# Time-Varying Risk Premiums and the Output Gap

**Ilan Cooper and Richard Priestley**
University of Haifa and Norwegian School of Management

The output gap, a production-based macroeconomic variable, is a strong predictor of U.S. stock returns. It is a prime business cycle indicator that does not include the level of market prices, thus removing any suspicion that returns are forecastable due to a "fad" in prices being washed away. The output gap forecasts returns both in-sample and out-of-sample, and it is robust to a host of checks. We show that the output gap also has predictive power for excess stock returns in other G7 countries and U.S. excess bond returns. (*JEL* E44, G12, G14)



Research on stock-return predictability is controversial for a number of reasons. First, due to statistical biases, unstable in-sample results, and weak out-of-sample performance, it is not clear that stock returns are predictable.[1] Second, there is little agreement as to why stock returns are predictable. For example, the predictability of stock returns by financial variables has been rationalized by the claim that they correlate with the business cycle (see, for example, Fama and French 1989). However, when using financial variables, it is possible that predictability is the result of a "fad" in prices being washed away. Furthermore, a host of so-called investor sentiment variables have been shown to have predictive power for stock returns (see, among others, Baker and Wurgler 2006). Third, traditional business cycle variables, such as GDP growth, that are the natural candidates for capturing variations in business conditions, have proven dismal in predicting returns (see Pena, Restoy, and Rodriguez 2002).

This paper demonstrates that a prime business cycle indicator, namely the output gap, predicts stock and bond market returns both in-sample and out-of-sample. Thus, we provide a direct line linking return predictability to economic

We would like to thank an anonymous referee, Mathew Spiegel (the Editor), Ravi Bansal, Kobi Boudoukh, Ian Garrett, Vito Gala, Francisco Gomes, Urban Jermann, Eric Jondeau, Boyan Jovanovic, Leonid Kogan, Jonathan Lewellen, Sydney Ludvigson, Thomas Philippon, Kjetil Storesletten, Raman Uppal, Annette Vissing-Jørgensen, and seminar participants at the AFA Meeting 2006, Econometric Society World Congress 2005, EFA Meeting 2007, University of Bath, University of Haifa, Hebrew University, HEC Lausanne, and University of Oslo for helpful comments and suggestions. We also thank John Cochrane, Michael Roberts, and Jeffrey Wurgler for making their data available to us. We express our gratitude to Todd Clark for generously providing his code that calculates the out-of-sample forecasting tests. Send correspondence to Richard Priestley, Norwegian School of Management, Nydalsveien 37, N-0444 Oslo, Norway; telephone: +47 46 410 515; fax: +47 23 264 723. E-mail: richard.priestley@bi.no.

[1] See, for example, Bossaerts and Hillion (1999); Stambaugh (1999); Goyal and Welch (2003); Valkanov (2003); Lewellen (2004); Boudoukh, Richardson, and Whitelaw (2005); Torous, Valkanov, and Yan (2005); Campbell and Yogo (2006); and Goyal and Welch (2006).


doi:10.1093/rfs/hhn087 Advance Access publication October 2, 2008

fundamentals. The predictability of stock and bond returns by a classical business cycle variable, such as the output gap, serves to enhance our understanding of the economics of time-varying risk premiums. The output gap has several a priori advantages over other predictive variables. First, in contrast to financial market variables, the output gap does not contain the level of asset prices. This is significant because, as Cochrane (2005) notes, using nonfinancial market-based variables removes any suspicion that stock-return predictability arises due to a "fad" in prices being washed away. That is, predictability of stock returns through the output gap is unlikely to stem from stock mispricing. Thus, we contribute to the debate over why stock returns are predictable. Second, while almost all other known macroeconomic predictor variables use consumption data, the output gap uses only production-related data and is a classical business cycle variable. Thus, the predictive power of the output gap constitutes independent evidence regarding the variation of the risk premiums over the business cycle.

We find that the output gap, as measured by the deviations of the log of industrial production from a trend that incorporates both a linear and a quadratic component, predicts stock returns at business cycle frequencies. The $\overline{R}^2$ is 2% at the one-month horizon and the point estimate is both economically and statistically significant. The finding that returns are predictable at the one-month horizon is particularly interesting, since Cochrane (1999) notes that "month-to-month returns are still strikingly unpredictable." At a quarterly horizon, the $\overline{R}^2$ rises to 5% and at the one-year horizon to 11%, which are substantial when compared to other predictor variables. At longer horizons, predictability is also evident and remains economically important, since the estimates and the $\overline{R}^2s$ are different from their implied estimates under the null hypothesis of no predictability with persistent regressors and sampling error (see Boudoukh, Richardson, and Whitelaw 2005). A Monte Carlo experiment shows that the size properties of the Newey-West $t$-statistics deteriorate with the horizon; however, the predictability of returns is shown to be statistically significant when using the empirical $t$-statistics.

The in-sample results we report are an important contribution to the literature on understanding the time variation in risk premiums, since Goyal and Welch (2006) provide a comprehensive analysis of in-sample predictability and conclude that "most models are no longer significant in-sample." Our in-sample results relate a direct macroeconomic, business cycle variable to expected returns and show that this relationship is statistically and economically significant across time, countries, and types of financial asset.

We perform an extensive set of out-of-sample tests using strictly data and parameter estimates that were known to investors at the time the forecast was made. The forecasts of returns based on the output gap are more accurate than forecasts based on the historical mean. These findings are based on encompassing tests, tests of the equality of mean-squared errors, out-of-sample $\overline{R}^2$, and investor utility gains from using gap to calculate portfolio weights.

The results are not confined to any particular time period: they are robust to whatever decade we start the out-of-sample forecasts. This is a notable result, since Goyal and Welch (2006) conduct a comprehensive examination of the out-of-sample predictability of many of the variables suggested by the literature as predictors of the equity premium, and conclude that by and large, these variables perform poorly out-of-sample, particularly in the past three decades. For example, perhaps the best known predictor of returns, the dividend yield, cannot predict returns better than the historical mean (see also Bossaerts and Hillion 1999; Goyal and Welch 2003).

We show that stock-return predictability by the output gap (which we intermittently refer to as *gap*) is robust to how it is measured, subsample analysis, and the inclusion of financial market-based predictor variables, including the net payout ratio, which performs particularly well in forecasting stock returns in-sample (see Robertson and Wright 2006; Boudoukh et al. 2007). What is particularly interesting is that the correlation coefficient between *gap* and the net payout ratio is –0.27, which suggests that they are capturing different factors that predict returns. While we are able to show that *gap* is correlated with a number of procyclical macroeconomic variables and NBER recession dates, this is not the case for the net payout ratio. In a regression of monthly excess returns on both *gap* and the net payout ratio, only *gap* is statistically significant. However, in the same regression using annual returns, both are statistically significant and the $\overline{R}^2$ is an impressive 16%. This suggests that the net payout ratio may be capturing long-run factors, while *gap* captures short-run business cycle frequency factors. This is important since it suggests that there may be two sources of time variation in expected returns.

The economic importance of *gap* as a variable that captures business-cycle-related risk premiums is further assessed by regressing the excess stock returns from the remaining G7 countries on their country-specific measures of the output gap. We find evidence that the output gap can predict stock returns at business cycle frequencies and at longer horizons in the majority of the remaining G7 countries. This part of the analysis should also guard against data-snooping biases that might arise from the focus on U.S. in-sample results.

We also find that the output gap can predict next year's excess returns on U.S. government bonds. *Gap* is negatively correlated with Cochrane and Piazzesi's (2005) forward rate factor and is robust to the inclusion of the part of Cochrane and Piazzesi's forward rate factor that is uncorrelated with *gap*, suggesting the possibility that the predictive power of Cochrane and Piazzesi's forward rate factor partly stems from its correlation with *gap*. Furthermore, to the best of our knowledge, this article is the first to show that a specific macroeconomic variable can predict bond returns. This is important since, referring to the predictability of excess bond returns by financial indicators, such as forward spreads and yield spreads, Ludvigson and Ng (2006) write that "An unanswered question is whether such movements in bond market risk premia bear any relation to the macroeconomy," as economic theory predicts.

The rest of the paper is organized as follows. Section 1 describes the data. Section 2 presents results from forecasting stock returns using the output gap. Section 3 addresses biases in forecasting regressions. Section 4 analyzes the predictability of government bond returns. Section 7 concludes.

## 1. Data

We analyze stock returns on the CRSP value-weighted index and the S&P 500 index, which are sampled from the first month of 1948 (1948:1) to the final month of 2005 (2005:12). Excess returns are formed by subtracting the return on the 30-day T-bill from the actual stock return. Due to findings in Rangvid (2006) that actual returns are more predictable than excess returns when using GDP scaled by price, we examine both excess and actual returns. For bond returns, we follow Cochrane and Piazzesi (2005) and use the Fama and Bliss data from CRSP to calculate annual excess bond returns at a monthly frequency over the sample 1952:6 to 2003:12. We obtain the annual return in a given month by borrowing at the one-year rate and buying either a two-, three-, four-, or five-year bond and then selling it after one year.

Output is measured from the total Industrial Production index, which is published by the Federal Reserve and given at a monthly frequency. The Federal Reserve has made a number of ex post revisions to the industrial production data. As the information in the ex post revisions is not available to investors, we have collected the unrevised data as they appear in the Federal Reserve Bulletins at the time of publication. We call these unrevised data the "vintage data" and they would have been available at the time the investor made the forecast, and therefore we use them in estimating *gap*.[2] As industrial production data for month $t$ are published by the Federal Reserve in the middle of month $t+1$, we use the second lag of *gap* when forecasting monthly stock returns. That is, the forecast of monthly returns at time $t$ is based on *gap* in time $t-2$.

We consider four ways to measure the output gap. The first technique, which serves as our main measure of *gap*, is widely used in macroeconomics (see, for example, Clarida, Gali, and Gertler 2000; Fuhrer and Rudebusch 2004) and allows for a slowly changing trend by employing a quadratic trend as well as a linear trend (we intermittently refer to this trend as the quadratic trend),

$$y_t = a + b \cdot t + c \cdot t^2 + v_t, \tag{1}$$

where $y_t$ is the log of industrial production, $t$ is a time trend, and $v_t$ is an error term, which is the output *gap*. The second technique extends a simple linear trend formulation to allow for a breakpoint to account for the well-known

[2] In unreported results, we are able to show that using the current (revised) output data produces very similar results to those that use the vintage data.

slowdown in economic growth in the early 1970s (see, for example, Perron 1989; Taylor 1993),

$$y_t = a + b \cdot t + v_t \quad \text{for} \quad t \leq t_1,$$
$$y_t = a + b \cdot t + c \cdot (t - t_1) + v_t \quad \text{for} \quad t > t_1. \tag{2}$$

While the actual slowdown in economic growth occurred in the early 1970s, we assume that the break is incorporated in the measure in 1977 (see the discussion in Council of Economic Advisers 1977; Clark 1978).[3]

We use a simple linear trend to estimate the third measure of the output gap,

$$y_t = a + b \cdot t + v_t. \tag{3}$$

When examining stock-return predictability in-sample, we estimate the deterministic trend coefficients using the full sample. This is consistent with in-sample predictability tests in, among others, Baker and Wurgler (2006), Lettau and Ludvigson (2001, 2005), Ludvigson and Ng (2006), Lustig and Van Nieuwerburgh (2005), and Julliard (2006). However, when we undertake the out-of-sample tests, we estimate the coefficients on the trend terms recursively, ensuring that the estimate of *gap* at time $t$ is using only data and parameter estimates available to the investor at time $t$. Thus, while for the in-sample tests we have single estimates for the parameters $a$, $b$, and $c$ above, in the out-of-sample tests we obtain new estimates of these parameters for each month after an initial estimation period.

The final measure of the output gap we employ is the deviation of actual GDP from potential GDP. Potential GDP data are published quarterly by the Congressional Budget Office (CBO). The CBO defines potential GDP as the maximum sustainable output which is estimated based on a neoclassical (Cobb-Douglas) production function.[4] We subtract potential GDP from actual GDP to obtain our final measure of *gap*. As we do not have vintage versions of the CBO data, we only examine in-sample predictability with this version of *gap*.

Table 1 investigates the contemporaneous relationship between the different measures of the output gap and measures of the business cycle by regressing variables that are known to correlate with the business cycle on the output gap. We consider the growth rates in aggregate corporate earnings, real GDP, industrial production, inventories, and real aggregate investment as business cycle variables. Most of these macroeconomic aggregates are measured at a quarterly frequency, and therefore these regressions use quarterly data. The findings in Table 1 show that there is a positive relationship between the output

[3] Clark (1978) argues that "The evidence for slower growth in the past decade is now incontrovertible." We also searched for unknown breaks in the trend using the methodology of Perron (1989) and uncovered up to three breaks in the trend. However, incorporating these did not change our empirical results reported below and therefore we report results using the simplest method.

[4] For a detailed description of the methodology for estimating the output gap, see CBO Macroeconomic Analysis paper (2001).

