IS STOCK RETURN PREDICTABILITY SPURIOUS?

Wayne E. Ferson[a,*], *Sergei Sarkissian*[b], *and Timothy Simin*[c]

Two problems, spurious regression bias and naïve data mining, conspire to mislead analysts about predictive models for stock returns. This article demonstrates the two problems, how they interact, and makes suggestions for what to do about it.

The interest in predicting stock returns is probably as old as the markets themselves. While there are many approaches to predicting returns, most rely on the relation of future returns to lagged variables, such as past returns, interest rates, payout-to-price ratios such as dividend yield, book-to-market, and yield spreads such as between low-grade and high-grade bonds or between long- and short-term bonds. Many important applications use predictive relations, including tactical asset allocation, active portfolio management, conditional performance evaluation, and market timing, to name a few.

Suppose that returns are generated by some model:

$$R_{t+1} = a_1 + b_1 Z_t^* + v_{t+1}, \qquad (1)$$

[a]Boston College, MA, USA.

[b]Faculty of Management, McGill University, Montreal, PQ, Canada.

[c]Department of Finance, Smeal College of Business, Pennsylvania State University, PA, USA.

[*]Corresponding author. Boston College, 140 Commonwealth Ave., Chestnut Hill, MA 02467, USA. Tel.: (617) 552-6431; e-mail: wayne.ferson@bc.edu

where Z_t^* has autocorrelation ρ^*, but instead we observe Z_t and estimate

$$R_{t+1} = a_2 + b_2 Z_t + u_{t+1}, \qquad (2)$$

and Z_t has autocorrelation, ρ. Because of Eq. (1) the *ex ante* expected return is autocorrelated and this means that Eq. (2) will show a statistically significant slope coefficient when this conclusion is uncalled for. If the researcher searches for predictability among a large set of variables, he is likely to find some, which, due to high autocorrelation, show predictability where none exists. Data mining becomes a more serious problem in the presence of autocorrelated independent variables.

In this paper, after examining the theory stated above in more detail, we examine a number of past studies and show that many of the results could be arrived at by a combination of data mining and spurious regression. We examine the interaction of these two problems. Spurious regression was studied by Yule (1926) and Granger and Newbold (1974) for economic data. These authors warned that if two variables were highly "persistent" over time, a regression of one on the other will likely produce a "significant" slope coefficient, evaluated by the

usual t-statistics, even if the variables are, in fact unrelated. Persistent variables are those that have large autocorrelations. Stock returns are not highly autocorrelated, so you might think that spurious regression would not be an issue for stock returns. However, think of a stock return as equal to the *ex ante* expected return plus the unexpected return. We show that, if the *expected* return is persistent, there is a risk of spurious regression.

The second issue is "naïve data mining." With so many analysts trying to build predictive models, and so many models back-tested on the same historical data, about 5% of the models will show "significant" predictive ability. If an analyst learns about one of the 5% models that appears to work, the chances are that he or she will "confirm" its performance when similar historical data are used. The problem, of course, is that the model will not work with fresh data, so it will be of no practical benefit.

By "naïve data mining," we refer to an analysis where a number of predictor variables or models are examined but the statistics do not correctly account for the number examined. Not all data mining is naïve. In fact, increasing computing power and data availability have allowed the development of some very sophisticated data mining (see Hastie *et al.* (2001) for the statistical foundations).

The main point of this article is that spurious regression and data mining interact in a most pernicious way. If analysts search for data that produces "significant" predictive regressions, they are more likely to find the spurious, persistent regressors. Data mining makes spurious regression more of a problem, and the possibility of spurious regression makes data mining seem more effective in finding predictive variables. Our simulations imply that virtually all of the predictive regressions in a set of prominent academic studies are consistent with a "spurious mining process." However, we can also reject the hypothesis that *ex ante* expected returns are constant over time.

Our results create an apparent conundrum. On the one hand, if the variables identified by previous studies of predictability are spurious, they will not work with fresh data in the future. On the other hand, if expected returns vary over time, there is an incentive to build predictive models for stock returns. The last section of this paper evaluates some suggestions for resolving these issues.

