Is Time-Series Based Predictability Evident in Real-time?[***]

Michael Cooper[*] and Huseyin Gulen[**]

May, 2002

[***] We would like to thank Mike Cliff, David Denis, Diane Denis, John Easterwood, Roberto Gutierrez Jr., Greg Kadlec, Raman Kumar, John McConnell, Raghu Rau, Vijay Singal, and seminar participants at the University of Georgia and Purdue for their helpful comments and suggestions. Also, we thank James Garret for his excellent research assistant work. We would like to give special thanks to Ken French for his many invaluable comments and suggestions.

[*] Krannert Graduate School of Management, Purdue University, West Lafayette, IN 47907-1310
765-494-4438, mcooper@mgmt.purdue.edu

[**] Virginia Tech, Department of Finance, 1016 Pamplin Hall – 0221, Blacksburg, VA 24061
540-231-4377, hgulen@vt.edu

Is Time-Series Based Predictability Evident in Real-time?

Abstract

A real-time investor is one who must base his portfolio decisions solely on information available *today*, not using information from the future. Academic predictability papers almost always violate this principle via exogenous specification of critical portfolio formation parameters used in the backtesting of investment strategies. We show that when the choice of parameters such as predictive variables, traded assets, and estimation periods are endogenized (thus making the tests more real-time), all evidence of predictability vanishes. However, an investor with the correct specific sets of priors on predictive variables, assets, and estimation periods *will* find evidence of predictability. But since no real theory exists to guide one on the choice of the correct priors, finding this predictability seems unlikely. Our results provide an explanation for the performance gap between mutual funds and the academic market predictability literature, and carry important implications for asset pricing models, cost-of-capital calculations, and portfolio management.

There now appears to be overwhelming evidence of stock market predictability. A large body of research shows that excess returns on the aggregate market are forecastable from the default spread, dividend yield, dividend payout, the term spread, consumption data, inflation, industrial production, wealth and labor income, to name but a few variables.[1] Yet, despite this seemingly overwhelming evidence, there appear to be few real-world investors capable of taking advantage of this predictability, especially at the levels of profits suggested by the academic predictability papers.[2] Cochrane (1999) states, "It is uncomfortable to note that fund returns still cluster around the (buy-and-hold) market Sharpe ratio." He suggests, "If the strategy is real and implementable, one must argue that funds simply failed to follow it." Thus, there appears to be a large gap between real-time investor performance and the high levels of predictability found in the literature.

We offer an explanation for this performance gap that is based on potential collective data-snooping biases on the part of researchers. This collective snooping may be inherent to the market predictability literature because (1) there is little explicit guidance from theory regarding the identity of the predictive variables used in these studies, hence making it a data-fitting exercise; (2) any new research endeavor is inherently conditioned on the collective knowledge built up to that point; and (3) there is a tendency in the literature and the profession at large to retain the findings that "work" and discard the ones that do not. Given these issues, it is feasible that a nontrivial proportion of the relations reported in the literature, and accepted as economically meaningful, are simply due to pure luck. As Black (1993a, b), Denton (1985), Lo and MacKinlay (1990), Foster, Smith and Whaley (1997), Ross (1989), and Sullivan, Timmermann, and White (1999) point out; we (usually out of sheer necessity) *collectively* condition our studies on existing empirical regularities with the unintended consequence of snooping the data.

We attempt to gauge the impact of potential data-snooping on empirical findings that are based on commonly used methodologies in the market forecasting literature and under specific

[1] A partial list of academic papers that document stock market predictability include: Breen, Glosten, and Jagannathan (1989), Campbell (1987), Campbell and Shiller (1988a, 1988b), Chen, Roll and Ross (1986), Cochrane (1991), Fama and French (1988, 1989), Ferson and Harvey (1991), Hodrick (1992), Keim and Stambaugh (1986), Lamont (1998), Lettau and Ludvigson (2001a), Lewellen (1999), Pontiff and Schall (1998), and Santos and Veronesi (2001).

[2] The current notion that the stock market is predictable stands in contrast to the well-documented inability of mutual funds to beat the market (see Carhart (1997), Wermers (2000)). It is interesting to note that in addition to mutual fund studies, nearly all other studies of real-time investment performances also fail to show that the market is clearly beatable. Barber and Odean (2000) find this for individual investors; Christopherson, Ferson, and Glassman (1998) find this for pension funds; Pirinsky (2001) finds this for banks, investment advisors, and insurance companies; Desai and Jain (1995) find this for "superstar"

plausible scenarios of snooping. Specifically, we use a computer-intensive methodology to (1) mimic what may be the inherent process of research on empirical asset pricing studies that attempt to relate lagged predictive variables to aggregate market returns and (2) test for the potential effects of data-snooping in light of this research process.

Specifically, to "mimic the inherent process of research," we start by examining the forecasting methodologies of the market predictability papers. Typically, these papers use "out-of-sample" tests. In this type of set up, researchers employ variations on a rolling forecast method in which the researcher estimates a model of expected returns from a prior data period and employs that information to create forecasts in a hold out period. The researcher rolls through the data, creating a time series of out-of-sample forecasts, and employs a variety of statistical and economic tests to evaluate the forecasts.[3]

However, a characteristic of many out-of-sample papers is that they are not truly out-of-sample, in the sense of using an independent holdout period.[4] Typically, researchers use the same, or substantially the same period to discover predictive relationships as to test them. If snooping occurs, then the use of full period information can result in a subtle, but very important test size problem emanating from a researcher's design of the out-of-sample forecasting algorithm. Upon closer inspection, it appears that not all out-of-sample forecasts are created equal; indeed, there are important differences regarding the degree of endogeneity (or lack there of) in choosing critical parameters used to create the forecasts. Many features of a researcher's out-of-sample experiment such as the choice of assets to forecast, the countries of the assets, the return horizon of the assets, the choice of predictive variables, how to control for regime shifts in the underlying return generating process (i.e., the length of the in-sample window used to obtain forecast parameters), the method of model selection, and other aspects, are typically *exogenously* determined by the researcher.[5] If one agrees with the view that there exists little theory to guide

---

money managers; Metrick (1999) finds this for newsletter recommendations; Barber et. al. (2000) find this for analysts' consensus recommendations.

[3] As numerous papers point out, an out-of-sample framework is preferable to an in-sample approach (in which a predictive model is estimated using the entire sample) because it minimizes false rejections of the null hypothesis of no predictability, and increases the "real-time" nature of the forecasting experiment.

[4] In the cross-sectional literature, when viewed on a case-by-case basis, there is *now* hold-out sample evidence that supports initial claims of predictability. For example, Fama and French (1992) find evidence of a value premium using data ending in 1990. In the 1990's and early 2000's, the value premium has continued to be statistically significant. Similarly, the momentum premium of Jegadeesh and Titman (1993) continues to be significant (Jegadeesh and Titman (2001)).

[5] Some researchers have examined endogenizing one or two of these forecasting aspects in the time-series predictability literature. Pesaran and Timmermann (1995) and Bossaerts and Hillion (1999) endogenize predictive variable selection by using various statistical model selection criteria; Pesaran and Timmermann (1995) develop a "hyper selectivity" forecasting model that endogenizes the statistical model selection criteria; and Pesaran and Timmermann (1999), hoping to solve issues related to model nonstationarity by

us on the proper selection of these parameters (henceforth referred to as "econometrician choice variables"), then the choice variables may potentially be chosen in either (1) an ad-hoc fashion, (2) to make the out-of-sample forecast "work," or (3) by conditioning on the collective knowledge built up to that point (which may emanate from (1) and/or (2)), or some combination of the three.[6]

Thus, we mimic the process of time-series predictability research using a computer intensive algorithm that loops over parameter values for groups of econometrician choice variables that are normally exogenously specified in the literature. Specifically, using a recursive forecasting method that is ubiquitous to the market forecasting literature, we explicitly snoop over a range of three commonly exogenously specified econometrician choice variables: predictive variables, assets, and in-sample window lengths.[7] On first consideration, it may appear that variations in just these three parameters would not be the source of too much concern about potential data-snooping problems. However, using just three data sets from recently published time-series studies, and examining plausible parameterizations over these three aspects, we find that the number of exogenously specified forecasts can easily result in close to *100,000* parameterizations of out-of-sample forecasts. The data we use come from three market predictability papers (Pesaran and Timmermann (1995), Bossaerts and Hillion (1999), and Lettau and Ludvigson (2001)). The data sets from these papers include time-series variables such as a consumption to wealth ratio, dividend yield, dividend payout ratio, various interest rate and term structure measures, a default risk measure, inflation, industrial production, and a number of other predictive variables, along with the excess returns of 13 countries' major indexes, covering 1953 to 1997. Thus, our data set spans the majority of variables used in the market predictability literature, providing us with a comprehensive and plausible set of predictive variables to carry out our "mimicking of the research process" simulations.

---

capturing shifts in factor/return relations, endogenize in-sample window length. Swanson and White (1997) endogenize variable selection and window length via linear models and artificial neural networks in an attempt to forecast macroeconomic variables. In the cross-sectional literature, Cooper, Gutierrez, and Marcum (2001) further explore such "real-time" issues inherent in out-of-sample tests by requiring the investor to endogenously determine in-sample the optimal predictor variables, rules relating those variables to future returns, and the dimensionality of the sort. Once they endogenize these portfolio investment parameters, it is difficult for an investor to outperform a passive buy-and-hold benchmark portfolio.

[6] An example of such snooping, which is not likely to be the result of an explicit search on the part of a researcher, is the practice of using the best subset of a group of predictive variables from in-sample tests in contemporaneous out-of-sample tests.

[7] It is worth emphasizing that these three appear to us to be the most obvious exogenously specified parameters. There are *many* other more subtle parameters a researcher must specify before implementing an out-of-sample test. These include return horizon, model selection criteria, asset allocation rules, forecast update frequency, test(s) of the null, learning features, transaction costs, and others. In a later section of the paper, we expand our analysis to endogenize model selection criteria and transaction costs.

We find in these recursive out-of-sample simulations, for which we consider exogenous combinations of various dimensions of the above econometrician choice variables, that approximately 1% to 80% of the forecasts yield evidence of predictability. Obviously, this is a *huge* variation, and it illustrates the striking differences in predictability across exogenously specified variable groups, assets, in-sample window lengths, and performance measures. We find that the economic and statistical significance of the best out-of-sample forecasts are startling; on US index data, many of the successful models beat their buy-and-hold benchmarks by a magnitude of 4 to 5 times on a terminal wealth basis and handily outperform using a battery of statistical tests commonly used in the literature. We also find plenty of evidence in favor of out-of-sample predictability on international indexes. For example, for the 12 countries we examine outside of the US, we find strong evidence of predictability in *all* 12 countries for *some* combination of the choice variables.

This process of explicitly snooping the data yields some interesting insights. First, the distributions of the successful forecasts' econometrician choice variables span the full spectrum of the choice variables' values, providing us with little guidance on the true values of these parameters. Second, rejections of the null hypothesis of no predictability are very sensitive to minor changes in values of the choice variables. Third, many of the related literature's forecasting models tend to be in the upper end of the snooped ex post distribution. Overall, the potential for serious data-snooping problems appears to be very high, especially considering that we find large variations in rejections of the null across minor changes in the econometrician choice variables, compounded by the fact that there appears to be minimal ex ante guidance from theory on the correct values of these variables. Obviously, the performances of the best out-of-sample forecasts are only attainable if an investor held a prior dictating a strategy *exactly* mimicking the parameters of the best forecasts.

Next, we test for the effects of potential data snooping in our simulations and the related literature. We follow Sullivan, Timmermann, and White (1999), who note that the effects of data-snooping, operating over time and across many researchers, can only be quantified provided that one considers the performance of the best trading rule in the context of the full universe of trading rules from which this rule is conceivably chosen. Therefore, we gauge the amount of data-snooping bias present in the best performing out-of-sample exogenous combinations by performing out-of-sample experiments in which we (1) remove the effects of parameter snooping via endogenizing the selection of the econometrician choice variables and (2) examine the difference in profitability between the endogenous forecasts (which reduce data-snooping biases) and the best ex post runs from the exogenous simulations (which explicitly contain snooping

biases). Thus, these tests tell us if time-series based predictability is evident in "real-time" or if it just an "ex post econometrician" induced phenomena.

We use two techniques to endogenize the econometrician choice variables for our "real-time" forecasts. First, we use the mutual fund literature's technique of testing for persistence (see Jensen (1969), Grinblatt and Titman (1992), Carhart (1997), and others). We treat each of the exogenously specified out-of-sample forecasts from above as "mutual funds." We test for persistence by ranking on prior performance of these "funds" and then examine performance in step-ahead periods. Our second approach is to develop a recursive forecasting method, building on approaches in Pesaran and Timmermann (1995) and Bossaerts and Hillion (1999), to endogenize the econometrician choice variables. This method employs an in-sample period to choose the best forecasting model from the universe of potential models, and then uses the optimal model to form portfolios in step-ahead periods. The mutual fund persistence method has the advantage of being relatively simple and intuitive, allows for a simple form of learning in the ranking period, and does not force the choice of a single best model, as typically does a recursive method. On the other hand, the recursive method has the advantage of being more commonly used in the market forecasting literature, which helps to facilitate comparison with previous papers' results.

Our results indicate that the degree of data-snooping in the market predictability literature is likely to be high. We find that the decrease in predictability between the exogenously specified "snooped" out-of-sample forecasts and the real-time, endogenized forecasts is quite large. The best performing snooped out-of-sample combinations are highly significant by all standards – large Jensen's alphas, large Fama-French three factor alphas, and significant evidence of market timing. In contrast, the real-time forecasts, i.e., the endogenized forecasts, have a hard time beating their buy-and-hold benchmarks in our recursive forecasts, and rarely create a meaningful dispersion between winners and losers in our "mutual fund" persistence tests. Thus, researchers do and will find market predictability from lagged macroeconomic variables, but much of it does not appear to be "real," that is, once ex post biases in the selection of predictive variables, assets, and estimation periods are reduced, evidence of predictability is largely eliminated.

Thus, the main contribution of this paper relative to previous data snooping papers (Black (1993a, b), Denton (1985), Lo and MacKinlay (1990), Foster, Smith and Whaley (1997), Ross (1989), and Sullivan, Timmermann, and White (1999)) is the incorporation of more aspects of uncertainty facing a real-time investor, and the resulting dramatic reduction in time-series based predictability.

Overall, our results suggest that to minimize potential data-snooping problems, it is critically important to endogenize investor choice parameters in out-of-sample forecasts. Thus, the results carry implications for the growing numbers of conditional asset pricing studies that exogenously choose lagged predictive variables based on their ability to forecast the general market, and employ the variables' loadings in cross-sectional tests (for example, see Ferson and Harvey (1999), Lettau and Ludvigson (2001), or Santos and Veronesi (2001)). Similarly, the results have implications for the rapidly growing Bayesian predictability literature that typically chooses an exogenous set of predictive variables and shows how "parameter uncertainty" can lead to important changes in investors' allocations to stocks (for example, see Kandel and Stambaugh (1996), or Barberis (2000)).

The remainder of the paper is organized as follows. In section I, we develop and discuss a reality spectrum of out-of-sample forecasts, guided by the underlying principle that the correct approach to an out-of-sample forecast should be to simulate as accurately as possible all of the uncertainties faced by a real investor. The reality spectrum provides a summary of many parameters in the market forecasting literature that are typically exogenously specified. We use this reality spectrum as a basis for the design of the exogenously specified "snooping" simulations in section II. In section III, we present the results of out-of-sample forecasts which endogenize the selection of predictive variables, assets, and estimation periods. Section IV contains a discussion of the results and our conclusion.

