Spurious Regression and Data Mining in Conditional Asset Pricing Models*

for the Handbook of Quantitative Finance, C.F. Lee, Editor, Springer Publishing

by:

Wayne Ferson, University of Southern California
Sergei Sarkissian, McGill University
Timothy Simin, the Pennsylvania State University

*This chapter is based in part on two published papers by the authors: (i) "Spurious Regressions in Financial Economics?," (Journal of Finance, 2003), and (ii) "Asset Pricing Models with Conditional Betas and Alphas: The Effects of Data Snooping and Spurious Regression," (Journal of Financial and Quantitative Analysis, 2007). In addition, we are grateful to Raymond Kan for a helpful suggestion.

OUTLINE

1. Introduction

Predictive models for common stock returns have long been a staple of financial economics. Early studies, reviewed by Fama (1970), used such models to examine market efficiency. Stock returns are assumed to be predictable, based on lagged instrumental variables, in the current conditional asset pricing literature.

The simplest predictive model is a regression for the future stock return, r_{t+1}, on a lagged predictor variable:

$$(1) \qquad r_{t+1} = a + bZ_t + v_{t+1}.$$

Standard lagged variables include the levels of short-term interest rates, payout-to-price ratios for stock market indexes, and yield spreads between low-grade and high-grade bonds or between long- and short-term bonds. Table 1 surveys major studies that propose predictor variables. Many of these variables behave as persistent, or highly autocorrelated, time series. We study the finite sample properties of stock return predictive regressions with persistent lagged regressors.

Regression models for stock or portfolio returns on contemporaneously-measured market-wide factors have also long been a staple of financial economics. Such factor models are used in event studies (e.g., Fama et al., 1969), in tests of asset pricing theories such as the Capital Asset Pricing Model (CAPM, Sharpe, 1964) and in other applications. For example, when the market return r_m is the factor, the regression model for the return r_{t+1} is:

$$(2) \qquad r_{t+1} = \alpha + \beta\, r_{m,t+1} + u_{t+1},$$

where $E(u_{t+1}) = E(u_{t+1}r_{m,t+1}) = 0$. The slope coefficients are the "betas," which measure the market-factor risk. When the returns are measured in excess of a reference asset like a risk-free Treasury bill return, the intercepts are the "alphas," which measure the expected abnormal return. For example, when r_m is the market portfolio excess return, the CAPM implies that $\alpha = 0$, and the model is evaluated by testing that null hypothesis.

Recent work in conditional asset pricing allows for time-varying betas modeled as linear functions of lagged predictor variables, following Maddala (1977). Prominent examples include Shanken (1990), Cochrane (1996), Ferson and Schadt (1996), Jagannathan and Wang (1996) and Lettau and Ludvigson (2001). The time-varying beta coefficient is $\beta_t = b_0 + b_1 Z_t$, where Z_t is a lagged predictor variable. In some cases, the intercept or conditional alpha is also time-varying, as $\alpha_t = \alpha_0 + \alpha_1 Z_t$ (e.g. Christopherson, Ferson and Glassman, 1998). This results in the following regression model:

$$(3) \qquad r_{t+1} = \alpha_0 + \alpha_1 Z_t + b_0\, r_{m,t+1} + b_1\, r_{m,t+1} Z_t + u_{t+1},$$

where $E(u_{t+1}) = E(u_{t+1}[Z_t\, r_{m,t+1}]) = 0$. The conditional CAPM implies that $\alpha_0 = 0$ and $\alpha_1 = 0$. This chapter also studies the finite-sample properties of asset pricing model regressions like (3) when there are persistent lagged regressors.

The rest of the chapter is organized as follows. Section 2 discusses the issues of data mining and spurious regression in the simple predictive regression (1). Section 3 discusses the impact of spurious regression and data mining on conditional asset pricing. Section 4 describes the data. Section 4 presents the models used in the simulation experiments. Section 6 presents the simulation results for predictive regressions. Section 7 presents the simulation results for various forms of conditional asset pricing models. Section 7 discusses and evaluates solutions to the problems of spurious regression and data mining. Section 9 examines the robustness of results. Section 10 concludes.

2. Spurious Regression and Data Mining in Predictive Regressions

In our analysis of regressions, like (1), that attempt to predict stock returns, we focus on two issues. The first is spurious regression, analogous to Yule (1926), and Granger and Newbold

(1974). These studies warned that spurious relations may be found between the levels of trending time series that are actually independent. For example, given two independent random walks, it is likely that a regression of one on the other will produce a "significant" slope coefficient, evaluated by the usual *t*-statistics.

Stock returns are not highly autocorrelated, so you might think that spurious regression would not be an issue for stock returns. Thus, one may think that spurious regression problems are unlikely. However, the returns may be considered as the sum of an unobserved expected return, plus unpredictable noise. If the underlying *expected* returns are persistent time series there is still a risk of spurious regression. Because the unpredictable noise represents a substantial portion of the variance of stock returns, the spurious regression will differ from the classical setting.

The second issue is "naïve data mining" as studied for stock returns by Lo and MacKinlay (1990), Foster, Smith, and Whaley (1997), and others. If the standard instruments employed in the literature arise as the result of a collective search through the data, they may have no predictive power in the future. Stylized "facts" about the dynamic behavior of stock returns using these instruments (e.g., Cochrane, 1999) could be artifacts of the sample. Such concerns are natural, given the widespread interest in predicting stock returns. Not all data mining is naïve. In fact, increasing computing power and data availability have allowed the development of some very sophisticated data mining (for statistical foundations, see Hastie, Tibshirani, and Friedman, 2001).

We focus on spurious regression and the interaction between data mining and spurious regression bias. If the underlying expected return is not predictable over time, there is no spurious regression bias, even if the chosen regressor is highly autocorrelated. This is because, under the null hypothesis that there is no predictability, the autocorrelation of the regression errors the same as that of the left hand side asset returns. In this case, our analysis reduces to pure data mining as studied by Foster, Smith, and Whaley (1997).

The spurious regression and data mining affects reinforce each other. If researchers have mined the data for regressors that produce high t-statistics in predictive regressions, then mining is more likely to uncover the spurious, persistent regressors. The standard regressors in the literature tend to be highly autocorrelated, as expected if the regressors result from this kind of a "spurious mining" process. For reasonable parameter values, all the regressions that we review from the literature are consistent with a spurious mining process, even when only a small number of instruments are considered in the mining.

While data mining amplifies the problem of spurious regressions, persistent lagged variables and spurious regression also magnify the impact of data mining. As a consequence, we show that standard corrections for data mining are inadequate in the presence of persistent lagged variables.

These results have profound potential implications for asset pricing regressions because the conditional asset pricing literature has, for the most part, used variables that were discovered based on predictive regressions like (1). It is important therefore to examine how data mining and spurious regression biases influence asset pricing regressions.

3. Spurious Regression, Data Mining and Conditional Asset Pricing

The conditional asset pricing literature using regressions like (3) has evolved from the literature on pure predictive regressions. First, studies identified lagged variables that appear to predict stock returns. Later studies, beginning with Gibbons and Ferson (1985), used the same variables to study asset pricing models. Thus, it is reasonable to presume that data mining is directed at the simpler predictive regressions. The question now is: How does this affect the validity of the subsequent asset pricing research that uses these variables in regressions like (3)?

Table 2 summarizes representative studies that use the regression model (3). It lists the sample period, number of observations and the lagged instruments employed. It also indicates whether the study uses the full model (3), with both time-varying betas and alphas, or restricted versions of the model in which either the time-varying betas or time-varying alphas are suppressed. Finally, the table summarizes the largest t-statistics for the coefficients α_1 and b_1 reported in each study. If we find that the largest t-statistics are insignificant in view of the joint effects of spurious regression and data mining, then none of the coefficients are significant. We return to this table later and revisit the evidence.

Using regression models like Equation (3), the literature has produced a number of "stylized facts." First, studies typically find that the intercept is smaller in the "conditional" model (3) than in the "unconditional" model (2): $|\alpha| > |\alpha_0|$. The interpretation of these studies is that the conditional CAPM does a better job of "explaining" average returns than the unconditional CAPM. Examples with this finding include Cochrane (1996), Ferson and Schadt (1996), Ferson and Harvey (1997, 1999), Lettau and Ludvigson (2001), and Petkova and Zhang (2005). Second, studies typically find evidence of time varying betas: The coefficient estimate for b_1 is statistically significant. Third, studies typically find that the conditional models fail to completely explain the dynamic properties of returns: The coefficient estimate for α_1 is significant, indicating a time-varying alpha. Our objective is to study the reliability of such inferences in the presence of persistent lagged instruments and data mining.

4. The Data

Table 1 surveys nine of the major studies that propose instruments for predicting stock returns. The table reports summary statistics for monthly data, covering various sub-periods of 1926 through 1998. The sample size and period depends on the study and the variable, and the table provides the details. We attempt to replicate the data series that were used in these studies as closely as possible. The summary statistics are from our data. Note that the first order autocorrelations of the predictor variables frequently suggest a high degree of persistence. For example, short-term Treasury-bill yields, monthly book-to-market ratios, the dividend yield of the S&P500 and some of the yield spreads have sample first order autocorrelations of 0.97 or higher.

[Table 1 about here]

Table 1 also summarizes regressions for the monthly return of the S&P500 stock index, measured in excess of the one-month Treasury-bill return from Ibbotson Associates, on the lagged instruments. These are OLS regressions using one instrument at a time. We report the slope coefficients, their *t*-ratios, and the adjusted *R*-squares. The *R*-squares range from less than one percent to more than seven percent, and eight of the 13 *t*-ratios are larger than 2.0. The *t*-ratios are based on the OLS slopes and Newey-West (1987) standard errors, where the number of lags is chosen based on the number of statistically significant residual autocorrelations.[1]

The small *R*-squares in Table 1 suggest that predictability represents a tiny fraction of the variance in stock returns. However, even a small *R*-squared can signal economically significant predictability. For example, Kandel and Stambaugh (1996) and Fleming, Kirby, and Ostdiek (2001) find that optimal portfolios respond by a substantial amount to small *R*-squares in standard models. Studies combining several instruments in multiple regressions report higher *R*-squares. For example, Harvey (1989), using five instruments, reports adjusted *R*-squares as high as 17.9 percent for size portfolios. Ferson and Harvey (1991) report *R*-squares of 5.8 percent to 13.7 percent for monthly size and industry portfolio returns. These

---

[1] Specifically, we compute 12 sample autocorrelations and compare their 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. The number of lags chosen is indicated in the far right column.

values suggest that the "true" R-squared, if we could regress the stock return on its time-varying conditional mean, might be substantially higher than we see in Table 1. To accommodate this possibility, we allow the true R-squared in our simulations to vary over the range from zero to 15 percent. For exposition we focus on an intermediate value of 10 percent.

To incorporate data mining, we compile a randomly selected sample of 500 potential instruments, through which our simulated analyst sifts to mine the data for predictor variables. 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 4,000 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 Consumer Price Index (CPI) measures from the city average listings, but include no finer subcategories. The Producer Price Index (PPI) measures include the aggregates and the two-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.

Because many of the data are reported in levels, we tested for unit roots using an augmented Dickey-Fuller test (with a zero order time polynomial). We could not reject the hypothesis of a unit root for 361 of the 500 series and we replaced these series with their first differences. The 500 series are randomly ordered, and then permanently assigned numbers between one and 500. When a data miner in our simulations searches through, say 50 series, we use the sampling properties of the first 50 series to calibrate the parameters in the simulations.

We also use our sample of potential instruments to calibrate the parameters that govern the amount of persistence in the "true" expected returns in the model. On the one hand, if the instruments we see in the literature, summarized in Table 1, arise from a spurious mining process, they are likely to be more highly autocorrelated than the underlying "true" expected stock return. On the other hand, if the instruments in the literature are a realistic representation of expected stock returns, the autocorrelations in Table 1 may be a good proxy for the persistence of the true expected returns.[2] The mean autocorrelation of our 500 series is 15 percent and the median is two percent. Eleven of the 13 sample autocorrelations in Table 1 are higher than 15 percent, and the median value is 95 percent. We consider a range of values for the true autocorrelation based on these figures, as described below.