1 Data

Table 1 surveys nine of the major academic studies that propose lagged variables for predicting stock returns. We study monthly data, covering various subperiods of 1926 through 1998. We attempt to replicate the data series that were used in the original studies as closely as possible. The summary statistics are from our data. Note that the first-order autocorrelations of the lagged variables are high, suggesting a high degree of persistence. For example, the short term Treasury bill yields, monthly book-to-market ratios, the dividend yield of the S&P 500 and some of the yield spreads have sample autocorrelations of 0.97 or higher. These high autocorrelations are a red flag for a spurious regression problem.

Table 1 also summarizes regressions for the monthly return of the S&P 500 stock index, measured in excess of the one-month Treasury bill return from Ibbotson Associates, on the lagged variables taken one at a time. We report the slope coefficients and their t-ratios. Eight of the 13 t-ratios are larger than two.[1]

To incorporate data mining in our analysis we compile a randomly selected sample of 500 potential lagged variables through which our simulated analyst sifts, to mine the data for predictor variables. We select the 500 series randomly from a much larger sample of 10,866 potential variables. The specifics are described in the Appendix. The degree of persistence in the variables is crucial. The mean

Table 1 Common predictor variables summary statistics and OLS regression results: This table summarizes variables used in the literature to predict stock returns. The first column indicates the published study. The second column denotes the lagged variable. The next two columns give the sample (Period) and the number of observations (Obs) on the stock returns. Subsequent columns report the autocorrelation (ρ_z) and the standard deviation (σ_z) of the lagged variable. The next columns report regression results for Standard & Poors 500 excess return on a lagged variable. The slope coefficient is β and the t-statistic is t. The method of Newey and West (1987) is used. The abbreviations in the table are as follows. TB1y is the yield on the one-month Treasury bill. Two–one, Six–one, and Lag(two)–one are computed as the spreads on the returns of the two and one-month bills, six and one-month bills, and the lagged value of the two-month and current one-month bill. AAAy is the yield on the AAA rated corporate bonds. UBAAy is the yield on corporate bonds with a below BAA rating. "Cay" is the linear function of consumption, asset wealth, and labor income. DJBM and SPDM are the book-to-market ratios for the Dow Jones Industrial Average and the S&P 500 respectively. ALLy denotes the yield on all corporate bonds.

Reference	Predictor	Period	Obs	ρ_z	σ_z	β	t
Breen *et al.* (1989)	TB1y	5404–8612	393	0.97	0.0026	−2.49	−3.58
Campbell (1987)	Two–one	5906–7908	264	0.32	0.0006	11.87	2.38
	Six–one	5906–7908	264	0.15	0.0020	2.88	2.13
	Lag(two)–one	5906–7908	264	0.08	0.0010	9.88	2.67
Fama (1990)	ALLy–AAAy	5301–8712	420	0.97	0.0040	0.88	1.46
Fama and French (1988a,b)	Dividend yield	2701–8612	720	0.97	0.0013	0.40	1.36
Fama and French (1989)	AAAy–TB1y	2601–8612	732	0.92	0.0011	0.51	2.16
Keim and Stambaugh (1986)	UBAAy	2802–7812	611	0.95	0.0230	1.50	0.75
	UBAAy–TB1y	2802–7812	611	0.97	0.0320	1.57	1.48
Kothari and Shanken (1997)	DJBM	1927–1992	66	0.66	0.2270	0.28	2.63
Lettau and Ludvigson (2000)	"Cay"	52Q4–98Q4	184	0.79	0.0110	1.57	2.58
Pontiff and Schall (1998)	DJBM	2602–9409	824	0.97	0.2300	2.96	2.16
	SPBM	5104–9409	552	0.98	0.0230	9.32	1.03

autocorrelation of our 500 series is 15% and the median is 2%. If the variables from the academic literature, summarized in Table 1, arise from a spurious mining process, our analysis suggests they are likely to be highly autocorrelated. Eleven of the 13 sample autocorrelations in Table 1 are higher than 15%, and the median value is 95%. On the other hand, if the variables in the literature are a realistic representation of expected stock returns, the autocorrelations in Table 1 may be a good proxy for the true persistence of expected stock returns. We consider a range of values for the autocorrelations, based on these figures.