I. A Reality Spectrum of Out-of-Sample Forecasts

It is a trivial exercise to find evidence of predictability using the same data to discover and validate relations between lagged predictive variables and stock returns (Foster, Smith and Whaley (1997) and Sullivan, Timmermann and White (1999)). Thus, many researchers have turned away from explicit in-sample tests in validating return anomalies and have instead focused on out-of-sample tests. However, as we discuss in the introduction, not all out-of-sample forecasts are created equal. A true out-of-sample forecast would involve choosing the best model(s) now, and then waiting years into the future to validate the model. Given the impracticality of this approach, financial economists typically perform recursive tests using the same or substantially the same period to discover predictive relationships as to test them. Thus, the level of realism in these "in-sample" out-of-sample tests hinge critically on the degree of exogenous/endogenous specification of parameters used to implement the experiments.

The main point of this paper is to test the extent to which the practice of exogenously specifying parameters may result in the false rejection of the null of no predictability. To accomplish this task, we first develop an idea of the extent and identity of exogenous parameter specification in the market forecasting literature. To facilitate this, Figure 1 provides a "reality spectrum" of out-of-sample forecasts. The left most column of Figure 1 provides a list of some of the more commonly exogenously specified econometrician choice variables. We separate the choice variables into "major" and "minor" categories based on our perception of their relative importance in the market predictability literature. The reality spectrum ranges from "NONE," in which a researcher employs an in-sample methodology, to "LOW," in which the researcher employs a recursive forecast but all of the econometrician choice parameters are exogenously determined, to "SOME," in which a few of the parameters are endogenized, up to "HIGH," in which most of the forecast parameters are endogenized. Clearly, even a "HIGH" level of realism in the modeling process is a simplified version of an actual investors' decision-making process.[8]

We start with the most obvious parameter. The choice of predictive variables is likely to be the winner, with many papers invoking the phrase "we focus on a common set of lagged instruments, shown to have worked in previous studies" as justification for their chosen set of predictive variables. Studies typically use a fixed set of three to five variables. Examples include Campbell (1987), who uses lagged returns, T-bill yield, change in yield, and a yield spread measure, Keim and Stambaugh (1986), who use the yield on Baa-rated bonds less the one-month T-bill yield, a ratio of the level of the S&P500 to the 45 year average of the S&P500 level, and a measure of share price, averaged equally across the quintile of smallest market cap firms on the NYSE, Ferson (1990), who uses lagged returns, yield on a short term T-bill, change in the yield on a short term T-bill, yield spread between the yield on an overnight fixed income security and the short term T-bill, and Ferson and Harvey (1993), who use lagged returns, yield on a short term T-bill, growth rate of industrial production, an inflation measure, and an unexpected inflation measure. Other prominent variables in the literature include dividend yields (Shiller (1984), Fama and French (1988)), dividend payout, or the ratio of dividends to earnings (Lamont (1998)), term spread measures (Fama and French (1989)), the level of consumption relative to income and wealth (Lettau and Ludvigson (2001)), and a dummy variable for the month of

[8] It is likely that any modeling attempt cannot possibly account for the myriad of uncertainties facing a real-time investor. For example, a more accurate depiction of the "real-world" uncertainties facing an investor might include a real-time expanding predictive variable set (likely numbering in the tens or hundreds of thousands of variables), a survivorship bias-free collection of assets within all countries and across all countries (Brown, Goetzmann, and Ross, 1995 and Jorion and Goetzmann, 1999), and a real-time expanding consideration of all possible model selection methods and computing technologies, to name just

January (Harvey (1991)). The above list is by no means all-inclusive. Some other more nontraditional variables include deseasonalized cloud cover, raininess, and snowiness (Hirshleifer and Shumway (2001)), ambient noise level (Coval and Shumway (2001)), and the distance of a trader from the corporate headquarters of the traded stock (Hau (1999)). A casual perusal of DataStream (a popular purveyor of world-wide financial data) reveals thousands of time series for the US and many other countries. Thus, as Foster, Smith and Whaley (1997) point out, "There are limitless possible linear and nonlinear transformations of these variables."

The next most commonly varied parameter is likely to be the choice of the predicted asset(s). Popular assets include excess US stock (EW and VW CRSP indexes, the S&P 500 Composite Index) and bond portfolios (for example, Keim and Stambaugh (1986), Campbell (1987), Harvey (1989), Lettau and Ludvigson (2001), Fama and French (1988), Ferson and Harvey (1991), Pesaran and Timmermann (1995), and Brandt (1999)). Also popular are industry-grouped portfolios (Ferson and Harvey (1991), Ferson and Korajczyk (1995), Lo and MacKinlay (1997)), size sorted portfolios (Ferson and Harvey (1991), Avramov (1999), Ferson and Korajczyk (1995), Lo and MacKinlay (1997)) and size and B/M sorted portfolios (Ferson and Harvey (1999), and Lettau and Ludvigson (2001)). Internationally, papers have used the popular Morgan Stanley Capital International indexes (Ferson and Harvey (1993), Ang and Bekaert (2001)). These indexes are typically value-weighted, and are available for a broad range of developed countries. Other papers employ the Datastream international indexes. Also, papers have used country specific indexes, for example, the Nikkei 225 in Japan, the Bovespa in Brazil, and the Madrid SE General for Spain, to name but a few. As with the predictive variable list, the above list of assets is by no means all-inclusive.

The next econometrician choice variable, in-sample window length, might be less of an obvious parameter as compared to predictive variables and asset choice. But, as we will show in the data-snooping simulations, it has a large bearing on how often one rejects the null of no predictability. In-sample window length refers to the sample period from which time series forecasting model parameters (betas) are estimated. These betas are then multiplied by the predictive variable realizations to form expected return estimates in step ahead out-of-sample periods. The choice of window length is not at all straightforward; if one believes regime shifts may have occurred across a given sample period, one may employ a relatively short window, or apply exponentially declining weights to past observations. If one believes that "the truth" only emerges from betas estimated over a long time series, one may employ an expanding window.

---

a few. Therefore, our conservative depiction of the number of econometrician choice variables serves to likely bias tests in favor of finding real-time predictability.

Examples of expanding windows include Pesaran and Timmermann (1995), Ferson and Harvey (1999), and Lettau and Ludvigson (2001). Examples of fixed windows include Bossaerts and Hillion (1999), Sullivan, Timmerman, and White (1999), Ferson and Harvey (1993, 1999), Lo and MacKinlay (1997), and Cremers (2000).

The goal of our paper is to evaluate potential data-snooping effects in the market predictability literature from the above "major" choice variables, with major being defined in large part from the range in values these variables exhibit in the literature. However, there are many other "minor" parameters that a researcher must calibrate in implementing out-of-sample forecasts, and in the interest of presenting a more complete reality spectrum, we discuss these next. We begin with model selection. This category is closely related to "choice of predictive variables." All time-series predictability papers must choose an overall group of conditioning variables. For papers that exogenously choose and hold fixed their predictive variables (which appear to us to be the vast majority of papers in the market predictability literature), this category is effectively removed, or is what we will refer to as "none." However, there are some papers that endogenize the choice of predictive variables from within a fixed exogenously specified universe of variables. For example, Pesaran and Timmermann (1995) endogenize model selection across a family of statistical and economic-based model selection criteria (for example, R^2, Akaike, Schwarz, "sign," "Sharpe," and "wealth" criteria). Bossaerts and Hillion (1999) endogenize variable selection across a number of exogenously specified statistical selection methods. Swanson and White (1997), Allen and Karjalainen (1997), and Brown, Goetzmann, and Kumar (1998), use various forms of nonlinear selection criteria (including neural networks and genetic algorithms) to endogenize variable selection. Ait-Sahalia and Brandt (2001) use sample analogues of the conditional Euler equations to select variables. In the Bayesian predictability literature, a growing number of papers, including Avramov (1999), Pastor (2000), Cremers (2000), start with a fixed set of portfolios (for example, SMB, HML, and the market) or fixed set of predictive variables, and endogenize them via estimating a prior distribution of model parameters and applying those parameter estimates to obtain a predictive distribution. Thus, the most common forms of model selection appear to be "none," followed by various statistical selection methods, while some have used "artificial intelligence" methods, and a small but growing group use a Bayesian model uncertainty approach.

Finally, in Figure 1 we list out other less obvious parameters that a researcher must decide upon in implementing an out-of-sample forecast. These include the trading rule used to translate an expected return forecasts into an asset allocation, return horizon of the predicted assets (potential values include monthly, quarterly, yearly, etc.), forecast update frequency (for

example, monthly, yearly), study period (typically a researcher will use data up to the point of their study, with the starting point being the exogenously determined aspect), test(s) of the null hypothesis (for example, parametric or nonparametric tests statistics, parameterization of a utility function if the tests are based on utility measures, method of standard error calculation, and number and identity of "risk factors" in estimating an alpha), forms of learning, and transaction costs. Lastly, the appropriate use of "technology" is an important issue that is rarely addressed in studies of predictability.[9] For example, it would be inappropriate to use a computer intensive genetic algorithm to uncover evidence of predictability before the algorithm or computer was available.

II. Mimicking the Inherent Process of Time-Series Predictability Research

Does the inherent process of research tend to make us converge on values of the econometrician choice variables that work the best, but are not known ex ante, in real-time? We address these questions in two parts; first, in this section, we use computer intensive simulations to explicitly search over combinations of assets, predictive variables and in-sample window lengths that researchers may have explored. These simulations provide us with the ex post distribution of out-of-sample predictability that emerges from a systematic search over plausible values of the three econometrician choice variables of predictive variables, assets, and in-sample window lengths. We want to stress that we do not believe that any one researcher actually conducted such a search, but that the process of research, across researchers and over time, may have implicitly resulted in such a search. In addition, the fact that many papers in the time-series predictability literature use a method of first finding the best predictive variables and assets via in-sample methods, and then use the exact, or closely related, combination of optimal parameters into contemporaneous out-of-sample tests, suggests that our ex post views of the best simulations in this section may not be too different than the "inherent process of research."

We examine our ex post, or snooped distribution from a number of standpoints; we examine what percentage of the exogenously specified out-of-sample runs reject the null of no predictability; we examine if there are any common characteristics of the successful forecasts, such as the identity of the assets, variables, or estimation lengths, to guide us in determining which values of the econometrician choice variables may be the "truth"; we judge how sensitive the best forecasts are to minor changes in the choice variables; we examine where the current best

[9] An exception is Pesaran and Timmermann (1995). In their real-time recursive study of predictability on the S&P500, they consider a subgroup of forecasting tools, which "use simple statistical and computing techniques that were clearly publicly available to any investor throughout the sample period analyzed in this paper." See page 1203.

factors in the market predictability literature fall in the snooped distribution; and we establish an upper bound on what researchers may find in the future using various exogenous combinations of these conditioning variables and assets. This analysis is the foundation for the most important question – is it real? Are the forecast combinations that "work" just ex post illusions, or are they attainable ex ante? To answer this question, in a later section we endogenize the econometrician choice variables and examine what happens to predictability once we explicitly control for uncertainty related to the selection of assets, variables, and estimation periods.

A. Predictive Variables and Assets

Table 1 describes the data. We use data from three recent market predictability papers, Pesaran and Timmermann (1995), Bossaerts and Hillion (1999), and Lettau and Ludvigson (2001).[10] The data sets from these papers include time-series variables such as a consumption to wealth ratio, dividend yield, dividend payout ratio, various interest rate and term structure measures, a default risk measure, inflation, industrial production, a January dummy, and a number of other predictive variables, along with the excess returns of 13 countries' major indexes (in US dollar returns) covering 1953 to 1997. The Lettau and Ludvigson data use quarterly returns, and the other two data sets use monthly returns. Each one of the variables from the three data sets has been shown to predict returns. Thus, our data set provides us with a comprehensive and plausible set of predictive variables to carry out our "mimicking of the research process" snooping simulations.

B. Exogenous "Snooping" Methodology

In this section, we construct out-of-sample forecasts using exogenous combinations of predictive variables, assets, and in-sample window lengths for the three data sets. Specifically, we follow these steps for each data set to construct the "snooped distribution":

1. For all possible variable combinations, I ($I=2^K-1$ models (each model includes an intercept), where K=the number of predictive variables in each data set), and all possible in-sample window lengths, W (W=10, 15, and 20 years of fixed moving windows and an *expanding* window for data set 1, and W=5, 6, 7, 8, 9, 10 years of fixed moving windows and an *expanding* window for data sets 2 and 3)[11], and all possible assets, A (A=1 for data sets 1 and 2, and A=13 for data set

[10] We thank Allan Timmerman, Peter Bossaerts, and Sydney Ludvigson for providing us with the data used in their studies.
[11] We use longer in-sample window lengths for data set 1 because of that data's quarterly return horizon.

3), we construct an out-of-sample time series of returns using the following recursive approach:

A. We estimate, using OLS, a linear model of the form $r_\tau = \beta_I' X_{\tau-1,I} + \varepsilon_{\tau,I}$ where $X_{\tau-1,I}$ is a $(n_I + 1)$ x 1 vector of predictive variables, including a vector of ones for the intercept term, and r_τ is the excess return for asset *A* during in-sample period τ. We estimate the model in the in-sample period *W*, and use the in-sample loadings on the predictive variables to form expected return forecasts in recursive, step ahead out-of-sample periods. For example, consider data set 2. The initial in-sample period is from 1954(1) to 1963(12). We estimate the linear model, obtain predictive variable loadings, and form an expected return estimate in the first out-of-sample period in 1964(1).

B. We then roll forward the in-sample end date by one period, re-estimate the model, and obtain a forecast for 1964(2). We repeat this process until the end of the out-of-sample period. Thus, for each data set, we obtain *W* x *A* x (2^K-1) out-of-sample forecast series.

2. For each out-of-sample forecast series, we obtain a series of realized returns from the following trading strategy: go long asset *A* if the expected excess return estimate for that period is great than zero, else invest in a t-bill. For each return series we estimate four performance measures; a forecast beta, Jensen's alpha, Fama-French three factor model alpha, and the Henriksson and Merton (1981) market timing measure.[12]

C. Results of the Exogenous Simulations

We present the results of the simulations in Table 2 Panel A for data set 1, Panel B for data set 2, and Panel C for data set 3. In each panel, we report the percentage of models which reject the null hypothesis of no predictability at the five percent level or better for each of the four performance measures. We also break down the rejection rates by in-sample window length *W* and the number of predictive variables *K* in a given model. In Table 3, Panels A, B, and C, we provide the details of the variable groups and in-sample window lengths for the top and bottom performing exogenous simulations for data set 1, 2, and 3, respectively. Considering all exogenous combinations of the three econometric choice variables of assets, predictive variables, and in-sample window lengths, results in 508 combination for data set 1 (1 asset* 4 windows * (2^7-1) models), 3577 combinations for data set 2 (1 asset * 7 windows * (2^9-1) models), and 93,093 combinations (13 assets * 7 windows * $(2^{10}-1)$ models) for data set 3. The large number of exogenous forecast combinations might seem extreme, but we maintain that when it is considered in light of the full reality spectrum of Figure 1 and in terms of the observed variations

[12] We thank Ken French for providing us with the monthly premiums for the Fama-French three factor model. The forecast beta (β_f) provides a measure of overall out-of-sample fit and is calculated by regressing the monthly realized return of the predicted asset on the forecasted return from each forecasting model: $r_\tau = \alpha + \beta_f r_{forecast,\tau} + \varepsilon_\tau$

of parameters in the published literature, it is not, but rather, likely represents a *smaller* number of combinations relative to the true distribution from which the best performing models in the literature have been drawn.

