Should We Expect Significant Out-of-Sample Results when Predicting Stock Returns?

K.7

Erik Hjalmarsson

International Finance Discussion Papers
Board of Governors of the Federal Reserve System

Number 855
February 2006

Board of Governors of the Federal Reserve System

International Finance Discussion Papers

Number 855

February 2006

Should We Expect Significant Out-of-Sample Results when Predicting Stock Returns?

Erik Hjalmarsson

NOTE: International Finance Discussion Papers are preliminary materials circulated to stimulate discussion and critical comment. References in publications to International Finance Discussion Papers (other than an acknowledgment that the writer has had access to unpublished material) should be cleared with the author or authors. Recent IFDPs are available on the Web at www.federalreserve.gov/pubs/ifdp/.

Should We Expect Significant Out-of-Sample Results when Predicting Stock Returns?

Erik Hjalmarsson*

Division of International Finance

Federal Reserve Board, Mail Stop 20, Washington, DC 20551, USA

February 2006

Abstract

Using Monte Carlo simulations, I show that typical out-of-sample forecast exercises for stock returns are unlikely to produce any evidence of predictability, even when there is in fact predictability and the correct model is estimated.

JEL classification: C15; C53; G14.

Keywords: Stock return predictability; Out-of-sample tests.

---

[*] Tel.: +1-202-452-2436; fax: +1-202-263-4850; email: erik.hjalmarsson@frb.gov. The views presented in this paper are solely those of the author and do not represent those of the Federal Reserve Board or its staff.

1 Introduction

Forecasting models play a central role in economics and finance, both for practitioners and academics. Although standard, in-sample, econometric methods are usually relied upon to judge the validity of a given model, out-of-sample tests are often deemed to provide the most robust assessment of any econometric forecasting model. Goyal and Welch (2003,2004) provide perhaps the most noted expression of this belief in recent times; they argue that virtually all variables that have been proposed as predictors of future stock returns provide no predictive gains in out-of-sample exercises.

Should one interpret these findings by Goyal and Welch as conclusive evidence that stock returns are not predictable? There are several reasons to adopt a more nuanced view. Inoue and Kilian (2004) show that out-of-sample tests typically have lower power than in-sample tests. They also argue that the widely held belief that out-of-sample tests are less susceptible to data-mining is not generally true. Campbell and Thompson (2005) take a different approach and show that by imposing some common sense restrictions when forming the out-of-sample forecasts, there is in fact fairly strong evidence of out-of-sample predictive ability in stock returns.

This note adds to the above strand of literature by reporting the results from simulated out-of-sample exercises. I show that the results of Goyal and Welch (2003,2004) do not imply that previous in-sample results are spurious. In fact, using Monte Carlo simulations, similar out-of-sample results to those of Goyal and Welch are found even when the postulated forecasting model is in fact the true data generating process. This result stems from the fact that any predictive component in stock returns must be small, if it does exist. Therefore, if the predictive relationship is estimated poorly, the conditional forecasting model may be outperformed by the unconditional benchmark model which assumes that expected returns are constant over time. Put in other words, in order to produce good forecasts when the slope coefficient in a linear regression is small, you are often better off setting it equal to zero, rather than using a noisy estimate of it. To accurately estimate a very small coefficient, large amounts of data are needed. The results in this paper show that when testing for stock return predictability, the sample sizes in most relevant cases are simply too small relative to the size of the slope coefficient for any predictive ability to show up in out-of-sample exercises.

The findings in this paper, of course, do not tell us that stock returns *are* predictable, but merely that we should not disregard the econometric in-sample results in favour of out-of-sample results. As shown here, when the level of predictability is low, it is quite possible to correctly specify and estimate the true predictive relationship without obtaining any improvement in out-of-sample forecasts. However, this raises the question of the practical use of identifying a predictive relationship if, in fact, that relationship cannot be used to improve upon forecasts. It is evident that much care must be used in forming the forecast if it is to perform better than a simple unconditional alternative. For instance, Campbell and Thompson (2005), manage to improve the out-of-sample performance substantially by imposing some simple economically motivated restrictions on the forecasts. On the econometric side, it is of course also important to form the best possible estimate of the relevant parameters in the model. At present, most econometric studies on stock return predictability have focused primarily on the issues of testing, rather than point estimation. Hopefully, there will be also be some advances in the estimation area over the next few years.

2 Simulation design

To create simulated samples of stock-returns and predictor variables, I rely on the standard data generating process (dgp) most often found in the stock return predictability literature (e.g. Campbell and Yogo, 2005). Let r_t denote the excess stock return in period $t = 1, ..., T$, and let x_t denote the corresponding value of some scalar predictor variable, such as the dividend- or earnings-price ratio. The variables r_t and x_t are generated according to

$$\begin{aligned} r_t &= \alpha + \beta x_{t-1} + u_t, & (1) \\ x_t &= \rho x_{t-1} + v_t, & (2) \end{aligned}$$

where the joint innovations, $w_t = (u_t, v_t)'$, are *iid* normal with unit variance and correlation δ. The autoregressive root ρ is modelled as local-to-unity and takes values $\rho = 1 + c/T$, where c is the so-called local-to-unity parameter.

