Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Financial Economics 81 (2006) 101–141
JOURNAL OF Financial ECONOMICS
www.elsevier.com/locate/jfec

Cross-sectional forecasts of the equity premium ☆

Christopher Polk[a,*], Samuel Thompson[b], Tuomo Vuolteenaho[c,d]

[a]*Kellogg School of Management, Northwestern University, Evanston, IL 60208, USA*
[b]*Department of Economics, Harvard University, Cambridge, MA 02138, USA*
[c]*Arrowstreet Capital L.P., Cambridge, MA 02138, USA*
[d]*National Bureau of Economic Research, Cambridge, MA 02138, USA*

Received 15 April 2004; received in revised form 14 December 2004; accepted 16 March 2005
Available online 18 January 2006

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.
© 2005 Published by Elsevier B.V.

JEL classification: C12; C15; C32; C53; G12; G14

Keywords: Equity premium; CAPM; Predicting returns; Conditional inference; Neural networks

☆We would like to thank Nick Barberis, John Campbell, Xiaohong Chen, Randy Cohen, Kent Daniel, Ken French, Ravi Jagannathan, Matti Keloharju, Jussi Keppo, Jonathan Lewellen, Stefan Nagel, Vesa Puttonen, Jeremy Stein, and seminar participants at the American Finance Association 2005 meeting, Boston University Economics Department, Brown Economics Department, 2004 CIRANO conference, Dartmouth Tuck School, Harvard Economics Department, Kellogg School of Management, MIT Sloan School, University of Michigan Industrial and Operations Engineering Department, NYU Economics Department, NYU Stern School, and University of Rochester Economics Department for useful comments. We are grateful to Ken French and Robert Shiller for providing us with some of the data used in this study. All errors and omissions remain our responsibility.

*Corresponding author.
E-mail address: c-polk@kellogg.northwestern.edu (C. Polk).

0304-405X/$ - see front matter © 2005 Published by Elsevier B.V.
doi:10.1016/j.jfineco.2005.03.013

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 equity-premium forecasts. We compute a number of cross-sectional association measures between stocks' expected-return proxies (including the book-to-market equity ratio, earnings yield, etc.) and 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 or her portfolio between high-beta and low-beta stocks. Given that 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. Because a high equity premium almost necessarily manifests itself with a low price for the market, negative correlation between our variables and the Standard and Poor's (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, our cross-sectional beta-premium measures forecast the equity premium. In the US data over 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. These predictive results are also robust to a number of alternative methods of constructing the cross-sectional beta-premium measure.

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 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 is not inconsistent with the theory. Campbell and Shiller (1988a, b) 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. Because our variables are based on an economic theory and a cross-sectional approach that is 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 post-1965 subsample, the predictive ability of our cross-sectional beta-premium measures is less strong than in the pre-1965 subsample. This is perhaps not surprising, given that we generate our cross-sectional forecasts using a model, the CAPM, that fails to empirically describe the cross-section of average returns in more recent subsamples (Fama and French, 1992, and others). An optimist, seeing our results, would point out that 95% confidence interval always covers a positive value for the forecasting coefficient in all of the subsample partitionings. A pessimist would counter that our cross-sectional measure is not statistically useful in predicting the equity-premium in the second half of the sample.

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 1980s. If the market's smoothed earnings yield is 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 show statistically significant forecastability of the 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 conditional Sharpe ratio on this zero-beta zero-investment portfolio was close to one in early 1982.

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 with shocks that 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 Nelson and Kim (1993) and 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 the Jansson and Moreira (2003) idea of conditioning the critical value of the test on a sufficient statistic of the data. Specifically, we map the sufficient statistics of the data to the critical value for the usual ordinary least squares (OLS) t-statistic using a neural network (essentially a fancy look-up table). 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) whether or not the data series follows a unit root process.

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:

$$\mathrm{E}_{t-1}(R_{i,t}) - R_{rf,t-1} = \beta_{i,t-1}[\mathrm{E}_{t-1}(R_{M,t}) - R_{rf,t-1}]. \tag{1}$$

In Eq. (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 $t-1$. $\mathrm{E}_{t-1}(R_{M,t}) - R_{rf,t-1}$ is the expected market premium. In our empirical implementation, we use the Center for Research in Securities Prices (CRSP) value-weight portfolio of stocks as our proxy for the market portfolio.[1]

Intuitively, a high expected return on stock i (caused by either a high beta of stock i or a high equity premium or both) 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 risk-premium forecast:

$$\frac{D_i}{P_i} - R_{rf} + \mathrm{E}(g_i) = \mathrm{E}(R_i) - R_{rf} \tag{2}$$

Eq. (2) states that the expected return on the stock equals the dividend yield (D_i/P_i) minus the interest rate plus the expected dividend growth $\mathrm{E}(g_i)$.

Reorganizing Eq. (2), substituting the Sharpe-Lintner CAPM's prediction for expected return, and assuming that betas and the risk-free rate are constant yields

$$\frac{D_{i,t}}{P_{i,t-1}} \approx \beta_i \mathrm{E}_{t-1}[R_{M,t} - R_{rf}] - \mathrm{E}(g_i - R_{rf}). \tag{3}$$

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 could have a high beta. Second, the premium per a unit of beta, that is, the expected equity premium, could be high. Third, and finally, the dividends of the stock could be expected to grow slowly in the future.

