NBER WORKING PAPER SERIES

BIASES IN LONG-HORIZON PREDICTIVE REGRESSIONS

Jacob Boudoukh
Ronen Israel
Matthew P. Richardson

Working Paper 27410
http://www.nber.org/papers/w27410

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2020

AQR Capital Management is a global investment management firm, which may or may not apply similar investment techniques or methods of analysis as described herein. The views expressed here are those of the authors and not necessarily those of AQR or the National Bureau of Economic Research. We would like to thank seminar participants at the AQR Research Colloquium and NYU Stern School.

At least one co-author has disclosed a financial relationship of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w27410.ack

Biases in Long-Horizon Predictive Regressions
Jacob Boudoukh, Ronen Israel, and Matthew P. Richardson
NBER Working Paper No. 27410
June 2020
JEL No. C01,C1,C22,C53,C58,G12,G17

ABSTRACT

Analogous to Stambaugh (1999), this paper derives the small sample bias of estimators in J-horizon predictive regressions, providing a plug-in adjustment for these estimators. A number of surprising results emerge, including (i) a higher bias for overlapping than nonoverlapping regressions despite the greater number of observations, and (ii) particularly higher bias for an alternative long-horizon predictive regression commonly advocated for in the literature. For large J, the bias is linear in (J/T) with a slope that depends on the predictive variable's persistence. The bias adjustment substantially reduces the existing magnitude of long-horizon estimates of predictability.

Jacob Boudoukh
Arison School of Business, IDC
Kanfei Nesharim St
Herzlia 46150
ISRAEL
jboudouk@idc.ac.il

Ronen Israel
AQR Capital
Two Greenwich Plaza--3rd floor
Greenwich, CT 06830
ronen.israel@aqr.com

Matthew P. Richardson
Stern School of Business
New York University
44 West 4th Street, Suite 9-190
New York, NY 10012
and NBER
mrichar0@stern.nyu.edu

I. Introduction

Much of modern empirical asset pricing has been devoted to documenting and testing whether expected asset returns vary through time. A typical predictive regression involves the researcher regressing asset returns, $R_{t:t+1}$, on some lagged predictive variable, X_t, using T periods of data. X_t is often a price-based measure of some underlying asset (such as a valuation ratio or yield), which itself is persistent and mean-reverting. For example:

$$
\begin{aligned}
R_{t,t+1} &= \alpha_1 + \beta_1 X_t + u_{t+1} \\
X_{t+1} &= \varpi + \rho X_t + v_{t+1}
\end{aligned}
\tag{1}
$$

In highly impactful research, Stambaugh (1999) shows that if $\sigma_{uv} \neq 0$ (which is common in the stock return prediction literature) then the OLS estimator $\hat{\beta}_1$ will be biased. Indeed, for the model in equation (1), he derives the result, $E\left[\hat{\beta}_1 - \beta_1\right] = \frac{\sigma_{uv}}{\sigma_v^2} E\left[\hat{\rho} - \rho\right] \approx -\frac{\sigma_{uv}}{\sigma_v^2}\frac{1+3\rho}{T}$. It is now standard to adjust the β_1 coefficient estimator for this bias (e.g., see also Amihud and Hurvich (2004)).

Once the above estimators have been biased-adjusted, predictability regressions are quite disappointing, exhibiting low R^2s and insignificant *t*-statistics. In an attempt to generate greater test power and motivated by theories of low frequency mean-reversion in expected returns (either behavioral or risk-based), researchers have looked to predict long-horizon returns. A typical long-horizon regression involves the researcher regressing J horizon returns of an asset, $R_{t:t+J}$, on some lagged predictive variable, X_t, using T periods of data:

$$
R_{t:t+J} = \alpha_J + \beta_J X_t + \varepsilon_{t:t+J} \tag{2}
$$

If the researcher estimates equation (2) by sampling every J^{th} period, using nonoverlapping sample length T/J (denote as *nol*), standard ordinary least squares (OLS) applies. For large J, however, the sample size is often small, leading researchers to estimate regression (2) using all available overlapping data (denote as *ol)*. Using more data increases the asymptotic efficiency of the estimators but also leads to the misspecification of OLS standard errors due to autocorrelated errors. As a consequence, researchers adjust the standard errors using one of the various heteroscedasticity and autocorrelation (HAC) adjusted estimators that have been developed, with Newey and West (1987) being the preferred choice in the finance literature.

Fama and French (1988, 1989) were the first to examine equation (2) using dividend yields as their original predictive variable. In his AFA presidential address, Cochrane (2011) chooses dividend-price ratios to highlight the large amounts of time-variation of discount rates. Indeed, he argues all price-dividend variation corresponds to expected return variation. As an illustration of his findings, using his sample period (1947-2009), Figure 1 below graphs $\hat{\beta}_J^{ol}$ (red) for forecasting horizons of one month (J=1) to five years (J=60). While recognizing issues with estimated standard errors, Cochrane (2011) points out that the regression coefficients are nevertheless *economically* large and increasing with the horizon. This view of the evidence is the prevailing one in the literature.

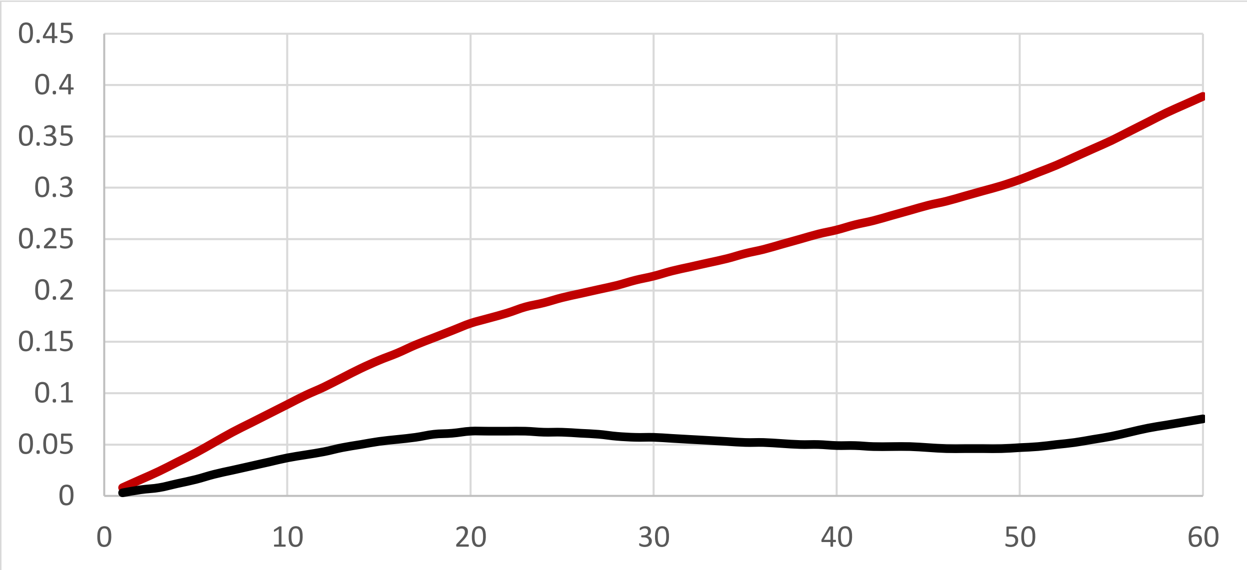
While it is known that long-horizon predictive regression coefficient estimators are biased in small samples, existing studies make this point using simulation evidence.[1] There is no analytical result analogous to Stambaugh (1999). Possibly because of the lack of a theoretical result, researchers rarely adjust long horizon regressions for the bias, an example being Cochrane (2011).

In this paper, we derive the analytical small-sample bias for long-horizon regressions. Similar to Stambaugh's formula, the bias of the coefficient estimator of the long-horizon overlapping regression is a function of the correlation between the innovation in returns and the predictive variable (σ_{uv}), the autocorrelation of the predictive variable ρ, and the sample size T, but now also the horizon J. Specifically, $E\left[\hat{\beta}_J^{ol} - \beta_J\right] = \frac{1}{T}\left[J(1+\rho) + 2\rho\left(\frac{1-\rho^J}{1-\rho}\right)\right]\frac{\sigma_{uv}}{\sigma_v^2}$. Using this formula to adjust $\hat{\beta}_J^{ol}$ for its small sample bias, the black line in Figure 1 below shows that these bias-adjusted coefficients are in fact *not economically significant,* but rather small in magnitude. As Cochrane (2011) puts it, "*discount rate variation is the central question of current asset pricing research*". Figure 1 suggests a rethink of arguably our most important stylized fact describing this variation.

[1] See, for example, Goetzmann and Jorion (1993), Nelson and Kim (1993) and Torous, Valkanov and Yan (2000).

Figure 1: Overlapping Regression Beta and Bias-Adjusted Beta for DP

Figure 1 graphs $\hat{\beta}_J^{ol}$ (red) and the bias-adjusted beta, $\hat{\beta}_J^{ol} - E[\hat{\beta}_J^{ol}]$ (black), for forecasting horizons of one month (J=1) to five years (J=60), where the bias is calculated under an AR1 Assumption for DP_t using the estimated $E[\hat{\beta}_J^{ol} - \beta_J] = \frac{1}{T}\left[J(1+\rho) + 2\rho\left(\frac{1-\rho^J}{1-\rho}\right)\right]\frac{\sigma_{uv}}{\sigma_v^2}$. Data is obtained from Amit Goyal's website (see Goyal Welch (2008)). Sample period is 1947-2009 (corresponding to the period used by Cochrane (2011)), T=732m, $\hat{\sigma}_{uv} = -0.00178$, $\hat{\sigma}_u = 0.042, \hat{\sigma}_v = 0.043$ and $\hat{\rho}^{adj}$=0.9996.



The above formula implies the bias is monotonically increasing, albeit nonlinearly, in J and ρ. It is fairly standard to show results for multiple long horizon regressions with increasing J. Note that, for large J, the bias is linearly increasing in J/T with a slope $(1+\rho)\frac{\sigma_{uv}}{\sigma_v^2}$ and, importantly, the bias never asymptotes.

Because of problems associated with estimation of HAC standard errors in small samples, such as Newey and West (1987) [2], researchers have used alternative long-horizon regression equations to those in equation (2), including nonoverlapping regressions; single period return regressions on lagged sums of the predictive variable $\sum_{j=1}^{J} X_{t-j}$ (e.g., see Jegadeesh (1991) and Hodrick (1992));

[2] See, for example, Richardson and Stock (1989), Andrews (1991), Nelson and Kim (1993), Goetzmann and Jorion (1993), Newey and West (1994), Bekaert, Hodrick and Marshall (1997), Valkanov (2003), Hjalmarsson (2011), Britten-Jones, Neuberger, and Nolte (2011), Chen and Tsang (2013) and Boudoukh, Israel and Richardson (2019) for the use of Newey and West (1987) standard errors.

and implied long-horizon coefficients using the structure of equation (1) (e.g., see Kandel and Stambaugh (1989), Campbell (1991) and Hodrick (1992)). In this paper, we derive the small sample biases of these variants of long-horizon regressions. Some interesting results emerge. Most surprising, for all ρ and J, the small sample bias is more severe for overlapping versus nonoverlapping regressions. In addition, we show that the popular Jegadeesh (1991)/Hodrick (1992) alternative long-horizon regression is severely biased. We explain why and relate it to common empirical methodologies employed in the finance and macroeconomics literature. The theoretical results suggest the need to reexamine common approaches to return forecasting.

What effect do these small sample biases have on existing evidence of long-horizon predictability? Focusing on a representative list of popular valuation ratio predictors, we study the impact of small sample biases on predictability evidence. Aside from a couple of surprising departures, the evidence in favor of predictability mostly disappears.

II. Small Sample Bias in Long-Horizon Predictive Regressions

The coefficient estimates from long-horizon regressions of equation (2) are rarely adjusted for small sample bias. Past researchers have provided Monte Carlo or bootstrapped p-values to capture the idea that the distribution of the estimators does not conform to the consistent asymptotic normal distributions implied by the theory (e.g., Goetzmann and Jorion (1993) and Kim and Nelson (1993)). This literature finds that the point estimates deserve less confidence (i.e., are less "statistically significant"). To employ as much of the data as possible, researchers often use overlapping observations. The use of overlapping observations requires estimates of HAC standard errors, and researchers frequently employ the procedure of Newey and West (1987). Yet, there is a body of simulation evidence that documents biases in small samples associated with HAC standard error calculations and their effect on "t-statistics" (e.g., see footnote 2). But, importantly, typically in this area of research, the point estimates remain untouched.

As an alternative to regression equation (2), researchers often choose to reverse the regression utilizing the fact that the covariance between J-period returns and a lagged predictor is identical to the covariance between a one-period return and a lagged J-sum of the same predictor (e.g., Jegadeesh (1991) and Hodrick (1992)):

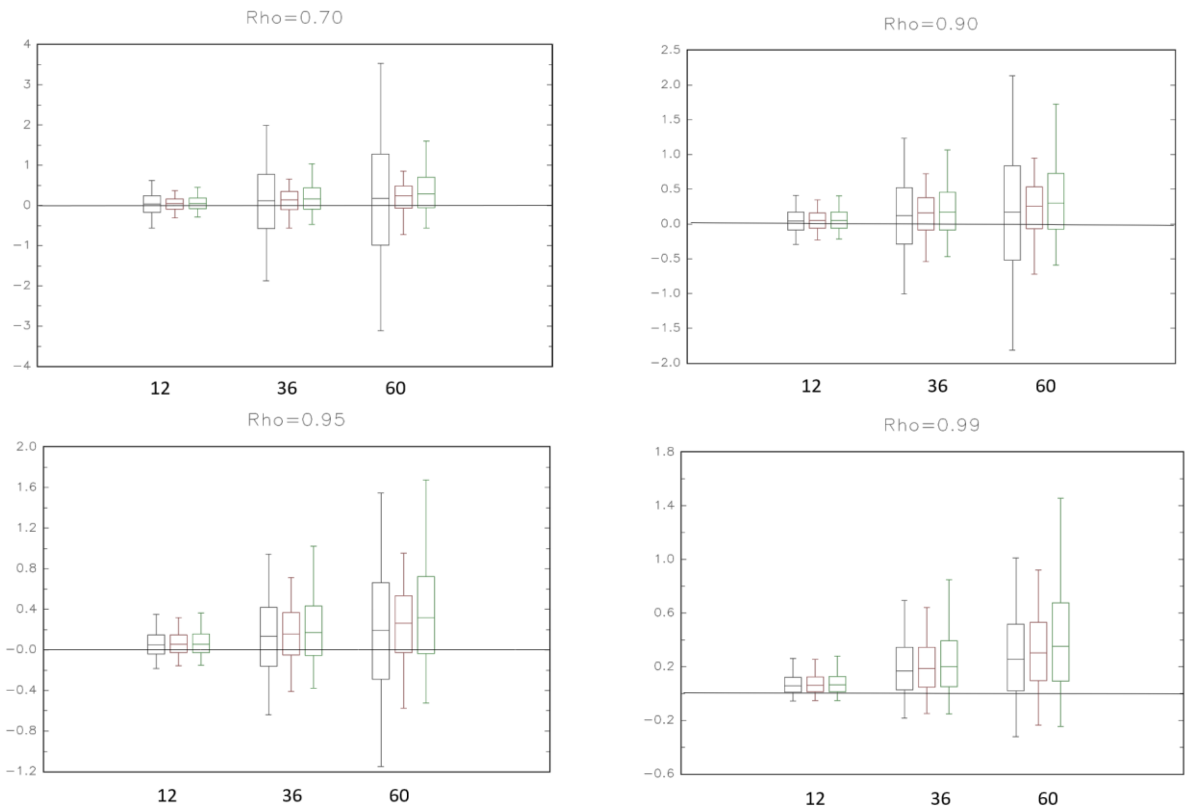
$$R_{t,t+1} = \mu_J + \gamma_J X_{t-j,t} + \xi_{t,t+1} \tag{3}$$

Since there are no overlapping errors, one advantage of equation (3) is that standard OLS applies. Boudoukh and Richardson (1994) show that the asymptotic efficiency of $\hat{\beta}_J^{ol}$ and $\hat{\gamma}_J$ are identical though they argue $\hat{\gamma}_J$ is likely to have worse small sample properties due to its required estimation of the long-horizon variance estimator, $\mathrm{var}\left(X_{t-J,J}\right)$. They show that, for small persistence in X_t, i.e., low ρ, both $\hat{\beta}_J^{ol}$ and $\hat{\gamma}_J$ are much more asymptotically efficient than $\hat{\beta}_J^{nol}$. However, for ρ close to 1, all the estimators are similar on the efficiency front. If this is the case in small samples, then the bias magnitude of these estimators should play an even larger role with respect to the researcher's choice of estimator.