In Table 2, there are large variations in predictability across variable groups, in-sample window lengths, data sets, and performance measures. Depending on which performance measure one wants to examine, we find evidence of out-of-sample predictability in approximately 2% to 25% of the exogenous combinations for data set 1, 30% to 82% of the exogenous combinations for data set 2, and 1% to 5% of the exogenous combinations for data set 3.[13] The level of predictability in the best performing models is striking; in data set 1 (see Table 3, Panel A), the best model (CAY and RREL, with a ten year window), as defined by terminal wealth, handily beats an S&P500 buy-and-hold benchmark ($40.03 versus $18.99), has a quarterly Jensen's alpha of 1.14% (p=0.004), a Fama and French three factor alpha of 1.1% (p=0.009), a forecast beta of 0.68 (p=0.02), and a market timing value of 1.15 (p=0.04). We observe similar performance for the best model combinations in data sets 2 and 3; large terminal wealths compared to buy and hold measures, and large and significant values of the other performance measures. In Table 3, Panel C, it appears that across countries, there is some combination of variable group and window that results in market beating performance in every country. Thus, on an ex post basis, there is a large degree of out-of-sample predictability evident in all three data sets; from Table 3, panels A, B, and C, the annualized spread in the Fama-French alphas between the best and worst model combination is approximately 9.4%, 8.7%, and 20.7% for data sets 1, 2, and 3, respectively.[14]

This process of explicitly snooping the data yields some interesting insights. First, it is not too surprising that these three data sets generate out-of-sample predictability when we consider that the particular predictive variables in each data set were selected from other successful papers that were to some extent contemporaneous to the studies from which we gathered our data. Second, what is more surprising is A) the distributions of the econometrician choice variables in the successful forecasts span the full spectrum of the choice variables' values and B) rejections of the null are very sensitive to minor changes in the values of the choice variables. Thus, the simulations do not offer us much guidance on the true values of these econometrician choice variables.

[13] As Fama (1991) points out, all tests of asset pricing models are conditional upon the model of risk adjustment used, and the results in this section dramatically demonstrate different rejection rates across commonly employed test statistics. We do not directly pursue this issue, but obviously the exogenous choice of test statistic could dramatically change conclusions of predictability for these three data sets.

[14] The spread for data set 3 is the average spread across countries. The highest (lowest) annual spread for any individual country occurs in Sweden (The Netherlands) at 39% (12%).

For point "A" above, first consider the number of variables in a model, as broken out in each panel of Table 2. Across the data sets and performance measures, there is no apparent consistent pattern; in some cases larger variable groups result in more rejections of the null (e.g., in Panel C, we observe more rejections across the four performance measures as we move from models with one variable up to models with ten variables), but in other cases we observe greater numbers of rejections for smaller variable models (e.g., in Panel A, we observe a greater rate of rejection for models with one to five variables, and then a sharp drop off for *models* with six and seven variables). We also calculate, but do not report in the tables, the inclusion rates of the predictive variables in the successful and unsuccessful forecasts. For data set 1 we obtain the following inclusion rates from the models in the top decile of Jensen's alpha: estimated trend deviation in consumption (CAY), 49% of the models, S&P 500 excess return (SPX), 45%, dividend yield (DY), 18%, dividend payout ratio (DP), 41%, 30-day t-bill rate minus its 12 month moving average (RREL), 67%, 10-year T-bond yield less 1-year T-bond yield (TRM), 43%, and yield difference between BAA and AAA corporate bonds (DEF), 14%. For data set 1 we obtain the following inclusion rates from the models in the bottom decile of Jensen's alpha: CAY, 49% of the models, SPX, 52%, DY, 94%, DP, 29%, RREL, 49%, TRM, 12%, and DEF, 67%. Thus, some variables, such as CAY, lagged market, and the dividend payout ratio have very similar inclusion rates across the successful and unsuccessful models. Others, such as dividend yield and default spread, show up much more frequently in the unsuccessful models, but do show up in a nontrivial number of models in the successful models.

Consider next the in-sample window length. In data set 1, from Table 2, Panel A, the 20 year window is best for the Forecast Beta criterion, the expanding window is best for the Jensen's alpha, the 10, 15, and 20 year tie for the most rejections of the null under the three-factor alpha, and the 20 year is best under the market timing measure. In data set 2, a 5-year window is best for the forecast beta measure, but an expanding window is best for the other three measures. Finally, in the data set 3, an expanding window is never the best – the window that results in the most rejections is often of intermediate length. However, for all data sets, we still observe rejections of the null for all window lengths. In addition, across the top and bottom decile of out-of-sample runs (as defined by Jensen's alpha), there is not much difference in the average length of in-sample window.

For point "B" above, that rejections of the null are sensitive to small changes in the econometrician choice variables, first consider the best model from data set 1 as reported at the bottom of Table 2, Panel A. The best performing out-of-sample model as defined by terminal wealth is CAY and RREL using a 10-year in-sample window. That model has a terminal wealth

of over $40.00 as compared to $18.99 for the S&P500 buy-and-hold strategy. If we vary the window to 15 years, the terminal wealth drops to $30.43. If we add DY to the CAY and RREL model, we see a maximum wealth of $28.55 for a 20-year window and a minimum wealth of $13.81 for a 15-year window. If we select the second best model, which is DY, RREL, TRM, and DEF with a 20-year window, that model earns $37.79. If we now change to a 10-year window, the terminal wealth drops to $9.46. For data set 2, as reported in the bottom of panel B of Table 2, the best model, which uses an expanding window, has a terminal wealth of $84.25. If we change the same set of predictive variables to use a 5-year window, the wealth drops to $22.01. The same types of patterns are evident for the international data, in panel C. Indeed, small changes to the predictive variable group and estimation period can result in dramatic changes to our inferences of predictability. Overall, the lack of dominate predictive variables and estimation periods, and the sensitivity of the forecasts to small changes in parameterization suggest at best that as a group, the successful models emanate from "richly complex processes" or, at worst, arise from some combination of luck and/or ex post snooping.

Consider that for the three data sets we use (see Table 1), two of the original papers, Pesaran and Timmermann (1995) and Lettau and Ludvigson (2001), find predictability, while the third, Bossaerts and Hillion (1999), does not. We show that it is possible to find a great degree of predictability, or none at all, in all three data sets by looping over relatively small ranges of just three econometrician choice variables from our reality spectrum (Figure 1). Minor changes in variables, assets, and estimation periods can result in strong rejections of the null; other minor changes in parameterization result in no rejections of the null.

D. The Use of the Best In-sample Model "Out-of-Sample"

If we consider that *all* of the variables in the three data sets have been used and continue to be used in the related literature, then the potential for data-snooping problems is large in light of the above evidence. As we mention in the introduction, there are multiple ways in which snooping may occur. One simple method, which carries with it no insidious implications whatsoever to the researcher involved, is the widespread practice of recursively testing the best model(s) from a series of in-sample tests using the same, or substantially the same data. In this section, we examine where the best in-sample predictive variables fall in the exogenous out-of-sample simulations' distributions. If the use of the best in-sample model in contemporaneous out-of-sample tests results in an upward bias, we would expect to observe that the best in-sample models fall in the upper right-hand tail of the exogenous simulations' profitability distribution.

The best in-sample models from the data used in the exogenous out-of-sample simulations, in terms of adjusted R^2, are CAY and RREL for data set 1 (R^2 of 11.2%) and dividend yield lagged once (DY), 1 month T-bill rate lagged once ($Tbill_{-1}$) and twice ($Tbill_{-2}$), 12 month T-bond rate lagged once ($Tbond_{-1}$), yearly inflation rate lagged twice (Π), change in industrial production lagged twice (ΔIP), and change in narrow money stock lagged twice (ΔM) for data set 2 (R^2 of 10.5%). For data set 3, the country with the highest in-sample R^2 is Japan, at 14.2%, from a model of monthly stock return of the local index in $US terms lagged once ($R_{i,-1}$), yield to maturity on a representative Treasury bond lagged once (YTM), price level of the country's market index lagged once (P_i), the stock market's dividend yield lagged once (DY_i), and the stock market's price-to-earnings ratio lagged once (PE_i).[15]

As we might expect, all three of these best in-sample models perform quite well in contemporaneous out-of-sample tests. For data set 1, the best in-sample model, across all window lengths, is always in the top 7% of the exogenous simulation distribution (based on Jensen's alpha). For data set 2, the best in-sample model, using an expanding window, is ranked 2nd out of 3577 runs on Jensen's alpha. Interestingly, when the best model is run out-of-sample for the other window lengths, they are at the 50th percentile or below for Jensen's alpha. For Japan, in data set 3, the best in-sample model, using a 10-year window, ranks 605th out of 7161 runs. None of the other window lengths for the best in-sample model are in the top 10% of the alpha distribution. For the other 12 countries, the results are similar. Thus, these results suggest that using variables that have "worked" over the entire sample period will bias the recursive out-of-sample performances of these variables towards providing evidence of predictability.

Overall, the fact that a subset of out-of-sample forecast combinations yield predictability when considered ex post is of little value in ascertaining if predictability could be discovered ex ante; in the real world, it is entirely possible that investors do not know ex ante the best predictive variables, the best in-sample window length, the most predictable return horizons, nor the most predictable assets. We explore these issues next.

III. Is it Real? Endogenizing the Econometrician Choice Variables

Our purpose in this section is to determine if the macro-economic based predictability documented in the previous section is genuine or not. To accomplish this task, we examine if predictability survives after recursively endogenizing the choice of the three parameters of

[15] These R^2's are determined from the single best predictive variable combination over the entire out-of-sample period for each dataset as reported in Table 1.

predictive variables, assets, and in-sample window length. The act of endogenizing the parameters makes the experiment more real-time – that is, it allows us to ask the question of whether an investor, operating without the benefit of full period information, can find the predictability that we now know exists ex post. We assume that one has no particularly strong priors concerning the identity of the optimal values of these three parameters, except for priors implicitly imposed by each data set. For example, in data sets 1 and 2 (Table 1), one has the prior to consider a specific group of variables along with only a single US asset. The prior on a fixed asset is relaxed in data set 3, when we include multiple assets. We use two techniques to endogenize the econometrician choice variables for our real-time forecasts. First, we use the mutual fund literature's technique of testing for persistence (see Jensen (1968), Grinblatt and Titman (1992), Carhart (1997), and others). Our second approach is to develop a recursive forecasting method, building on approaches in Pesaran and Timmermann (1995) and Bossaerts and Hillion (1999), to endogenize the econometrician choice variables. This method employs an in-sample period to choose the best forecasting model from the universe of potential models, and then uses the optimal model to form portfolios in step-ahead periods.

A. Endogenizing Via a Persistence Strategy

To implement the persistence strategy, we treat each of the exogenously specified out-of-sample forecasts from the snooping simulations above as "mutual funds." We test for persistence by ranking on prior performance of these funds and then examine performance in step-ahead periods.

Specifically, for each data set we use all out-of-sample return combinations from the exogenous simulations. We use a five-year ranking period to group the "funds" into deciles based on mean return in the ranking period. Within each decile, we equally weight each return stream and then track the performance of each decile portfolio for the next year. At the end of the year, we then re-rank, reform the portfolios, and track their performance for another year. We repeat this process until the end of the out-of-sample period. We report absolute and relative performance tests. For the absolute measures, we simply report the means, terminal wealths, Sharpe ratios, and alphas of each decile portfolio. Of course, the absolute measures may be upward biased within a given data set if the group of predictive variables, or assets, were themselves snooped. To address this, we also examine relative performance measures based on the percent change in performance across the ex post decile rankings of the exogenous simulations relative to the decile rankings of the persistence strategies.

The results of the persistence strategies are reported in Table 4. Panel A contains the results for data set 1, Panel B contains the results for data set 2, and Panel C contains the results for data set 3. The left hand side of each panel contains the results of a single ex post decile ranking of all exogenously specified out-of-sample runs from section 2.3, and the right hand side of each panel contains the results to the persistence strategy.

Across the three data sets, there is very little evidence of persistence in the raw returns of the decile portfolios. The spread in means between the lowest and highest decile in the persistence strategy is a quarterly 0.38%, monthly 0.13%, and monthly –0.14% for data sets 1, 2, and 3, respectively. The spreads are only statistically significant for data set 2, and only significant in one of the two means tests; we find a p-value of 0.06 for a chi-square test on the difference in average monthly returns between decile 1 and decile 10. The other chi-square test, a test on the spread in average monthly returns across all deciles is insignificant, with a p-value of 0.21[16] Similarly, on a risk-adjusted basis, only data set 2 shows evidence of predictability, with significant Jensen's and Fama-French alphas in Panel B of between 21 and 33 basis points per month for each decile, and statistically significant spreads across decile 1 and 10 for the two alpha measures. Interestingly, *all* deciles in Panel B have significant alphas, suggesting that if one had a prior to consider just the group of variables that make up data set 2, then we have found some genuine predictability. Another view, reinforced by the fact that the spread in alphas *across*

---

[16] We use a χ_1^2-statistic to test the null hypothesis of equality of monthly returns across the persistence strategy decile portfolios. Specifically, we use GMM with the following moment conditions to form the chi-square statistic:

$$\begin{Bmatrix} \varepsilon_1 = R_{p1} - \mu * 1 \\ \varepsilon_2 = R_{p2} - \mu * 1 \\ . \\ \varepsilon_{10} = R_{p10} - \mu * 1 \end{Bmatrix}$$

, where R_{pn} is a t x 1 time series of trades from the portfolio formed from the equally weighted average of "funds" in decile *n,* 1 is a column vector of ones, and μ is the mean return parameter to be estimated. The system of moment conditions is overidentified, with ten moment conditions and only one parameter to estimate. Thus, the resulting χ_9^2 statistic tests the null hypothesis of $\overline{R}_{p1} = \overline{R}_{p2} = ... = \overline{R}_{p10}$, where $\overline{R}_{pn}$ is the mean return to sort decile *n* for a given one-way sort. We also test the null hypothesis of equality of means for decile 1 and decile 10 using the moment conditions:

$$\begin{Bmatrix} \varepsilon_1 = R_{p1} - \mu * 1 \\ \varepsilon_{10} = R_{p10} - \mu * 1 \end{Bmatrix}$$

, where R_{pn} is a t x 1 time series of trades from the portfolio formed from the equally weighted average of "funds" in decile 1 or 10, 1 is a column vector of ones, and μ is the mean return parameter to be estimated. Thus, the resulting χ_1^2 statistic tests the null hypothesis of $\overline{R}_{p1} = \overline{R}_{p10}$. The statistics are robust to heteroskedasticity and autocorrelation (Gallant, 1987).

the deciles is relatively low (0.11 percent to 0.12%), is that this entire group of variables may in part be subject to a hindsight bias. To some extent this is supported in Panel B of Table 2 (the exogenous simulations) where we observe high rejection rates of the null across all models. For example, approximately 63% of the models experience significant positive forecast betas, 58% significant Jensen's alphas, 30% significant Fama-French Alphas, and 82% significant market timing. But nonetheless, an investor with a firm prior of conditioning on the predictive variables in data set 2, and trading the S&P500, would find evidence of risk-adjusted predictability in the alphas between the decile 1 and decile 10 persistence portfolios from Panel B, Table 4, of approximately 1.5% per annum.