Eq. (3) leads to a natural cross-sectional measure of the equity premium. Simply regress the cross-section of dividend yields on betas and expected dividend growth,

$$\frac{D_{i,t}}{P_{i,t-1}} \approx \lambda_{0,t-1} + \lambda_{1,t-1}\beta_i + \lambda_{2,t-1}\mathrm{E}(g_i). \tag{4}$$

If expected excess returns on the market are constant, $\lambda_{1,t-1}$ recovers the expected excess market return. The central idea in our paper is to measure $\lambda_{1,t-1}$ for each period using

[1] Roll (1977) argues that this proxy is too narrow, because it excludes many assets such as human capital, real estate, and corporate debt. Although Stambaugh (1982) shows some evidence that inference about the CAPM is insensitive to exclusion of less risky assets, a reader who is concerned about the omission of assets from our market proxy can choose to interpret our subsequent results within the arbitrage pricing theory (APT) framework of Ross (1976).

purely cross-sectional data, and then use that measurement to forecast the next period's equity premium.[2]

The CAPM does a poor job describing the cross-section of stock returns in the post-1963 sample. However, that failure does not necessarily invalidate our approach. First, Kothari et al. (1995) and Ang and Chen (2004) find a positive univariate relation between average returns and CAPM betas. Given that both studies use a long sample as we do, their evidence indicates that the CAPM is an adequate model for our purposes, at least for our full-period tests. If the CAPM is a poor model in the second subsample, then it is reasonable to expect our predictor to also perform poorly in the post-1963 sample. Second, Cohen et al. (2003, 2005b) show that although the CAPM perhaps does a poor job describing cross-sectional variation in average returns on dynamic portfolios, that model does a reasonable job describing the cross-section of stock prices, which is essentially our left-hand-side variable in Eq. (4). Third, for our method to work, we do not need the CAPM to be a perfect model. All we need is that a higher expected equity premium (relative to its time-series mean) results in a more positive relation between various price-level yield measures and CAPM betas. Fourth, and most important, we do not simply assume that $\lambda_{1,t-1}$ is the equity premium but test its predictive ability in our subsequent time-series tests.

Our methodology can be easily extended to multi-factor models that also include a market factor, such as the Merton (1973) intertemporal capital asset pricing model (ICAPM) and many arbitrage pricing theory (APT) specifications of Ross (1976). For such models, one can regress the expected-return proxies on multi-factor betas (including the loading on the market in a multiple regression). The partial regression coefficient on the market-factor loading is again related to the expected excess return on the market.

Neither the theory we rely on (the CAPM) nor our empirical tests provide insight into why the expected equity premium and cross-sectional beta premium vary over time. The hypothesis we test is whether the pricing of risk is 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 the result of 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; Cohen et al., 2005a), or some unmodeled hedging demand beyond our myopic framework (Merton, 1973; Fama, 1998) remains an unanswered question.

3. Data and construction of variables

We construct a number of alternative proxies for the cross-sectional risk premium. In construction of all these cross-sectional risk-premium measures, we avoid any look-ahead bias so that all of our proxies are valid variables in regressions forecasting the equity premium.

[2]The Gordon model has the limitation that expected returns and expected growth must be constant, and thus using the Gordon model to infer time-varying expected returns is in principle internally inconsistent. Interpretating Eq. (4) in the context of the Campbell and Shiller (1988a, b) log-linear dividend discount model that allows for time-varying expected returns alleviates this concern. If one repeats the above steps using the Campbell–Shiller model and assumes that the expected one-period equity premium $E_{t-1}[R_{M,t} - R_{rf}]$ follows a first-order autoregressive process, the expected one-period equity premium is then linearly related to the multiple regression coefficient $\lambda_{1,t-1}$.

The first set of proxies, λ^{SRC}, λ^{REG}, and λ^{MSCI}, is 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 not only being robust to outliers in the underlying data but also of never generating extreme values themselves. This robustness comes at a cost, however, because the ordinal measures have the disadvantage of throwing away some of the information in the magnitude of the cross-sectional spread in valuation multiples.

The second set of cross-sectional risk-premium proxies, λ^{DP}, λ^{DPG}, λ^{BM}, and λ^{BMG}, is measured on a ratio scale and thus relates more closely to Eq. (4). To alleviate the outlier problem associated 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 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.

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. Selecting appropriate proxies for a firm's valuation multiple is the main challenge of our empirical implementation. Because 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 Eq. (4). Instead, we 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, stemming from 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 to price (E/P), and cash flow to price (C/P). The raw cross-sectional data come from the merger of three databases. The first of these, the 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 US stocks. The Compustat accounting information is supplemented by the third database, Moody's book equity information for industrial firms as collected by Davis et al. (2000). Detailed data definitions are as follows. 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, set to zero if unavailable), plus post-retirement benefit liabilities (data item 330, set to zero if unavailable), minus the book value of preferred stock. Depending on availability of preferred stock data, 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. (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 and 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 accounting 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 re-ranked 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 et al. (1994), Reinganum (1981), and Rosenberg et al. (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 years of monthly returns in an OLS regression on a constant and the contemporaneous return on the value-weight NYSE-Amex-Nasdaq portfolio. We skip those months in which a firm is missing returns. However, we require all observations to occur within a four-year window. As we sometimes estimate beta using only 12 returns, we censor each firm's individual monthly return to the range $(-50\%, 100\%)$ 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 we 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 resulting from simplicity and robustness. First, averaging the ranks on available multiples conveniently deals with missing data for one or more of our valuation 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.

This proxy also has the following disadvantages. First, in computing λ^{SRC} we do not control for expected growth and profitability that could be cross-sectionally related to betas, causing an omitted-variables bias in the estimates. This omitted-variable bias can be significant, if expected growth and profitability are correlated with betas. 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} could cause us to throw away some significant information related to expansions and contractions of the cross-sectional spread in betas and valuation multiples. Appendix A shows via a calibration exercise that the latter disadvantage is unlikely to be a significant concern in our sample.