**Table 1**
**Relating the output gap to macroeconomic variables**

| Business cycle variable | *gap_q* | *gap_b* | *gap_l* | *gap_cbo* |
|---|---|---|---|---|
| Growth rate of aggregate corporate earnings | 0.139 | 0.155 | 0.160 | 0.373 |
| | (1.14) | (1.26) | (1.74) | (2.60) |
| Growth rate of inventories | 0.074 | 0.155 | 0.049 | 0.153 |
| | (10.01) | (10.26) | (8.14) | (9.71) |
| Growth rate of real GDP | 0.028 | 0.030 | 0.014 | 0.084 |
| | (2.40) | (2.53) | (1.57) | (3.43) |
| Growth rate of industrial production | 0.112 | 0.117 | 0.068 | 0.145 |
| | 3.65) | (3.77) | (2.87) | (2.36) |
| Growth rate of real investment | 0.086 | 0.092 | 0.055 | 0.223 |
| | (1.35) | (1.32) | (1.13) | (1.78) |
| Term structure of interest rates | −10.577 | −10.796 | −6.988 | −25.585 |
| | (9.02) | (9.11) | (7.45) | (9.45) |
| | Correlation matrix | | | |
| *gap_q* | 1.000 | | | |
| *gap_b* | 0.976 | 1.000 | | |
| *gap_l* | 0.766 | 0.755 | 1.000 | |
| *gap_cbo* | 0.816 | 0.862 | 0.619 | 1.000 |

This table reports results from regressing business cycle variables on four measures of the contemporaneous output gap. The macroeconomic variables are growth rates in corporate earnings, inventories, real GDP, industrial production, real investment, and the term structure of interest rates. *gap_q* is the deviation of the logarithm of total industrial production from a trend that includes both a linear component and a quadratic component. *gap_b* is the deviation of the logarithm of industrial production from a trend that includes a linear component that contains a break in the linear trend in 1977, *gap_l* is the deviation of the logarithm of industrial production from a trend that includes a linear component. *gap_cbo* is the difference between potential GDP and actual GDP collected from the CBO. Total industrial production data is vintage data. $t$-statistics are in parentheses. All data are sampled quarterly from 1948Q1 to 2005Q4.

gap, as measured by the quadratic trend, and these macroeconomic aggregates, which is generally statistically significant. As these macroeconomic variables are highly procyclical, this result confirms the strong link between the output gap and the business cycle: the output gap is high during expansions and low during recessions. We also regress the term structure of interest rates, a financial market-based measure that has been used as a business cycle indicator, on the output gap. We find that there is a negative and statistically significant relationship between the two: a widening of the term structure is evident in recessions, which is also related to a fall in the output gap. Figure 1 plots the monthly measure of the output gap based on the quadratic trend along with shaded NBER recessions dates. The output gap begins to fall before and throughout every recession; the largest falls in the output gap are through these recession periods. After a recession, the output gap rises. Thus, there appears to be a clear relationship between the output gap and the business cycles, as defined by the NBER.

Table 1 also shows that a very similar relationship between the macroeconomic variables and *gap* is observed when using the breaking trend measure of *gap*. This finding is not particularly surprising, given that the bottom of Table 1 reports a high correlation coefficient between the quadratic and breaking trend measures of *gap*. When considering the linear trend measure of *gap*, we find

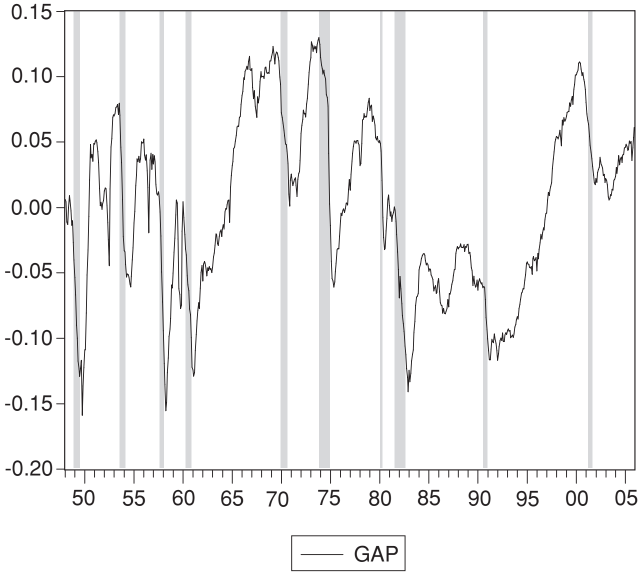


**Figure 1**
**The output gap**
The output gap is measured as deviation from a linear and a quadratic trend. NBER recession dates are in shaded areas.

a similar positive relationship between the procyclical macroeconomic variables and *gap*; however, it is slightly weaker than the previous two measures of *gap*. The final measure of *gap* based on the CBO's measure of potential GDP has a similar relationship to the macroeconomic variables as the quadratic and breaking trend measures of *gap*. The main part of the analysis in our paper is conducted using the quadratic version of *gap*, given its close relationship with procyclical variables in Table 1. However, we do report some findings using the other measures to illustrate that they are robust to how *gap* is measured.

In the empirical analysis, we assess how robust the output gap is as a predictor of stock returns relative to a set of instruments that have been used in the extant literature. The dividend yield, $dy$, measured as the difference between the log of the 12-month moving sum of dividends and the log of lagged price, has been shown to have predictive power by Campbell and Shiller (1988), and Fama and French (1988), among many others. The term spread, *term*, measured as the difference between the ten-year Treasury bond and the thirty-day T-bill rate, and the default spread, *def*, measured as the difference between AAA and Baa corporate bonds, have been shown to have predictive power for returns by, among others, Fama and French (1989). The risk-free rate has been shown to be able to predict returns by, for example, Campbell (1987), and Ang and Bekaert (2007). We use the change in the risk-free rate, *drf*, because its level is highly correlated with default risk.

**Table 2**
**Predicting stock returns in-sample at business cycle frequencies**

| | Excess returns | | | Actual returns | | |
|---|---|---|---|---|---|---|
| | Constant | *gap* | $\overline{R}^2$ | Constant | *gap* | $\overline{R}^2$ |
| | | | Panel A: Monthly returns | | | |
| CRSP | 0.006 | −0.114 | 0.02 | 0.010 | −0.107 | 0.02 |
| | (3.94) | (4.08) | | (6.46) | (3.85) | |
| S&P500 | 0.003 | −0.111 | 0.02 | 0.007 | −0.103 | 0.02 |
| | (2.12) | (4.14) | | (4.73) | (3.90) | |
| | | | Panel B: Quarterly returns | | | |
| CRSP | 0.018 | −0.316 | 0.05 | 0.031 | −0.295 | 0.05 |
| | (4.31) | (4.39) | | (7.43) | (4.14) | |
| S&P500 | 0.009 | −0.308 | 0.06 | 0.022 | −0.288 | 0.05 |
| | (2.38) | (4.76) | | (5.43) | (4.54) | |
| | | | Panel C: Annual returns | | | |
| CRSP | 0.074 | −0.925 | 0.11 | 0.122 | −0.848 | 0.10 |
| | (4.59) | (3.65) | | (7.91) | (2.79) | |
| S&P500 | 0.039 | −0.947 | 0.12 | 0.087 | −0.871 | 0.11 |
| | (2.37) | (3.53) | | (5.62) | (3.25) | |

This table reports estimates from OLS regressions of current stock returns on the second lag of the output gap. *gap* is measured as the deviations of the log of total industrial production from a quadratic and linear trend. Panel A reports results using monthly returns, Panel B reports results using quarterly returns, and Panel C reports results using annual returns. We report results using the excess returns and actual returns on the CRSP value-weighted index and the S&P500 index. $\overline{R}^2$ is the adjusted $R^2$. The Newey-West corrected $t$-statistics are reported in parentheses. The data are sampled monthly over the period 1948:1 to 2005:12.

## 2. Predicting Stock Returns at Business Cycle Frequencies

We start by assessing the in-sample predictive ability of *gap* for stock returns at business cycle frequencies.[5] We report Newey-West standard errors that correct for serial correlation and heteroscedasticity using lags equal to two times $k-1$, where $k$ is the horizon. We estimate the following univariate regression:

$$r_t = \alpha + \gamma \text{gap}_{t-2} + e_t, \tag{4}$$

where $r_t$ is the excess return, *gap* is measured using a linear and a quadratic trend, and $e_t$ is an error term. Table 2 reports predictability results using one-month (Panel A), three-month (Panel B), and 12-month (Panel C) excess and actual returns on the CRSP value-weighted index and the S&P 500 index.[6] Panel A shows that for one-month excess returns, the estimated coefficients on *gap* are negative, implying that a fall in *gap* today predicts higher future expected returns. The estimated coefficients are highly statistically significant and the $\overline{R}^2$s from the one-month excess return regressions are 2% for both indices. When predicting actual returns (the right-hand side of Panel A), results are very similar to those of excess returns.

[5] The average length of NBER contractions in the 1945–2001 period is 10 months. The NBER counts 10 cycles in this period.

[6] Unreported results using real returns are very similar.

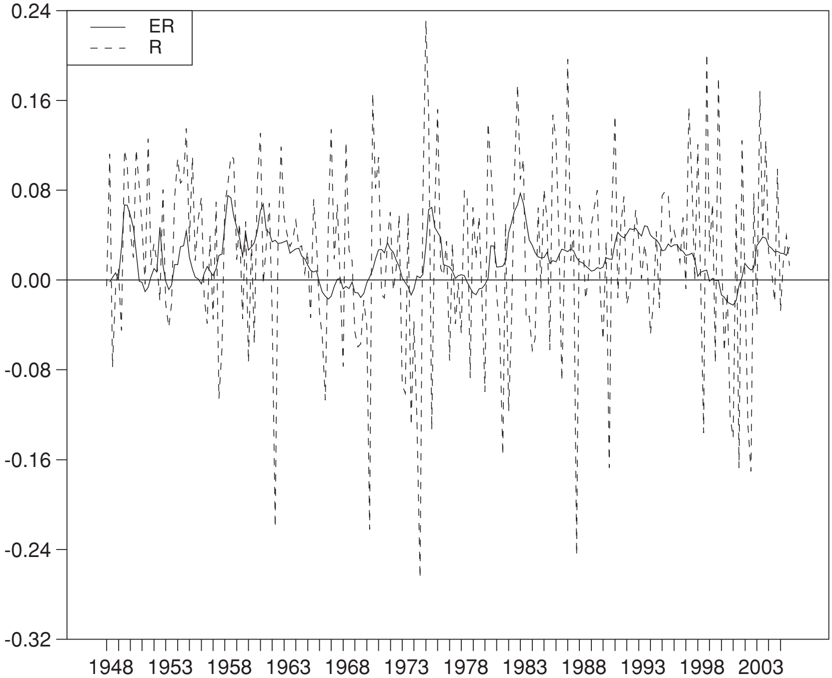


**Figure 2**
**Actual returns (R) and expected returns (ER)**

Panel B of Table 2 reports results from using quarterly returns formed from overlapping monthly observations. Both CRSP and S&P excess and actual returns are strongly predictable with $t$-statistics around 4 and the $\overline{R}^2$s are 5% to 6%. These results are impressive when set against the extant literature. Panel C of Table 2 examines predictability using annual returns formed from overlapping monthly observations. The estimated coefficients are statistically significant and the $\overline{R}^2$s are 11% to 12%. Unreported results using nonoverlapping quarterly and annual returns, rather than those obtained from overlapping monthly returns, are very similar. In contrast to Rangvid (2006), who finds that actual returns are predictable but excess returns are only marginally predictable using GDP scaled by price, we find no difference in the predictability of actual and excess returns; interestingly, this does not stem from the inability of *gap* to predict the risk-free rate of return because the risk-free rate is also highly predictable.

The economic impact of *gap* on returns is large but reasonable; for example, the estimated coefficient on *gap* from the annual predictability regression for the CRSP value-weighted index is –0.925, implying that a one-standard-deviation fall in *gap* leads to an increase in expected annual excess stock returns of 5.01%. This magnitude is larger than that of the dividend yield (3.60% per annum) but smaller than *cay* (7.39% per annum) and the net payout ratio (10.2%).[7] Figure 2 plots the expected return (fitted values) from the regression that forecasts with *gap* along with the actual excess return, at a quarterly frequency. The expected return series is much less variable than actual returns and the economic magnitude appears reasonable.

In summary, at business cycle frequencies, we find that actual and excess stock returns are predictable using *gap*. This in-sample predictability is strong both statistically and economically and provides evidence that stock returns vary with business cycle conditions: expected returns rise as economic conditions worsen and fall when economic conditions improve. Importantly, *gap* does not include any price variables. Therefore, we provide independent evidence lending support to the notion that predictability is not due to behavioral biases that could lead to a fall in prices, but rather is due to time variation in the required compensation for risk.