When we consider the relative measures, that is, the percent change between the spread in the ex post runs versus the spreads in the persistence portfolios, we see big drops across all three data sets. For example, in data set 1, the ex post spread in raw quarterly returns between decile 1 and decile 10 is 1.12%, and the spread for the persistence portfolios is 0.38%, a 66% decrease. We also see large percent decreases for the other performance measures for data set 1, with this pattern continuing for data sets 2 and 3.[17] In data set 2, the one for which we do find significant positive alphas from the persistence strategy, we see a drop of approximately 69% (73%) for the Jensen's (Fama-French) alphas from the ex post runs to the persistence results. We see even larger decreases for data set 3, with the spread in alphas decreasing from an average of 1.37% and 1.35% for the Jensen's and Fama-French alphas in the ex post simulations, to –0.12 for both measures in the persistence results. Thus, relative to the best ex post exogenous out-of-sample combinations, which we might believe now in the year 2001 provide us with robust evidence of predictability, the persistence strategies show that endogenizing the choice of predictive variables, assets, and in-sample window lengths results in large decreases in predictability. In addition, the only data set for which we do find significant risk adjusted alphas, data set 2, requires specific priors on predictive variables and the S&P 500. If we relax that prior, and consider a different group of predictive variables (data set 1) or expand to 13 assets, as in data set 3, we see that all evidence of persistence disappears, suggesting that the strong evidence of market predictability from the snooping simulations are not evident in real-time.

A.1 The Curse of Implementation

[17] For data set 3, we do not report spreads in mean or terminal wealth for the ex post exogenous simulations since the 13 different assets do not cover the same period. Instead, as an ad hoc measures to compare across periods, we report the Sharpe ratio, Jensen's alpha, and Fama-French three-factor alphas.

An irony of our persistence strategies is that we must exogenously parameterize certain aspects of the experiment in order to implement the very experiments we use to endogenize the three econometrician choice variables! Thus, we fall prey to our critique of possible biases from exogenous parameter specification. For example, we must decide on the exact criteria used to group the exogenous out-of-sample "funds" into deciles. These criteria include both the length of the ranking period and the identity of the objective function used to rank. For our results in Table 4, we exogenously parameterize the ranking length at 5 years, and the ranking function as the mean return over the 5 year ranking period. Of course, those are not the only possible values, and in this section, we examine the results of the persistence strategies over an expanded range of values, including 1 and 3 year ranking periods, and terminal wealth and Sharpe ratio ranking functions.

In table 5 we report the spread in returns between decile 10 and decile 1 for the persistence strategies for the nine combinations of three ranking lengths (1, 3, and 5 years) and three ranking functions (mean, terminal wealth, and Sharpe ratio). For data set 1, there are two significant decile spreads, both at the one-year ranking length, for the mean and Sharpe ratio criteria. For data set 2, there are also two significant spreads, both at the 5-year ranking length for the mean and Sharpe ratio ranking functions. For the international data of data set 3, there are no significant spreads across the nine exogenous combinations. These results suggest the same message as before: if one works hard enough, one can find some evidence of predictability. If one has a prior of data set 1, a 1 year ranking period, and either a mean or Sharpe ratio ranking criteria, or of data set 2, a 5 year ranking criteria and either a mean or Sharpe ratio ranking criteria, then there is evidence of predictability, at a maximum of about 2.3% per annum. If one has no particular prior for these parameters, then the averages reported in the bottom row of table 5, obtained from averaging across the ranking length and ranking functions, suggest little evidence of predictability across all three data sets.

B. Endogenizing Via a Recursive Strategy

In this section we use variations on commonly employed recursive out-of-sample techniques to endogenize the assets, predictive variables, and in-sample window lengths. The main difference between the recursive approach in this section and the persistence strategies from the last section is that the recursive approach is based on identifying the best single strategy across variable groups, assets, and estimation lengths via optimizing an *in-sample* objective

function based on in-sample expected return estimates. In contrast, the persistence strategies use *out-of-sample* realized returns in the ranking period to identify the optimal strategies.[18]

To implement the recursive strategy, we identify the best combination of assets, predictive variables, and estimation lengths from an in-sample period, and then apply those optimal parameters in step-ahead periods to form an out-of-sample portfolio.

Specifically, we construct a single out-of-sample time series of returns for each dataset using the following recursive approach:

1. For all possible variable combinations, I (I=2^K-1 models (each model includes an intercept), where K=the number of predictive variables in each data set), all possible in-sample window lengths, W (W=10, 15, and 20 years of fixed moving windows and an *expanding* window for data set 1, and W=5, 6, 7, 8, 9, 10 years of fixed moving windows and an *expanding* window for data sets 2 and 3), and all possible assets, A (A=1 for data sets 1 and 2, and A=13 for data set 3), we estimate, using OLS, a linear model of the form $r_\tau = \beta_I' X_{\tau-1,I} + \varepsilon_{\tau,I}$ where $X_{\tau-1,I}$ is a (n_I+1) x 1 vector of predictive variables, including a vector of ones for the intercept term, and r_τ is the excess return for asset A during in-sample period τ. We estimate the model in the in-sample period W, and use the loadings on the predictive variables to form expected return estimates during the in-sample period. For each forecast series, we obtain a series of realized returns from the following trading strategy: go long asset A if the expected excess return estimate for that period is greater than zero, else invest in a t-bill. We then choose the best W, A, and I combination from the W x A x (2^K-1) total combinations based on the average in-sample terminal wealth, standardized by the number of periods in W.

2. Using the optimal model from above, we form a step ahead out-of-sample forecast using the in-sample intercept and predictive variable loadings.

3. We then roll forward the in-sample end date by one period, repeat steps 1 and 2, and obtain a forecast for the next out-of-sample period. We repeat this process until the end of the out-of-sample period. Thus, for each data set, we obtain a single out-of-sample forecast series.

4. For the out-of-sample forecast series, we obtain a series of realized returns for the "active" portfolio from the following trading strategy: go long in the optimal asset A if the expected excess return estimate for that period is great than zero, else invest in a t-bill.

The results are presented in Table 6.[19] Across the three data sets, and similar to the persistence results, we only find statistically significant predictability in data set 2. The results for

[18] Another approach to analyze the general robustness of time-series predictability would be to estimate full-period in-sample regressions and analyze subperiod stability of the predictive variables' betas. However, this approach would not endogenize the econometrician choice variables, that is, it would not allow for real-time competition of predictive variables, assets, and estimation lengths.

[19] We also examine mean and Sharpe ratio as the objective function. The results (not reported, but available from the authors) are very similar to the terminal wealth objective function.

data set 2, in Panel B, are about the same as in the persistence strategies, with the active portfolio averaging 1.05% per month, which is 13 basis points per month greater than a buy-and-hold position in the S&P500. Out of 348 months in the out-of-sample period, the active strategy trades 196 months. Although the active portfolio's raw mean is not that much greater then the S&P500, the standard deviation is lower, resulting in almost double the Sharpe ratio. In addition, the active portfolio exhibits a significant forecast beta, market-timing statistic and Jensen and Fama-French alphas.

We also endogenize various fixed transaction costs for the recursive experiment. We do this by altering the in-sample trading rule to "go long in asset *A* if the expected excess return is greater than zero plus the one-way transaction cost." Thus, under this setting, the optimal in-sample combination of assets, predictive variables, and window lengths is determined accounting for transaction costs, and then applied to the step-ahead out-of-sample period. To form the active out-of-sample portfolio, we also require the expected return estimate to be greater than zero plus the transaction costs. We consider one-way transaction costs of 10, 30, and 50 basis points. The economic evidence of predictability for data set 2 is much less strong after accounting for even the lowest level of one-way transaction costs of 10 basis points. At this transaction costs level, the active portfolio now has a lower Sharpe ratio relative to the no-transaction costs portfolio, and has statistically insignificant alphas, but retains a significant forecast beta and market timing measure. For data sets 1 and 3, endogenizing transaction costs does not help in improving the performance of the active portfolio.

Overall, the recursive method of endogenizing the econometrician choice variables provides us with results similar to the persistence method; finding predictability requires a prior for data set 2's predictive variables and the S&P500. If one does not have this prior, then there is little evidence of out-of-sample predictability from the other two data sets.

B.1 How Much Endogenizing is Enough?

In light of the large number of potential parameters that researchers typically exogenously specify in order to conduct time series predictability tests – we list 12 and test just 3 from our Reality Spectrum (Figure 1) – we ask the question in this section of how much endogeneity is enough? If one finds predictability after endogenizing say one, or two aspects, is that enough? If the goal were to be real "real-time," as in modeling the full spectrum of uncertainty that an actual investor faces, then likely all aspects would need to be endogenized. Obviously, this is not practical and probably impossible to implement. However, a recent positive

(in our opinion) trend in the time series literature has been to attempt to reduce ex post biases by endogenizing one or two aspects of real-time uncertainty. For example, as we discuss in section I, some papers have endogenized predictive variable selection (Pesaran and Timmermann (1995), Bossaerts and Hillion (1999), Avramov (1999), Ait-Sahalia and Brandt (2001), Pastor (2000), and Cremers (2000)), in-sample window length (Pesaran and Timmermann (1999)), and statistical model selection (Pesaran and Timmermann (1995)).[20] We will refer to these types of experiments as "univariate endogeneity." Obviously, if the conclusions from these univariate experiments are robust, we should see that variations over other *reasonable* econometrician choice parameters do not materially alter the outcomes.

In Table 7 we report the percentage of out-of-sample forecasts for which the null is rejected in experiments in which we endogenize one or more parameters, while exogenously looping over other parameter values. In panel A we endogenize variable selection via 6 statistical model selection criteria and loop over exogenous values of estimation windows and assets[21], in panel B we endogenize window length and loop over exogenous values of model selection criteria and assets, in panel C we endogenize model selection criteria and loop over exogenous values of estimation windows and assets, and in panel D, we endogenize both statistical model selection and windows, and examine variations across assets. We use the same recursive methodology as in section III.B, but now we first endogenize one aspect via a terminal wealth objective function, and then generate other out-of-sample portfolios by varying the values of the exogenous choice variables.[22]

The results in panels A, B, and C suggest exogenous parameter selection has a large effect on how often one rejects the null *even when other aspects are endogenized.* For example, in panel A, after endogenizing variable selection, data set 2 experiences significant rejections of the null in 16% to 74% of the exogenous model selection and window length specifications, depending on which performance measure is used. We observe similar patterns for all three data sets in panels A, B, and C. Across the data sets and panels, there does not appear to be any consistent pattern in which type of model selection or window length rejects the null. Lastly, in

[20]APT papers such as Roll and Ross (1980) and Dhrymes, Friend, and Gultekin (1984) use factor analysis to extract priced factors from historical returns. Thus, these papers can also be viewed as endogenizing predictive variables.

[21] The six criteria are Akaike's Information Criterion (AIC), Schwarz's Bayesian Information Criterion (SBIC), Sawa's Bayesian Information Criterion (BIC), Amemiya's Prediction Criteria (PC), Adjusted-R^2, and a model that uses all predictive variables, ALL.

[22] For example, consider Panel A of Table 7 in which we endogenize window length for each selection criteria and asset. During the in-sample period, for each selection criteria, we find the variable combination with the highest value of the selection criteria for each window length. Using the best model for each

panel D, we report the results from endogenizing both windows and model selection criteria. For data sets 1 and 2 this results in a single out-of-sample portfolio, and for data set 3 it results in one portfolio for each of the 13 countries. For data sets 1 and 2, the only evidence of predictability shows up in the market timing measure for data set 2; the other three measures show no evidence of significant predictability. In the last row of panel D, for data set 3, there is one country (France) that survives this process, generating significant alphas and market timing, but not a significant forecast beta.

Thus, similar to the snooping simulations reported earlier in the paper, it appears that in experiments in which one aspect is endogenized, the rejection of the null is heavily dependent upon the value of other exogenously specified parameters. And again, since these parameters vary across the range of possible values, it appears unlikely that one would posses an ex ante prior on the successful forecast combinations.

5. Discussion and Conclusion

Time-series based predictability is not evident in real-time, at least not for the commonly used predictive variables and asset combinations that we examine. We show that market predictability is largely an ex post phenomenon emanating from the exogenous specification of many aspects of uncertainty facing a real-time investor, such as predictive variables selection, length of the estimation period, assets, and other aspects.

Thus, the dangers of data-snooping are very real in the time-series-based market predictability literature; a researcher *will* find evidence of out-of-sample predictability by searching over exogenous combinations of predictive variables, in-sample estimation lengths, and assets. And it does not take an inordinate amount of searching; we present examples in which a researcher using the best subset of a group of predictive variables from in-sample tests always finds predictability in contemporaneous out-of-sample tests. Or viewed slightly differently, an investor with the correct specific sets of priors on predictive variables, assets, and estimation periods will find evidence of predictability. But since no real theory exists to guide one on the choice of the correct priors, finding this predictability seems unlikely.

Once we endogenize the choice of predictive variables, in-sample estimation lengths, and assets, using both a persistence and a recursive strategy, all predictability disappears. Overall, our results suggest that in order to minimize false rejections of the null hypothesis of no

---

window, we then find the window combination that results in the highest average terminal wealth. Thus, we arrive at the best window length for each of the six selection criteria and asset(s).

predictability, researchers should employ an out-of-sample methodology that endogenizes critical portfolio formation choice variables.

Our research is most closely related to the recent work of Foster, Smith and Whaley (1997).[23] Foster, Smith and Whaley also focus on tests of asset pricing models, but examine the biases in R^2 measures associated with a particular researcher choosing *k* fixed predictors from a larger set of *m* possible variables. They propose variations in the traditional in-sample tests that researchers use to assess whether a particular variable group can predict a single asset. Our work also warns about ad hoc selection of predictive variables, but we expand on their work to include the effects of uncertainty of other common parameters, as well as variable selection, in a real-time economic-significance out-of-sample setting. Our consideration of multiple sources of uncertainty is likely very important. For example, Foster, Smith and Whaley find that a few previous studies (pg. 603, Table IV) are likely to survive their data-snooping controls. But their tests do not allow for snooping effects across researchers, assets, variations in predictor variables, or estimation periods. In our tests, we examine many of the same predictive variables as the papers cited in Foster, Smith and Whaley, but we find that endogenizing these other aspects of uncertainty results in not finding significant predictability.

Our work is also closely related to Pesaran and Timmermann (1995) and Bossaerts and Hillion (1999). Both papers endogenize predictive variable selection using various statistical and economic based criteria. Pesaran and Timmermann find predictability in the US market, whereas Bossaerts and Hillion find none across 13 countries. Since we examine both of these papers' exact data sets, we are in a unique position to examine the robustness of their results in the face of additional real-time sources of uncertainty. We find that their results change across reasonable variations in exogenous parameters, making it unlikely that a real-time investor would converge on the successful parameter combinations.