Figure 2 below provides simulated boxplots of the distribution of the long-horizon estimators - $\hat{\beta}_J^{ol}$ (red), $\hat{\beta}_J^{nol}$ (black) and $\hat{\gamma}_J$ (green). The boxplots are shown at the 5%, 25%, 50%, 75% and 95% levels for ρ=0.7, 0.90, 0.95 and 0.99 for J=12, 36 and J=60 and T=600. While the distribution of $\hat{\beta}_J^{nol}$ is wider than $\hat{\beta}_J^{ol}$ for less persistent predictors (e.g., ρ=0.7 and 0.9), there is little difference between $\hat{\beta}_J^{nol}$ and $\hat{\beta}_J^{ol}$ for highly persistent variables (e.g., ρ=0.99) such as commonly found for valuation ratios. Interestingly, $\hat{\gamma}_J$ is a particularly poor estimator in terms of efficiency for large ρ. Figure 2 also demonstrates the underlying bias of the estimators. The median of the distribution is substantially above the zero line (yellow) for each horizon J and persistence ρ. Moreover, the medians are increasing in the horizon and can generally be ordered in magnitude from $\hat{\beta}_J^{nol}$ to $\hat{\beta}_J^{ol}$ to $\hat{\gamma}_J$ for different ρ's and J's. In this section, we derive the analytical small sample bias of the above estimators in long-horizon predictive regressions.

Figure 2: Box Plot of Long Horizon Estimators

Figure 2 below provides simulated boxplots of the distribution of the long-horizon estimators - $\hat{\beta}_J^{ol}$ (red), $\hat{\beta}_J^{nol}$ (black) and $\hat{\gamma}_J$ (green). The boxplots are shown at the 5%, 25%, 50%, 75% and 95% levels for ρ=0.7, 0.90, 0.95 and 0.99 for *J*=12, 36 and *J*=60 and *T*=600.



A. Analytical Bias Calculations for Long-Horizon Estimators

Given the process for X_t in equation (1) and the long-horizon regression of equation (2), we derive the following proposition for the small sample bias of the nonoverlapping and overlapping OLS estimators, β_J^{nol} and β_J^{ol}, under the null $\beta_1 = 0$:

Proposition 1: $E\left[\hat{\beta}_J^{nol} - \beta_J\right] = -\frac{J}{T}\frac{(1+\rho)(1+3\rho^J)}{1+\rho^J}\frac{\sigma_{uv}}{\sigma_v^2}$

Proof (see Appendix)

Proposition 2: $E\left[\hat{\beta}_J^{ol} - \beta_J\right] = -\frac{1}{T}\left[J(1+\rho) + 2\rho\left(\frac{1-\rho^J}{1-\rho}\right)\right]\frac{\sigma_{uv}}{\sigma_v^2}$

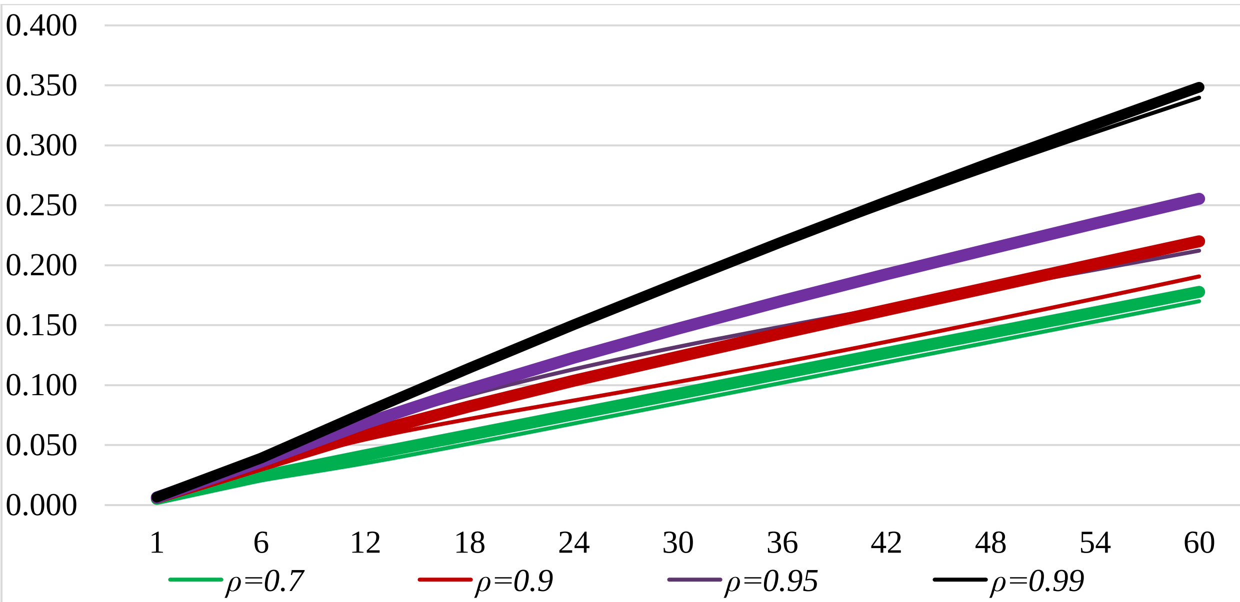
Proof (see Appendix)

Figure 3 graphs the small sample bias terms, $\frac{J(1+\rho)(1+3\rho^J)}{1+\rho^J}$ and $J(1+\rho)+2\rho\left(\frac{1-\rho^J}{1-\rho}\right)$, as a function of the horizon J for different levels of persistence, ρ, namely 0.70, 0.9 0.95, and 0.99. The reason for choosing high values of ρ reflects the high persistence level of most stock return predictive variables. The theoretical $\hat{\beta}_J^{nol}$ biases are depicted as thin lines in Figure 3, while the theoretical $\hat{\beta}_J^{ol}$ biases are given by thick lines.

Several observations are in order. First, the *ol* and *nol* biases are always increasing in the horizon J and persistence ρ. Second, this increase (at least theoretically) is generally nonlinear in shape, starting out as a concave function and eventually (as J increases) turning linear. This conversion from a concave to linear function depends on how quickly ρ^J goes to zero. Third, to this point, for values of ρ and large J such that $\rho^J \approx 0$, the biases are approximately $\frac{J(1+\rho)}{T}\frac{\sigma_{uv}}{\sigma_v^2}$ and $\frac{J(1+\rho)+2\left(\frac{\rho}{1-\rho}\right)}{T}\frac{\sigma_{uv}}{\sigma_v^2}$ for the nonoverlapping and overlapping regresssions respectively. Note that for both types of regressions the biases are increasing in $\frac{J}{T}$ with slope $\frac{(1+\rho)\sigma_{uv}}{\sigma_v^2}$. In other words, the slope varies between 1 to 2 times $\frac{\sigma_{uv}}{\sigma_v^2}$ depending on the value of ρ. Fourth, the aforementioned approximation (for $\rho^J \approx 0$) and exact calculations provided in Figure 3 show that the *ol* estimator is everywhere more biased than the *nol* estimator for a given ρ and *J*. This result is surprising given that researchers estimate equation (2) using all available overlapping data. While the *ol* estimator in theory improves the asymptotic efficiency of the estimator though less so for highly persistent variables (as shown in Figure 2), little is known about its small sample bias relative to the nonoverlapping case. It may have been logical to believe, however, that the use of "*J*"-times the data should reduce the bias, but propositions 1 and 2 show this is not the case.

Figure 3: Analytical Bias of Estimators in Nonoverlapping and Overlapping Long-Horizon Return Regressions

Figure 3 depicts the small sample bias terms of the coefficient estimators for nonoverlapping and overlapping *J*-horizon return regressions on a variable X_t with AR(1) coefficient ρ: $E\left[\hat{\beta}_J^{nol}-\beta_J\right]=\frac{J}{T}\frac{(1+\rho)(1+3\rho^J)}{1+\rho^J}\frac{\sigma_{uv}}{\sigma_v^2}$ (*nol*, thin lines) and $E\left[\hat{\beta}_J^{ol}-\beta_J\right]=\frac{1}{T}\left[J(1+\rho)+2\rho\left(\frac{1-\rho^J}{1-\rho}\right)\right]\frac{\sigma_{uv}}{\sigma_v^2}$ (*ol*, thick lines). For simplicity of interpretation, we use $\frac{\sigma_{uv}}{\sigma_v^2}=1$, *T*=600, and different levels of persistence, ρ, namely 0.70, 0.9 0.95, and 0.99.



Richardson and Stock (1989) and Valkanov (2003), among others, show that standard fixed *J* asymptotic theory $\left(\frac{J}{T}\to 0\right)$ provides a poor approximation to the true distribution of $\hat{\beta}_J$, instead arguing for an alternative asymptotic theory based on $\left(\frac{J}{T}\to\delta\right)$.[3] As Richardson and Stock (1989) argue, the point of asymptotic theory is to provide an approximation to the small sample distribution, so it is irrelevant whether the econometrician's choice of *J* is influenced by δ. Practically, for large *J* relative to T, the $\left(\frac{J}{T}\to\delta\right)$ theory provides a better approximation to the sampling distribution of $\hat{\beta}_J$ than the fixed *J*-asymptotics $\left(\frac{J}{T}\to 0\right)$ theory, resulting in more accurate critical values. The results of Propositions 1 and 2 provide an explanation. While the slope of this

[3] See also Campbell and Yogo (2006) and Hjalmarsson (2011). Most of the asymptotic theory is derived under local-to-unity asymptotics in which $\rho=1-\frac{c}{T}$, so that ρ approaches one as T goes to infinity (e.g., Elliott and Stock (1994)).

linear relationship varies across ρ, the bias is nevertheless increasing in $\frac{J}{T}$, so δ is a sufficient statistic. In other words, the small sample bias is similar for $\frac{J}{T} = \frac{J^*}{T^*} = \delta$. Importantly, this result has nothing to do with *J* being large relative to T; it holds for all *J* and T.

As pointed out by Stambaugh (1999), his small sample bias is an approximation, so there will be a departure between the small sample bias and the empirical sampling distribution. For his application, Stambaugh (1999) shows the difference is nonzero but relatively small. Stambaugh (1999) relies on Kendall's (1954) and Marriott and Pope's (1954) approximation for sample autocorrelation estimators, which is valid only up to order $\frac{1}{T}$. Sawa (1978) and Nankervis and Savin (1988) derive the exact distribution of the autocorrelation estimators.[4] They also document a relatively small error between the exact distribution and the Kendall (1954) approximation except for relatively low *T* and/or ρ close to 1 (e.g., see also MacKinnon and Smith (1988)). Nevertheless, it is important to analyze how well these approximations work for the long-horizon predictive regression studied here.

Table 1 documents the empirical small sample bias for different values of ρ (0.70, 0.9, 0.95 and 0.99), horizon *J* (1,12,60) and sample size *T* (300, 600 and 1200) in regression equation (2) using overlapping and nonoverlapping data. The simulation involves 100,000 replications of the time-series process described in equations (1) and (2). Table 1 compares the simulated values to the analytical bias calculations of Propositions 1 and 2. For the most part, the analytical and simulated results are quite similar. For example, consider *J*-period horizons of 12 and 60 and ρ =0.95 across the four sample sizes. For *J=12*, the ratios of simulated to analytical bias for nonoverlapping and overlapping data for sample size 300 is 12.00/11.95=1.00 and 12.72/12.26=1.04 respectively. Across *T*'s the ratios are (1.00, 1.01, and 1.01) and (1.04, 1.02, and 1.01), respectively again. Moreover, the ratio of simulated bias of overlapping to nonoverlapping varies from 1.06 (for *T*=300) to 1.02 (for *T*=1200) compared to the analytical ratio of 1.03. As a comparison, for *J*=60, the ratios of simulated to analytical bias for nonoverlapping and overlapping data respectively are

---

[4] De Gooijer (1980) and Shaman and Stine (1988) study the bias properties of autocorrelation estimators under more general ARMA processes than equation (1) and the ones studied in the aforementioned papers. Kiviet and Phillips (2012) extend Kendall (1954) and Marriott and Pope (1954) to approximations of order $\frac{1}{T^2}$.

(0.95, 0.98, and 1.02) and (1.02, 1.02, and 1.01). The ratio of simulated bias of overlapping to nonoverlapping varies from 1.29 (for T=300) to 1.19 (for T=1200) compared to the analytical ratio of 1.20. As expected, the bias declines as the horizon increases. Consistent with the above papers documenting the autocorrelation bias, the differences between the theoretical and simulated also decline with the horizon. Indeed, Table 1 suggests that the only real differences between the theory and simulated exist for very small samples such as T=300, which for J=60 represents just 5 nonoverlapping observations, particularly for highly persistent variables such as ρ=0.99.

