9115	0.9109	0.9110
0.8	250	0.8875	0.8920	0.9017	0.9043	0.9057	0.9062	0.9042	0.9050	0.9054
	500	0.8960	0.8908	0.8978	0.9020	0.9029	0.9039	0.9018	0.9042	0.9034
	50	0.8490	0.8921	0.9081	0.9051	0.9225	0.9273	0.9036	0.9126	0.9188
	100	0.8618	0.8860	0.9006	0.9228	0.9275	0.9302	0.9205	0.9229	0.9318
0.9	250	0.8809	0.8898	0.9026	0.9131	0.9168	0.9168	0.9128	0.9172	0.9160
	500	0.8952	0.8888	0.9018	0.9108	0.9121	0.9132	0.9108	0.9128	0.9130
	50	0.8434	0.9016	0.9096	0.8851	0.9123	0.9218	0.8836	0.9023	0.9112
	100	0.8493	0.8924	0.9163	0.9194	0.9333	0.9411	0.9199	0.9287	0.9421
0.95	250	0.8663	0.8895	0.9018	0.9227	0.9344	0.9346	0.9232	0.9322	0.9329
	500	0.8878	0.8904	0.9071	0.9210	0.9245	0.9265	0.9207	0.9238	0.9254
	50	0.8229	0.9064	0.9026	0.8395	0.8939	0.9154	0.8400	0.8831	0.9010
	100	0.8340	0.9099	0.9081	0.8697	0.9117	0.9276	0.8659	0.9036	0.9238
0.99	250	0.8476	0.9056	0.9039	0.9003	0.9303	0.9447	0.8976	0.9264	0.9397
	500	0.8552	0.9007	0.9095	0.9202	0.9391	0.9515	0.9196	0.9373	0.9506

whereas the AH and the REML estimates perform better in view of the MSE. However, the MSE of the estimates based on the jackknife method can be reduced if we use larger number of groups. Finally, although the plug-in estimates based on the Cauchy estimates do not necessarily perform well in view of the bias and the MSE, the confidence intervals based on them have correct coverage rates and hence, the latter estimates could be useful for the hypothesis testing.

IV. *Predictability of U.S. Stock Returns*

In this section, we implement the estimation methods considered in the previous section on U.S. equity data for model (1)–(2), taking the previous findings of Campbell and Yogo (2006) as the benchmark of the comparison.

1. Description of Data

We use five different series of the U.S. stock returns. The first three series are the returns on the annual S&P 500 index, the monthly and annual Center for Research in Security Prices (CRSP) data. These return data, along with the financial ratio variables — the dividend-price ratio and the earnings-price ratio — are used as the predictor variables. The series and the sample period are the same as in Campbell and Yogo (2006)[6]. The estimation results using

[6] We use the 1880-2002 and 1880-1994 samples for the annual S&P 500 index; the 1926M12-2002M12, 1926M12-1994M12, and 1952M1-2002M12 samples for the monthly CRSP data; and the 1926-2002, 1926-1994, and 1952-2002

these data series are presented in Panels A, B, and C of Table 4, respectively. In addition, we also use the updated annual return on the S&P 500 index and the monthly data[7]. The estimation results using these data series are presented in Panels D and E of Table 4, respectively.

To compute the excess returns of stocks over risk-free return, we use the one-month T-bill rate for monthly series and roll over the one-month T-bill rate for the annual series. In our analysis, we use two additional predictor variables, the three month T-bill rate and the long-short yield spread, following Campbell and Yogo (2006). As in Fama and French (1989) and other previous researches, the long yield used to compute the yield spread is Moody's seasoned Aaa corporate bond yield. The short rate is the one-month T-bill rate[8]. Following the usual convention, the excess returns and the predictor variables are in logs.

2. Persistence of the Predictor Variables

In the fifth column of Table 4, we report the estimated coefficient for the autoregressive root $\rho = \phi_1 + \cdots + \phi_p$ with the standard error in parentheses for the log dividend-price ratio (d-p), the log earnings-price ratio (e-p), the three-month T-bill rate (3my), and the long-short yield spread (ys) using the (direct) jackknife method with the number of groups equal to 3 for annual data and 4 for monthly data. From Subsection 2.2, the estimated coefficient is denoted by $\hat{\rho}_{j,m} = \hat{\phi}_{1,j,m} + \cdots + \hat{\phi}_{p,j,m}$. The autoregressive lag length $p \in [1, \bar{p}]$ for the predictor variable is determined by the Bayesian information criterion (BIC). We set the maximum lag length $\bar{p}$ as four for annual data and eight for monthly data. The estimated lag lengths are reported in the fourth column of Table 4.