Table 3 reports results from performing various robustness tests of the predictability of returns using *gap*. While the univariate regressions have shown that *gap* has in-sample forecasting ability, we do not know if this is robust to the inclusion of predictor variables that have been used in earlier work. To assess this, we add the financial market-based predictor variables (*dy*, *drf*, *term*, and *def*) defined by the vector **Z** to the regression that includes *gap* and estimate

$$r_t = \alpha + \gamma \mathrm{gap}_{t-2} + \theta' \mathbf{Z}_{t-1} + u_t, \tag{5}$$

where $\theta'$ is a $1 \times 4$ vector of coefficient estimates on the variables in $\mathbf{Z}_{t-1}$, and $u_t$ is an error term. Panel A of Table 3 examines the robustness of predictability with *gap* to this set of alternative predictor variables. *gap* retains its forecasting power with roughly the same coefficient size and same level of statistical significance when compared to the financial market-based variables. Two recent papers, Boudoukh et al. (2007) and Robertson and Wright (2006), have documented that payout yields derived from dividends, repurchases, and issuances, as opposed to the simple dividend yield, are robust predictors of excess returns. In particular, Boudoukh et al. (2007) show that the net payout ratio (dividends plus repurchases minus issuances) performs particularly well. Panel B reports results regarding the predictability of returns with the net payout ratio, *net*. Using monthly returns, *net* is marginally statistically significant, but when using annual returns formed from overlapping observations it is highly statistically significant and the $\overline{R}^2$ is 10%. The next row of Panel B reports a regression of returns on both *net* and *gap* and shows that with monthly returns *net* is not statistically significant, but *gap* is. However, in predicting annual returns, both *net* and *gap* are statistically significant and the $\overline{R}^2$ is an impressive 16%.

Interestingly, the correlation coefficient between *gap* and *net* is –0.27, which suggests that *net* and *gap* are capturing different factors that predict returns.

[7] See Table 6 in Lettau and Ludvigson (2001) for the estimate of the annual coefficient on *cay*, and Table II in Boudoukh et al. (2007) for the estimate of the annual coefficient on the net payout ratio.

**Table 3**
**Predicting stocks returns: control variables and robustness**

Panel A: Financial variables

| | *Constant* | *gap* | *dy* | term | *def* | *drf* | $\overline{R}^2$ |
|---|---|---|---|---|---|---|---|
| CRSP | 0.036 | −0.097 | 0.008 | 0.089 | −0.376 | −14.376 | 0.03 |
| | (2.46) | (3.11) | (1.93) | (0.62) | (0.85) | (2.53) | |
| S&P | 0.025 | −0.101 | 0.006 | 0.078 | −0.493 | −13.476 | 0.03 |
| | (1.78) | (3.28) | (1.36) | (0.55) | (1.14) | (2.59) | |

Panel B: CRSP; Net payout ratio

| Monthly | | | | Annual | | | |
|---|---|---|---|---|---|---|---|
| Constant | *gap* | *net* | $\overline{R}^2$ | Constant | *gap* | *net* | $\overline{R}^2$ |
| 0.042 | | 0.016 | 0.01 | 0.435 | | 0.216 | 0.10 |
| (2.28) | | (1.93) | | (3.41) | | (3.53) | |
| 0.027 | −0.103 | 0.009 | 0.02 | 0.435 | −0.733 | 0.166 | 0.16 |
| (1.48) | (3.69) | (1.15) | | (3.41) | (2.73) | (2.84) | |
| Orthogonalized *net* | | | | | | | |
| 0.006 | −0.114 | 0.009 | 0.02 | 0.073 | −0.919 | 0.158 | 0.16 |
| (3.95) | (4.09) | (1.11) | | (4.80) | (3.47) | (2.67) | |

Panel C: Subsample analysis: 1970

| | *Constant* | *gap* | $\overline{R}^2$ |
|---|---|---|---|
| 1948:1 to 1970:12 | | | |
| CRSP | 0.009 | −0.079 | 0.01 |
| | (4.08) | (2.38) | |
| S&P | 0.006 | −0.078 | 0.01 |
| | (2.50) | (2.21) | |
| 1970:1 to 2005:12 | | | |
| CRSP | 0.005 | −0.138 | 0.02 |
| | (2.15) | (3.09) | |
| S&P | 0.002 | −0.131 | 0.02 |
| | (0.95) | (3.10) | |

Panel D: Subsample analysis: 1975

| | *Constant* | *gap* | $\overline{R}^2$ |
|---|---|---|---|
| 1948:1 to 1974:12 | | | |
| CRSP | 0.007 | −0.124 | 0.03 |
| | (3.23) | (3.78) | |
| S&P | 0.004 | −0.122 | 0.03 |
| | (1.82) | (3.77) | |

*continued overleaf*

**Table 3**
**(Continued)**

Panel D: Subsample analysis: 1975

| | *Constant* | *gap* | $\overline{R}^2$ |
|---|---|---|---|
| | | 1975:1 to 2005:12 | |
| CRSP | 0.006 | −0.103 | 0.01 |
| | (2.52) | (2.22) | |
| S&P | 0.003 | −0.099 | 0.01 |
| | (1.37) | (2.28) | |

Panel A reports estimates from OLS regressions of current stock returns on the second lag of *gap* and a set of control variables: the dividend yield, *dy*, which is defined as the difference between the log of dividends and the log of lagged prices; the term spread, *term*, which is defined as the yield on 10-year government bonds minus the yield on three-month T-bills; the default spread, *def*, which is the difference between the yield on BAA corporate bonds and the yield on AAA corporate bonds; the change in the risk-free rate, *drf*, where the risk-free rate is the one-month T-bill; *gap* is measured as the deviations of the log of total industrial production from a quadratic and linear trend. Panel B reports OLS regressions of current stock returns on the second lag of *gap* and the net payout ratio, *net*. Orthogonalized *net* is derived from the residuals of a regression *net* on *gap*. The data used in Panels A and B are sampled 1948:1 to 2005:12. Panels C and D report subsample analysis from OLS regressions of current stock returns on the second lag of *gap*. $\overline{R}^2$ is the adjusted $R^2$. The Newey-West-corrected *t*-statistics are reported in parentheses.

Recall from Table 1 and Figure 1 that *gap* is correlated with a number of procyclical macroeconomic variables and NBER recession dates. In contrast, unreported results show that in regressions using quarterly data of these macroeconomic variables and the term structure on *net*, *net* is never statistically significant. Therefore, it seems that *net* is capturing something other than business cycle frequency macroeconomic fluctuations. Due to the considerable negative correlation between *gap* and *net*, it makes sense to look at the relative performance of *gap* and *net* in predicting returns when *net* is orthogonalized relative to *gap*. We are then asking if *net* contributes anything above that which is related to *gap*. The final row of Panel B of Table 3 shows that at monthly horizons there is no additional information in the orthogonalized version of *net*. At the annual horizon, however, both *gap* and orthogonalized *net* are statistically significant and the $\overline{R}^2$ is 16%.[8] That the $\overline{R}^2$ is the same for annual returns irrespective of whether *net* is orthogonalized suggests the *gap* and *net* contain independent information regarding risk premiums.

We investigate further the predictability of returns with *net* by looking at various subsamples, in particular in the pre- and post-1975 sample and in a sample that omits the 1990s where the dividend yield is known to lose its predictive ability for returns (results are unreported). On its own and together with *gap*, *net* has forecasting power for monthly excess returns in the 1948:1 to 1974:12 subsample. Using annual returns, *net* is also statistically significant in that period. In the 1975:1 to 2005:12 subsample, *net* is not statistically significant until horizons of two years or greater. When we omit data from

[8] Unreported results indicate that *net* is not statistically significant at the quarterly horizon when included along with *gap*. At horizons longer than one year, both *gap* and *net* are statistically significant.

1990 onward, *net* is not statistically significant using monthly returns, although it is using annual returns. Interestingly, if we omit data in the aftermath of the run-up in prices (2000–2005), *net* cannot predict returns at any horizon, starting either in 1948:1 or 1975:1.

Panel C of Table 3 addresses the issue of the robustness of predictability by *gap* to subsample analysis by splitting the data into two roughly equal periods: one starting in 1948 and ending in 1969 and the other starting in 1970 and ending in 2005. We choose this split because of the decline in output growth in the 1970s, as well as providing similar-sized subsamples. The predictability of stock returns by the output gap is statistically significant in each subperiod.

Results in Goyal and Welch (2006) show that a number of predictor variables lose their forecasting power after the oil price crisis in the 1970s. To assess the sensitivity of in-sample predictability to this shock, we compute a Chow test that tests for a break in the estimated coefficient. We run two regressions: one over the period 1948:1 to 1974:12 and the other over the period 1975:1 to 2005:12. The Chow test has a value of 0.133 with an associated probability value of 0.94. Panel D of Table 3 reports the estimated coefficient over the 1948:1 to 1974:12 and the 1975:1 to 2005:12 subperiods and shows that predictability is largely unaffected by splitting the sample in 1975. Therefore, unlike for some other predictor variables, the oil price crisis in the mid-1970s does not have a significant impact on the in-sample predictability using *gap*. Our finding of in-sample predictability in the 1975–2005 period is important in view of Goyal and Welch's finding that most predictor variables cease to be significant in-sample in this period. In unreported results, we are able to show that the predictability of returns by *gap* is unaffected by ending the sample in 1989, the period where the dividend yield performs well, or in 1999, the period before the fall in prices related to technology stocks.

So far, we have focused on a version of *gap* measured using a quadratic trend. Table 4 reports results of predictability at business cycle frequencies using the three other measures of *gap* discussed in Table 1. We find evidence of predictability using all three measures at all three horizons considered. The predictability is stronger at the 12-month horizon when using the linear trend version of *gap* than when using the other measures of *gap*. However, irrespective of how we measure *gap*, there is strong evidence of predictability at all horizons for all measures.

### 2.1 International evidence

In order to guard against data-snooping biases, and to serve as an out-of-sample test of the in-sample predictability of U.S. stock returns by *gap*, we also report results using data from the remaining G7 countries: Canada, France, Germany, Italy, Japan, and the UK. The excess stock returns are calculated from the Morgan Stanley capital market total return index minus the local short-term interest rate (the short-term rates are the three-month T-bill rate in Canada, France, and the UK; the three-month Euro-Mark rate in Germany; the

**Table 4**
**Alternative measures of the output gap**

| | Monthly returns | | Quarterly returns | | Annual returns | |
|---|---|---|---|---|---|---|
| | Estimate | $\overline{R}^2$ | Estimate | $\overline{R}^2$ | Estimate | $\overline{R}^2$ |
| *gap_b* | −0.112 | 0.02 | −0.309 | 0.05 | −0.902 | 0.10 |
| | (4.01) | | (4.24) | | (3.06) | |
| *gap_l* | −0.089 | 0.02 | −0.251 | 0.06 | −0.824 | 0.14 |
| | (4.32) | | (4.60) | | (3.71) | |
| *gap_cbo* | | | −0.454 | 0.02 | 1.424 | 0.05 |
| | | | (2.48) | | (2.30) | |

This table reports estimates from OLS regressions of current stock returns on the second lag of different measures of the output gap, *gap*. *gap_b* is the deviation of the logarithm of industrial production from a trend that includes a linear component that contains a break in the linear trend in 1977, *gap_l*is the deviation of the logarithm of industrial production from a trend that includes a linear component. *gap_cbo*is the difference between potential GDP and actual GDP collected from the CBO. We report results using the excess returns on the CRSP value-weighted index at a monthly, quarterly, and annual frequency. $\overline{R}^2$ is the adjusted $R^2$.The Newey-West-corrected $t$-statistics are reported in parentheses. The data are sampled monthly over the period 1948:1 to 2005:12.

**Table 5**
**International evidence**

| | Monthly returns | | | Quarterly returns | | | Annual returns | | |
|---|---|---|---|---|---|---|---|---|---|
| | Const. | *gap* | $\overline{R}^2$ | Const. | *gap* | $\overline{R}^2$ | Const. | *gap* | $\overline{R}^2$ |
| Canada | 0.004 | −0.064 | 0.00 | 0.011 | −0.221 | 0.02 | 0.044 | −0.883 | 0.06 |
| | (1.55) | (1.36) | | (1.65) | (1.66) | | (1.70) | (1.87) | |
| France | 0.005 | −0.101 | 0.00 | 0.013 | −0.265 | 0.01 | 0.056 | −0.901 | 0.03 |
| | (1.58) | (1.70) | | (1.62) | (1.57) | | (1.66) | (1.34) | |
| Germany | 0.004 | −0.204 | 0.02 | 0.011 | −0.625 | 0.06 | 0.047 | −2.075 | 0.16 |
| | (1.42) | (3.32) | | (1.52) | (4.15) | | (1.75) | (4.39) | |
| Italy | 0.011 | −0.097 | 0.00 | 0.031 | −0.270 | 0.01 | 0.131 | −0.976 | 0.02 |
| | (3.17) | (1.23) | | (3.11) | (1.02) | | (2.74) | (0.75) | |
| Japan | 0.004 | −0.081 | 0.01 | 0.012 | −0.269 | 0.03 | 0.046 | −1.107 | 0.09 |
| | (1.66) | (1.75) | | (1.61) | (2.21) | | (1.43) | (2.47) | |
| U.K. | 0.005 | −0.128 | 0.01 | 0.014 | −0.355 | 0.02 | 0.062 | −1.075 | 0.04 |
| | (1.73) | (2.16) | | (1.88) | (2.36) | | (2.49) | (1.69) | |

This table reports estimates from OLS regressions of current stock returns at the one-month, one-quarter, and one-year horizon on lags of the output *gap*. In each country, *gap* is measured as the deviation of the logarithm of total industrial production from a trend that includes both a linear component and a quadratic component. We report results using the excess returns on the MSCI-value-weighted indices for Canada, France, Germany, Italy, Japan, and the UK. $\overline{R}^2$ is the adjusted $R^2$. Const. is constant. The Newey-West-corrected $t$-statistics are reported in parentheses. The sample period is 1970:1 to 2005:12.

three-month interbank deposit rate in Italy; and the overnight money market rate in Japan). Industrial production is production in the total economy in Canada, Italy, and the UK, and production in total manufacturing in France, Germany, and Japan. The data are sampled from 1970:1 to 2005:12 and are collected from Datastream. As in the case of the U.S. data, in order to account for publication lags, we use the second lag of *gap* in our tests. Note that for these six countries we only have current data, which have been revised ex post. Therefore, we are only able to focus on in-sample predictability.