5. The Models

5.1. Predictive Regressions

In the model for the predictive regressions, the data are generated by an unobserved latent variable, Z_t^*, as:

$$r_{t+1} = \mu + Z_t^* + u_{t+1}, \tag{4}$$

where u_{t+1} is white noise with variance, σ_u^2. We interpret the latent variable, Z_t^* as the deviations of the conditional mean return from the unconditional mean, μ, where the

[2] There are good reasons to think that expected stock returns may be persistent. Asset pricing models like the consumption model of Lucas (1978) describe expected stock returns as functions of expected economic growth rates. Merton (1973) and Cox, Ingersoll, and Ross (1985) propose real interest rates as candidate state variables, driving expected returns in intertemporal models. Such variables are likely to be highly persistent. Empirical models for stock return dynamics frequently involve persistent, auto-regressive expected returns (e.g., Lo and MacKinlay, 1988; Conrad and Kaul, 1988; Fama and French, 1988b; or Huberman and Kandel, 1990).

expectations are conditioned on an unobserved "market" information set at time t. The predictor variables follow an autoregressive process:

$$(5) \qquad \left(Z_t^*, Z_t\right)' = \begin{Bmatrix} \rho^* & 0 \\ 0 & \rho \end{Bmatrix} \left(Z_{t-1}^*, Z_{t-1}\right)' + \left(\varepsilon_t^*, \varepsilon_t\right)',$$

where Z_t is the measured predictor variable and ρ is the autocorrelation. The assumption that the true expected return is autoregressive (with parameter ρ^*) follows previous studies such as Lo and MacKinlay (1988), Conrad and Kaul (1988), Fama and French (1988b), and Huberman and Kandel (1990).

To generate the artificial data, the errors $\left(\varepsilon_t^*, \varepsilon_t\right)$ are drawn randomly as a normal vector with mean zero and covariance matrix, Σ. We build up the time-series of the Z and Z^* through the vector autoregression equation (3), where the initial values are drawn from a normal with mean zero and variances, $Var(Z)$ and $Var(Z^*)$. The other parameters that calibrate the simulations are $\{\mu, \sigma_u^2, \rho, \rho^*, \text{and } \Sigma\}$.

We have a situation in which the "true" returns may be predictable, if Z_t^* could be observed. This is captured by the *true* R-squared, $Var(Z^*)/[Var(Z^*) + \sigma_u^2]$. We set $Var(Z^*)$ to equal the sample variance of the S&P500 return, in excess of a one-month Treasury-bill return, multiplied by 0.10. When the true R-squared of the simulation is 10 percent, the unconditional variance of the r_{t+1} that we generate is equal to the sample variance of the S&P500 return. When we choose other values for the true R-squared, these determine the values for the parameter σ_u^2. We set μ to equal the sample mean excess return of the S&P500 over the 1926 through 1998 period, or 0.71 percent per month.

The extent of the spurious regression bias depends on the parameters, ρ and ρ^*, which control the persistence of the measured and the true regressor. These values are determined by reference to Table 1 and from our sample of 500 potential instruments. The specifics differ across the special cases, as described below.

While the stock return could be predicted if Z_t^* could be observed, the analyst uses the measured instrument Z_t. If the covariance matrix Σ is diagonal, Z_t and Z_t^* are independent, and the true value of δ in the regression (1) is zero.

To focus on spurious regression in isolation, we specialize equation (3) as follows. The covariance matrix Σ is a 2 x 2 diagonal matrix with variances (σ^{*2}, σ^2). For a given value of ρ^* the value of σ^{*2} is determined as $\sigma^{*2} = (1- \rho^{*2})Var(Z^*)$. The measured regressor has $Var(Z) = Var(Z^*)$. The autocorrelation parameters, $\rho^* = \rho$ are allowed to vary over a range of values. (We also allow ρ and ρ^* to differ from one another, as described below.)

Following Granger and Newbold (1974), we interpret a spurious regression as one in which the "*t*-ratios" in the regression (1) are likely to indicate a significant relation when the variables are really independent. The problem may come from the numerator or the denominator of the *t*-ratio: The coefficient or its standard error may be biased. As in Granger and Newbold, the problem lies with the standard errors.[3] The reason is simple to understand. When the null hypothesis that the regression slope $\delta = 0$ is true, the error term u_{t+1} of the regression equation (1) inherits autocorrelation from the dependent variable. Assuming stationarity, the slope coefficient is consistent, but standard errors that do not account for the serial dependence correctly, are biased.

Because the spurious regression problem is driven by biased estimates of the standard error, the choice of standard error estimator is crucial. In our simulation exercises, it is possible to find an efficient unbiased estimator, since we know the "true" model that describes the regression error. Of course, this will not be known in practice. To mimic the practical reality, the analyst in our simulations uses the popular autocorrelation-heteroskedasticity-consistent (HAC) standard errors from Newey and West (1987), with an automatic lag selection procedure. The number of lags is chosen by computing the autocorrelations of the

---

[3] 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 (not reported in the tables) 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.

estimated residuals, and truncating the lag length when the sample autocorrelations become "insignificant" at longer lags. (The exact procedure is described in Footnote 1, and modifications to this procedure are discussed below.)

This setting is related to Phillips (1986) and Stambaugh (1999). Phillips derives asymptotic distributions for the OLS estimators of the regression (1), in the case where $\rho = 1$, $u_{t+1} \equiv 0$, and $\left\{\varepsilon_t^*, \varepsilon_t\right\}$ are general independent mean zero processes. We allow a nonzero variance of u_{t+1} to accommodate the large noise component of stock returns. We assume $\rho < 1$ to focus on stationary, but possibly highly autocorrelated, regressors.

Stambaugh (1999) studies a case where the errors $\left\{\varepsilon_t^*, \varepsilon_t\right\}$ are perfectly correlated, or equivalently, the analyst observes and uses the correct lagged stochastic regressor. A bias arises when the correlation between u_{t+1} and ε^*_{t+1} is not zero, related to the well-known small sample bias of the autocorrelation coefficient (e.g., Kendall (1954)). In the pure spurious regression case studied here, the observed regressor Z_t is independent of the true regressor Z_t^*, and u_{t+1} is independent of ε^*_{t+1}. The Stambaugh bias is zero in this case. The point is that there remains a problem in predictive regressions, in the absence of the bias studied by Stambaugh, because of spurious regression.

5.2. Conditional Asset Pricing Models

The data in our simulations of conditional asset pricing models are generated according to:

$$(6) \qquad \begin{aligned} r_{t+1} &= \beta_t \, r_{m,t+1} + u_{t+1}, \\ \beta_t &= 1 + Z_t^*, \\ r_{m,t+1} &= \mu + k\, Z_t^* + w_{t+1}. \end{aligned}$$

Our artificial analyst uses the simulated data to run the regression model (3), focusing on the t-statistics for the coefficients $\{\alpha_0, \alpha_1, b_0, b_1\}$. The variable Z_t^* in equation (6) is an unobserved latent variable that drives both expected market returns and time-varying betas.

The term β_t in Equation (6) is a time-varying beta coefficient. As Z_t^* has mean equal to zero, the expected value of beta is 1.0. When $k \neq 0$ there is an interaction between the time variation in beta and the expected market risk premium. A common persistent factor drives the movements in both expected returns and conditional betas. Common factors in time-varying betas and expected market premiums are important in asset pricing studies such as Chan and Chen (1988), Ferson and Korajczyk (1995), and Jagannathan and Wang (1996), and in conditional performance evaluation, as in Ferson and Schadt (1996). There is a zero intercept, or "alpha," in the data generating process for r_{t+1}, consistent with asset pricing theory.

The market return data, $r_{m,t+1}$, are generated as follows. The parameter μ was described earlier. The variance of the error is $\sigma_w^2 = \sigma_{sp}^2 - k^2\ Var(Z^*)$, where $\sigma_{sp} = 0.057$ matches the S&P500 return and $Var(Z^*) = 0.055$, is the estimated average monthly variance of the market betas on 58 randomly selected stocks from CRSP over the period 1926-1997.[4] The predictor variables follow the autoregressive process (3).

6. Results for Predictive Regressions

6.1. Pure Spurious Regression

---

[4] We calibrate the variance of the betas to actual monthly data by randomly selecting 58 stocks with complete CRSP data for January 1926 through December 1997. Following Fama and French (1997), we estimate simple regression betas for each stock's monthly excess return against the S&P500 excess return, using a series of rolling 5-year windows, rolling forward one month at a time. For each window we also compute the standard error of the beta estimate. This produces a series of 805 beta estimates and standard error estimates for beta for each firm. We calibrate the variance of the true beta for each firm to equal the sample variance of the rolling beta estimates minus the average estimated variance of the estimator. Averaging the result across firms, the value of $Var(Z^*)$ is 0.0550. Repeating this exercise with firms that have data from January of 1926 through the end of 2004 increases the number of months used from 864 to 948 but decreases the number of firms from 58 to 46. The value of $Var(Z^*)$ in this case is 0.0549.

Table 3 summarizes the results for the case of pure spurious regression, with no data mining. We record the estimated slope coefficient in regression (1), its Newey-West *t*-ratio, and the coefficient of determination at each trial and summarize their empirical distributions. The experiments are run for two sample sizes, based on the extremes in Table 1. These are $T = 66$ and $T = 824$ in Panels A and B, respectively. In Panel C, we match the sample sizes to the studies in Table 1. In each case, 10,000 trials of the simulation are run; 50,000 trials on a subset of the cases produce similar results.

[Table 3 about here]

The rows of Table 3 refer to different values for the true *R*-squares. The smallest value is 0.1 percent, where the stock return is essentially unpredictable, and the largest value is 15 percent. The columns of Table 3 correspond to different values of ρ^*, the autocorrelation of the true expected return, which runs from 0.00 to 0.99. In these experiments we set $\rho = \rho^*$. The sub-panels labeled *Critical t-statistic* and *Critical estimated* R^2 report empirical critical values from the 10,000 simulated trials, so that 2.5 percent of the *t*-statistics or five percent of the *R*-squares, lie above these values.

The sub-panels labeled *Mean* δ report the average slope coefficients over the 10,000 trials. The mean estimated values are always small, and very close to the true value of zero at the larger sample size. This confirms that the slope coefficient estimators are well behaved, so that bias due to spurious regression comes from the standard errors.

When $\rho^* = 0$, and there is no persistence in the true expected return, the table shows that spurious regression phenomenon is not a concern. This is true even when the measured regressor is highly persistent. (We confirm this with additional simulations, not reported in the tables, where we set $\rho^* = 0$ and vary ρ.) The logic is that when the slope in Equation (1) is zero and $\rho^* = 0$, the regression error has no persistence, so the standard errors are well behaved. 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.

Table 3 shows that spurious regression bias does not arise to any serious degree, provided ρ^* is 0.90 or less, and provided that the true R^2 is one percent or less. For these parameters the empirical critical values for the t-ratios are 2.48 (T = 66, Panel A), and 2.07 (T = 824, Panel B). The empirical critical R-squares are close to their theoretical values. For example, for a five percent test with T = (66, 824) the F distribution implies critical R-squared values of (5.9 percent, 0.5 percent). The values in Table 3 when $\rho^* = 0.90$ and true R^2 = one percent, are (6.2 percent, 0.5 percent); thus, the empirical distributions do not depart far from the standard rules of thumb.

Variables like short-term interest rates and dividend yields typically have first order sample autocorrelations in excess of 0.95, as we saw in Table 1. We find substantial biases when the regressors are highly persistent. Consider the plausible scenario with a sample of T = 824 observations where $\rho = 0.98$ and true R^2 = 10 percent. In view of the spurious regression phenomenon, an analyst who was not sure that the correct instrument is being used and who wanted to conduct a five percent, two-tailed t-test for the significance of the measured instrument, would have to use a t-ratio of 3.6. The coefficient of determination would have to exceed 2.2 percent to be significant at the five percent level. These cutoffs are substantially more stringent than the usual rules of thumb.

Panel C of Table 3 revisits the evidence from the literature in Table 1. The critical values for the t-ratios and R-squares are reported, along with the theoretical critical values for the R-squares, implied by the F-distribution. We set the true R-squared value equal to 10 percent and $\rho^* = \rho$ in each case. We find that seven of the 17 statistics in Table 1 that would be considered significant using the traditional standards, are no longer significant in view of the spurious regression bias.

