NEW FORECASTS OF THE EQUITY PREMIUM

Christopher Polk
Samuel Thompson
Tuomo Vuolteenaho

Working Paper **10406**

NEW FORECASTS OF THE EQUITY PREMIUM

Christopher Polk
Samuel Thompson
Tuomo Vuolteenaho

New Forecasts of the Equity Premium
Christopher Polk, Samuel Thompson, and Tuomo Vuolteenaho
NBER Working Paper No. 10406
March 2004
JEL No. G12, G14

ABSTRACT

If investors are myopic mean-variance optimizers, a stock's expected return is linearly related to its beta in the cross section. The slope of the relation is the cross-sectional price of risk, which should equal the expected equity premium. We use this simple observation to forecast the equity-premium time series with the cross-sectional price of risk. We also introduce novel statistical methods for testing stock-return predictability based on endogenous variables whose shocks are potentially correlated with return shocks. Our empirical tests show that the cross-sectional price of risk (1) is strongly correlated with the market's yield measures and (2) predicts equity-premium realizations especially in the first half of our 1927-2002 sample.

Christopher Polk
Kellogg School of Management
Northwestern University
Evanston, IL 60208
c-polk@kellogg.northwestern.edu

Samuel Thompson
Department of Economics
Harvard University
Cambridge, MA 02138
sthompson@harvard.edu

Tuomo Vuolteenaho
Department of Economics
Harvard University
Cambridge, MA 02138
and NBER
t_vuolteenaho@harvard.edu

1 Introduction

The Capital Asset Pricing Model (CAPM) predicts that risky stocks should have lower prices and higher expected returns than less risky stocks (Sharpe, 1964, Lintner, 1965, Black, 1972). The CAPM further specifies the beta (the regression coefficient of a stock's return on the market portfolio's return) as the relevant measure of risk. According to the Sharpe-Lintner CAPM, the expected-return premium per one unit of beta is the expected equity premium, or the expected return on the value-weight market portfolio of risky assets less the risk-free rate.

We use this CAPM logic to construct a class of simple variables to forecast the equity premium. We compute a number of simple cross-sectional association measures between the stocks' expected-return proxies (including the book-to-market equity ratio, earnings yield, etc.) and the stocks' estimated betas. Low values of the cross-sectional association measures should on average be followed by low realized equity premia and high values by high realized equity premia. Should this not be the case, there would be an incentive for a myopic mean-variance investor to dynamically allocate his/her portfolio between high-beta and low-beta stocks. Since not all investors can overweight either high-beta or low-beta stocks in equilibrium, prices must adjust such that the cross-sectional price of risk and the expected equity premium are consistent.

Our cross-sectional beta-premium variables are empirically successful, as evident from the following two results. First, the variables are highly negatively correlated with the price level of the stock market. Since a high equity premium almost necessarily manifests itself with a low price for the market, negative correlation between our variables and the S&P 500's valuation multiples is reassuring. In particular, our cross-sectional measures have a correlation as high as 0.8 with "the Fed model's" ex-ante equity-premium forecast (defined by us as the smoothed earnings yield minus the long-term Treasury bond yield.)

Second, the cross-sectional beta-premium measures forecast the excess returns on the CRSP value-weight index. For the 1927:5-2002:12 period, most of our cross-sectional beta premium variables are statistically significant predictors at a better than 1% level of significance, with the predictive ability strongest in the pre-1965 subsample. Although predictability is less strong later in the sample, a positive association between future market returns and our cross-sectional beta-premium measures is always covered within the 95% confidence interval in all sample partitionings we

1

have tried. As with most prediction exercises, the predictability is more extreme for the CRSP equal-weight index, which gives a heavier weight to low-capitalization stocks. These predictive results are also robust to a number of alternative methods of constructing the measure of cross-sectional price of risk: Any cross-sectional measure that compares the betas of value stocks to those of growth stocks will provide predictability results that are consistent to those we report in this paper.

We obtain similar predictive results in an international sample. Because of data constraints (we only have portfolio-level data for our international sample), we define our cross-sectional risk premium measure as the difference in the local-market beta between value and growth portfolios. If the expected equity premium is high (and the CAPM holds), a sort on valuation measures will sort a disproportionate number of high-beta stocks into the value portfolio and low-beta stocks into the growth portfolio. Thus a high beta of a value-minus-growth portfolio should forecast a high equity premium, holding everything else constant. In a panel of 22 countries, the past local-market beta of value-minus-growth is a statistically significant predictor of the future local-market equity premium, consistent with our alternative hypothesis. In individual country regressions, the sign of the coefficient is as predicted for 17 of the 22 countries we study. Nine of these 22 coefficients are statistically significant at the 10% level. In a pooled regression with a homogenous predictive coefficient across countries, we always reject the null at better than the 5% level of significance.

In multiple regressions forecasting the equity premium, the cross-sectional beta premium beats the term yield spread (for all measures), but the horse race between the market's smoothed price-earnings ratio and the cross-sectional beta premium is a draw. This, of course, is not inconsistent with the theory. Campbell and Shiller (1988a) show that if growth in a cash-flow measure is nearly unpredictable, the ratio of price to the cash-flow measure is mechanically related to the long-run expected stock return, regardless of the economic forces determining prices and expected returns. Since our variables are based on an economic theory and cross-sectional measures that are not mechanically linked to the market's expected return, the fact that the two different types of variables track a common predictable component in the equity premium is not surprising if the logic underlying our variables is correct.

In the second post-1965 subsample, the predictive ability of our cross-sectional beta-premium measures is less strong than in the first subsample. Furthermore, the market's smoothed earnings yield and our cross-sectional beta-premium measures are much less correlated in the second subsample than in the first subsample, strongly

diverging in the early 1980's. If the market's smoothed earnings yield is indeed a good predictor of the market's excess return and the cross-sectional beta premium a good predictor of the return of high-beta stocks relative to that of low-beta stocks, the divergence of the two types of equity-premium measures implies a trading opportunity. Consistent with this hypothesis, we document statistically significant forecastability of returns on a hedged market portfolio, constructed by buying the market portfolio and "beta hedging" it by selling high-beta and buying low-beta stocks. According to our point estimates, the annualized Sharpe ratio on this zero-beta zero-investment portfolio was close to one in early 1982. To summarize, most of the time the cross-sectional beta-premium measures closely track the market's smoothed earnings yield and the expected equity premium. In the rare occasions when our measure diverges from the market's smoothed earnings yield, the divergence implies an attractive cross-sectional investment opportunity.

Of course, neither the theory we rely on (the CAPM) or our empirical tests provide insight into why the expected equity premium and cross-sectional beta premium vary over time. Our claim is only that, for the most part of our sample period, the pricing of risk appears consistent enough between the cross section and time series to yield a useful variable for forecasting the equity premium. Whether the expected equity premium is due to time-varying risk aversion (Campbell and Cochrane, 1999), investor sentiment (Shiller, 1981, 2000), investor confusion about expected real cash-flow growth (Modigliani and Cohn, 1979, Ritter and Warr, 2002), or some unmodeled hedging demand beyond our myopic framework (Merton, 1973, Fama, 1998) remains an unanswered question.

We also tackle a statistical question that is important to financial econometrics. In many time-series tests of return predictability, the forecasting variable is persistent and shocks to the forecasting variable are correlated with return shocks. It is well known that in this case the small-sample p-values obtained from the usual student-t test can be misleading (Stambaugh, 1999, Hodrick, 1992, and others). Even in the Gaussian case, complex Monte-Carlo simulations such as those performed by Ang and Bekaert (2001) have been the main method of reliable inference for such problems.

We describe a method for computing the small-sample p-values for the Gaussian error distributions in the presence of a persistent and correlated forecasting variable. Our method is an implementation of Jansson and Moreira's (2002) idea of conditioning the critical value of the test on a sufficient statistic of the data. Specifically, we implement a function with a neural network (essentially a fancy look-up table) that

3

maps the sufficient statistics of the data to the critical value for the usual OLS t-statistic. Our Monte Carlo experiments show that this conditional-critical-value function produces a correctly sized test (i.e., the error is less than the Monte Carlo computational accuracy). While calibration of this function is a computationally intensive task, once the function is calibrated the user only needs to compute some simple statistics of the data to find the correct critical values for the t-statistic.

The organization of the paper is as follows. In Section 2, we recap the CAPM and the link between the cross-sectional beta premium and the expected equity premium. In Section 3, we describe the construction of our cross-sectional beta-premium measures. Section 4 describes the statistical method. In Section 5, we present and interpret our empirical results. Section 6 concludes.

2 CAPM can link the time series and cross section

According to the Sharpe-Lintner CAPM, the expected-return premium per one unit of beta is the expected equity premium, or the expected return on the value-weight market portfolio of risky assets less the risk-free rate:

$$E_{t-1}(R_{i,t}) - R_{rf,t-1} = \beta_{i,t-1}[E_{t-1}(R_{M,t}) - R_{rf,t-1}] \qquad (1)$$

In the above equation (1), $R_{i,t}$ is the simple return on asset i during the period t. $R_{rf,t-1}$ is the risk-free rate during the period t known at the end of period $t-1$. $R_{M,t}$ is the simple return on the value-weight market portfolio of risky assets. $\beta_{i,t-1}$, or "beta" of stock i, is the conditional regression coefficient of $R_{i,t}$ on $R_{M,t}$, known at time $t-1$.

$E_{t-1}(R_{M,t}) - R_{rf,t-1}$ is the expected market premium, which we assume to be well approximated by the equity premium. In our empirical implementation, we use the CRSP value-weight portfolio of stocks as our proxy for the market porfolio. Roll (1977) argues that this proxy is too narrow, since it excludes many assets such as human capital, real estate, and corporate debt. Although Stambaugh (1982) shows some evidence that the inference about CAPM is insensitive to exclusion of less risky assets, a cautious reader may choose to interpret our subsequent results within the arbitrage-pricing-theory framework of Ross (1976).

Intuitively, a high expected return on stock i (caused by either a high beta of stock i or a high equity premium) should translate into a low price for the stock. Consistent with this intuition, Gordon (1962) proposes a stock-valuation model that can be inverted to yield an ex-ante equity-premium forecast:

$$\frac{D_{i,t}}{P_{i,t-1}} - R_{rf} + E(g_i) = E(R_i) - R_{rf} \tag{2}$$

The above equation states that the expected return on the stock equals the dividend yield $(D_{i,t}/P_{i,t-1})$ minus the interest rate plus the expected dividend growth $E(g_i)$. The Gordon model (2) has the limitation that expected returns and expected growth must be constant.

Campbell and Shiller (1988a) develop a loglinear approximate present-value relation that allows us to make this intuitive link more formal in the presence of time-varying expected returns and expected growth. Campbell and Shiller approximate the definition of log return on a dividend-paying asset, $r_{i,t+1} \equiv \log(P_{i,t+1} + D_{i,t+1}) - \log(P_{i,t})$, around the mean log dividend-price ratio, $\overline{(d-p)}$, using a first-order Taylor expansion. Above, P denotes price, D dividend, and lower-case letters log transforms. The resulting approximation is $r_{i,t+1} \approx k + \overline{\rho}p_{i,t+1} + (1 - \overline{\rho})d_{i,t+1} - p_{i,t}$, where $\overline{\rho}$ and k are parameters of linearization defined by $\overline{\rho} \equiv 1/(1 + \exp\overline{(d-p)})$ and $k \equiv -\log(\overline{\rho}) - (1 - \overline{\rho})\log(1/\overline{\rho} - 1)$. A typical value for $\overline{\rho}$ is 0.97. When the dividend-price ratio is constant, then $\overline{\rho} = P/(P + D)$, the ratio of the ex-dividend to the cum-dividend stock price.

Solving forward iteratively, imposing the "no-infinite-bubbles" terminal condition that $\lim_{j\to\infty} \overline{\rho}^j(d_{i,t+j} - p_{i,t+j}) = 0$, taking expectations, and subtracting the current dividend, one gets

$$p_{i,t-1} - d_{i,t-1} \approx \frac{k}{1 - \overline{\rho}} + \sum_{j=0}^{\infty} \overline{\rho}^j [\mathrm{E}_{t-1}\Delta d_{i,t+j} - \mathrm{E}_{t-1}r_{i,t+j}] , \tag{3}$$

where Δd denotes log dividend growth. This equation says that the log price-dividend ratio is high when dividends are expected to grow rapidly, or when stock returns are expected to be low. The equation should be thought of as an accounting identity rather than a behavioral model; it has been obtained merely by approximating an identity, solving forward subject to a terminal condition, and taking expectations. If the price-dividend ratio cannot permanently diverge, the investors must then expect some combination of high dividends and low stock returns if their expectations are

5

to be consistent with the observed (high) price, and there are no exceptions to this constraint.

Reorganizing equation (3), substituting the Sharpe-Lintner CAPM's prediction for expected return, assuming that betas and the risk-free rate are constant, and ignoring the time and cross-sectional variation due to the difference between expected log and simple returns yields:

$$d_{i,t-1} - p_{i,t-1} \approx c_{t-1} + \beta_i \sum_{j=0}^{\infty} \overline{\rho}^j E_{t-1}[R_{M,t+j} - R_{rf,t+j-1}] - E_{t-1} \sum_{j=0}^{\infty} \overline{\rho}^j \Delta d_{i,t+j}, \quad (4)$$

where c_{t-1} denotes the collection of constant terms that includes, among other things, the risk-free rate term. In the reasonable cases in which the expected equity premium is positive, the dividend yield on stock i can be high for three reasons. First, the stock may have a high beta. Second, the premium per a unit of beta, that is, the expected equity premium, may be high. Third, and finally, the dividends of the stock may be expected to grow slowly in the future.

Equation (4) leads to a natural cross-sectional measure of the equity premium. Simply regressing the cross-section of dividend yields on betas and expected dividend growth recovers the long-horizon expected equity premium:

$$d_{i,t-1} - p_{i,t-1} \approx \lambda_{0,t-1} + \lambda_{1,t-1}\beta_i + \lambda_{2,t-1} \left[E_{t-1} \sum_{j=0}^{\infty} \overline{\rho}^j \Delta d_{i,t+j} \right], \quad (5)$$

If the expected one-period equity premium $E_{t-1}[R_{M,t+j} - R_{rf,t+j-1}]$ follows a first-order autoregressive process, then the expected one-period equity premium is linearly related to the multiple regression coefficient $\lambda_{1,t-1}$. The central idea in our paper is to measure $\lambda_{1,t-1}$ for each period using purely cross-sectional data, and then use the estimate of $\lambda_{1,t-1}$ to forecast the next period's equity premium.

Of course, the regression of valuation multiple on betas need not be based on dividend yields and dividend growth rates. A similar logic can be applied to the log-linear book-to-market model by Vuolteenaho (2000, 2002). In the log-linear book-to-market model, the current log book-to-market ratio is equal to the discounted sum of future log returns minus the discount sum of future log returns on equity. (Return on equity, or ROE, is defined as clean-surplus earnings divided by the beginning of the period book equity). Applying the above steps to the log-linear book-to-market

6

model will give an analogous regression equation for recovering the cross-sectional beta premium:

$$b_{i,t-1} - p_{i,t-1} \approx \lambda_{0,t-1} + \lambda_{1,t-1}\beta_i + \lambda_{2,t-1} \left[\mathrm{E}_{t-1} \sum_{j=0}^{\infty} \overline{\rho}^j (roe_{i,t+j}) \right], \qquad (6)$$

where b is log book value of equity and roe is $\log(1 + ROE)$.

It is well known that the CAPM does a poor job describing the cross section of stock returns. However, that failure does not invalidate our approach. First, Cohen, Polk, and Vuolteenaho (2002) show that though the CAPM may do a poor job describing cross-sectional variation in average returns, that model does a reasonable job describing the cross-section of stock prices, which is essentially our left-hand side variable in equations (5) and (6). Second, even if market beta does a poor job of completely describing the cross section of stock returns because other state variables are priced (Merton, 1973), investors will still demand a premium to be exposed to market bets. Our method does not depend on whether that premium only reflects compensation for endowment risk or also includes additional compensation because market beta is correlated with sensitivity to hedging demands.[2] The approach only relies on the pricing of beta being consistent between the cross section and time series and our ability to extract that pricing information.