As discussed in the literature, the high persistence of these typical predictor variables suggests that the first-order asymptotics, that is, the *t*-statistic based on the OLS estimate being approximately normal in large samples, can possibly lead to misleading results. It should also be noted that whether or not the conventional inference based on the *t*-test is reliable depends on the correlation (δ) between the innovations to the excess returns and to the predictor variable, in addition to the true value of ρ. We report the point estimates of δ in the sixth column of Table 4. The correlations of the innovations to stock returns with the financial ratios are negative and large, but those with the interest rate variables (3my and ys) are much smaller. The former result indicates that the movements in stock returns and in financial ratios mostly come from the movements in the stock price. The latter finding suggests that for the interest rate variables, the conventional t-test perhaps provides approximately valid inference. These findings are essentially the same as those obtained by Campbell and Yogo (2006).

3. Estimating and Testing the Predictability of the U.S. Stock Returns

Following the methodology discussed in Section 2, we calculate the point estimate of β by the direct application of the jackknife method ($\hat{\beta}_{j,m}$) to see the magnitude of the predictability

---

samples for the annual CRSP data. These data are taken from MotohiroYogo's website: http://www.nber.org/˜myogo/

[7] These data and the financial ratio variables are taken from Robert Shiller's website : http://www.econ.yale.edu/˜shiller/data.htm

[8] The T-bill rates and the Moody's seasoned Aaa corporate bond yield are from Yogo's and FRB's websites: http://www.federalreserve.gov/releases/h15/data.htm