While Panels A and B of Table 3 show that spurious regression can be a problem in stock return regressions, Panel C finds that accounting for spurious regression changes the inferences about specific regressors that were found to be significant in previous studies. In particular, we question the significance of the term spread in Fama and French (1989), on the basis of either the t-ratio or the R-squared of the regression. Similarly, the book-to-market ratio of the Dow Jones index, studied by Pontiff and Schall (1998) fails to be significant with either statistic. Several other variables are marginal, failing on the basis of one but not both statistics. These include the short-term interest rate (Fama and Schwert, 1977; using the more recent sample of Breen, Glosten, and Jagannathan, 1989), the dividend yield (Fama and French, 1988a), and the quality-related yield spread (Keim and Stambaugh, 1986). All of these regressors would be considered significant using the standard cutoffs.

It is interesting to note that the biases documented in Table 2 do not always diminish with larger sample sizes; in fact, the critical t-ratios are larger in the lower right corner of the panels when $T = 824$ than when $T = 66$. The mean values of the slope coefficients are closer to zero at the larger sample size, so the larger critical values are driven by the standard errors. A sample as large as $T = 824$ is not by itself a cure for the spurious regression bias. This is typical of spurious regression with a unit root, as discussed by Phillips (1986) for infinite sample sizes and nonstationary data.[5] It is interesting to observe similar patterns, even with stationary data and finite samples.

Phillips (1986) shows that the sample autocorrelation in the regression studied by Granger and Newbold (1974) converges in limit to 1.0. However, we find only mildly inflated residual autocorrelations (not reported in the tables) for stock return samples as large as $T = 2000$, even when we assume values of the true R^2 as large as 40 percent. Even in these extreme cases, none of the empirical critical values for the residual autocorrelations are larger

---

[5] Phillips derives asymptotic distributions for the OLS estimators of equation (1), in the case where $\rho = 1$, $u_{t+1} \equiv 0$. He shows that the t-ratio for δ diverges for large T, while $t(\delta)/\sqrt{T}$, δ, and the coefficient of determination converge to well-defined random variables. Marmol (1998) extends these results to multiple regressions with partially integrated processes, and provides references to more recent theoretical literature. Phillips (1998) reviews analytical tools for asymptotic analysis when nonstationary series are involved.

than 0.5. Since u_{t+1} = 0 in the cases studied by Phillips, we expect to see explosive autocorrelations only when the true R^2 is very large. When R^2 is small the white noise component of the returns serves to dampen the residual autocorrelation. Thus, we are not likely to see large residual autocorrelations in stock return regressions, even when spurious regression is a problem. The residuals-based diagnostics for spurious regression, such as the Durbin-Watson tests suggested by Granger and Newbold, are not likely to be very powerful in stock return regressions. For the same reason, typical application of the Newey-West procedure, where the number of lags is selected by examining the residual autocorrelations, is not likely to resolve the spurious regression problem.

Newey and West (1987) show that their procedure is consistent for the standard errors when the number of lags used grows without bound as the sample size T increases, provided that the number of lags grows no faster than $T^{1/4}$. The lag selection procedure in Table 3 examines 12 lags. Even though no more than nine lags are selected for the actual data in Table 1, more lags would sometimes be selected in the simulations, and an inconsistency results from truncating the lag length.[6] However, in finite samples an increase in the number of lags can make things worse. When "too many" lags are used the standard error estimates become excessively noisy, which thickens the tails of the sampling distribution of the t-ratios. This occurs for the experiments in Table 2. For example, letting the procedure examine 36 autocorrelations to determine the lag length (the largest number we find mentioned in published studies) the critical t-ratio in Panel A, for true R^2 = 10 percent and ρ^* = 0.98, increases from 2.9 to 4.8. Nine of the 17 statistics from Table 1 that are significant by the usual rules of thumb now become insignificant. The results calling these studies into question are therefore even stronger than before. Thus, simply increasing the number of lags in the

[6] At very large sample sizes, a huge number of lags can control the bias. We verify this by examining samples as large as T = 5000, letting the number of lags grow to 240. With 240 lags the critical t-ratio when the true R^2 = 10 percent and ρ = 0.98 falls from 3.6 in Panel B of Table 2, to a reasonably well-behaved value of 2.23.

Newey-West procedure is not likely to resolve the finite sample, spurious regression bias.[7] We discuss this issue in more detail in Section 8.1.

We draw several conclusions about spurious regression in stock return predictive regressions. Given persistent expected returns, spurious regression can be a serious concern well outside the classic setting of Yule (1926) and Granger and Newbold (1974). Stock returns, as the dependent variable, are much less persistent than the levels of most economic time series. Yet, when the *expected* returns are persistent, there is a risk of spurious regression bias. The regression residuals may not be highly autocorrelated, even when spurious regression bias is severe. Given inconsistent standard errors, spurious regression bias is not avoided with large samples. Accounting for spurious regression bias, we find that seven of the 17 *t*-statistics and regression *R*-squares from previous studies of predictive regressions that would be significant by standard criteria, are no longer significant.

6.2. Spurious Regression and Data Mining

We now consider the interaction between spurious regression and data mining in the predictive regressions, where the instruments to be mined are independent as in Foster, Smith, and Whaley (1997). There are L measured instruments over which the analyst searches for the "best" predictor, based on the *R*-squares of univariate regressions. In equation (5), Z_t becomes a vector of length L, where L is the number of instruments through which the analyst sifts. The error terms $\left(\varepsilon_t^*, \varepsilon_t\right)$ become an $L + 1$ vector with a diagonal covariance matrix; thus, ε^*_t is independent of ε_t.

The persistence parameters in equation (5) become an $(L + 1)$-square, diagonal matrix, with the autocorrelation of the true predictor equal to ρ^*. The value of ρ^* is either the average

[7] We conduct several experiments letting the number of lags examined be 24, 36, or 48, when $T = 66$ and $T = 824$. When $T = 66$ the critical *t*-ratios are always larger than the values in Table 2. When $T = 824$ the effects are small and of mixed sign. The most extreme reduction in a critical *t*-ratio, relative to Table 2, is with 48 lags, true $R^2 = 15$ percent, and $\rho^* = 0.99$, where the critical value falls from 4.92 to 4.23.

from our sample of 500 potential instruments, 15 percent, or the median value from Table 1, 95 percent. The remaining autocorrelations, denoted by the L-vector, ρ, are set equal to the autocorrelations of the first L instruments in our sample of 500 potential instruments.[8] When $\rho^* = 95$ percent, we rescale the autocorrelations to center the distribution at 0.95 while preserving the range in the original data.[9] The simulations match the unconditional variances of the instruments, $Var(Z)$, to the data. The first element of the covariance matrix Σ is equal to σ^{*2}. For a typical *i-th* diagonal element of Σ, denoted by σ_i, the elements of $\rho(Z_i)$ and $Var(Z_i)$ are matched to the data, and we set $\sigma_i^2 = [1 - \rho(Z_i)^2]Var(Z_i)$.

Table 4 summarizes the results. The columns correspond to different numbers of potential instruments, through which the analyst sifts to find the regression that delivers the highest sample R-squared. The rows refer to the different values of the true R-squared.

[Table 4 about here]

The rows with true $R^2 = 0$ refer to data mining only, similar to Foster, Smith and Whaley (1997). The columns where $L = 1$ correspond to pure spurious regression bias. We hold fixed the persistence parameter for the true expected return, ρ^*, while allowing ρ to vary depending on the measured instrument. When $L = 1$, we set $\rho = 15$ percent. We consider two values for ρ^*, 15 percent or 95 percent.

---

[8] We calibrate the true autocorrelations in the simulations to the sample autocorrelations, adjusted for first-order finite-sample bias as: $\hat{\rho} + (1 + 3\hat{\rho})/T$, where $\hat{\rho}$ is the OLS estimate of the autocorrelation and T is the sample size.

[9] The transformation is as follows. In the 500 instruments, the minimum bias-adjusted autocorrelation is -0.571, the maximum is 0.999, and the median is 0.02. We center the transformed distribution about the median in Table 1, which is 0.95. If the original autocorrelation ρ is less than the median, we transform it to:

$$0.95 + (\rho\text{-}0.02)\{(0.95+0.571)/(0.02+0.571)\}.$$

If the value is above the median, we transform it to:

$$0.95 + (\rho\text{-}0.02)\{(0.999\text{-}0.95)/(0.999\text{-}0.02)\}.$$

Panels A and B of Table 4 show that when $L = 1$ (there is no data mining) and $\rho^* = 15$ percent, there is no spurious regression problem, consistent with Table 2. The empirical critical values for the *t*-ratios and *R*-squared statistics are close to their theoretical values under normality. For larger values of L (there is data mining) and $\rho^* = 15$ percent, the critical values are close to the values reported by Foster, Smith, and Whaley (1997) for similar sample sizes.[10] There is little difference in the results for the various true *R*-squares. Thus, with little persistence in the true expected return there is no spurious regression problem, and no interaction with data mining.

Panels C and D of Table 4 tell a different story. When the underlying expected return is persistent ($\rho^* = 0.95$) there is a spurious regression bias. When $L = 1$ we have spurious regression only. The critical *t*-ratio in Panel C increases from 2.3 to 2.8 as the true *R*-squared goes from zero to 15 percent. The bias is less pronounced here than in Table 2, with $\rho = \rho^* = 0.95$, which illustrates that for a given value of ρ^*, spurious regression is worse for larger values of ρ.

Spurious regression bias interacts with data mining. Consider the extreme corners of Panel C. Whereas, with $L = 1$ the critical *t*-ratio increases from 2.3 to 2.8 as the true *R*-squared goes from zero to 15 percent, with $L = 250$, the critical *t*-ratio increases from 5.2 to 6.3 as the true *R*-squared is increased. Thus, data mining magnifies the effects of the spurious regression bias. When more instruments are examined, the more persistent ones are likely to be chosen, and the spurious regression problem is amplified. The slope coefficients are centered near zero, so the bias does not increase the average slopes of the selected regressors. Again, spurious regression works through the standard errors.

We can also say that spurious regression makes the data mining problem worse. For a given value of L the critical *t*-ratios and R^2 values increase moving down the rows of Table 4.

---

[10] Our sample sizes, T, are not the same as in Foster, Smith, and Whaley (1997). When we run the experiments for their sample sizes, we closely approximate the critical values that they report.

For example, with $L = 250$ and true $R^2 = 0$, we can account for pure data mining with a critical t-ratio of 5.2. However, when the true R-squared is 15 percent, the critical t-ratio rises to 6.3. The differences moving down the rows are even greater when $T = 824$, in Panel D. Thus, in the situations where the spurious regression bias is more severe, its impact on the data mining problem is also more severe.

Finally, Panel E of Table 4 revisits the studies from the literature in view of spurious regression and data mining. We report critical values for L, the number of instruments mined, sufficient to render the regression t-ratios and R-squares insignificant at the five percent level. We use two assumptions about persistence in the true expected returns: (1) ρ^* is set equal to the sample values from the studies, as in Table 1, or (2) $\rho^* = 95$ percent. With only one exception, the critical values of L are 10 or smaller. The exception is where the instrument is the lagged one-month excess return on a two-month Treasury-bill, following Campbell (1987). This is an interesting example because the instrument is not very autocorrelated, at eight percent, and when we set $\rho^* = 0.08$ there is no spurious regression effect. The critical value of L exceeds 500. However, when we set $\rho^* = 95$ percent in this example, the critical value of L falls to 10, illustrating the strong interaction between the data mining and spurious regression effects.

7. Results for Conditional Asset Pricing Models

7.1. Cases with Small Amounts of Persistence

We first consider a special case of the model where we set $\rho^* = 0$ in the data generating process for the market return and true beta, so that Z^* is white noise and $\sigma^2(\varepsilon^*) = Var(Z^*)$. In this case the predictable (but unobserved by the analyst) component of the stock market return and the betas follow white noise processes. We allow a range of values for the autocorrelation, ρ, of the measured instrument, Z, including values as large as 0.99. For a given value of ρ, we choose $\sigma^2(\varepsilon) = Var(Z^*)(1 - \rho^2)$, so the measured instrument and the unobserved beta have the same variance. We find in this case that the critical values for all of the coefficients are well behaved. Thus, when the true expected returns and betas are not persistent, the use of even a highly persistent regressor does not create a spurious regression bias in the asset pricing regressions of equation (3).