Recent Bayesian papers have shown that portfolio allocations can depend critically on the level of investor uncertainty about the parameters of a given forecasting model. Kandel and Stambaugh (1996) were the first to highlight the importance of parameter uncertainty on portfolio allocations in a short-horizon, predictable return environment. Barberis (2000) shows that as the investment horizon increases, investors will give less weight to equities in an uncertain parameter world than in a world with parameter certainty. In both the Kandel and Stambaugh and the Barberis papers, predictability emanates from the dividend yield, making their results conditional

[23] In related work, Ferson, Sarkissian, and Simin (2000) examine the problem of spurious regression biases for predictive regressions. They find bias problems to be the greatest in the context of model selection, especially when the underlying expected return is highly autocorrelated.

on that specific model.[24] Cremers (2000) and Avramov (1999) incorporate investor uncertainty into beta estimation and predictive variable selection. Cremers finds that after endogenizing variable selection, he finds "some, albeit small" evidence of out-of-sample predictability. Avramov finds that the model-uncertainty component has more of an effect on optimal portfolio choices than does beta parameter uncertainty. Both the Cremers and the Avramov papers hint at our results in section III.B.1 *How Much Endogenizing is Enough?*; the more uncertainty one attempts to incorporate into a real-time forecasting model, the less sure the results become. Thus, an implication from our paper for these types of Bayesian studies, or any study that attempts to model time-series market predictability is that it is critically important to endogenize all possible aspects of real-time decision making into the experiment.[25]

In summary, the real-time methodology used in this paper does not suggest an alternative model of the factors that drive aggregate market returns. The power to detect real-time market predictability may be increased by incorporating other aspects of uncertainty that we have not considered, such as other predictive variables, different assets, multiple return horizons, non-linear models, different forms of learning, and other changes. But again, the parameterization of these features should be endogenized in a recursive manner, and not exogenously specified. Our results provide an explanation for the performance gap between mutual funds and the academic market predictability literature, and carry important implications for asset pricing models, cost-of-capital calculations, and portfolio management.

[24] In related *non*-Bayesian work, Goyal and Welch (1999) document substantial in-sample predictability in the time series of stock index returns based on dividend yields, but find no evidence of out-of-sample forecastability. They attribute the difference in performance between in-and out-of-sample predictability to parameter instability, i.e., a time-varying correlation between expected returns and dividend yield.

[25] In related work, Lewellen and Shanken (2001) argue that the Bayesian learning of economic agents can generate ex post predictable patterns that are ex ante rational and therefore not real-time tradable opportunities. In this case, predictability is just an ex post illusion. For example, suppose you *know* that the time-series of stock returns is mean-reverting. In real time, you still do not know if stock prices will be higher or lower next period because you do not know the true mean of the distribution. Nonetheless, a pattern of mean reversion is easily detected ex post relative to the *sample* mean.

REFERENCES

Ait-Sahalia, Y. and M. Brandt, 2001, "Variable Selection for Portfolio Choice," *Journal of Finance* 56, 1297-1351.

Allen, F. and R. Karjalainen, 1999, "Using Genetic Algorithms to Find Technical Trading Rules," *Journal of Financial Economics* 51, 245-271.

Andrews, D. 1991, "Heteroskedastic and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica* 59, 817-858.

Ang, A. and G. Bekaert, 2001, "Stock Return Predictability: Is it There?," working paper, Columbia University and NBER.

Avramov, D., 1999, "Stock Return Predictability and Model Uncertainty," Working Paper, University of Maryland.

Barber, B., R. Lehavy, M. McNichols, and B. Trueman, 2000, "Can Investors Profit from the Prophets?: Security Analysts' Recommendations and Stock Returns," forthcoming *Journal of Finance*.

Barber, B. and T. Odean, 2000, "Trading is Hazardous to Your Wealth: The Common Stock Investment Performance of Individual Investors," *Journal of Finance* 55, 773-806.

Barberis, N., 2000, "Investing for the Long Run When Returns are Predictable," *Journal of Finance* 55, 225-264.

Bekaert, G. and R. J. Hodrick, 1992, "Characterizing Predictable Components in Excess Returns on Equity and Foreign Exchange Markets," *Journal of Finance* 47, 467-509.

Black, F., 1993a, "Beta and return," *Journal of Portfolio Management* 20, 8-18.

Black, F., 1993b, "Estimating expected return," *Financial Analyst Journal* 49, 36-38.

Black, F., M. Jensen, and M. Scholes, 1972, "The capital asset pricing model: Some empirical tests," in M. Jensen. Ed.: *Studies in the Theory of Capital Markets* (Praeger).

Bossaerts, P., and P. Hillion, 1998, "IPO Post-Issue Markets: Questionable Predilections but Diligent Learners," unpublished manuscript.

Bossaerts, P., and P. Hillion, 1999, "Implementing Statistical Criteria To Select Return Forecasting Models: What do we learn?" *Review of Financial Studies* 12, 405-428.

Brandt, M., 1999, "Estimating Portfolio and Consumption Choice: A Conditional Euler Equations Approach," *Journal of Finance* 54, 1609-1646

Breen, W., L.Glosten, and R. Jagannathan, 1989, "Economic Significance of Predictable Variations in Stock Index Returns," *Journal of Finance* 44, 1177-1189.

Brown, S., W. Goetzmann, and A. Kumar, 1998, "The Dow Theory: William Peter Hamilton's Track Record Reconsidered," Journal of Finance 53, 1311-1333.

Brown, S., W. Goetzmann, S. Ross, 1995, "Survival," *Journal of Finance* 50, 853-873.

Campbell, J., 1987, "Stock Returns and the Term Structure," *Journal of Financial Economics* 18, 373-399.

Campbell, J. and R. Shiller, 1988a, "Stock Prices, Earnings, and Expected Dividends," *Journal of Finance* 43, 661-676.

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

Carhart, M., 1997, "On Persistence in Mutual Fund Performance," *Journal of Finance* 52, 57-82.

Chan, K., Y. Hamao, and J. Lakonishok, 1991, "Fundamentals and Stock Returns in Japan," *Journal of Finance* 46, 1739-1764.

Chan, K., H. Chen, and J. Lakonishok, 1999, "On Mutual Fund Investment Styles," NBER working paper 7215.

Chen, N., R. Roll, and R. Ross, 1986, "Economic Forces and the Stock Market," *Journal of Business*, 59, 383-403.

Christopherson, J., W. Ferson, and D. Glassman, 1998, "Conditioning Manager Alphas on Economic Information: Another Look at the Persistence of Performance," *Review of Financial Studies* 11, 111-142.

Cochrane, J., 1991, "Production Based Asset Pricing and the Link Between Stock Returns and Macroeconomic Fluctuations," *Journal of Finance* 46, 209-238.

Cochrane, J., 1999, "Portfolio Advice for a Multifactor World," *Economic Perspectives* Federal Reserve Bank of Chicago 23 (3) 59-78

Cooper, Michael, Roberto Gutierrez, and William Marcum, 2001, "On the Predictability of Stock Returns in Real Time," forthcoming *The Journal of Business.*

Coval, J. and T. Shumway, 2001, "Is sound just noise?" forthcoming *Journal of Finance*

Cremers, M., 2000, "Stock Return Predictability: A Bayesian Model Selection Perspective," forthcoming *Review of Financial Studies*.

Daniel, K. and S. Titman, 1997, "Evidence on the Characteristics of Cross Sectional Variation in Stock Returns," *Journal of Finance* 52, 1-33.

Denton, F., 1985, "Data mining as an industry," *Review of Economics and Statistics* 67, 124-127.

Desai, H. and P. Jain, 1995, "An Analysis of the Recommendations of the 'Superstar' Money Managers at Barron's Annual Roundtable, *Journal of Finance* 50, 1257-1274.

Dhrymes, P., I. Friend, and N. Gultekin, 1984, "A Critical Reexamination of the Empirical Evidence on the Arbitrage Pricing Theory," *Journal of Finance* 39, 323- 346.

Fama, E., 1991, "Efficient Capital Markets II," *Journal of Finance* 46, 1575-1643

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

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

Fama, E., and K. French, 1992, "The Cross Section of Expected Stock Returns," *Journal of Finance* 47, 427-465.

Fama, E., and K. French, 1993, "Common risk factors in the returns of stocks and bonds," *Journal of Financial Economics* 33, 3-56.

Fama, E., and K. French, 1996, "Multifactor Explanations of Asset Pricing Anomalies," *Journal of Finance* 51, 55-84.

Fama, E., and K. French, 1998, "Value versus Growth: The International Evidence," *Journal of Finance* 53, 1975-1999.

Ferson, W. 1990, "Are the Latent Variables in Time-Varying Expected Returns Compensation for Consumption Risk?," *Journal of Finance* 45, 397-430.

Ferson, W. and C. Harvey, 1991, "The variation of economic risk premiums," *Journal of Political Economy* 99, 385-415.

Ferson, W. and C. Harvey, 1993, "The Risk and Predictability of International Equity Returns," *Review of Financial Studies* 6, 527-566.

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

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

Ferson, W., S. Sarkissian, and T. Simin, 2000, "Spurious Regressions in Financial Economics?," working paper, University of Washington.

Foster, D., T. Smith, and R. Whaley, 1997, "Assessing the Goodness-of-Fit of Asset Pricing Models: The Distribution of the Maximal R^2," *Journal of Finance* 52, 591-607.

Gallant, R., 1987, *Nonlinear Statistical Models*, New York: Wiley.

Goetzmann, W. and P. Jorion, "Global Stock Markets in the Twentieth Century," *The Journal of Finance*

Goyal, A. and I. Welch, 1999, "The Myth of Predictability: Does the Dividend Yield Forecast the Equity Premium?" unpublished manuscript.

Grinblatt, M. and S. Titman, 1992, “The Persistence of Mutual Fund Performance,” *Journal of Finance* 47, 1977-1984.

Harvey, C., 1989, “Time-varying conditional covariance in tests of asset pricing models,” *Journal of Financial Economics* 24, 289-317.

Harvey, C., 1991, “The World Price of Covariance Risk,” *Journal of Finance* 46, 111-157.

Hau, H., 1999, “Information and Geography: Evidence from the German Stock Market,” working paper, ESSEC Graduate Business School CEPR.

Henriksson, R. and R. Merton, 1981, “On market timing and investment performance. II. Statistical procedure for evaluating forecasting skills,” *Journal of Business* 54, 513-533.

Hirshleifer, D., and T. Shumway, 2001, “Good Day Sunshine: Stock Returns and the Weather,” forthcoming *Journal of Finance*.

Hodrick, R., 1992, “Dividend Yields and Expected Stock Returns: Alternative Procedures for Inference and Measurement,” *Review of Financial Studies* 5, 357-386.

Jegadeesh, N. and S. Titman, 1993, “Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency,” *Journal of Finance* 48, 65-91.

Jegadeesh, N. and S. Titman, 2001, “Profitability of Momentum Strategies: An Evaluation of Alternative Explanations,” forthcoming *Journal of Finance*

Jensen, M., 1968, “The Performance of Mutual Funds in the Period 1945-1964,” *Journal of Finance* 23, 389-416.

Kandel, S. and R. Stambaugh, (1996) "On the Predictability of Stock Returns: An Asset Allocation Perspective," *Journal of Finance* 51, 385-424.

Keim, D., and A. Madhavan, 1997, “Execution Costs and Investment Performance: An Empirical Analysis of Institutional Equity Trades,” *Journal of Financial Economics*, 46, 265-292.

Keim, D. and R. Stambaugh, 1986, “Predicting Returns in the Stock and Bond Markets,” *Journal of Financial Economics*, 17, 357-390.

Lamont, O., 1998, “Earnings and Expected Returns,” *Journal of Finance* 53, 1563-1587.

Lettau, M. and S. Ludvigson, 2001, “Consumption, aggregate wealth and expected stock returns,” *Journal of Finance*, 56, 815-849.

Lewellen, J. 1999, "The time-series relations among expected return, risk, and book-to-market, *Journal of Financial Economics*, 54, 5-43.

Lewellen, J. and J. Shanken, 2001, "Learning, Asset-Pricing Tests, and Market Efficiency," forthcoming *Journal of Finance*.

Lo, A. and C. MacKinlay, 1990, “Data-snooping biases in tests of financial asset pricing models,” *Review of Financial Studies*, 3, 431-467.

Lo, A., and A. MacKinlay, 1997, "Maximizing Predictability in the Stock and Bond Markets," *Macroeconomic Dynamics* 1, 102-134.

Metrick, A., 1999, “Performance Evaluation with Transactions Data: The Stock Selection of Investment Newsletters,” *Journal of Finance* 54, 1743-1775.

Odean, P., 1999, "Do Investors Trade Too Much?", *American Economic Review* 89, 1279-1298.

Pastor, L., 2000, “Portfolio selection and asset pricing models,” *Journal of Finance* 55, 179-223

Pastor, L. and R. Stambaugh, 2000, "Comparing Asset Pricing Models: an Investment Perspective," *Journal of Financial Economics* 56, 335-381.

Pesaran, M. and A. Timmermann, 1995, “Predictability of Stock Returns: Robustness and Economic Significance,” *Journal of Finance* 50, 1201-1228.

Pesaran, M. and A. Timmermann, 1999, “Model Instability and Choice of Observation Window,” working paper, University of California, San Diego

Pirinsky, C., 2001, “Are Financial Institutions Better Investors,” manuscript, Ohio State University.

Pontiff, J., and L. Schall, 1998, "Book-to-Market Ratios as Predictors of Market Returns," *Journal of Financial Economics* 49, 141-160.

Poterba, J. and L. Summers, 1988, “Mean Reversion in Stock Prices: Evidence and Implications,” *Journal of Financial Economics* 22, 27-59.

Roll, R. and S. Ross, 1980, “An Empirical Investigation of the Arbitrage Pricing Theory,” *Journal of Finance* 35, 1073-1103

Ross, S., 1989, “Regression to the max,” Working Paper, Yale School of Organization and Management.

Santos, T. and P. Veronesi, 2001, “Labor Income and Predictable Stock Returns,” working paper, University of Chicago.

Shiller, R., 1984, “Stock prices and social dynamics,” *Brookings Papers on Economic Activity* 2, 457-498.

Shleifer, A. and R. Vishny, 1997, “The Limits of Arbitrage,” *Journal of Finance* 52, 32-55.

Sullivan, R., A. Timmermann, and H. White, 1999, “Data-Snooping, Technical Trading Rule Performance, and the Bootstrap,” *Journal of Finance* 54, 1647-1691.

Swanson, N., and H. White, 1997, “A Model Selection Approach to Real-Time Macroeconomic Forecasting Using Linear Models and Artificial Neural Networks,” *Review of Economics and Statistics* 79, 540-550.

Wermers, R., 2000, "Mutual Fund Performance: An Empirical Decomposition into Stock-Picking Talent, Style, Transactions Costs, and Expenses," *Journal of Finance* 55, 1655-1695.

Figure 1
The reality spectrum of out-of-sample forecasts

	Reality Level:				
	HIGH (Out-of-sample methodology)	MORE (Out-of-sample methodology)	SOME (Out-of-sample methodology)	LOW (Out-of-sample methodology)	NONE (In-sample methodology)
Econometrician Choice Variables					
Major					
Predictive Variables:	Endog. Real time expanding	Endog. Fixed	Endog. Fixed	Exog. Fixed	Exog. Fixed
Assets:	Endog. Real time expanding including failed assets	Endog. Multiple	Exog. Multiple	Exog. Single	Exog. Single or multiple
In-sample Window Lengths:	Endog. Many	Endog. Multiple	Exog. Multiple	Exog. Single	Exog. Single
Minor					
Model Selection:	Endog. Many Competing criteria	Endog. Competing criteria	Endog. Single criterion	Exog. None	Exog. None
Trading rule:	Endog. Many rules	Endog. Multiple rules	Exog. Single rule	Exog. Single rule	Exog. Single rule
Return Horizon:	Endog. Multiple	Exog. Multiple	Exog. Single	Exog. Single	Exog. Single
Forecast Update Frequency:	Endog. Recursive	Exog. Recursive	Exog. Single holdout period	Exog. Single holdout period	Exog. In-sample method
Study Period:	Endog. Many Subperiods	Endog. Subperiods	Exog. Subperiods	Exog. Fixed	Exog. Fixed
Test of Null:	Multiple tests	Multiple tests	Single test	Single test	Single test
Learning:	Yes	Yes	No	No	No
Transaction costs:	Endog. Time-varying	Endog. Fixed	Exog. Fixed	No	No
Technology:	Endog. Real-time Expanding	Endog. Multiple	Exog. Multiple	Exog. Fixed	Exog. Fixed

"Exog." implies that the researcher exogenously specifies this feature. "Endog." implies that this feature is endogenously determined by the data.