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. To control for growth opportunities, we need proxies for expected future growth Eq. (4) 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 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 zero for dividend payers and one 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) show 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 standard industrial classification (SIC) code industry. Low concentration within industry should signal intense competition and thus lower profitability. Because the selection of growth proxies is a judgment call, it is fortunate that our main subsequent results are insensitive to the inclusion or exclusion of these 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:

$$VALRANK_{i,t} = \lambda_{0,t} + \lambda_t^{REG}\widehat{\beta}_{i,t} + \sum_{g=1}^{6} \lambda_t^g GROWTHRANK_{i,t}^g + \varepsilon_{i,t} \tag{5}$$

$GROWTHRANK_{i,t}^g$ is the corresponding percentile rank for six firm-level profitability controls. Given that Cohen et al. (2003, 2005b) show that the majority of the cross-sectional variation in valuation ratios across firms is the result of 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. Because 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 website. We take the top 30% and bottom 30% portfolios sorted on four Morgan Stanley Capital International (MSCI) 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 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 compute the value-weight one-year dividend growth for the portfolios. λ^{DPG} is the multiple regression coefficient of the portfolios' dividend yields on their betas, controlling for one-year past dividend growth rates.

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 portfolios' BE/MEs on their betas, controlling for one-year past value-weight ROEs.

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 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 -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}.[3] 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^e_M), computed as the difference between the simple return on the 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^e_m) 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 an 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 compare their predictive ability against that of 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 (1988a, b, 1998) advocate averaging earnings over several years to avoid temporary spikes in the price-earnings ratio caused by cyclical declines in earnings. We follow the Campbell and Vuolteenaho (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 10-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.

[3]The variables are cross-sectionally demeaned, because only cross-sectional variation in expected stock returns matters for our premium estimates and because demeaning reduces the noise and adds to the precision of the estimates.

In our informal illustrations, we also use the dividend-price ratio, computed as the ratio of trailing 12-month dividends and the price for the S&P 500 index. We also use the 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 in which the forecasting variable is persistent and shocks to the forecasting variable are potentially correlated with shocks to the variable being forecast.

4.1. Inference in univariate regressions

Consider the one-period prediction model

$$\begin{aligned} y_t &= \mu_1 + \theta x_{t-1} + u_t, \quad \text{and} \\ x_t &= \mu_2 + \rho x_{t-1} + v_t, \end{aligned} \tag{6}$$

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. Because 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 the Stambaugh (1999) example that the dividend yield does not predict stock returns. The usual t-statistic for this hypothesis is

$$\widehat{t} = \widehat{\sigma}_u^{-1} \sqrt{\sum (x_{t-1} - \overline{x})^2}\, \widehat{\theta}, \tag{7}$$

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. However, this is a poor approximation of the true sampling distribution of $\widehat{t}$ in small samples. 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.

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, $\Pr[\widehat{t} > k(\rho)] = 0.05$. 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, Jansson and Moreira (2003) have proposed a solution to this problem. Suppose that the covariance parameters σ_u^2, σ_v^2 and γ are known. Under the null that $\theta = 0$,

the statistics

$$S = \left\{ \frac{\sum (x_{t-1} - \overline{x})(x_t - \sigma_v \gamma y_t / \sigma_u)}{\sum (x_{t-1} - \overline{x})^2}, \sum (x_{t-1} - \overline{x})^2, \overline{x}, \overline{y}, x_1, y_1 \right\} \tag{8}$$

are sufficient statistics for the parameter ρ, where $\overline{x} = (T-1)^{-1} \sum_{t=2}^{T} x_{t-1}$ and $\overline{y} = (T-1)^{-1} \sum_{t=2}^{T} y_t$. The definition of a sufficient statistic is as follows: 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. The idea in their method is to 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 $S = s$:

$$\Pr[\widehat{t} \leqslant q(s, \alpha) \mid S = s, \ \theta = 0] = \alpha. \tag{9}$$

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

Jansson and Moreira (2003) 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 with q^{nn}, a neural network:

$$\begin{aligned} q(S, \alpha) &\approx q_\alpha^{nn}(X, \widehat{\psi}, \widehat{\xi}), \\ q_\alpha^{nn}(X, \psi, \xi) &\equiv \operatorname{sign}(\widehat{\gamma}) \mu(X) + \sigma(X) \Phi^{-1}(\alpha). \end{aligned} \tag{10}$$

$\Phi^{-1}(\alpha)$ is the quantile function for a standard normal variable, so $\Pr[N(0,1) \leqslant \Phi^{-1}(\alpha)] = \alpha$. The mean and variance are neural networks in the sufficient statistics:

$$\begin{aligned} \mu(X) &= \xi_0^\mu + \sum_{j=1}^{4} \xi_j^\mu g(\psi_j' e^X) \quad \text{and} \quad \sigma(X) = \exp\left(\xi_0^\sigma + \sum_{j=1}^{4} \xi_j^\sigma g(\psi_j' e^X) \right), \\ X &= \left(0, T(\widehat{\rho}_R - 1)/50, -T^{-2} \sum (x_{t-1} - \overline{x})^2 / \widehat{\sigma}_v^2, \log|\widehat{\gamma}|, -T/100 \right)'. \end{aligned} \tag{11}$$

ψ_j is a five-dimensional parameter vector. The hatted variables are the usual least-squares estimators of the covariance parameters. $\operatorname{sign}(\widehat{\gamma})$ is $+1$ if $\widehat{\gamma}$ is positive, -1 otherwise. $\widehat{\rho}_R$ is the constrained maximum likelihood estimate for ρ, given that the null is true and the covariance parameters are known:

$$\widehat{\rho}_R = \frac{\sum (x_{t-1} - \overline{x})(x_t - \widehat{\sigma}_v \widehat{\gamma} y_t / \widehat{\sigma}_u)}{\sum (x_{t-1} - \overline{x})^2}. \tag{12}$$

g is called the activation function. We use the tanh activation function

$$g(x) = \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}. \tag{13}$$

We choose ψ and ξ to closely approximate the critical function q. The parameter values used are

$$\begin{aligned}
\xi^{\mu} &= (-1.7383 \quad 4.5693 \quad 2.7826 \quad -0.0007 \quad 3.9894)', \\
\xi^{\sigma} &= (0.1746 \quad -0.4631 \quad -0.1955 \quad 0.0210 \quad -0.4641)', \\
\psi_1 &= (1.8702 \quad 2.1040 \quad -3.4355 \quad 0.5738 \quad 0.0119)', \\
\psi_2 &= (3.7744 \quad -2.5565 \quad -1.9475 \quad -0.8120 \quad -0.0262)', \\
\psi_3 &= (49.9034 \quad 2.9268 \quad -52.7576 \quad -5.0194 \quad 4.4890)', \\
\psi_4 &= (-1.6534 \quad -1.0395 \quad 2.8437 \quad -0.2264 \quad 0.0084)'.
\end{aligned} \tag{14}$$

We provide an algorithm for choosing the parameters in Appendix B.