It seems intuitive that there should be no spurious regression problem when there is no persistence in Z^*. Since the true coefficient on the measured instrument, Z, is zero, the error term in the regression is unaffected by the persistence in Z under the null hypothesis. When there is no spurious regression problem there can be no interaction between spurious regression and data mining. Thus, standard corrections for data mining (e.g. White, 2000) can be used without concern in these cases.

In our second experiment the measured instrument and the true beta have the same degree of persistence, but their persistence is not extreme. We fix $Var(Z) = Var(Z^*)$ and choose, for a given value of $\rho^* = \rho$, $\sigma^2(\varepsilon) = \sigma^2(\varepsilon^*) = Var(Z^*)(1 - \rho^2)$. For values of $\rho < 0.95$ and all values of the true predictive R-squared, R_p^2 the regressions seem generally well-specified, even at sample sizes as small as $T = 66$. These findings are similar to the findings for the predictive regression (1). Thus, the asset pricing regressions (3) also appear to be well specified when the autocorrelation of the true predictor is below 0.95.

7.2. Cases with Persistence

Table 5 summarizes simulation results for a case that allows data mining and spurious regression. In this experiment, the true persistence parameter ρ^* is set equal to 0.95. The table summarizes the results for time-series samples of $T = 66$, $T = 350$ and $T = 960$. The number of variables over which the artificial agent searches in mining the data, ranges from one to 250. We focus on the two abnormal return coefficients, $\{\alpha_0, \alpha_1\}$ and on the time-varying beta coefficient, b_1.

[Table 5 about here]

Table 5 shows that the means of the coefficient α_0, the fixed part of the alpha, are close to zero, and they get closer to zero as the number of observations increases, as expected of a consistent estimator. The 5% critical t-ratios for α_0 are reasonably well specified at the larger sample sizes, although there is some bias at $T = 66$, where the critical values rise with the extent of data mining. Data mining has little effect on the intercepts at the larger sample sizes. Since the lagged instrument has a mean of zero, the intercept is the average conditional alpha. Thus, the issue of data mining for predictive variables appears to have no serious implications for measures of average abnormal performance in the conditional asset pricing regressions, provided $T > 66$. This justifies the use of such models for studying the cross-section of average equity returns.

The coefficients α_1, which represent the time-varying part of the conditional alphas, present a different pattern. We would expect a data mining effect, given that the data are mined based on the coefficients on the lagged predictor in the simple predictive regression. The presence of the interaction term, however, would be expected to attenuate the bias in the standard errors, compared with the simple predictive regression. The table shows only a small effect of data mining on the α_1 coefficient, but a large effect on its t-ratio. The overall effect is the greatest at the smaller sample size ($T = 66$), where the critical t-ratios for the intermediate R_p^2 values (10% predictive R^2) vary from about 2.4 to 5.2 as the number of variables mined increases from one to 250. The bias diminishes with T, especially when the number of mined variables is small, and for $L = 1$ there is no substantial bias at $T = 360$ or $T = 960$ months.

The results on the α_1 coefficient are interesting in three respects. First, the critical t-ratios vary by only small amounts across the rows of the table. This indicates very little interaction between the spurious regression and data mining effects. Second, the table shows a smaller data mining effect than observed on the pure predictive regression. Thus, standard data mining corrections for predictive regressions will overcompensate in this setting. Third, the critical t-ratios for α_1 become smaller in Table 5 as the sample size is increased. This is just the opposite of what is found for the simple predictive regressions, where the inconsistency in the standard errors makes the critical t-ratios larger at larger sample sizes. Thus, the sampling distributions for time-varying alpha coefficients are not likely to be well approximated by simple corrections.[11]

Table 5 does not report the t-statistics for b_0, the constant part of the beta estimate. These are generally unbiased across all of the samples, except that the critical t-ratios are slightly inflated at the smaller sample size ($T = 66$) when data mining is not at issue ($L = 1$).

Finally, Table 5 shows results for the b_1 coefficients and their t-ratios, which capture the time-varying component of the conditional betas. Here, the average values and the critical t-ratios are barely affected by the number of variables mined. When $T = 66$ the critical t-ratios stay in a narrow range, from about 2.5 to 2.6, and they cluster closely around a value of 2.0 at the larger sample sizes. There are no discernible effects of data mining on the distribution of the time-varying beta coefficients except when the R^2 values are very high. This is an important result in the context of the conditional asset pricing literature, which we characterize as having mined predictive variables based on the regression (1). Our results suggest that the empirical evidence in this literature for time-varying betas, based on the regression model (3), is relatively robust to the data mining.

7.3. Suppressing Time-Varying Alphas

Some studies in the conditional asset pricing literature use regression models with interaction terms, but without the time-varying alpha component (e.g. Cochrane (1996), Ferson and Schadt (1996), Ferson and Harvey, 1999). Since the time-varying alpha component is the

[11] We conducted some experiments in which we applied a simple local-to-unity correction to the t-ratios, dividing by the square root of the sample size. We found that this correction does not result in a t-ratio that is approximately invariant to the sample size.

most troublesome term in the presence of spurious regression and data mining effects, it is interesting to ask if regressions that suppress this term may be better specified. Table 6 presents results for models in which the analyst runs regressions without the α_1 coefficient. The results suggest that the average alpha coefficient, α_0, and its t-statistic remain well specified regardless of data mining and potential spurious regression. Thus, once again we find little cause for concern about the inferences on average abnormal returns using the conditional asset pricing regressions, even though they use persistent, data mined lagged regressors.

The distribution of the average beta estimate, b_0, is not shown in Table 6. The results are similar to those obtained in a factor model regression where no lagged instrument is used. The coefficients and standard errors generally appear well specified. However, we find that the coefficient measuring the time-varying beta is somewhat more susceptible to bias than in the regression that includes α_1. The b_1 coefficient is biased, especially when $T = 66$, and its mean varies with the number of instruments mined. The critical t-ratios are inflated at the higher values of R_p^2 and when more instruments are mined.

[Table 6 about here]

These experiments suggest that including the time-varying alpha in the regression (3) helps “soak up” the bias so that it does not adversely effect the time varying beta estimate. We conclude that if one is interested in obtaining good estimates of conditional betas, then in the presence of potential data mining and persistent lagged instruments, the time-varying alpha term should be included in the regression.

7.4. Suppressing Time-Varying Betas

There are examples in the literature where the regression is run with a linear term for a time-varying conditional alpha but no interaction term for a time varying conditional beta (e.g.

Jagannathan and Wang, 1996). Table 7 considers this case.

[Table 7 about here]

First, the coefficient for the average beta in the regression with no b_1 term (not shown in the table) is reasonably well specified and largely unaffected by data mining on the lagged instrument. We find that the coefficients for alpha, α_0 and α_1, behave similarly to the corresponding coefficients in the full regression model (3). The estimates of the average alpha are reasonably well behaved, and only mildly affected by the extent of data mining at smaller sample sizes. The bias in α_1 is severe. The bias leads the analyst to overstate the evidence for a time-varying alpha, and the bias is worse as the amount of data mining increases. Thus, the evidence in the literature for time-varying alphas, based on these asset-pricing regressions, is likely to be overstated.

7.5. A Cross-Section of Asset Returns

We extend the simulations to study a cross-section of asset returns. We use five book-to-market (BM) quintile portfolios, equally weighted across the size dimension, as an illustration. The data are courtesy of Kenneth French. In these experiments the cross-section of assets features cross-sectional variation in the true conditional betas. Instead *of* $\beta_t = 1 + Z_t^*$, the betas are $\beta_t = \beta_0 + \beta_1 Z_t^*$, where the coefficients β_0 and β_1 are the estimates obtained from regressions of each quintile portfolio's excess return on the market portfolio excess return and the product of the market portfolio with the lagged value of the dividend yield. The set of β_0's is {1.259, 1.180, 1.124, 1.118, 1.274}, the set of β_1's is {-1.715, 1.000, 3.766, 7.646, 8.970}.[12]

[12] The β_1 coefficient for the BM2 portfolio is 1.0, replacing the estimated value of 0.047. When the β_1 coefficient is 0.047 the simulated return becomes nearly perfectly correlated with r_m and the simulation is uninformative. The dividend yield is demeaned and multiplied by 10. The dividend yield has the largest average sample correlation with the five BM portfolios among the standard instruments we examine.

The true predictive *R*-squared in the artificial data generating process is set to 0.5 percent. This value matches the smallest R-squared from the regression of the market portfolio on the lagged dividend yield with a window of 60 months.

Table 8 shows simulation results for the conditional model with time-varying alphas and betas. The means of the b_0 and b_1 coefficients are shown in excess of their true values in the simulations. The critical t-statistics for both α_1 and b_1 are generally similar to the case where $R_p^2 = 0.5$ percent in Table 5. As before, there is a large bias in the t-statistic for α_1 that increases with data mining but decreases somewhat with the sample size. The t-statistics for the time-varying betas are generally well specified.

[Table 8 about here]

We conduct additional experiments using the cross section of asset returns, where the conditional asset pricing regression suppresses either the time-varying alphas or the time-varying betas. The results are similar to those in Table 8. When the time-varying betas are suppressed there is severe bias in α_1 that diminishes somewhat with the sample size. When time-varying alphas are suppressed there is a mild bias in b_1.

7.6. Revisiting Previous Evidence

In this section, we explore the impact of the joint effects of data mining and spurious regression bias on the asset pricing evidence based on regression (3). First, we revisit the studies listed in Table 2. Consider the models with both time-varying alphas and betas. If the data mining searches over 250 variables predicting the test asset return and $T = 350$, the 5% cut-off value to apply to the t-statistic for α_1 is larger than 3.8 in absolute value. For smaller sample sizes, the cut-off value is even higher. Note from Table 2 that the largest t-statistic for α_1 in Shanken (1990) with a sample size of 360 is -3.57 on the T-bill volatility, while the largest t-statistic for α_1 in Christopherson, Ferson and Glassman (1998) with a sample size of

144 is 3.72 on the dividend yield. This means that the significance of the time-varying alphas in both of these studies is questionable. However, the largest t-statistic for b_1 in Shanken (1990) exceeds the empirical 5% cut-off, irrespective of spurious regression and data mining adjustments. This illustrates that the evidence for time-varying beta is robust to the joint effects of data mining and spurious regression bias, while the evidence for time-varying alphas is fragile.

Now consider the model with no time-varying alpha. If the data mining searches over 250 variables to predict the test asset return, the 5% cut-off value to apply to the t-statistic on b_1 is less than 3.5 in absolute value. Cochrane (1996) reports a t-statistic of -4.74 on the dividend yield in a time-varying beta, with a sample of $T = 186$. Thus, we find no evidence to doubt the inference that there is a time-varying beta. (However, the significance of the term premium in the time-varying beta, with a t-statistic of -1.76, is in doubt at the 10% level.)

Finally, consider the model with no time-varying beta. If the data mining searches over 25 variables to predict the test asset return, then the 5% cut-off value to apply to the t-statistic on α_1 is larger than 3.1 in absolute value. The largest t-statistic in Jagannathan and Wang (1996) with a sample size of 330 is 3.1. Therefore, their evidence for a time-varying alpha does not survive even with a modest amount of data mining.

We conclude that some aspects of the conditional asset pricing regression (3) are robust to data mining over persistent predictor variables, while others are not. The regression delivers reliable estimates of average abnormal returns and betas. However, the estimates of time-varying alphas may have vastly overstated statistical significance when standard tests are used.

8. Solutions to the problems of Spurious Regression and Data Mining

8.1. Solutions in Predictive Regressions

The essential problem in dealing with the spurious regression bias is to get the right standard errors. We examine the Newey-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 footnote 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 9 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 consistent 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 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 find that in realistic sample sizes, more lags do not help the spurious regression problem. The fourth column of Table 9 (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.

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 9 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 two 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 doesn't work very well is 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.[13]

[Table 9 about here]

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

---

[13] More formally, consider a case where the *ex ante* return is an AR(1) process, in 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.

(7) $$X_t = Z_t - (1/\tau) \Sigma_{j=1,\ldots,\tau} Z_{t-j}.$$

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 9. 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.