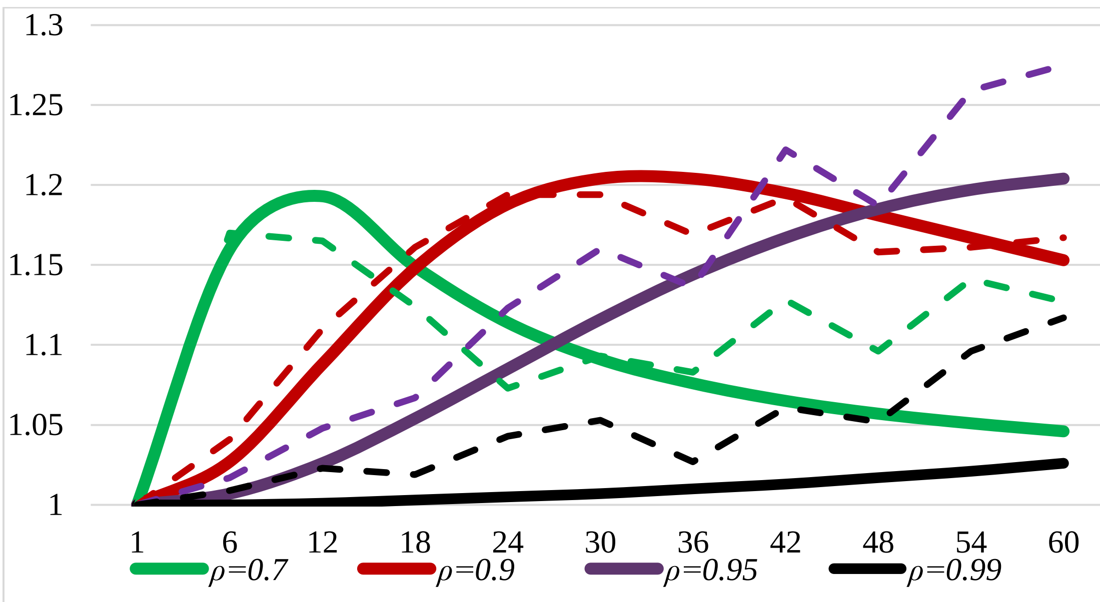
In order to pin down the comparison between the analytical and simulated results, consider one of our key findings, namely that the overlapping bias exceeds that of the nonoverlapping bias for a given J and for ρ=0.7, 0.9, 0.95 and 0.99 (as can be seen in Table 1). Figure 4 graphs the ratio of the overlapping to nonoverlapping bias for both the theoretical values from Proposition 1 and 2 and the simulated values described above for T=600. The theoretical lines are graphed as solids while the simulated ones are represented as dashed lines. Consider first the theoretical (solid) lines. In comparing Proposition 1 versus 2, the functional forms are clearly not the same - the small sample biases are different and, most surprising, the bias in the overlapping regression is everywhere greater than that of the nonoverlapping regression. The increase ranges from between 0% to 20%, depending on J and ρ. Moreover, the ratio is hump-shaped, starting at 1 for J=1, increasing with J until it eventually turns and then forever decreasing with J, eventually going back to 1. Of some note, this pattern is true across all ρ though the humped shape itself varies with ρ. For lower values of ρ the shape is tight, increasing rapidly and then, with relatively low J, the ratio begins to asymptote back towards 1. In contrast, for very persistent values of ρ, the relative bias of the overlapping estimator increases more slowly but is present at many more horizons. Importantly, at least based on this small sample bias metric, the results here put into question the use of overlapping data.

Now consider the simulated dashed lines of Figure 4. The ratio of the *ol* and nol estimates are graphed as a function of the horizon J for different levels of persistence, ρ, at T=600. The simulated small sample biases of the *ol* and *nol* estimators follow a very similar pattern to those implied by theory. The bias in the overlapping regression is greater than that of the nonoverlapping regression for different J and ρ, and similarly hump-shaped with the shape tighter (wider) for low (high) ρ.

However, as J gets large relative to T, the analytical and simulated biases do begin to diverge. This result is consistent with that of Table 1 and the previous literature on autocorrelation bias. This point aside, Table 1 and Figure 4 suggest a close link between the theory and simulated values. Importantly, the large biases of long-horizon estimators (for large J) that depend on ρ are generally worse for overlapping versus nonoverlapping regressions.

Figure 4: The Ratio of Biases Between Overlapping and Nonoverlapping OLS Estimators in Long-Horizon Return Regressions

Figure 4 depicts the ratio of biases, overlapping to non-overlapping for various forecast horizons J using analytical formulae in Proposition 1 (solid lines) as well as simulations (averages of 100,000 simulations, dashed lines), across various persistence parameters ρ of the predictive variable. The sample size for both the analytical calculations as well as for the simulation is T=600. The forecast horizon J is on the X axis and the ratio of the biases is on the Y axis.



B. Analytical Bias Calculations for Alternative Long-Horizon Estimators

Given the process for X_t in equation (1) and the alternative long-horizon regression of equation (3), we can derive the small sample bias of the OLS estimator, $\hat{\gamma}_J$. In comparing regression equations (1) and (3), note that $\mathrm{cov}\left(\sum_{i=1}^{J} R_{t+i}, X_t\right) = \mathrm{cov}\left(R_{t,t+1}, \sum_{i=1}^{J} X_{t+1-i}\right)$. Thus, the regression coefficient γ_J is related to β_J via:

$$\gamma_J = \frac{\mathrm{cov}\left(R_{t+1}, \sum_{i=1}^{J} X_{t+1-J}\right)}{\mathrm{var}\left(\sum_{i=1}^{J} X_{t+1-J}\right)} \equiv \frac{\mathrm{cov}\left(R_{t+1}, \sum_{i=1}^{J} X_{t+1-J}\right) \Big/ \mathrm{var}\left(X_t\right)}{\mathrm{var}\left(\sum_{i=1}^{J} X_{t+1-J}\right) \Big/ \mathrm{var}\left(X_t\right)} = \frac{\beta_J}{VR_J\left(X\right)}$$

where $VR_J\left(X\right)$ is the *J*-period variance ratio of X. In effect, γ_J equals β_J scaled down by $VR_J\left(X\right)$. The expected value of the estimator, $\hat{\gamma}_J$, is given by $E\left[\hat{\gamma}_J\right] = E\left[\frac{\hat{\beta}_J}{V\hat{R}_J\left(X\right)}\right]$. In the appendix, under the null of no predictability, we prove:

Proposition 3: $E\left[\hat{\gamma}_J\right] \approx \frac{E\left[\hat{\beta}_J\right]}{E\left[V\hat{R}_J\left(X\right)\right]}\left(1 - \frac{\mathrm{cov}\left(\hat{\beta}_J, V\hat{R}_J\left(X\right)\right)}{E\left[\hat{\beta}_J\right]E\left[V\hat{R}_J\left(X\right)\right]} + \frac{\mathrm{var}\left(V\hat{R}_J\left(X\right)\right)}{E\left[V\hat{R}_J\left(X\right)\right]^2}\right)$

where $E\left[\hat{\beta}_J\right] = -\frac{1}{T}\left[J(1+\rho) + 2\rho\left(\frac{1-\rho^J}{1-\rho}\right)\right]\frac{\sigma_{uv}}{\sigma_v^2}$

$$EV\hat{R}_J(X) = J + 2\sum_{i=1}^{J-1}(J-i)\left[\rho^j - \frac{1}{T}\left[(1+\rho)\frac{1-\rho^j}{1-\rho} + 2j\rho^j\right]\right]$$

$$\mathrm{cov}\left(\hat{\beta}_J, V\hat{R}_J\left(X\right)\right) = \frac{\sigma_{uv}}{\sigma_v^2}\left[2\sum_{i=1}^{J-1}\left(J-i\right)\mathrm{cov}(\hat{\rho}_i, \hat{\rho}_J) + 2(1-\rho)\sum_{i=1}^{J-1}\sum_{k=1}^{J-1}\left(J-i\right)\mathrm{cov}\left(\hat{\rho}_i, \hat{\rho}_k\right)\right]$$

$$\mathrm{var}\left(V\hat{R}_J\left(X\right)\right) = 4\sum_{i=1}^{J-1}\sum_{k=1}^{J-1}\left(J-i\right)\left(J-k\right)\mathrm{cov}\left(\hat{\rho}_i, \hat{\rho}_k\right)$$

$$\mathrm{cov}\left(\hat{\rho}_i, \hat{\rho}_k\right) = \frac{1}{T}\rho^{k-i}\left[\frac{\left(1+\rho^2\right)\left(1-\rho^{2i}\right)}{1-\rho^2} + \left(k-i\right)\right] - \frac{1}{T}\left(i+k\right)\rho^{i+k}, \quad k \geq i$$

and $\mathrm{var}\left(\hat{\rho}_i\right) = \frac{1}{T}\left[\frac{\left(1+\rho^2\right)\left(1-\rho^{2i}\right)}{1-\rho^2}\right] - \frac{2}{T}i\rho^{2i}$

Proof (see Appendix)

As unwieldly as Proposition 3 looks, note that $E\left[\hat{\gamma}_J\right]$ is closed form and is a specific function of ρ, *J* and *T*.

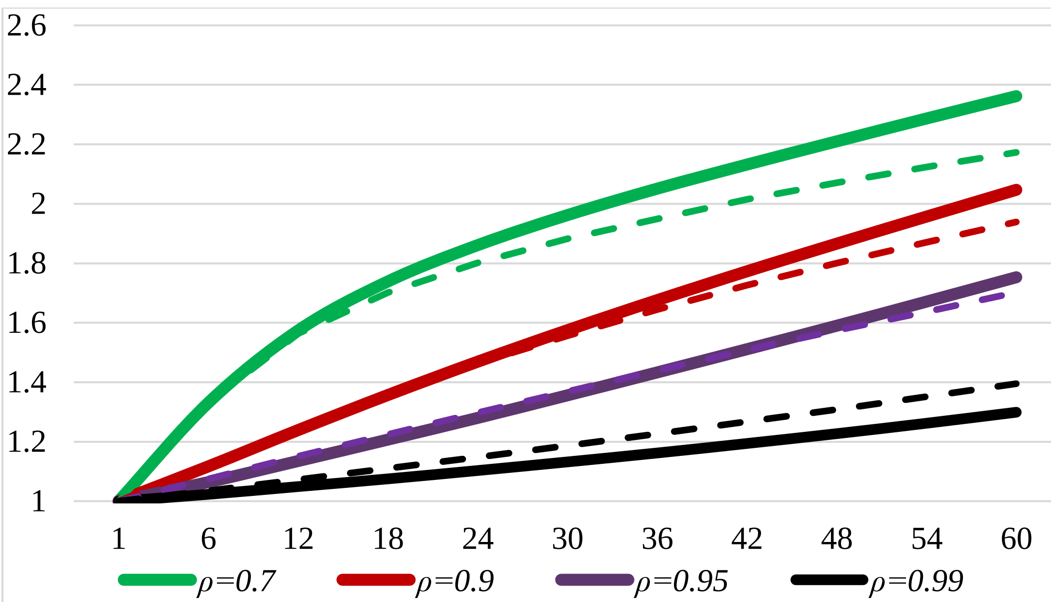
Figure 5 graphs the ratio of the bias of the alternative long-horizon estimator $\hat{\gamma}_J$ to the standard long-horizon estimator $\hat{\beta}_J^{ol}$ for both the theoretical values from Proposition 2 and 3 and the simulated values for T=600 and ρ=0.7, 0.9, 0.95 and 0.99. The theoretical lines are graphed as solids while the simulated ones are represented as dashed lines. In order to make the estimators comparable, note that $\hat{\gamma}_J$ is scaled up by the true $VR_J\left(X\right)$ which is a known function of ρ and *J*.

Consider first the theoretical (solid) lines. In comparing Propositions 2 versus 3, the bias of the alternative regression estimator $\hat{\gamma}_J$ is everywhere greater than that of the standard overlapping long-horizon estimator $\hat{\beta}_J^{ol}$. Some consistent patterns emerge. First, the ratio of the biases is increasing in the horizon *J*. This is bad news for this estimator; its main purpose is precisely for large *J* when HAC estimators, like those of Newey and West (1987), have particularly poor properties. Second, though still worse, the ratio of the biases is generally closer for high levels of persistence ρ. However, this finding is not because the alternative long-horizon estimator is in some sense getting "less biased" but rather the bias of the standard long-horizon estimators, both overlapping or $\hat{\beta}_J^{ol}$ and nonoverlapping $\hat{\beta}_J^{nol}$, are getting considerably worse (e.g., see Figure 3). Finally, consider the simulated dashed lines of Figure 5. The ratio of the simulated bias of $\hat{\gamma}_J$ to $\hat{\beta}_J^{ol}$ maps closely to the analytical small sample biases of these estimators. This is true both in terms of magnitudes and the underlying patterns across different ρ and *J*.

The bottom line from this analysis is that if the researcher wants to estimate long-horizon regression equation (2), the regression equation (3) is not a very good alternative. While the methodology avoids HAC standard calculations, Figure 2 shows that it is likely not a very efficient estimator. Its small sample distribution is wide relative to the overlapping regression estimator. Even worse, the bias of this estimator is a magnitude greater than that of the overlapping regression estimator which is already not superior to its nonoverlapping counterpart. For example, for *J*=60, these biases range from 25% (for high ρ, such as 0.99) to 100% (for lower ρ, such as 0.90). As shown in Figure 1, the long-horizon biases are already sufficiently large to effectively reduce the magnitude of the estimates. These results should not be a surprise. This alternative regression estimator $\hat{\gamma}_J$ is effectively the long horizon estimator $\hat{\beta}_J^{ol}$ scaled by the long-horizon variance ratio estimator of the predictive variable, $V\hat{R}_J(X)$. The variance-ratio estimator suffers from similar biases, thus compounding the problem in estimating $\hat{\gamma}_J$.

Figures 5: The Ratio of the Bias of Long-Horizon Estimators ($\hat{\gamma}_J$ to $\hat{\beta}_J^{ol}$)

Figure 5 depicts the ratio of biases of $\hat{\gamma}_J$ regressions (see (3)) relative to $\hat{\beta}_J^{ol}$ regressions (see (2)) for various forecast horizons *J* using analytical formulae in propositions 3 and 2 respectively (solid lines) as well as simulations (averages of 100,000 simulations, dashed lines), across various persistence parameters ρ of the predictive variable. The sample size for both the analytical calculations as well as for the simulation is T=600. The forecast horizon *J* is on the X axis and the ratio of the biases is on the Y axis.