Fitting the neural network is a computationally demanding task, but we should emphasize that the applied researcher does not need to fit the net. An applied researcher can use our parameter values to easily calculate the exact small-sample critical value for any quantile α and any sample size T, under the assumptions of data-generating process Eq. (6) and i.i.d. Gaussian errors.

q^{nn} can approximate the critical function q to arbitrary accuracy. q^{nn} implies that the t-statistic has a conditional normal distribution with mean and standard deviation given by neural networks. This is a special case of the mixture of experts net (see Bishop, 1995, pp. 212–222), which approximates a conditional distribution with mixtures of normal distributions whose parameters are neural nets in the conditioning variables. The mixture of experts net is a universal approximator: Given enough activation functions and enough mixture distributions, the net can approximate any conditional distribution to arbitrary accuracy (see Chen and White, 1999). We fit our simple net Eq. (10) with a single mixture distribution and also fit larger nets with more mixture distributions. While the larger models are a bit more accurate, for practical purposes the simple net is accurate enough. Furthermore, the net with a single distribution leads to convenient expressions both for the conditional quantile q and the p-value of the test. For testing the null that $\theta = 0$ against the one-sided alternative $\theta > 0$, the p-value is

$$\text{pval}(\widehat{t}) \approx 1 - \Phi\left(\frac{\widehat{t} - \text{sign}(\widehat{\gamma})\mu(X)}{\sigma(X)}\right). \tag{15}$$

The vector X differs from the sufficient statistics in several ways for computational convenience. X transforms some of the sufficient statistics and omits the statistics $\overline{x}$, $\overline{y}$, x_1, and y_1. X also uses parameter estimates $\widehat{\sigma}_u$, $\widehat{\gamma}$, and $\widehat{\sigma}_v$ in place of the known covariance parameters. Fortunately, the omitted statistics $\overline{x}$, $\overline{y}$, x_1 and y_1 are not particularly informative about the nuisance parameter ρ. Size distortions caused by omitting these statistics are very small.

The Jansson–Moreira theory delivers an exact test when the covariance parameters are known. In practice one must use parameter estimates. We design 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 could be the case that no exact test exists in this model. Again, however, any size distortions caused by unknown covariance

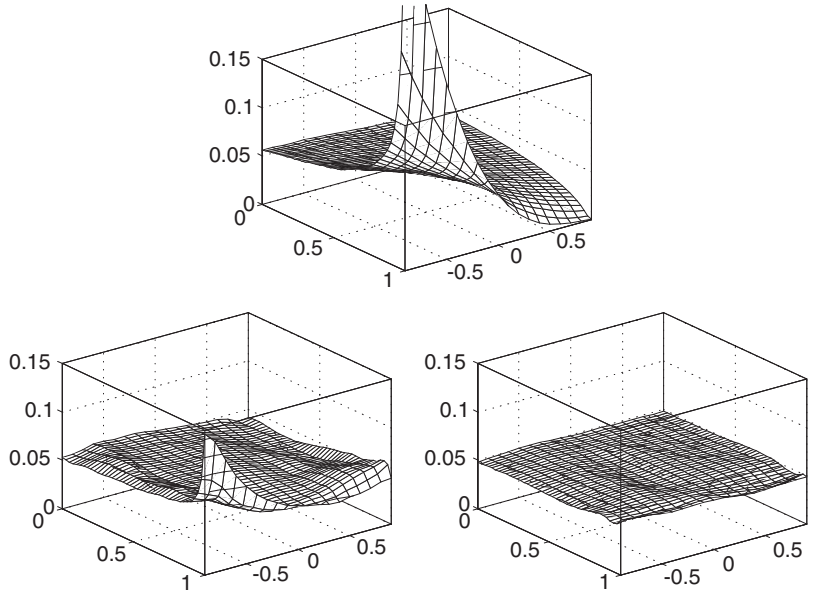

Fig. 1. Size of the test in a Monte Carlo experiment. Consider the data-generating process: $y_t = \mu_1 + \theta x_{t-1} + u_t$; $x_t = \mu_2 + \rho x_{t-1} + v_t$, with i.i.d. normal error vectors. We are interested in testing the hypothesis $\theta = 0$ against $\theta > 0$. On the top are empirical rejection frequencies from using the usual critical value of 1.65 for a one-tailed t test. On the bottom left are results for the bootstrap, and on the bottom right the conditional critical function q_α^{nn} is used. The grid values range over $\text{Corr}(u, v) \in \{-.9, -.8, \ldots, .8, .9\}$ and $\rho \in \{0, .025, \ldots, .975, 1\}$. For each grid point in the top and bottom right pictures there are 40 thousand Monte Carlo trials of $T = 120$ observations. The bootstrap is calculated as follows. For each grid point we simulate 30 thousand sample data sets, and for each simulated sample we bootstrap one thousand 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.

parameters are small. Furthermore, estimation error for the covariance parameters is asymptotically negligible, whether the x_t process is stationary or not.[4]

A Monte Carlo experiment demonstrates the accuracy of our approximation. Fig. 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 many 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 at the top reports results for the classical critical value of 1.65. The plot at the bottom left reports results from using the bootstrap, and the bottom right gives results for the conditional critical function q_α^{nn}. The bootstrap algorithm is described in the notes to the Fig. 1.