To capture the often large negative correlation between u_t and v_t, the parameter δ is set equal to -0.9. The local-to-unity-parameter is set equal to either $c = -20$ or $c = -2$ and the sample size T is set equal to either 600 or 1200 to represent sample sizes of 50 or 100 years of monthly data.[1] The intercept α is set equal to 0.005 and I let the slope-coefficient vary between zero and 0.05. Campbell and Yogo (2005), who analyze predictability in aggregate U.S. stock returns, present their empirical results in a format standardized to conform with a unit variance in the innovations u_t and v_t and show that in most cases the OLS estimate of β is between 0.01 and 0.02 for monthly data.

The aim of the Monte Carlo study is to compare the out-of-sample forecasts of r_t based on an estimate of equation (1) to those based on a model of constant expected returns (i.e. $\beta = 0$). These forecasts will be referred to as the conditional and unconditional forecasts, respectively.

The simulated out-of-sample exercises are performed in the following manner. The first half of the sample is used to form the initial estimates of the conditional and unconditional models. The estimate of the unconditional model is, of course, merely the mean of the returns observed up to that point in time. For each time-period in the remaining half of the sample, the one-step-ahead forecasts based on the conditional and unconditional models are calculated and the estimates of the forecast models are updated using the additional data that become available to the forecaster every period. The slope coefficients are estimated using standard OLS. The mean squared errors (MSE) from every forecast are calculated, as well as the corresponding Diebold and Mariano (1995) (DM) statistic, which tests the null hypothesis of no additional predictive accuracy in the conditional forecast compared to the unconditional one. In order to assess the impact of poorly estimated βs, I also form conditional forecasts using the true value of β. The value of α is still estimated, however. All simulation results are based on 10,000 repetitions.

---

[1] Simulations based on annual parameter values, using sample sizes of 50 and 100 years were also performed, but not reported here. These simulations delivered qualitatively identical results to the monthly ones presented here.

3 Results

The results of the paper are presented in Figures 1 and 2, which show the Monte Carlo evidence from samples with sizes $T = 600$ and $T = 1200$, respectively. The top two panels in both figures give the results corresponding to $c = -20$, and the two lower panels correspond to $c = -2$.

Panel (a) in Figure 1 shows the ratios between the MSEs for the unconditional and conditional forecasts, when $c = -20$. When, on average, the conditional forecast outperforms the unconditional one, this ratio is greater than one, and vice versa. As is evident from the plot, when the conditional forecast is based on the OLS estimate of β, the true value of β needs to be greater than 0.015 for the conditional forecast to outperform the unconditional one on average. The conditional forecast based on the true value of β does, of course, always outperform the unconditional one. The same results for $c = -2$ are shown in Panel (c). The conditional forecast, based on the OLS estimates, now perform better relative to the unconditional one; a true value of β greater than 0.06 is sufficient for the conditional forecast to beat the unconditional one on average. Panels (b) and (d) show the rejection rates for a one-sided 5% DM test of higher accuracy in the conditional forecast versus the unconditional forecast, based on the MSE.[2] Clearly, when using OLS estimates of β to form the conditional forecasts, the DM test lacks power to reject the null hypothesis of equal forecasting ability in the relevant regions of the parameter space. For $\beta = 0.015$, the rejection rates are 5.8% and 20.5% for $c = -20$ and $c = -2$, respectively. The chance of detecting predictive ability through the DM test is thus extremely limited when β is small. Indeed, even when the true value of β is used in forming the conditional forecasts the rejection rates remain low.

In Figure 2, the Monte Carlo results from samples with sizes $T = 1200$, representing 100 years of monthly data, are reported. As expected, the conditional forecasts based on OLS estimates of β perform substantially better than in the $T = 600$ case. A value of β greater than 0.07, for $c = -20$, and 0.04 for $c = -2$, is sufficient for the conditional forecast to outperform the unconditional one, on average. For $c = -20$, the DM test statistic is still not very powerful, although for $c = -2$ there is now a fair chance of rejecting the null for reasonable parameter values; for $\beta = 0.015$ the rejection rates are 15.3% and 69.3%, for $c = -20$ and $c = -2$, respectively.

---

[2] The DM statistic is calculated using the long-run variance estimator of Andrews and Monahan (1992) with a quadratic spectral kernel.