3   Data and construction of variables

We construct a number of alternative proxies for the cross-sectional risk premium. The first set of proxies, λ^{SRC}, λ^{REG}, and λ^{MSCI} are based on various ordinal association measures between a stock's or portfolio's beta and its valuation ratios. These ordinal measures have the advantage of being robust to outliers in the underlying data and they also never take extreme values themselves. This robustness comes at a cost, however, since the ordinal measures have the disadvantage of throwing away some of the information in the magnitude of the cross-sectional spread in valuation

---

[2]In related work, Polk (2003) shows how a general version of the CAPM holds once assets are orthogonalized to Merton (1973)'s hedging demands. Polk uses this insight to measure the degree of risk aversion implicit in multifactor models by splitting the market premium into the piece due to pure aggregate endowment risk and the piece due to the market portfolio's sensitivity to state variable hedging demands.

multiples. The second set of cross-sectional risk-premium proxies, λ^{DP}, λ^{DPG}, λ^{BM}, and λ^{BMG}, are measured on a ratio scale and thus relate more closely to equations (5) and (6). To alleviate the outlier problem with firm-level regressions, these ratios are computed from cross-sections of value-weight portfolios sorted on valuation multiples. The third type of proxy that we use, λ^{ER}, is perhaps most directly connected to the CAPM market premium but perhaps the least robust to errors in data. This proxy pre-estimates the function that maps the various firm characteristics into expected returns and then regresses the current fitted values on betas, recovering the market premium implied by the firm-level return forecasts. In construction of these proxies, we avoid any look-ahead bias and all of these proxies are thus valid forecasting variables in our equity-premium forecasting regressions.

3.1 λ^{SRC} measure of the cross-sectional price of risk

We construct our first measure of the cross-sectional price of risk, λ^{SRC}, in three steps. First, we compute a number of valuation ratios for all stocks. In our empirical implementation, the main challenges are related to selecting appropriate proxies for the valuation multiple. Since dividend policy is largely arbitrary at the firm level, it would be ill-advised to use firm-level dividend yield directly as the only variable on the left-hand-side of regression (5). Instead, we will use a robust composite measure of multiple different valuation measures. An additional complication in construction of the left-hand-side variable is that there are likely structural breaks in the data series, due to changes in dividend policy, accounting rules, and sample composition. To avoid these pitfalls, we use an ordinal composite measure of the valuation multiple by transforming the valuation ratios into a composite rank, with a higher rank denoting higher expected return.

We calculate four raw firm-level accounting ratios, dividend-to-price ratio (D/P), book-to-market equity (BE/ME, the ratio of the book value of common equity to its market value), earnings/price (E/P) and cash flow/price (C/P). The raw cross-sectional data comes from the merger of three databases. The first of these, the Center for Research in Securities Prices (CRSP) monthly stock file, provides monthly prices; shares outstanding; dividends; and returns for NYSE, AMEX, and NASDAQ stocks. The second database, the COMPUSTAT annual research file, contains the relevant accounting information for most publicly traded U.S. stocks. The COMPUSTAT accounting information is supplemented by the third database, Moody's book equity

8

information for industrial firms as collected by Davis, Fama, and French (2000). Detailed data definitions are the following. We measure D as the total dividends paid by the firm from June year $t-1$ to May year t. We define BE as stockholders' equity, plus balance sheet deferred taxes (COMPUSTAT data item 74) and investment tax credit (data item 208) (if available), plus post-retirement benefit liabilities (data item 330) (if available), minus the book value of preferred stock. Depending on availability, we use redemption (data item 56), liquidation (data item 10), or par value (data item 130) (in that order) for the book value of preferred stock. We calculate stockholders' equity used in the above formula as follows. We prefer the stockholders' equity number reported by Moody's, or COMPUSTAT (data item 216). If neither one is available, we measure stockholders' equity as the book value of common equity (data item 60), plus the book value of preferred stock. (Note that the preferred stock is added at this stage, because it is later subtracted in the book equity formula.) If common equity is not available, we compute stockholders' equity as the book value of assets (data item 6) minus total liabilities (data item 181), all from COMPUSTAT. We calculate E as the three-year moving average of income before extraordinary items (data item 18). Our measure of C is the three-year moving average of income before extraordinary items plus depreciation amd amortization (data item 14). In both the calculation of E and C, we require data to be available for the last three consecutive years. We match D along with the BE, E, and C for all fiscal year ends in calendar year $t-1$ (1926-2001) with the firm's market equity at the end of May year t to compute D/P, BE/ME, E/P, and C/P.

Next, we transform these accouting ratios into a single annual ordinal composite measure of firm-level valuation. Specifically, each year we independently transform each ratio into a percentile rank, defined as the rank divided by the number of firms for which the data are available. After computing these four relative percentile rankings, we average the available (up to four) accounting-ratio percentile ranks for each firm. This average is then reranked across firms (to spread the measure for each cross section over the interval from zero to one), resulting in our expected return measure, $VALRANK_{i,t}$. High values of $VALRANK$ correspond to low prices and – according to the logic of Graham and Dodd (1934) and the empirical findings of Ball (1978), Banz (1981), Basu (1977, 1983), Fama and French (1992), Lakonishok, Shleifer, and Vishny (1994), Reinganum (1981), and Rozenberg, Reid, and Lanstein (1985) – also to high expected subsequent returns.

Second, we measure betas for individual stocks. Our monthly measure of risk is estimated market beta, $\widehat{\beta}_{i,t}$. We estimate the betas using at least one and up to three

9

years of monthly returns in an OLS regression on a constant and the contemporaneous return on the value-weight NYSE-AMEX-NASDAQ portfolio.[3] As we sometimes estimate beta using only twelve returns, we censor each firm's individual monthly return to the range (-50%,100%) in order to limit the influence of extreme firm-specific outliers. In contrast to the value measures, we update our beta estimate monthly. Our results are insensitive to small variations in the beta-estimation method.

Third, we compute the association between valuation rank and beta, and use this association measure as our measure of the cross-sectional beta premium. Our first proxy is the Spearman rank correlation coefficient, λ_t^{SRC}, at time t between $VALRANK_{i,t}$ and $\widehat{\beta}_{i,t}$. The resulting monthly series for the proxies begins in May 1927 and ends in December 2002.

The λ^{SRC} proxy has the following advantages mostly due to simplicity and robustness: First, missing data for some valuation multiples are dealt conveniently by averaging the ranks on available multiples. Second, the use of ranks eliminates any hardwired link between the level of the market's valuation and the magnitude of the cross-sectional spread in valuation levels. Third, ranks are a transformation of the underlying multiples that is extremely robust to outliers. The proxy also has the following disadvantages. First, in computing λ^{SRC} we do not control for expected growth and profitability that may be cross-sectionally related to betas, causing an omitted-variables bias in the estimates. Second, if the independent variation in expected firm-level growth and profitability explains a small fraction of the cross-sectional spread in valuation multiples, the ordinal nature of λ^{SRC} may cause us to throw away some significant information related to expansions and contractions of the cross-sectional spread in betas and valuation multiples.

3.2 λ^{REG} measure of the cross-sectional price of risk

Our second measure, λ^{REG}, modifies λ^{SRC} to control for growth opportunities. In order to control for growth opportunities, we need proxies for expected future growth (5) and profitability (6) to serve as control variables in our empirical implementation. A textbook treatment of the Gordon growth model shows that two variables, return on equity and dividend payout ratio, drive a firm's long-term growth. Thus, we use

---

[3] We skip those months in which a firm is missing returns. However, we require all observations to occur within a four-year window.

as our primary profitability controls those selected by Fama and French (1999) to predict firm level profitability, excluding variables that have an obvious mechanical link to our valuation measures.

Our first profitability control is D/BE, the ratio of dividends in year t to year $t-1$ book equity, for those firms with positive book equity. Fama and French motivate this variable by the hypothesis that firms target dividends to the permanent component of earnings (Lintner, 1956, Miller and Modigliani, 1961, and others). We censor each firm's D/BE ratio to the range (0,0.15) to limit the influence of near-zero book equity firms. Following Fama and French (1999), our second profitability control is a non-dividend-paying dummy, DD, that is 0 for dividend payers and 1 for those firms not paying dividends. Including DD in the regression in addition to D/BE helps capture any nonlinearity between expected profitability and dividends. As Fama and French (1999) document substantial mean reversion in profitability, our third and fourth profitability controls are past long-term profitability and transitory profitability. We calculate long-term profitability as the three-year average clean-surplus profitability, $\overline{ROE} \equiv (BE_t - BE_{t-3} + D_{t-2} + D_{t-1} + D_t)/(3 \times BE_{t-3})$. We define transitory profitability as $ROE - \overline{ROE}$, where ROE is current profitability and is equal to $(BE_t - BE_{t-1} + D_t)/(BE_{t-1})$. Our fifth profitability control is a loss dummy. Firms losing money typically continue to do poorly in the future. We motivate our final profitability control from the extensive Industrial Organization literature on product market competition. This proxy is the Herfindahl index of equity market capitalizations for the top five firms in the two-digit SIC-code industry. Low concentration within industry should signal intense competition and thus lower profitability. Since the selection of growth proxies is a judgement call, it is fortunate that our main subsequent results are insensitive to the inclusion or exclusion of expected-growth measures.

λ_t^{REG} is the cross-sectional regression coefficient, λ_t^{REG} of $VALRANK_{i,t}$ on $\widehat{\beta}_{i,t}$ and growth/profitability controls, estimated with OLS:

$$VALRANK_{i,t} = \lambda_{0,t} + \lambda_t^{REG}\widehat{\beta}_{i,t} + \sum_{g=1}^{6} \lambda_t^g GROWTHRANK_{i,t}^g + \varepsilon_{i,t} \qquad (7)$$

$GROWTHRANK_{i,t}^g$ is the corresponding percentile rank for six firm-level profitability controls. Since Cohen, Polk, and Vuolteenaho (2003) show that the majority of the cross-sectional variation in valuation ratios across firms is due to differences in expected future profitability, not differences in future expected returns, these controls

have the potential to improve our measurement of the cross-sectional beta premium significantly.

3.3 λ^{MSCI} measure of the cross-sectional price of risk

We also measure the cross-sectional price of risk for an international sample of 22 countries using an ordinal measure. Since we do not have security-level data for our international sample, only portfolio returns, we work with value and growth portfolios constructed by Kenneth French and available on his web site. We take the top-30% and bottom-30% portfolios sorted on four of Morgan Stanley Capital International's value measures: D/P, BE/ME, E/P, and C/P. We then estimate the betas for these portfolios using a three-year rolling window, and define the predictor variable λ^{MSCI} as the average beta of the four value portfolios minus the average beta of the four growth portfolios. The subsequent international results are insensitive to changing the beta-estimation window to four or five years (longer windows actually improve the results), and to selecting a subset of value measures for constructing λ^{MSCI}.

3.4 λ^{DP} and λ^{DPG} measures of the cross-sectional price of risk

We also construct cross-sectional risk premium measures that use valuation multiples on a ratio scale. The first two such measures, λ^{DP} and λ^{DPG}, are implemented using five value-weight dividend-yield sorted portfolios. We sort stocks into five portfolios on the end-of-May dividend yield. Then, for each portfolio we measure value-weight average dividend yield (computed as aggregate dividends over aggregate market value) and the value-weight average past estimated beta using the rolling betas updated each month. We then regress these five portfolio-level dividend yields in levels on the portfolios' betas, and denote the regression coefficient by λ^{DP}.

λ^{DPG} modifies λ^{DP} by controlling for past dividend growth. In addition to the dividend yield, we also compute the value-weight one-year dividend growth for the portfolios. λ^{DPG} is the multiple regression coefficient of the portfolio's dividend yield on its beta, controlling for its one-year dividend growth.

3.5 λ^{BM} and λ^{BMG} measures of the cross-sectional price of risk

We construct book-to-market based proxies λ^{BM} and λ^{BMG} analogously to λ^{DP} and λ^{DPG}. We sort stocks into five portfolios based on end-of-May BE/ME. Then, for each portfolio we measure value-weight average BE/ME (computed as aggregate book value of equity over aggregate market value) and the value-weight average past estimated beta using the rolling betas updated each month. We then regress these five portfolio-level book-to-market ratios in levels on the portfolios' betas, and denote the regression coefficient by λ^{BM}. λ^{BMG} is the multiple regression coefficient of the portfolio's BE/ME on its beta, controlling for the portfolios one-year past value-weight ROE.

3.6 λ^{ER} measure of the cross-sectional price of risk

In contrast to our other measures of cross-sectional risk premium that relate price levels to betas, we also measure the cross-sectional price of risk based on how well betas explain estimates of one-period expected returns. We extract this measure using a two-stage approach. Our first stage is as follows. Each month, using a rolling ten-year panel of data over the period $t-120$ to $t-1$, we regress cross-sectionally demeaned firm-level returns on lagged cross-sectionally demeaned characteristics: $VALRANK$; $\widehat{\beta}$; the raw valuation multiples D/P, BE/ME, E/P, and C/P; and the raw profitability controls used in construction of λ^{REG}. Note that in this regression we replace missing values with cross-sectional means and drop E/P and C/P from the specification in subperiods in which data for those measures are not available for any firm. The resulting coefficient estimates in conjunction with the time t observations on the associated characteristics produce forecasts of firm-level expected returns at time t. In our second stage, we regress these forecasts on our beta estimates as of time t. We repeat this process each month, generating our λ^{ER} series as the coefficients of these cross-sectional regressions.

3.7 Other variables

We use two measures of the realized equity premium. The first measure is the excess return on the value-weight market portfolio (R_M^e), computed as the difference between the simple return on the Center for Research in Securities Prices (CRSP) value-weight stock index (R_M) and the simple risk-free rate. The risk-free-rate data are constructed by CRSP from Treasury bills with approximately three months to maturity. The second measure (R_m^e) is the excess return on the CRSP equal-weight stock index. For the international sample, we use an equity-premium series constructed from MSCI's stock-market data and interest-rate series from Global Financial Data.

We also construct variables that should logically predict the market return if the expected equity premium is time varying. Previous research shows that scaled price variables and term-structure variables forecast market returns. We pick the smoothed earnings yield and term yield spreads as examples of such variables, and horse race them against our variables.

The log earnings-price ratio (ep) is from Shiller (2000), constructed as a ten-year trailing moving average of aggregate earnings of companies in the S&P 500 index divided by the price of the S&P 500 index. Following Graham and Dodd (1934), Campbell and Shiller (1988b, 1998) advocate averaging earnings over several years to avoid temporary spikes in the price-earnings ratio caused by cyclical declines in earnings. We follow Campbell and Vuolteenaho's (2003) method of constructing the earnings series to avoid any forward-looking interpolation of earnings. This ensures that all components of the time t earnings-price ratio are contemporaneously observable by time t. The ratio is log transformed.

The term yield spread (TY) is provided by Global Financial Data and is computed as the yield difference between ten-year constant-maturity taxable bonds and short-term taxable notes, in percentage points. The motivation of the term yield spread as a forecasting variable, suggested by Keim and Stambaugh (1986) and Campbell (1987), is the following: TY predicts excess returns on long-term bonds. As stocks are also long-term assets, it should also forecast excess stock returns, if the expected returns of long-term assets move together.

In our informal illustrations, we also use the dividend-price ratio, computed as the ratio of trailing twelve-month dividends and the price for the S&P 500 index. We also use simple (not log) smoothed earnings yield, which is defined simply as $\exp(ep)$. In

the Gordon (1962) model computations, any interest rate adjustments are performed using the same ten-year constant-maturity taxable bond yield ($Y10$) as is used in the computation of the term yield spread.

4   Conditional tests for predictive regressions

This section describes the statistical methodology for computing the correct small-sample critical values of the usual t-statistic in those situations where the forecasting variable is persistent and shocks to the forecasting variable are potentially correlated with shocks to the variable being forecast.

Consider the one-period prediction model

$$
\begin{aligned}
y_t &= \mu_1 + \theta x_{t-1} + u_t \\
x_t &= \mu_2 + \rho x_{t-1} + v_t
\end{aligned}
\tag{8}
$$

with $\mathrm{E}\, u_t = \mathrm{E}\, v_t = 0$, $\mathrm{E}\, u_t^2 = \sigma_u^2$, $\mathrm{E}\, v_t^2 = \sigma_v^2$ and $\mathrm{Corr}\,(u_t, v_t) = \gamma$. In a practical example introduced by Stambaugh (1999), y is the excess stock return on a stock market index and x is the index dividend yield.   Since dividends are smooth and returns cumulate to price, we have strong a priori reasons to expect the correlation γ to be negative.

