

Data Appendix for “In-Sample vs. Out-of-Sample Tests of Stock Return Predictability in the Context of Data Mining”

20th Century Annual Data

- *S&P 500 log real returns.* Real returns are calculated using the S&P 500 nominal stock price index for January of each year and the annual S&P 500 nominal dividend index available from Robert Shiller’s web page at <http://www.econ.yale.edu/~shiller>. The nominal return for year t is calculated as the sum of the nominal stock price in January of year $t + 1$ and dividends for year t divided by the nominal stock price in January of year t . The real return is the nominal return divided by the ratio of the consumer price index in January of year $t + 1$ to the consumer price index in January of year t . We compute the natural logarithm and multiply by 100.
- *CRSP equal-weighted log real returns.* The real return for year t is the Center for Research on Security Prices (CRSP) equal-weighted annual nominal return index with dividends for year t divided by the ratio of the consumer price index in January of year $t + 1$ to the consumer price index in January of year t . We compute the natural logarithm and multiply by 100.
- *Dividend-price ratio, log-level.* We calculate the dividend-price ratio corresponding to S&P 500 and CRSP equal-weighted returns. The dividend-price ratio corresponding to S&P 500 returns is based on the Shiller data. The dividend-price ratio in year t is the ratio of nominal dividends during year t to the nominal stock price in January of year $t + 1$. We construct a dividend-price ratio corresponding to CRSP equal-weighted returns by first backing out nominal end-of-year stock price and annual nominal dividend indexes from the annual CRSP equal-weighted nominal returns with and without dividends. The dividend-price ratio

- in year t is the ratio of the dividends index for year t to the nominal stock price index for the end of year t . For both annual dividend-price ratios, we compute the natural logarithm.
- *Price-earnings ratio, log-level.* The price-earnings ratio in year t is the ratio of the S&P 500 nominal stock price in January of year $t + 1$ to the annual S&P 500 nominal earnings index in year t . The annual S&P 500 nominal earnings index is available from Robert Shiller's web page. We compute the natural logarithm.
 - *Book-to-market ratio, log-level.* This is the ratio of the annual book value of the Dow Jones Industrial Average (DJIA) year t to the DJIA index in year t . The data are from the Value Line publication, "A Long-Term Perspective." We compute the natural logarithm.
 - *Fed q, log-level.* This is the ratio of the total market value of equities outstanding to corporate net worth. Robertson and Wright (2000) obtain this data from the Federal Reserve. This ratio is available from Stephen Wright's web page at <http://www.econ.bbk.ec.uk/faculty/wright> and is used in Robertson and Wright (2000). They note that the Federal Reserve does not refer to this series as "Fed q," and, while it does not correspond exactly with Tobin's q, the statistical properties of the two series are very similar. More specifically, the definition for Tobin's q used in most academic studies is $(S \cdot E + L) / K$, while the Fed q is $S \cdot E / (K - L)$, where S is the price of stocks, E is the number of shares, L is the market value of liabilities, and K is corporate assets. Robertson and Wright (2000) note that the two definitions have the same properties in terms of the sign of the derivative with respect to each of the components. We compute the natural logarithm.
 - *Payout ratio, log-level.* This is the ratio of S&P 500 nominal dividends to S&P 500 nominal earnings. We compute the natural logarithm.
 - *Term spread.* This is the difference between long-term and short-term government bond yields. The long-term government bond yield data are from Ibbotson Associates (2001, Table A9) and are based on a maturity of approximately 20 years. The short-term yield is the

3-month Treasury bill rate (secondary market) and is from the Federal Reserve Economic Database available from the Federal Reserve Bank of St. Louis web page at <http://www.stls.frb.org/fred/>. The original Treasury bill rates are monthly. We obtain annual observations by averaging over the twelve months comprising a calendar year. As the Treasury bill rate data begin in 1934, we use the short-term Treasury certificate yield data from Homer (1963, Table 49) for the 1927-1933 period.