To sum up, in a monthly sample spanning 50 years it is often difficult to detect *any* predictive ability, when β is small, even under such perfect circumstances as in a controlled Monte Carlo experiment where the true functional form of the model is known and completely stable over time. In reality, the model is likely to be only, at best, a decent approximation of the true data generating process, which is also not likely to remain unchanged over a 50 year time span. This point, of course, is even more valid for the 100 year sample, where a stable model for the entire time span seems even less probable. Practical limitations on data availability also often restrict researchers to time spans of around 50 years or shorter. For instance, use of the short interest rate as a predictor is typically only considered for data after 1952 when the Fed unpegged the interest rate, and accounting variables, such as the book-to-market ratio, are often only available under an even shorter period.

The overall interpretation of these results must be that, in practice, it should be difficult to detect any out-of-sample predictability in stock returns, even without any demands for statistical significance but merely as evaluated by the MSE.

References

Andrews, D.W.K., and C.J. Monahan, 1992. An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator, Econometrica 60, 953–966.

Campbell., J.Y., and S.B. Thompson, 2005. Predicting the Equity Premium Out of Sample: Can Anything Beat the Historical Average?, Working Paper, Harvard University.

Campbell, J.Y., and M. Yogo, 2005. Efficient Tests of Stock Return Predictability, forthcoming Journal of Financial Economics.

Diebold, F.X., and R.S., Mariano, 1995. Comparing predictive accuracy, Journal of Business and Economics Statistics 13, 253-263.

Goyal, A., and I. Welch, 2003. Predicting the equity premium with dividend ratios, Management Science 49, 639-654.

Goyal, A., and I. Welch, 2004. A comprehensive look at the empirical performance of equity premium prediction, NBER Working Paper 10483.

Inoue, A., and L. Kilian, 2004. In-sample or out-of-sample tests of predictability: which one should we use?, forthcoming Econometric Reviews.

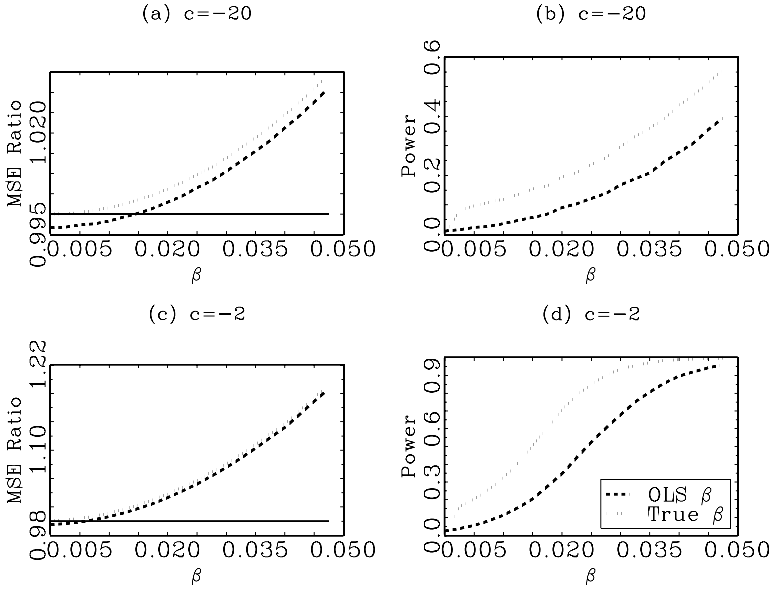

Figure 1: Results from Monte Carlo simulations with a simulated monthly sample of 600 observations. The top two panels show the results for $c = -20$ and the bottom panels for $c = -2$. The left panels, (a) and (c), display the ratios between the mean squared errors (MSE) for the unconditional and conditional forecasts; a ratio greater than one implies that the conditional forecast outperforms the unconditional one. The right hand panels show the 5% rejection rates for the Diebold and Mariano (DM) test of higher accuracy in the conditional forecast versus the unconditional forecast, based on the MSE. The dashed lines show the results for conditional forecasts based on the OLS estimate of β and the dotted lines the results for conditional forecasts based on the true value of β. The flat lines in the left hand graphs indicate a value of one.