TABLE 4. EMPIRICAL RESULTS FOR STOCK RETURNS

r_t	x_t	sample	p	$\hat{\rho}_{j,m}$ (std)	$\hat{\delta}$	$\hat{\beta}_{j,m}$ (p-value)	CI by $\hat{\beta}_{j,m}$	CI by $\tilde{\beta}_{c,rgls}$	CI by BF-Q
Panel A CY(SP&500,Y)	d-p	1880-2002	3	1.004 (0.088)	-0.851	0.072 (0.078)	(-0.012, 0.155)	(0.035, 0.248)	(-0.020, 0.142)
		1880-1994	3	0.921 (0.125)	-0.842	0.119 (0.036)	(0.010, 0.227)	(0.047, 0.301)	(-0.017, 0.236)
	e-p	1880-2002	1	0.886 (0.053)	-0.962	0.116 (0.010)	(0.034, 0.197)	(0.055, 0.266)	(0.043, 0.225)
		1880-1994	1	0.813 (0.061)	-0.958	0.195 (0.000)	(0.099, 0.291)	(0.079, 0.332)	(0.092, 0.323)
Panel B CY(CRSP, M)	d-p	1926.12-2002.12	2	0.992 (0.006)	-0.951	0.007 (0.096)	(-0.002, 0.016)	(-0.006, 0.012)	(-0.005, 0.011)
		1926.12-1994.12	2	0.978 (0.007)	-0.949	0.011 (0.069)	(-0.001, 0.023)	(-0.011, 0.012)	(-0.007, 0.017)
		1952.01-2002.12	1	0.998 (0.008)	-0.967	0.007 (0.175)	(-0.005, 0.019)	(-0.002, 0.011)	(-0.004, 0.010)
	e-p	1926.12-2002.12	1	0.991 (0.005)	-0.987	0.011 (0.034)	(0.001, 0.020)	(0.000, 0.019)	(0.001, 0.018)
		1926.12-1994.12	2	0.980 (0.007)	-0.984	0.020 (0.002)	(0.009, 0.032)	(0.000, 0.026)	(0.006, 0.029)
		1952.01-2002.12	1	0.995 (0.006)	-0.982	0.005 (0.237)	(-0.006, 0.015)	(-0.006, 0.011)	(-0.004, 0.012)
	3my	1952.01-2002.12	2	0.993 (0.013)	-0.071	-0.005 (0.636)	(-0.028, 0.018)	(-0.030, -0.006)	(-0.029, -0.006)
	ys	1952.01-2002.12	1	0.933 (0.017)	-0.066	0.035 (0.037)	(0.003, 0.067)	(0.018, 0.070)	(0.020, 0.072)
Panel C CY(CRSP,Y)	d-p	1926-2002	1	1.019 (0.100)	-0.721	0.051 (0.281)	(-0.094, 0.197)	(0.035, 0.211)	(0.013, 0.185)
		1926-1994	1	0.905 (0.101)	-0.693	0.176 (0.013)	(0.046, 0.306)	(0.047, 0.291)	(0.059, 0.333)
		1952-2002	1	1.001 (0.081)	-0.749	0.062 (0.223)	(-0.072, 0.197)	(-0.037, 0.139)	(-0.006, 0.179)
	e-p	1926-2002	1	0.919 (0.088)	-0.957	0.112 (0.089)	(-0.025, 0.249)	(0.080, 0.332)	(0.041, 0.272)
		1926-1994	1	0.798 (0.088)	-0.959	0.229 (0.008)	(0.074, 0.385)	(0.138, 0.468)	(0.131, 0.455)
		1952-2002	1	0.989 (0.119)	-0.955	-0.008 (0.522)	(-0.234, 0.219)	(-0.029, 0.226)	(-0.032, 0.225)
	3my	1952-2002	1	0.863 (0.088)	0.006	-0.081 (0.837)	(-0.217, 0.055)	(-0.228, 0.040)	(-0.228, 0.039)
	ys	1952-2002	1	0.777 (0.219)	-0.243	-0.051 (0.592)	(-0.416, 0.313)	(-0.043, 0.356)	(-0.073, 0.354)
Panel D US(M)	d-p	1926.01-2009.12	2	0.998 (0.005)	-0.560	0.039 (0.000)	(0.028, 0.050)	(0.026, 0.037)	(0.026, 0.038)
	e-p	1926.01-2009.12	3	0.994 (0.004)	-0.556	0.035 (0.000)	(0.025, 0.046)	(0.022, 0.033)	(0.021, 0.034)
	3my	1926.12-2009.12	2	0.995 (0.006)	0.011	-0.005 (0.859)	(-0.012, 0.003)	(-0.010, 0.003)	(-0.010, 0.003)
	ys	1926.01-2009.12	8	0.987 (0.003)	0.018	0.137 (0.000)	(0.084, 0.191)	(0.100, 0.110)	(0.100, 0.110)
Panel E US(Y)	d-p	1881-2009	3	0.967 (0.069)	-0.835	0.023 (0.303)	(-0.050, 0.095)	(0.016, 0.199)	(-0.055, 0.115)
	e-p	1881-2009	1	0.903 (0.044)	-0.932	0.057 (0.103)	(-0.017, 0.132)	(-0.023, 0.143)	(-0.007, 0.156)
	3my	1926-2009	3	0.988 (0.092)	0.171	-0.038 (0.789)	(-0.117, 0.040)	(-0.103, 0.052)	(-0.099, 0.045)
	ys	1919-2009	2	0.738 (0.141)	-0.268	-0.044 (0.699)	(-0.184, 0.095)	(-0.169, 0.070)	(-0.164, 0.089)

on the U.S. stock returns, because it is less biased compared to the other estimates. On the other hand, we check the statistical significance of the predictability by obtaining the 90% confidence intervals (CI) based on not only the direct jackknife estimate ($\hat{\beta}_{j,m}$) but also the plug-in estimate with the Cauchy estimate using the GLS detrending method ($\tilde{\beta}_{c,rgls}$), because the latter has a correct coverage rate irrespective of the strength of the persistence. As shown by the Mote Carlo simulations in the previous section, the confidence interval based on $\hat{\beta}_{j,m}$ tends to be conservative (the coverage rate tends to be larger than expected) while the latter is more accurate, so that we expect that, in some cases, the confidence interval based on $\hat{\beta}_{j,m}$ includes $\beta=0$ whereas the lower bound of the confidence interval based on $\tilde{\beta}_{c,rgls}$ is greater than 0. For the purpose of comparison, we also calculate the confidence interval based on the Bonferroni procedure (BF-Q test) proposed by Campbell and Yogo (2006). The point estimate and the confidence intervals are multiplied by $\hat{\sigma}_e/\hat{\sigma}_u$ for the normalization.