8.2. Solutions in Conditional Asset Pricing Models

Since detrending works relatively well in simple predictive regressions, one would think of using it also in conditional asset pricing tests to correct the inflated t-statistics on the time-varying alpha coefficient. However, as we observed above, the bias in the t-statistic on a_1 is largely due to data mining rather than spurious regression. As a result, high t-statistics on a_1 for large number of data mining searches come not from the high autocorrelation of Z_t but rather from its high cross-correlation with the asset return. Therefore, simple detrending does not work in this case because chosen instruments may not necessarily have high persistency.[14]

9. Robustness of the Asset Pricing Results

This section summarizes the results of a number of additional experiments. We extend the

[14] Note that the more the asset return volatility resembles that of the entire market (i.e., if it is lower than in our simulations) the higher is the likelihood of finding more evidence of spurious regression bias. Then a simple detrending will help adjusting the t-statistics of both a_1 and b_1 just as it did in the predictive regression case.

simulations of the asset pricing models to consider examples with more than a single lagged instrument. We consider asset pricing models with multiple factors, motivated by Merton's (1973) multiple-beta model. We also examine models where the data mining to select the lagged instruments focuses on predicting the market portfolio return instead of the test asset returns.

9.1. Multiple Instruments

The experiments summarized above focus on a single lagged instrument, while many studies in the literature use multiple instruments. We modify the simulations, assuming that the researcher mines two independent instruments with the largest absolute t-statistics and then uses both of them in the conditional asset pricing regression (3) with time-varying betas and alphas. (Thus, there are two a_1 coefficients and two b_1 coefficients.) These simulations reveal that the statistical behavior of both coefficients are similar to each other and similar to our results as reported in Table 5.

9.2. Multiple-Beta Models

We extend the simulations to study models with three state variables or factors. In building the three-factor model, we make the following assumptions. All three risk premiums are linear functions of one instrument, Z^*. The factors differ in their unconditional means and their disturbance terms, which are correlated with each other. The variance-covariance matrix of the disturbance terms matches that of the residuals from regressions of the three Fama-French (1993, 1996) factors on the lagged dividend yield. The true coefficients for the asset return on all three factors and their interaction terms with the correct lagged instrument, Z^*, are set to unity. Thus, the true conditional betas on each factor are equal to $1 + Z^*$. We find that the bias in the *t*-statistic for α_1 remains and is similar to the simulation in Table 5. There are no biases in the *t*-statistics associated with the b_1's for the larger sample sizes.

9.3. Predicting the Market Return

Much of the previous literature looked at more than one asset to select predictor variables. For the examples reported in the previous tables, the data mining is conducted by attempting to predict the excess returns of the tests assets. But a researcher might also choose instruments to predict the market portfolio return. We examine the sensitivity of the results to this change in the simulation design. The results for the conditional asset pricing model with both time-varying alphas and betas are re-examined. Recall that when the instrument is mined to predict the test asset return, there is an upward bias in the t-statistic for α_1. The bias increases with data mining and decreases somewhat with T. When the instruments are mined to predict the market, the bias in α_1 is small and is confined to the smaller sample size, $T = 66$. Mining to predict the market return has little impact on the sampling distribution of b_1.

9.4. Simulations under the Alternative Hypothesis

Note that the return generating process (6) does not include an intercept or alpha, consistent with asset pricing theory. Thus, the data are generated under the null hypothesis that an asset pricing model holds exactly. However, no asset pricing model is likely to hold exactly in reality. We therefore conduct experiments in which the data generating process allows for a nonzero alpha. We modify equation (6) as follows:

(8)
$$r_{t+1} = a_1 Z_t^* + \beta_t \, r_{m,t+1} + u_{t+1},$$
$$\beta_t = 1 + Z_t^*,$$
$$r_{m,t+1} = \mu + k \, Z^*_t + w_{t+1}.$$

In the system (8), there is a time-varying alpha, proportional to Z_t^*. We set the coefficient $a_1 = k$ and estimate the model (3) again. With this modification, the bias in the time-varying alpha coefficient , α_1, is slightly worse at the larger R^2 values and larger values of L than it was before. The overall patterns, including the reduction in bias for larger values of T, are similar. We also run the model with no time-varying beta, and the results are similar to those reported above for that case.

10. 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 two 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, 2003). 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.

We also study regression models for conditional asset pricing models in which lagged variables are used to model conditional betas and alphas. The conditional asset pricing literature has, for the most part used the same variables that were discovered based on simple predictive regressions, and our analysis characterizes the problem by assuming the data mining occurs in this way. Our results relate to several stylized facts that the literature on conditional asset pricing has produced.

Previous studies find evidence that the intercept, or average alpha, is smaller in a conditional model than in an unconditional model, suggesting, for example, that the conditional CAPM does a better job of explaining average abnormal returns. Our simulation evidence finds that the estimates of the average alphas in the conditional models are reasonably well specified in the presence of spurious regression and data mining, at least for samples larger than $T = 66$. Some caution should be applied in interpreting the common 60-month rolling regression estimator, but otherwise we take no issue with the stylized fact that conditional models deliver smaller average alphas.

Studies typically find evidence of time varying betas based on significant interaction terms. Here again we find little cause for concern. The coefficient estimator for the interaction term is well specified in larger samples, and largely unaffected by data mining in the presence of persistent lagged regressors. There is an exception when the model is estimated without a linear term in the lagged instrument. In this case, the coefficient measuring the time-varying beta is slightly biased. Thus, when the focus of the study is to estimate accurate conditional betas, we recommend that a linear term be included in the regression model.

Studies also find that even conditional models fail to explain completely the dynamic properties of stock returns. That is, the estimates indicate time-varying conditional alphas. We find that this result is the most problematic. The estimates of time variation in alpha inherit biases similar to, if somewhat smaller than, the biases in predictive regressions. We use our simulations to revisit the evidence of several prominent studies. Our analysis suggests that the evidence for time-varying alphas in the current literature should be viewed with some suspicion. Perhaps, the current generation of conditional asset pricing models do a better job of capturing the dynamic behavior of asset returns than existing studies suggest.

Finally, we think that our study, as summarized in this chapter, represents the beginning of what could be an important research direction at the nexus of econometrics and financial economics. The literature in this area has arrived at a good understanding of a number of econometric issues in asset pricing research; the two issues that we take on are only part of a much longer list that includes stochastic regressor bias, unit roots, cointegration, overlapping data, time aggregation, structural change, errors-in-variables, and many more. But what is less understood is how these econometric issues interact with each other. We have seen that the interaction of data mining and spurious regression is likely to be a problem of practical importance. Many of other econometric issues also occur in combination in our empirical practice. We need to study these other interactions in future research.

References:

Breen, W., L.R. Glosten., and R. Jagannathan, 1999, "Economic Significance of Predictable Variations in Stock Index returns," *Journal of Finance* 44, 1177-1190.

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

Chan, C.K., and N.F. Chen, 1988, "An Unconditional Asset Pricing Test and the Role of Size as an Instrumental Variable for Risk," *Journal of Finance* 43, 309-325.

Christopherson, J.A., W.E. Ferson, and D. Glassman, 1988, "Conditioning Manager Alpha on Economic Information: Another Look at the Persistence of Performance," *Review of Financial Studies* 11, 111-142.

Cochrane, J.H., 1996, "A Cross-Sectional Test of an Investment-Based Asset Pricing Model," *Journal of Political Economy* 104, 572-621.

Cochrane, J.H., 1999, "New Facts in Finance," Working Paper, University of Chicago.

Conrad, J. and G. Kaul, 1988, "Time-Variation in Expected Returns," *Journal of Business* 61, 409-425.

Cox, J.C., J.E. Ingersoll Jr., and S.A. Ross, 1985, "A Theory of the Term Structure of Interest Rates," *Econometrica* 53, 363-384.

Fama, E.F., 1970, "Efficient Capital Markets: A Review of Theory and Empirical Work," *Journal of Finance* 25, 383-417.

Fama, E.F., L. Fisher, M.C. Jensen, and R. Roll. 1969, "The Adjustment of Stock Prices to New Information," *International Economic Review* 10, 1-21.

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

Fama, E.F., and K.R. French., 1988b, "Permanent and Temporary Components of Stock Prices," *Journal of Political Economy* 96, 246-273.

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

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

Ferson, W.E., and C.R. Harvey, 1991, "Sources of Predictability in Portfolio Returns," *Financial Analysts Journal* 3, 49-56

Ferson, W.E., and C.R. Harvey, 1997, "The Fundamental Determinants of International Equity Returns: A Perspective on Country Risk and Asset Pricing," *Journal of Banking and Finance* 21, 1625-1665.

Ferson, W.E., and C.R. Harvey, 1999, "Conditioning Variables and the Cross-Section of Stock Returns," *Journal of Finance* 54, 1325-1360.

Ferson, W.E., and R. Korajczyk, 1995, "Do Arbitrage Pricing Models Explain the Predictability of Stock Returns?" *Journal of Business* 68, 309-349.

Ferson, W.E., and R. Schadt, 1996, "Measuring Fund Strategy and Performance in Changing Economic Conditions," *Journal of Finance* 51, 425-462.

Fleming, J., C. Kirby, and B. Ostdiek, 2001, "The Economic Value of Volatility Timing," *Journal of Finance* 61, 329-352.

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

Gibbons, M.R., and W.E. Ferson, 1995, "Testing Asset Pricing Models With Changing Expectations and an Unobservable Market Portfolio," *Journal of Financial Economics* 14, 216-236.

Goyal, A., and I. Welch, 2003, "Predicting the Equity Premium with Dividend Ratios," *Management Science* 49, 639–654.

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

Harvey, C.R., 1989, "Time-Varying Conditional Covariances in Tests of Asset Pricing Models," *Journal of Financial Economics* 24, 289-318.

Hastie, T., R. Tibshirani, and J. Friedman, 2001, *The Elements of Statistical Learning*, Springer.

Huberman, G., and S. Kandel, 1990, Market effciency and Value Line's record, Journal of Business 63, 187-216.

Jagannathan, R., and Z. Wang., 1996, "The Conditional CAPM and the Cross-Section of Expected Returns," *Journal of Finance* 51, 3-54.

Kandel, S., and R.F. Stambaugh, 1990, "Expectations and Volatility of Consumption and Asset Returns," *Review of Financial Studies* 3, 207-232.

Kendall, M.G., 1954, "A Note on the Bias in the Estimation of Autocorrelation," *Biometrica*

41, 403-404.

Keim, D.B., and R.F. Stambaugh, 1996, "Predicting Returns in the Bond and Stock Markets," *Journal of Financial Economics* 17, 357-390.

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

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

Lee, C., J. Myers, and B. Swaminathan, 1999, "What is the Intrinsic Value of the Dow?" *Journal of Finance* 54, 1693-1742.

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

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

Lucas, R.E. Jr., 1978, "Asset Prices in an Exchange Economy," *Econometrica* 46, 1429-1445.

Maddala, G.S. *Econometrics*, McGraw-Hill, New York, 1977.

Marmol, F., 1998, "Spurious Regression Theory with Non-Stationary Fractionally Integrated Processes," *Journal of Econometrics* 84, 233-250.

Merton, R.C., 1973, "An Intertemporal Capital Asset Pricing Model," *Econometrica* 41, 867-887.

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

Petkova, R. and L. Zhang, 2005, "Is Value Riskier Than Growth?" *Journal of Financial Economics* 78, 187-202.

Phillips, P.C.B., 1986, "Understanding Spurious Regressions in Econometrics," *Journal of Econometrics* 33, 311-340.

Phillips, P.C.B., 1998, "New Tools for Understanding Spurious Regressions," *Econometrica* 66, 1299-1326.

Pontiff, J., and L. Schall, 1998, "Book-to-Market as a Predictor of Market Returns," *Journal of Financial Economics* 49, 141-60.

Shanken, J., 1990, "Intertemporal Asset Pricing: An Empirical Investigation," *Journal of*

Econometrics 45, 99-120.

Sharpe, W.F., 1964, "Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk," *Journal of Finance* 19, 425-442.

Schwert, G.W., 2002, "Anomalies and market efficiency," in G.M. Constantinides, M. Harris, R.M. Stulz, eds.: *Handbook of the Economics of Finance*, North Holland: Amsterdam.

Stambaugh, R.S., 1999, "Predictive Regressions," *Journal of Financial Economics* 54, 315-421.