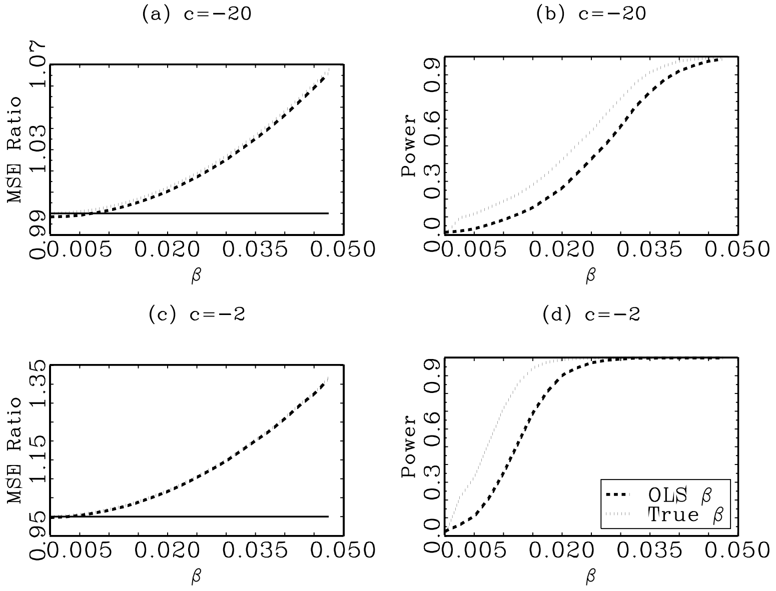

Figure 2: Results from Monte Carlo simulations with a simulated monthly sample of 1200 observations. The top two panels show the results for $c = -20$ and the bottom panels for $c = -2$. The left panels, (a) and (c), display the ratios between the mean squared errors (MSE) for the unconditional and conditional forecasts; a ratio greater than one implies that the conditional forecast outperforms the unconditional one. The right hand panels show the 5% rejection rates for the Diebold and Mariano (DM) test of higher accuracy in the conditional forecast versus the unconditional forecast, based on the MSE. The dashed lines show the results for conditional forecasts based on the OLS estimate of β and the dotted lines the results for conditional forecasts based on the true value of β. The flat lines in the left hand graphs indicate a value of one.

International Finance Discussion Papers

IFDP Number	Titles	Author(s)
	2006	
851	Exchange-Rate Pass-Through in the G-7 Countries	Jane E. Ihrig Mario Marazzi Alexander D. Rothenberg
850	The Adjustment of Global External Imbalances: Does Partial Exchange Rate Pass-Through to Trade Prices Matter?	Christopher Gust Nathan Sheets
	2005	
849	Interest Rate Rules, Endogenous Cycles and Chaotic Dynamics in Open Economies	Marco Airaudo Luis-Felipe Zanna
848	Fighting Against Currency Depreciation Macroeconomic Instability and Sudden Stops	Luis-Felipe Zanna
847	The Baby Boom Predictability in House Prices and Interest Rates	Robert F. Martin
846	Explaining the Global Pattern of Current Account Imbalances	Joseph W. Gruber Steven B. Kamin
845	DSGE Models of High Exchange-Rate Volatility and Low Pass-Through	Giancarlo Corsetti Luca Dedola Sylvain Leduc
844	The Response of Global Equity Indexes to U.S. Monetary Policy Announcements	Jon Wongswan
843	Accounting Standards and Information: Inferences from Cross-Listed Financial Firms	John Ammer Nathanael Clinton Gregory P. Nini
842	Alternative Procedures for Estimating Vector Autoregressions Identified with Long-Run Restrictions	Lawrence J Christiano Martin Eichenbaum Robert J. Vigfusson

---

Please address requests for copies to International Finance Discussion Papers, Publications, Stop 127, Board of Governors of the Federal Reserve System, Washington, DC 20551.
Email: publications-bog@frb.gov. Fax (202) 728-5886.

International Finance Discussion Papers

IFDP Number	Titles	Author(s)
841	Monetary Policy and House Prices: A Cross-Country Study	Alan G. Ahearne John Ammer Brian M. Doyle Linda S. Kole Robert F. Martin
840	International Capital Flows and U.S. Interest Rates	Francis E. Warnock Veronica C. Warnock
839	Effects of Financial Autarky and Integration: The Case of the South Africa Embargo	Brahima Coulibaly
838	General-to-specific Modeling: An Overview and Selected Bibliography	Julia Campos Neil R. Ericsson David F. Hendry
837	Currency Crashes and Bond Yields in Industrial Countries	Joseph E. Gagnon
836	Estimating Elasticities for U.S. Trade in Services	Jaime Marquez
835	SIGMA: A New Open Economy Model for Policy Analysis	Christopher Erceg Luca Guerrieri Christopher Gust
834	Optimal Fiscal and Monetary Policy with Sticky Wages and Sticky Prices	Sanjay K. Chugh
833	Exchange Rate Pass-through to U.S. Import Prices: Some New Evidence	Mario Marazzi Nathan Sheets Robert J. Vigufsson And Others
832	A Flexible Finite-Horizon Identification of Technology Shocks	Neville Francis Michael T. Owyang Jennifer E. Roush
831	Adjusting Chinese Bilateral Trade Data: How Big is China's Surplus	John W. Schindler Dustin H. Beckett
830	Order Flow and Exchange Rate Dynamics in Electronic Brokerage System Data	David W. Berger Alain P. Chaboud Sergey V. Chernenko Edward Howorka Raj S. Krishnasami Iyer David Liu Jonathan H. Wright