We wish to test the null hypothesis $\theta = 0$, indicating that x does not predict y, or in Stambaugh's (1999) example that the dividend yield does not predict stock returns. The usual t-statistic for this hypothesis is

$$
\widehat{t} = \widehat{\sigma}_u^{-1} \sqrt{\sum (x_{t-1} - \overline{x})^2}\, \widehat{\theta},
\tag{9}
$$

where $\widehat{\theta}$ is the least squares estimate of θ and $\widehat{\sigma}_u^2$ is an estimator of σ_u^2.   Classical asymptotic theory states that in a large sample the t-statistic is approximately distributed standard normal. It is well known, however, that this is a poor approximation to the true sampling distribution of $\widehat{t}$. For example, Stambaugh (1999) shows that when x is the dividend yield and y is the market excess return, the null distribution of $\widehat{t}$ is centered at a positive number, leading to over-rejection of a true null hypothesis.

In order to get the size of the test right, we want a critical value q equal to the 95% quantile of the null distribution of $\widehat{t}$. When the errors are normal, the exact null

distribution of $\widehat{t}$ depends on the parameter ρ. Thus there exists a function $k(\rho)$ so that under the null, $\Pr[\widehat{t} > k(\rho)] = .05$. One can calculate $k(\rho)$ by the bootstrap or using methods described by Imhof (1961). We cannot directly use $k(\rho)$ as a critical value because we do not know ρ, and evaluating $k(\rho)$ at the least squares estimate $\widehat{\rho}$ leads to size distortions.

Recently Michael Jansson and Marcelo Moreira have proposed a solution to this problem (Jansson and Moreira, 2002). Suppose that the covariance parameters σ_u^2, σ_v^2 and γ are known. Under the null that $\theta = 0$, the statistics

$$S = \left\{ \frac{\sum (x_{t-1} - \overline{x})(x_t - \sigma_v \gamma y_t / \sigma_u)}{\sum (x_{t-1} - \overline{x})^2}, \ \sum (x_{t-1} - \overline{x})^2, \ \overline{x}, \ \overline{y}, \ x_1, \ y_1 \right\} \tag{10}$$

are sufficient statistics for the parameter ρ, where $\overline{x} = (T-1)^{-1} \sum_{t=2}^{T} x_{t-1}$ and $\overline{y} = (T-1)^{-1} \sum_{t=2}^{T} y_t$. Recall the definition of a sufficient statistic: a statistic S is sufficient for a parameter ρ if the conditional distribution of the data given S is independent of ρ. While the unconditional distribution of $\widehat{t}$ depends on the unknown ρ, the conditional distribution does not. So set the critical value to a quantile of the conditional distribution. Let $q(s, \alpha)$ denote the α-quantile of the conditional null distribution of $\widehat{t}$ given $S = s$:

$$\Pr[\widehat{t} > q(s, \alpha) \mid S = s, \ \theta = 0] = \alpha. \tag{11}$$

When the covariance parameters are known, a test which rejects the null when $\widehat{t} > q(S, \alpha)$ will have the correct null rejection probability in any sample size and for any value of ρ.

Jansson and Moreira (2002) do not provide a closed form expression for the conditional distribution of t given the sufficient statistics. Our contribution is to devise a computationally feasible implementation of their procedure. We approximate the critical function q_α with q_α^{nn}, a neural network:

$$q_\alpha(S, \alpha) \ \approx \ q_\alpha^{nn}(X, \widehat{\psi}, \widehat{\phi}), \tag{12}$$

$$q_\alpha^{nn}(X, \psi, \phi) \ \equiv \ \psi_0 + \sum_{j=1}^{5} \psi_j g\left(\phi_j' X\right),$$

16

with $\phi_j = (\phi_{j1}, \phi_{j2}, \phi_{j3}, \phi_{j4})'$ and

$$X = \left(1, \, T\left(\frac{\sum (x_{t-1} - \overline{x})(x_t - \widehat{\sigma}_v\widehat{\gamma}y_t/\widehat{\sigma}_u)}{\sum (x_{t-1} - \overline{x})^2} - 1\right), \, \log\left(T^{-2}\sum (x_{t-1} - \overline{x})^2 / \widehat{\sigma}_v^2\right), \, \widehat{\gamma}\right)'. \tag{13}$$

The hatted variables are the usual least-squares estimators of the covariance parameters. g is called the *activation function*. We use the tanh activation function

$$g(x) = \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}. \tag{14}$$

ψ and ϕ are parameters of the neural net. As White (1992) has shown, this network is a universal approximator: given enough activation functions, q_α^{nn} can approximate any nonlinear function to an arbitrary degree of accuracy.

We choose ψ and ϕ to closely approximate the critical function q_α. Fitting the net is a computationally demanding task. We describe the algorithm used to fit the net in the Appendix. Once the net has been estimated, it is very easy to apply, since the approximate critical values are known in closed form. As an example, we provide one these functions in Appendix. Matlab code for calculating the others will be made available from the authors upon request.[4]

The vector X differs from the sufficient statistics in several ways. X transforms some of the sufficient statistics, and omits the statistics $\overline{x}$, $\overline{y}$, x_1 and y_1. X uses parameter estimates $\widehat{\sigma}_v$, $\widehat{\gamma}$ and $\widehat{\sigma}_u$ in place of the known covariance parameters. The X vector was chosen for computational convenience. The algorithm that fits the net searches over all relevant parameter values. The test statistic $\widehat{t}$ and each element of X is exactly invariant to σ_u^2 and σ_v^2, so the fitting algorithm does not need to search over all possible variances. Furthermore, the omitted statistics $\overline{x}$, $\overline{y}$, x_1 and y_1 are not particularly informative about the nuisance parameter ρ. We will see that size distortions caused by omitting the statistics are within the simulation error of our Monte Carlo study.

The Jansson-Moreira theory delivers an exact test when the covariance parameters are known. In practice one must use parameter estimates. We designed the neural net training algorithm to correct for estimation error in the covariance parameters. This is not a completely clean application of the statistical theory. It may be the case

---

[4]Contact Samuel Thompson at sthompson@harvard.edu.

that no exact test exists in this model. However, we will see that any size distortions caused by unknown covariance parameters are quite small and within the simulation error of our Monte Carlo study. Furthermore, estimation error for the covariance parameters is asymptotically negligible, whether the x_t process is stationary or not. While classical asymptotic theory requires stationarity, the conditional testing theory is not sensitive to "unit root" problems. See the argument by Jansson and Moreira (2002) for details.

A Monte Carlo experiment demonstrates the accuracy of our approximation. Figure 1 reports empirical rejection frequencies over a range of values for ρ and $\text{Corr}(u,v)$. For each $(\rho, \text{Corr}(u,v))$ pair, we simulate 50000 samples of 120 observations each, and perform a t-test of the null $\theta = 0$ against the alternative $\theta > 0$. Nominal test size is 5%. The plot on the left reports results for the classical critical value of 1.65, and the plot on the right uses the conditional critical function q_α^{nn}.

When ρ is close to 1, the classical critical value underrejects for positive $\text{Corr}(u,v)$ and overrejects for negative $\text{Corr}(u,v)$. When $\rho = 1$ and the correlation is $-.9$ the usual critical value rejects a true null about 38% of the time. Our conditional critical function leads to accurate rejection frequencies over the entire range of ρ and $\text{Corr}(u_t, v_t)$ values, with rejection rates ranging from 4.70% to 5.18%. X and the t-statistic are exactly invariant to σ_u^2 and σ_v^2, so these results hold for any variance parameters.

The conditional critical values are also superior to critical values generated by the parametric bootstrap. When x is the dividend yield, Stambaugh estimated $\text{Corr}(u,v)$ to be $-.90$ and ρ to be $.972$. We "bootstrap the bootstrap" at similar parameter values: we simulated 10000 samples of 120 observations each, setting $\text{Corr}(u,v) = -.90$ and varying ρ over the range $.95$, $.975$, and 1.00. For each simulated sample we bootstraped 5000 new data sets from the model with normal errors, setting $\theta = 0$ and the other parameters to their least squares estimates. We set the bootstrapped critical value equal to the 95th percentile of bootstrapped t-statistics. The rejection frequencies were 7.93, 9.97 and 15.55% for ρ equal to .95, .975 and 1.00. Rejection frequencies from using the conditional quantile function were essentially equal to 5%; the differences were within the Monte Carlo error.

Although the above experiments show that size of the test is correct, we make no guarantees about its power. It is possible that tests designed by Lewellen (2002) and Campbell and Yogo (2002) may have more power under some circumstances.
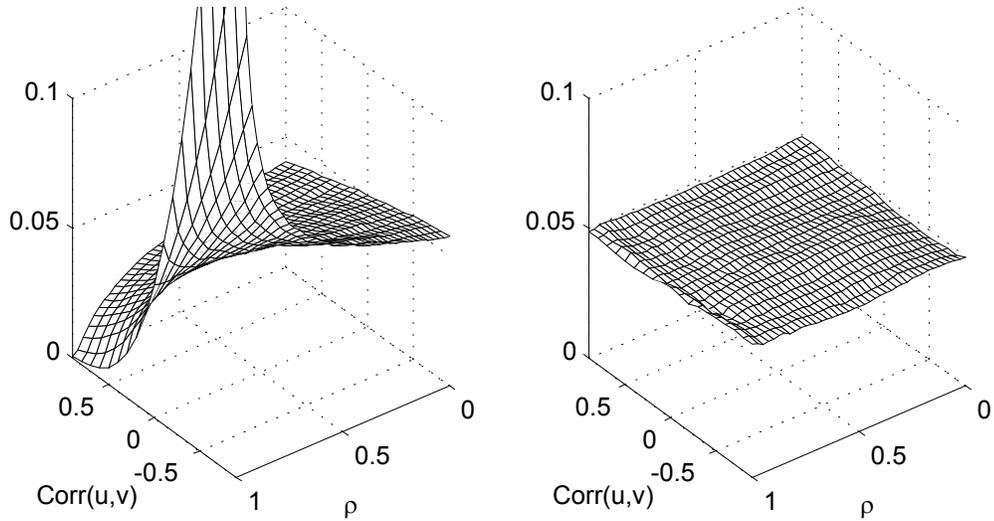
Figure 1: Size of the test in a Monte Carlo experiment.

Consider the data-generating process
$y_t = \mu_1 + \theta x_{t-1} + u_t$
$x_t = \mu_2 + \rho x_{t-1} + v_t.$
We are interested in testing the hypothesis $\theta = 0$ against $\theta > 0$. On the left we see empirical rejection frequencies from using the usual critical value of 1.65 for a one-tailed t test. On the right we see results from the conditional critical function q_α^{nn}. The conditional critical function leads to rejection frequencies very close to the nominal 5%. The grid values range over $\mathrm{Corr}(u, v) \in \{-.9, -.8, ..., .8, .9\}$ and $\rho \in \{0, .025, ..., .975, 1\}$. For each grid point there are 50000 Monte Carlo trials of $T = 120$ observations.

4.1 Constructing confidence intervals

Confidence intervals consist of all the nulls we fail to reject. We construct confidence intervals for θ by inverting a sequence of hypothesis tests. Let $\Phi = (\mu_1, \mu_2, \theta, \rho, \sigma_u^2, \sigma_v^2, \gamma)'$ denote the parameters of the model. A $100\alpha\%$ confidence set C for θ has the property that it contains the true parameter value with probability at least α:

$$\inf_{\Phi} \Pr\left[\theta \in C; \Phi\right] \geq \alpha \text{ for all } \Phi. \tag{15}$$

C is a random interval, since it is a function of the data, and $\Pr\left[\theta \in C; \Phi\right]$ denotes the probability that θ is in C given the parameters Φ. Suppose that, for each point $\overline{\theta}$ in the parameter space, we carry out the conditional t test of size $1 - \alpha$ for the hypothesis $\theta = \overline{\theta}$. We define C as the set of all $\overline{\theta}$ that we fail to reject. C is a valid interval because it contains the true θ with probability equal to α:

$$\Pr\left[\overline{\theta} \in C; \Phi\right] \equiv \Pr\left[\text{fail to reject null } \theta = \overline{\theta} \text{ when null is true}; \Phi\right] = 1 - \alpha \text{ for all } \Phi. \tag{16}$$

Thus we have an algorithm for constructing confidence intervals. We 1) construct a grid of J null hypotheses $\theta_1 < \theta_2 < \cdots < \theta_J$, 2) test each null $\theta = \theta_j$ versus the two-sided alternative $\theta \neq \theta_j$ and 3) take the confidence interval to be all the θ_j's that are not rejected.[5]

The conditional tests we have described so far are designed to test the null that θ is zero. To test the general null $\theta = \theta_j$, transform the model so that the null is again zero. Create the variable $\widetilde{y}_t = y_t - \theta_j x_{t-1}$, so the first equation becomes

$$\widetilde{y}_t = \mu_1 + \widetilde{\theta} x_{t-1} + u_t, \tag{17}$$

with $\widetilde{\theta} = \theta - \theta_j$. Then compute a conditional test of the null $\widetilde{\theta} = 0$.

---

[5]Throughout the paper a size $1 - \alpha$ test rejects the null $\theta = \theta_j$ in favor of $\theta \neq \theta_j$ when $\widehat{t} > q\left(S, (1 + \alpha)/2\right)$ or $\widehat{t} < q(S, (1 - \alpha)/2)$.

4.2 Inference in multivariate regressions

This section extends the Janssen-Moreira methodology to a simple vector autoregression. Consider the bivariate regression

$$
\begin{aligned}
y_t &= \mu_1 + \theta' \mathbf{x}_{t-1} + u_t \\
\mathbf{x}_t &= \boldsymbol{\mu}_2 + K \mathbf{x}_{t-1} + V_t,
\end{aligned}
\tag{18}
$$

where $\mathbf{x}_t$, $\boldsymbol{\mu}_2$ and θ are 2-dimensional column vectors, K is a 2×2 matrix, and V_t is a 2-dimensional vector of mean zero errors. For example we could take the elements of $\mathbf{x}_t$ to be the index dividend yield and price earnings ratio, in which case the coefficient vector θ determines the predictive content of each variable controlling for the other. We wish to test the null hypothesis that the first element of θ is zero. The usual approach is to run a multivariate regression and reject the null for large values of the t-statistic

$$
\widehat{t} = \frac{\widehat{\theta}_1}{\sqrt{V_{11}}}.
\tag{19}
$$

$\widehat{\theta}_1$ is the ordinary least squares estimate of θ_1, the first element of θ, and V_{11} is an estimate of the variance of $\widehat{\theta}_1$ - it is the $(1,1)$ element of $V = \widehat{\sigma}_u^2 \left(\sum \mathbf{x}_{t-1} \mathbf{x}_{t-1}' \right)^{-1}$. Classical asymptotic theory approximates the null distribution of $\widehat{t}$ with a standard normal variable. It is well known that this is a poor approximation when the elements of $\mathbf{x}_t$ are highly serially correlated. In many cases of interest, classical theory leads to over rejection of a true null hypothesis.