We use the quadratic trend to measure *gap* from each individual country's industrial production index. Table 5 presents the cross-country evidence regarding predictability at business cycle frequencies using one-, three-, and

12-month excess returns. At the one-month horizon, we find a negative relationship between *gap* and future excess stock returns in every country, and in every country except Canada and Italy, the estimated coefficient is statistically significant (marginally so in France and Japan, although if we consider that the test is one-sided given that the estimate of *gap* under the alternative is negative, then this evidence is no longer marginal). The statistical significance and the $\overline{R}^2$s are a little smaller than in the United States, but this is to be expected given the shorter sample available in these countries, which covers considerably fewer recessions and expansions. Table 5 also reports results using quarterly and annual returns and shows that there is a negative relationship between excess returns and *gap* in all countries and that in many countries this relationship is statistically significant.

The evidence presented using data from six additional countries that excess stock returns are predictable by *gap* should go a long way toward guarding against the potential explanation that the findings of in-sample predictability using U.S. data are the result of data mining.[9] Further evidence against data mining can be obtained from undertaking out-of-sample forecasting.

### 2.2 Out-of-sample evidence

In the extant literature there is some evidence that variables that can predict in-sample cannot predict out-of-sample better than a constant. For example, Bossaerts and Hillion (1999), and Goyal and Welch (2003) find no evidence that the dividend yield can predict out-of-sample better than a constant. More recently, Goyal and Welch (2006) examine the out-of-sample predictive ability of a large number of predictor variables and find little evidence that they can predict out-of-sample better than a constant.[10] Campbell and Thompson (2007) show that sensible restrictions on forecasting models leads to the finding that a number of predictor variables have out-of-sample forecasting ability.

In order to provide out-of-sample forecasts that could actually have been made by an investor, it is necessary to use only information that was available to the investor at the time the forecast was made. Recall that we are using vintage data, that is, data that have not been subsequently revised, so at each point in time the data are available to the investor. In the in-sample analysis, where the parameters on *gap* are estimated over the entire sample, we regressed returns at time $t$ on *gap* at time $t-2$ to account for the publication lag in industrial production. However, in the out-of-sample analysis, we recursively estimate the parameters on *gap* every month. Therefore, in the out-of-sample analysis, if we were to regress returns at time $t$ on *gap* at time $t-2$, where *gap* is estimated

[9] Similar results are obtained when using actual and real returns.

[10] While out-of-sample evidence is important, interpreting the finding that returns are not predictable out-of-sample is problematic. See, for example, Inoue and Kilian (2004), Cochrane (2006), Hjalmarsson (2006), and Campbell and Thompson (2007), who show that evidence of predictability based on in-sample forecasting is likely to be more reliable than that based on out-of-sample evidence. Their findings show that uncovering out-of-sample evidence of predictability may not be possible in small samples even when returns are actually predictable.

using industrial production data from the beginning of the sample up to time $t-1$, it would require knowledge of the parameters that provide the estimate of *gap* at time $t-1$. This information is clearly not available to investors because of the publication lag. Therefore, we proceed slightly differently in the out-of-sample tests.

Instead of using the second lag of *gap*, we adjust the data to take account of the publication lag by replacing February 1948 industrial production with January 1948 industrial production and replacing March 1948 industrial production with February 1948 industrial production, and so on to the end of the sample where November 2005 industrial production replaces December 2005 industrial production (this procedure implies that our first observation of the output gap is 1948:2). We then forecast out-of-sample using

$$r_t = \alpha + \gamma \mathrm{gap}_{t-1} + e_t. \tag{6}$$

Note that this procedure still implies that we are using *gap* at month $t-2$ to predict month $t$'s excess returns, *gap* at month $t-4$ to predict the excess returns in the quarter that starts at the beginning of month $t-2$ and ends at the end of month $t$, and *gap* at month $t-13$ to predict the excess returns in the year that starts at the beginning of month $t-11$ and ends at the end of month $t$.

Given that we have adjusted the data for the publication lag, we want to ensure that the parameter estimates of the trend are also available to the investor. Therefore, for the out-of-sample tests, we estimate the trend coefficients recursively as follows: starting in 1948:2 we estimate until 1951:12 to get the first estimate of the parameters and *gap*

$$y_t = a_\tau + b_\tau \cdot t + c_\tau \cdot t^2 + v_t, \tag{7}$$

where $\tau$ =1951:12, $t = 1, 2, 3, \ldots, \tau$, and the residual $v_t$ is the measure of *gap* over the period 1948:2 to 1951:12. Note the subscript $\tau$ for the three parameters, which indicates that they are updated with each ending month. Next, we update the estimates of the trend by one month by estimating over the period 1948:2 to 1952:1:

$$y_t = a_{\tau+1} + b_{\tau+1} \cdot t + c_{\tau+1} \cdot t^2 + v_t, \tag{8}$$

where month $\tau + 1$ is 1952:1 and $t = 1, 2, 3, \ldots, \tau + 1$). We add on the estimate of *gap* in 1952:1 to the time series of *gap* estimated previously over the period 1948:2 to 1951:12. We then repeat this, month-by-month, estimating new trend coefficients and values of *gap* until the end of the sample. All three trend-based measures of *gap* use this recursive procedure which ensures the investors have access to both the data and parameter estimates that we use.

Given an initial estimation period of 1948:2 to 1951:12, we then estimate (6) and form an out-of-sample forecast of returns for 1952:1. We then add on one month and re-estimate (6), forming a new out-of-sample forecast for 1952:2. We repeat this process, month-by-month, to the end of the sample.

We conduct several out-of-sample tests. The benchmark model that we want to compare *gap* to is one where excess returns are regressed on a constant, month-by-month, to provide forecasts at each month of excess returns based on the historical mean. This constant expected return model is a restricted, nested version of a model of time-varying expected returns that includes a constant and *gap*. The assessment of out-of-sample predictability involves four metrics. The first is a test that asks if the forecasts from one model *encompass* the forecasts from another. If the forecasts from the constant expected return model do not encompass the forecasts from the time-varying expected return model, then the latter model has some information that is useful for forecasting out-of-sample. Clark and McCracken (2001) extend the encompassing test of Harvey, Leybourne, and Newbold (1998) by deriving the nonstandard asymptotic distribution of a test statistic for forecast that encompassing, which is termed ENC-NEW. Clark and McCracken show that the encompassing test has more power than tests of the equality of mean-squared forecast errors. We employ the ENC-NEW test to examine whether the forecasts from the constant expected return model encompass the forecasts from the time-varying expected return model that includes a constant and *gap*. The test is given as

$$\mathrm{ENC-NEW} = \frac{T-h+1}{T} \cdot \frac{\Sigma_{t=1}^{T}\left(\varepsilon_t^2 - \varepsilon_t \cdot e_t\right)}{\mathrm{MSE}_e}, \tag{9}$$

where $T$ is the number of observations, $h$ is the degree of overlap and is equal to one when there is no overlap, $\varepsilon_t$ is the vector of rolling out-of-sample errors from the historical mean model, $e_t$ is the vector of rolling out-of-sample errors from the forecasting model including *gap*, and $\mathrm{MSE}_e$ is the mean-squared error from the forecasting model that includes *gap*.

The second test statistic we report tests for the equality of the mean-squared forecasting errors of one forecast relative to another. To do this, we use the MSE-$F$ test developed by McCracken (2007), which tests the null hypothesis that the constant expected return model has a mean-squared forecasting error that is less than, or equal to, that of the time-varying expected return model. The alternative hypothesis is that the time-varying expected return model has a lower MSE. The test statistic is given as

$$\mathrm{MSE}-F = (T-h+1) \cdot \left(\frac{\mathrm{MSE}_\varepsilon - \mathrm{MSE}_e}{\mathrm{MSE}_e}\right), \tag{10}$$

where $\mathrm{MSE}_\varepsilon$ is the mean-squared error from the model that includes just a constant. For both the ENC-NEW and MSE-$F$ tests, we follow the methodology in Clark and McCracken (2005), which provides bootstrapped critical values for these nonstandard tests.

Further analysis of the out-of-sample performance of *gap* in predicting stock returns is obtained from calculating the out-of-sample $R^2$, $R^2_{\mathrm{oos}}$, which following Campbell and Thompson (2007) is defined as

$$R^2_{\text{oos}} = 1 - \frac{\sum_{t=1}^{T}(r_t - \widehat{r}_t)^2}{\sum_{t=1}^{T}(r_t - \overline{r}_t)^2}, \tag{11}$$

where $\widehat{r}_t$ is the forecast of excess return based on data up to $t-1$, and $\overline{r}_t$ is the historical average excess return estimated using data up to $t-1$. The $R^2_{\text{oos}}$ is measured in units that are comparable to the in-sample $R^2$. If the out-of-sample $R^2$ is positive, then the predictive regression has lower average mean-squared prediction error than the historical average return.

As a final measure of the out-of-sample performance of *gap*, we follow Campbell and Thompson (2007) and calculate the utility gains for a mean-variance investor from using the time-varying expected returns model relative to using the historical mean return forecast. As in Campbell and Thompson, we assume that the investor's objective function is expected portfolio return less ($\gamma$/2) portfolio variance, where $\gamma$ can be interpreted as the coefficient of relative risk aversion. As in Campbell and Thompson, $\gamma$ is assumed to be 3. We impose realistic portfolio constraints, preventing the investor from shorting stocks or taking more than 50% leverage. Thus, the fraction of the investor's wealth invested in stocks is between 0 and 150%.

In order to rule out any one particular period driving the out-of-sample results, we present findings where the initial estimation periods are 1948:2 to 1951:12, 1948:2 to 1959:12, 1948:2 to 1969:12, 1948:2 to 1974:12, 1948:2 to 1979:12, 1948:2 to 1989:12, and 1948:2 to 1999:12. This provides out-of-sample evidence for an investor who started investing in 1952, at the beginning of each decade from 1960 to 2000, and at the beginning of 1975, the period Goyal and Welch (2006) focus on. This guards against the out-of-sample evidence being affected by any particularly good or bad performance in a short period. In particular, Goyal and Welch (2006) show that out-of-sample predictability is not evident in the last three decades, although Campbell and Thompson (2007) do find evidence that some variables have positive out-of-sample $R^2$s and provide positive utility gains in more recent periods.

Panel A of Table 6 reports the results of comparing forecasts based on a constant to those based on *gap* as measured by the quadratic trend, using CRSP excess returns. Considering the initial estimation period of 1948:2 to 1951:12, we find that the one-, three-, and 12-month return forecasts based on *gap* are better than those using a constant. For example, over all horizons, the ENC-NEW encompassing tests reject the null hypothesis that the forecasts from the constant expected return model encompass the forecasts from the time-varying expected return model. The MSE-*F* tests all reject the null hypothesis in favor of the alternative that the MSEs from the forecasts that use *gap* are smaller than those based on the historical average return. The out-of-sample $R^2$ is 0.5% at the one-month horizon, 1.8% at the three-month horizon, and 3.6% at the