When ρ is close to one, the classical critical value under-rejects for positive $\text{Corr}(u, v)$ and over-rejects for negative $\text{Corr}(u, v)$. When $\rho = 1$ and the correlation is -0.9 the usual critical value rejects a true null about 38% of the time. The bootstrap improves on this

[4]While classical asymptotic theory requires stationarity, the conditional testing theory is not sensitive to unit root problems. See the argument by Jansson and Moreira (2003) for details.

result but remains flawed: For $\rho = 1$ and $\gamma = -0.9$, the rejection frequency of a nominal 5% size test is 17%. Our conditional critical function leads to accurate rejection frequencies over the entire range of ρ and $\mathrm{Corr}(u, v)$ values, with rejection rates ranging from 4.46% to 5.66%. 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 above experiments show that the size of the test is correct. The test is also powerful. In Appendix B, we discuss the optimality properties of the test. Our conclusion from these power considerations is that it is difficult to devise a test with a significant power advantage relative to our conditional test, at least as long as we have no additional information about the parameters (such as $\rho = 1$).

The conditional testing procedures described above assume homoskedastic normal errors. Our conditional testing procedures can be modified to be more robust to heteroskedasticity. Under the so-called local-to-unity asymptotic limit used by Campbell and Yogo (2006) and Torous et al. (2005), heteroskedasticity does not alter the large sample distribution of the t-statistic. The local-to-unity limit takes $\rho = 1 + c/T$ for fixed c and increasing T, i.e., it takes ρ to be in a shrinking neighborhood of unity. This is in contrast to the traditional asymptotics, which fix ρ as T becomes large. Under the traditional (fixed ρ) asymptotic limit, heteroskedasticity changes the null distribution of the t-statistic. However, under the local-to-unity limit, heteroskedasticity is asymptotically irrelevant.

In interpreting the asymptotic local-to-unity results, one should note that it is a large sample result that holds only when ρ is very close to one. In a small sample, or when ρ is small enough so that traditional asymptotics work, heteroskedasticity matters. In the empirical sections of the paper, we also carry out conditional inference based on t-statistics computed with Eicker-Huber-White (White, 1980) standard errors. We calibrate separate critical-value functions analogous to Eq. (10) for this test statistic. This calibration process is analgous to the calibration process for the test using the usual OLS t-statistic, and thus we omit the details here to conserve space.

Although the combination of Eicker-Huber-White standard errors and conditional inference appears sensible, this test comes 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. However, while we have not proven any formal analytical results, unreported Monte Carlo experiments suggest that the Eicker-Huber-White t-statistic is much more robust to heteroskedasticity in small samples than the uncorrected t-statistic. Also, we do know that under homoskedasticity the size of this modified test is correct.

4.2. Inference in multivariate regressions

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

$$\begin{aligned} y_t &= \mu_1 + \theta' \mathbf{x}_{t-1} + u_t, \quad \text{and} \\ \mathbf{x}_t &= \boldsymbol{\mu}_2 + K\mathbf{x}_{t-1} + V_t, \end{aligned} \tag{16}$$

where $\mathbf{x}_t$, $\boldsymbol{\mu}_2$ and θ are two-dimensional column vectors, K is a 2×2 matrix, and V_t is a two-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

$$\widehat{t} = \frac{\widehat{\theta}_1}{\sqrt{\Omega_{11}}}. \tag{17}$$

$\widehat{\theta}_1$ is the ordinary least squares estimate of θ_1, the first element of θ, and Ω_{11} is an estimate of the variance of $\widehat{\theta}_1$, the (1, 1) element of $\Omega = \widehat{\sigma}_u^2(\sum \mathbf{x}_{t-1}\mathbf{x}'_{t-1})^{-1}$.

Classical asymptotic theory approximates the null distribution of $\widehat{t}$ with a standard normal variable. It is well known that this could be 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 i.i.d. mean zero normal variables with known covariance matrix $\Sigma = \mathrm{E}[(u_t, V_t')'(u_t, V_t')]$. 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 $\Sigma^{1/2}(\Sigma^{1/2})' = \Sigma$. The sufficient statistics for K are

$$S = \left\{\widetilde{K}, \sum(\mathbf{x}_{t-1} - \overline{\mathbf{x}})(\mathbf{x}_{t-1} - \overline{\mathbf{x}})', \overline{\mathbf{x}}, \overline{y}, \mathbf{x}_1, y_1\right\}, \tag{18}$$

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 Eq. (6) 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.

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 we 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. We also carry out heteroskedasticity-robust conditional inference using the same conditional-bootstrap procedure based on t-statistics computed with Eicker-Huber-White (White, 1980) standard errors. Details of these procedures are given in Appendix B.

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 Fig. 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, in which 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, because the Federal Reserve Board supposedly uses a similar model to judge the level of equity prices.[5]

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 arguments why the earnings yield should be augmented by subtracting the interest rate. First, if one is interested in the equity premium instead of the total equity return, subtracting the interest rate from the earnings yield is natural. 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 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.

[5]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." The Federal Reserve has not officially endorsed any stock-valuation model.