In principle it is easy to extend the Janssen-Moreira methodology to this model. Suppose that the errors $(u_t, V_t')'$ are iid mean zero normal variables with known covariance matrix $\Sigma = \mathrm{E}\left[(u_t, V_t')' (u_t, V_t') \right]$. The null distribution of $\widehat{t}$ depends on the unknown matrix K. However the conditional null distribution of $\widehat{t}$ given sufficient statistics for K does not depend on unknown parameters. To construct the sufficient statistics, define the transformed variables $(\widetilde{y}_t, \widetilde{\mathbf{x}}_t')' = \Sigma^{-1/2} (y_t, \mathbf{x}_t')'$, where $\Sigma^{1/2}$ is the lower diagonal choleski decomposition of Σ and satisfies $\Sigma^{1/2} \left(\Sigma^{1/2} \right)' = \Sigma$. The sufficient statistics for K are

$$
S = \left\{ \widetilde{K}, \ \sum (\mathbf{x}_{t-1} - \overline{\mathbf{x}}) (\mathbf{x}_{t-1} - \overline{\mathbf{x}})', \ \overline{\mathbf{x}}, \ \overline{y}, \ \mathbf{x}_1, \ y_1 \right\},
\tag{20}
$$

where $\overline{\mathbf{x}} = (T-1)^{-1} \sum_{t=2}^{T} \mathbf{x}_{t-1}$, $\overline{y} = (T-1)^{-1} \sum_{t=2}^{T} y_t$, and $\widetilde{K}$ is the 2×2 matrix of least squares estimates from regressing $\widetilde{\mathbf{x}}_t$ on $\mathbf{x}_{t-1}$ and a constant, and premultiplying

the result by $\Sigma^{1/2}$. The t-test will have correct size for any sample size if we reject the null when $\widehat{t}$ is bigger than the $1 - \alpha$ quantile of the conditional null distribution of $\widehat{t}$ given S.

Computing the quantiles of the conditional null distribution for a multivariate system is a daunting computational problem. In the univariate model (8) with just one regressor, the t-statistic has a null distribution that depends on the two parameters ρ and γ. Our neural net approximation q_α^{nn} "learns" the conditional quantile function by searching over a grid of ρ and γ values. In the two dimensional case, it is computationally feasible to search over all grid points that are close to empirically relevant cases. In the multivariate setting the null distribution depends on the four elements of K as well as the correlation terms in Σ. It does not appear to be computationally feasible for our neural net to learn all possible cases of this high dimensional parameter space. We experimented with different algorithms for fitting the neural net but were unable to achieve the accuracy attained for the univariate model.

In order to carry out conditional inference in the multivariate setting, we propose a modified version of the usual parametric bootstrap. If we could simulate from the conditional distribution of $\widehat{t}$ given S, we could use the empirical quantile of the simulated $\widehat{t}$ draws as the critical value. While we cannot directly simulate from the distribution of $\widehat{t}$ given S, it is straightforward to simulate from their joint distribution - for fixed parameter values simulate data sets from the model and compute $\widehat{t}$ and S. We simulate from the conditional null of $\widehat{t}$ given S using a nearest neighbor estimator: we simulate B draws of $\widehat{t}$ and S, and construct a sample of N conditional draws by choosing the $\widehat{t}$ statistics corresponding to the N draws of S that are closest to the sufficient statistics observed in the data. We call this procedure the "conditional bootstrap." Details are given in the appendix.

A small Monte Carlo experiment suggests that the conditional bootstrap leads to smaller size distortions than the usual parametric bootstrap. We simulated 5000 samples of 120 observations each, setting $\theta_1 = 0$ and the rest of the model parameters equal to the unrestricted least squares estimates when y is R_M^e, the value-weighted CRSP excess return, and $\mathbf{x}_t$ contains the two predictors λ_t^{SRC} and ep_t. For each simulated sample we tested the null $\theta_1 = 0$ against the one-sided alternative $\theta_1 > 0$. We computed critical values using the parametric bootstrap and the conditional bootstrap. For the parametric bootstrap we simulated 20000 new data sets from the model with normal errors, setting $\theta_1 = 0$ and the other parameters to their unrestricted least squares estimates. The conditional bootstrap was computed taking

$B = 20000$ and $N = 1000$.

The above experiment yields the following results. When θ_1 is the coefficient on λ_t^{SRC}, the parametric bootstrap rejected the null 5.48% of the time and the conditional bootstrap rejected 3.84% of the time. When θ_1 is the coefficient on ep_t the rejection frequencies were 11.46% and 4.78%. We then simulated from the model with $K = I$, to see how the bootstraps perform when the predictors follow unit roots. When θ_1 is the coefficient on λ_t^{SRC}, the parametric bootstrap rejected 6.36% of the time and the conditional bootstrap rejected 3.32% of the time. When θ_1 is the coefficient on ep_t the rejection frequencies were 15.64% and 7.80%.

4.3 Generated regressors in predictive regressions

It is well known that the usual OLS coefficient estimate will be biased and the usual OLS standard errors will be incorrect, if the regressor is generated with another regression from the same sample of data. To alleviate potential confusion about the genereated-regressors problem, in this section we explain why our hypothesis tests seeking to reject the null of no predictability are valid, as long as our generated regressor only uses past information in its construction.

Put aside for a moment the problem of bias, and focus instead on the standard error of $\widehat{\theta}$. Let I_t denote the set of all data available at time t and at all preceeding times. Assume that x_t is a function only of data in I_t. For example, if x_t depends on an estimated vector $\widehat{\beta}$, then the estimates $\widehat{\beta}$ must be constructed only from data available up to time t. To clarify this point we write $x_t = x_t(\widehat{\beta}_t)$. Therefore, x_{t-1} is a function only of information available at time $t - 1$, and the usual assumption $\mathrm{E}(u_t|I_{t-1}) = 0$ implies that $\mathrm{E}\left(u_t|x_{t-1}\right) = 0$. This is the only necessary assumption for OLS to deliver the asymptotically correct standard errors for $\widehat{\theta}$. Under stationarity and conditional homoskedasticity, so $\mathrm{E}_{t-1}\, u_t^2 = \sigma^2$, the variance of the OLS estimate is

$$T \operatorname{Var}\left[\widehat{\theta}\right] = T \operatorname{Var}\left[\frac{\sum (x_{t-1} - \overline{x})u_t}{\sum (x_{t-1} - \overline{x})^2}\right] \approx \frac{T^{-1}\sum u_t^2}{T^{-1}\sum (x_{t-1} - \overline{x})^2}, \tag{21}$$

which is consistently estimated by the usual OLS formula per the standard GMM asymptotics.

Why does the predictive problem differ from other problems with generated regres-

sors? Outside of predictive regressions, generated regressors are usually correlated with the regression errors, so $\mathrm{E}\left(u_t | x_{t-1}\right) \neq 0$. For example, if $\widehat{\beta}$ is calculated from data available from time $t = 1$ to T, then $\widehat{\beta}$ will usually be correlated with all the errors $u_1, ..., u_T$ and therefore $x_{t-1}(\widehat{\beta})$ will be correlated with u_t. This correlation becomes small as the sample size grows, so the estimator remains consistent. However, this correlation usually affects the standard error, even asymptotically.

Let us return to the problem of bias caused by generated regressors in the predictive equation of (8) with $\mathrm{E}_{t-1} u_t = 0$. x_{t-1} is a generated regressor, but is generated only from data available up to time $t - 1$, thus $\mathrm{E}(u_t | x_{t-1}) = 0$. OLS delivers a consistent estimate of θ, the coefficient of the best linear predictor of the market return y_t with the (generated) predictive regressor x_{t-1}. Of course, in many cases x_{t-1} is a proxy variable for $E_{t-1}^* y_t$, which is market participants' expected market return next period given information available today. Ideally, we would like to run the regression

$$y_t = \mu_1^* + \theta^* \left(E_{t-1}^* y_t \right) + u_t^*. \tag{22}$$

There is no reason to expect the OLS estimate $\widehat{\theta}$ to converge to θ^* even in large samples. Hence, while $\widehat{\theta}$ is a biased estimate of the coefficient on the true, unknown market expectation, it is a consistent estimator of the coefficient on the proxy variable x_{t-1}.

Finally, note that our test of the null hypothesis of constant expected returns remains valid even if our forecasting variable is a generated regressor as long as the forecasting variable is generated from past information only. (Naturally, regressing future equity-premium realizations on x_{t-1} may lead to a less accurate forecast and thus to a lower power than regressing them on $E_{t-1}^* y_t$, if we wish to test the null hypothesis that the future returns are unpredictable.)

5 Empirical results

Our empirical results can be summarized with two findings. First, the cross-sectional price of risk is highly negatively correlated with the market price level and highly positively correlated with popular ex-ante equity-premium measures derived from the Gordon (1962) growth model, such as the smoothed earnings yield minus the long-term Treasury bond yield.

Second, the cross-sectional beta-premium forecasts future excess-return realizations on the CRSP value-weight index. For the 1927:5-2002:12 period, the cross-sectional beta premium is statistically significant at a level better than 1%, with most of the predictive ability coming from the pre-1965 subsample. We also detect predictability in a largely independent international sample, indicating that our results are not sample specific.

5.1   Correlation with ex-ante equity-premium measures

As an informal illustration, we graph the time-series evolution of popular ex-ante equity-premium measures and our first cross-sectional measure, λ_t^{SRC}, in Figure 2. (We focus on λ_t^{SRC} in these illustrations to save space, but similar results can be obtained for our other cross-sectional variables.)   One popular ex-ante measure is based on the comparison deemed the "Fed model," where the equity risk premium equals the equity yield (either dividend yield or smoothed earnings yield) minus the long-term Treasury bond yield. This measure is often called the Fed model, since the Federal Reserve Board supposedly uses a similar model to judge the level of equity prices.[6]

The Fed model and its variations provide an intuitive estimator of the forward-looking equity risk premium.   The earnings-yield component of the Fed model is easily motivated with the Gordon (1962) growth model.   As for the interest-rate component, there are two related logical bases to the argument that low interest rates should coincide with low earnings yields.   First, if one is interested in the equity premium instead of the total equity return, one has to subtract the interest rate from earnings yield to compare apples to apples.   Second, many argue that an environment of low interest rates is good for the economy and thus raises the expected future earnings growth.

Asness (2002) points out that, while seeming plausible, these arguments are flawed in the presence of significant and time-varying inflation. In the face of inflation, cash

---

[6]The Federal Reserve Board's Monetary Policy Report to the Congress of July 1997 argues: "Still, the ratio of prices in the S&P 500 to consensus estimates of earnings over the coming twelve months has risen further from levels that were already unusually high. Changes in this ratio have often been inversely related to changes in long-term Treasury yields, but this year's stock price gains were not matched by a significant net decline in interest rates."   Of course, the Federal Reserve has not officially endorsed any stock-valuation model.

flows for the stock market should act much like a coupon on a real bond, growing with inflation. Holding real growth constant, low inflation should forecast low nominal earnings growth. In a sense, stocks should be a long-term hedge against inflation. (Modigliani and Cohn, 1979, and Ritter and Warr, 2002, argue that the expected real earnings growth of levered firms in fact increases with inflation.) Thus, in the presence of time-varying inflation, the Fed model of equity premium should be modified to subtract the real (instead of nominal) bond yield, for which data unfortunately do not exist for the majority of our sample period.

An alternative to the implicit constant-inflation assumption in the Fed model is to assume that the real interest rate is constant. If the real interest rate is constant and earnings grow at the rate of inflation, the earnings yield is a good measure of the the forward-looking expected real return on equities. Under this assumption, the earnings yield is also a good measure of the forward-looking equity premium. Figure 2 also plots the smoothed earnings yield without the interest-rate adjustment.

The three variables in Figure 2 are demeaned and normalized by the sample standard deviation. Our sample period begins only two years before the stock market crash of 1929. This event is clearly visible from the graph in which all three measures of the equity premium shoot up by an extraordinary five sample standard deviations from 1929 to 1932. Another striking episode is the 1983-1999 bull market, during which the smoothed earnings yield decreased by four sample standard deviations. However, in 1983 both the smoothed earnings yield less the bond yield (i.e., the Fed model) and our cross-sectional beta-premium variable are already low, and thus diverged from the earnings yield.

It is evident from the figure that our cross-sectional risk premium tracks the Fed model's equity-premium forecast with an incredible regularity. This relation is also shown in Table 1, in which we regress the cross-sectional premium λ^{SRC} on $\exp(ep)$ and $\exp(ep) - Y10$. Essentially, the regression fits extremely well with an R^2 of 72% where the explanatory power is entirely due to the Fed model ($\exp(ep) - Y10$). (Please note that the OLS t-statistics in the table do not take into account the persistence of the variables and errors, and are thus unreliable.) Our conclusion from Table 1 and Figure 2 is that the market prices the cross-sectional beta premium to be consistent with the equity premium implied by the Fed model.

There is potentially a somewhat mechanical link between the market's earnings yield and our cross-sectional measure. Recall that our λ measures are cross-sectional regression coefficients of earnings yields (and other such multiples) on betas. If
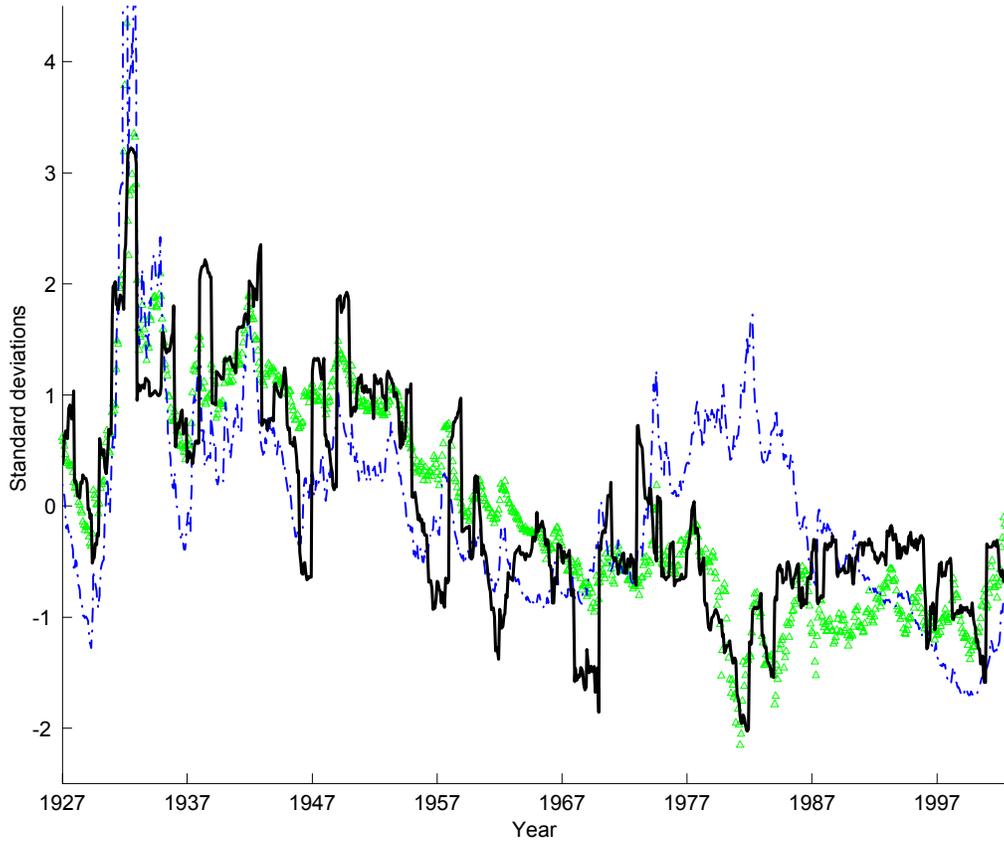
Figure 2: Time-series evolution of the ex-ante equity-premium forecasts.

This figure plots the time-series of three equity-premium measures: (1) λ^{SRC}, the cross-sectional Spearman rank correlation between valuation levels and estimated betas, marked with a thick solid line; (2) $\exp(ep)$, the ratio of a ten-year moving average of earnings to price for S&P500, marked with a dash-dotted line; and (3) $\exp(ep) - Y10$, the ratio of a ten-year moving average of earnings to price for S&P500 minus the long-term Government bond yield, marked with triangles. All variables are demeaned and normalized by their sample standard deviations. The sample period is 1927:5-2002:12.

the market has recently experienced high past returns, high-beta stocks should have also experienced high past returns relative to low-beta stocks. The high return on high-beta stocks will imply a lower yield on those stocks, if earnings do not adjust immediately. Therefore, high returns on the market cause low values of our cross-sectional beta premium, which might explain the strong link between market's valuation multiples and our cross-sectional measures.

Unreported experiments confirm that our results are not driven by this link. We first regressed λ^{SRC} on five annual lags of the annual compound return on the CRSP value-weight index. The coefficients in this regression are negative, but the R^2 is low at 12%. Then, we took the residuals of this regression, and compared them to the earnings yield and Fed model's forecast. Even after filtering out the impact of past market returns, the residuals of λ^{SRC} plot almost exactly on top of the Fed model's forecast, with a correlation of approximately 0.8. Furthermore, using the residuals of λ^{SRC} in place of λ^{SRC} in the subsequent predictability tests does not alter our conclusions. Thus, we conclude that our results are not driven by a mechanical link between the market's past returns and our cross-sectional measures.