While our results comment on the viability of the popular transformation of the long-horizon equation (2) to equation (3), our findings also allow us to comment on a much larger literature in finance and macroeconomics that is not interested in long-horizon regressions per se. It is quite common to regress single period changes in the variable of interest, like stock returns, on a predictive variable, constructed from a long-run smoothed out series. This long-run series often takes the form of a moving-average or a stochastic trend. Examples of popular predictors include cyclically adjusted price earnings (CAPE) ratio (e.g., Campbell and Shiller (1988)); long-term (i.e., 5-year) reversals as a measure of value (e.g., De Bondt and Thaler (1985) and Fama and French (1988), among many others); stock return momentum (i.e., 1 year) (e.g., Jegadeesh and Titman (1993), Asness (1994) and Carhart (1997), among many others); moving averages of inflation (e.g., Cieslak and Povala (2015) and Bauer (2017) in fixed income); volume (e.g., in microstructure); risk-return regressions using measures of volatility and beta; and so forth. Two points of note are (i) many of these analyses are effectively long-horizon regressions (i.e., going from equation (3) to equation (2)) and thus subject to the large biases documented in this paper, and (ii) to the extent these regressions are long-horizon regressions, the methodological approach of using equations like (3) are especially problematic.

Given the efficiency issues underlying the estimators in equations (2) and (3), researchers have proposed a structural modelling approach to long-horizon predictability. In particular, the estimation strategy is to jointly estimate the short-horizon return process and autoregressive process for the predictive variable. Given this joint estimation of $\left(R_{t,t+1}, X_t\right)$ based on $\left(X_t, X_{t-1}, \ldots X_{t-m}\right)$ where *m* is small relative to *J*, the researcher can infer a long-horizon *J*-period return forecast. (See, for example, Kandel and Stambaugh (1989), Campbell (1991), Hodrick (1992) and Boudoukh and Richardson (1994)). Given equation (1), the researcher estimates β_J from β_1 and ρ_X. Boudoukh and Richardson (1994) show that a consistent estimator is $\hat{\beta}_J^{imp} = \hat{\beta}_1 \frac{1-\hat{\rho}^J}{1-\hat{\rho}}$ where *imp* refers to the *J*-period estimator implied from the nonlinear function of $\hat{\beta}_1$ and $\hat{\rho}$. They show that in comparison to the other aforementioned long-horizon estimators, $\hat{\beta}_J^{ol}, \hat{\beta}_J^{nol}$ and $\hat{\gamma}_J$, the asymptotic variance of $\hat{\beta}_J^{imp}$ can be magnitudes lower, especially for less persistent X_t.

Putting aside the important issue that, in contrast to the estimators $\hat{\beta}_J^{ol}, \hat{\beta}_J^{nol}$ and $\hat{\gamma}_J$, $\hat{\beta}_J^{imp}$ will be an inconsistent estimator of β_J if equation (1) is misspecified, there has been no analysis to date of the small sample bias of $\hat{\beta}_J^{imp}$. On the one hand, the bias might be small since the estimation requires estimates of only $\hat{\beta}_1$ and $\hat{\rho}$ which are much less biased than their long-horizon counterparts. On the other hand, $\hat{\beta}_J^{imp}$ is a nonlinear function of these estimators and thus any small sample bias could be amplified. In the appendix, we derive the small sample bias of $\hat{\beta}_J^{imp}$:[5]

Proposition 4: $E\left[\hat{\beta}_J^{imp}-\beta_J\right]\approx\frac{1-\rho^J}{1-\rho}\left(-\frac{1+3\rho}{T}\right)-\frac{1}{2}\left[\frac{J\rho^{J-1}}{1-\rho}-\frac{1-\rho^J}{(1-\rho)^2}\right]\left(\frac{1-\rho^2}{T}\right)$

Proof (see Appendix)

Figure 6 below graphs the ratio of the bias of the implied long-horizon estimator $\hat{\beta}_J^{imp}$ to the standard long-horizon estimator $\hat{\beta}_J^{ol}$ for both the theoretical values from Proposition 2 and 4 and the simulated values for T=600 and ρ=0.7, 0.9, 0.95 and 0.99. The theoretical lines are graphed as solids while the simulated ones are represented as dashed lines.

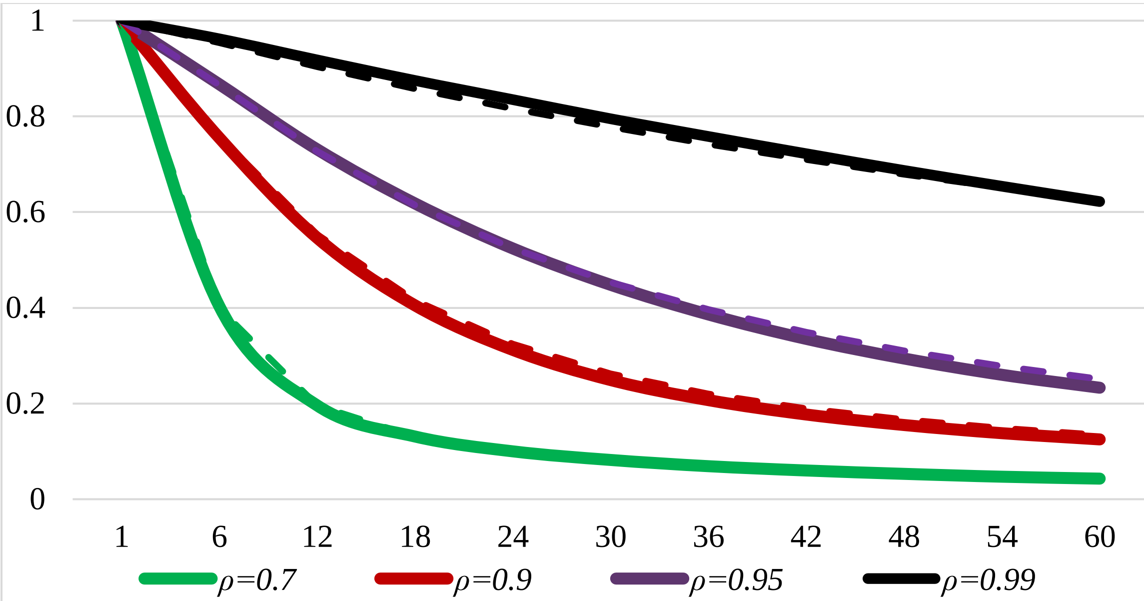
Some observations are in order. First, in comparing Propositions 2 versus 4, the bias of the implied long horizon regression estimator $\hat{\beta}_J^{imp}$ is everywhere smaller than that of the standard overlapping long-horizon estimator $\hat{\beta}_J^{ol}$. Second, this ratio declines with J and ρ. This is mixed news for this estimator in terms of its applications to finance. Given that the biases are problematic for long-horizon estimators for large J, $\hat{\beta}_J^{imp}$ provides a viable option to these more standard estimators. However, as Figure 6 shows, the bias improvement is considerably smaller for $\hat{\beta}_J^{imp}$ for ρ close to 1, a common feature of stock return predictors. Finally, the differences between the analytical (solid lines) and simulated (dashed lines) demonstrate the efficacy of the small sample bias approximations. In other words, given equation (1), Proposition 4 can be used to adjust $\hat{\beta}_J^{imp}$ in small samples.

---

[5] Marriott and Pope (1954) derive the asymptotic variance of the autocorrelation estimator to order T, i.e., $\left(\frac{1-\rho^2}{T}\right)$. Subsequent to Marriott and Pope (1954), a number of authors have provided small sample approximations to this variance (e.g., see White (1961), Shenton and Johnson (1965), Sawa (1978) and De Gooijer (1980)). For the simulation results to follow, we use Shenton and Johnson's (1965) approximation which performs better for the ρ, J and T faced in our problems, i.e., $\mathrm{var}(\hat{\rho})=\left(\frac{1-\rho^2}{T}\right)\approx\left(\frac{1-\rho^2}{T}\right)-\left(\frac{1-14\rho^2}{T^2}\right)+\left(\frac{5-78\rho^2+76\rho^4}{T^3(1-\rho^2)}\right)$.

Figures 6: The Ratio of the Bias of Long-Horizon Estimators ($\hat{\beta}_J^{imp}$ to $\hat{\beta}_J^{ol}$)

Figure 6 depicts the ratio of biases of $\hat{\beta}_J^{imp}$ relative to $\hat{\beta}_J^{ol}$ regressions for various forecast horizons J using analytical formulae in propositions 4 and 2 respectively (solid lines) as well as simulations (averages of 100,000 simulations, dashed lines), across various persistence parameters ρ of the predictive variable. The sample size for both the analytical calculations as well as for the simulation is T=600. The forecast horizon J is on the X axis and the ratio of the biases is on the Y axis.



III. Extensions

The above theoretical results for the biases of various estimators in long-horizon predictive regressions provide closed form solutions as a function of $\frac{\sigma_{uv}}{\sigma_v^2}$ ρ, J and T. Simulation results above show that these small sample bias formulas approximate well in small samples. The formulas are derived assuming the model structure of equation (1) under the null hypothesis of $\beta_1 = 0$. This latter assumption may seem innocuous given Stambaugh's (1999) result, $E\left[\hat{\beta}_1 - \beta_1\right] = -\frac{\sigma_{uv}}{\sigma_v^2}\left(\frac{1+3p}{T}\right)$. That is, the bias is fixed for all β_1. It turns out, however, that this result just holds for the single horizon case, J=1. For $J > 1$, the bias has two components, one shrinking the magnitude of β_1 and the other a fixed bias along the lines of Propositions 1 and 2.

Another key assumption that may be violated in the data is that the bias depends only on contemporaneous correlations of the innovation terms, $\frac{\sigma_{uv}}{\sigma_v^2}$, that is, assuming no lead-lag effects, $\operatorname{cov}(u_t, v_{t-k}) = 0 \ \ \forall k \neq 0$. Finally, equation (1) describes a univariate regression of returns on a predictive variable following an AR(1). It may be of interest to consider multivariate models with more elaborate AR(p) representations of the predictive variables. For example, Shaman and Stine (1988) provide an extension to Kendall (1954) and Marriott and Pope (1954) for the case of AR(p), and Nicholls and Pope (1988) and Pope (1990) consider the multivariate case. Amihud and Hurvich (2008, 2010) apply some of these results to the single horizon framework of equation (1). In theory, these results also extend to the multiple horizon case albeit with a fair degree of complication. We leave this extension to future research, but below explicitly solve for the $\beta_1 \neq 0$ case and describe procedures for dealing with $\operatorname{cov}(u_t, v_{t-k}) \neq 0$.

A. Analytical Bias Calculations for Long-Horizon Estimators (assuming $\beta_1 \neq 0$)

The proofs of Propositions 1 and 2 in the appendix derive formulas for the bias of estimators $\hat{\beta}_J^{ol}$ and $\hat{\beta}_J^{nol}$ under the alternative hypothesis of predictability, i.e., $\beta_1 \neq 0$ in equation (1).

Specifically,

$$E[\hat{\beta}_{JNOL} - \beta_J] = -\frac{1+3\rho^J}{T/_J}\left[\frac{1+\rho}{1+\rho^J}\ \frac{\sigma_{uv}}{\sigma_v^2} + \beta_J\,\frac{\rho-\rho^J}{(1-\rho^{2J})}\right] \tag{4}$$

$$E[\hat{\beta}_{JOL} - \beta_J] = -\frac{1}{T}\left[J(1+\rho) + 2\rho\left(\frac{1-\rho^J}{1-\rho}\right)\right]\left[\frac{\sigma_{uv}}{\sigma_v^2} + \beta_J\left(\frac{1}{1-\rho^J} - \frac{1}{J(1-\rho)}\right)\right]$$

The first term of equation (4) is a fixed bias adjustment as a function of $\frac{\sigma_{uv}}{\sigma_v^2}$ ρ, *J* and *T*. The second term is the adjustment under the alternative hypothesis of predictability. Interestingly, this second term scales down the magnitude of β_J by a factor, independent of β_J. Thus, the larger the β_J the greater is the magnitude in adjustment to β_J. For $J = 1$, the biases in equation (4) reduce to the Stambaugh bias, and, for $\beta_J = 0$, the bias equals those given in Propositions 1and 2. Note that if β_J and σ_{uv} are of the same sign, then the bias gets amplified. In contrast, if β_J and σ_{uv} are of different signs, then the bias gets offset, reducing the overall effect. For many finance applications, e.g., those involving valuation ratios as predictors, the latter case is more relevant. Thus, the estimation bias will be less for $\beta_1 \neq 0$.

Table 2 documents the empirical small sample bias of $\hat{\beta}_{JOL}$ for different values of ρ (0.70, 0.9, 0.95 and 0.99), horizon *J* (12, 36, and 60) and sample size *T* (300, 600 and 1200) in regression equation (2) using overlapping data for different values of β_1 in equation (4). The simulation involves 100,000 replications of the time-series process described in equations (1) and (2). One question is what values of β_1 should be chosen to illustrate the bias of $\hat{\beta}_{JOL}$ for nonzero β_1? We choose R^2s of 0.25% to 0.75% (in the first column) for the single horizon regression in equation (1) to match those documented empirically at long horizons (in the last column). This range of short horizon R^2 values then correspond to a range of β_1 (for different ρ) used in the simulation. Before analyzing the bias calculations, it is important to point out that Table 2's very high R^2s at long horizons across all β_1s (even zero) can be explained by the well-known small sample bias in R^2 (e.g., see Cramer (1987)). As an illustration, consider the $\beta_1 = 0$ case for ρ=0.95, *J*=60 and *T*=300, 600, and 1200. The simulated average R^2s are 29.6%, 14.7% and 7.3% even though there is no predictability by construction. The bias of long horizon R^2 is driven by both the bias of the coefficient estimators documented in Section II and the variation of $\hat{\beta}_{JOL}$. Because R^2 is a squared measure and thus truncated at 0%, $\text{var}\left(\hat{\beta}_{JOL}\right)$ plays an important role in the R^2 bias. The simulations of Table 1 and Figure 3 demonstrate the large bias of $\hat{\beta}_{JOL}$ for large *J* and Figure 2 similarly shows the large small sample variance of $\hat{\beta}_{JOL}$. For $\beta_1 \geq 0$, the average R^2s increase, but only marginally compared to the $\beta_1 = 0$ case. For example, using ρ=0.95 and *J*=60, for $\beta_1 = 0.25,\ 0.5 \text{ and } 0.75$, the R^2s are respectively 31.6%, 33.0% and 34.7% for *T*=300; 17.0%, 19.2% and 21.8% for *T*=600; and 9.6%, 12.2% and 15.2% for *T*=1200. In other words, a significant fraction of reported R^2s are likely bias-related.