**Table 6**
**Predicting excess stock returns out-of-sample**

| | ENC-NEW | | MSE-*F* | | | |
|---|---|---|---|---|---|---|
| | Statistic | Boot p.v. | Statistic | Boot p.v. | $R^2_{oos}$ | U Gain |
| **Panel A: Quadratic trend** | | | | | | |
| Forecasting from 1952 | | | | | | |
| 1 month | 4.69 | 0.01 | 3.60 | 0.01 | 0.005 | 0.15 |
| 3 month | 10.04 | 0.02 | 7.08 | 0.01 | 0.018 | 0.10 |
| 12 month | 23.93 | 0.05 | 14.61 | 0.03 | 0.036 | 0.23 |
| Forecasting from 1960 | | | | | | |
| 1 month | 4.64 | 0.01 | 3.74 | 0.01 | 0.006 | 0.12 |
| 3 month | 8.72 | 0.02 | 6.05 | 0.01 | 0.017 | 0.12 |
| 12 month | 21.38 | 0.02 | 13.26 | 0.04 | 0.035 | 0.27 |
| Forecasting from 1970 | | | | | | |
| 1 month | 2.69 | 0.01 | 1.55 | 0.04 | 0.003 | 0.17 |
| 3 month | 5.28 | 0.02 | 2.88 | 0.07 | 0.011 | 0.32 |
| 12 month | 14.01 | 0.04 | 6.79 | 0.10 | 0.020 | 0.42 |
| Forecasting from 1975 | | | | | | |
| 1 month | 2.99 | 0.01 | 2.68 | 0.01 | 0.007 | 0.24 |
| 3 month | 5.86 | 0.01 | 4.97 | 0.02 | 0.019 | 0.36 |
| 12 month | 14.01 | 0.03 | 10.41 | 0.07 | 0.037 | 0.34 |
| Forecasting from 1980 | | | | | | |
| 1 month | 1.90 | 0.02 | 1.58 | 0.04 | 0.005 | 0.19 |
| 3 month | 4.03 | 0.02 | 3.68 | 0.04 | 0.016 | 0.30 |
| 12 month | 8.42 | 0.07 | 6.91 | 0.09 | 0.030 | 0.35 |
| Forecasting from 1990 | | | | | | |
| 1 month | 1.29 | 0.01 | 1.39 | 0.03 | 0.008 | 0.09 |
| 3 month | 2.98 | 0.01 | 3.53 | 0.03 | 0.026 | 0.15 |
| 12 month | 7.51 | 0.04 | 9.95 | 0.05 | 0.085 | 0.57 |
| Forecasting from 2000 | | | | | | |
| 1 month | 1.88 | 0.00 | 3.12 | 0.00 | 0.041 | 0.45 |
| 3 month | 4.12 | 0.00 | 7.26 | 0.00 | 0.103 | 1.63 |
| 12 month | 7.99 | 0.01 | 14.54 | 0.00 | 0.187 | 1.89 |
| **Panel B: Breaking trend** | | | | | | |
| Forecasting from 1952 | | | | | | |
| 1 month | 8.34 | 0.00 | 7.00 | 0.00 | 0.010 | 0.16 |
| 3 month | 18.53 | 0.01 | 14.58 | 0.00 | 0.029 | 0.09 |
| 12 month | 49.74 | 0.00 | 35.55 | 0.01 | 0.049 | −0.35 |
| Forecasting from 1960 | | | | | | |
| 1 month | 6.95 | 0.00 | 5.65 | 0.00 | 0.009 | 0.13 |
| 3 month | 14.51 | 0.00 | 10.33 | 0.01 | 0.023 | 0.11 |
| 12 month | 36.19 | 0.00 | 22.31 | 0.02 | 0.042 | −0.40 |
| Forecasting from 1970 | | | | | | |
| 1 month | 4.39 | 0.00 | 2.91 | 0.01 | 0.006 | 0.15 |
| 3 month | 9.47 | 0.00 | 5.53 | 0.02 | 0.018 | 0.23 |
| 12 month | 21.44 | 0.01 | 7.75 | 0.09 | 0.040 | −0.18 |
| Forecasting from 1975 | | | | | | |
| 1 month | 3.85 | 0.00 | 2.12 | 0.02 | 0.006 | 0.19 |
| 3 month | 8.23 | 0.00 | 3.80 | 0.04 | 0.015 | 0.22 |
| 12 month | 20.32 | 0.01 | 3.60 | 0.15 | 0.021 | −0.56 |
| Forecasting from 1980 | | | | | | |
| 1 month | 2.50 | 0.01 | 1.39 | 0.04 | 0.005 | 0.11 |
| 3 month | 5.57 | 0.01 | 3.28 | 0.04 | 0.015 | 0.10 |
| 12 month | 12.57 | 0.02 | 5.12 | 0.11 | 0.025 | −0.77 |
| Forecasting from 1990 | | | | | | |
| 1 month | 1.80 | 0.01 | 1.84 | 0.02 | 0.010 | 0.09 |
| 3 month | 4.37 | 0.01 | 4.86 | 0.01 | 0.031 | 0.20 |
| 12 month | 12.54 | 0.01 | 15.73 | 0.02 | 0.105 | 0.34 |

(*continued overleaf*)

**Table 6**
**(Continued)**

| | ENC-NEW | | MSE-*F* | | | |
|---|---|---|---|---|---|---|
| | Statistic | Boot p.v. | Statistic | Boot p.v. | $R^2_{\text{oos}}$ | U Gain |
| | | | Forecasting from 2000 | | | |
| 1 month | 1.87 | 0.00 | 2.89 | 0.00 | 0.042 | 0.48 |
| 3 month | 4.87 | 0.00 | 7.82 | 0.00 | 0.110 | 1.88 |
| 12 month | 11.57 | 0.01 | 19.64 | 0.00 | 0.224 | 2.70 |
| | | | Panel C: Linear trend | | | |
| | | | Forecasting from 1952 | | | |
| 1 month | 8.58 | 0.00 | 5.85 | 0.00 | 0.008 | 0.29 |
| 3 month | 18.93 | 0.00 | 11.78 | 0.00 | 0.024 | 0.41 |
| 12 month | 51.88 | 0.00 | 31.73 | 0.01 | 0.044 | 0.53 |
| | | | Forecasting from 1960 | | | |
| 1 month | 7.19 | 0.00 | 4.58 | 0.00 | 0.007 | 0.27 |
| 3 month | 14.88 | 0.00 | 7.74 | 0.01 | 0.019 | 0.47 |
| 12 month | 38.54 | 0.01 | 18.70 | 0.03 | 0.036 | 0.61 |
| | | | Forecasting from 1970 | | | |
| 1 month | 4.61 | 0.00 | 1.93 | 0.02 | 0.004 | 0.33 |
| 3 month | 9.63 | 0.00 | 3.16 | 0.05 | 0.013 | 0.70 |
| 12 month | 23.74 | 0.01 | 4.38 | 0.12 | 0.034 | 1.10 |
| | | | Forecasting from 1975 | | | |
| 1 month | 4.09 | 0.00 | 1.08 | 0.07 | 0.003 | 0.41 |
| 3 month | 8.62 | 0.01 | 1.21 | 0.12 | 0.008 | 0.77 |
| 12 month | 23.07 | 0.01 | −0.43 | 0.27 | 0.011 | 0.94 |
| | | | Forecasting from 1980 | | | |
| 1 month | 2.78 | 0.01 | 0.46 | 0.12 | 0.002 | 0.38 |
| 3 month | 6.10 | 0.01 | 1.07 | 0.13 | 0.008 | 0.78 |
| 12 month | 15.81 | 0.02 | 2.76 | 0.16 | 0.016 | 1.11 |
| | | | Forecasting from 1990 | | | |
| 1 month | 1.92 | 0.01 | 1.06 | 0.05 | 0.005 | 0.22 |
| 3 month | 4.68 | 0.01 | 3.14 | 0.04 | 0.020 | 0.38 |
| 12 month | 15.28 | 0.00 | 15.32 | 0.02 | 0.091 | 0.38 |
| | | | Forecasting from 1980 | | | |
| 1 month | 2.78 | 0.01 | 0.46 | 0.12 | 0.002 | 0.38 |
| 3 month | 6.10 | 0.01 | 1.07 | 0.13 | 0.008 | 0.78 |
| 12 month | 15.81 | 0.02 | 2.76 | 0.16 | 0.016 | 1.11 |
| | | | Forecasting from 1990 | | | |
| 1 month | 1.92 | 0.01 | 1.06 | 0.05 | 0.005 | 0.22 |
| 3 month | 4.68 | 0.01 | 3.14 | 0.04 | 0.020 | 0.38 |
| 12 month | 15.28 | 0.00 | 15.32 | 0.02 | 0.091 | 0.38 |
| | | | Forecasting from 2000 | | | |
| 1 month | 1.04 | 0.01 | 1.11 | 0.02 | 0.015 | 0.20 |
| 3 month | 2.49 | 0.01 | 2.85 | 0.02 | 0.046 | 0.74 |
| 12 month | 5.65 | 0.02 | 7.87 | 0.02 | 0.111 | 0.53 |

This table reports the results of out-of-sample forecast comparisons of the excess return on the CRSP Index. The comparisons are of forecasts of excess returns based on a constant (the restricted model) and forecasts based on a constant and *gap* (the unrestricted model). We report comparisons based on forecasting one month ahead using one-, three-, and 12-month returns. The column labeled "ENC-NEW" provides the Clark and McCracken (2001) encompassing test statistic. The column labeled "MSE-*F*" gives the *F*-test of McCracken (2007), which tests the null hypothesis of equal MSEs against the alternative that the MSE from the unrestricted model is smaller. The column labeled "Statistic" provides the test statistics and the column labeled Boot p.v. provides the bootstrapped probability value using the methodology in Clark and McCracken (2005). $R^2_{\text{oos}}$ is the out-of-sample $R^2$. The column headed "U Gain" reports the utility gain (in basis points) an investor would have obtained from using the predictability results as compared to using the historical mean excess returns when forming portfolio weights. Panel A reports results using $gap\_q$, which is measured as the deviation of the log of total industrial production from a trend that includes a quadratic and a linear component. Panel B reports results using $gap\_b$, which is the deviation of the logarithm of industrial production from a trend that includes a linear component that contains a break in the linear trend in 1977. Panel C reports results using $gap\_l$, which is the deviation of the logarithm of industrial production from a trend that includes a linear component.

12-month horizon. The final column of the table reports the utility gain from forecasting with *gap* relative to a constant expected return model. Over all three horizons, the utility gains for a mean-variance investor are positive and are large in magnitude when compared to the findings concerning some of the predictor variables in Campbell and Thompson (2007).[11] Thus, based on forecasts from 1952:1 to the end of the sample, according to all four metrics that judge the out-of-sample forecasting performance, it appears that *gap* can forecast excess stock returns out-of-sample better than a constant.

The remainder of Panel A of Table 6 repeats the out-of-sample analysis where the forecasting starts in a different period. The encompassing tests reject the null hypothesis that the constant expected return model encompasses the time-varying expected return model in all forecasting periods at the 5% level, except when the forecasting starts in 1980 and 12-month returns are used, where in this case the *p*-value is 7%; however, note that the utility gains are large when forecasting starts in 1980, when compared to most of the variables in Campbell and Thompson (2007) in that period. Regarding the MSE-*F* test for the null hypothesis that the MSE from the constant expected return model are less than or equal to those of the time-varying expected return model, only in four cases are the *p*-values greater than 5% and in all these cases they are 10% or less. When considering the out-of-sample $R^2$s they are always positive, irrespective of the forecasting period, as are the utility gains.

The results in Panel A of Table 6 show that even in periods after the 1970s, where evidence in Goyal and Welch (2006) shows that many other predictor variables do not perform well, we find substantial evidence in favor of predictability with *gap*. Even if the initial estimation period is from 1948 to 1975 and forecasts are made after this period, we find evidence in favor of out-of-sample predictability. For example, for this particular period, using one-month returns, the out-of-sample $R^2$ is 0.7%, the utility gain is positive and relatively large from forecasting with *gap*, and both the MSE-*F* test and the ENC-NEW test find evidence in favor of predictability at the 1% level.

Panels B and C of Table 6 assess whether *gap* can forecast out-of-sample better than a constant when *gap* is measured using a breaking trend and a linear trend, respectively. Considering Panel B first, the results reflect very closely those of the quadratic trend: the ENC-NEW test of forecast encompassing finds evidence in favor of *gap* and the MSE-*F* tests reveal that the MSE is smaller from the forecasts that use *gap* relative to a constant. Panel B also reports out-of-sample $R^2$s and utility gains similar to those reported in Panel A (with the exception of annual returns forecasts where some of the utility gains are negative). In Panel C, the linear trend version of *gap* also finds strong support for *gap* being able to forecast out-of-sample. Notably, the utility gains when using this measure are very large and are about 1% annually in the post-1975

[11] Our findings in this column should be compared to the column titled “Unconstrained” of Table 4 in Campbell and Thompson (2007).

period (the utility gains are in units of expected annual returns). There are only minor differences in the out-of-sample performance of the different measures of *gap*. For example, the average out-of-sample $R^2$ using the quadratic *gap* is 3.38%, for the breaking trend *gap* it is 4.4%, and for the linear *gap* it is 2.5%. Thus, it appears that the linear trend measure of *gap* does not do quite as good a job at forecasting out-of-sample as the other two measures when looking at the out-of-sample $R^2$, although it still does considerably better than a constant. However, in terms of utility gains, the linear trend version of *gap* does very well.

In summary, when we measure *gap* and whatever decade an investor started using *gap* to forecast returns, we find that *gap* can predict future returns out-of-sample better than a constant. These results are in contrast to Goyal and Welch (2006), who accentuate the findings that predictor variables have been unsuccessful out-of-sample in the last few decades. However, our findings are in line with those of Campbell and Thompson (2007), who show that certain predictor variables, based on out-of-sample $R^2$ and utility gains, do have out-of-sample forecasting power in recent decades.