Figure 2 also casts light on Franzoni's (2002), Adrian and Franzoni's (2002), and Campbell and Vuolteenaho's (2003) result that the betas of value stocks have declined relative to betas of growth stocks during our sample period. This trend has a natural explanation if the CAPM is approximately true and the expected equity premium has declined, as suggested by Fama and French (2002), Campbell and Shiller (1998), and others. Value stocks are by definition stocks with low prices relative to their cash-flow generating ability. On the one hand, if the market premium is large, it is natural that many high beta stocks have low prices, and thus end up in the value portfolio. On the other hand, if the market premium is near zero, there is no obvious reason to expect high beta stocks to have much lower prices than low beta stocks. If anything, if growth options are expected to have high CAPM betas, then growth stocks should have slightly higher betas. Thus, the downward trend in the market premium we document provides a natural explanation to the seemingly puzzling behavior of value and growth stocks' betas identified by Franzoni (2002) and others.

5.2 Univariate tests of predictive ability in the US sample

While the above illustrations show that the cross-sctional price of risk is highly correlated with reasonable ex-ante measures of the equity premium, it remains for us to

Table 1: Explaining the cross-sectional risk premium with the Fed model's equity premium forecast and smoothed earnings yield variables

The table shows the OLS regression of cross-sectional risk-premium measure, λ^{SRC}, on $\exp(ep)$ and $\exp(ep) - Y10$. λ^{SRC} is the Spearman rank correlation between valuation rank and estimated CAPM beta. Higher than average values of λ^{SRC} imply that high-beta stocks have lower than average prices and higher than average expected returns, relative to low-beta stocks. ep is the log ratio of S&P 500's ten-year moving average of earnings to S&P 500's price. $Y10$ is the nominal yield on ten-year constant-maturity taxable bonds in fractions. The OLS t-statistics (which do not take into account the persistence of the variables and regression errors) are in parentheses, and R^2 is adjusted for the degrees of freedom. The regression is estimated from the full sample period 1927:5-2002:12, 908 monthly observations.

Variable		Const.	$\exp(ep_t)$	$\exp(ep_t) - Y10_t$	adj. R^2
λ_t^{SRC}	$=$	-.1339	.0628	4.6292	71.87%
		(-10.54)	(.30)	(36.54)	
λ_t^{SRC}	$=$	-.3996	5.561		30.45%
		(-24.39)	(19.95)		
λ_t^{SRC}	$=$	-.1303		4.6536	71.90%
		(-34.02)		(48.19)	

29

show that our variable actually forecasts equity-premium realizations. Below, we use the new statistical tests introduced in Section 4 to conclusively reject the hypothesis that the equity premium is unforcastable based on our variables.

Table 2 shows descriptive statistics for the variables used in our formal predictability tests. To save space we only report the descriptive statistics for one cross-sectional risk-premium measure, λ^{SRC}. Recall that a high cross-sectional beta premium suggests that at that point in time high-beta stocks were cheap and low-beta stocks expensive. The correlation matrix in Table 2 shows clearly that the variation in the cross-sectional measure, λ^{SRC}, appears positively correlated with the log earnings yield, high overall stock prices coinciding with low cross-sectional beta premium. The term yield spread (TY) is a variable that is known to track the business cycle, as discussed by Fama and French (1989). The term yield spread is very volatile during the Great Depression and again in the 1970's. It also tracks λ^{SRC} with a correlation of .31 over the full sample.

Table 3 presents the univariate prediction results for the excess CRSP value-weight index return, and Table 4 for the excess CRSP equal-weight index return. The first panel of the tables forecasts the equity premium with the cross-sectional risk-premium measure λ^{SRC}. The second panel uses the log smoothed earnings yield (ep) and the third panel the term yield spread (TY) as the forecasting variable. The fourth panel shows regressions using alternative cross-sectional risk-premium measures. While the first three panels also show subperiod estimates, the fourth panel omits the subperiod results to save space.

The regressions of value-weight equity premium in Table 3 reveal that our cross-sectional risk-premium measures do forecast future market returns. For all measures except λ^{DPG}, we can reject the null hypothesis of a zero coefficient in favor of a positive coefficient with a p-value better than 1% in full-sample tests. Comparing the small-sample p-values to the usual critical values for t-statistics, it is clear that the usual t-test would perform adequately in this case. This is not surprising, since the correlation between equity-premium shocks and our cross-sectional forecasting-variable shocks is small in absolute value.

The subperiod results for λ^{SRC} show that the predictability is stronger in the first half of the sample than in the second half. The coefficient on λ^{SRC} drops from .0368 for the 1927:5-1965:2 period to 0.0088 for the 1965:2-2002:12 period. The similar drop is observed for other cross-sectional measures, except for λ^{ER}, which performs well in all subsamples (results unreported). However, the 95%-confidence intervals

Table 2: Descriptive statistics of the VAR state variables

The table shows the descriptive statistics estimated from the full sample period 1927:5-2002:12, 908 monthly observations. R_M^e is the excess simple return on the CRSP value-weight index. R_m^e is the excess simple return on the CRSP equal-weight index. λ^{SRC} is the Spearman rank correlation between valuation rank and estimated CAPM beta. Higher than average values of λ^{SRC} imply that high-beta stocks have lower than average prices and higher than average expected returns, relative to low-beta stocks. ep is the log ratio of S&P 500's ten-year moving average of earnings to S&P 500's price. TY is the term yield spread in percentage points, measured as the yield difference between ten-year constant-maturity taxable bonds and short-term taxable notes. "Stdev." denotes standard deviation and "Autocorr." the first-order autocorrelation of the series.

Variable	Mean	Median	Stdev.	Min	Max
$R_{M,t}^e$	.0062	.0095	.0556	-.2901	.3817
$R_{m,t}^e$	.0097	.0114	.0758	-.3121	.6548
λ_t^{SRC}	-.0947	-.1669	.2137	-.5272	.5946
ep_t	-2.8769	-2.8693	.3732	-3.8906	-1.4996
TY_t	.6232	.5500	.6602	-1.3500	2.7200

Correlations	$R_{M,t}^e$	$R_{m,t}^e$	λ_t^{SRC}	ep_t	TY_t
$R_{M,t}^e$	1	.9052	.1078	.0305	.0474
$R_{m,t}^e$	.9052	1	.1333	.0658	.0798
λ_t^{SRC}	.1078	.1333	1	.5278	.3120
ep_t	.0305	.0658	.5278	1	.2223
TY_t	.0474	.0798	.3120	.2223	1
$R_{M,t-1}^e$	.1048	.2052	.0825	-.0475	.0428
$R_{m,t-1}^e$	.1070	.2059	.1075	-.0010	.0726
λ_{t-1}^{SRC}	.0930	.1321	.9748	.5196	.3011
ep_{t-1}	.1140	.1509	.5359	.9923	.2279
TY_{t-1}	.0469	.0812	.3219	.2188	.9131

Table 3: Univariate predictors of excess value-weight CRSP return (R_M^e)
These are results from the model

$$R_{M,t}^e = \mu_1 + \theta x_{t-1} + e_t; \; x_t = \mu_2 + \rho x_{t-1} + u_t$$

with $\mathrm{E}\, e_t = \sigma_1^2$, $\mathrm{E}\, u_t = \sigma_2^2$, $\mathrm{Corr}\,(e_t, u_t) = \gamma$. "t-stat" is the usual t-statistic for testing the null that $\theta = 0$. The p-values and confidence intervals are based on the conditional critical functions. The p-value tests the null $\theta = 0$ against the one-sided alternative $\theta > 0$. The confidence interval is a two-sided interval for θ. The hatted variables are unrestricted OLS estimates.

Specification	$\widehat{\theta}$	t-stat	p-value	95% conf int	$\widehat{\rho}$	$\widehat{\gamma}$	$\widehat{\sigma}_1$	$\widehat{\sigma}_2$
Prediction by the cross-sectional beta premium, $x_t = \lambda_t^{SRC}$								
1927:5-2002:12	.0242	2.811	<.01	[.008, .041]	.975	.0773	.0553	.0477
1927:5-1965:2	.0368	2.450	<.01	[.008, .066]	.960	-.0152	.0633	.0562
1965:2-2002:12	.0088	.413	.309	[-.031, .054]	.931	.278	.0460	.0368
1927:5-1946:3	.0663	1.967	.030	[-.001,.131]	.934	-.0413	.0823	.0585
1946:3-1965:2	.0395	3.113	<.01	[.016, .065]	.957	.080	.0348	.0534
1965:2-1984:1	.0147	.6027	.240	[-.03, .065]	.942	.188	.0458	.0429
1984:1-2002:12	-.0190	-.4181	>.5	[-.099, .089]	.885	.416	.0462	.0292
Prediction by log smoothed earnings/price, $x_t = ep_t$								
1927:5-2002:12	.0170	3.454	.014	[.003, .024]	.993	-.669	.0552	.0464
1927:5-1965:2	.0317	3.282	.018	[.003, .046]	.987	-.671	.0630	.0549
1965:2-2002:12	.00756	1.319	>.5	[-.009, .011]	.996	-.668	.0459	.0359
1927:5-1946:3	.0410	2.670	.096	[-.006, .061]	.981	-.659	.0817	.0707
1946:3-1965:2	.0294	2.344	.168	[-.009, .043]	.994	-.727	.0351	.0322
1965:2-1984:1	.0204	1.817	.291	[-.012, .028]	.987	-.662	.0455	.0362
1984:1-2002:12	.0105	1.251	>.5	[-.012, .013]	.990	-.668	.0460	.0352
Prediction by term yield spread, $x_t = TY_t$								
1927:5-2002:12	.00396	1.413	.075	[-.001, .009]	.917	.0111	.0555	.269
1927:5-1965:2	.00489	1.015	.178	[-.005, .013]	.968	-.156	.0636	.151
1965:2-2002:12	.00270	.862	.150	[-.003, .008]	.871	.111	.0460	.346
1927:5-1946:3	.00497	.711	.316	[-.010, .017]	.969	-.174	.0829	.184
1946:3-1965:2	.0201	1.978	.030	[.000, .039]	.886	-.0707	.0352	.108
1965:2-1984:1	.00868	1.677	.043	[-.001, .019]	.765	.218	.0455	.378
1984:1-2002:12	-.00221	-.521	>.5	[-.010, .005]	.918	.00463	.0462	.301
Full sample predictive results for alternative cross-sectional measures:								
$x_t = \lambda_t^{REG}$	.0908	3.605	<.01	[.042, .141]	.937	.0644	.0552	.0255
$x_t = \lambda_t^{DP}$	.03539	2.53	<.01	[.007, .062]	.926	-.167	.0554	.0498
$x_t = \lambda_t^{DPG}$	.02419	1.75	.044	[-.003, .051]	.917	-.107	.0555	.0531
$x_t = \lambda_t^{BM}$	.001121	2.63	<.01	[.0003, .0019]	.942	-.236	.0554	1.440
$x_t = \lambda_t^{BMG}$	.001449	2.98	<.01	[.0005, .0023]	.919	-.222	.0553	1.500
$x_t = \lambda_t^{ER}$	2.175	3.15	<.01	[.80, 3.53]	.979	-.0331	.0459	.0005

Table 4: Univariate predictors of excess equal-weight CRSP return (R_m^e)
These are results from the model

$$R_{m,t}^e = \mu_1 + \theta x_{t-1} + e_t;\ x_t = \mu_2 + \rho x_{t-1} + u_t$$

with $\mathrm{E}\,e_t = \sigma_1^2$, $\mathrm{E}\,u_t = \sigma_2^2$, $\mathrm{Corr}\,(e_t, u_t) = \gamma$.   "t-stat" is the usual t-statistic for testing the null that $\theta = 0$. The p-values and confidence intervals are based on the conditional critical functions. The p-value tests the null $\theta = 0$ against the one-sided alternative $\theta > 0$. The confidence interval is a two-sided interval for θ. The hatted variables are unrestricted OLS estimates.

Specification	$\widehat{\theta}$	t-stat	p-value	95% conf int	$\widehat{\rho}$	$\widehat{\gamma}$	$\widehat{\sigma}_1$	$\widehat{\sigma}_2$
Prediction by the cross-sectional beta premium, $x_t = \lambda_t^{SRC}$								
1927:5-2002:12	.0469	4.012	<.01	[.025, .070]	.975	.0202	.0752	.0477
1927:5-1965:2	.0786	3.755	<.01	[.037, .119]	.960	-.0613	.0882	.0562
1965:2-2002:12	.0345	1.260	.100	[-.017, .092]	.931	.226	.0589	.0368
1927:5-1946:3	.147	3.041	<.01	[.048, .238]	.934	-.0960	.118	.0585
1946:3-1965:2	.0466	3.256	<.01	[.020, .075]	.957	.0603	.0392	.0534
1965:2-1984:1	.0457	1.357	.076	[-.017, .115]	.942	.139	.0633	.0429
1984:1-2002:12	.00571	.107	.405	[-.089, .132]	.885	.379	.0542	.0292
Prediction by log smoothed earnings/price, $x_t = ep_t$								
1927:5-2002:12	.0307	4.594	<.01	[.011, .040]	.993	-.683	.0750	.0464
1927:5-1965:2	.0662	4.943	<.01	[.026, .086]	.987	-.683	.0872	.0549
1965:2-2002:12	.0104	1.420	.470	[-.011, .015]	.996	-.689	.0589	.0359
1927:5-1946:3	.0839	3.833	.014	[.016, .112]	.981	-.683	.117	.0707
1946:3-1965:2	.0285	2.004	.250	[-.015, .045]	.993	-.707	.0398	.0322
1965:2-1984:1	.0290	1.866	.300	[-.017, .039]	.987	-.705	.0631	.0362
1984:1-2002:12	.00131	.132	>.5	[-.026, .004]	.990	-.684	.0542	.0352
Prediction by term yield spread, $x_t = TY_t$								
1927:5-2002:12	.00935	2.452	<.01	[.002, .016]	.915	.0140	.0756	.269
1927:5-1965:2	.0106	1.572	.074	[-.003, .023]	.968	-.151	.0893	.151
1965:2-2002:12	.00774	1.933	.026	[.00, .015]	.871	.124	.0588	.346
1927:5-1946:3	.00867	.857	.243	[-.013, .026]	.969	-.171	.120	.186
1946:3-1965:2	.0208	1.808	.043	[-.002, .043]	.886	-.0699	.0398	.108
1965:2-1984:1	.0172	2.410	<.01	[.004, .032]	.765	.184	.0628	.378
1984:1-2002:12	.00420	.844	.138	[-.005, .013]	.918	.0809	.0541	.301
Full sample predictive results for alternative cross-sectional measures:								
$x_t = \lambda_t^{REG}$	.165	4.811	<.01	[.098, .232]	.937	.0238	.0749	.0255
$x_t = \lambda_t^{DP}$	.07916	4.17	<.01	[.041, .115]	.926	-.206	.0751	.0498
$x_t = \lambda_t^{DPG}$	.06813	3.63	<.01	[.031, .104]	.917	-.136	.0753	.0531
$x_t = \lambda_t^{BM}$	.002609	4.51	<.01	[.0015, .0037]	.942	-.245	.075	1.440
$x_t = \lambda_t^{BMG}$	.003129	4.76	<.01	[.0018, .0043]	.919	-.242	.0749	1.500
$x_t = \lambda_t^{ER}$	3.371	3.74	<.01	[1.55, 5.12]	.979	-.0697	.0599	.0005

suggest that one should not read too much into these subperiod estimates. The point estimate for the first subperiod is contained within the confidence interval of the second subperiod, and the point estimate of the second subperiod within the confidence interval of the first subperiod. Furthermore, for every subperiod we have examined, a positive coefficient is contained within the 95%-confidence intervals.