Table 2 also compares the simulated values to the analytical bias calculations of equation 4 above. First, for the most part, the analytical and simulated biases are quite similar. For example, consider $\beta_1 = 0.5$, ρ=0.95 and *J*-period horizons of 12 and 60 across the three sample sizes. For *J=12*

(J=60), the ratios of simulated to analytical bias for sample size 300, 600 and 1200 are respectively 11.04/10.31=1.07 (0.89), 5.49/5.29=1.04 (0.95) and 2.69/2.68=1.00 (0.95). Second, as implied by equation (4), the biases are greater for the $\beta_1 = 0$ versus $\beta_1 \neq 0$ cases given β_J and σ_{uv} are of the opposite sign in our simulation. As an illustration, for J=60 and T=600, the ratio of the analytical bias for $\beta_1 = 0.5$ against $\beta_1 = 0$ for ρ = 0.70, 0.9, 0.95 and 0.99 are 14.5/17.6=0.82, 15.2/21.2=0.72, 15.8/23.8=0.66 and 20.3/28.3=0.72. Consistent with the first point above, the simulated bias ratios of the $\beta_1 = 0$ versus $\beta_1 = 0.5$ cases are almost identical to those implied by the analytical formulas, i.e., 0.82, 0.70, 0.64 and 0.69 respectively for ρ = 0.70, 0.9, 0.95 and 0.99. Finally, as expected, while the bias declines with β_1 and with the number of observations T, the bias increases with the horizon J and persistence ρ. Consistent with Table 1, the differences between the theoretical and simulated also decline with T. The simulation results here provide comfort for researchers requiring a plug-in bias adjustment like equation (4) above. If the researcher is not focused solely on the null of no predictability, and instead has priors over a range of β_1, then equation (4) allows the researcher to infer the possible ex ante bias of the regression estimators by integrating over possible values of β_1.

B. Analytical Bias Calculations for Long-Horizon Estimators (assuming $\text{cov}(u_t, v_{t+k}) \neq 0$ for k ≥ 1)

Stambaugh's (1999) model given by equation (1) assumes $\text{cov}(u_t, v_{t+k}) = 0$ for k ≥ 1. For some applications in finance, this assumption may be a poor one. As an illustration, consider equation (1) with $X_t = Z_{t-k}$, i.e., a lagged predictive variable. In this case, $\text{cov}(u_t, v_{t+k}) = \sigma_{uv}$, and the usual assumption no longer holds. Thus, even if $\text{cov}(u_t, v_t) = 0$, the coefficient estimators will still be biased. (Note that we explore this case empirically in Section IV below when comparing long-horizon return predictability using dividend-price ratios rather than dividend yields.)

In particular, in the appendix, for $X_t = Z_{t-k}$, we derive the following result:

Proposition 5: $E\left[\hat{\beta}_J^{ol} - \beta_J\right] = -\frac{1}{T}\left[J(1+\rho) + 2\rho^{k+1}\left(\frac{1-\rho^J}{1-\rho}\right)\right]\frac{\sigma_{uv}}{\sigma_v^2}$

Proof (see Appendix)

When *J*=1, Proposition 5 extends the Stambaugh (1999) bias to lagged predictors with the formula: $E\left[\hat{\beta}_1\right] = -\frac{1}{T}\left(1+\rho+2\rho^{k+1}\right)$. The difference between the bias for Stambaugh's (1999) $k = 0$ versus the $k \geq 1$ case is $\frac{-2\rho\left(1-\rho^k\right)}{T}$. Note that this difference is not monotonic in ρ for large *k*. For very high ρ, there is little difference in bias because Z_{t-k} and Z_t are effectively the same. For low ρ, the horizon *k* gets drowned out and the difference is effectively $\frac{-2\rho}{T}$. In contrast, for high ρ, there is a mix of these two effects, and the difference may no longer be small. The same intuition carries through for all *J* and thus the long-horizon estimator given by Proposition 5, i.e., for the case $J >> 1$.

The above example serves to illustrate the importance of modeling assumptions for the magnitude of the bias resulting from equations (1) and (2). As an illustration, assume $v_{t+1} = \sum_{k=0}^{K-1} \lambda_k u_{t+1-k} + \eta_{K,t+1}$ in equation (1). This equation is a representative model if asset return innovations forecast future realizations of the predictive variable. For example, future changes in valuation ratios based on cash flows (CF_t) and the corresponding asset price (P_t), $\ln\left(\frac{CF_{t+k}}{P_{t+k}}\right) - \ln\left(\frac{CF_{t+1}}{P_{t+k}}\right) \equiv \ln\left(\frac{CF_{t+k}}{CF_{t+1}}\right) - \ln\left(\frac{P_{t+k}}{P_{t+1}}\right)$, may be predictable because although current returns, $\ln\left(\frac{P_{t+1}}{P_t}\right)$, do not forecast future returns, $\ln\left(\frac{P_{t+k}}{P_{t+1}}\right)$, $\ln\left(\frac{P_{t+1}}{P_t}\right)$ does have news for future cash flow growth, $\ln\left(\frac{CF_{t+k}}{CF_{t+1}}\right)$. (See, for example, Kothari, Lewellen and Warner (2006), Sadka and Sadka (2009), and He and Hu (2014).) In Section IV below, we empirically discuss the case of earnings-to-price ratios which fall into the above class of predictors.

In addition, it is a common perception that the Stambaugh (1999) bias is less relevant for macroeconomic predictors due to the lower contemporaneous correlation between asset returns and macro innovations. The above discussion puts this view into question. If returns on assets, such as stocks and bonds, are leading indicators for future macroeconomic realizations, then macroeconomic shocks (implied by univariate time-series models) will not be uncorrelated with past returns. The mechanics of the problem correspond to those of Proposition 5 above albeit with a different specification. The bias of the estimator $\hat{\beta}_J$ using macroeconomic predictors will equal the cumulative sum of the biases associated with each $\mathrm{cov}(u_t, v_{t+k})$ and then summed up over the multiple horizons for *J*-period return regressions. As with the AR(1) representation in equation (1), the plug-in formulas will be tied to the specified model as $E[\hat{\rho}_j] - \rho_j$ will change with the model (e.g., see Shaman and Stine (1988)). Importantly, for future research, the calculation of the bias of $\hat{\beta}_J$ follows the methodological approach used for Propositions 2 and 5.

IV. Empirical Application

Sections II above documents large small-sample biases of the coefficient estimators in long-horizon predictive regressions. The asymptotic distribution of the *t*-statistic, $\frac{1}{\sqrt{T}}\frac{\hat{\beta}_J}{\hat{\sigma}(\hat{\beta}_J)}$, in the long-horizon regression equation (2) is normally distributed with mean zero and variance 1. In terms of the small sample properties of this statistic, however, it is natural that the aforementioned biases distort the distribution of this test statistic. In this section, we analyze the bias-adjusted long-horizon coefficients and corresponding t-statistics for stock return predictors based on valuation ratios. The question is: what implications do the bias results have for long horizon stock return predictability regressions?

A large literature has emerged since Fama and French's (1988) original long-horizon stock return regression on dividend yields. This literature is partially summarized by Campbell, Lo and Mackinlay (1997), Ang and Bekaert (2007) and Cochrane (2011). A particularly well-known paper in this literature is Welch and Goyal (2007) who perform both short- and long-horizon return regressions using various predictive variables dating back to 1965. While Welch and Goyal (2007) evaluate the performance of these predictors out-of-sample, we focus on in-sample results. We document the results for valuation ratio-based predictors taken from Amit Goyal's website, including dividend-to-price, dividend yield, earnings yield, cyclically adjusted earnings yield using nominal and real earnings, and book to market. Note that each predictor comes with its own unique $\hat{\rho}$ and $\frac{\hat{\sigma}_{uv}}{\hat{\sigma}_v^2}$ given the specification in equation (1). Furthermore, in terms of the discussion in Section III.b, each predictor also may have a lead-lag structure with its innovations, i.e., $\text{cov}(u_t, v_{t+k})$ for $k \geq 1$.

Table 3 documents the results for each of these predictors for $\hat{\beta}_{JOL}$ (i.e., regression equation (2)). To coincide with Welch and Goyal (2007), the results are reported for horizons of J=1, 12, 36 and 60 months over the period 1968 to 2017 (i.e., 600 monthly observations). We report $\hat{\beta}_{JOL}$ and $\left(\hat{\beta}_{JOL} - E\left[\hat{\beta}_{JOL}\right]\right)$, and the t-statistics associated with these estimates, $\frac{1}{\sqrt{T}} \frac{\hat{\beta}_{JOL}}{\hat{\sigma}(\hat{\beta}_J)}$ and $\frac{1}{\sqrt{T}} \frac{\hat{\beta}_{JOL} - E[\hat{\beta}_{JOL}]}{\hat{\sigma}(\hat{\beta}_J)}$. In the simulations of Sections II and III, we knew the true ρ for approximating the bias. In practice, we follow Amihud and Hurvich (2004) and plug-in the bias-adjusted value, $\hat{\rho} + \frac{1+3\hat{\rho}}{T}$.[6] As reported elsewhere, typical HAC standard error calculations of $\hat{\sigma}\left(\hat{\beta}_J\right)$ have poor small sample properties (see footnote 2). In order to correct for these small sample problems, we use the analytical asymptotic value under an AR(1) in equation model (1) (see Boudoukh, Richardson and Whitelaw (2008)), $\sigma^2\left(\hat{\beta}_J^{OL}\right) = \frac{J\,\text{var}(R_{t,t+1})}{T\,\text{var}(X_t)} \sqrt{1 + \frac{2}{J}\frac{\rho}{1-\rho}\left((J-1) - \frac{\rho}{1-\rho}\left(1-\rho^{J-1}\right)\right)}$. Table 3 also documents the empirical p-value of the $\hat{\beta}_{JOL}$ estimate under two different simulation models: (i) equation (1) with parameters to match those of the predictors, and (ii) equation (1) with the additional assumption $v_{t+1} = \sum_{k=0}^{K-1} \lambda_k u_{t+1-k} + \eta_{K,t+1}$, again parameters chosen to match those of the data. The last three columns of Table 3 correspond to this latter assumption.

---

[6] Note that this bias adjusted estimate could be iterated down further, e.g., $\hat{\rho} + \frac{1+3\left(\hat{\rho} + \frac{1+3\hat{\rho}}{T}\right)}{T}$ and so forth. Amihud and Hurvich (2004) find that these adjustments make little difference for the sample sizes used in finance.

Several observations are in order. First, and foremost, the headline result of Figure 1, namely that the bias-adjusted coefficient on dividend price ratios is essentially zero, comes through. For J=1, 12, 36 and 60, the coefficient (and t-value) drop from 0.005 (1.33), 0.074 (1.54), 0.203 (1.44) and 0.354 (1.54) to -0.001 (-0.28), -0.004 (-0.081), -0.030 (-0.23) and -0.034 (-0.15). The simulated p-values range from 0.38 to 0.42, that is, in the center of the distribution. Second, the standard overlapping estimator does not produce any significant t-statistics for the six predictors. Of course, part of the explanation is due to overlapping data providing little benefit for highly persistent regressors (i.e., Figure 2). In effect, the sample sizes are too small to generate statistically significant results. What we also document here, however, is that the coefficients are small once they are bias adjusted. Indeed, the simulated p-values are mostly towards the center of the distribution of the $\hat{\beta}_{JOL}$ across J. And, importantly, the evidence for predictability does not show up more strongly at the 60-month horizon. This finding is in contrast to the common view in the literature on long-horizon predictability.

That said, there are three interesting, and surprising, results that emerge from the empirical analysis of Table 3. First, Welch and Goyal (2007) differentiate dividend yields from dividend-to-price ratios via the lag of the price variable, in other words, $DY_t \approx DP_{t-12}$. Consider the horizons J=12 and 60. The unadjusted coefficient estimates $\hat{\beta}_{12}$ and $\hat{\beta}_{60}$ are respectively 0.069 and 0.311 with t-statistics 1.40 and 1.33, and, when adjusting for contemporaneous correlation σ_{uv}, barely changes to 0.064 and 0.291 with t-statistics 1.32 and 1.24. However, when we use Proposition 5 to correctly adjust the bias because $cov(u_t, v_{t+k}) = \sigma_{uv}$ for $k = 12$ and not $k = 0$, the bias-adjusted estimates $\hat{\beta}_{12} - E[\hat{\beta}_{12}]$ and $\hat{\beta}_{60} - E[\hat{\beta}_{60}]$ (in column 7) are now -0.028 and -0.101 with t-statistics -0.58 and -0.43. Second, as described in Section III.b, stock returns are known to forecast future earnings growth. Taking equation (1) and estimating the lead-lag structure of the innovations of R_t and X_t (i.e., λ_k in the above model) for each of the predictors, only EP shows sufficient structure. For example, for k=0 to 6, $\text{corr}(v_{t+1}, u_{t+1-k})$ = -0.58, 0.11, 0.16, 0.12, 0.11, 0.06 and 0.14 (not reported in Table 3). Now consider the horizons J=12 and 60. The unadjusted coefficient estimates $\hat{\beta}_{12}$ and $\hat{\beta}_{60}$ are respectively 0.044 and 0.140, and, when adjusting for contemporaneous correlation σ_{uv}, are 0.014 and -0.006, but, when including the lead-lag structure, increase to 0.046 and 0.152 (in column 7).[7] In words, estimates using *EP* have effectively little bias given the structure $v_{t+1} = \sum_{k=0}^{K-1} \lambda_k u_{t+1-k} + \eta_{K,t+1}$. Nevertheless, the estimates themselves are still insignificant at conventional levels, partly due to the magnitude of the coefficients being smaller for *EP* and also the large standard errors due to large *J* relative to *T*. Of all the predictors, however, *EP* estimates are furthest out in the distribution, ranging from 0.72 to 0.81. Finally, a quick look at Table 3 shows that without any adjustments the "weakest" predictor is *B/M* in term of t-statistics. Ironically, when adjusting for the bias, the *B/M* results flip sign from positive to negative coefficients and t-statistics. Though still insignificant, B/M is now the "strongest" of the predictors (the above lead-lag adjustment for *EP* aside) but with the opposite sign to the conventional estimates. This finding illustrates how the magnitude of the bias, and thus ignoring the bias, can very much steer the researcher away from the potentially interesting results.

V. Conclusion

Expected return variation is at the center of modern financial asset pricing. Much of this literature hinges on the large documented regression coefficients at long horizons. This paper extends Stambaugh's (1999) famous bias result for stock return predictability to long-horizons. We provide a plug-in estimator which is a closed-form function of only the parameters $\frac{\sigma_{uv}}{\sigma_v^2}$ ρ, *J* and *T*. The biases increase with ρ and linearly in *J* (for large *J*). While there is a long literature debating the statistical significance of long horizon estimators given the typical horizons and sample sizes employed, our point is very different. Our analytical calculations put into serious question the magnitude of the return predictability.