2 The models

Consider a situation in which an analyst runs a time-series regression like Eq. (2), but the "true" unobserved regression (1) uses a variable, Z^*, that the analyst cannot observe. Both Z and Z^* are assumed to be autocorrelated over time, and we denote their autocorrelations by ρ and ρ^*, respectively. While the stock return could be predicted if Z^* could be observed, the analyst can only use Z. Since Z and Z^* are independent, the true value of b_2 in the regression (2) is zero. Spurious regression occurs if the analyst finds a significant coefficient.

If the analyst uses a t-ratio, the spurious regression problem could come from the numerator or the denominator of the t-ratio: The coefficient or its standard error may be biased. We find that the problem lies with the standard error.[2] The reason is fairly simple to understand. When the hypothesis that the regression slope $b_2 = 0$ is true, the error term of the regression Eq. (2) inherits autocorrelation directly from the dependent variable. The slope coefficient would be correct if you had an infinite sample size, because regression slopes are known to be "consistent" in the presence of auto-correlated residuals. However, standard errors that do not account for the serial dependence correctly are biased, no matter how large the sample is.

To capture the interaction between spurious regression and data mining, we model an example where the analyst searches over L variables for the "best" predictor, based on the t-ratios and R-squares in the regressions. The L variables have autocorrelations, denoted by $\rho_1, \ldots, \rho_L$. We set these equal to the sample autocorrelations of L randomly selected variables in our data set of 500 potential predictor variables.[3] In our model, the analyst naively mines, or searches through the data for significant predictive relations. Since there none by design, he or she finds them either by luck or because they are spurious. The spurious regressors are more likely to be discovered. This is the essence of the spurious data mining process. When there are highly persistent variables included in the data mining process, the process is more likely to turn up "significant" predictive relations. At the same time, the predictor variables discovered by a spurious mining processes are likely to be more persistent than a randomly chosen variable.

3   Results

Table 2 summarizes the results for the case of pure spurious regression. Here, the analyst does

Table 2 Simulation results for spurious regression.

Obs	ρ	Critical t-statistic
393	0.97	3.25
264	0.32	2.06
264	0.15	2.06
264	0.08	2.03
420	0.97	3.27
720	0.97	3.20
732	0.92	2.40
611	0.95	2.88
611	0.97	3.25
66	0.66	2.42
184	0.79	2.27
824	0.97	3.16
552	0.98	3.68

not search for the best variables to predict returns, but uses a single randomly selected predictor. The unobserved *ex ante* expected return has the autocorrelation, ρ^*, taken from the studies in Table 1. The *Critical t-statistic* in Table 2 is the value from 10,000 simulated trials, so that 2.5% of the t-statistics lie above these values. This corresponds to a 5%, two-tailed test of the hypothesis that the slope is zero.

Comparing the t-ratios in Table 1 with their critical values in Table 2 shows that six of the eight regressions that appeared significant in Table 1 remain so, when you allow for the possibility of spurious regression but do not consider data mining. The simulations do cause us to question the significance of the term spread in Fama and French (1989) and the book-to-market ratio, studied by Pontiff and Schall (1998). Several of the other variables are marginal, with t-ratios within 10% of the cut-off point for significance. These include the short-term interest rate [Fama and Schwert (1977), using the more recent sample of Breen *et al.* (1989)], and the consumption–wealth ratio from Lettau and

Ludvigson (2001). All these regressors would be considered significant using the standard cutoffs.

It is interesting to note that using larger sample sizes does not solve the spurious regression problem; in fact, the critical t-ratios can be larger when the sample sizes are larger. We can see that the critical t-ratios in Table 2 are similar, for the examples with $\rho^* = 97\%$, whether the sample size is $T = 393$, 611, or 824 months. This result is driven by the fact that the standard errors of the t-ratios are too small, regardless of the sample size.

We do not show it in the tables, but our simulations confirm that when $\rho^* = 0$, meaning there is no persistence in the *ex ante* expected return, the spurious regression phenomenon does not arise. This is true even when the predictor variable used in the regression is highly persistent. The logic is that when the true slope in Eq. (1) is zero and $\rho^* = 0$, the regression error has no persistence, so the standard errors are well behaved. Using a highly persistent predictor variable does not change that. This implies that spurious regression is not a problem from the perspective of testing the null hypothesis that expected stock returns are unpredictable, even if a highly autocorrelated regressor is used. In particular, the eight significant t-ratios in Table 1 can be interpreted as rejecting the hypothesis that the *ex ante* expected return is constant over time. Thus, spurious regression by itself may or may not be a problem, depending on the purpose of the predictive regression model. This somewhat subtle distinction has important practical implications, which are discussed below. First, we show that the studies in Table 1 do not hold up to scrutiny when we also consider data mining.