## 2.3 Long-horizon forecasts

We are interested in finding out whether *gap* can predict stock returns in-sample at longer horizons, where the extant literature has found the strongest evidence of predictability. However, it is important to recognize that some caution should be placed on interpreting the results from long-horizon regressions especially when using overlapping observations. Boudoukh, Richardson, and Whitelaw (2005) show that for many persistent predictor variables, under the null hypothesis of no predictability, the coefficient estimates and $\overline{R}^2$s are highly correlated across horizons. For example, they show that for the dividend yield, under the null of no predictability, the estimators are over 98% correlated. Sampling error that is present at the one-period horizon appears in every horizon. Under the null hypothesis of no predictability and assuming the estimators are asymptotically distributed as multivariate normal with a mean of zero, Boudoukh, Richardson, and Whitelaw (2005) calculate the implied estimate at the $k$th horizon, given the one-period estimate as

$$E(\widehat{\gamma}_k|\widehat{\gamma}_1) = \left(1 + \frac{\rho(1-\rho^{k-1})}{1-\rho}\right)\widehat{\gamma}_1, \tag{12}$$

where $\widehat{\gamma}_1$ is the actual estimate of the one-period coefficient, $\rho$ is the autocorrelation coefficient of the predictor variable, and $k$ is the horizon. Similarly, the implied $R^2$ can be calculated as

$$E\left(R_k^2\middle|R_1^2\right) = \frac{\left(1 + \frac{\rho(1-\rho^{k-1})}{1-\rho}\right)^2}{k} R_1^2. \tag{13}$$

**Table 7**
**Long-horizon predictability**

| | 24 month | | 36 month | | 48 month | | 60 month | |
|---|---|---|---|---|---|---|---|---|
| | *gap* | $\overline{R}^2$ | *gap* | $\overline{R}^2$ | *gap* | $\overline{R}^2$ | *gap* | $\overline{R}^2$ |
| | | | | Panel A: Actual estimates | | | | |
| CRSP | −1.119 | 0.09 | −1.449 | 0.11 | −1.752 | 0.13 | −2.345 | 0.17 |
| | (2.33) | | (2.66) | | (2.49) | | (2.89) | |
| Canada | −1.550 | 0.11 | −2.219 | 0.20 | −2.329 | 0.21 | −2.363 | 0.19 |
| | (2.76) | | (5.19) | | (4.23) | | (2.74) | |
| France | −1.676 | 0.06 | −2.421 | 0.10 | −2.162 | 0.06 | −1.501 | 0.02 |
| | (1.57) | | (1.74) | | (1.05) | | (0.65) | |
| Germany | −3.548 | 0.24 | −4.318 | 0.26 | −3.957 | 0.18 | −2.690 | 0.08 |
| | (3.04) | | (3.00) | | (3.08) | | (2.73) | |
| Italy | −1.967 | 0.04 | −2.659 | 0.05 | −1.634 | 0.01 | −0.159 | 0.00 |
| | (1.09) | | (1.56) | | (1.05) | | (0.08) | |
| Japan | −1.780 | 0.12 | −2.156 | 0.12 | −2.830 | 0.17 | −3.162 | 0.17 |
| | (2.52) | | (2.37) | | (2.53) | | (2.49) | |
| U.K. | −1.799 | 0.07 | −2.357 | 0.10 | −2.862 | 0.13 | −3.044 | 0.15 |
| | (2.59) | | (2.08) | | (2.33) | | (2.21) | |
| | | | | Panel B: Implied estimates from one-month estimate | | | | |
| CRSP | −2.076 | 0.28 | −2.727 | 0.32 | −3.207 | 0.33 | −3.562 | 0.33 |
| Canada | −1.130 | 0.01 | −1.464 | 0.01 | −1.701 | 0.01 | −1.870 | 0.01 |
| France | −1.403 | 0.01 | −1.658 | 0.01 | −1.793 | 0.01 | −1.864 | 0.01 |
| Germany | −2.047 | 0.08 | −2.212 | 0.07 | −2.266 | 0.05 | −2.284 | 0.04 |
| Italy | −0.561 | 0.00 | −0.567 | 0.00 | −0.567 | 0.00 | −0.567 | 0.01 |
| Japan | −1.491 | 0.14 | −1.967 | 0.16 | −2.323 | 0.17 | −2.589 | 0.17 |
| U.K. | −1.681 | 0.07 | −1.950 | 0.06 | −2.082 | 0.06 | −2.146 | 0.05 |

Panel A reports results from long-horizon regressions using *gap_q*, which is measured as the deviation of the log of total industrial production from a quadratic and linear trend. We report results using the excess returns on the CRSP-value-weighted index for the United States and the excess returns in the other G7 countries using each country's Morgan Stanley Capital Market Index. $\overline{R}^2$ is the adjusted $R^2$. The Newey-West-corrected $t$-statistics are reported in parentheses. Panel B of the table reports implied estimates from the one-month actual estimates under the null of no long-horizon predictability using the methodology of Boudoukh, Richardson, and Whitelaw (2005). The data are sampled 1948:1 to 2005:12.

For a large estimate of $\rho$, we would expect the estimates and $R^2$s to increase monotonically with the horizon under the null of no predictability. Therefore, we report the implied estimates of the coefficients and $\overline{R}^2$s to compare with the actual estimates.

Panel A of Table 7 reports the long-horizon results for the United States and all other G7 countries. The sign on the estimate of *gap* is reassuringly always negative, which indicates that economically the relationship between excess returns and *gap* is of the correct sign. Statistically, there is evidence of predictability at longer horizons: with the exception of France and Italy, all countries' excess returns are predictable by the output gap at horizons as long as 60 months.

To assess whether the predictability found in Panel A of Table 7 is economically significant, Panel B reports the implied estimates from expressions (14) and (15) using the estimated autocorrelation coefficient of *gap* for each country. Focusing first on the United States, given the one-month estimates in Table 2, the implied estimates and $\overline{R}^2$s increase with the horizon. However, as Panel A indicates, the actual estimates and $\overline{R}^2$s are a long way from the implied estimates, suggesting that our results that *gap* can predict stock returns at long horizons are not driven by the bias discussed in Boudoukh, Richardson, and Whitelaw (2005). This is clearly illustrated if we consider the 24-month horizon $\overline{R}^2$, for example, which is actually smaller than the 12-month horizon $\overline{R}^2$, while the implied estimates are never decreasing.

For other G7 countries, there is evidence that predictability is economically important as the estimates and, in particular, the $\overline{R}^2$s are far from their implied values under the null of no predictability. For example, in Canada, the actual $\overline{R}^2$s at long horizons are approximately 20%, a far cry from the implied $\overline{R}^2$s of roughly 1% under the null of no predictability. Similarly, there are differences in the magnitudes and the pattern of the estimated and implied $\overline{R}^2$s across Germany, Japan, and the UK, the other countries where we find evidence of long-horizon predictability.

## 3. Biases in Forecasting Regressions

There is a growing literature based on both simulation and analytical work illustrating the weakness of relying on asymptotic distribution theory when using standard $t$-statistics to interpret predictability regressions. In particular, large sample theory can provide a poor approximation to the finite sample distribution of test statistics when the regressor used to do the predicting is persistent and its errors from an autoregressive regression are highly correlated with the variable being predicted (see, for example, Mankiw and Shapiro 1986; Stambaugh 1999).

Recent papers in finance have focused on how these problems affect return predictability (see, for example, Valkanov 2003; Lewellen 2004; Torous, Valkanov, and Yan 2005; Campbell and Yogo 2006). In particular, the problems have the potential to be particularly severe when the predictor variables are scaled by price. This is because the innovations in the autoregressive model of the predictor regression will be highly correlated with returns by construction. Of course, for *gap*, which does not include prices, this should be less of a problem.[12]

Work on local-to-unit root processes has been used to provide a more accurate approximation to the actual finite distribution of $t$-statistics (see Elliot and Stock 1994). Using this framework, Torous, Valkanov, and Yan (2005) find evidence of short-horizon predictability, but not, surprisingly given the extant literature, at the long horizon. Lewellen (2004), on the other hand, finds evidence of predictability at long horizons using financial ratios. Valkanov (2003) reports evidence that predictability using financial variables is reduced when corrections to $t$-statistics are made to account for the persistence in the predictor

[12] We calculated the Campbell and Yogo (2005) pretest regarding the applicability of asymptotic $t$-statistics in predictability regressions. We found that *gap* passed this test and hence the $t$-statistics should be fine asymptotically.

variables and the correlation between returns and the residuals from a regression of the predictor on its own lag. Ang and Bekaert (2007) show that there are substantial size distortions with the Newey-West $t$-statistics when forecasting stock returns using the dividend yield, a highly persistent regressor. They show that the empirical size of the Newey-West $t$-statistic is 23% against a nominal 5% and hence there is an obvious tendency to overreject the null of no predictability.

We now turn to addressing this issue when predicting stock returns with *gap*. We perform a Monte Carlo experiment to investigate whether inferences on the statistical significance of the parameter estimates are affected by size distortions when using Newey-West $t$-statistics. The data for the Monte Carlo experiment are generated under the null hypothesis of no predictability:

$$r_t = \gamma_0 + \upsilon_t, \tag{14}$$

where $\gamma_0$ is a constant and $\upsilon$ is drawn from a normal distribution.

To complete the data generation process, we need to specify a data generating equation for *gap*. Based on the properties of the data, we specify the following:

$$\text{gap}_t = \gamma_1 + \rho_1 \text{gap}_{t-1} + \eta_t, \tag{15}$$

where $\eta$ is drawn from a normal distribution. The values we use for $\rho_1$ and $\gamma_1$ are those estimated from an AR(1) regression using the actual data for *gap*.

We generate 100,000 samples with $100 + T$ observations, where $T$ is the sample size for the relevant regression. We then discard the first 100 observations and estimate Equation (4) 100,000 times with the remaining $T$ observations. This gives us the distribution of the $t$-statistics testing the null hypothesis that $\gamma = 0$ in (4), along with the distribution of the $\overline{R}^2$. To assess whether there are any size distortions with the Newey-West $t$-statistics, we compare its empirical size generated from the Monte Carlo experiment against a 5% nominal size. The empirical size is the percentage of times the relevant null hypothesis is rejected at the 5% level of significance. If the empirical size of the $t$-statistics is greater than 5%, the Newey-West $t$-statistics have a tendency to overreject the null hypotheses, that is, they find predictability when it is not likely to be there.

Table 8 reports the results of the Monte Carlo experiments for the Newey-West $t$-statistics using *gap* in all G7 countries. We report the empirical size of the test, the $t$-statistic, and the 95% confidence interval of the $\overline{R}^2$ for each forecasting horizon. We find that the Newey-West $t$-statistic testing the null that $\gamma = 0$, that is, that *gap* cannot predict returns, has good size properties for the one-month-ahead forecasting regressions (in the United States it is 5.12%, as opposed to the nominal 5% value, and it is never higher than 5.6% when considering the other six countries). However, this steadily disappears for the remaining regressions that use a greater number of months. In these cases, the size properties deteriorate with the forecast horizon, being 10% at the one-year horizon in the United States and as high as 13.5% in Japan. The

**Table 8**
**Newey-West size properties, *t*-statistics, and $\overline{R}^2$ with simulated unpredictable returns**

| | 1 month | 3 months | 12 months | 24 months | 36 months | 48 months | 60 months |
|---|---|---|---|---|---|---|---|
| | | | | U.S. | | | |
| Size | 5.12 | 8.91 | 10.98 | 14.88 | 17.18 | 19.81 | 22.66 |
| $t$ | –1.89 (–1.56) | –2.10 (–1.76) | –2.32 (–1.90) | –2.61 (–2.09) | –2.84 (–2.24) | –3.04 (–2.47) | –3.31 (–2.62) |
| $\overline{R}^2$ | –0.00, 0.01 | –0.00, 0.03 | –0.00, 0.08 | –0.00, 0.14 | –0.00, 0.19 | –0.00, 0.23 | –0.00, 0.27 |
| | | | | Canada | | | |
| Size | 5.09 | 9.04 | 12.51 | 18.04 | 21.32 | 24.60 | 27.71 |
| $t$ | –1.93 (–1.62) | –2.32 (–1.90) | –2.50 (–2.05) | –3.00 (–2.48) | –3.35 (–2.67) | –3.61 (–2.90) | –4.15 (–3.19) |
| $\overline{R}^2$ | –0.00, 0.01 | –0.00, 0.04 | –0.00, 0.11 | –0.00, 0.19 | –0.00, 0.26 | –0.00, 0.29 | –0.00, 0.32 |
| | | | | France | | | |
| Size | 4.99 | 9.65 | 12.70 | 16.00 | 20.08 | 22.14 | 25.07 |
| $t$ | –1.90 (–1.60) | –2.26 (–1.89) | –2.58 (–2.09) | –2.83 (–2.26) | –3.09 (–2.48) | –3.46 (–2.74) | –3.44 (–2.82) |
| $\overline{R}^2$ | –0.00, 0.01 | –0.00, 0.04 | –0.00, 0.11 | –0.00, 0.16 | –0.00, 0.20 | –0.00, 0.23 | –0.00, 0.25 |
| | | | | Germany | | | |
| Size | 5.31 | 9.67 | 11.42 | 14.24 | 17.48 | 19.40 | 22.73 |
| $t$ | –1.94 (–1.64) | –2.28 (–1.89) | –2.46 (–1.99) | –2.63 (–2.10) | –2.92 (–2.31) | –2.95 (–2.38) | –3.17 (–2.53) |
| $\overline{R}^2$ | –0.00, 0.00 | –0.00, 0.04 | –0.00, 0.09 | –0.00, 0.13 | –0.00, 0.16 | –0.00, 0.17 | –0.00, 0.19 |
| | | | | Italy | | | |
| Size | 5.51 | 8.86 | 11.06 | 12.97 | 14.90 | 16.05 | 18.66 |
| $t$ | –2.02 (–1.68) | –2.24 (–1.85) | –2.43 (–1.99) | –2.60 (–2.13) | –2.74 (–2.23) | –2.78 (–2.26) | –3.06 (–2.46) |
| $\overline{R}^2$ | –0.00, 0.01 | –0.00, 0.03 | –0.00, 0.07 | –0.00, 0.09 | –0.00, 0.10 | –0.00, 0.010 | 0.002, 0.10 |
| | | | | Japan | | | |
| Size | 5.36 | 9.64 | 13.50 | 17.67 | 21.50 | 26.08 | 29.43 |
| $t$ | –1.94 (–1.60) | –2.24 (–1.89) | –2.56 (–2.07) | –2.94 (–2.36) | –3.40 (–2.66) | –3.74 (–2.91) | –4.06 (–3.11) |
| $\overline{R}^2$ | –0.00, 0.01 | –0.00, 0.04 | –0.00, 0.12 | –0.00, 0.20 | –0.00, 0.26 | –0.00, 0.30 | –0.00, 0.34 |
| | | | | U.K. | | | |
| Size | 5.60 | 9.69 | 12.30 | 15.80 | 18.89 | 22.06 | 24.19 |
| $t$ | –1.95 (–1.64) | –2.22 (–1.84) | –2.48 (–2.06) | –2.73 (–2.22) | –3.00 (–2.43) | –3.39 (–2.66) | –3.55 (–2.78) |
| $\overline{R}^2$ | –0.00, 0.01 | –0.00, 0.04 | –0.00, 0.10 | –0.00, 0.16 | –0.00, 0.20 | –0.00, 0.24 | –0.00, 0.25 |

The table reports the results of a Monte Carlo experiment to investigate the empirical properties of the Newey-West *t*-statistics. The data are generated under the null hypothesis of no predictability. The parameters in the data generation process are their empirical counterparts. We use the moments of the excess returns on each country's excess stock return to simulate the unpredictable returns. The row "Size" reports the percentage of times $H_0 : \gamma = 0$ is rejected against a nominal significance level of 5%. The row "*t*" reports the Monte Carlo-generated 5% *t*-statistics testing $H_0 : \gamma = 0$ against $H_1 : \gamma \neq 0$. The *t*-statistics in parenthesis report one-sided critical values. The row "$\overline{R}^2$" is the value of the 95% confidence interval for the Monte Carlo-generated $\overline{R}^2$. *gap_q* is used in each country and is measured as the deviation of the log of total industrial production from a quadratic and linear trend.

empirical size is approximately 23% for the five-year forecasting horizon in the United States and peaks at 29% in Japan. This is consistent with the findings in Ang and Bekaert (2007) that Newey-West $t$-statistics are unreliable in long-horizon regressions and suggests that predictability is found more often than it is actually there.