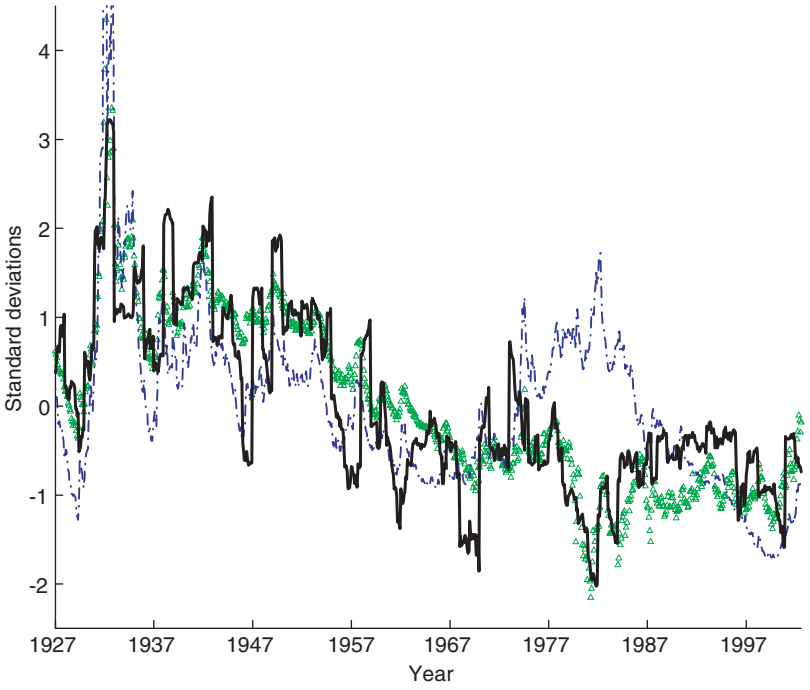

Fig. 2. Time-series evolution of the ex ante equity-premium forecasts. This figure plots the time-series of three equity-premium measures: (1) λ^{SRC}, the cross-sectional Spearman rank correlation between valuation levels and estimated betas, marked with a thick solid line; (2) exp(*ep*), the ratio of a ten-year moving average of earnings to price for Standard and Poor's (S&P) 500, marked with a dash-dotted line; and (3) exp(*ep*) − *Y*10, the ratio of a ten-year moving average of earnings to price for S&P 500 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.

In a sense, stocks should be a long-term hedge against inflation. (Modigliani and Cohn, 1979; Ritter and Warr, 2002, argue that the expected real earnings growth of levered firms 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 (plus perhaps a constant), the earnings yield is a good measure of 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. Fig. 2 also plots the smoothed earnings yield without the interest rate adjustment.

The three variables in Fig. 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

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

Variable		Constant	$\exp(ep_t)$	$\exp(ep_t) - Y10_t$	Adjsuted R^2
λ_t^{SRC}	=	−0.1339	0.0628	4.6292	71.87%
		(−10.54)	(0.30)	(36.54)	
λ_t^{SRC}	=	−0.3996	5.561		30.45%
		(−24.39)	(19.95)		
λ_t^{SRC}	=	−0.1303		4.6536	71.90%
		(−34.02)		(48.19)	

The table shows the ordinary least squares (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 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 Standard and Poor's (S&P) 500's 10-year moving average of earnings to S&P 500's price. $Y10$ is the nominal yield on 10-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 with 908 monthly observations.

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%, and the explanatory power is entirely due to the Fed model ($\exp(ep) - Y10$). (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 Fig. 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. Our λ measures are cross-sectional regression coefficients of earnings yields (and other such multiples) on betas. If 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 implies 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 the market's valuation multiples and our cross-sectional measures. Unreported experiments confirm that our results are not driven by this link.[6]

[6]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 with 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.

Fig. 2 also casts light on the Franzoni (2002), Adrian and Franzoni (2002), and Campbell and Vuolteenaho (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 ability to generate cash flow. 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 detect 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-sectional price of risk is highly correlated with reasonable ex ante measures of the equity premium, it remains for us to show that our variable forecasts equity-premium realizations. 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 report only the descriptive statistics for one cross-sectional risk-premium measure, λ^{SRC}. A high cross-sectional beta premium suggests that at that 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 1970s. It also tracks λ^{SRC}, with a correlation of 0.31 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 each table forecasts the equity premium with the cross-sectional risk-premium measure λ^{SRC}.[7] 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

[7]At first, it could appear that our statistical tests are influenced by the so-called generated regressor problem. However, because our proxy variables for the expected market premium is a function only of information available at $t-1$, our predictability tests do not over-reject. While $\widehat{\theta}$ is a biased estimate of the coefficient on the true, unknown market expectation of the equity premium, it is a consistent estimator of the coefficient on the proxy variable x_{t-1}. Further details are available from the authors on request.

Table 2
Descriptive statistics of the vector autoregression state variables

Variable	Mean	Median	S.D.	Min	Max
$R^e_{M,t}$	0.0062	0.0095	0.0556	−0.2901	0.3817
$R^e_{m,t}$	0.0097	0.0114	0.0758	−0.3121	0.6548
λ^{SRC}_t	−0.0947	−0.1669	0.2137	−0.5272	0.5946
ep_t	−2.8769	−2.8693	0.3732	−3.8906	−1.4996
TY_t	0.6232	0.5500	0.6602	−1.3500	2.7200
Correlations	$R^e_{M,t}$	$R^e_{m,t}$	λ^{SRC}_t	ep_t	TY_t
$R^e_{M,t}$	1	0.9052	0.1078	0.0305	0.0474
$R^e_{m,t}$	0.9052	1	0.1333	0.0658	0.0798
λ^{SRC}_t	0.1078	0.1333	1	0.5278	0.3120
ep_t	0.0305	0.0658	0.5278	1	0.2223
TY_t	0.0474	0.0798	0.3120	0.2223	1
$R^e_{M,t-1}$	0.1048	0.2052	0.0825	−0.0475	0.0428
$R^e_{m,t-1}$	0.1070	0.2059	0.1075	−0.0010	0.0726
$\lambda^{\mathrm{SRC}}_{t-1}$	0.0930	0.1321	0.9748	0.5196	0.3011
ep_{t-1}	0.1140	0.1509	0.5359	0.9923	0.2279
TY_{t-1}	0.0469	0.0812	0.3219	0.2188	0.9131

The table shows the descriptive statistics estimated from the full sample period 1927:5-2002:12 with 908 monthly observations. R^e_M is the excess simple return on the Center for Research in Securities Prices (CRSP) value-weight index. R^e_m is the excess simple return on the CRSP equal-weight index. λ^{SRC} is the Spearman rank correlation between valuation rank and estimated 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 Standard and Poor's (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 10-year constant-maturity taxable bonds and short-term taxable notes. S.D. denotes standard deviation.