For annual S&P 500 index, we see from Panel A that the 90% confidence intervals based on the jackknife method in the eighth column and the BF-Q test in the last column include $\beta=0$ for the log dividend-price ratio (d-p) in the whole sample (1880-2002) whereas the confidence interval based on $\tilde{\beta}_{c,rgls}$ implies the predictability. In this case, the point estimate of β is 0.072 so that one point increase in the dividend-price ratio can predict the 7.2% increase in the future stock return. When the sample period is shortened up to 1994, the confidence interval based $\hat{\beta}_{j,m}$, in addition to the confidence interval based on $\tilde{\beta}_{c,rgls}$, also shows the predictability and the point estimate becomes 0.119, which is greater than 0.072 in the whole sample period. This implies that the predictability is more evident before 1994 and the predictive power decreases recently, as is often pointed out in the existing literature. For the log earnings-price ratio (e-p), all the three confidence intervals imply the consistent results; the annual S&P 500 is predictable by the earnings-price ratio and again, it is more strongly predictable before 1994.

For the monthly CRSP series with the Campbell and Yogo (2006) sample, reported in Panel B, all the three methods result in similar confidence intervals and only the log earnings-price ratio and the log-short yield spread (ys) has return predictability at the 5% significance level. These findings are almost the same as those observed by Campbell and Yogo (2006). As in the case of annual data, the estimate of β for the log earnings-price ratio before 1994 is greater and almost double compared with the estimate in the whole sample period; it is 0.020 and significant at the 1% significance level.

In panel C, we report the result for the annual CRSP series with the Campbell and Yogo (2006) sample. We see that the log dividend-price ratio and the log earnings-price ratio have return predictability from the confidence intervals based on $\tilde{\beta}_{c,rgls}$ and the BF-Q test for the 1926-2002 and the 1926-1994 samples. In particular, we observe the strong predictability from the fact that the p-values based on $\hat{\beta}_{j,m}$ are almost 1% and the estimated coefficients in the subsample are almost triple and double compared with those in the whole sample for the log dividend-price ratio and the log earnings-price ratio, respectively. However, both the predictor variables do not have predictive power for the 1952-2002 sample.

For the longer and more recent monthly sample, reported in Panel D, we reject the null hypothesis of no predictability at the 1% level based on the jackknife point estimates for the log dividend-price ratio (d-p), the log earnings-price ratio (e-p), and the long-short yield spread (ys). According to the three 90% confidence intervals, we also reject the null hypothesis for these three series. Note that there is no difference between the confidence intervals based on $\tilde{\beta}_{c,\,rgls}$ and the BF-Q test. Judging from these findings, we observe that the log dividend-price ratio (d-p) and the log earnings-price ratio (e-p) are more likely to have predictability on this monthly sample. However, as to the long-short yield spread (ys), there is room for reconsidering the interpretation, as stated above.

For the updated longer annual sample reported in Panel E, we reject the null at the 5% level by the confidence interval based on $\tilde{\beta}_{c,\,rgls}$ for the log dividend-price ratio whereas the BF-Q test fails to reject the null. We cannot find the significant predictive power for the other three predictor variables.

V. *Conclusion*

In this paper, we have investigated several point estimates for predictive regressions. Most of them are easily obtained by the least squares and the instrumental variable methods. Those estimates have the nice finite sample properties such as small biases and the correct coverage rates of the confidence intervals. In particular, the direct jackknife estimate has an advantage over the existing plug-in estimates and the REML estimate in that the former does not require the specification of the model for the predictor variables. Based on our unified framework, we have investigated the U.S. stock returns and found that some variables, which have not been statistically detected as useful predictors in the literature, are able to predict stock returns. Because of their nice properties, the methods considered in this paper complement the existing statistical tests for predictability to investigate the relations between stock returns and economic variables.