To gauge how important the size distortions are, the next row of the Table reports the Monte Carlo-generated critical values for the $t$-statistic testing $\gamma = 0$. If the usual asymptotic critical values for the $t$-statistics were being used, they would be 1.96 at all horizons if we were to use a two-sided test. However, we might argue that a one-sided test is appropriate given that we would expect the coefficient on *gap* only to be negative. The empirical critical values for the $t$-statistics, at horizons of three months and greater, are above the asymptotic values and for the United States rise to 3.31 (2.62) for the five-year horizon using a two-(one-)sided test. The $t$-statistic is as high as 4.15 (3.19) in Canada at the five-year horizon. The next row of the table provides the lower and upper value of the 95% confidence intervals for the Monte Carlo-generated $\overline{R}^2$. These indicate that, for U.S. returns at the five-year horizon, an $\overline{R}^2$ of 27% is not outside the realms of possibility, even under the null of no predictability.

Taken together, the results show that for horizons greater than one month, inference in predictive regressions using Newey-West $t$-statistics and statements about predictive power based on the $\overline{R}^2$ can be hazardous if regard to the small-sample properties is ignored. In our case, however, the findings in Table 8 do not alter our conclusions about predictability reached on the basis of the results presented earlier: the forecasting model based on *gap* is still capable of forecasting stock returns at business cycle horizons even when using the empirically derived $t$-statistics presented in Table 8. For example, in the United States, the $t$-statistics are around three to four at the business cycle frequencies compared to the simulations that produce two-sided empirical $t$-statistics of 1.89, 2.10, and 2.32 at the one-, three-, and 12-month horizons.

## 4. Bond Returns

In this section of the paper, we ask whether *gap* can predict the excess returns on government bonds. It is well known that the expectation hypothesis is rejected (see, for example, Fama and Bliss 1987; Campbell and Shiller 1988; Cochrane and Piazzesi 2005) in favor of bond returns being predictable by forward or yield spreads. This literature finds evidence of a time-varying risk premia in bond returns, which is, as Cochrane and Piazzesi (2005) note, "suggestively correlated with the business cycle"; however, evidence based on forecasting with term spreads and forward spreads "does not tie the time-varying premia to macroeconomic or monetary fundamentals." Ludvigson and Ng (2006) note that there is little evidence linking the macroeconomy and bond risk premia. These authors use factor analysis to show that real and inflation factors, along with financial factors, are important in forecasting government bond returns and document that the yield curve has a strong countercyclical component. However, because they use factor analysis of a large number of macroeconomic variables, they do not identify a specific business cycle variable as being the source of the predictability.

It is interesting to assess whether *gap*, a specific business cycle variable, can predict bond returns. If it can, this would add further economic support to the notion that *gap* is an important predictor of risk premia, and would provide a direct link between the macroeconomy and bond risk premia. To ascertain this, we regress each excess bond return on *gap*. We also regress the excess bond return on Cochrane and Piazzesi's (2005) forward rate predictor variable ($CP$) along with *gap*. However, because these two predictor variables are highly correlated (the correlation coefficient is −0.46 and a regression of $CP$ on *gap*

**Table 9**
**Bond excess returns**

| Two-year bond | | | Three-year bond | | | Four-year bond | | | Five-year bond | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| *gap* | $CP$ | $\overline{R}^2$ | *gap* | $CP$ | $\overline{R}^2$ | *gap* | $CP$ | $\overline{R}^2$ | *gap* | $CP$ | $\overline{R}^2$ |
| | | | | | Quadratic gap: $gap\_q$ | | | | | | |
| −0.044 | | 0.02 | −0.077 | | 0.02 | −0.106 | | 0.02 | −0.127 | | 0.02 |
| (3.24) | | | (3.09) | | | (3.12) | | | (3.03) | | |
| −0.041 | 0.336 | 0.20 | −0.071 | 0.661 | 0.23 | 0.097 | 0.978 | 0.26 | −0.116 | 1.111 | 0.23 |
| (3.31) | (10.74) | | (3.17) | (11.90) | | (3.23) | (12.54) | | (3.06) | (11.07) | |
| | | | | | Breaking gap: $gap\_b$ | | | | | | |
| −0.038 | | 0.01 | −0.060 | | 0.01 | −0.083 | | 0.01 | −0.099 | | 0.01 |
| (2.76) | | | (2.43) | | | (2.43) | | | (2.39) | | |
| −0.034 | 0.346 | 0.20 | −0.051 | 0.685 | 0.24 | 0.083 | 1.024 | 0.28 | −0.085 | 1.150 | 0.23 |
| (2.63) | (10.98) | | (2.26) | (12.27) | | (2.43) | (13.714) | | (2.18) | (11.42) | |
| | | | | | Linear gap: $gap\_l$ | | | | | | |
| −0.057 | | 0.04 | −0.105 | | 0.04 | −0.143 | | 0.04 | −0.174 | | 0.04 |
| (4.82) | | | (4.99) | | | (4.90) | | | (4.90) | | |
| −0.052 | 0.307 | 0.21 | −0.096 | 0.596 | 0.24 | 0.129 | 0.882 | 0.27 | −0.159 | 0.999 | 0.23 |
| (5.01) | (10.00) | | (5.30) | (11.90) | | (5.27) | (11.50) | | (5.17) | (10.13) | |

This table reports estimates from OLS regressions of current bond excess returns on the thirteenth lag of measures of the output gap and the twelfth lag of Cochrane and Piazzesi's (2005) factor, $CP$, which is orthogonalized relative to *gap*. *gap* is measured as follows: $gap\_q$ is the deviation of the logarithm of total industrial production from a trend that includes both a linear component and a quadratic component. $gap\_b$ is the deviation of the logarithm of industrial production from a trend that includes a linear component that contains a break in the linear trend in 1977. $gap\_l$ is the deviation of the logarithm of industrial production from a trend that includes a linear component. We report results using the excess returns constructed from Fama and Bliss (1987) data from CRSP. We calculate annual excess bond returns at a monthly frequency over the sample 1952:6 to 2003:12. We obtain the annual return in a given month by borrowing at the one-year rate and buying either a two-, three-, four-, or five-year bond and then selling it after one year. $\overline{R}^2$ is the adjusted $R^2$. The Newey-West-corrected $t$-statistics are reported in parentheses.

produces an $\overline{R}^2$ of 22%), we first orthogonalize them by regressing $CP$ on *gap*. We then regress the excess bond returns on *gap* and the orthogonalized component of $CP$.

Table 9 shows that, on its own, *gap* has predictive power for excess bond returns. The coefficients on *gap* are highly statistically significant for all bonds irrespective of their maturity and the $\overline{R}^2$ ranges form 1% to 4% depending on which measure of *gap* is used. Interestingly, the next row of the table shows that both *gap* and orthogonalized $CP$ are statistically significant. This suggests that *gap* is capturing risk that is independent of the financial market-based variable $CP$. The predictability of bond returns with a direct macroeconomic variable is a novel finding and suggests that affine term structure models that attribute all bond return predictability to yields or forward rates are unlikely to fully describe bond price movements.

Table 10 reports results of out-of-sample forecasting of excess bond returns using *gap* and orthogonal $CP$. We use the same approach as in testing stock-return predictability out-of-sample. We start with an initial estimation period of 1953:6 to 1964:12. Out-of-sample forecasts are then formed in 1965:1. This process is repeated recursively until the end of the sample. We repeat this process with the following initial estimation periods: 1953:6 to 1974:12, 1953:6

**Table 10**
**Predicting bond excess returns out-of-sample**

| | ENC-NEW | | MSE-$F$ | | $R^2_{oos}$ | U Gain |
|---|---|---|---|---|---|---|
| | Statistic | Boot p.v. | Statistic | Boot p.v. | | |
| | Panel A: Quadratic gap , $gap\_q$ | | | | | |
| | Forecasting from 1965 m1 | | | | | |
| 2 Year | 83.449 | 0.00 | 114.659 | 0.00 | 0.196 | 4.35 |
| 3 Year | 103.811 | 0.00 | 142.222 | 0.00 | 0.224 | 4.66 |
| 4 Year | 121.045 | 0.00 | 169.816 | 0.00 | 0.260 | 5.38 |
| 5 Year | 94.129 | 0.00 | 134.210 | 0.00 | 0.221 | 5.08 |
| | Forecasting from 1975 m1 | | | | | |
| 2 Year | 81.235 | 0.00 | 115.731 | 0.00 | 0.247 | 4.70 |
| 3 Year | 93.444 | 0.00 | 131.584 | 0.00 | 0.273 | 5.12 |
| 4 Year | 109.566 | 0.00 | 156.346 | 0.00 | 0.313 | 6.07 |
| 5 Year | 87.768 | 0.00 | 128.881 | 0.00 | 0.270 | 6.11 |
| | Forecasting from 1985 m1 | | | | | |
| 2 Year | 66.996 | 0.00 | 96.101 | 0.00 | 0.291 | 1.77 |
| 3 Year | 83.090 | 0.00 | 120.652 | 0.00 | 0.348 | 1.83 |
| 4 Year | 99.069 | 0.00 | 146.060 | 0.00 | 0.386 | 3.56 |
| 5 Year | 80.876 | 0.00 | 123.080 | 0.00 | 0.348 | 4.85 |
| | Forecasting from 1995 m1 | | | | | |
| 2 Year | 11.363 | 0.00 | 10.767 | 0.00 | 0.083 | −2.48 |
| 3 Year | 17.754 | 0.00 | 23.647 | 0.00 | 0.176 | −3.29 |
| 4 Year | 25.128 | 0.00 | 35.762 | 0.00 | 0.241 | −1.98 |
| 5 Year | 22.181 | 0.00 | 32.930 | 0.00 | 0.237 | −0.15 |
| | Panel B: Breaking trend | | | | | |
| | Forecasting from 1965 m1 | | | | | |
| 2 Year | 104.872 | 0.00 | 104.909 | 0.00 | 0.238 | 4.77 |
| 3 Year | 128.762 | 0.00 | 174.178 | 0.00 | 0.262 | 5.11 |
| 4 Year | 152.949 | 0.00 | 207.554 | 0.00 | 0.301 | 5.75 |
| 5 Year | 122.297 | 0.00 | 170.503 | 0.00 | 0.266 | 5.30 |
| | Forecasting from 1975 m1 | | | | | |
| 2 Year | 98.146 | 0.00 | 137.769 | 0.00 | 0.286 | 4.95 |
| 3 Year | 113.171 | 0.00 | 154.422 | 0.00 | 0.303 | 5.21 |
| 4 Year | 134.329 | 0.00 | 183.867 | 0.00 | 0.343 | 6.19 |
| 5 Year | 109.776 | 0.00 | 155.343 | 0.00 | 0.305 | 6.20 |
| | Forecasting from 1985 m1 | | | | | |
| 2 Year | 93.305 | 0.00 | 129.147 | 0.00 | 0.364 | 2.12 |
| 3 Year | 108.033 | 0.00 | 149.204 | 0.00 | 0.389 | 2.32 |
| 4 Year | 125.799 | 0.00 | 175.511 | 0.00 | 0.431 | 4.02 |
| 5 Year | 105.991 | 0.00 | 152.919 | 0.00 | 0.397 | 5.53 |
| | Forecasting from 1995 m1 | | | | | |
| 2 Year | 19.114 | 0.00 | 24.085 | 0.00 | 0.145 | −1.86 |
| 3 Year | 22.836 | 0.00 | 31.635 | 0.00 | 0.224 | −2.44 |
| 4 Year | 29.170 | 0.00 | 41.381 | 0.00 | 0.277 | −1.34 |
| 5 Year | 26.515 | 0.00 | 39.201 | 0.00 | 0.259 | 0.73 |
| | Forecasting from 1965 m1 | | | | | |
| 2 Year | 83.637 | 0.00 | 111.258 | 0.00 | 0.193 | 5.30 |
| 3 Year | 98.277 | 0.00 | 134.029 | 0.00 | 0.221 | 5.32 |
| 4 Year | 113.192 | 0.00 | 158.301 | 0.00 | 0.274 | 5.69 |
| 5 Year | 88.630 | 0.00 | 125.588 | 0.00 | 0.211 | 5.36 |
| | Forecasting from 1975 m1 | | | | | |
| 2 Year | 81.772 | 0.00 | 113.006 | 0.00 | 0.240 | 5.59 |
| 3 Year | 89.063 | 0.00 | 123.927 | 0.00 | 0.263 | 5.67 |
| 4 Year | 102.075 | 0.00 | 145.408 | 0.00 | 0.293 | 6.26 |
| 5 Year | 82.571 | 0.00 | 120.670 | 0.00 | 0.253 | 6.25 |