- *Default spread.* This is the difference between the Moody's seasoned Baa corporate bond yield and the Moody's seasoned Aaa corporate bond yield. The corporate bond yield data are also from FRED. The original corporate bond yield data are monthly. We obtain annual observations by averaging over the twelve months comprising a calendar year.
- *Short-term interest rate.* This is the short-term government bond yield described above.
- *Equity share.* This is the annual total of equity issues (common and preferred) divided by the sum of the annual total of equity issues and the annual total of long-term debt issues (public and private). The data are from Baker and Wurgler (2000), which they compile from the *Federal Reserve Bulletin*. The data are available from Jeffrey Wurgler's web page at <http://www.stern.nyu.edu/~jwurgler/>.

Postwar Quarterly Data

- *S&P 500 log real returns.* The nominal return for quarter t is the sum of the nominal stock price at the end of quarter t and the flow of dividends over quarter t and the previous three quarters (divided by four) divided by the nominal stock price at the end of the quarter $t - 1$. The real return is the nominal return divided by the ratio of the consumer price index in the first month of quarter $t + 1$ to the consumer price index in the first month of quarter t . We compute the natural logarithm and multiply by 400.

- *CRSP equal-weighted log real returns.* The real return for quarter t is the CRSP equal-weighted annual nominal return index with dividends for quarter t divided by the ratio of the consumer price index in the first month of quarter $t + 1$ to the consumer price index in the first month of quarter t . We compute the natural logarithm and multiply by 400.
- *Dividend-price ratio, log-level.* For the S&P 500 index, the dividend-price ratio in quarter t is the flow of nominal dividends during quarter t and the three previous quarters to the nominal stock price at the end of quarter t . We construct a dividend-price ratio corresponding to CRSP equal-weighted returns by first backing out nominal end-of-quarter stock price and nominal dividend indexes from quarterly CRSP equal-weighted nominal returns with and without dividends. The dividend-price ratio in quarter t is the ratio of the dividends index for quarter t and the previous three quarters to the nominal stock price index for the end of quarter t . For both quarterly dividend-price ratios, we compute the natural logarithm.
- *Price-earnings ratio, log-level.* The price-earnings ratio in quarter t is the ratio of the S&P 500 nominal stock price index at the end of quarter t to the flow of nominal earnings over quarter t and the previous three quarters. We compute the natural logarithm.
- *Fed q , log-level.* This is the ratio of the total market value of equities outstanding to corporate net worth. The quarterly data are available from <http://www.valuingwallstreet.com/updates.shtml>. We compute the natural logarithm.
- *Payout ratio, log-level.* This is the ratio of the S&P 500 nominal dividends index to the S&P 500 nominal earnings index. We compute the natural logarithm.
- *Consumption-wealth ratio.* This is calculated as $cay = c - 0.2985 \cdot a - 0.597 \cdot y$, where c is real per-capita consumption of nondurables and services, a is financial wealth, and y is labor income, all measured in natural logarithms. The coefficients are based on the estimated

cointegrating vector in Lettau and Ludvigson (2002). The data are available from Martin Lettau's web page at <http://www.stern.nyu.edu/~mlettau.htm>.

- *Term spread.* This is the difference between long-term and short-term government bond yields. The long-term government bond yield is the 10-year Treasury bond yield, and the short-term government bond yield is the 3-month Treasury bill rate (secondary market). The data are from FRED. The original data are monthly. We obtain quarterly observations by averaging over the three months comprising a quarter.
- *Default spread.* This is the difference between the Moody's seasoned Baa corporate bond yield and the Moody's seasoned Aaa corporate bond yield. The corporate bond yield data are from FRED. The original corporate bond yield data are monthly. We obtain quarterly observations by averaging over the three months comprising a quarter.
- *Short-term interest rate.* This is the quarterly short-term government bond yield described above. Following the convention in the literature, we express the quarterly short-term interest rate in deviations from a four-quarter moving average consisting of the current and three previous quarters.