Appendix: Derivation of the Estimate of the Variance-covariance Matrix

For the case of the jackknife estimate, we first note that

$$\begin{aligned}\sum_{t=1}^{T}\tilde{x}_{t-1}\tilde{e}_{j,\,m,\,t} &= \sum_{t=1}^{T}\tilde{x}_{t-1}(\tilde{x}_t-\hat{\varphi}'_{j,\,m}\tilde{z}_{t-1})\\ &= \sum_{t=1}^{T}\tilde{x}_{t-1}\{e_t-(\hat{\varphi}_{j,\,m}-\varphi)'\tilde{z}_{t-1}\}\\ &= \sum_{t=1}^{T}\tilde{x}_{t-1}e_t-\left(\sum_{t=1}^{T}\tilde{x}_{t-1}\tilde{z}'_{t-1}\right)(\hat{\varphi}_{j,\,m}-\varphi).\end{aligned}$$

Then, from the definition of $\tilde{\beta}$ given in (7), we can see that

$$\begin{aligned}\tilde{\beta}_{j,\,m} &= \beta+\frac{\sum_{t=1}^{T}\tilde{x}_{t-1}\left(u_t-\dfrac{\hat{\sigma}_{ue}}{\hat{\sigma}_e^2}\tilde{e}_{j,\,m,\,t}\right)}{\sum_{t=1}^{T}\tilde{x}_{t-1}^2}\\ &= \left\{\beta+\frac{\sum_{t=1}^{T}\tilde{x}_{t-1}\left(u_t-\dfrac{\hat{\sigma}_{ue}}{\hat{\sigma}_e^2}e_t\right)}{\sum_{t=1}^{T}\tilde{x}_{t-1}^2}\right\}+\frac{\hat{\sigma}_{ue}}{\hat{\sigma}_e^2}\frac{\sum_{t=1}^{T}\tilde{x}_{t-1}\tilde{z}'_{t-1}}{\sum_{t=1}^{T}\tilde{x}_{t-1}^2}(\hat{\varphi}_{j,\,m}-\varphi)\end{aligned}$$

$$=\tilde{\beta}^*+\frac{\hat{\sigma}_{ue}}{\hat{\sigma}_e^2}K'(\hat{\varphi}_{j,m}-\varphi),$$

where $\tilde{\beta}^*$ is defined as $\hat{\beta}$ by replacing σ_{ue}/σ_e^2 with $\hat{\sigma}_{ue}/\hat{\sigma}_e^2$ and then $\tilde{\beta}^*$ is the asymptotically efficient estimate. Since $\{u_t-(\sigma_{ue}/\sigma_e^2)e_t\}$ is an uncorrelated sequence with $\{e_t\}$, we can see that $\tilde{\beta}^*$ is asymptotically independent of $(\hat{\varphi}_{j,m}-\varphi)$ while K converges in probability. Hence, the variance estimate of $\tilde{\beta}_{j,m}$ is given by

$$\widehat{Var(\tilde{\beta}_{j,m})}=\widehat{Var(\hat{\beta}^*)}+\left(\frac{\hat{\sigma}_{ue}}{\hat{\sigma}_e^2}\right)^2 J^{*\prime}\hat{\Sigma}_{j,m}J^*,$$

where $\widehat{Var(\hat{\beta}^*)}$ is the variance estimate of the infeasible efficient estimate, which can be easily shown to be

$$Var(\hat{\beta}^*)=\frac{\hat{\sigma}_e^2\hat{\sigma}_u^2-\hat{\sigma}_{ue}^2}{\hat{\sigma}_e^2\sum_{t-1}^T\tilde{x}_{t-1}^2}$$

and $\hat{\Sigma}_{j,m}$ is the estimate of the variance-covariance matrix of $\hat{\varphi}_{j,m}$.

Similarly, for the case of the Cauchy estimate, we can see that

$$\begin{aligned}\sum_{t=1}^T\tilde{x}_{t-1}\tilde{e}_{c,rols,t}&=\sum_{t=1}^T\tilde{x}_{t-1}(\tilde{x}_t-\hat{\rho}_{c,rols}\tilde{x}_{t-1}-\hat{\varphi}_{1,c,rols}\Delta x_{t-1}-\cdots-\hat{\varphi}_{p-1,c,rols}\Delta x_{t-p+1})\\&=\sum_{t=1}^T\tilde{x}_{t-1}e_t-\sum_{t=1}^T\tilde{x}_{t-1}\{(\hat{\rho}_{c,rols}-\rho)\tilde{x}_{t-1}+(\hat{\varphi}_{1,c,rols}-\psi_1)\Delta x_{t-1}\\&\qquad-\cdots-(\hat{\varphi}_{p-1,c,rols}-\psi_{p-1})\Delta x_{t-p+1})\}.\end{aligned}$$

Then, we obtain expression (12) in exactly the same man ner as $\widehat{Var(\tilde{\beta}_{j,m})}$.