The applicability of our method is widespread. We show the link between the typical long-horizon regression estimator and one based on short-horizons with moving-averages of the predictors. We discuss bias approximations under the null and alternative of return predictability and describe how more general models of the econometric structure between return innovations and those of

[7] Recall from Section II.b that, given the model structure of the errors, the analytical bias calculations require a new formula for $\hat{\rho}_j - E[\hat{\rho}_j]$. The lead-lag bias adjustments therefore in Table 3 use simulated rather than analytical values.

the predictive variable may impact the results. We provide empirical examples using valuation ratios for forecasting stock returns. Many of the issues and results brought up in this paper are especially applicable to predictability results in the fixed income and exchange rate area. We hope to document important findings in future research.

Appendix - Proofs

Proposition 1:

Based on equation (1), we can write the two nonoverlapping *J*-period equations for $\sum_{i=1}^{J} R_{t+i}$ and X_t as:

$$\sum_{i=1}^{J} R_{t+i} = \alpha_J + \beta_J X_t + \varepsilon_{t:t+J} \qquad (A1)$$
$$X_{t+j} = \mu_J + \rho^J X_t + \left(1 v_{t+J} + \rho v_{t+J-1} + \rho^2 v_{t+J-2} + \cdots + \rho^{J-1} v_{t+1}\right)$$

where

$$\beta_J = \beta_1 \tfrac{1-\rho^J}{1-\rho}$$
$$\varepsilon_{t,t+J} = \sum_{i=1}^{J} u_{t+i} + \beta_1 \sum_{i=1}^{J-1} \left(\tfrac{1-\rho^i}{1-\rho} v_{t+J-i} \right) \qquad (A2)$$
$$v_{t,t+J} \equiv \sum_{i=1}^{J} p^{(i-1)} v_{t+J-(i-1)}$$

Regression equation (A1) is run every *J* sampling periods. Following Stambaugh (1999), assume $b_{1:J} = \left(\alpha_J, \beta_J\right)$ and $b_{2:J} = \left(\mu_J, \rho_J\right)$, and $X = \left(1 \quad X_t\right)$, $t = 1, J+1, \ldots, T-J$, then

$$\hat{b}_{1:J}^{NOL} - b_{1:J} = (X'X)^{-1} X' \varepsilon_{t:t+J} \qquad (A3)$$
$$\hat{b}_{2:J}^{NOL} - b_{2:J} = (X'X)^{-1} X' v_{t:t+J} \qquad (A4)$$

Decompose $\varepsilon_{t,t+J}$ into a function of $v_{t,t+J}$ and $\eta_{t,t+J}$, i.e.,

$$\varepsilon_{t,t+J} = \frac{cov\left(\varepsilon_{t,t+J}, v_{t,t+J}\right)}{var\left(v_{t,t+J}\right)} v_{t,t+J} + \eta_{t,t+J}$$

Rewrite $\hat{b}_{1:J} - b_{1:J}$ in terms of the above equation and taking expectations yields:

$$E\left[\hat{b}_{1:J}^{NOL} - b_{1:J}\right] = \frac{cov\left(\varepsilon_{t,t+J}, v_{t,t+J}\right)}{var\left(v_{t,t+J}\right)} \; E\left[\hat{b}_{2:J}^{NOL} - b_{2:J}\right]$$

Under the model in equation (1),

$$cov\left(\varepsilon_{t,t+J}, v_{t,t+J}\right) = \frac{1-\rho^J}{1-\rho} \sigma_{uv} + \frac{\rho - \rho^J}{(1-\rho^2)} \sigma_v^2$$
$$var\left(v_{t,t+J}\right) = \left(1 + \rho^2 + \rho^4 + \cdots + \rho^{2(J-1)}\right)\sigma_v^2 = \frac{1-\rho^{2J}}{1-\rho^2} \sigma_v^2$$

Substituting these covariances and variances into the above equation results in:

$$E\left[\hat{\beta}_J^{NOL} - b_{1:J}\right] = E\left[\hat{b}_{2:J}^{NOL} - b_{2:J}\right] \left[\frac{1+\rho}{1+\rho^J} \, \frac{\sigma_{uv}}{\sigma_v^2} + b_{1:J} \frac{\rho - \rho^J}{(1-\rho^{2J})} \right]$$

Applying the autocorrelation bias for a first order autocorrelation from Kendall (1954) and Marriott and Pope (1954), and again under the null of $\beta_1 = 0$,

$$E\left[\hat{\beta}_J^{NOL}\right] = -\frac{1+\rho}{1+\rho^J} \, \frac{1+3\rho^J}{T/J} \frac{\sigma_{uv}}{\sigma_v^2}$$

Proposition 2:

Based on equation (A1) in Proposition 1, we can run the regression of the *J*-period return $\sum_{i=1}^{J} R_{t+i}$ on X_t using overlapping data, that is, by sampling every period. In other words, we run the regressions using $X = \begin{pmatrix} 1 & X_t \end{pmatrix},\ t = 1, 2, \dots, T-J$:

$$\hat{b}_{1:J}^{OL} - b_{1:J} = (X'X)^{-1} X' \varepsilon_{t:t+J} \qquad (A5)$$

$$(\hat{b}_{2:j} - \rho\,\hat{b}_{2:j-1}) - (b_{2:j} - \rho b_{2:j-1}) = (X'X)^{-1} X' v_{t+j} \qquad (A6)$$

Note that (A6) therefore implies $\left(X'X\right)^{-1} X' \left(\sum_{i=1}^{J} v_{t+i} \right) = \left[\hat{b}_{2:J} - b_{2:J} \right] + (1-\rho) \sum_{i=1}^{J-1} \left[\hat{b}_{2:i} - b_{2:i} \right]$. We can then decompose $\varepsilon_{t,t+J}$ into a function of $\left(\sum_{i=1}^{J} v_{t+i} \right)$ and $\eta_{t,t+J}$:

$$\varepsilon_{t,t+J} = \frac{cov\left(\varepsilon_{t,t+J}, \left(\sum_{i=1}^{J} v_{t+i} \right) \right)}{Jvar(v_t)} \left(\sum_{i=1}^{J} v_{t+i} \right) + \eta_{t,t+J}\,,$$

with $\dfrac{\text{cov}\left(\varepsilon_{t,t+J}, \sum_{i=1}^{J} v_{t+i} \right)}{J\ \text{var}\left(v_t\right)} = \dfrac{\sigma_{uv}}{\sigma_v^2} + \beta_J \left[\dfrac{1}{1-\rho^J} - \dfrac{1-\rho^J}{J\left(1-\rho\right)} \right]$.

Rewriting $\hat{b}_{1:J} - b_{1:J}$ in terms of the above equation and taking expectations yields:

$$E\left[\hat{b}_{1:J}^{OL} - b_{1:J}\right] = \frac{cov\left(\varepsilon_{t,t+J}, \left(\sum_{i=1}^{J} v_{t+i} \right) \right)}{Jvar(v_t)} E\left[\left[\hat{b}_{2:J} - b_{2:J} \right] + (1-\rho) \sum_{i=1}^{J-1} \left[\hat{b}_{2:iOL} - b_{2:i} \right] \right]$$

Applying the *i*th order autocorrelation bias from Kendall (1954) and Marriott and Pope (1954), that is, $E\left[\hat{\rho}_i - \rho_i\right] = -\dfrac{1}{T}\left[(1+\rho)\dfrac{1-\rho^i}{1-\rho} + 2i\rho^i \right]$,

$$E\left[\hat{\beta}_J^{OL} - \beta_J^{OL}\right] = -\frac{1}{T}\left[J(1+\rho) + 2\rho\left(\frac{1-\rho^J}{1-\rho}\right) \right] \left[\frac{\sigma_{uv}}{\sigma_v^2} + \beta_J^{OL}\left(\frac{1}{1-\rho^J} - \frac{1}{J(1-\rho)} \right) \right]$$

Assuming the null of $\beta_1 = 0$, yields the desired result:

$$E\left[\hat{\beta}_J^{OL}\right] = -\frac{1}{T}\left[J(1+\rho) + 2\rho\left(\frac{1-\rho^J}{1-\rho}\right) \right] \frac{\sigma_{uv}}{\sigma_v^2}$$

Proposition 3:

From equations (1) to (3), the expected value of then alternative long horizon estimator $\hat{\gamma}_J$ can be written as $E\left[\hat{\gamma}_J\right] = E\left[\frac{\hat{\beta}_J}{V\hat{R}_J(X)}\right]$. Taking a second-order Taylor series expansion of the ratio yields:

$$E\left[\frac{\hat{\beta}_J}{V\hat{R}_J(X)}\right] = \frac{E\left[\hat{\beta}_J\right]}{E\left[V\hat{R}_J(X)\right]}\left(1 - \frac{\operatorname{cov}\left(\hat{\beta}_J, V\hat{R}_J(X)\right)}{E\left[\hat{\beta}_J\right]E\left[V\hat{R}_J(X)\right]} + \frac{\operatorname{var}\left(V\hat{R}_J(X)\right)}{E\left[V\hat{R}_J(X)\right]^2}\right)$$

Note that $E\left[\hat{\beta}_J\right]$ is given by Proposition 2 above. The estimator of the *J*-period variance-ratio of X_t can be written as $V\hat{R}_J(X) = J + 2\sum_{i=1}^{J-1}(J-i)\hat{\rho}_i$. Using the results of Marriott and Pope (1954), w can write the mean and variance of the autocorrelation estimator as:

$$E\left[\hat{\rho}_j\right] = \rho^j - \frac{1}{T}\left[(1+\rho)\frac{1-\rho^j}{1-\rho} + 2j\rho^j\right] \text{ and } \quad \operatorname{var}\left[\hat{\rho}_j\right] = \frac{1}{T}\left[\left(1+\rho^2\right)\frac{1-\rho^{2j}}{1-\rho^2} - 2j\rho^{2j}\right].$$

Note that the closed form $VR_J(X) = J + 2\sum_{i=1}^{J-1}(J-i)\rho^i = J - \frac{2p(1+J(-1+p)-p^J)}{(-1+p)^2}$. Thus, we can derive the expected value of the estimator of $VR_J(X)$:

$$EV\hat{R}_J(X) = J + 2\sum_{i=1}^{J-1}(J-i)\left[\rho^j - \frac{1}{T}\left[(1+\rho)\frac{1-\rho^j}{1-\rho} + 2j\rho^j\right]\right]$$

$$= J - \frac{2p(1+J(-1+p)-p^J)}{(-1+p)^2} + \frac{2}{T}\frac{J^2(-1+p)^2(1+p)+2p(1+p)(-1+p^J)+J(-1+p)(1-2p+p^2-4p^{1+J})}{2(-1+p)^3}$$

Note also we can write the covariance between $\hat{\beta}_J(X)$ and $V\hat{R}_J(X)$, as well as the variance of $V\hat{R}_J(X)$:

$$\operatorname{cov}\left(\hat{\beta}_J(X), V\hat{R}_J(X)\right) = \frac{\sigma_{uv}}{\sigma_v^2}\operatorname{cov}\left(\left[\rho_J + (1-\rho)\sum_{i=1}^{J-1}\hat{\rho}_i\right], \left[2\sum_{i=1}^{J-1}(J-i)\hat{\rho}_i\right]\right)$$

$$= \frac{\sigma_{uv}}{\sigma_v^2}\left[2\sum_{i=1}^{J-1}(J-i)\operatorname{cov}(\hat{\rho}_i, \hat{\rho}_J) + 2(1-\rho)\sum_{i=1}^{J-1}\sum_{k=1}^{J-1}(J-i)\operatorname{cov}\left(\hat{\rho}_i, \hat{\rho}_k\right)\right]$$

$$\operatorname{var}\left(V\hat{R}_J(X)\right) = \operatorname{var}\left(J + 2\sum_{i=1}^{J-1}(J-i)\hat{\rho}_i\right)$$

$$= 4\sum_{i=1}^{J-1}\sum_{k=1}^{J-1}(J-i)(J-k)\operatorname{cov}\left(\hat{\rho}_i, \hat{\rho}_k\right)$$

where the correlation between $\hat{\rho}_s$ and $\hat{\rho}_{s+t}$ is (e.g., Bartlett (1946))[8]:

$$\operatorname{cov}\left(r_s, r_{s+t}\right)=\frac{1}{T}\rho^t\left[\frac{\left(1+\rho^2\right)\left(1-\rho^{2s}\right)}{1-\rho^2}+t\right]-\frac{1}{T}(2s+t)\rho^{2s+t}$$

$$\Rightarrow \operatorname{var}\left(r_s\right)=\frac{1}{T}\left[\frac{\left(1+\rho^2\right)\left(1-\rho^{2s}\right)}{1-\rho^2}\right]-\frac{2}{T}s\rho^{2s}$$

Putting all of these results together yields the closed-form solution of Proposition 3.

Proposition 4:

Taking a 2nd order bivariate Taylor expansion of the long-horizon implied estimator from equation (1), i.e.,

$$\begin{aligned}
\hat{\beta}_{J:imp} &= f\left(\hat{\beta}_1, \hat{\rho}\right) \\
&= \hat{\beta}_1 \frac{1-\hat{\rho}^J}{1-\hat{\rho}} \\
&\approx f\left(\beta_1, \rho\right) + \tfrac{\partial f}{\partial \beta_1}\left(\hat{\beta}_1-\beta_1\right) + \tfrac{\partial f}{\partial \rho}\left(\hat{\rho}-\rho\right) \\
&\quad + \tfrac{1}{2}\left[\tfrac{\partial f}{\partial^2 \beta_1}\left(\hat{\beta}_1-\beta_1\right)^2 + \tfrac{\partial f}{\partial \beta_1 \rho}\left(\hat{\beta}_1-\beta_1\right)\left(\hat{\rho}-\rho\right) + \tfrac{\partial f}{\partial^2 \rho}\left(\hat{\rho}-\rho\right)^2\right] + \ldots
\end{aligned}$$

$$\begin{aligned}
E\left[\hat{\beta}_{J:imp}\right] &\approx 0 \text{ (under null)} + \frac{1-\rho^J}{1-\rho} E\left(\hat{\beta}_1-\beta_1\right) + 0 \text{ (under null)} + 0 \\
&\quad - \frac{1}{2}\left[\frac{J\rho^{J-1}}{1-\rho}-\frac{1-\rho^J}{(1-\rho)^2}\right] E\left[\left(\hat{\beta}_1-\beta_1\right)\left(\hat{\rho}-\rho\right)\right] + 0 \text{ (under null)} \\
&\approx \frac{1-\rho^J}{1-\rho}\left(-\frac{1+3\rho}{T}\right) - \frac{1}{2}\left[\frac{J\rho^{J-1}}{1-\rho}-\frac{1-\rho^J}{(1-\rho)^2}\right]\left(\frac{1-\rho^2}{T}\right)
\end{aligned}$$

where we can replace

$$\operatorname{var}(\hat{\rho})=\left(\frac{1-\rho^2}{T}\right)\approx\left(\frac{1-\rho^2}{T}\right)-\left(\frac{1-14\rho^2}{T^2}\right)+\left(\frac{5-78\rho^2+76\rho^4}{T^3(1-\rho^2)}\right)$$

[8] Note that Bartlett (1946) contains some errors which are corrected in an errata contained in the same journal in 1949. The formulas provided above are correct.