Of the two extant instruments we study, the log smoothed earnings yield is the stronger forecaster of the equity premium, while the term yield spread has only weak predictive ability. Consistent with economic logic, the coefficient on ep is positive for all subsamples, and the t-statistic testing the null of no predictability is 3.45 for the full sample. Our new statistical methodology maps this t-statistic to a one-sided p-value 1.4%. Notice that while the t-statistic on ep is higher than on our first cross-sectional measure λ^{SRC} (2.81 vs. 3.45), the p-value for ep is higher than the p-value for λ^{SRC}. Of course this is the motivation for our econometric work in Section 4; the earnings-yield is quite persistent, and its shocks are strongly negatively correlated with equity-premium shocks, making standard statistical inference misleading.

As with most return-prediction exercises, the results for equal-weight index are a more extreme version of those for value-weight index. Table 4 shows that our main cross-sectional measure, λ^{SRC}, forecasts monthly excess equal-weight returns with a t-statistic of 4.01. Similarly high t-statistics are obtained for the earnings yield (4.59) and alternative cross-sectional mesures (ranging from 3.63 to 4.81), while the term yield spread's is slightly lower (2.45). All t-statistic imply rejection of the null at a better-than-1% level, even after accounting for the problems due to persistent and correlated regressors.

In unreported tests, we also examined the robustness of our results to heteroskedasticity. To correct the critical vlaues for possible heteroskedasticity, we carried out conditional inference based on t-statistics computed with Eicker-Huber-White (White, 1980) standard errors. The Eicker-Huber-White t-statistics weaken the case for predictability. When predicting the value-weight CRSP excess return over the entire sample, λ^{SRC} is significant for a one-side test but not a two-sided test (the one-sided p-value is .032). ep and TY are not statistically significant predictors, with one-sided p-values of .162 and .095. Although the combination of Eicker-Huber-White standard errors and conditional inference appears sensible, these tests come with a caveat: The conditional distribution of the Eicker-Huber-White t-statistic has not been studied, and it is not known whether the conditional Eicker-Huber-White t-statistic is robust to heteroskedasticity.

Another way of addressing the issue of heteroskedasticity is to note that stock returns were very volatile during the Great Depression. A simple check for undue influence of heteroskedasticity is to simply omit this volatile period from estimation. When we estimate the model and p-values from the 1946-2002 sample, λ^{SRC} remains statistically significant predictor at a better than 1% level, while the log earnings yield is no longer significant even at the 10% level. Thus, these two experiments both suggest that our cross-sectional forecasting variables have predictive ability, and that our results are not just an illusion created by heteroskedasticity.

5.3 Univariate tests of predictive ability in the international sample

We also examine the predictive ability of cross-sectional risk-premium measures in an international sample and obtain similar predictive results as in the US sample. Because of data constraints (we only have portfolio-level data for our international sample), we define our cross-sectional risk premium measure as the difference in the local-market beta between value and growth portfolios. We work with value and growth portfolios constructed by Kenneth French and available on his web site, focusing on the top-30% and bottom-30% portfolios sorted on four of Morgan Stanley Capital International's value measures: D/P, BE/ME, E/P, and C/P. We then estimate the betas for these portfolios using a 36-month rolling window, and define the predictor variable λ^{MSCI} as the average beta of the four value portfolios minus the average beta of the four growth portfolios.

If for example the CAPM holds, the beta difference between two dynamic trading strategies, a low-multiple value portfolio and a high-multiple growth portfolio, is a natural measure of the expected equity premium. The underlying logic is perhaps easiest to explain in a simple case in which individual stocks' growth opportunities and betas are constant for each stock and cross-sectionally uncorrelated across stocks. During years when the expected equity premium is high, the high beta stocks will have low prices (relative to current cash-flow generating ability) and will thus mostly be sorted into the value portfolio. Symmetrically, low beta stocks will have relatively high prices and those stocks will mostly end up in the "growth" or high-multiple portfolio. Consequently, a high expected equity premium will cause the value portfolio's beta to be much higher than that of the growth portfolio. In contrast, during years when the expected equity premium is low, multiples are determined primarily

by growth opportunities. The high beta and low beta stocks will have approximately the same multiples and are thus approximately equally likely to end up in either the low-multiple value portfolio or the high-multiple growth portfolio. Thus during years when the expected equity premium is low, the beta difference between value and growth portfolio should be small. This simple logic allows us to construct a cross-sectional risk-premium proxy without security-level data.

We find that the past local-market beta of value-minus-growth is generally a statistically significant predictor of the future local-market equity premium. In the individual country regressions of Table 5, 17 out of 22 countries have the correct sign in the associated local-market equity premium prediction regression, with nine out of 22 estimates statistically significant at the 10% level. Moreover, the five negative estimates are not measured precisely. Finally, the parameter estimates for all the countries are similar to those obtained for the US (with the exception of Mexico with its extremely short sample).

In addition to the country-by-country regressions, we also pool the data and thus constrain the regression coefficients to be equal across countries. Fortunately, our pooled regression specification does not suffer significantly from the usual problems associated with equity-premium prediction regressions. This is because of two reasons: First, the shocks to the predictor variable are largely uncorrelated with the return shocks. In fact, the correlation point estimates are close to 0.05, suggesting that the usual asymptotic test is slightly conservative. Second, even if the shocks for a given country were negatively correlated, the cross-sectional dimension in the data set lowers the pooled correlation between the predictor variable and past return shocks.

However, the usual OLS standard errors (and hypothesis tests based on them) suffer from another problem. The OLS standard errors ignore the potential cross-correlation between the residuals. To deal with this problem, we compute standard errors that cluster by cross-section. Our Monte Carlo experiments show that for our parameter values, clustered standard errors provide a slightly conservative hypothesis test.

Table 6 shows that we can reject the null hypothesis of no predictability in favor of the alternative that the betas of the country-specific value-minus-growth portfolios are positively related to the country-specific expected equity premiums. This conclusion is robust to inclusion/exclusion of the US data and inclusion/exclusion of country fixed effects in the pooled regression. All p-values are under 5%. Thus we conclude

Table 5: Predicting the equity premium, country-by-country regressions
These are results from the model

$$R^e_{M,t,i} = \mu_{1,i} + \theta_i x_{t-1,i} + e_{t,i};\ x_{t,i} = \mu_{2,i} + \rho_i x_{t-1,i} + u_{t,i}$$

with $\text{Corr}(e_t, u_t) = \gamma$. $x_{t,i} = \lambda^{MSCI}_{t,i}$ for country i in year t. $\lambda^{MSCI}_{t,i}$ is constructed by taking the top-30% and bottom-30% portfolios sorted on four of Morgan Stanley Capital International's value measures: D/P, BE/ME, E/P, and C/P. We then estimate the betas for these portfolios using a three-year rolling window, and define the predictor variable λ^{MSCI} as the average beta of the four value portfolios minus the average beta of the four growth portfolios. The dependent variable in the regressions is the local-market equity premium, for which the stock market returns are from Kenneth French's files and the local risk-free returns are from Global Financial Data. The regressions are estimated using country-by-country OLS regressions. The "OLS t-stat" is the homoskedastic t-statistic for testing the null that $\theta = 0$. "White t-stat" is the t-statistic robust to heteroskedasticity. The p-values in parentheses are based on the conditional critical functions and test the null $\theta = 0$ against the one-sided alternative $\theta > 0$. The hatted variables are unrestricted OLS estimates.

Country	Time period	Obs.	$\widehat{\theta}$	OLS t-stat		White t-stat		$\widehat{\rho}$	$\widehat{\gamma}$
Australia	1975:1 - 2001:12	324	.0237	1.70	(.111)	1.60	(.132)	.989	-.229
Austria	1987:1 - 2001:12	180	-.0251	-.61	(.584)	-.48	(.540)	.935	.241
Belgium	1975:1 - 2001:12	324	.0234	1.36	(.064)	1.43	(.056)	.972	.196
Denmark	1989:1 - 2001:12	156	.0149	.78	(.240)	.78	(.238)	1.02	.027
Finland	1988:1 - 2001:12	168	-.0150	-.81	(.710)	-.78	(.701)	1.00	.151
France	1975:1 - 2001:12	324	.0444	2.08	(.028)	2.08	(.028)	1.00	.033
Germany	1975:1 - 2001:12	324	.0226	1.32	(.078)	1.25	(.089)	.985	.018
Hong Kong	1975:1 - 2001:12	324	.0200	.50	(.278)	.49	(.282)	.977	.11
Ireland	1991:1 - 2001:12	132	.0100	.39	(.374)	.36	(.386)	.911	-.061
Italy	1975:1 - 2001:12	324	.0268	.92	(.233)	.92	(.234)	1.01	.081
Japan	1975:1 - 2001:12	324	.0172	1.40	(.095)	1.66	(.058)	.992	-.037
Malaysia	1994:1 - 2001:10	94	.0418	.81	(.089)	.76	(.096)	.918	.306
Mexico	1982:1 - 1987:12	72	.3490	1.41	(.044)	1.50	(.036)	.844	.311
Netherland	1975:1 - 2001:12	324	-.0061	-.37	(.596)	-.30	(.573)	.984	.105
New Zealand	1988:1 - 2001:12	168	.0456	1.95	(.018)	1.98	(.017)	.959	.023
Norway	1975:1 - 2001:12	324	-.0053	-.54	(.708)	-.50	(.697)	.994	-.024
Singapore	1975:1 - 2001:12	324	.0159	.76	(.201)	.66	(.229)	.977	.094
Spain	1975:1 - 2001:12	324	.0366	2.76	(.004)	2.79	(.004)	.986	-.051
Sweden	1975:1 - 2001:12	324	.0177	1.57	(.046)	1.37	(.068)	1.01	.019
Switzerland	1975:1 - 2001:12	324	-.0025	-.15	(.552)	-.15	(.552)	.974	.030
UK	1975:1 - 2001:12	324	.0115	.53	(.341)	.44	(.372)	.971	-.153
US	1926:7 - 2002:12	918	.0166	2.41	(.008)	2.02	(.019)	.993	.089

Table 6: Predicting the equity premium, pooled international regressions
These are results from the model

$$R^e_{M,t,i} = \mu_{1,t,i} + \theta x_{t-1,i} + e_{t,i}; \; x_{t,i} = \mu_{2,t,i} + \rho x_{t-1,i} + u_{t,i}$$

with Corr $(e_{t,i}, u_{t,i}) = \gamma$. $x_{t,i} = \lambda^{MSCI}_{t,i}$ for country i in year t. $\lambda^{MSCI}_{t,i}$ is constructed by taking the top-30% and bottom-30% portfolios sorted on four of Morgan Stanley Capital International's value measures: D/P, BE/ME, E/P, and C/P. We then estimate the betas for these portfolios using a three-year rolling window, and define the predictor variable λ^{MSCI} as the average beta of the four value portfolios minus the average beta of the four growth portfolios. The dependent variable in the regressions is the local-market equity premium, for which the stock market returns are from Kenneth French's files and the local risk-free returns are from Global Financial Data. "FE" denotes fixed effects, meaning we estimate different intercepts $\mu_{1,t,i}$ and $\mu_{2,t,i}$ for each country or each country and time point. "No FE" indicates that we estimate a common intercept for all countries and time points. "tstat, homoskedastic" and "tstat, heteroskedastic" indicate the usual ols t-statistic and the White t-statistic which is robust to heteroskedasticity. "tstat, clustering by year" indicates that we calculated standard errors robust to correlations between firms, but assumed independence over time. The p-values in parentheses test the null $\theta = 0$ against the one-sided alternative $\theta > 0$. p-values are based on the usual standard normal approximation to the null distribution of a t-statistic. The hatted variables are unrestricted OLS estimates.

	No FE		Country FE		Country, time FE	
	All	Excl. US	All	Excl. US	All	Excl. US
$\widehat{\theta}$	.0102	.0090	.0132	.0123	.00961	.00756
t-stat, homoskedastic	3.21	2.53	3.76	3.09	3.32	2.34
p-val, homoskedastic	(.0007)	(.0057)	(.0001)	(.0010)	(.0004)	(.0010)
t-stat, heteroskedastic	2.69	2.12	3.31	2.73	2.97	2.14
p-val, heteroskedastic	(.0036)	(.0171)	(.0005)	(.0032)	(.0015)	(.016)
t-stat, clustering by year	2.08	1.65	2.31	1.89	2.57	1.78
p-val, clustering by year	(.0189)	(.0494)	(.0105)	(.0295)	(.0051)	(.0376)
$\widehat{\rho}$	.992	.992	.990	.990	.988	.987
$\widehat{\gamma}$	.0519	.0469	.0545	.0497	.0578	.0483

that our simple proxy, λ^{MSCI}, predicts equity premium realizations in a sample largely independent of our main US sample, as well as in the US sample.

5.4   Multivariate predictability tests

The above tests demonstrate that our new cross-sectional variables can forecast the equity premium. In this section, we perform multivariate tests to see whether the predictive information in our new variables subsume or is subsumed by that in the earnings yield and term yield spread. We show the results from these "horse races" for the value-weight index in Table 7. Unreported results for the equal-weight index are similar but statistically stronger.

The horse race between λ^{SRC} and ep is a draw. In regressions forecasting the value-weight return over the full period, we fail to reject at the 5% level of significance the hypothesis that λ^{SRC} has no predictive ability independent of ep (p-value 15.8%). Likewise, we cannot reject the hypothesis that ep has no predictive ability controlling for λ^{SRC} (p-value 10.8%). Since these p-values are relatively close to 10% for both the earnings yield and our cross-sectional measures, we are cautious about drawing clear conclusions about the independent predictive ability of these variables.

Though the horse race between λ^{SRC} and ep is a draw, many of our alternative cross-sectional measures wins their respective races with ep. When λ^{REG}, λ^{BMG}, and λ^{ER} are raced against ep, the above conclusions change: We now fail to reject the hypothesis that ep has no independent predictive power (p-values ranging from 7.8% to 28.4%), but do reject the hypothesis that λ^{REG}, λ^{BMG}, and λ^{ER} have no independent predictive power (p-values ranging from 1.5% to 5.0%). These conclusions change slightly in unreported equal-weighted regressions, in which the dependent variable is the future excess return on an equal-weight portfolio of stocks. For all combinations, both the cross-sectional risk premium and the market's earnings yield are statistically significant. Our result that equal-weight returns are more predictable is consistent with results in the previous literature.

The term yield spread is unimpressive in multiple regressions. All other variables beat the term yield spread, and TY is insignificant even in most regressions that forecast the equal-weight equity premium.

Table 7: Multivariate predictors of excess value-weight CRSP return (R_M^e)
$\widehat{\theta}_i$ is the OLS estimate of θ_i, with $\theta_i = (\theta_1 \; \theta_2)'$ in the model

$$R_{M,t}^e = \mu_1 + \theta' \mathbf{x}_{t-1} + u_t; \; \mathbf{x}_t = \boldsymbol{\mu}_2 + K\mathbf{x}_{t-1} + V_t.$$

"t-stat" is the usual t-statistic for testing the null $\theta_i = 0$. "p-val" denotes the p-value for testing the null $\theta_i = 0$ against $\theta_i > 0$. "95% conf" denotes a two-sided confidence interval. "F p-val" denotes the p-value for the F-test of the null $\theta_1 = \theta_2 = 0$. All p-values are computed using the conditional bootstrap described in Appendix B.