4   Spurious regression and data mining

Spurious regression interacts with data mining. When more instruments are examined, the more persistent ones are likely to be chosen, and the spurious regression problem is amplified. Spurious regression also makes data mining seem more effective at uncovering predictive relations, because the spurious ones are easier to find.

Table 3 revisits the academic studies summarized in Table 1, in view of spurious regression and data mining. We report critical values for L, the number of instruments mined, sufficient to render the regressions no longer significant at the 5% level. We use two assumptions about persistence in the true expected returns: ρ^* is set equal to the sample values from the studies, as in Table 1, or $\rho^* = 95\%$. For a given value of L the critical t-ratios increase with the persistence of the true expected return, ρ^*, so the critical value of L is smaller when ρ^* is larger.[4] This says that you do not have to mine as many variables to find a large t-ratio when spurious regression is afoot. For these academic studies the critical values of L are 5 or smaller. The case where $L = 5$ is the lagged excess return on a two-month Treasury bill, following Campbell (1987). This is an interesting

Table 3 Simulation results for spurious regression and data mining.

Obs	ρ	Critical L	ρ^*	Critical L
393	0.97	2	0.95	4
264	0.32	2	0.95	1
264	0.15	2	0.95	1
264	0.08	5	0.95	1
420	0.97	1	0.95	1
720	0.97	1	0.95	1
732	0.92	1	0.95	1
611	0.95	1	0.95	1
611	0.97	1	0.95	1
66	0.66	2	0.95	1
184	0.79	2	0.95	1
824	0.97	1	0.95	1
552	0.98	1	0.95	1

example because the lagged variable is not very auto-correlated, at 8%, and when we set $\rho^* = 8\%$ there is no spurious regression effect. The critical value of $L = 5$ thus reflects essentially pure data mining, without spurious regression.[5] However, when we set $\rho^* = 95\%$ in this example, the critical value of L falls to one, further illustrating the interaction between the data mining and spurious regression effects. Overall, none of the regressions in Table 1 are significant when you allow for the possibility of a spurious data mining process.

5 Solutions

The results of this article put many analysts between a rock and a hard place. On the one hand, our results *do not* imply that stock returns are unpredictable. It is quite the opposite, in fact. Stock returns should be unpredictable when expected returns are constant over time. In this case, $\rho^* = 0$ and our simulations show that the regressions in Table 1 are not spurious. So, there is still a strong motivation to build predictive models. The trouble is, it is hard to build them without biases. We consider some alternative solutions.[6]

We are concerned with the interaction of two problems, data mining and spurious regression bias. On the former, the important point is to correctly account for the amount of data that is mined. The more predictive models you examine, the higher should be the statistical hurdle. Lo and MacKinlay (1990) and Foster *et al.* (1997) provide examples of this approach when spurious regression is not at issue, and the statistics literature contains a number of useful tools for addressing the problem (see Hastie *et al.*, 2001). When spurious regression enters the picture things get more complicated. Even when the statistics correctly account for the amount of data that is mined, the statistical hurdle will still be too low because of spurious regression bias. However, if you can get rid of spurious

regression bias it cannot interact with data mining. We, therefore, focus on dealing with the spurious regression bias.

The essential problem in dealing with the spurious regression bias is to get the right standard errors. We examine the Newey–West (Newey and West, 1987) style standard errors that have been popular in recent studies. These involve a number of "lag" terms to capture persistence in the regression error. We use the automatic lag selection procedure described in Note 1, and we compare it to a simple ordinary least squares (OLS) regression with no adjustment to the standard errors, and to a heteroskedasticity-only correction due to White (1980). Table 4 shows the critical t-ratios you would have to use in a 5%, two-tailed test, accounting for the possibility of spurious regression. Here, we consider an extreme case with $\rho^* = 99\%$, because if we can find a solution that works in this case it should also work in most realistic cases. The critical t-ratios range from 2.24 to 6.12 in the first three columns. None of the approaches delivers the right critical value, which should be 1.96. The table shows that a larger sample size is no insurance against spurious regression. In fact, the problem is the worst at the largest sample size.