Proposition 5:

Consider the following extension to regression equation (1):

$$R_{t,t+1} = \alpha_1 + \beta_1 X_{t-k} + u_{t+1}^k$$
$$X_{t+1} = \varpi + \rho X_t + v_{t+1} \tag{A7}$$

Using similar logic to that of Proposition 2, we can run the regression of $R_{t,t+1}$ on X_{t-k}. Defining $X = \begin{pmatrix}1 & X_{t-k}\end{pmatrix}$, the regression coefficients for (A7) are:

$$\widehat{b}_1 - b_1 = (X'X)^{-1}X'u_{t+1}^k$$
$$(\widehat{b}_{2:k} - \rho\,\widehat{b}_{2:k-1}) - (b_{2:k} - \rho b_{2:k-1}) = (X'X)^{-1}X'v_{t+1}$$

Assuming no predictability (i.e., $u_{t+1}^k \equiv u_{t+1}$) and, as with Propositions 1 and 2, decomposing u_{t+1} into a function of v_{t+1} and η_{t+1}, we can rewrite $\widehat{b}_1 - b_1$ in terms of $(\widehat{b}_{2:k} - \rho\,\widehat{b}_{2:k-1}) - (b_{2:k} - \rho b_{2:k-1})$. Taking expectations, we derive an extension to Stambaugh (1989) related to equation (A7):

$$E\left[\hat{\beta}_1\right] = \frac{\sigma_{uv}}{\sigma_v^2} E\left[\ \rho_{k+1} - \rho_{k+1}) - \rho\left(\hat{\rho}_k - \rho_k\right)\right]$$
$$= -\frac{1}{T}\frac{\sigma_{uv}}{\sigma_v^2}\left(1 + \rho + 2\rho^{k+1}\right)$$

where the kth order autocorrelation bias is $E\left[\hat{\rho}_k - \rho_k\right] = -\frac{1}{T}\left[\left(1+\rho\right)\frac{1-\rho^k}{1-\rho} + 2k\rho^k\right]$ (e.g., see Kendall (1954) and Marriott and Pope (1954)). In terms of the analogous regression to equation (3), $R_{t,t+J} = \alpha_J + \beta_J X_{t-k} + \varepsilon_{t:t+J}^k$, summing up over the $J\,E\left[\hat{\beta}_j\right]s$ yields the desired result:

$$E\left[\hat{\beta}_J^{ol} - \beta_J\right] = -\frac{1}{T}\left[J(1+\rho) + 2\rho^{k+1}\left(\frac{1-\rho^J}{1-\rho}\right)\right]\frac{\sigma_{uv}}{\sigma_v^2}$$

References

Amihud, Yakov, and Clifford M. Hurvich. "Predictive regressions: A reduced-bias estimation method." *Journal of Financial and Quantitative Analysis* 39.4 (2004): 813-841.

Amihud, Yakov, Clifford M. Hurvich, and Yi Wang, 2008, "Multiple-predictor regressions: Hypothesis testing." *The Review of Financial Studies* 22, no. 1:13-434.

Amihud, Yakov, Clifford M. Hurvich, and Yi Wang, 2010, "Predictive regression with order-p autoregressive predictors." *Journal of Empirical Finance* 17, no. 3: 513-525.

Andrews, Donald WK. 1991. "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation." *Econometrica: Journal of the Econometric Society*, 817-858.

Andrews, Donald WK, and J. Christopher Monahan. 1992. "An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator." *Econometrica: Journal of the Econometric Society*, 953-966.

Ang, Andrew, and Geert Bekaert. 2007. "Stock return predictability: Is it there?." *Review of Financial studies* 20.3: 651-707.

Bekaert, Geert, Robert J. Hodrick, and David A. Marshall. 1997. "On biases in tests of the expectations hypothesis of the term structure of interest rates." *Journal of Financial Economics* 44.3: 309-348.

Boudouk, Jacob, and Matthew Richardson. "THE STATISTICS OF LONG-HORIZON REGRESSIONS REVISITED 1." *Mathematical Finance* 4.2 (1994): 103-119.

Boudoukh, Jacob, Matthew Richardson, and Robert F. Whitelaw. "The myth of long-horizon predictability." *The Review of Financial Studies* 21.4 (2008): 1577-1605.

Britten-Jones, Mark, Anthony Neuberger, and Ingmar Nolte. 2011. "Improved inference in regression with overlapping observations." *Journal of Business Finance & Accounting* 38.5-6: 657-683.

Campbell, John Y, 1991, "A variance decomposition for stock returns." *The economic journal* 101, no. 405: 157-179.

Campbell, John Y., Anrew Lo, and Craig MacKinlay, 1997, *The Econometrics of Financial Markets*, Princeton: Princeton University Press.

Campbell, John Y., and Motohiro Yogo. 2006 "Efficient tests of stock return predictability." *Journal of financial economics* 81, no. 1: 27-60.

Chen, Yu-chin, and Kwok Ping Tsang, 2013. "What does the yield curve tell us about exchange rate predictability?." *Review of Economics and Statistics* 95.1: 185-205.

Cochrane, John H. 2011. "Presidential address: Discount rates." *The Journal of finance* 66, no. 4: 1047-1108.

De Gooijer, Jan G. 1980 "Exact moments of the sample autocorrelations from series generated by general ARIMA processes of order (p, d, q), d= 0 or 1." *Journal of Econometrics* 14, no. 3: 365-379.

Elliott, Graham, and James H. Stock. 1994. "Inference in time series regression when the order of integration of a regressor is unknown." *Econometric theory* 10, no. 3-4: 672-700.

Fama, Eugene F., and Kenneth R. French. 1988. "Dividend yields and expected stock returns." *Journal of financial economics*22, no. 1: 3-25.

Goetzmann, William N., and Philippe Jorion. "Testing the predictive power of dividend yields." *The Journal of Finance* 48.2 (1993): 663-679.

Hansen, Lars Peter, and Robert J. Hodrick. "Forward exchange rates as optimal predictors of future spot rates: An econometric analysis." *Journal of political economy* 88.5 (1980): 829-853.

He, Wen, and Maggie Rong Hu, 2014, "Aggregate earnings and market returns: International evidence." *Journal of Financial and Quantitative Analysis* 49, no. 4: 879-901.

Hjalmarsson, Erik. 2011. "New methods for inference in long-horizon regressions." *Journal of Financial and Quantitative Analysis* 46.03: 815-839.

Hodrick, Robert J. "Dividend yields and expected stock returns: Alternative procedures for inference and measurement." *The Review of Financial Studies* 5.3 (1992): 357-386.

Kendall, Maurice G. "Note on bias in the estimation of autocorrelation." *Biometrika* 41.3-4 (1954): 403-404.

Nelson, Charles R., and Myung J. Kim. "Predictable stock returns: The role of small sample bias." *The Journal of Finance* 48.2 (1993): 641-661.

Kiviet, Jan F., and Garry DA Phillips. 2012. "Higher-order asymptotic expansions of the least-squares estimation bias in first-order dynamic regression models." *Computational Statistics & Data Analysis* 56, no. 11: 3705-3729.

Kothari, S. P., Jonathan Lewellen, and Jerold B. Warner, 2006, "Stock returns, aggregate earnings surprises, and behavioral finance." *Journal of Financial Economics* 79, no. 3 (2006): 537-568.

MacKinnon, James G., and Anthony A. Smith Jr. 1998 "Approximate bias correction in econometrics." *Journal of Econometrics* 85, no. 2: 205-230.

Marriott, F. H. C., and J. A. Pope. 1954. "Bias in the estimation of autocorrelations." *Biometrika* 41, no. 3/4 (1954): 390-402.

Nankervis, John C., and N. Eugene Savin. 1988. "The exact moments of the least-squares estimator for the autoregressive model corrections and extensions." *Journal of Econometrics*37, no. 3: 381-388.

Newey, Whitney K., and Kenneth D.West, 1987, A simple, positive definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703-708.

Newey, Whitney K., and Kenneth D.West, 1994, Automatic lag Selection in Covariance Matrix Estimation, A *Review of Economic Studies*, vol. 61, issue 4, 631-653

Richardson, Matthew, and James H. Stock. "Drawing inferences from statistics based on multiyear asset returns." *Journal of Financial Economics* 25.2 (1989): 323-348.

Sadka, Gil, and Ronnie Sadka, 2009, "Predictability and the earnings–returns relation." *Journal of Financial Economics* 94, no. 1: 87-106.

Sawa, Takamitsu. 1978. "The exact moments of the least squares estimator for the autoregressive model." *Journal of Econometrics* 8, no. 2: 159-172.

Shaman, Paul, and Robert A. Stine. 1988. "The bias of autoregressive coefficient estimators." *Journal of the American Statistical Association* 83, no. 403: 842-848.

Stambaugh, Robert F. "Predictive regressions." *Journal of Financial Economics* 54.3 (1999): 375-421.

Torous, Walter, Rossen Valkanov, and Shu Yan. 2004. "On predicting stock returns with nearly integrated explanatory variables." *The Journal of Business* 77, no. 4: 937-966.

Valkanov, Rossen. "Long-horizon regressions: theoretical results and applications." *Journal of Financial Economics* 68.2 (2003): 201-232.

Welch, Ivo, and Amit Goyal. 2007. "A comprehensive look at the empirical performance of equity premium prediction." *The Review of Financial Studies* 21, no. 4: 1455-1508.

Table 1: Analytical vs. Simulated Bias for Long-Horizon Regression Estimators

Table 1 presents analytical betas and simulation-based mean betas under the null. Sample size is *T*=300, 600 and 1200. AB_J^{nol} denotes the analytical *nol* bias $\frac{J}{T}\frac{(1+\rho)(1+3\rho^J)}{1+\rho^J}\frac{\sigma_{uv}}{\sigma_v^2}$ (Proposition 1) while SB_J^{nol} is its simulated counterpart using 100,000 simulations. The corresponding *ol* counterparts are AB_J^{ol} ***for*** the analytical bias using overlapping data $\frac{1}{T}\left[J(1+\rho)+2\rho\left(\frac{1-\rho^J}{1-\rho}\right)\right]\frac{\sigma_{uv}}{\sigma_v^2}$ (Proposition 2) and SB_J^{ol} for the mean *ol* simulated beta. Other simulation parameters are $\sigma_u=\sigma_v=1$, $\rho_{uv}=-0.9$. All numbers are multiplied by 100.

		J=12				*J*=36				*J*=60			
ρ	T	AB_J^{nol}	SB_J^{nol}	AB_J^{ol}	SB_J^{ol}	AB_J^{nol}	SB_J^{nol}	AB_J^{ol}	SB_J^{ol}	AB_J^{nol}	SB_J^{nol}	AB_J^{ol}	SB_J^{ol}
0.70	300	6.29	6.44	7.50	7.62	18.36	18.85	19.76	21.10	30.60	30.32	32.00	35.05
0.70	600	3.14	3.04	3.75	3.72	9.18	9.45	9.88	10.17	15.30	13.86	16.00	16.73
0.70	1200	1.57	1.59	1.88	1.96	4.59	4.82	4.94	5.09	7.65	7.28	8.00	8.19
0.90	300	9.85	9.50	10.72	10.69	21.42	22.01	25.80	26.46	34.32	33.83	39.59	41.76
0.90	600	4.93	4.83	5.36	5.33	10.71	10.88	12.90	13.10	17.16	16.92	19.80	20.52
0.90	1200	2.46	2.52	2.68	2.72	5.36	5.46	6.45	6.57	8.58	8.68	9.90	10.19
0.95	300	11.95	12.00	12.26	12.72	26.80	26.47	30.66	31.05	38.19	36.42	45.98	47.12
0.95	600	5.97	6.01	6.13	6.27	13.40	13.48	15.33	15.51	19.10	18.76	22.99	23.42
0.95	1200	2.99	3.03	3.07	3.10	6.70	6.72	7.67	7.67	9.55	9.74	11.49	11.57
0.99	300	13.90	15.12	13.91	15.69	39.14	37.73	39.53	40.29	61.16	50.44	62.72	58.62
0.99	600	6.95	7.54	6.96	7.69	19.57	20.35	19.76	20.88	30.58	28.79	31.36	32.01
0.99	1200	3.47	3.65	3.48	3.69	9.78	10.01	9.88	10.28	15.29	14.99	15.68	16.10

Table 2: Analytical vs. Simulated Bias for Long-Horizon Regression Estimators Under the Alternative

Table 2 presents analytical and simulation-based bias adjustment comparison under the alternative. The columns record the ***'R^2'*** in percent terms (for example 0.50 corresponds to a simulated ½% per month) used for calibrating β_1, the single period beta, where parentheses '' are used since the finite sample R^2 , denoted ***ER^2***, is higher due to the finte sample bias of ***R^2***. The serial correlation of X_t is ρ , and T is the simulated sample length. The "true" *J*-period beta given nonzero β_1, denoted ***$Tru\beta_J^{olp}$***, is compared with its finite sample biased mean simulated counterpart, ***$Sim\beta_J^{olp}$***. The difference, = ***Diff= $Sim\beta_J^{olp}$-$Tru\beta_J^{olp}$*** is the finite sample bias, to be compared with ***Bias*** calculated using Proposition 2. The first block is a set of simulations under the null of no predictability, then each block uses a larger R^2, all else equal, hence increasing the degree of predictability. Other simulation parameters are $\sigma_u=\sigma_v=1$, $\rho_{uv}=-0.9$. All numbers are multiplied by 100.