White, H., 2000, "A Reality Check for Data Snooping," *Econometrica* 68, 1097-1126.

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.

Table 1

Common Instrumental Variables: Sources, 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 instrument. The next two columns give the sample (*Period*) and the number of observations (*Obs*) on the stock returns. Columns five and six report the autocorrelation (ρ_Z) and the standard deviation of the instrument (σ_Z), respectively. The next three columns report regression results for Standard & Poors 500 excess return on a lagged instrument. The slope coefficient is β, the t-statistic is t, and the coefficient of determination is R^2. The last column (HAC) reports the method used in computing the standard errors of the slopes. The method of Newey-West (1987) is used with the number of lags given in parentheses. 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. The yield on all corporate bonds is denoted as *ALLy*. The yield on AAA rated corporate bonds is *AAAy* and *UBAAy* is the yield on corporate bonds with a below BAA rating. The variable "*Cay"* is the linear function of consumption, asset wealth, and labor income. The book-to-market ratios for the Dow Jones Industrial Average and the S&P500 are respectively *DJBM* and *SPBM*.

(1) Reference	**(2) Predictor**	**(3) Period**	**(4) Obs**	**(5)** ρ_Z	**(6)** σ_Z	**(7)** β	**(8) t**	**(9)** R^2	**(10) HAC**
Breen, Glosten & Jagannathan (89)	TB1y	5404-8612	393	0.97	0.0026	-2.49	-3.58	0.023	NW(5)
Campbell (87)	Two–one	5906-7908	264	0.32	0.0006	11.87	2.38	0.025	NW(0)
	Six–one	5906-7908	264	0.15	0.0020	2.88	2.13	0.025	NW(0)
	Lag(two) – one	5906-7908	264	0.08	0.0010	9.88	2.67	0.063	NW(6)
Fama (90)	ALLy–AAAy	5301-8712	420	0.97	0.0040	0.88	1.46	0.005	MA(0)
Fama & French (88a)	Dividend yield	2701-8612	720	0.97	0.0013	0.40	1.36	0.007	MA(9)
Fama & French (89)	AAAy–TB1y	2601-8612	732	0.92	0.0011	0.51	2.16	0.007	MA(9)
Keim & Stambaugh (86)	UBAAy	2802-7812	611	0.95	0.0230	1.50	0.75	0.002	MA(9)
	UBAAy–TB1y	2802-7812	611	0.97	0.0320	1.57	1.48	0.007	MA(9)
Kothari & Shanken (97)	DJBM	1927-1992	66	0.66	0.2270	0.28	2.63	0.078	MA(0)
Lettau & Ludvigson (01)	"Cay"	52Q4-98Q4	184	0.79	0.0110	1.57	2.58	0.057	MA(7)
Pontiff & Schall (98)	DJBM	2602-9409	824	0.97	0.2300	2.96	2.16	0.012	MA(9)
	SPBM	5104-9409	552	0.98	0.0230	9.32	1.03	0.001	MA(5)

Table 2

Representative Studies on Conditional Asset Pricing Models

This table summarizes representative conditional asset pricing and performance evaluation studies. The first column indicates the published study. The second column specifies the lagged instruments used. The next two columns give the sample (*Period*) and the number of observations (*Obs*) on the stock returns. Columns five and six indicate whether the conditional model includes time-varying alpha (α_t) and the time-varying beta (β_t), respectively. The last five columns summarize the regression results. Column seven shows the ratio of an intercept (a pricing error) in the conditional model to that of the unconditional. Columns 8 and 9 report the point estimates of the time-varying alpha, α_1, and their corresponding t-statistics, $t(\alpha_t)$. Columns 10 and 11 report the point estimates of the time-varying beta coefficient, b_1, and their corresponding t-statistics, $t(b_1)$. For each predictor, the table reports regression estimates corresponding to the largest in absolute value t-statistics. The abbreviations in the table are as follows. *TB1y* and *TB1vol* are the yield and volatility on the one-month Treasury bill, respectively. *Three-one* is the difference between the lagged returns of a three-month and a one-month T-bill. The variable "*Cay*" is the linear function of consumption, asset wealth, and labor income. *DY* is the dividend yield of the CRSP index. *Term* is a spread between long-term and short-term bonds. *Default* is a spread between low-grade and high-grade corporate bonds. "None" stands parameter values not used in the study, "NA" stands for results not reported.

(1) **Reference**	**(2)** **Predictor**	**(3)** **Period**	**(4)** **Obs**	**(5)** α_t	**(6)** β_t	**(7)** $\lvert\alpha_0/\alpha\rvert$	**(8)** α_1	**(9)** $t(\alpha_1)$	**(10)** b_1	**(11)** $t(b_1)$
Shanken (90)	TB1y TB1vol	5301-8212	360	Yes	Yes	NA	-0.48 -5.70	-1.17 -3.56	1.42 -8.40	5.92 -4.42
Cochrane (96)	DY Term	47Q1-93Q4	188	No	Yes	NA	None	None	-0.53 -0.31	-4.74 -1.76
Ferson & Schadt (96)	TB1y DY Term Default	6801-9012	276	No	Yes	0.72	None	None	NA	NA
Jagannathan & Wang (96)	Default	6307-9012	330	Yes	No	1.53	-65.7	-3.10	None	None
Christopherson et al. (98)	TB1y DY Term	7901-9012	144	Yes	Yes	0.77	-0.21 1.22 -0.21	-2.01 3.72 -1.85	NA	NA
Lettau & Ludvigson (01)	"Cay"	63Q3-98Q3	144	Yes	Yes	0.84	NA	NA	NA	NA
Petkova & Zhang (05)	TB1y DY Term Default	2701-0112	900	No	Yes	0.97	None	None	NA	NA

Table 3

The Monte Carlo Simulation Results for Regressions with a Lagged Predictor Variable

The table reports the 97.5 percentile of the Monte Carlo distribution of 10,000 Newey-West *t*-statistics, the 95 percentile for the estimated coefficients of determination, and the average estimated slopes from the regression

$$r_{t+1} = \alpha + \delta Z_t + v_{t+1},$$

where r_{t+1} is the excess return, Z_t is the predictor variable, and t=1,...,T. The parameter $\rho*$ is the autocorrelation coefficient of the predictors, Z_t^* and Z_t. The $\boldsymbol{R^2}$ is the coefficient of determination from the regression of excess returns r_{t+1} on the unobserved, true instrument Z_t^*. Panel A depicts the results for T = 66 and Panel B for T = 824. Panel C gives the simulation results for the number of observations and the autocorrelations in Table 1. In Panel C, the true R^2 is set to 0.1. The theoretical critical R^2 is from the F-distribution.

Panel A: 66 Observations

R^2/ ρ*	**0**	**0.5**	**0.9**	**0.95**	**0.98**	**0.99**
			Means: δ			
0.001	-0.0480	-0.0554	-0.0154	-0.0179	-0.0312	-0.0463
0.005	-0.0207	-0.0246	-0.0074	-0.0088	-0.0137	-0.0193
0.010	-0.0142	-0.0173	-0.0055	-0.0066	-0.0096	-0.0129
0.050	-0.0055	-0.0075	-0.0029	-0.0037	-0.0040	-0.0042
0.100	-0.0033	-0.0051	-0.0023	-0.0030	-0.0026	-0.0021
0.150	-0.0024	-0.0040	-0.0020	-0.0026	-0.0020	-0.0012
			Critical t-statistics			
0.001	2.1951	2.3073	2.4502	2.4879	2.4746	2.4630
0.005	2.2033	2.3076	2.4532	2.5007	2.5302	2.5003
0.010	2.2121	2.3123	2.4828	2.5369	2.5460	2.5214
0.050	2.2609	2.3335	2.6403	2.7113	2.7116	2.6359
0.100	2.2847	2.3702	2.8408	2.9329	2.9043	2.7843
0.150	2.2750	2.3959	3.0046	3.1232	3.0930	2.9417
			Critical estimated R^2			
0.001	0.0593	0.0575	0.0598	0.0599	0.0610	0.0600
0.005	0.0590	0.0578	0.0608	0.0607	0.0616	0.0604
0.010	0.0590	0.0579	0.0619	0.0623	0.0630	0.0612
0.050	0.0593	0.0593	0.0715	0.0737	0.0703	0.0673
0.100	0.0600	0.0622	0.0847	0.0882	0.0823	0.0766
0.150	0.0600	0.0649	0.0994	0.1032	0.0942	0.0850

Table 3 (continued)

Panel B: 824 Observations

R^2/ ρ*	**0**	**0.5**	**0.9**	**0.95**	**0.98**	**0.99**
			Means: δ			
0.001	0.0150	0.0106	0.0141	0.0115	0.0053	-0.0007
0.005	0.0067	0.0049	0.0069	0.0055	0.0021	-0.0011
0.010	0.0048	0.0035	0.0052	0.0040	0.0014	-0.0012
0.050	0.0021	0.0017	0.0029	0.0021	0.0003	-0.0014
0.100	0.0015	0.0013	0.0023	0.0016	0.0001	-0.0014
0.150	0.0012	0.0011	0.0021	0.0014	-0.0000	-0.0014
			Critical t-statistics			
0.001	1.9861	2.0263	2.0362	2.0454	2.0587	2.0585
0.005	1.9835	2.0297	2.0429	2.1123	2.1975	2.2558
0.010	1.9759	2.0279	2.0655	2.1479	2.3578	2.4957
0.050	1.9878	2.0088	2.2587	2.5685	3.1720	3.7095
0.100	1.9862	2.0320	2.3758	2.7342	3.6356	4.4528
0.150	2.0005	2.0246	2.4164	2.8555	3.8735	4.9151
			Critical estimated R^2			
0.001	0.0046	0.0047	0.0047	0.0047	0.0049	0.0049
0.005	0.0046	0.0047	0.0048	0.0051	0.0056	0.0059
0.010	0.0046	0.0047	0.0050	0.0054	0.0065	0.0073
0.050	0.0046	0.0047	0.0066	0.0085	0.0132	0.0183
0.100	0.0047	0.0049	0.0084	0.0125	0.0220	0.0316
0.150	0.0046	0.0050	0.0104	0.0166	0.0308	0.0450

Panel C: Table 1 simulation

Obs	$\rho*$	Critical theoretical R^2	Critical t-statistic	Critical estimated R^2
393	0.97	0.0098	3.2521	0.0311
264	0.32	0.0146	2.0645	0.0151
264	0.15	0.0146	2.0560	0.0151
264	0.08	0.0146	2.0318	0.0146
420	0.97	0.0092	3.2734	0.0304
720	0.97	0.0053	3.2005	0.0194
732	0.92	0.0053	2.3947	0.0103
611	0.95	0.0063	2.8843	0.0167
611	0.97	0.0063	3.2488	0.0219
66	0.66	0.0586	2.4221	0.0656
184	0.79	0.0209	2.2724	0.0270
824	0.97	0.0047	3.1612	0.0173
552	0.98	0.0070	3.6771	0.0293

Table 4

The Monte Carlo Simulation Results of Regressions with Spurious Regression and Data Mining, with Independent Regressors

The table reports the 97.5 percentile of the Monte Carlo distribution of 10000 Newey-West *t*-statistics, the 95 percentile for the estimated coefficients of determination, and the average estimated slopes from the regression

$$r_{t+1} = \alpha + \delta Z_t + v_{t+1},$$

where r_{t+1} is the excess return, Z_t is the predictor variable, and t=1,...,T. The R^2 is the coefficient of determination from the regression of excess returns r_{t+1} on the unobserved, true instrument Z_t^*, which has the autocorrelation ρ*. The parameter L is the number of instruments mined, where the one with the highest estimated R^2 is chosen. Panels A and B depict the results for T = 66 and T = 824 respectively, when the autocorrelation of the true predictor, ρ* = 0.15. Panels C and D depict the results for T = 66 and T = 824, respectively, when the autocorrelation of the true predictor, ρ* = 0.95, the median autocorrelation in Table I. In Panel E, the true R^2 is set to 0.1 and the original distribution of instruments is transformed so that their median autocorrelation is set at 0.95. The left-hand-side of Panel E gives the critical L for the given number of observations and autocorrelation that is sufficient to generate critical *t*-statistics or R^2's in excess of the corresponding statistics in Table 1. The right-hand-side of Panel E gives the critical L that is sufficient to generate critical t-statistics or R^2's in excess of the corresponding statistics in Table 1 when ρ* = 0.95.