The Newey–West approach is consistent, which means that by letting the number of lags grow when you have longer samples, you should eventually get the right standard error and solve the spurious regression problem. So, the first potential solution we examine is simply to use more lags in the Newey–West standard errors. Unfortunately, it is hard to know how many lags to use. The reason is that in stock return regressions the large unexpected part of stock returns is in the regression error, and this "noise" masks the persistence in the expected part of the return. If you use too few lags the standard errors are biased and the spurious regression remains. The "White" example in column two is an illustration of a case where the number of lags is zero. If you use

too many lags the standard errors will be inefficient and inaccurate, except in the largest sample sizes. We use simulations to evaluate the strategy of letting the number of lags grow large. We found that in realistic sample sizes, more lags did not help the spurious regression problem. The fourth column of Table 4 (denoted NW(20)) shows an example of this where 20 lags are used in monthly data. The critical t-ratios are still much larger than two. In the smaller sample size ($T = 60$) it is actually better to use the standard procedure, without any adjustments. Increasing the number of lags in the Newey–West standard errors is not a practical cure for the spurious regression problem.

A second potential solution to the spurious regression problem is to include a lagged value of the dependent variable as an additional right-hand side variable in the regression. The logic of this approach is that the spurious regression problem is caused by autocorrelation in the regression residuals, which is inherited from the dependent variable. Therefore, logic suggests that putting a lagged dependent variable in the regression should "soak up" the autocorrelation, leaving a clean residual. The columns of Table 4 labeled "lagged return" evaluate this approach. It helps a little bit, compared with no adjustment, but the critical t-ratios are still much larger than 2 at the larger sample sizes. For a hypothetical monthly sample with 350 observations, a t-ratio of 3.7 is needed for significance. The reason that this approach does not work very well is essentially the same reason that increasing the number of lags in the Newey–West method fails to work in finite samples. It is peculiar to stock return regressions, where the *ex ante* expected return may be persistent but the actual return includes a large amount of unpredictable noise. Spurious regression is driven in this case by persistence in the *ex ante* return, but the noise makes the lagged return a poor instrument for this persistence.[7]

Of the various approaches we tried, the most practically useful insurance against spurious regression seems to be a form of "stochastic detrending" of the lagged variable, advocated by Campbell (1991). The approach is very simple. Just transform the lagged variable by subtracting off a trailing moving

Table 4 Possible solutions to the spurious regression problem: critical t-ratios. Each cell contains the critical t-ratios at the 97.5 percentiles of 10,000 Monte Carlo simulations. OLS contains the critical t-ratios without any adjustment to the standard errors, in the White column the t-stats are formed using White's standard errors, the NW(auto) t-stats use Newey–West standard errors based on the automatic lag selection, the NW(20) t-stats use the Newey–West procedure with 20 lags. The regression model of stock returns in columns two-to-five has one independent variable—the lagged instrument; in columns six and seven, two independent variables—the lagged instrument and the lagged return; in the last two columns the only independent variable, the lagged instrument, is stochastically detrended using a trailing 12 month moving average. The autocorrelation parameter of the *ex ante* expected return and the lagged predictor variable is set to 99% and the *ex ante* return variance is 10% of the total return variance.

Observations	OLS	White	NW(auto)	NW(20)	Lagged return		Detrended (12)	
					OLS	NW(auto)	OLS	NW(auto)
60	2.24	2.36	2.71	3.81	2.19	2.67	2.06	2.46
350	4.04	4.10	3.87	3.77	3.74	3.73	2.28	2.21
2000	6.08	6.12	4.62	4.17	5.49	4.58	2.33	1.94

average of its own past values. Instead of regressing returns on Z_t, regress them on

$$X_t = Z_t - \frac{1}{\tau} \sum_{j=1,\ldots,\tau} Z_{t-j}. \qquad (3)$$

While different numbers of lags could be used in the detrending, Campbell uses 12 monthly lags, which seems natural for monthly data. We evaluate the usefulness of his suggestion in the last two columns of Table 4. With this approach, the critical t-ratios are less than 2.5 at all sample sizes, and much closer to 1.96 than any of the other examples. The simple detrending approach works pretty well. Detrending lowers the persistence of the transformed regressor, resulting in autocorrelations that are below the levels where spurious regression becomes a problem. Stochastic detrending can do this without destroying the information in the data about a persistent *ex ante* return, as would be likely to occur if the predictor variable is simply first differenced. Overall, we recommend stochastic detrending as a simple method for controlling the problem of spurious regression in stock returns.