Table 1
Data

Table 1 provides a summary of the data used in the out-of-sample forecasts. Using Data set 1, we forecast the quarterly excess returns of the S&P 500 using seven predictors: estimated trend deviation in consumption (CAY), S&P excess return lagged once (SPX), dividend yield lagged once (DY), dividend payout ratio lagged once (DP), relative T-bill rate, calculated as the 30-day T-bill rate minus its 12-month moving average lagged once (RREL), the term spread (10-year T-bond yield less 1-year T-bond yield) lagged once (TRM), and the default spread, calculated as the yield difference between BAA and AAA corporate bonds lagged once (DEF).

Using Data set 2, we forecast the monthly excess returns of the S&P 500 using nine predictors: dividend yield lagged once (DY), S&P 500 aggregate earnings-to-price ratio lagged once (EP), 1 month T-bill rate lagged once ($Tbill_{-1}$) and twice ($Tbill_{-2}$), 12 month T-bond rate lagged once ($Tbond_{-1}$) and twice ($Tbond_{-2}$), yearly inflation rate lagged twice (Π), change in industrial production lagged twice (ΔIP), and change in narrow money stock lagged twice (ΔM).

Using Data set 3, we forecast the monthly excess returns, in $US terms, of 13 countries' monthly indices, using ten predictors specific to each country. The 13 indices are: the S&P 500 (US) and the country indices reported by MSCI for the remaining 12 countries. The ten predictors are: a January dummy (JAN), monthly stock return of the local index in $US terms lagged once ($R_{i,-1}$), monthly stock return of the local index in $US terms lagged twice ($R_{i,-2}$), monthly bond excess return lagged once ($R_{Bi,-1}$), monthly bond excess return lagged twice ($R_{Bi,-2}$) yield to maturity on a representative Treasury bond lagged once (YTM), price level of the country's market index lagged once (P_i), the yield-to-maturity on a three-month Treasury bill lagged once ($Tbill_i$), the stock market's dividend yield lagged once (DY_i), and the stock market's price-to-earnings ratio lagged once (PE_i). For each dataset, we report the initial in-sample, final in-sample, and out-of-sample periods.

Data Set	Index	Data Period			Variables
		In-Sample		Out-of-Sample	
		Initial Period	Final Period		
1	S&P 500	1953(9)-1973(6)	1953(9)-1997(12)	1973(9)-1998(3)	CAY, SPX, DY, DP, RREL, TRM, DEF
2	S&P 500	1954(1)-1963(12)	1954(1)-1992(11)	1964(1)-1992(12)	DY, EP, $Tbill_{-1}$, $Tbill_{-2}$, $Tbond_{-1}$, $Tbond_{-2}$, Π, ΔIP, ΔM
3	Australia	1971(4)-1981(3)	1971(4)-1995(4)	1981(4)-1995(5)	JAN, $R_{i,-1}$, $R_{i,-2}$, $R_{Bi,-1}$, $R_{Bi,-2}$, $YTM_{Bi,}$ P_i, $Tbill_i$, DY_i, PE_i
	Belgium	1981(3)-1991(2)	1981(3)-1995(4)	1991(3)-1995(5)	
	Canada	1980(1)-1979(12)	1970(1)-1995(4)	1980(1)-1995(5)	
	France	1979(3)-1989(2)	1979(3)-1995(4)	1989(3)-1995(5)	
	Germany	1970(4)-1980(3)	1979(4)-1995(4)	1980(4)-1995(5)	
	Italy	1973(4)-1983(3)	1973(4)-1995(4)	1983(4)-1995(5)	
	Japan	1981(4)-1991(3)	1981(4)-1995(4)	1991(4)-1995(5)	
	Netherlands	1971(4)-1981(3)	1971(4)-1995(4)	1981(4)-1995(5)	
	Spain	1978(2)-1988(1)	1978(2)-1995(4)	1988(2)-1995(5)	
	Sweden	1982(4)-1992(3)	1982(4)-1995(4)	1992(4)-1995(5)	
	Switzerland	1980(1)-1989(12)	1980(1)-1995(4)	1990(1)-1995(5)	
	UK	1970(9)-1989(8)	1970(9)-1995(4)	1980(9)-1995(5)	
	US	1970(1)-1979(12)	1970(1)-1995(4)	1980(1)-1995(5)	

Table 2
Out-of-Sample Simulation Results with Exogenously Specified Predictor Variables and In-Sample Window Lengths

This table presents the percentage of out-of-sample forecasts rejecting the null hypothesis of no predictability under various performance measures. The out-of-sample forecasts are formed from all exogenous combinations of predictive variables and in-sample window lengths for three data sets and are based on a recursive methodology. For each out-of-sample forecast combination, we obtain a series of realized returns from the following trading strategy: go long in the traded asset if the expected excess return estimate is great than zero, else invest in a t-bill. Panel A reports results for dataset 1, Panel B reports results for dataset 2, and Panel C reports results for dataset 3. The performance measures are the forecast beta (β_f), Jensen's alpha, Fama French (1993) three-factor alpha (FF alpha), and the market timing statistics of Henriksson and Merton (1981) (HM p and HM $_{p1+p2}$). Rejection rates are reported at a 5% or better significance level. The rejection rates reported for the Jensen's Alpha and FF alpha are based on two conditions; the alpha of the out-of-sample portfolio must be greater than the alpha of the buy-and-hold portfolio and the alpha of the out-of-sample portfolio must be significant at the 5% or better level. The coefficient estimate of the slope (β_f) provides a measure of overall out-of-sample fit and is calculated by regressing the monthly realized return on the forecasted return: $r_\tau = \alpha + \beta_f r_{forecast,\tau} + \varepsilon_\tau$. For β_f we report the percentage of forecasts with positive betas and significance betas at the 5% or better level. At the bottom of panels A and B we report the model specifications that result in highest and lowest value of Terminal Wealth (TW) along with the corresponding terminal wealth values for the buy-and-hold strategies (TW_{bh}).

Panel A: Data set 1. All $\binom{7}{K}$ Forecasting Model Combinations of Seven Predictive Variables and Four Window Lengths (10 years, 15 years, 20 years, and Expanding)

We examine all $\binom{7}{K}$ forecasting model combinations, where K=1,2,...,7, of the seven predictive variables and 4 window lengths of 10 years, 15 years, 20 years, and EXPANDING, for a total of $4*(2^7-1)=508$ out-of-sample return series from 1973(9)-1998(3). The traded asset is the S&P500, using quarterly returns.

		Percent of out-of-sample forecasts rejecting the null under the following criteria:			
Prespecified Variable	Number of Specifications	Forecast Beta	Jensen's Alpha	FF Alpha	Market Timing
		$\beta_f>0$ ($p_{\beta f} \le 0.05$)	$\alpha_j > \alpha_{j,bh}$ ($p_\alpha \le 0.05$)	$\alpha_{ff} > \alpha_{ff,bh}$ ($p_{\alpha ff} \le 0.05$)	$HM_{p1+p2}>1$ ($HM_p \le 0.05$)
1 variable	28	28.6%	25.0%	14.3%	32.1%
2 variables	84	27.4%	21.4%	4.8%	19.0%
3 variables	140	28.6%	13.6%	0.7%	6.4%
4 variables	140	28.6%	7.9%	0.7%	3.6%
5 variables	84	19.0%	6.0%	0.0%	0.0%
6 variables	28	3.6%	3.6%	0.0%	0.0%
7 variables	4	0.0%	0.0%	0.0%	0.0%
10 years	127	14.2%	7.1%	2.4%	8.7%
15 years	127	26.0%	8.7%	2.4%	9.4%
20 years	127	40.9%	14.2%	2.4%	10.2%
Expanding	127	19.7%	18.1%	0.8%	2.4%
	508	25.2%	12.0%	2.0%	7.7%

Model with highest TW: Variables: CAY, RREL; Window: 10 years; TW: $40.03; TW_{bh}: $18.99
Model with lowest TW: Variables: DY, DEF; Window: 10 years; TW: $5.68; TW_{bh}: $18.99

Table 2, Continued

Panel B: Data set 2. All $\binom{9}{K}$ Forecasting Model Combinations of Seven Predictive Variables and Seven Window Lengths (5 years, 6 years, 7 years, 8 years, 9 years, 10 years, and Expanding)

We examine all $\binom{9}{K}$ forecasting model combinations, where K=1,2,...,9, of the nine predictive variables and 7 window lengths of 5 years, 6 years, 7 years, 8 years, 9 years, 10 years, and EXPANDING, for a total of $7*(2^9-1)=3{,}577$ out-of-sample return series from 1964(1)-1992(12). The traded asset is the S&P500, using monthly returns.

		Percent of out-of-sample forecasts rejecting the null under the following criteria:			
Prespecified Variable	Number of Specifications	Forecast Beta	Jensen's Alpha	FF Alpha	Market Timing
		$\beta_f>0$ ($p_{\beta f} \leq 0.05$)	$\alpha_j > \alpha_{j,bh}$ ($p_\alpha \leq 0.05$)	$\alpha_{ff} > \alpha_{ff,bh}$ ($p_{\alpha ff} \leq 0.05$)	$HM_{p1+p2}>1$ ($HM_p \leq 0.05$)
1 variable	63	9.5%	28.6%	3.2%	66.7%
2 variables	252	29.4%	46.4%	23.8%	77.4%
3 variables	588	48.0%	57.1%	32.8%	80.3%
4 variables	882	61.5%	57.4%	32.8%	81.6%
5 variables	882	71.4%	53.9%	31.2%	82.9%
6 variables	588	77.2%	53.2%	27.7%	85.9%
7 variables	252	82.9%	48.8%	25.4%	86.5%
8 variables	63	90.5%	38.1%	17.5%	84.1%
9 variables	7	85.7%	14.3%	14.3%	85.7%
5 years	511	91.6%	37.2%	14.1%	51.9%
6 years	511	51.7%	55.0%	23.5%	75.5%
7 years	511	68.1%	51.9%	21.7%	83.8%
8 years	511	65.9%	55.0%	22.9%	84.5%
9 years	511	71.8%	63.8%	26.6%	90.8%
10 years	511	71.4%	51.9%	22.9%	92.2%
Expanding	511	71.6%	88.8%	75.3%	97.1%
	3577	63.1%	57.6%	29.6%	82.2%

Model with highest TW: Variables: EP, $Tbill_{-1}$, $Tbond_{-1}$, $Tbond_{-2}$, Π, ΔIP, ΔM; Window: Expanding; TW: \$84.25; TW_{bh}: \$17.55
Model with lowest TW: Variables: EP, IP, ΔM; Window: 10 years; TW: \$5.58; TW_{bh}: \$17.55

Table 2, Continued

Panel C: Data set 3. All $\binom{10}{K}$ Forecasting Model Combinations Ten Predictive Variables and Seven Window Lengths (5 years, 6 years, 7 years, 8 years, 9 years, 10 years, and EXPANDING) for 13 Countries.

We examine all $\binom{10}{K}$ forecasting model combinations, where K=1,2,...,10, of the ten predictive variables and 7 window lengths of 5 years, 6 years, 7 years, 8 years, 9 years, 10 years, and EXPANDING, for 13 countries, for a total of $13*7*(2^{10}-1)=93{,}093$ (or 7,161 for each country) out-of-sample return series from 1980(1)-1995(5). The traded assets are the S&P500 monthly returns for the US, and the $US denominated MSCI monthly returns for each of the twelve other countries.

		Percent of out-of-sample forecasts rejecting the null under the following criteria:			
Prespecified Variable	Number of Specifications	Forecast Beta	Jensen's Alpha	FF Alpha	Market Timing
		$\beta_f>0$ ($p_{\beta f} \leq 0.05$)	$\alpha_j > \alpha_{j,bh}$ ($p_\alpha \leq 0.05$)	$\alpha_{ff} > \alpha_{ff,bh}$ ($p_{\alpha ff} \leq 0.05$)	$HM_{p1+p2}>1$ ($HM_p \leq 0.05$)
1 variable	910	0.1%	2.3%	0.2%	2.6%
2 variables	4095	0.5%	3.0%	0.4%	3.3%
3 variables	10920	0.9%	2.8%	0.7%	3.9%
4 variables	19110	1.1%	2.7%	0.8%	4.3%
5 variables	22932	1.1%	2.5%	1.0%	4.7%
6 variables	19440	1.0%	2.6%	1.0%	5.0%
7 variables	10920	0.9%	2.7%	1.1%	5.5%
8 variables	4095	1.0%	2.9%	1.0%	6.6%
9 variables	910	1.5%	3.1%	1.0%	7.9%
10 variables	91	2.2%	3.3%	1.1%	6.6%
5 years	13299	0.1%	2.2%	1.1%	4.8%
6 years	13299	0.2%	2.4%	1.1%	4.2%
7 years	13299	0.8%	2.9%	0.9%	4.2%
8 years	13299	0.9%	3.0%	1.1%	4.5%
9 years	13299	2.1%	3.2%	1.0%	6.7%
10 years	13299	1.5%	2.7%	0.6%	4.9%
Expanding	13299	1.4%	2.3%	0.5%	3.7%
Total	93093	1.0%	2.7%	0.9%	4.7%

Table 3
Out-of-Sample Performances of the Top 20 and Bottom 20 Exogenously Specified Models as Defined by Terminal Wealth

Panel A: Data set 1.

Top 20 and bottom 20 terminal wealth results of all $\binom{7}{K}$ forecasting model combinations, where K=1,2,...,7, of the seven predictive variables and 4 window lengths of 10 years, 15 years, 20 years, and EXPANDING, for a total of $4*(2^7-1)=508$ out-of-sample return series from 1973(9)-1998(3). The traded asset is the S&P500, using quarterly returns. The regression specifications (Model) of the top and bottom 20 models are represented as a sequence of zeros and ones. If a variable is included in the model it is represented with 1, otherwise it takes a value of 0. The predictor variables are in the following order: an intercept, CAY, SPX, DY, DP, RREL, TRM, and DEF.