Specification	$\widehat{\theta}_1$	t-stat	p-val	95% conf	$\widehat{\theta}_2$	t-stat	p-val	95% conf	F p-val
Prediction equation: $R_{M,t}^e = \theta_0 + \theta_1 \lambda_{t-1}^{SRC} + \theta_2 ep_{t-1} + u_t$									
1927:5-2002:12	.012	1.17	.158	[-.009,.031]	.013	2.32	.108	[-.004,.021]	.020
1927:5-1965:2	.001	.045	.532	[-.043,.041]	.031	2.17	.091	[-.010,.049]	.056
1965:2-2002:12	.011	.527	.285	[-.030,.057]	.008	1.36	.447	[-.009,.013]	.592
1927:5-1946:3	-.002	-.030	.573	[-.118,.098]	.042	1.79	.176	[-.029,.072]	.139
1946:3-1965:2	.037	2.03	.038	[.001,.071]	.004	.225	.798	[-.050,.023]	.086
1965:2-1984:1	.027	1.08	.183	[-.027,.074]	.023	2.03	.172	[-.013,.037]	.210
1984:1-2002:12	-.034	-.738	.778	[-.126,.057]	.012	1.39	.341	[-.014,.021]	.527
Prediction equation: $R_{M,t}^e = \theta_0 + \theta_1 \lambda_{t-1}^{SRC} + \theta_2 TY_{t-1} + u_t$									
1927:5-2002:12	.023	2.49	.006	[.006,.040]	.002	.564	.289	[-.003,.007]	.017
1927:5-1965:2	.038	2.23	.017	[.005,.068]	-.001	-.132	.604	[-.011,.008]	.051
1965:2-2002:12	.005	.233	.357	[-.036,.052]	.003	.791	.184	[-.003,.009]	.670
1927:5-1946:3	.069	1.84	.043	[-.008,.138]	-.001	-.158	.634	[-.017,.012]	.160
1946:3-1965:2	.035	2.55	.006	[.009,.060]	.010	.912	.201	[-.012,.030]	.005
1965:2-1984:1	.010	.398	.280	[-.035,.063]	.008	1.61	.038	[.000,.020]	.229
1984:1-2002:12	-.012	-.257	.524	[-.098,.095]	-.002	-.403	.633	[-.010,.007]	.849
Full sample results for alternative cross-sectional risk premium measures									
Prediction equation: $R_{M,t}^e = \theta_0 + \theta_1 x_{t-1} + \theta_2 ep_{t-1} + u_t$									
$x_t = \lambda_t^{REG}$	.066	2.36	.015	[.010,.124]	.012	2.12	.151	[-.005,.019]	.003
$x_t = \lambda_t^{DP}$	.0200	1.32	.107	[-.012,.052]	.0143	2.69	.056	[-.001,.022]	.005
$x_t = \lambda_t^{DPG}$	.0076	.516	.321	[-.022,.039]	.0160	3.02	.027	[.001,.024]	.010
$x_t = \lambda_t^{BM}$	.0005	1.03	.174	[-.001,.002]	.0140	2.46	.062	[-.002,.023]	.007
$x_t = \lambda_t^{BMG}$	.0009	1.71	.050	[.000,.002]	.0131	2.44	.078	[-.003,.021]	.003
$x_t = \lambda_t^{ER}$	1.766	2.40	.019	[.082,3.37]	.0081	1.57	.284	[-.007,.015]	.011
Full sample results for alternative cross-sectional risk premium measures									
Prediction equation: $R_{M,t}^e = \theta_0 + \theta_1 x_{t-1} + \theta_2 TY_{t-1} + u_t$									
$x_t = \lambda_t^{REG}$	.088	3.35	.000	[.038,.138]	.001	.474	.316	[-.003,.007]	.001
$x_t = \lambda_t^{DP}$	.0335	2.38	.013	[.006,.064]	.0032	1.13	.136	[-.003,.009]	.017
$x_t = \lambda_t^{DPG}$	.0215	1.53	.066	[-.007,.049]	.0032	1.13	.129	[-.003,.009]	.121
$x_t = \lambda_t^{BM}$	.0010	2.39	.012	[.001,.002]	.0026	.910	.184	[-.003,.008]	.028
$x_t = \lambda_t^{BMG}$	.0014	2.82	.003	[.001,.002]	.0029	1.03	.159	[-.003,.008]	.006
$x_t = \lambda_t^{ER}$	2.154	3.08	.001	[.700,3.73]	.0005	.172	.400	[-.005,.006]	.006

5.5 Implications of premia divergence in the 1980's

Across specifications, our cross-sectional beta-premium variables show their poorest performance as predictors of the equity premium in the second subsample, especially in the 1980's. Curiously, as Figure 2 shows, the second subsample also exhibits occasionally large divergences between the market's smoothed earnings yield and the cross-sectional beta premium. For example, in 1982 both our cross-sectional measures and the Fed model forecast a low equity premium, while the smoothed earnings yield forecasts a high equity premium.

If ep is indeed a good predictor of market's excess return and λ^{SRC} of the return of high-beta stocks relative to that of low-beta stocks, the divergence implies a trading opportunity. In 1982, an investor could have bought the market portfolio of stocks (which had a high expected return), and then hedged this investment by a zero-investment portfolio long low-beta stocks and short high-beta stocks (which had a low expected return). At this time, this hedged market portfolio should have had a high expected return relative to both its systematic and unsystematic risk.

We test this hypothesis by constructing a zero-investment portfolio consisting of 1.21 times the CRSP VW excess return, minus the return difference between the highest-beta (10) and lowest-beta (1) deciles. The beta-decile portfolios are formed on past estimated betas, value weighted, and rebalanced monthly. We picked the coefficient 1.21 to give the portfolio an approximately zero unconditional beta. The excess return on this beta-hedged market portfolio is denoted by R^e_{arb}.

Table 8 confirms this implication of premia difference. When we forecast the beta-hedged market return with λ_{SRC} and ep, the former has a negative coefficient and the latter a positive coefficient (although ep's t-statistic is only 1.13.) The variables are jointly significant for the full period as well as for both subperiods. However, since λ_{SRC} and ep are so highly correlated in the first subsample, the identification for the partial regression coefficients must come from the second sample. Consistent with this conjecture, the nulls for both variables are rejected at a better than 10% level in the second subsample, while the p-values are consderably higher in the first subsample. Similar conclusions can be drawn from regressions that use other measures of cross-sectional beta premium.

Even a cursory examination of the fitted values suggests that these predictability results are also economically significant. In the beginning of year 1982, the pre-

Table 8: Multivariate predictors of hedged value-weight CRSP return (R^e_{arb})
$\widehat{\theta}_i$ is the OLS estimate of θ_i, with $\theta_i = (\theta_1\ \theta_2)'$ in the model

$$R^e_{arb,t} = \mu_1 + \theta' \mathbf{x}_{t-1} + u_t; \ \mathbf{x}_t = \boldsymbol{\mu}_2 + K\mathbf{x}_{t-1} + V_t.$$

R^e_{arb} is the return on a zero-investment portfolio consisting of 1.21 times the CRSP VW excess return, minus the return difference between the highest-beta (10) and lowest-beta (1) deciles. "t-stat" is the usual t-statistic for testing the null $\theta_i = 0$. "p-val" denotes the p-value for testing the null against alternatives $\theta_1 < 0$ and $\theta_2 > 0$. "95% conf" denotes a two-sided confidence interval. "F p-val" denotes the p-value for the F-test of the null $\theta_1 = \theta_2 = 0$. All p-values are computed using the conditional bootstrap described in Appendix B.

Specification	$\widehat{\theta}_1$	t-stat	p-val	95% conf	$\widehat{\theta}_2$	t-stat	p-val	95% conf	F p-val
Prediction equation: $R^e_{arb,t} = \theta_0 + \theta_1 \lambda^{SRC}_{t-1} + \theta_2 ep_{t-1} + u_t$									
1927:6-2002:12	-.030	-2.87	.002	[-.049, -.011]	.007	1.13	.140	[-.004, .017]	.013
1927:6-1965:2	-.018	-.965	.161	[-.056, .018]	-.014	-1.10	.877	[-.037, .008]	.018
1965:2-2002:12	-.063	-2.29	.016	[-.114, -.004]	.010	1.34	.108	[-.004, .023]	.022
1927:6-1946:3	.004	.087	.522	[-.082, .088]	-.018	-.907	.821	[-.059, .021]	.440
1946:3-1965:2	-.016	-1.05	.150	[-.046, .014]	.009	.600	.486	[-.026, .032]	.605
1965:2-1984:1	-.070	-2.75	.006	[-.111, -.017]	.007	.599	.300	[-.017, .029]	.012
1984:1-2002:12	-.083	-1.19	.121	[-.213, .065]	.025	1.89	.028	[.000, .052]	.126
Full sample results for alternative cross-sectional risk premium measures									
Prediction equation: $R^e_{arb,t} = \theta_0 + \theta_1 x_{t-1} + \theta_2 ep_{t-1} + u_t$									
$x_t = \lambda^{REG}$	-.062	-2.32	.011	[-.110, -.012]	.001	.467	.305	[-.003, .007]	.062
$x_t = \lambda^{DP}$	-.072	-4.66	.000	[-.103,-.042]	.0073	1.34	.099	[-.004,.018]	.000
$x_t = \lambda^{DPG}$	-.066	-4.36	.000	[-.095,-.036]	.0064	1.19	.130	[-.005,.017]	.000
$x_t = \lambda^{BM}$	-.0029	-5.80	.000	[-.004,-.002]	.0146	2.53	.007	[.002,.026]	.000
$x_t = \lambda^{BMG}$	-.0029	-5.31	.000	[-.004,-.002]	.0098	1.79	.041	[-.001,.020]	.000
$x_t = \lambda^{ER}$	-2.21	-2.61	.005	[-3.82,-.615]	.0114	1.94	.036	[.000,.023]	.019

dicted value for R_{arb}^e is over 20% annualized in the regression that uses λ_{SRC} and ep as forecasting variables. Since the unconditional volatility of R_{arb}^e is under 20% annualized (and various conditional volatility estimates even lower), the fitted values imply a conditional annualized Sharpe ratio of over one at the extreme point of the divergence. In summary, the evidence in Table 8 clearly shows that divergence of λ_{SRC} and ep creates a both economically and statistically significant trading opportunity for an investor who can borrow at the Treasury-bill rate. An alternative but equivalent way to describe our results is that the zero-beta rate in the universe of stocks deviates predictably from the Treasury-bill rate.

6   Conclusions

This paper tells a coherent story connecting the cross-sectional properties of expected returns to the variation of expected returns through time. We use the simplest risk model of modern portfolio theory, the Sharpe-Lintner CAPM, to relate the cross-sectional beta premium to the equity premium. When the cross-sectional beta premium is high, Sharpe-Lintner CAPM predicts that the equity premium should also be expected to be high.

We construct a class of cross-sectional beta-premium variables by measuring the cross-sectional association between valuation multiples (book-to-price, earnings yield, etc.) and estimated betas. Consistent with the Sharpe-Lintner CAPM, our time-series tests show that the cross-sectional beta premium is highly correlated with the market's yield measures. Furthermore, the cross-sectional variable forecasts the equity premium, both on its own and in a multiple regression with smoothed earnings yield, although the high correlation between the two variables makes the multiple-regression results less conclusive. Results obtained from an international sample support our main conclusions drawn from the US sample.

Since equity-premium realizations are very noisy, forecasting the equity premium with univariate methods is a nearly impossible task. Fortunately, simple economic logic makes predictions about the equity premium, such as high stock prices should imply a low equity premium (Campbell and Shiller, 1988a, Fama and French, 1989), the equity premium should usually be positive because of risk aversion (Merton, 1980), and the cross-sectional pricing of risk should be consistent with the time-series pricing of risk. We join others in arguing that imposing such economically reasonable

guidelines can be of great practical utility in formulating reasonable equity-premium forecasts.

Beyond simply forecasting the equity premium, our results provide insight into the process by which the market prices the cross-section of equities. According to our estimates, the stock market prices one unit of beta in the cross-section with a premium that is equal to the equity premium derived from "the Fed model," the earnings yield minus the long-term bond yield. In our sample, the Fed model explains 72% of the time-series variation in our main cross-sectional risk-price measure. Of course, our claim is not that one should use the CAPM and the Fed model for relative valuation of stocks. We merely document that the cross-section prices are set approximately as if the market participants did so.

We also provide a practical solution to a long-standing inference problem in financial econometrics. A volume of studies have asked the question whether the equity premium can be predicted by financial variables such as the dividend or earnings yield (Rozeff, 1984, Keim and Stambaugh, 1986, Campbell and Shiller, 1988ab, Fama and French, 1988 and 1989, Hodrick, 1992, and others). Although the usual asymptotic p-values indicate a statistically reliable predictability, Stambaugh (1999) notes that the small-sample inference is complicated by two issues. First, the predictor variable is often very persistent, and, second, the shocks to the predictor variable are correlated with the unexpected component of the realized equity premium. Together, these two issues can cause large small-sample size distortions in the usual tests. Consequently, elaborate simulation schemes (e.g., Ang and Bekaert, 2001) have been necessary for finding reasonably robust p-values even in the case of Gaussian errors.

We solve for the exact small-sample p-values using a novel method. The method is based on Jansson and Moreira's (2002) idea of reducing the data to a sufficient statistic, and then creating the nonlinear mapping from the sufficient statistic to the correct critical value for the OLS t-statistic. For a single forecasting variable and the now-usual setup proposed by Stambaugh (1999), we provide the community with a function that enables an applied finance researcher to implement a correctly-sized test of predictability in seconds.

7 Technical Appendix A: Algorithm for computing q_α^{nn}

In this appendix we describe the algorithm for computing q_α^{nn}, the neural network approximation to the critical values. We choose the parameters of the neural net to minimize the sum of squared size distortions over a grid of (ρ, γ) values:

$$\left(\widehat{\psi}, \widehat{\phi}\right) = \operatorname{argmin}_{(\psi,\phi)} \sum_{i=1}^{N} \left(\alpha - B^{-1} \sum_{b=1}^{B} 1_h \left[\widehat{t}_{b,i} > q_\alpha^{nn}(X_{b,i}, \psi, \phi)\right]\right)^2. \qquad (23)$$

$i = 1, ..., N$ indexes a grid of (ρ_i, γ_i) pairs. For each i, we simulate B data sets from the null model, with $\theta = 0$, iid normal errors, $\mu_1 = \mu_2 = 0$, and $\sigma_u^2 = \sigma_v^2 = 1$. $\widehat{t}_{b,i}$ is the t-statistic for the bth simulated sample generated from (ρ_i, γ_i), and $X_{b,i}$ is the X vector generated from this sample. We can estimate the rejection frequency based on the critical value q_α^{nn} by averaging over the simulated draws:

$$\operatorname{Pr}\left[\widehat{t} > q_\alpha^{nn}(X, \psi, \phi) \; ; \; \rho_i, \gamma_i\right] \approx B^{-1} \sum_{b=1}^{B} 1\left[\widehat{t}_{b,i} > q_\alpha^{nn}(X_{b,i}, \psi, \phi)\right]. \qquad (24)$$

$1(x)$ is the indicator function, equal to 1 when $x \geq 0$ and zero otherwise. Thus our minimization problem is a simulation-based way to minimize the sum of squared size distortions. Since the indicator function is not differentiable, we replace it with the differentiable function

$$1_h(x) = \left(1 + e^{-x/h}\right)^{-1}. \qquad (25)$$

As h goes to zero, 1_h converges pointwise to the indicator function. Since our objective function is differentiable in ψ and ϕ, we can use efficient minimization methods.

The neural net critical values used in this paper were computed setting $B = 20000$ and $h = .01$. The grid points were all possible combinations of $\rho \in \boldsymbol{\rho}$ and $\gamma \in \Gamma$, where

$$\boldsymbol{\rho} = (-100, -75, -50, -30, -20, -15, -12, -10, -8, -6, -4, -2, 0)/T \qquad (26)$$
$$\Gamma = (0, -.2, -.4, -.5, -.6, -.7, -.8, -.85, -.86, -.88, -.90, -.92, -.94, -.96, -.98).$$

We do not need to simulate over different values of μ_1, μ_2, σ_u^2 or σ_v^2 because both $\widehat{t}_{b,i}$ and $X_{b,i}$ are exactly invariant to these parameters.

Since we only simulated over negative correlations, q_α^{nn} is valid only when $\gamma < 0$. When $\widehat{\gamma} > 0$, we replace x_t by $-x_t$, thus reversing the sign of the estimated correlation. This transformation also reverses the sign of θ, so instead of testing the null $\theta = 0$ against the positive alternative $\theta > 0$, we instead test the null $\theta = 0$ against the negative alternative $\theta < 0$. Instead of rejecting when the t-statistic is greater than the 95% quantile of the conditional null distribution, we reject when the transformed t-statistic is less than the 5% quantile of the conditional null.