6   Conclusions

Our results have distinct implications for tests of predictability and model selection. In tests of predictability, the researcher chooses a lagged variable and regresses future returns on the variable. The hypothesis is that the slope coefficient is zero. Spurious regression presents no problem from this perspective, because under the null hypothesis the expected return is not actually persistent. If this characterizes the academic studies of Table 1, the eight t-ratios larger than 2 suggest that *ex ante* stock returns are not constant over time.

The more practical problem is model selection. In model selection, the analyst chooses a lagged instrument to predict returns, for purposes such as implementing a tactical asset allocation strategy,

active portfolio management, conditional performance evaluation or market timing. Here is where the spurious regression problem rears its ugly head. You are likely to find a variable that appears to work on the historical data, but will not work in the future. A simple form of stochastic detrending lowers the persistence of lagged predictor variables, and can be used to reduce the risk of finding spurious predictive relations.

The pattern of evidence for the lagged variables in the academic literature is similar to what is expected under a spurious data mining process with an underlying persistent *ex ante* return. In this case we would expect instruments to be discovered, then fail to work with fresh data. The dividend yield rose to prominence in the 1980s, but apparently fails to work for post-1990 data [Goyal and Welch, 2003; Schwert (in press)]. The book-to-market ratio also seems to have weakened in recent data. When more data are available, new instruments appear to work (e.g. Lettau and Ludvigson, 2001; Lee *et al.*, 1999). Analysts should be wary that the new instruments, if they arise from the spurious mining process that we suggest, are likely to fail in future data, and thus fail to be practically useful.

Appendix: The sample of 500 instruments

All the data come from the web site Economagic.com: Economic Time Series Page, maintained by Ted Bos. The sample consists of all monthly series listed on the main homepage of the site, except under the headings of LIBOR, Australia, Bank of Japan, and Central Bank of Europe. From the Census Bureau we exclude Building Permits by Region, State, and Metro Areas (more than 4000 series). From the Bureau of Labor Statistics we exclude all non-civilian Labor force data and State, City, and International Employment (more than 51,000 series). We use the CPI measures from the

city average listings, but include no finer subcategories. The PPI measures include the aggregates and the 2-digit subcategories. From the Department of Energy we exclude data in Section 10, the International Energy series.

We first randomly select (using a uniform distribution) 600 out of the 10,866 series that were left after the above exclusions. From these 600 we eliminated series that mixed quarterly and monthly data and extremely sparse series, and took the first 500 from what remained.

Acknowledgments

Notes

[1] The *t*-ratios are based on the OLS slopes and the Newey–West (Newey and West, 1987) standard errors, where the number of lags is chosen based on the number of statistically significant residual autocorrelations. We compute 12 sample autocorrelations and compare the values with a cutoff at two approximate standard errors: $2/\sqrt{T}$, where T is the sample size. The number of lags chosen is the minimum lag length at which no higher order autocorrelation is larger than two standard errors.

[2] While Granger and Newbold (1974) do not study the slopes and standard errors to identify the separate effects, our simulations designed to mimic their setting confirm that their slopes are well behaved, while the standard errors are biased. Granger and Newbold use OLS standard errors, while we focus on the heteroskedasticity and autocorrelation-consistent standard errors that are more common in recent studies.

[3] For details on the simulation design, see Ferson *et al.* (2003).

[4] In simulations that we do not report, we learned that spurious regression is not much of a problem for ρ^* less than 95%. Of course, no one knows what the persistence

of *ex ante* expected returns really is. But if expectations evolve slowly over time, updating in proportion to new information, they may be highly persistent.

[5] For a detailed analysis of pure data mining effects, see Foster *et al.* (1997).

[6] There is an active stream of academic research concerned with measuring and testing for return predictability in the presence of persistent predictor variables. Examples include Torous *et al.* (in press), Torous and Valkanov (2003), Campbell and Yogo (2002), and Chapman *et al.* (2003). However, none of these studies addresses the interaction between spurious regression and data mining.