except λ^{DPG}, we can reject the null hypothesis of a zero coefficient in favor of a positive coefficient at better than a 1% level of significance in full-sample tests assuming homoskedasticity. Using the version of conditional test that is robust to heteroskedasticity, the *p*-values are slightly higher and the evidence slightly less uniform: The measures based on firm-level data continue to be significant at better than 1%, but the measures based on portfolio-level data are no longer significant. Comparing the small-sample *p*-values to the usual critical values for these *t*-statistics, it is clear that the usual *t* tests would perform adequately in these cases. This is not surprising, given that 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 0.0368 for the 1927:5-1965:2 period to 0.0088 for the 1965:2-2002:12 period. A similar drop is observed for the other cross-sectional measures, except for λ^{ER}, which performs well in all subsamples (results unreported). However, the 95% confidence intervals 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 examine, a positive coefficient is contained within the

Table 3
Univariate predictors of the excess value-weight market return (R^e_M)

Specification	$\widehat{\theta}$	t-statistic	p	p^W	95% CI	$\widehat{\rho}$	$\widehat{\gamma}$	$\widehat{\sigma}_1$	$\widehat{\sigma}_2$
Prediction by the cross-sectional beta premium, $x_t = \lambda_t^{\mathrm{SRC}}$									
1927:5-2002:12	0.0242	2.811	<0.01	0.030	[0.008, 0.041]	0.975	0.0773	0.0553	0.0477
1927:5-1965:2	0.0368	2.450	<0.01	0.043	[0.008, 0.066]	0.960	−0.0152	0.0633	0.0562
1965:2-2002:12	0.0088	0.413	0.309	0.297	[−0.031, 0.054]	0.931	0.278	0.0460	0.0368
1927:5-1946:3	0.0663	1.967	0.030	0.091	[−0.001, 0.131]	0.934	−0.0413	0.0823	0.0585
1946:3-1965:2	0.0395	3.113	<0.01	<0.01	[0.016, 0.065]	0.957	0.080	0.0348	0.0534
1965:2-1984:1	0.0147	0.6027	0.240	0.249	[−0.03, 0.065]	0.942	0.188	0.0458	0.0429
1984:1-2002:12	−0.0190	−0.4181	>0.5	>0.5	[−0.099, 0.089]	0.885	0.416	0.0462	0.0292
Prediction by log smoothed earnings/price, $x_t = ep_t$									
1927:5-2002:12	0.0170	3.454	0.014	0.215	[0.003, 0.024]	0.993	−0.669	0.0552	0.0464
1927:5-1965:2	0.0317	3.282	0.018	0.349	[0.003, 0.046]	0.987	−0.671	0.0630	0.0549
1965:2-2002:12	0.00756	1.319	>0.5	>0.5	[−0.009, 0.011]	0.996	−0.668	0.0459	0.0359
1927:5-1946:3	0.0410	2.670	0.096	0.374	[−0.006, 0.061]	0.981	−0.659	0.0817	0.0707
1946:3-1965:2	0.0294	2.344	0.168	0.204	[−0.009, 0.043]	0.994	−0.727	0.0351	0.0322
1965:2-1984:1	0.0204	1.817	0.291	0.380	[−0.012, 0.028]	0.987	−0.662	0.0455	0.0362
1984:1-2002:12	0.0105	1.251	>0.5	0.483	[−0.012, 0.013]	0.990	−0.668	0.0460	0.0352
Prediction by term yield spread, $x_t = TY_t$									
1927:5-2002:12	0.00396	1.413	0.075	0.095	[−0.001, 0.009]	0.917	0.0111	0.0555	0.269
1927:5-1965:2	0.00489	1.015	0.178	0.214	[−0.005, 0.013]	0.968	−0.156	0.0636	0.151
1965:2-2002:12	0.00270	0.862	0.150	0.179	[−0.003, 0.008]	0.871	0.111	0.0460	0.346
1927:5-1946:3	0.00497	0.711	0.316	0.300	[−0.010, 0.017]	0.969	-0.174	0.0829	0.184
1946:3-1965:2	0.0201	1.978	0.030	0.030	[0.000, 0.039]	0.886	−0.0707	0.0352	0.108
1965:2-1984:1	0.00868	1.677	0.043	0.045	[−0.001, 0.019]	0.765	0.218	0.0455	0.378
1984:1-2002:12	−0.00221	−0.521	>0.5	>0.5	[−0.010, 0.005]	0.918	0.00463	0.0462	0.301
Full sample predictive results for alternative cross-sectional measures									
$x_t = \lambda_t^{\mathrm{REG}}$	0.0908	3.605	<0.01	<0.01	[0.042, 0.141]	0.937	0.0644	0.0552	0.0255
$x_t = \lambda_t^{\mathrm{DP}}$	0.03539	2.53	<0.01	0.138	[0.007, 0.062]	0.926	−0.167	0.0554	0.0498
$x_t = \lambda_t^{\mathrm{DPG}}$	0.02419	1.75	0.044	0.204	[−0.003, 0.051]	0.917	−0.107	0.0555	0.0531
$x_t = \lambda_t^{\mathrm{BM}}$	0.001121	2.63	<0.01	0.156	[0.0003, 0.0019]	0.942	−0.236	0.0554	1.440
$x_t = \lambda_t^{\mathrm{BMG}}$	0.001449	2.98	<0.01	0.146	[0.0005, 0.0023]	0.919	−0.222	0.0553	1.500
$x_t = \lambda_t^{\mathrm{ER}}$	2.175	3.15	<0.01	<0.01	[0.80, 3.53]	0.979	−0.0331	0.0459	0.0005

This table shows results from the model

$$R^e_{M,t} = \mu_1 + \theta x_{t-1} + u_t;\ x_t = \mu_2 + \rho x_{t-1} + v_t,$$

with E $u_t = \sigma_1^2$, E $v_t = \sigma_2^2$, and $\mathrm{Corr}(u_t, v_t) = \gamma$. The p-value tests the null $\theta = 0$ against the one-sided alternative $\theta > 0$, p denoting the p-value computed using the regular t-statistic and p^W using heteroskedasticity-robust White t-statistic. The confidence interval is a two-sided interval for θ computed assuming homoskedasticity. The hatted variables are unrestricted ordinary least squares estimates. 95% CI is the 95% confidence interval.