Panel A: 66 Observations; ρ* = 0.15

R^2/L	1	5	10	25	50	100	250
				Means: δ			
0	-0.0004	0.0002	-0.0002	0.0004	-0.0001	0.0001	0.0005
0.001	-0.0114	0.0044	-0.0069	0.0208	-0.0078	0.0012	0.0162
0.005	-0.0050	0.0017	-0.0017	0.0113	-0.0014	-0.0031	0.0109
0.010	-0.0035	0.0008	-0.0014	0.0076	-0.0002	-0.0011	0.0098
0.050	-0.0014	0.0004	-0.0004	0.0018	-0.0023	-0.0013	0.0063
0.100	-0.0009	0.0006	-0.0004	0.0014	-0.0013	-0.0007	0.0044
0.150	-0.0007	0.0007	-0.0002	0.0009	-0.0010	-0.0010	0.0035
				Critical t-statistics			
0	2.2971	3.2213	3.5704	4.1093	4.4377	4.8329	5.2846
0.001	2.2819	3.2105	3.5418	4.1116	4.4351	4.8238	5.2803
0.005	2.2996	3.2250	3.5466	4.1190	4.4604	4.7951	5.2894
0.010	2.2981	3.2109	3.5492	4.1198	4.4728	4.7899	5.2900
0.050	2.2950	3.2416	3.5096	4.0981	4.4036	4.8803	5.2527
0.100	2.3175	3.2105	3.5316	4.1076	4.4563	4.8772	5.2272
0.150	2.3040	3.2187	3.5496	4.0644	4.5090	4.8984	5.2948
				Critical estimated R^2			
0	0.0594	0.0974	0.1153	0.1387	0.1548	0.1738	0.1944
0.001	0.0589	0.0969	0.1149	0.1386	0.1546	0.1739	0.1944
0.005	0.0591	0.0972	0.1151	0.1383	0.1545	0.1734	0.1948
0.010	0.0592	0.0967	0.1158	0.1386	0.1544	0.1733	0.1950
0.050	0.0596	0.0970	0.1163	0.1390	0.1557	0.1738	0.1955
0.100	0.0608	0.0969	0.1165	0.1392	0.1570	0.1738	0.1954
0.150	0.0612	0.0975	0.1165	0.1397	0.1577	0.1745	0.1967

Table 4 (continued)

Panel B: 824 Observations; ρ* = 0.15

R^2/L	**1**	**5**	**10**	**25**	**50**	**100**	**250**
				Means: δ			
0	0.0000	0.0000	0.0000	0.0000	-0.0001	-0.0002	0.0000
0.001	-0.0004	0.0032	-0.0017	0.0000	-0.0028	-0.0058	0.0015
0.005	-0.0002	0.0012	-0.0004	0.0000	-0.0020	-0.0031	0.0007
0.010	-0.0001	0.0009	-0.0004	-0.0003	-0.0015	-0.0020	0.0004
0.050	-0.0001	0.0005	0.0000	-0.0005	-0.0006	-0.0009	0.0004
0.100	0.0000	0.0005	-0.0001	-0.0003	-0.0001	-0.0002	0.0003
0.150	0.0000	0.0003	-0.0003	-0.0003	0.0001	-0.0002	0.0002
				Critical t-statistics			
0	2.0283	2.5861	2.8525	3.1740	3.3503	3.5439	3.8045
0.001	2.0369	2.6000	2.8534	3.1785	3.3616	3.5443	3.7928
0.005	2.0334	2.6043	2.8565	3.1769	3.3625	3.5440	3.7906
0.010	2.0310	2.6152	2.8694	3.1782	3.3544	3.5477	3.7917
0.050	2.0272	2.6229	2.8627	3.1846	3.3450	3.5552	3.8039
0.100	2.0115	2.6304	2.8705	3.1807	3.3648	3.5673	3.8041
0.150	2.0044	2.6327	2.8618	3.1766	3.3691	3.5723	3.7965
				Critical estimated R^2			
0	0.0047	0.0079	0.0096	0.0116	0.0130	0.0145	0.0166
0.001	0.0047	0.0079	0.0096	0.0116	0.0130	0.0145	0.0166
0.005	0.0047	0.0080	0.0096	0.0116	0.0129	0.0145	0.0166
0.010	0.0047	0.0080	0.0096	0.0115	0.0129	0.0145	0.0166
0.050	0.0047	0.0081	0.0096	0.0116	0.0130	0.0145	0.0167
0.100	0.0047	0.0081	0.0097	0.0117	0.0131	0.0146	0.0168
0.150	0.0047	0.0082	0.0096	0.0117	0.0130	0.0146	0.0168

Table 4 (continued)

Panel C: 66 Observations; ρ* = 0.95

R^2/L	**1**	**5**	**10**	**25**	**50**	**100**	**250**
				Means: δ			
0	-0.0005	0.0002	0.0006	-0.0001	-0.0006	-0.0003	0.0017
0.001	-0.0140	0.0069	0.0212	-0.0105	-0.0134	-0.0112	0.0557
0.005	-0.0060	0.0042	0.0082	-0.0068	-0.0024	-0.0033	0.0240
0.010	-0.0042	0.0031	0.0051	-0.0029	-0.0018	-0.0027	0.0145
0.050	-0.0016	0.0006	0.0035	-0.0023	-0.0016	-0.0019	0.0012
0.100	-0.0010	-0.0002	0.0021	-0.0013	-0.0017	-0.0005	0.0028
0.150	-0.0007	-0.0005	0.0015	-0.0008	-0.0011	-0.0001	0.0013
				Critical t-statistics			
0	2.3446	3.3507	3.6827	4.1903	4.4660	4.9412	5.2493
0.001	2.3641	3.3547	3.6776	4.1756	4.5157	4.9201	5.2441
0.005	2.4030	3.3864	3.7013	4.1984	4.5625	4.9381	5.2760
0.010	2.3939	3.4197	3.7308	4.1952	4.6039	4.9718	5.3083
0.050	2.5486	3.5482	3.9676	4.4703	4.9512	5.2027	5.5539
0.100	2.6955	3.7336	4.1899	4.7485	5.2335	5.5027	5.9006
0.150	2.8484	3.9724	4.4329	4.9748	5.5547	5.8256	6.2563
				Critical estimated R^2			
0	0.0579	0.0974	0.1140	0.1374	0.1515	0.1689	0.1885
0.001	0.0587	0.0981	0.1143	0.1376	0.1518	0.1692	0.1884
0.005	0.0596	0.0987	0.1153	0.1385	0.1530	0.1699	0.1895
0.010	0.0604	0.1002	0.1166	0.1402	0.1543	0.1711	0.1910
0.050	0.0691	0.1113	0.1307	0.1552	0.1711	0.1859	0.2057
0.100	0.0802	0.1265	0.1508	0.1774	0.1952	0.2099	0.2307
0.150	0.0911	0.1451	0.1728	0.2021	0.2209	0.2370	0.2587

Table 3 (continued)

Panel D: 824 Observations; ρ* = 0.95

R^2/L	1	5	10	25	50	100	250
				Means: δ			
0	-0.0001	0.0000	0.0000	0.0000	0.0001	0.0002	0.0001
0.001	-0.0027	-0.0016	-0.0007	0.0005	0.0015	0.0072	0.0039
0.005	-0.0012	-0.0004	0.0003	0.0006	-0.0008	0.0029	0.0026
0.010	-0.0009	-0.0005	0.0000	0.0003	-0.0008	0.0013	0.0006
0.050	-0.0004	-0.0005	0.0001	-0.0002	0.0007	-0.0006	0.0001
0.100	-0.0003	-0.0002	-0.0001	-0.0003	0.0000	0.0001	-0.0004
0.150	-0.0003	0.0000	0.0000	-0.0002	0.0001	0.0002	-0.0002
				Critical t-statistics			
0	1.9807	2.6807	2.8535	3.1579	3.3640	3.5673	3.8103
0.001	1.9989	2.6876	2.8758	3.1745	3.3702	3.5792	3.8252
0.005	2.0406	2.7588	2.9269	3.2218	3.4497	3.6493	3.9075
0.010	2.1108	2.8538	3.0150	3.3500	3.5548	3.7836	4.0351
0.050	2.4338	3.3118	3.6292	4.1202	4.3685	4.6795	4.9741
0.100	2.6274	3.6661	4.0003	4.5660	4.9129	5.2567	5.6937
0.150	2.7413	3.8720	4.2048	4.8481	5.2200	5.5846	6.0420
				Critical estimated R^2			
0	0.0045	0.0080	0.0096	0.0113	0.0129	0.0145	0.0164
0.001	0.0046	0.0082	0.0097	0.0115	0.0130	0.0146	0.0167
0.005	0.0048	0.0086	0.0102	0.0121	0.0137	0.0153	0.0176
0.010	0.0050	0.0092	0.0108	0.0131	0.0146	0.0163	0.0187
0.050	0.0077	0.0145	0.0173	0.0216	0.0244	0.0273	0.0314
0.100	0.0113	0.0216	0.0264	0.0331	0.0374	0.0421	0.0482
0.150	0.0151	0.0293	0.0356	0.0446	0.0508	0.0568	0.0647

Panel E: Table 1 Simulation

Obs	ρ*	Critical L (t-statistic)	Critical L (R^2)	ρ*	Critical L (t-statistic)	Critical L (R^2)
393	0.97	2	1	0.95	4	2
264	0.32	2	5	0.95	1	1
264	0.15	2	5	0.95	1	1
264	0.08	5	>500	0.95	1	10
420	0.97	1	1	0.95	1	1
720	0.97	1	1	0.95	1	1
732	0.92	1	1	0.95	1	1
611	0.95	1	1	0.95	1	1
611	0.97	1	1	0.95	1	1
66	0.66	2	2	0.95	1	2
184	0.79	2	7	0.95	1	3
824	0.97	1	1	0.95	1	2
552	0.98	1	1	0.95	1	1

Table 5

Simulating a Conditional Asset Pricing Model

The table shows the results of 10,000 simulations of the estimates from the conditional asset pricing model, allowing for possible data mining of the lagged instruments. The regression model is:

$$r_{t+1} = \alpha_0 + \alpha_1 Z_t + b_0\, r_{m,t+1} + b_1\, r_{m,t+1} Z_t + u_{t+1}.$$

T is the sample size, L is the number of lagged instruments mined, R_p^2 is the true predictive R^2 in the artificial data generating process.