[7] More formally, consider a case where the *ex ante* return is an AR(1) process, in the Box–Jenkins notation. The realized return is distributed as an AR(1) plus noise, which is ARMA(1, 1). Regressing the return on the lagged return, the residual may still be highly persistent because of the moving average component.

References

Breen, W., Glosten, L. R., and Jagannathan, R. (1989). "Economic Significance of Predictable Variations in Stock Index Returns." *Journal of Finance* **44**, 1177–1190.

Bossaerts, P. and Hillion, P. (1999). "Implementing Statistical Criteria to Select Return Forecasting Models: What do we Learn?" *Review of Financial Studies* **12**, 405–428.

Campbell, J. Y. (1987). "Stock Returns and the Term Structure." *Journal of Financial Economics* **18**, 373–400.

Campbell, J. Y. and Shiller, R. (1988). "The Dividend Ratio and Small Sample Bias." *Economics Letters* **29**, 325–331.

Campbell, J. Y. and Yogo, M. (2002). "Efficient Tests of Stock Return Predictability." Manuscript, Harvard University.

Chapman, D. A., Simin, T., and Yan, H. (2003). "Stock Returns and Dividend Yields: Some New Evidence." Working Paper, the Pennsylvania State University.

Fama, E. F. (1990). "Stock Returns, Expected Returns, and Real Activity." *Journal of Finance* **45**, 1089–1108.

Fama, E. F. and Schwert, G. W. (1977). "Asset Returns and Inflation." *Journal of Financial Economics* **5**, 115–146.

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

Fama, E. and French, K. R. (1988b). "Permanent and temporary components of stock prices." *Journal of Political Economy* **96**, 246–273.

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

Foster, F. D., Smith, T. and Whaley, R. E. (1997). "Assessing Goodness-of-Fit of Asset Pricing Models: The Distribution of the Maximal R-Squared." *Journal of Finance* **52**, 591–607.

Goyal, A. and Welch, I. "Predicting the Equity Premium with Dividend Ratios." *Management Science* (in press).

Granger, C. W. J. and Newbold, P. (1974). "Spurious Regressions in Economics." *Journal of Econometrics* **4**, 111–120.

Hastie, T., Tibshirani, R., and Friedman, J. (2001). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, New York: Springer Verlag.

Keim, D. B. and Stambaugh, R. F. (1986). "Predicting Returns in the Bond and Stock Markets." *Journal of Financial Economics* **17**, 357–390.

Kothari, S. P. and Shanken, J. (1997). "Book-to-Market Time Series Analysis." *Journal of Financial Economics* **44**, 169–203.

Lee, C., Myers, J., and Swaminathan, B. (1999). "What is the Intrinsic Value of the Dow?" *Journal of Finance* 1693–1742.

Lettau, M. and Ludvigson, S. (2001). "Consumption, Aggregate Wealth and Expected Stock Returns." *Journal of Finance* **56**, 815–849.

Lo, A. W. and MacKinlay, A. C. (1988). "Stocks Prices do not Follow Random Walks." *Review of Financial Studies* **1**, 41–66.

Lo, A. W. and MacKinlay, A. C. (1990). "Data Snooping in Tests of Financial Asset Pricing Models." *Review of Financial Studies* **3**, 431–467.

Newey, W. and West, K. D. (1987). "A Simple, Positive Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica* **55**, 703–708.

Pontiff, J. and Schall, L. (1998). "Book-to-Market as a Predictor of Market Returns." *Journal of Financial Economics* **49**, 141–160.

Schwert, G. W. *Anomalies and Market Efficiency*, Chapter 15. In George M. Constantinides, Milton Harris and Rene' M. Stulz (eds) *Handbook of the Economics of Finance*. Amsterdam: North Holland (in press).

Torous, W., Valkanov, R., and Yan, S. "On Predicting Stock Returns with Nearly Integrated Explanatory Variables." *Journal of Business* (in press).

Torous, W. and Valkanov, R. (2003). "Boundaries of Predictability: Noisy Predictive Regressions." Working Paper, University of California at Los Angeles.

Yule, G. U. (1926). "Why do we Sometimes get Nonsense Correlations Between Time Series? A Study in Sampling and the Nature of Time Series." *Journal of the Royal Statistical Society* **89**, 1–64.