It is well known that minimizing objective functions in neural networks is computationally demanding. The objective function is not convex in the parameters and has many local minima. We used the following algorithm, which draws on suggestions in Bishop (1995) (chapter 7), White (1992) (chapter 11) and Masters (1993) (chapter 9). After generating all the X and $\widehat{t}$ values, we standardize them to have zero sample means and unit variances. Following Bishop (1995) (page 262), we randomly draw each element of ϕ from an independent Normal$(0, 1/2)$ distribution. Given ϕ, the neural network is a linear function of ψ. Therefore we generate sensible initial estimates of ψ by linear quantile regression of $\widehat{t}_{b,i}$ on $(1,\ g(\phi_1' X_{b,i}),\ ...,\ g(\phi_5' X_{b,i}))$ (see Koenker and Portnoy (1997)). We then iterate from the starting values for ϕ and ψ using the Broyden-Fletcher-Goldfarb-Shanno optimization algorithm. All computations were done using Ox 3.00, a programming language described in Doornik (2001), and quantile regression programs provided by Roger Koenker and Daniel Morillo.

We repeated this algorithm for many different randomly drawn starting values for ϕ. Most of the starting values led to solutions with small size distortions - the rejection frequencies were visually quite similar to those in figure 2. A few of the starting values converged at parameters that did not lead to accurate solutions.

We have fit neural nets for various sample sizes n and quantiles α. Please contact Samuel Thompson at sthompson@harvard.edu for Matlab code that uses the fitted nets to carry out conditional testing. As an example we provide the nets used to construct Figure 1, where $n = 120$. The approximation to the .95 conditional quantile

46

is

$$q_{.95}^{nn}(X, \widehat{\psi}, \widehat{\phi}) \equiv .6513 - 1.0909g \left[\left(\begin{array}{cccc} -.7415 & -.0046 & -.7617 & 1.7743 \end{array} \right) X \right]$$
$$-1.8609g \left[\left(\begin{array}{cccc} .3212 & .0209 & .0738 & .4704 \end{array} \right) X \right]$$
$$+1.9257g \left[\left(\begin{array}{cccc} 2.7021 & -.0015 & .6348 & .1664 \end{array} \right) X \right]$$
$$-1.7405g \left[\left(\begin{array}{cccc} .2764 & -.0030 & .4126 & -.7780 \end{array} \right) X \right]$$
$$+.9304g \left[\left(\begin{array}{cccc} -.1779 & .0011 & .0905 & -1.0458 \end{array} \right) X \right].$$

The approximation to the .05 conditional quantile is

$$q_{.05}^{nn}(X, \widehat{\psi}, \widehat{\phi}) \equiv -2.0121 - 1.9305g \left[\left(\begin{array}{cccc} -.1226 & -.0000 & .3373 & .4493 \end{array} \right) X \right]$$
$$+1.4384g \left[\left(\begin{array}{cccc} -.0586 & -.0085 & .6109 & .4962 \end{array} \right) X \right]$$
$$+1.6166g \left[\left(\begin{array}{cccc} -.0519 & .0021 & .2616 & -.3020 \end{array} \right) X \right]$$
$$+1.7194g \left[\left(\begin{array}{cccc} 1.7480 & -.0158 & 1.1703 & .5165 \end{array} \right) X \right]$$
$$+2.7767g \left[\left(\begin{array}{cccc} -.9747 & .0037 & -.8420 & -.8134 \end{array} \right) X \right].$$

Recall that these functions are only valid when the correlation between the innovations is negative. So when the estimated correlation $\widehat{\gamma}$ is negative, we reject the null that $\theta = 0$ versus the alternative $\theta > 0$ when the t-statistic is greater than $q_{.95}^{nn}(X, \widehat{\psi}, \widehat{\phi})$. When $\widehat{\gamma}$ is positive, we replace x_t with $-x_t$ and again calculate the sufficient statistics X and the t statistic $\widehat{t}$. In this case we reject the null when $-\widehat{t}$ is greater than $-q_{.05}^{nn}(X, \widehat{\psi}, \widehat{\phi})$.

8 Technical Appendix B: conditional bootstrap algorithm

In this section we describe the conditional bootstrap used to carry out inference in the multivariate model.

1. Compute $\widehat{\Sigma}$, the unrestricted regression estimate of Σ. Compute the transformed vector $(\widetilde{y}_t \; \widetilde{\mathbf{x}}_t')' = \widehat{\Sigma}^{-1/2} (y_t \; \mathbf{x}_t')'$, where $\widehat{\Sigma}^{1/2}$ is the lower diagonal choleski decomposition of $\widehat{\Sigma}$ and satisfies $\widehat{\Sigma}^{1/2} \left(\widehat{\Sigma}^{1/2} \right)' = \widehat{\Sigma}$. Compute $\widehat{\theta}_{2,R}$ by regressing $\widetilde{y}_t$ on the second element of $\widetilde{\mathbf{x}}_{t-1}$ and a constant. Compute $\widehat{K}_R$ by regressing $\widetilde{\mathbf{x}}_t$ on $\mathbf{x}_{t-1}$

and a constant, and premultiplying the result by $\widehat{\Sigma}^{1/2}$. $\widehat{\theta}_{2,R}$ and $\widehat{K}_R$ are the maximum likelihood estimators for θ_2 and K when $\widehat{\Sigma}$ is the known covariance matrix and the null $\theta_1 = 0$ is imposed. Define the vector

$$X = \left(\text{vec} \left(\widehat{K}_R \right)' \ \text{se} \left(x_1 \right) \ \text{se} \left(x_2 \right) \ \widehat{\text{Corr}} \left(x_1, x_2 \right) \right)', \tag{27}$$

where $\mathbf{x}_t = (x_{1,t}, x_{2,t})$, $[\text{se} \left(x_1 \right)]^2 = \frac{1}{T-1} \sum \left(x_{1,t-1} - \overline{x}_1 \right)^2$ is the estimated variance of the first element of, $[\text{se} \left(x_2 \right)]^2$ is the estimated variance of the second element, and $\widehat{\text{Corr}} \left(x_1, x_2 \right)$ is their estimated covariance.

2. Simulate B data sets from the parameter values $\theta_1 = 0$, $\widehat{\theta}_{2,R}$, $\widehat{K}_R$, and $\widehat{\Sigma}$. Let t_b denote the t-statistic for the bth simulated data set, and let X_b denote the X vector for the bth sample.

3. Create the variable $d_b = \max_i \left| \left(X_i - X_{b,i} \right) / s_i \right|$, where X_i and $X_{b,i}$ are the ith elements of X and X_b, and $s_i^2 = (B-1)^{-1} \sum_b (X_{b,i} - \overline{X}_i)^2$, the standard deviation of $X_{b,i}$. d_b is a measure of the distance between the sufficient statistics computed from the actual and the simulated data.

4. Let $d_{(b)}$ denote the bth sorted d value, sorted in ascending order, so $d_{(1)} \leq d_{(2)} \leq \cdots \leq d_{(B)}$. Let $\mathcal{D}$ denote the set of $\widehat{t}_b$ where the corresponding X_b is among the N which are nearest to the actual sufficient statistics:

$$\widehat{t}_b \in \mathcal{D} \text{ iff } d_{(b)} \leq d_{(N)}. \tag{28}$$

5. The set of draws $\mathcal{D}$ are treated as draws from the conditional distribution of $\widehat{t}$ given S. We estimate the $100\alpha^{th}$ quantile of the conditional distribution with the $100\alpha^{th}$ empirical quantile of the sample of draws $\mathcal{D}$.

This bootstrap procedure computes a nonparametric nearest neighbor estimate of the conditional quantile of $\widehat{t}$ given X. Chauduri (1991) shows that as B and N increase to infinity, with N becoming large at a slower rate than B, the boostrapped quantile converges in probability to the true conditional quantile. However, since X is a high-dimensional vector the curse of dimensionality requires B to be extraordinarily large, possibly in the billions. Thus if we take Chauduri's (1991) theory literally it is not computationally feasible to precisely estimate the conditional quantile. The

48

Monte Carlos in section 4.2 suggest that the conditional bootstrap accomplishes the more modest goal of improving on the parametric bootstrap.

We choose the N and B used in Tables 5 and 6 to match the Monte Carlos in section 4.2. For the p-values that test the nulls $\theta_1 = 0$, $\theta_2 = 0$, and $\theta_1 = \theta_2 = 0$, we set $N = 200000$ and $B = 10000$. For the confidence intervals we chose $N = 20000$ and $B = 1000$.

References

Adrian, Tobias and Francesco Franzoni, 2002. Learning about beta: An explanation of the value premium. Unpublished paper, MIT.

Ang, Andrew and Geert Bekaert, 2001. Stock return predictability: Is it there? Unpublished paper, Graduate School of Business, Columbia University.

Asness, Clifford S., 2002. Fight the Fed model: the relationship between stock market yields, bond market yields, and future returns. Unpublished paper, AQR Capital Management, LLC.

Ball, R., 1978. Anomalies in relationships between securities' yields and yield-surrogates. *Journal of Financial Economics* 6, 103–126.

Banz, Rolf W., 1981. The relation between return and market value of common stocks. *Journal of Financial Economics* 9, 3–18.

Basu, Sanjoy, 1977. Investment performance of common stocks in relation to their price-earnings ratios: A test of the efficient market hypothesis. *Journal of Finance* 32, 663–682.

Basu, Sanjoy, 1983. The relationship between earnings yield, market value, and return for NYSE common stocks: Further evidence. *Journal of Financial Economics* 12, 129–156.

Bishop, Christopher M., 1995. *Neural Networks for Pattern Recognition.* Oxford University Press, New York, NY.

Black, Fischer, 1972. Capital market equilibrium with restricted borrowing. *Journal of Business* 45, 444–454.

Campbell, John Y., 1987. Stock returns and the term structure. *Journal of Financial Economics* 18, 373–399.

Campbell, John Y., and John H. Cochrane, 1999. Force of habit: a consumption-based explanation of aggregate stock market behavior. *Journal of Political Economy* 107, 205–251.

Campbell, John Y. and Robert J. Shiller, 1988a. The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies* 1, 195–228.

Campbell, John Y. and Robert J. Shiller, 1988b. Stock prices, earnings, and expected dividends. *Journal of Finance* 43, 661–676.

Campbell, John Y. and Robert J. Shiller, 1998. Valuation ratios and the long-run stock market outlook. *Journal of Portfolio Management* 24(2), 11–26.

Campbell, John Y. and Motohiro Yogo, 2002. Efficient tests of stock return predictability, unpublished paper. Harvard University.

Campbell, John Y. and Tuomo Vuolteenaho, 2003. Bad beta, good beta. Unpublished paper, Harvard University.

Chaudhuri, P., 1991. Nonparametric quantile regression. *Annals of Statistics* 19, 760–777.

Cohen, Randolph, Christopher Polk, and Tuomo Vuolteenaho, 2002. Does risk or mispricing explain the cross section of stock prices? Unpublished paper, Northwestern University and Harvard University.

Cohen, Randolph, Christopher Polk, and Tuomo Vuolteenaho, 2003. The value spread. *Journal of Finance* 58, 609–641.

Davis, J. L., Eugene F. Fama, and Kenneth R. French, 2000. Characteristics, covariances, and average returns: 1929 to 1997. *Journal of Finance* 55, 389-406.

Doornik, Jurgen A., 2001. *Object-Oriented Matrix Programming using Ox 3.0*. London: Timberlake Consultants Ltd and Oxford: www.nuff.ox.ac.uk/Users/Doornik.

Fama, Eugene F., 1998. Determining the number of priced state variables in the ICAPM. *Journal of Financial and Quantitative Analysis* 33, 217-231.

Fama, Eugene F. and Kenneth R. French, 1988. Dividend yields and expected stock returns. *Journal of Financial Economics* 22, 3–27.

Fama, Eugene F. and Kenneth R. French, 1989. Business conditions and expected returns on stocks and bonds. *Journal of Financial Economics* 25, 23–49.

Fama, Eugene F. and Kenneth R. French, 1992. The cross-section of expected stock returns. *Journal of Finance* 47, 427–465.

Fama, Eugene F. and Kenneth R. French, 1999. Forecasting profitability and earnings. *Journal of Business* 73, 161–176.

Fama, Eugene F., and Kenneth R. French, 2002. The equity premium. *Journal of Finance* 57, 637-659.

Franzoni, Francesco, 2002. Where is beta going? The riskiness of value and small stocks. Unpublished paper, MIT.

Gordon, Myron, 1962. *The Investment, Financing, and Valuation of the Corporation.* Irwin, Homewood, IL.

Graham, Benjamin and David L. Dodd, 1934. *Security Analysis.* First edition, McGraw Hill, New York.

Hodrick, Robert J., 1992. Dividend yields and expected stock returns: alternative procedures for inference and measurement. *Review of Financial Studies* 5, 357–386.

Imhof, J. P., 1961. Computing the distribution of quadratic forms in normal variables. *Biometrika* 48, 419–426.

Jansson, Michael and Marcelo Moreira, 2002. Conditional inference in models with nearly nonstationary regressors. Unpublished paper, Harvard University.

Keim, Donald and Robert Stambaugh, 1986. Predicting returns in the stock and bond markets. *Journal of Financial Economics* 17, 357–390.

Koenker, Roger and Portnoy, Stephen, 1997. The Gaussian hare and Laplacian tortoise: computability of squared-error versus absolute-error estimators. *Statistical Science* 12, 279–296.

Lakonishok, Josef, Andrei Shleifer, and Robert W. Vishny, 1994. Contrarian investment, extrapolation, and risk. *Journal of Finance* 49, 1541–1578.

Lewellen, Jonathan, 2002. Predicting returns with financial ratios. Unpublished paper, Sloan School of Management, MIT.

Lintner, John, 1956. Distribution of incomes of corporations among dividends, retained earnings, and taxes. *American Economic Review* 61, 97-113.

Lintner, John, 1965. The valuation of risky assets and the selection of risky investments in stock portfolios and capital budgets. *Review of Economics and Statistics* 47, 13–37.

Masters, Timothy, 1993. *Practical Neural Network Recipes in C++*. Academic Press, Boston, MA.

Merton, Robert C., 1973. An intertemporal capital asset pricing model. *Econometrica* 41, 867–87.

Merton, Robert C., 1980. On the estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics* 8, 323–361.

Miller, Merton and Franco Modigliani, 1961. Dividend policy, growth, and the valuation of shares. *Journal of Business* 34, 411-433.

Modigliani, Franco and Richard A. Cohn. Inflation, rational valuation, and the market. *Financial Analysts Journal*, March–April 1979, 24–44.

Polk, Christopher, 2003. The market as a hedge. Unpublished paper, Northwestern University.

Reinganum, Mark R., 1981. Misspecification of capital asset pricing: Empirical anomalies based on yields and market values. *Journal of Financial Economics* 9, 19–46.

Ritter, Jay R., and Richard Warr, 2002. The decline of inflation and the bull market of 1982-1999. *Journal of Financial and Quantitative Analysis* 37, 29-61.

Roll, Richard, 1977. A critique of the asset pricing theory's tests: Part I. *Journal of Financial Economics* 4, 129–176.

Rosenberg, Barr, Kenneth Reid, and Ronald Lanstein, 1985. Persuasive evidence of market inefficiency. *Journal of Portfolio Management* 11, 9–17.

Ross, Stephen A., 1976. The arbitrage theory of capital asset pricing. *Journal of Economic Theory* 13, 341–360.

Rozeff, M., 1984. Dividend yields are equity risk premiums. *Journal of Portfolio Management* 11, 68-75.

Sharpe, William, 1964. Capital asset prices: a theory of market equilibrium under conditions of risk. *Journal of Finance* 19, 425–442.

Shiller, Robert J., 1981. Do stock prices move too much to be justified by subsequent changes in dividends? *American Economic Review* 71, 421–436.

Shiller, Robert J., 2000. *Irrational Exuberance.* Princeton University Press, Princeton, NJ.

Stambaugh, Robert F., 1982. On the exclusion of assets from tests of the two parameter model. *Journal of Financial Economics* 10, 235–268.

Stambaugh, Robert F., 1999. Predictive regressions. *Journal of Financial Economics* 54, 375–421.

Vuolteenaho, Tuomo, 2000. Understanding the aggregate book-to-market ratio and its implications to current equity-premium expectations. Unpublished paper, Harvard University.

Vuolteenaho, Tuomo, 2002. What drives firm-level stock returns. *Journal of Finance* 57, 233-264.

White, Halbert, 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48, 817-838.

White, Halbert , 1992. *Artificial Neural Networks: Approximation and Learning Theory.* Blackwell Publishers, Cambridge, MA.