*continued overleaf*

**Table 10**
**(Continued)**

| | ENC-NEW | | MSE-$F$ | | $R^2_{oos}$ | U Gain |
|---|---|---|---|---|---|---|
| | Statistic | Boot p.v. | Statistic | Boot p.v. | | |
| | | | Panel C. Linear gap, *gap_l* | | | |
| | | | Forecasting from 1985 m1 | | | |
| 2 Year | 75.677 | 0.00 | 105.224 | 0.00 | 0.311 | 3.50 |
| 3 Year | 83.348 | 0.00 | 118.943 | 0.00 | 0.332 | 3.50 |
| 4 Year | 93.339 | 0.00 | 136.603 | 0.00 | 0.374 | 5.06 |
| 5 Year | 78.518 | 0.00 | 118.487 | 0.00 | 0.340 | 6.27 |
| | | | Forecasting from 1995 m1 | | | |
| 2 Year | 10.830 | 0.00 | 9.829 | 0.00 | 0.082 | −0.04 |
| 3 Year | 14.610 | 0.00 | 18.319 | 0.00 | 0.141 | −1.02 |
| 4 Year | 19.092 | 0.00 | 25.629 | 0.00 | 0.193 | −0.37 |
| 5 Year | 17.802 | 0.00 | 25.386 | 0.00 | 0.186 | 1.11 |

This table reports the results of out-of-sample forecast comparisons of the annual excess bond returns at a monthly frequency over the sample 1953:6 to 2003:12. We obtain the annual return in a given month by borrowing at the one-year rate and buying either a two-, three-, four-, or five-year bond and then selling it after one year. The comparisons are of forecasts of excess returns based on a constant (the restricted model) and forecasts based on a constant and *gap* (the unrestricted model). We report comparisons based on forecasting one month ahead using one-, three-, and 12-month returns. The column labeled "ENC-NEW" provides the Clark and McCracken (2001) encompassing test statistic. The column labeled "MSE-$F$" gives the $F$-test of McCracken (2007) that tests the null hypothesis of equal MSEs against the alternative that the MSE from the unrestricted model is smaller. The column labeled "Statistic" provides the test statistics and the column labeled Boot p.v. provides the bootstrapped probability value using the methodology in Clark and McCracken (2005). $R^2_{\mathrm{oos}}$ is the out-of-sample $R^2$. The column headed "U Gain" reports the utility gain (in annualized percentage points) an investor would have obtained from using the predictability results as compared to using the historical mean excess returns when forming portfolio weights. Panel A reports results using *gap_q*, which is measured as the deviation of the log of total industrial production from a quadratic and linear trend. Panel B reports results using *gap_b*, which is the deviation of the logarithm of industrial production from a trend that includes a linear component that contains a break in the linear trend in 1977. Panel C reports results using *gap_l*, which is the deviation of the logarithm of industrial production from a trend that includes a linear component.

to 1984:12, and 1953:6 to 1994:12. According to the ENC-NEW test, whatever the initial estimation period, we always reject at the 1% level the hypothesis that the forecasts from the constant expected return model encompass those from the time-varying expected return model. Similarly, the MSE-$F$ test always finds strong evidence in favor of the time-varying expected return model. The out-of-sample $R^2$s are always positive, and the utility gains are generally large.

In summary, we find evidence that excess bond returns are predictable both in-sample and out-of-sample. The predictability adds support to our earlier findings using both U.S. and international stock returns that *gap* has an important role to play in describing the time variation of expected returns and risk premia.

## 5. Conclusion

This paper provides evidence that a prime business cycle variable, namely the output gap, is a strong in-sample predictor of stock returns. In addition, the output gap predicts international stock returns and U.S. excess bond returns. These results are particularly important as other macroeconomic business cycle variables have proven dismal as predictive variables and even financial market-based variables have been found to perform poorly in the past three decades. As the output gap does not contain the level of stock market prices, its predictive ability is unlikely to stem from stock mispricing. Our findings lend support to the notion that stock and bond return predictability is the rational response to changing business conditions rather than market inefficiency. This is consistent with the view that the variation in expected returns reflects the rational response of agents to time-varying investment opportunities or time-varying risk aversion.

The predictive ability of *gap* is robust to the inclusion of a set of financial market-based predictor variables and the net payout ratio, subsample analysis, and how *gap* is measured. We perform a battery of out-of-sample predictive tests based on forecast encompassing, comparisons of MSE, out-of-sample $R^2$s, and utility gains from forming optimal portfolios based on *gap*. The results show that *gap* provides out-of-sample forecasts better than the historical mean, even in the post-1975 period during which most predictor variables perform poorly out-of-sample.

We undertake a Monte Carlo experiment under the null hypothesis of unpredictable returns to assess the bias in Newey-West $t$-statistics that arises when predictor variables have a near-unit root. While the $t$-statistics have good size properties when forecasting with *gap* at the one-month horizon, there is substantial bias in the $t$-statistics at longer horizons in that they overreject the null hypothesis of no predictability. Despite the bias in the $t$-statistics, *gap* still has predictive power especially at business cycle frequencies when we compare the actual $t$-statistics to the empirically derived $t$-statistics.

The findings fill an important void in the literature on return predictability. Namely, we have identified a channel where an economically fundamental, production-based business cycle variable can track variations in expected returns. This constitutes important independent evidence regarding the predictability of returns that does not use price or consumption data.

**References**

Ang, A., and G. Bekaert. 2007. Stock Return Predictability: Is it There? *Review of Financial Studies* 20:651–707.

Baker, M., and J. Wurgler. 2006. Investor Sentiment and the Cross-Section of Stock Returns. *Journal of Finance* 61:1645–80.

Baker, M., J. Stein, and J. Wurgler. 2003. When Does the Market Matter? Stock Prices and the Investment of Equity-Dependent Firms. *Quarterly Journal of Economics* 118:969–1006.

Bossaerts, P., and P. Hillion. 1999. Implementing Statistical Criteria to Select Return Forecasting Models: What Do We Learn? *Review of Financial Studies* 12:405–28.

Boudoukh, J., R. Michaely, M. Richardson, and M. G. Roberts. 2007. On the Importance of Measuring Payout Yield: Implications for Empirical Asset Pricing. *Journal of Finance* 62:877–915.

Boudoukh, J., M. Richardson, and R. F. Whitelaw. 2005. The Myth of Long-Horizon Predictability. *Review of Financial Studies*, forthcoming.

Campbell, J. Y. 1987. Stock Returns and the Term Structure. *Journal of Financial Economics* 18:373–99.

Campbell, J. Y., and R. Shiller. 1988. The Dividend Price Ratio and Expectations of Future Dividends and Discount Factors. *Review of Financial Studies* 1:195–227.

Campbell, J. Y., and M. Yogo. 2006. Efficient Tests of Stock Return Predictability. *Journal of Financial Economics* 81:27–60.

Campbell, J. Y., and S. B. Thompson. 2007. Predicting the Equity Premium Out of Sample: Can Anything Beat the Historical Average? *Review of Financial Studies*, forthcoming.

Clarida, R., J. Gali, and M. Gertler. 2000. Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory. *Quarterly Journal of Economics* 115:147–80.

Clark, P. K. 1978. Capital Formation and the Recent Productivity Slowdown. *Journal of Finance* 33:965–75.

Clark, T. E., and M. W. McCracken. 2001. Tests of Forecast Accuracy and Encompassing for Nested Models. *Journal of Econometrics* 105:85–110.

Clark, T. E., and M. W. McCracken. 2005. Evaluating Direct Multi-Step Forecasts. *Econometric Reviews* 24:369–404.

Cochrane, J. H. 1999. New Facts in Finance. Economic Perspectives. *Federal Reserve Bank of Chicago* 23:36–58.

Cochrane, J. H. 2005. Financial Markets and the Real Economy, in R. Roll (ed.), *The International Library of Critical Writings in Financial Economics*. Edward Elgar Series, forthcoming.

Cochrane, J. H. 2006. The Dog That Did Not Bark: A Defense of Return Predictability. *Review of Financial Studies*, forthcoming.

Cochrane, J. H., and M. Piazzesi. 2005. Bond Risk Premia. *American Economic Review* 95:138–60.

Congressional Budget Office. 2001. CBO's Method for Estimating Potential Output: An Update. Congressional Budget Office Macroeconomic Analysis Publications. Washington, DC.

Council of Economic Advisers. 1977. Economic Report of the President. Washington, DC: U.S. Printing Office.

Elliot, G., and J. H. Stock. 1994. Inference in Time Series Regressions when the Order of Integration of a Regressor Is Unknown. *Econometric Theory* 10:672–700.

Fama, E. F., and R. R. Bliss. 1987. The Information in Long-Maturity Forward Rates. *American Economic Review* 77:680–92.

Fama, E. F., and K. R. French. 1988. Dividend Yields and Expected Stock Returns. *Journal of Financial Economics* 22:3–27.

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

Fuhrer, C. J., and G. D. Rudebusch. 2004. Estimating the Euler Equation for Output. *Journal of Monetary Economics* 51:1133–53.

Goyal, A., and I. Welch. 2003. Predicting the Equity Premium with Dividend Ratios. *Management Science* 49:639–54.

Goyal, A., and I. Welch. 2006. A Comprehensive Look at the Empirical Performance of Equity Premium Prediction. *Review of Financial Studies*, forthcoming.

Harvey, D., S. Leybourne, and P. Newbold. 1998. Tests for Forecast Encompassing. *Journal of Business and Economic Statistics* 16:254–59.

Hjalmarsson, E. 2006. Should We Expect Significant Out-of-Sample Results When Predicting Stock Returns? International Finance Discussion Papers 855, Board of Governors of the Federal Reserve System. Washington, DC.

Inoue, A., and L. Kilian. 2004. In-Sample or Out-of-Sample Tests of Predictability: Which One Should We Use? *Econometric Reviews* 23:371–402.

Julliard, C. 2006. Labor Income and Asset Returns. Working Paper, London School of Economics.

Lettau, M., and S. Ludvigson. 2001. Consumption, Aggregate Wealth, and Expected Stock Returns. *Journal of Finance* 56:815–49.

Lettau, M., and S. Ludvigson. 2005. Expected Returns and Expected Dividend Growth. *Journal of Financial Economics* 76:583–626.

Lewellen, J. 2004. Predicting Returns with Financial Ratios. *Journal of Financial Economics* 74:209–35.

Ludvigson, S., and S. Ng. 2006. Macro Factors in Bond Risk Premia. Working Paper, NYU.

Lustig, H. N., and S. G. Van Nieuwerburgh. 2005. Housing Collateral, Consumption Insurance, and Risk Premia: An Empirical Perspective. *Journal of Finance* 60:1167–219.

Mankiw, N. G., and M. D. Shapiro. 1986. Do We Reject Too Often? Small Sample Properties of Tests of Rational Expectations Models. *Economics Letters* 20:139–45.

McCracken, M. W. 2007. Asymptotics for Out-of-Sample Tests of Granger Causality. *Journal of Econometrics* 140:719–52.

Pena, I. J., F. Restoy, and R. Rodriguez. 2002. Can Output Explain the Predictability and Volatility of Stock Returns? *Journal of International Money and Finance* 21:163–82.

Perron, P. 1989. The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis. *Econometrica* 57:1361–401.

Rangvid, J. 2006. Output and Expected Returns. *Journal of Financial Economics* 81:595–624.

Robertson, D., and S. Wright. 2006. Dividends, Total Cash Flow to Shareholders, and Predictive Return Regressions. *Review of Economics and Statistics* 88:1–9.

Stambaugh, R. F. 1999. Predictive Regressions. *Journal of Financial Economics* 54:375–421.

Taylor, J. B. 1993. Discretion versus Policy Rules in Practice. *Carnegie-Rochester Conference Series on Public Policy* 39:195–214.

Torous, W., R. Valkanov, and S. Yan. 2005. On Predicting Stock Returns with Nearly Integrated Variables. *Journal of Business* 77:937–66.

Valkanov, R. 2003. Long-Horizon Regressions: Theoretical Results and Applications. *Journal of Financial Economics* 68:201–32.