Rank on TW	Model	Window (Years)	TW ($)	Quarterly Mean (%)	Sharpe Ratio	Forecast Beta	Jensen's Alpha (%)	FFalpha (%)	HMp	HMp1+p2
1	11000100	10	40.03	3.97	0.38	0.68**	1.14***	1.1***	0.045	1.154
2	10010111	20	37.79	3.89	0.38	0.26	1.15***	1.11***	0.031	1.192
3	11000101	Expanding	37.61	3.90	0.36	0.57**	1.08***	1.07**	0.067	1.126
4	11000100	Expanding	37.61	3.90	0.36	0.65***	1.08***	1.07**	0.067	1.126
5	10010110	20	37.49	3.88	0.38	0.35*	1.14***	1.1***	0.074	1.133
6	10001000	10	37.33	3.85	0.41	0.43**	1.31***	1.64***	0.031	1.192
7	10100100	20	35.76	3.84	0.36	0.53*	1.06***	1.01**	0.006	1.268
8	11000110	10	35.43	3.82	0.38	0.42*	1.14***	1**	0.008	1.254
9	11000110	Expanding	35.13	3.83	0.35	0.50**	1.02**	0.95**	0.088	1.111
10	11100100	Expanding	35.13	3.83	0.35	0.62**	1.02**	0.95**	0.088	1.111
11	11000111	Expanding	35.03	3.83	0.35	0.44**	1.01**	0.95**	0.109	1.096
12	11000000	20	34.95	3.83	0.35	0.60**	0.97**	0.92**	0.115	1.093
13	10100010	Expanding	33.95	3.84	0.32	0.34	0.87**	0.88**	0.013	1.200
14	11100111	Expanding	33.88	3.79	0.35	0.42**	0.99**	0.95**	0.092	1.109
15	10110110	20	33.68	3.77	0.36	0.32	1.04**	1**	0.105	1.103
16	11100101	Expanding	33.09	3.77	0.34	0.54**	0.96**	0.9**	0.109	1.095
17	10110111	20	33.07	3.75	0.36	0.23	1.02**	0.98**	0.076	1.132
18	10011110	20	32.88	3.75	0.35	0.39**	0.99**	0.99**	0.089	1.118
19	10111110	20	32.88	3.75	0.35	0.38**	0.99**	0.99**	0.089	1.118
20	10000100	20	32.68	3.75	0.34	0.58**	0.93**	0.89**	0.023	1.209
Closest to benchmark:										
273	10000001	Expanding	18.99	3.36	0.20	-0.03	-0.27*	-0.25**	NA	1.000
489	10010001	Expanding	10.22	2.64	0.13	-0.02	-0.61**	-0.44	0.018	0.793
490	10010000	Expanding	10.22	2.64	0.13	-0.06	-0.61**	-0.44	0.018	0.793
491	11111101	Expanding	10.21	2.60	0.13	0.33	-0.41**	-0.38	0.037	0.820
492	10010011	10	10.21	2.57	0.14	0.13	-0.29**	-0.5	0.047	0.831
493	10111101	15	10.11	2.62	0.13	0.25	-0.61**	-0.49	0.036	0.821
494	11111101	20	10.03	2.60	0.13	0.36*	-0.52**	-0.42	0.014	0.776
495	10110100	10	9.83	2.60	0.12	0.19	-0.56**	-0.76*	0.168	1.010
496	10010111	10	9.46	2.50	0.12	0.08	-0.44**	-0.64	0.101	0.891
497	11111101	15	9.28	2.52	0.11	0.29	-0.63**	-0.53	0.012	0.778
498	10010100	10	8.77	2.49	0.11	0.21	-0.68**	-0.87**	0.161	0.966
499	11011101	15	8.65	2.45	0.10	0.32	-0.72**	-0.69*	0.002	0.718
500	11110000	10	8.63	2.46	0.11	0.06	-0.62**	-1.02**	0.169	0.982
501	10110000	Expanding	8.60	2.46	0.10	0.09	-0.77**	-0.69*	0.010	0.761
502	10110001	Expanding	8.60	2.46	0.10	0.09	-0.77**	-0.69*	0.010	0.761
503	10010000	10	7.99	2.41	0.09	-0.57	-0.90**	-1***	0.136	0.925
504	10110101	10	7.74	2.36	0.09	0.08	-0.83**	-0.93**	0.168	0.995
505	10110000	10	7.60	2.36	0.08	-0.05	-0.96**	-1.12***	0.074	0.865
506	10010101	10	6.99	2.25	0.07	0.09	-0.95**	-1.09***	0.142	0.936
507	10110001	10	5.69	2.04	0.04	0.02	-1.13**	-1.22***	0.084	0.875
508	10010001	10	5.69	2.04	0.04	-0.04	-1.17**	-1.25***	0.072	0.861

Table 3, Continued

Panel B: Data Set 2

Top 20 and bottom 20 terminal wealth results of all $\binom{9}{K}$ forecasting model combinations, where K=1,2,...,9, of the nine predictive variables and 7 window lengths of 5 years, 6 years, 7 years, 8 years, 9 years, 10 years, and EXPANDING, for a total of $7*(2^9-1)=3,577$ out-of-sample return series from 1964(1)-1992(12). The traded asset is the S&P500, using monthly returns. The regression specifications (Model) of the top and bottom 20 models are represented as a sequence of zeros and ones. If a variable is included in the model it is represented with 1, otherwise it takes a value of 0. The predictor variables are in the following order: an intercept, DY, EP, $Tbill_{-1}$, $Tbill_{-2}$, $Tbond_{-1}$, $Tbond_{-2}$, Π, ΔIP, and ΔM.

Rank on TW	Model	Window (Years)	TW ($)	Monthly Mean (%)	Sharpe Ratio	Forecast Beta	Jensen's Alpha (%)	FF alpha (%)	HMp	HMp1+p2
1	1011011111	Expanding	84.25	1.33	0.27	0.62***	0.63***	0.61***	0	1.2392
2	1101110111	Expanding	77.47	1.30	0.26	0.59***	0.61***	0.59***	0.00001	1.2304
3	1011101111	Expanding	71.00	1.28	0.24	0.61***	0.57***	0.59***	0.00002	1.2127
4	1110101101	Expanding	70.40	1.28	0.25	0.60***	0.58***	0.60***	0.00004	1.2091
5	1101100001	Expanding	69.58	1.27	0.25	0.72***	0.58***	0.59***	0	1.2564
6	1101101001	Expanding	68.59	1.27	0.25	0.68***	0.57***	0.58***	0.00001	1.2312
7	1111110001	Expanding	67.12	1.26	0.24	0.64***	0.56***	0.59***	0.00008	1.1979
8	1101100101	Expanding	67.11	1.26	0.24	0.68***	0.56***	0.57***	0.00002	1.2147
9	1101100111	Expanding	65.49	1.26	0.24	0.62***	0.56***	0.55***	0.00004	1.2047
10	1110110101	Expanding	65.45	1.25	0.24	0.58***	0.56***	0.57***	0.00014	1.1899
11	1011100111	Expanding	65.15	1.27	0.22	0.63***	0.52***	0.54***	0.00001	1.2256
12	1101110001	Expanding	64.41	1.25	0.25	0.64***	0.56***	0.56***	0	1.2372
13	1100100101	Expanding	64.04	1.25	0.24	0.65***	0.55***	0.56***	0.00011	1.1947
14	1111101101	Expanding	63.23	1.25	0.23	0.64***	0.53***	0.56***	0.00022	1.1791
15	1101110101	Expanding	62.00	1.24	0.23	0.64***	0.54***	0.55***	0.00001	1.2224
16	1110101111	Expanding	61.66	1.23	0.24	0.57***	0.55***	0.54***	0.00014	1.1899
17	1110100101	Expanding	61.51	1.24	0.23	0.61***	0.54***	0.55***	0.00021	1.1847
18	1100110101	Expanding	61.29	1.24	0.24	0.61***	0.54***	0.55***	0.00007	1.2011
19	1110111101	Expanding	61.17	1.23	0.24	0.56***	0.54***	0.56***	0.00019	1.1859
20	1110110001	Expanding	60.90	1.23	0.24	0.59***	0.55***	0.56***	0.00002	1.2188
Closest to benchmark:										
2785	1100001010	5	17.55	0.88	0.11	0.20	0.152	0.13	0.07018	0.9643
3558	1000000011	6	8.28	0.67	0.05	-0.20	-0.07	-0.05	0.07741	0.9724
3559	1010010010	6	8.25	0.66	0.05	-0.12	-0.04	-0.04	0.05631	0.9503
3560	1010000011	5	8.19	0.64	0.05	-0.15	0	0.04	0.05352	0.9475
3561	1010000010	6	8.13	0.66	0.04	-0.19	-0.05	-0.004	0.06182	0.9559
3562	1010010111	5	8.02	0.64	0.04	-0.17	-0.04	-0.07	0.04058	0.9339
3563	1010000011	9	8.02	0.65	0.04	-0.41**	-0.06	-0.03	0.08277	0.9832
3564	1110000011	9	8.01	0.66	0.04	-0.15	-0.07	-0.05	0.08325	0.9792
3565	1000000011	10	7.92	0.66	0.04	-0.23	-0.09	-0.06	0.09072	0.9888
3566	1101000001	5	7.77	0.66	0.04	-0.07	-0.11	-0.12	0.06803	0.9611
3567	1110010111	8	7.71	0.64	0.04	0.02	-0.07	-0.13	0.07994	0.9796
3568	1110010111	7	7.70	0.64	0.04	-0.10	-0.08	-0.15	0.0756	0.9724
3569	1110010011	8	7.54	0.63	0.04	0.06	-0.07	-0.11	0.08549	1.0056
3570	1110010111	5	7.53	0.62	0.04	-0.11	-0.05	-0.03	0.0754	0.9724
3571	1010010011	5	7.20	0.61	0.03	-0.17	-0.06	-0.11	0.04503	0.9387
3572	1110000010	5	7.18	0.61	0.03	-0.01	-0.07	-0.05	0.03949	0.9327
3573	1110000011	6	7.12	0.62	0.03	-0.14	-0.11	-0.13	0.08332	0.9824
3574	1000000011	5	6.93	0.62	0.03	-0.17	-0.12	-0.09	0.05247	0.9463
3575	1010010111	7	6.78	0.60	0.03	-0.20	-0.09	-0.14	0.07928	0.9784
3576	1010010011	7	6.69	0.60	0.03	-0.20	-0.1	-0.16	0.08472	0.9912
3577	1010000011	10	5.58	0.54	0.01	-0.36	-0.16	-0.12	0.05843	0.9523

Table 3, Continued

Panel C: Data Set 3

Single best and worst terminal wealth result of all $\binom{10}{K}$ combinations, where K=1,2,...,10, of the nine predictive variables on each of the 13 country indices. This is interacted with 7 window lengths of 5 years, 6 years, 7 years, 8 years, 9 years, 10 years, and EXPANDING, for a total of $13*7*(2^{10}-1)=93{,}093$ (or 7,161 for each country) active trading rule out-of-sample return series from 1980(1)-1995(5). The model specifications that result in highest and lowest value of Terminal Wealth are reported for each country. The regression specification (Model) of the best and worst performing model is represented as a sequence of zeros and ones. If a variable is included in the model it is represented with 1, otherwise it takes a value of 0. The predictor variables are in the following order: an intercept, JAN, $R_{i,-1}$, $R_{i,-2}$, $R_{Bi,-1}$, $R_{Bi,-2}$, $YTM_{Bi,}$ P_i, $Tbill_i$, DY_i, and PE_i.

	Buy-and-Hold	Model with Highest TW			Model with Lowest TW		
	TW_{bh} ($)	Window (Years)	Model	TW ($)	Window (Years)	Model	TW ($)
Australia	3.75	10	10001010101	8.49	5	11111110010	0.85
Belgium	1.62	Expanding	11010111000	2.01	5	11100010100	0.96
Canada	3.36	5	10011110101	10.05	Expanding	11110001100	1.17
France	2.10	9	10000111110	3.98	8	11110001011	0.95
Germany	9.23	9	11000111001	16.17	8	10111011010	2.04
Italy	2.01	9	11011101110	12.67	Expanding	11010000011	0.73
Japan	1.16	8	10110000101	1.83	6	11010000000	0.57
Netherlands	14.57	5	11011110101	20.93	Expanding	11101100110	4.65
Spain	1.41	6	11010101001	2.64	5	10101100101	0.71
Sweden	1.54	8	10110000000	1.85	5	10001001101	0.59
Switzerland	2.39	6	10110111011	2.73	9	10011010110	1.05
UK	7.94	Expanding	10011010010	13.80	9	11111011111	1.63
US	8.90	9	10011100101	18.09	8	11100111000	2.29

Table 4
Endogenizing Predictive Variables and In-Sample Window Lengths Via a Persistence Strategy

This table reports the results of strategies that endogenize predictive variables and in-sample window lengths via a persistence strategy. For each data set we use all out-of-sample return combinations from the exogenous simulations from Table 2. We use a five-year ranking period to group the out-of-sample portfolios into deciles based on average mean return in the ranking period. Within each decile, we equally weight each portfolio and then track the performance of each decile portfolio for the next year. At the end of the year, we re-rank, reform the portfolios, and track their performance for another year. We repeat this process until the end of the out-of-sample period for each data set. The results from this strategy are reported below as the "Persistence Results." We also report the ex post spreads from the simulations of Table 2. Panel A reports persistence results for dataset 1 over the period 1973(9)-1997(6), Panel B reports results for dataset 2 over the period 1964(1)-1992(12), and Panel C reports results for dataset 3 over the period 1980(1)-1994(12). For both the simulations and the persistence strategy, we report the terminal wealth (TW), Sharpe Ratio (SR), Jensen's alpha, and the Fama French (1993) three-factor alpha (FF alpha). In Panel C, for data set 3, we do not report the ex post mean or terminal wealth from the exogenous simulations since the 13 different assets do not cover the same period. Instead, as an ad hoc measures to compare across periods, we report the Sharpe ratio, Jensen's alpha, and Fama French three-factor alphas. In the next-to-last row (last row), we report the p-value of a χ_9^2-statistic (χ_1^2-statistic) to test the null hypothesis of equality of performance measures across the 10 sort decile portfolios (between decile 1 and 10) for the mean return, Jensen's alpha, and FF alpha. The χ^2-statistics are robust to heteroskedasticity and autocorrelation (Gallant, 1987).

Panel A: Data Set 1

	Ex Post Results From Simulations					Persistence Results				
Decile	Quarterly Mean (%)	TW ($)	SR	Jensen's Alpha (%)	FF Alpha (%)	Quarterly Mean (%)	TW ($)	SR	Jensen's Alpha (%)	FF Alpha (%)
Losers 1	2.62	10.56	0.14	-0.41	-0.44	2.84	12.64	0.19	-0.08	-0.11
2	2.89	13.80	0.18	-0.06	-0.07	2.89	13.32	0.20	0.03	0.04
3	3.00	15.52	0.21	0.11	0.12	3.10	16.50	0.25	0.37	0.37
4	3.09	17.04	0.22	0.23	0.23	3.02	15.12	0.22	0.20	0.18
5	3.19	18.90	0.24	0.36	0.34	3.20	18.03	0.26	0.40	0.37
6	3.27	20.30	0.25	0.43	0.42	3.23	18.63	0.27	0.53	0.57
7	3.34	21.84	0.27	0.52	0.52	3.25	18.81	0.27	0.49	0.50
8	3.43	23.72	0.28	0.59	0.56	3.34	19.99	0.27	0.49	0.38
9	3.55	26.71	0.31	0.74	0.73	3.31	19.55	0.27	0.45	0.40
Winners 10	3.74	32.19	0.34	0.92	0.89	3.22	18.32	0.27	0.49	0.56
10-1	1.12	21.63	0.20	1.33	1.33	0.38	5.68	0.08	0.57	0.67
	P-values:					P-values:				
χ_9^2 Spread Across Deciles	0.00			0.00	0.00	0.27			0.50	0.52
χ_1^2 Decile 1 vs. Decile 10	0.00			0.00	0.00	0.15			0.36	0.25

*, **, *** The null hypothesis is rejected at the 10%, 5%, and 1% levels, respectively.