				J=12					*J*=36					*J*=60				
'R^2'	β_1	ρ	***T***	***$Tru\beta_J^{olp}$***	***$Sim\beta_J^{olp}$***	***Diff***	***Bias***	***ER^2***	***$Tru\beta_J^{olp}$***	***$Sim\beta_J^{olp}$***	***Diff***	***Bias***	***ER^2***	***$Tru\beta_J^{olp}$***	***$Sim\beta_J^{olp}$***	***Diff***	***Bias***	***ER^2***
0.00	0.00	0.70	300	0.00	7.61	7.61	7.63	5.51	0.00	20.71	20.71	21.96	13.95	0.00	34.78	34.78	39.12	22.21
0.00	0.00	0.70	600	0.00	3.74	3.74	3.78	2.75	0.00	9.99	9.99	10.39	6.97	0.00	16.46	16.46	17.57	11.08
0.00	0.00	0.70	1200	0.00	1.83	1.83	1.88	1.37	0.00	4.98	4.98	5.07	3.45	0.00	8.21	8.21	8.38	5.49
0.00	0.00	0.90	300	0.00	10.80	10.80	10.47	6.86	0.00	26.32	26.32	27.31	16.95	0.00	41.91	41.91	46.20	25.92
0.00	0.00	0.90	600	0.00	5.42	5.42	5.29	3.43	0.00	13.12	13.12	13.23	8.48	0.00	20.47	20.47	21.22	12.88
0.00	0.00	0.90	1200	0.00	2.76	2.76	2.66	1.70	0.00	6.65	6.65	6.53	4.20	0.00	10.20	10.20	10.24	6.39
0.00	0.00	0.95	300	0.00	12.52	12.52	11.69	7.69	0.00	30.95	30.95	30.50	19.55	0.00	46.87	46.87	50.14	29.58
0.00	0.00	0.95	600	0.00	6.22	6.22	5.98	3.75	0.00	15.41	15.41	15.23	9.67	0.00	23.31	23.31	23.81	14.70
0.00	0.00	0.95	1200	0.00	3.10	3.10	3.03	1.85	0.00	7.72	7.72	7.63	4.82	0.00	11.60	11.60	11.67	7.32
0.00	0.00	0.99	300	0.00	15.62	15.62	12.77	10.01	0.00	40.10	40.10	33.71	26.46	0.00	58.43	58.43	53.52	39.81
0.00	0.00	0.99	600	0.00	7.72	7.72	6.66	4.65	0.00	21.05	21.05	18.08	12.89	0.00	32.39	32.39	28.31	20.14
0.00	0.00	0.99	1200	0.00	3.75	3.75	3.41	2.18	0.00	10.42	10.42	9.44	6.18	0.00	16.26	16.26	14.83	9.80
0.25	3.58	0.70	300	11.75	18.61	6.86	6.90	5.86	11.92	30.26	18.34	19.33	14.18	11.92	42.23	30.31	34.24	22.43
0.25	3.58	0.70	600	11.75	15.12	3.37	3.42	3.04	11.92	20.77	8.85	9.15	7.07	11.92	26.55	14.63	15.38	11.09
0.25	3.58	0.70	1200	11.75	13.47	1.72	1.70	1.65	11.92	16.25	4.33	4.46	3.57	11.92	18.98	7.06	7.33	5.55
0.25	2.18	0.90	300	15.66	25.38	9.72	9.43	8.19	21.33	42.88	21.55	22.47	18.11	21.78	54.68	32.90	36.84	26.82
0.25	2.18	0.90	600	15.66	20.44	4.78	4.77	4.67	21.33	31.95	10.62	10.90	9.49	21.78	38.03	16.25	16.96	13.71
0.25	2.18	0.90	1200	15.66	18.05	2.39	2.40	2.96	21.33	26.58	5.25	5.38	5.22	21.78	29.70	7.92	8.18	7.16
0.25	1.56	0.95	300	14.37	25.78	11.41	10.73	9.55	26.33	50.82	24.48	24.82	21.91	29.82	64.60	34.78	38.23	31.58
0.25	1.56	0.95	600	14.37	20.08	5.71	5.49	5.70	26.33	38.65	12.32	12.40	12.28	29.82	47.17	17.34	18.15	16.99
0.25	1.56	0.95	1200	14.37	17.24	2.87	2.78	3.82	26.33	32.60	6.27	6.21	7.44	29.82	38.65	8.83	8.90	9.62
0.25	0.71	0.99	300	8.02	23.02	14.99	12.21	12.49	21.44	56.18	34.74	29.28	31.31	31.98	77.43	45.44	42.46	44.70
0.25	0.71	0.99	600	8.02	15.42	7.39	6.38	7.30	21.44	39.65	18.21	15.79	18.95	31.98	57.25	25.27	22.67	27.77
0.25	0.71	0.99	1200	8.02	11.59	3.57	3.26	4.90	21.44	30.46	9.02	8.27	12.84	31.98	44.84	12.86	11.93	18.89

0.50	5.06	0.70	300	16.64	23.20	6.56	6.59	6.19	16.87	34.14	17.26	18.23	14.36	16.87	45.30	28.43	32.19	22.56
0.50	5.06	0.70	600	16.64	19.82	3.18	3.27	3.37	16.87	25.22	8.34	8.63	7.24	16.87	30.47	13.60	14.47	11.24
0.50	5.06	0.70	1200	16.64	18.25	1.60	1.63	1.96	16.87	21.05	4.18	4.20	3.71	16.87	23.60	6.73	6.89	5.66
0.50	3.09	0.90	300	22.17	31.39	9.22	9.01	9.45	30.20	49.59	19.39	20.48	19.08	30.84	60.25	29.41	32.99	27.52
0.50	3.09	0.90	600	22.17	26.77	4.60	4.56	6.01	30.20	39.84	9.63	9.93	10.56	30.84	45.27	14.43	15.17	14.45
0.50	3.09	0.90	1200	22.17	24.53	2.36	2.29	4.34	30.20	34.99	4.79	4.89	6.28	30.84	37.95	7.11	7.30	7.90
0.50	2.21	0.95	300	20.35	31.39	11.04	10.31	11.43	37.28	59.28	22.00	22.39	24.10	42.23	71.82	29.59	33.16	33.00
0.50	2.21	0.95	600	20.35	25.84	5.49	5.29	7.69	37.28	48.41	11.12	11.22	14.95	42.23	57.24	15.01	15.79	19.22
0.50	2.21	0.95	1200	20.35	23.04	2.69	2.68	5.86	37.28	42.72	5.43	5.64	10.33	42.23	49.63	7.40	7.76	12.20
0.50	1.00	0.99	300	11.36	26.02	14.65	11.98	14.88	30.36	62.70	32.34	27.48	35.94	45.28	85.03	39.75	37.96	49.41
0.50	1.00	0.99	600	11.36	18.58	7.22	6.27	9.93	30.36	47.34	16.98	14.85	24.98	45.28	67.63	22.35	20.34	35.41
0.50	1.00	0.99	1200	11.36	14.85	3.49	3.20	7.63	30.36	38.81	8.45	7.78	19.57	45.28	56.73	11.45	10.73	28.13
0.75	6.21	0.70	300	20.41	26.57	6.17	6.36	6.51	20.69	37.08	16.38	17.39	14.55	20.69	47.71	27.01	30.63	22.72
0.75	6.21	0.70	600	20.41	23.51	3.11	3.15	3.71	20.69	28.68	7.99	8.23	7.42	20.69	33.70	13.01	13.76	11.38
0.75	6.21	0.70	1200	20.41	21.94	1.54	1.57	2.30	20.69	24.61	3.92	4.01	3.85	20.69	27.06	6.37	6.56	5.76
0.75	3.79	0.90	300	27.19	35.90	8.71	8.69	10.72	37.04	54.55	17.52	18.95	20.01	37.82	64.00	26.18	30.03	28.09
0.75	3.79	0.90	600	27.19	31.59	4.40	4.40	7.41	37.04	45.86	8.82	9.19	11.70	37.82	50.79	12.96	13.81	15.26
0.75	3.79	0.90	1200	27.19	29.36	2.17	2.21	5.75	37.04	41.39	4.36	4.53	7.50	37.82	44.25	6.42	6.65	8.78
0.75	2.71	0.95	300	24.95	35.71	10.76	10.00	13.38	45.72	65.65	19.93	20.56	26.47	51.79	77.23	25.45	29.34	34.68
0.75	2.71	0.95	600	24.95	30.25	5.30	5.13	9.75	45.72	55.82	10.09	10.31	17.88	51.79	64.75	12.97	13.98	21.75
0.75	2.71	0.95	1200	24.95	27.57	2.62	2.60	7.99	45.72	50.72	5.00	5.17	13.54	51.79	58.29	6.51	6.86	15.15
0.75	1.23	0.99	300	13.93	28.36	14.43	11.81	17.34	37.23	67.82	30.59	26.09	40.50	55.53	90.95	35.42	34.46	53.91
0.75	1.23	0.99	600	13.93	21.04	7.11	6.18	12.52	37.23	53.30	16.07	14.12	30.73	55.53	75.64	20.11	18.54	42.46
0.75	1.23	0.99	1200	13.93	17.36	3.43	3.16	10.32	37.23	45.24	8.01	7.40	25.99	55.53	65.88	10.35	9.80	36.62

Table 3: Empirical Application of Long-Horizon Stock Return Predictability

This table presents equity excess return forecast regressions using overlapping monthly observations for horizons of one month and one, three and five years. Data is from Amit Goyal's website. The Sample period is January 1968 to December 2017, for a total of 600 monthly observations. We use six common predictors: (i) dividend price ratio: **DP**=$D_{t-12,t}/P_t$, (ii) dividend yield: **DY**=$D_{t-12,t}/P_{t-12}$, (iii) earnings to price ratio: **EP**=$E_{t-12,t}/P_t$, (iv) Shiller's CAPE, the cyclically-adjusted PE: **CapeN**=$(\Sigma_j log(E12_{t-j})/log(P_t))/120$ averaging over nominal earnings, (v) Real CAPE: **CapeR**=$(\Sigma_j log(e12_{t-j}\, \pi_{t-j,t}) /log(P_t))/120$ where earnings in year *t-j* are adjusted by inflation from *t-j* to *t*, and (vi) book to market ratio: **B/M**=$Book_t/Mkt_t$.

The regression estimate β_J^{ol} is the OLS estimate using overlapping data and no bias adjustment, ***t-AR1*** is the t-statistic using analytical AR1 standard errors (***seAR1***$=T^{-1/2}(1-\rho^2)^{1/2}\,\sigma_u/\sigma_v\{j+[2\rho/(1-\rho)][j-1-\rho((1-\rho^{j-1})/(1-\rho))]\}^{1/2}$), $\boldsymbol{\beta_J^{Adj}}$ adjusts the OLS beta using the analytical overlapping bias under the null bias, $E[\beta_J^{ol}]=(\rho_{uv}\sigma_u/\sigma_v)\,[j(1+\rho)+2\rho(1-\rho^j)/(1-\rho)]/T$), and ***SimPval*** is the simulated probability value of the empirical β_J^{ol} over the distribution of simulated $\beta_J^{ol}s$ under the null using the estimated adjusted parameters. The $\boldsymbol{\beta_J^{Adj12}}$ adjusts not only for the contemporaneous ρ_{uv} but also for the $\rho_{uv}(i)$'s up to lag 12 using simulations to match the actual parameters, with the following specification $v_{t+1}=\sum_{i=0}^{k}\rho_{uv}(i)\,u_{t+1-i}+\varepsilon_{t+1}$ where $\sigma_\varepsilon^2=(1-\sum_{i=0}^{k}\rho_{uv}(i)^2)\sigma_v^2$.

(i) **DP**

J	β_J	*t-AR1*	β_J^{Adj}	*t-AR1*	*SimPval*	β_J^{Adj12}	*t-AR1*	*SimPval12*
1	0.005	*1.329*	-0.001	*-0.280*	*0.379*	-0.003	*-0.785*	*0.392*
12	0.074	*1.544*	-0.004	*-0.081*	*0.436*	-0.024	*-0.508*	*0.449*
36	0.203	*1.444*	-0.030	*-0.213*	*0.401*	-0.069	*-0.493*	*0.411*
60	0.354	*1.541*	-0.034	*-0.147*	*0.422*	-0.066	*-0.287*	*0.439*

(ii) **DY**

J	β_J	*t-AR1*	β_J^{Adj}	*t-AR1*	*SimPval*	β_J^{Adj12}	*t-AR1*	*SimPval12*
1	0.006	*1.574*	0.006	*1.491*	*0.881*	-0.002	*-0.462*	*0.454*
12	0.069	*1.404*	0.064	*1.320*	*0.861*	-0.028	*-0.580*	*0.401*
36	0.185	*1.289*	0.173	*1.204*	*0.839*	-0.083	*-0.579*	*0.359*
60	0.311	*1.325*	0.291	*1.238*	*0.848*	-0.101	*-0.430*	*0.370*

(iii) **EP**

J	β_J	*t-AR1*	β_J^{Adj}	*t-AR1*	*SimPval*	β_J^{Adj12}	*t-AR1*	*SimPval12*
1	0.004	*1.003*	0.001	*0.330*	*0.633*	0.004	*1.051*	*0.807*
12	0.044	*0.990*	0.014	*0.310*	*0.629*	0.046	*1.040*	*0.804*
36	0.107	*0.834*	0.018	*0.140*	*0.566*	0.113	*0.885*	*0.766*
60	0.140	*0.679*	-0.006	*-0.027*	*0.503*	0.151	*0.732*	*0.724*

(iv) **CapeR**

J	β_J	*t-AR1*	β_J^{Adj}	*t-AR1*	*SimPval*	β_J^{Adj12}	*t-AR1*	*SimPval12*
1	0.096	*1.329*	0.001	*0.017*	*0.468*	-0.026	*-0.355*	*0.490*
12	1.274	*1.488*	0.136	*0.159*	*0.513*	-0.130	*-0.152*	*0.539*
36	2.932	*1.172*	-0.479	*-0.191*	*0.401*	-0.955	*-0.382*	*0.423*
60	5.487	*1.349*	-0.195	*-0.048*	*0.464*	-0.500	*-0.123*	*0.488*

(v) **CapeN**

J	β_J	*t-AR1*	β_J^{Adj}	*t-AR1*	*SimPval*	β_J^{Adj12}	*t-AR1*	*SimPval12*
1	0.151	*1.576*	0.039	*0.405*	*0.613*	0.013	*0.131*	*0.630*
12	1.951	*1.731*	0.616	*0.546*	*0.655*	0.359	*0.319*	*0.674*
36	4.465	*1.366*	0.506	*0.155*	*0.546*	0.054	*0.017*	*0.562*
60	8.004	*1.517*	1.480	*0.280*	*0.593*	1.207	*0.229*	*0.619*

(vi) **B/M**

J	β_J	*t-AR1*	β_J^{Adj}	*t-AR1*	*SimPval*	β_J^{Adj12}	*t-AR1*	*SimPval12*
1	0.003	*0.484*	-0.005	*-0.753*	*0.242*	-0.007	*-1.200*	*0.245*
12	0.056	*0.789*	-0.033	*-0.459*	*0.334*	-0.060	*-0.846*	*0.334*
36	0.096	*0.455*	-0.171	*-0.815*	*0.235*	-0.227	*-1.082*	*0.237*
60	0.233	*0.679*	-0.211	*-0.614*	*0.302*	-0.263	*-0.767*	*0.303*