References

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

Amihud, Y., C. M. Hurvich and Y. Wang (2008), "Multiple-Predictor Regressions: Hypothesis Testing," *Review of Financial Studies* 22, pp.413-434.

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

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

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

Chen, W. W. and R. S. Deo (2009), "Bias Reduction and Likelihood-Based Almost Exactly Sized Hypothesis Testing in Predictive Regressions Using the Restricted Likelihood," *Econometric Theory* 25, pp.1143-1179.

Chiquoine, B. and E. Hjalmarsson (2009), "Jackknifing Stock Return Predictions," *Journal of Empirical Finance* 16, pp.793-803.

Choi, I. (1993), "Asymptotic Normality of the Lest-Squares Estimates for Higher Order Autoregressive Integrated Processes with Some Applications," *Econometric Theory* 9, pp.263-282.

Eliasz, P. (2004), "Optimal Median Unbiased Estimation of Coefficients on Highly Persistent Regressions," Manuscript, Princeton University.

Elliott, G., T. J. Rothenberg and J. Stock (1996), "Efficient Tests for an Autoregressive Unit Root," *Econometrica* 64, pp.813-836.

Elliott, G. and J. H. Stock (1994), "Inference in Time Series Regression When the Order of Integration of a Regressor is Unknown," *Econometric Theory* 10, pp.672-700.

Fama, E. F. and K. R. French (1989), "Business Conditions and Expected Returns on Stocks and Bonds," *Journal of Financial Economics* 25, pp.23-49.

Hausman, J. A. (1978), "Specification Tests in Econometrics," *Econometrica* 46, pp.1251-1271.

Jansson, M. and M. J. Moreira (2006), "Optimal Inference in Regression Models with Nearly Integrated Regressors," *Econometrica* 74, pp.681-714.

Kurozumi, E. and T. Yamamoto (2000), "Modified Lag Augmented Vector Autoregressions," *Econometric Reviews* 19, pp.207-231.

Lanne, M. (2002), "Testing the Predictability of Stock Returns," *Review of Economics and Statistics* 84, pp.407-415.

Lewellen, J. (2004), "Predicting Returns with Financial Ratios," *Journal of Financial Economics* 74, pp.209-235.

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

Park, H. J. and W. A. Fuller (1995), "Alternative Estimators and Unit Root Tests for the Autoregressive Process," *Journal of Time Series Analysis* 16, pp.415-429.

Phillips, P. C. B. and J. Yu (2005), "Jackknifing Bond Option Prices," *Review of Financial Studies* 18, pp.707-742.

Quenouille, M. H. (1949), "Approximate Tests of Correlation in Time Series," *Journal of the Royal Statistical Society (B)* 11, pp.68-84.

Quenouille, M. H. (1956), "Notes on Bias in Estimation," *Biometrika* 43, pp.353-360.

Silverman, B. W. (1986), "*Density Estimation for Statistics and Data Analysis*," Chapman and Hall, London.

So, B. S. and D. W. Shin (1999), "Cauchy Estimators for Autoregressive Processes with Applications to Unit Root Tests and Confidence Intervals," *Econometric Theory* 15, pp.165-176.

Stambaugh, R. F. (1999), "Predictive Regressions," *Journal of Financial Economics* 54, pp.375-421.

Toda, H. Y. and T. Yamamoto (1995), "Statistical Inference in Vector Autoregressions with Possibly Integrated Processes," *Journal of Econometrics* 66, pp.225-250.

Torous, W., R. Valkanov and S. Yan (2004), "On Predicting Stock Returns with Nearly Integrated Explanatory Variables," *Journal of Business* 77, pp.937-966.

Yamamoto, T. and E. Kurozumi (2005), "Lag Augmentation in Regression Models with Possibly Integrated Regressors," *Hitotsubashi Journal of Economics* 46, pp.159-175.