Table 4, Continued

Panel B: Data Set 2

	Ex Post Results From Simulations					Persistence Results				
Decile	Monthly Mean (%)	TW ($)	SR	Jensen's Alpha (%)	FF Alpha (%)	Monthly Mean (%)	TW ($)	SR	Jensen's Alpha (%)	FF Alpha (%)
Losers 1	0.76	11.94	0.08	0.05	0.05	0.91	21.16	0.16	0.21***	0.21***
2	0.84	15.71	0.10	0.13	0.13	0.97	26.24	0.19	0.28***	0.27***
3	0.89	18.26	0.12	0.18	0.17	0.98	26.66	0.19	0.29***	0.29***
4	0.92	20.34	0.13	0.21	0.20	0.97	25.86	0.19	0.28***	0.28***
5	0.95	22.50	0.14	0.24	0.23	0.96	24.89	0.18	0.27***	0.26***
6	0.98	24.81	0.15	0.27	0.26	0.95	24.48	0.18	0.26***	0.26***
7	1.00	27.29	0.16	0.30	0.29	0.97	25.85	0.18	0.27***	0.28***
8	1.03	30.36	0.17	0.33	0.33	1.02	30.24	0.20	0.32***	0.33***
9	1.07	34.66	0.18	0.37	0.37	1.02	30.19	0.20	0.33***	0.33***
Winners 10	1.15	45.86	0.20	0.45	0.45	1.04	32.09	0.20	0.33***	0.32***
10-1	0.39	33.92	0.13	0.39	0.40	0.13	10.93	0.04	0.12	0.11
P-values:						P-values:				
χ^2_9 Spread Across Deciles	0.00			0.00	0.00	0.21			0.15	0.08
χ^2_1 Decile 1 vs. Decile 10	0.00			0.00	0.00	0.06			0.04	0.05

*, **, *** The null hypothesis is rejected at the 10%, 5%, and 1% levels, respectively.

Table 4, Continued

Panel C: Data Set 3

	Ex Post Results From Simulations			Persistence Results				
Ranking period decile:	SR	Jensen's Alpha (%)	FF Alpha (%)	Monthly Mean (%)	TW ($)	SR	Jensen's Alpha (%)	FF Alpha (%)
Losers 1	-0.06	-0.68	-0.74	1.06	5.91	0.12	0.16	0.04
2	0.01	-0.34	-0.41	1.13	6.87	0.16	0.22	0.11
3	0.04	-0.18	-0.24	1.09	6.35	0.15	0.16	0.06
4	0.06	-0.06	-0.10	1.09	6.44	0.15	0.16	0.06
5	0.09	0.05	0.00	1.10	6.53	0.15	0.17	0.09
6	0.11	0.14	0.09	1.12	6.69	0.18	0.19	0.12
7	0.12	0.24	0.19	1.08	6.21	0.14	0.17	0.08
8	0.14	0.35	0.28	1.02	5.55	0.11	0.13	0.02
9	0.16	0.48	0.41	0.96	4.82	0.09	0.05	-0.05
Winners 10	0.21	0.69	0.61	0.92	4.38	0.07	0.04	-0.08
10-1	0.27	1.37	1.35	-0.14	-1.53	-0.05	-0.12	-0.12
	P-values:			P-values:				
χ^2_9 Spread Across Deciles		0.00	0.00	0.81			0.88	0.49
χ^2_1 Decile 1 vs. Decile 10		0.00	0.00	0.61			0.56	0.23

*, **, *** The null hypothesis is rejected at the 10%, 5%, and 1% levels, respectively.

Table 5
Variations of Ranking Period (1, 3, and 5 years) and Objective Function (Mean, Terminal wealth, and Sharpe ratio) for the Persistence Strategies

This table reports variations on ranking period and objective function for the persistence strategies that endogenize predictive variables and in-sample window length. For each data set we vary the ranking period over 1, 3, and 5 years, and vary the objective function over mean, terminal wealth, and Sharpe ratio. We report the spread in means between winners (decile 10) minus losers (decile 1) for the nine combinations of ranking period and objective function.

		Persistence Results		
		Dataset 1 Decile 10-1 Quarterly Mean (%)	Dataset 2 Decile 10-1 Monthly Mean (%)	Dataset 3 Decile 10-1 Monthly Mean (%)
Ranking period	Objective function			
1 year	Mean	0.58*	0.05	0.11
1 year	Terminal Wealth	0.48	0.02	0.33
1 year	Sharpe Ratio	0.58*	0.04	0.06
3 years	Mean	0.43	0.07	-0.08
3 years	Terminal Wealth	0.28	0.05	0.08
3 years	Sharpe Ratio	0.39	0.05	0.00
5 years	Mean	0.39	0.13*	-0.14
5 years	Terminal Wealth	0.14	0.08	-0.09
5 years	Sharpe Ratio	0.35	0.14**	-0.14
	Averages:	0.38	0.07	0.03

*, **, *** The null hypothesis of equality of average returns between decile 10 and decile 1 is rejected at the 10%, 5%, and 1% levels, respectively.

Table 6
Endogenizing Predictive Variables and In-Sample Window Lengths Via a Recursive Strategy

This table reports the results of strategies that endogenize predictive variables and in-sample window lengths via a recursive strategy. For each data set, we estimate, using OLS, a linear model of expected returns within an in-sample period for each predictive variable combination and window length. We use the loadings on the predictive variables and the estimated intercept to form expected return estimates during the in-sample period. We obtain a series of realized returns from the following trading strategy: go long asset *A* if the expected excess return estimate for that period is great than zero, else invest in a t-bill. We then choose the best combination from the *W* x *A* x (2^K-1) total models (where *W* = in-sample window length, *A* = assets, and *K* = the number of predictive variables) based on the average in-sample terminal wealth, standardized by the number of periods in *W*. We examine 508 combination for data set 1 (1 asset* 4 windows * (2^7-1) models), 3577 combinations for data set 2 (1 asset * 7 windows * (2^9-1) models), and 93,093 combinations (13 assets * 7 windows * (2^{10}-1) models) for data set 3. Using the optimal in-sample combination, we form a step ahead out-of-sample forecast using the in-sample intercept and predictive variable loadings. We then roll forward the in-sample end date by one period, find again the best in-sample combination, and obtain a forecast for the next out-of-sample period. We repeat this process until the end of the out-of-sample period. For the out-of-sample forecast series, we obtain a series of realized returns for the "active" portfolio from the following trading strategy: go long in the optimal asset *A* if the expected excess return estimate for that period is great than zero, else invest in a t-bill.

Panel A reports results for dataset 1, Panel B reports results for dataset 2, and Panel C reports results for dataset 3.We report the out-of-sample mean and standard deviation of returns for the active portfolio and the buy-and-hold benchmark strategy. For data sets 1 and 2, the buy-and-hold is a constant position in the S&P500. For data set 3, the buy-and-hold is an equally weighted portfolio of the 13 country assets. Terminal wealth is the total wealth at the end of the out-of-sample period of investing one dollar at the beginning. We also report a Sharpe ratio, Jensen's alpha, and Fama French (1993) three-factor alpha (FF alpha). For data sets 1 and 2, we also report the Henriksson and Merton (1981) market timing statistics (HM p-value and HM p_1+p_2) and the forecast beta (β_f), which provides a measure of overall out-of-sample fit and is calculated by regressing the monthly realized return on the forecasted return: $r_\tau = \alpha + \beta_f r_{forecast,\tau} + \varepsilon_\tau$. "Active trades" reports the number of periods that a specific strategy invests in the risky asset. N shows the number of out-of-sample periods. We report active portfolio results for one-way transaction costs scenarios; 0.0, 0.1, 0.3 and 0.5%.

Panel A: Quarterly Out-of-Sample Performance Results for Data Set 1, 1973(9) – 1998(3)

	Mean Return (%)	Standard Deviation	Terminal Wealth ($)	Sharpe Ratio	β_f	Jensen's Alpha (%)	FF Alpha (%)	HM p-value	HM p_1+p_2	Active Trades	N
Buy-Hold	3.36	8.08	18.99	0.20		-0.27	-0.25*				99
Active (tc=0%)	3.05	5.46	17.08	0.24	0.24	0.44	0.33	0.16	0.97	56	99
Active (tc=0.1%)	3.03	5.62	16.53	0.23	0.18	0.37	0.31	0.16	0.97	56	99
Active (tc=0.3%)	2.49	5.47	9.95	0.14	-0.06	-0.27	-0.03	0.03	0.80	57	99
Active (tc=0.5%)	3.06	5.84	16.94	0.23	0.23	0.31	0.19	0.17	0.98	54	99

Table 6, continued

Panel B: Monthly Out-of-Sample Performance Results for Data Set 2, 1964(1) – 1992(12)

	Mean Return (%)	Standard Deviation	Terminal Wealth ($)	Sharpe Ratio	β_f	Jensen's Alpha (%)	FF Alpha (%)	HM p-value	HM p_1+p_2	Active Trades	N
Buy-Hold	0.92	4.36	17.55	0.09		0.03	0.06				348
Active (tc=0%)	1.05	3.05	32.30	0.17	0.23**	0.36***	0.32**	0.001	1.16	196	348
Active (tc=0.1%)	0.87	3.24	17.06	0.11	0.23**	0.16	0.14	0.003	1.14	198	348
Active (tc=0.3%)	0.83	3.07	14.85	0.10	0.23**	0.13	0.12	0.036	1.07	181	348
Active (tc=0.5%)	0.71	3.06	9.89	0.06	0.19**	0.01	-0.04	0.034	1.07	159	348

Panel C: Monthly Out-of-Sample Performance Results for Data Set 3, 1980(1) – 1995(5)

	Mean Return (%)	Standard Deviation	Terminal Wealth ($)	Sharpe Ratio	Jensen's Alpha (%)	FF Alpha (%)	Average Number of Assets	Active Trades	N
Buy-Hold	1.28	4.58	8.70	0.15	0.20	0.08			185
Active (tc=0%)	1.40	4.88	10.53	0.16	0.32	0.20	4.69	182	185
Active (tc=0.1%)	1.47	5.02	11.78	0.17	0.38	0.21	4.50	184	185
Active (tc=0.3%)	1.32	5.06	8.85	0.14	0.22	0.07	4.41	183	185
Active (tc=0.5%)	1.40	5.15	10.11	0.15	0.29	0.18	4.15	179	185

*Denotes significance at 10% level, ** denotes significance at 5% level, *** denotes significance at 1% level.

Table 7
Out-of-Sample Simulation Results with Exogenously and Endogenously Specified Parameters

This table presents the number of cases for which the null hypothesis of no predictability is rejected using various performance measures for monthly/quarterly out-of-sample forecasts based on combinations of exogenous and endogenously specified parameters. In panel A we endogenize variable selection via 6 statistical model selection criteria and loop over exogenous values of estimation windows and assets, in panel B we endogenize window length and loop over exogenous values of model selection criteria and assets, in panel C we endogenize model selection criteria and loop over exogenous values of estimation windows and assets, and in panel D, we endogenize both statistical model selection and windows, and examine variations in assets. The performance measures are the forecast beta (β_f), Jensen's alpha, Fama French (1993) three-factor alpha (FF alpha), and the market timing statistics of Henriksson and Merton (1981) (HM p). Rejection rates are reported at a 5% or better significance level. The rejection rates reported for the Jensen's Alpha and FF alpha are based on two conditions; the alpha of the out-of-sample portfolio must be greater than the alpha of the buy-and-hold portfolio, and, the alpha of the out-of-sample portfolio must be significant at a 5% or better greater level. The coefficient estimate of the slope (β_f) provides a measure of overall out-of-sample fit and is calculated by regressing the monthly realized return on the forecasted return: $r_\tau = \alpha + \beta_f r_{forecast,\tau} + \varepsilon_\tau$

Panel A: Endogenized Variable Selection

				Percent of out-of-sample forecasts rejecting the null			
	Endogenized Parameter	Specifications	Total Number of Specifications	Forecast Beta	Jensen's Alpha	FF Alpha	Market Timing
				$\beta_F>0$ ($p_{\beta f} \leq 0.05$)	$\alpha_j > \alpha_{j,bh}$ ($p_\alpha \leq 0.05$)	$\alpha_{ff} > \alpha_{ff,bh}$ ($p_{\alpha ff} \leq 0.05$)	$HM_{p1+p2}>1$ ($HM_p \leq 0.05$)
Data set 1	Variable Selection	6 selection criteria X 4 windows	24	20.8%	12.5%	0%	4.2%
Data set 2	Variable Selection	6 selection criteria X 7 windows	42	73.8%	42.8%	16.7%	45.2%
Data set 3	Variable Selection	6 selection criteria X 7 windows X 13 assets	546	0.91%	4.0%	0.55%	7.3%

Panel B: Endogenized Estimation Windows

				Percent of out-of-sample forecasts rejecting the null			
	Endogenized Parameter	Specifications	Total Number of Specifications	Forecast Beta	Jensen's Alpha	FF Alpha	Market Timing
				$\beta_F>0$ ($p_{\beta f} \leq 0.05$)	$\alpha_j > \alpha_{j,bh}$ ($p_\alpha \leq 0.05$)	$\alpha_{ff} > \alpha_{ff,bh}$ ($p_{\alpha ff} \leq 0.05$)	$HM_{p1+p2}>1$ ($HM_p \leq 0.05$)
Data set 1	In-sample Window Length	6 selection criteria	6	0%	33.3%	0%	0%
Data set 2	In-sample Window Length	6 selection criteria	6	33.3%	16.7%	0%	50%
Data set 3	In-sample Window Length	6 selection criteria X 13 assets	78	0%	5.1%	0%	7.7%

Table 7, continued

Panel C: Endogenized Statistical Model Selection Criteria

				Percent of out-of-sample forecasts rejecting the null			
	Endogenized Parameter	Specifications	Total Number of Specifications	Forecast Beta	Jensen's Alpha	FF Alpha	Market Timing
				$\beta_F>0$ ($p_{\beta f} \leq 0.05$)	$\alpha j > \alpha j,bh$ ($p_\alpha \leq 0.05$)	$\alpha_{ff} > \alpha_{ff,bh}$ ($p_{\alpha ff} \leq 0.05$)	$HM_{p1+p2}>1$ ($HM_p \leq 0.05$)
Data set 1	Variable Selection Criteria	4 Windows	4	0%	0%	0%	0%
Data set 2	Variable Selection Criteria	7 Windows	7	42.8%	14.3%	14.3%	85.7%
Data set 3	Variable Selection Criteria	7 Windows X 13 assets	91	1.1%	4.4%	0%	6.6%

Panel D: Endogenized Window and Statistical Model Selection Criteria

				Percent of out-of-sample forecasts rejecting the null			
	Endogenized Parameter	Specifications	Total Number of Specifications	Forecast Beta	Jensen's Alpha (%)	FF Alpha (%)	Market Timing
				$\beta_F>0$ ($p_{\beta f} \leq 0.05$)	$\alpha_j > \alpha_{j,bh}$ ($p_\alpha \leq 0.05$)	$\alpha_{ff} > \alpha_{ff,bh}$ ($p_{\alpha ff} \leq 0.05$)	$HM_{p1+p2}>1$ ($HM_p \leq 0.05$)
Data set 1	Window and model selection	1	1	-0.01	0.10	-0.09	0.83
Data set 2	Window and model selection	1	1	0.14	0.18	0.17	1.11***
Data set 3	Window and model selection	13 assets	13	0%	7.7%	7.7%	7.7%