	T=66			T=350			T=960		
R_p^2	L=1	L=25	L=250	L=1	L=25	L=250	L=1	L=25	L=250
				Means: α_0					
0.001	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.005	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.01	-0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.05	-0.001	-0.001	-0.001	0.000	-0.001	0.000	0.000	0.000	0.000
0.1	-0.002	-0.002	-0.002	0.000	-0.001	-0.001	0.000	0.000	0.000
0.15	-0.002	-0.002	-0.002	-0.001	-0.001	-0.001	0.000	0.000	0.000
				Critical 5% t-statistics for α_0					
0.001	2.280	2.603	2.855	1.999	2.061	2.146	1.996	2.005	2.115
0.005	2.266	2.540	2.792	1.994	2.058	2.135	2.013	2.002	2.104
0.01	2.253	2.508	2.759	2.000	2.045	2.125	2.016	2.002	2.098
0.05	2.153	2.408	2.728	1.974	1.998	2.094	2.021	1.991	2.100
0.1	2.088	2.388	2.652	1.977	2.000	2.030	2.058	2.008	2.073
0.15	2.065	2.382	2.597	1.968	1.960	1.987	2.069	2.031	2.041
				Means: α_1					
0.001	0.001	0.007	-0.003	-0.001	-0.002	0.003	-0.001	-0.002	-0.001
0.005	0.001	0.005	-0.001	-0.001	-0.001	0.002	-0.001	-0.002	-0.001
0.01	0.001	0.005	0.000	-0.001	-0.001	0.002	-0.001	-0.002	-0.001
0.05	0.001	0.005	-0.002	-0.001	-0.001	0.002	-0.001	-0.001	-0.002
0.1	0.001	0.005	0.000	-0.001	-0.001	0.002	-0.001	-0.001	0.000
0.15	0.001	0.004	0.000	-0.001	0.000	0.002	-0.001	-0.001	0.001
				Critical 5% t-statistics for α_1					
0.001	2.392	3.992	5.305	2.023	3.240	3.891	1.910	3.097	3.748
0.005	2.390	3.961	5.252	2.024	3.206	3.874	1.905	3.092	3.719
0.01	2.387	3.912	5.198	2.025	3.198	3.872	1.902	3.091	3.712
0.05	2.412	3.924	5.237	2.039	3.172	3.855	1.902	3.062	3.706
0.1	2.426	3.912	5.163	2.036	3.159	3.837	1.912	3.040	3.690
0.15	2.423	3.913	5.086	2.024	3.155	3.785	1.909	3.024	3.629
				Means: b_1					
0.001	-0.041	-0.017	0.026	-0.003	-0.002	0.011	0.010	0.008	-0.009
0.005	-0.038	-0.019	0.061	-0.003	0.001	0.002	0.009	0.007	-0.005
0.01	-0.038	-0.012	0.055	-0.003	0.000	0.004	0.008	0.004	-0.006
0.05	-0.039	-0.014	0.077	-0.003	0.006	-0.003	0.006	0.001	0.001
0.1	-0.040	-0.018	0.058	-0.003	0.010	0.003	0.005	0.000	0.004
0.15	-0.041	-0.015	0.062	-0.003	0.014	0.007	0.004	0.001	0.000
				Critical 5% t-statistics for b_1					
0.001	2.576	2.534	2.634	2.122	2.098	2.175	1.996	2.013	2.075
0.005	2.574	2.579	2.611	2.116	2.138	2.210	2.022	2.036	2.075
0.01	2.583	2.574	2.597	2.114	2.133	2.219	2.027	2.071	2.126
0.05	2.603	2.588	2.597	2.149	2.212	2.336	2.027	2.121	2.219
0.1	2.612	2.614	2.596	2.157	2.297	2.451	2.024	2.188	2.475
0.15	2.610	2.601	2.657	2.156	2.361	2.614	2.018	2.322	2.722

Table 6

Simulating a Conditional Asset Pricing Model with no Time-Varying Alpha

The table shows the results of 10,000 simulations of the estimates from the conditional asset pricing model with no time-varying alpha, allowing for the possibility of data mining for the lagged instruments. The regression model is:

$$r_{t+1} = \alpha_0 + b_0\, r_{m,t+1} + b_1\, r_{m,t+1} Z_t + u_{t+1}.$$

T is the sample size, L is the number of lagged instruments mined, R_p^2 is the true predictive R^2 in the artificial data generating process.

	T=66			T=350			T=960		
R_p^2	L=1	L=25	L=250	L=1	L=25	L=250	L=1	L=25	L=250
				Means: α_0					
0.001	0.000	0.000	-0.001	0.000	0.000	0.000	0.000	0.000	0.000
0.005	-0.001	-0.001	-0.001	0.000	0.000	0.000	0.000	0.000	0.000
0.01	-0.001	-0.001	-0.001	0.000	0.000	0.000	0.000	0.000	0.000
0.05	-0.002	-0.002	-0.002	0.000	0.000	-0.001	0.000	0.000	0.000
0.1	-0.002	-0.002	-0.003	0.000	-0.001	-0.001	0.000	0.000	0.000
0.15	-0.003	-0.003	-0.003	-0.001	-0.001	-0.001	0.000	0.000	-0.001
				Critical 5% t-statistics for α_0					
0.001	2.165	2.146	2.123	1.951	1.988	1.947	1.990	1.965	1.934
0.005	2.138	2.114	2.089	1.949	1.981	1.929	1.983	1.961	1.922
0.01	2.132	2.102	2.054	1.944	1.966	1.918	1.992	1.956	1.931
0.05	2.042	2.020	1.986	1.907	1.921	1.917	1.983	1.970	1.921
0.1	1.984	1.956	1.929	1.896	1.910	1.885	1.988	1.962	1.904
0.15	1.949	1.922	1.893	1.900	1.884	1.829	1.993	1.965	1.869
				Means: b_1					
0.001	0.007	-0.006	-0.043	-0.008	0.004	-0.001	-0.003	0.013	-0.012
0.005	0.008	-0.017	-0.060	-0.006	-0.003	0.000	-0.003	0.013	-0.008
0.01	0.008	-0.015	-0.064	-0.006	-0.003	-0.004	-0.003	0.010	-0.007
0.05	0.010	-0.031	-0.047	-0.003	-0.003	-0.003	-0.002	0.007	-0.001
0.1	0.010	-0.020	-0.035	-0.002	-0.005	-0.001	-0.002	0.000	0.001
0.15	0.010	-0.029	-0.042	0.000	-0.003	-0.003	-0.002	0.002	-0.002
				Critical 5% t-statistics for b_1					
0.001	2.630	2.605	2.639	2.157	2.128	2.218	1.987	2.136	2.147
0.005	2.636	2.646	2.631	2.156	2.145	2.246	1.991	2.162	2.170
0.01	2.661	2.665	2.643	2.163	2.150	2.256	1.987	2.154	2.216
0.05	2.656	2.748	2.739	2.146	2.257	2.441	1.988	2.267	2.476
0.1	2.629	2.811	2.861	2.175	2.378	2.618	1.994	2.395	2.639
0.15	2.607	2.857	3.001	2.201	2.466	2.755	2.008	2.479	2.828

Table 7

Simulating a Conditional Asset Pricing Model with no Time-Varying Beta

The table shows the results of 10,000 simulations of the estimates from the conditional asset pricing model with no time-varying beta, allowing for the possibility of data mining for the lagged instruments. The regression model is:

$$r_{t+1} = \alpha_0 + \alpha_1 Z_t + b_0\, r_{m,t+1} + u_{t+1}.$$

T is the sample size, L is the number of lagged instruments mined, R_p^2 is the true predictive R^2 in the artificial data generating process.

	T=66			T=350			T=960		
R_p^2	L=1	L=25	L=250	L=1	L=25	L=250	L=1	L=25	L=250
				Means: α_0					
0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.005	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.01	-0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.05	-0.001	-0.001	-0.001	0.000	0.000	0.000	0.000	0.000	0.000
0.1	-0.002	-0.001	-0.001	0.000	0.000	0.000	0.000	0.000	0.000
0.15	-0.002	-0.002	-0.002	-0.001	-0.001	0.000	0.000	0.000	0.000
				Critical 5% t-statistics for α_0					
0.001	2.239	2.592	2.794	1.991	2.080	2.182	2.004	2.047	2.092
0.005	2.219	2.533	2.764	1.982	2.070	2.147	2.014	2.036	2.084
0.01	2.206	2.511	2.731	1.988	2.069	2.150	2.013	2.018	2.089
0.05	2.124	2.429	2.656	1.970	2.056	2.148	2.019	2.012	2.130
0.1	2.065	2.389	2.607	1.968	2.034	2.158	2.035	2.010	2.160
0.15	2.015	2.356	2.546	1.978	2.033	2.177	2.072	2.010	2.167
				Means: α_1					
0.001	0.001	0.000	0.002	-0.001	0.001	0.000	0.000	-0.001	0.000
0.005	0.001	0.000	0.001	-0.001	0.001	0.001	0.000	-0.001	0.001
0.01	0.001	0.000	0.001	-0.001	0.001	0.001	0.000	-0.001	0.001
0.05	0.000	0.001	0.003	-0.001	0.001	0.001	0.000	-0.001	0.002
0.1	0.000	0.001	0.005	-0.001	0.000	0.000	0.000	0.000	0.001
0.15	0.000	0.000	0.003	-0.001	0.000	-0.001	0.000	-0.001	0.000
				Critical 5% t-statistics for α_1					
0.001	2.307	4.015	5.298	2.068	3.172	3.912	2.020	3.073	3.794
0.005	2.311	4.005	5.169	2.066	3.149	3.890	2.017	3.061	3.761
0.01	2.312	3.998	5.142	2.061	3.148	3.891	2.017	3.058	3.754
0.05	2.322	3.968	5.040	2.054	3.156	3.885	2.003	3.029	3.739
0.1	2.323	3.908	5.041	2.056	3.138	3.849	2.012	3.001	3.713
0.15	2.323	3.901	5.075	2.056	3.124	3.783	2.005	3.003	3.659

Table 8

Conditional Asset Pricing Models with a Cross-section of Returns

The table shows the results of 10,000 simulations from a conditional asset pricing model, allowing for the possibility of data mining for the lagged instruments. The dependent variables are book-to-market quintile portfolios. The regression model is:

$$r_{t+1} = \alpha_0 + \alpha_1 Z_t + b_0\, r_{m,t+1} + b_1\, r_{m,t+1} Z_t + u_{t+1}.$$

T is the sample size and L is the number of lagged instruments mined. The true predictive R^2 in the artificial data generating process is 0.005.

	T=66			T=350			T=960		
BM quintile	L=1	L=25	L=250	L=1	L=25	L=250	L=1	L=25	L=250
				Means: α_0					
BM1 (low)	-0.002	-0.001	-0.001	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002
BM2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
BM3	0.000	0.002	0.001	0.002	0.002	0.002	0.002	0.002	0.003
BM4	0.001	0.003	-0.001	0.004	0.005	0.006	0.005	0.006	0.007
BM5 (high)	0.004	0.002	0.006	0.006	0.006	0.005	0.008	0.007	0.006
				Critical 5% t-statistics for α_0					
BM1 (low)	2.157	2.593	2.705	1.691	1.847	1.914	1.526	1.618	1.651
BM2	2.297	2.523	2.742	1.916	2.067	2.156	1.960	2.034	2.056
BM3	2.296	2.681	2.916	2.093	2.218	2.380	2.143	2.206	2.301
BM4	2.343	2.686	3.060	2.059	2.253	2.399	2.132	2.233	2.368
BM5 (high)	2.369	2.659	3.099	2.135	2.233	2.326	2.269	2.284	2.359
				Means: α_1					
BM1 (low)	0.002	0.000	0.001	-0.001	-0.004	0.002	0.000	0.000	0.000
BM2	0.000	0.000	0.002	0.001	0.000	0.001	0.000	0.000	0.001
BM3	0.009	0.006	-0.008	-0.001	-0.005	-0.008	0.001	0.007	0.001
BM4	0.019	-0.027	-0.004	0.007	-0.003	0.012	-0.003	-0.003	0.005
BM5 (high)	0.021	0.028	-0.068	0.005	0.028	0.027	0.002	-0.006	-0.009
				Critical 5% t-statistics for α_1					
BM1 (low)	2.381	4.088	5.382	2.037	3.243	3.917	1.962	3.115	3.813
BM2	2.390	3.884	4.956	2.025	3.145	3.793	1.971	3.044	3.637
BM3	2.418	4.146	5.720	1.972	3.264	3.999	1.952	3.148	3.804
BM4	2.403	4.263	5.705	2.078	3.240	3.934	1.999	3.179	3.807
BM5 (high)	2.417	4.227	5.594	2.005	3.271	4.076	2.021	3.134	3.817
				Means: b_1					
BM1 (low)	0.018	0.007	-0.054	-0.007	0.002	0.051	0.021	-0.001	0.004
BM2	-0.003	0.032	-0.015	0.000	-0.004	-0.009	-0.010	0.000	-0.002
BM3	0.108	0.050	0.128	0.066	-0.033	-0.042	-0.041	-0.004	0.034
BM4	-0.230	-0.050	-0.062	-0.016	-0.054	0.087	0.028	-0.068	0.032
BM5 (high)	-0.479	0.075	-0.389	0.041	0.113	-0.006	-0.136	0.032	0.058
				Critical 5% t-statistics for b_1					
BM1 (low)	2.612	2.522	2.548	2.065	2.168	2.181	2.045	2.115	2.061
BM2	2.521	2.591	2.559	2.146	2.126	2.149	2.056	2.056	2.067
BM3	2.573	2.501	2.585	2.083	2.159	2.089	2.103	2.023	2.061
BM4	2.624	2.556	2.552	2.048	2.114	2.091	2.130	2.004	2.035
BM5 (high)	2.628	2.640	2.536	2.177	2.149	2.068	2.084	2.050	1.982

Table 9

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.

					Lagged return		Detrended (12)	
Observations	OLS	White	NW(auto)	NW(20)	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