95% confidence intervals. Again, these conclusions are not altered even if we focus our attention on the heteroskedasticity robust version of the conditional test.

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

Table 4
Univariate predictors of the excess equal-weight market return (R_m^e)

Specification	$\widehat{\theta}$	t-stat	p	p^W	95% CI	$\widehat{\rho}$	$\widehat{\gamma}$	$\widehat{\sigma}_1$	$\widehat{\sigma}_2$
Prediction by the cross-sectional beta premium, $x_t = \lambda_t^{\mathrm{SRC}}$									
1927:5-2002:12	0.0469	4.012	<0.01	<0.01	[0.025, 0.070]	0.975	0.0202	0.0752	0.0477
1927:5-1965:2	0.0786	3.755	<0.01	<0.01	[0.037, 0.119]	0.960	−0.0613	0.0882	0.0562
1965:2-2002:12	0.0345	1.260	0.100	0.098	[−0.017, 0.092]	0.931	0.226	0.0589	0.0368
1927:5-1946:3	0.147	3.041	<0.01	0.030	[0.048, 0.238]	0.934	−0.0960	0.118	0.0585
1946:3-1965:2	0.0466	3.256	<0.01	<0.01	[0.020, 0.075]	0.957	0.0603	0.0392	0.0534
1965:2-1984:1	0.0457	1.357	0.076	0.082	[−0.017, 0.115]	0.942	0.139	0.0633	0.0429
1984:1-2002:12	0.00571	0.107	0.405	0.393	[−0.089, 0.132]	0.885	0.379	0.0542	0.0292
Prediction by log smoothed earnings/price, $x_t = ep_t$									
1927:5-2002:12	0.0307	4.594	<0.01	0.234	[0.011, 0.040]	0.993	−0.683	0.0750	0.0464
1927:5-1965:2	0.0662	4.943	<0.01	0.406	[0.026, 0.086]	0.987	−0.683	0.0872	0.0549
1965:2-2002:12	0.0104	1.420	0.470	>0.5	[−0.011, 0.015]	0.996	−0.689	0.0589	0.0359
1927:5-1946:3	0.0839	3.833	0.014	0.284	[0.016, 0.112]	0.981	−0.683	0.117	0.0707
1946:3-1965:2	0.0285	2.004	0.250	0.299	[−0.015, 0.045]	0.993	−0.707	0.0398	0.0322
1965:2-1984:1	0.0290	1.866	0.300	0.381	[−0.017, 0.039]	0.987	−0.705	0.0631	0.0362
1984:1-2002:12	0.00131	0.132	>0.5	>0.5	[−0.026, 0.004]	0.990	−0.684	0.0542	0.0352
Prediction by term yield spread, $x_t = TY_t$									
1927:5-2002:12	0.00935	2.452	<0.01	<0.01	[0.002, 0.016]	0.915	0.0140	0.0756	0.269
1927:5-1965:2	0.0106	1.572	0.074	0.089	[−0.003, 0.023]	0.968	−0.151	0.0893	0.151
1965:2-2002:12	0.00774	1.933	0.026	0.020	[0.00, 0.015]	0.871	0.124	0.0588	0.346
1927:5-1946:3	0.00867	0.857	0.243	0.220	[−0.013, 0.026]	0.969	−0.171	0.120	0.186
1946:3-1965:2	0.0208	1.808	0.043	0.029	[−0.002, 0.043]	0.886	−0.0699	0.0398	0.108
1965:2-1984:1	0.0172	2.410	<0.01	<0.01	[0.004, 0.032]	0.765	0.184	0.0628	0.378
1984:1-2002:12	0.00420	0.844	0.138	0.192	[−0.005, 0.013]	0.918	0.0809	0.0541	0.301
Full sample predictive results for alternative cross-sectional measures									
$x_t = \lambda_t^{\mathrm{REG}}$	0.165	4.811	<0.01	<0.01	[0.098, 0.232]	0.937	0.0238	0.0749	0.0255
$x_t = \lambda_t^{\mathrm{DP}}$	0.07916	4.17	<0.01	0.044	[0.041, 0.115]	0.926	−0.206	0.0751	0.0498
$x_t = \lambda_t^{\mathrm{DPG}}$	0.06813	3.63	<0.01	0.051	[0.031, 0.104]	0.917	−0.136	0.0753	0.0531
$x_t = \lambda_t^{\mathrm{BM}}$	0.002609	4.51	<0.01	0.043	[0.0015, 0.0037]	0.942	−0.245	0.075	1.440
$x_t = \lambda_t^{\mathrm{BMG}}$	0.003129	4.76	<0.01	0.052	[0.0018, 0.0043]	0.919	−0.242	0.0749	1.500
$x_t = \lambda_t^{\mathrm{ER}}$	3.371	3.74	<0.01	<0.01	[1.55, 5.12]	0.979	−0.0697	0.0599	0.0005

This table shows results from the model

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

with E $u_t = \sigma_1^2$, E $v_t = \sigma_2^2$, and Corr$(u_t, v_t) = \gamma$. The p-value tests the null $\theta = 0$ against the one-sided alternative $\theta > 0$, p denoting the p-value computed using the regular t-statistic and p^W using heteroskedasticity-robust White t-statistic. The confidence interval is a two-sided interval for θ computed assuming homoskedasticity. The hatted variables are unrestricted ordinary least squares estimates. 95% CI is the 95% confidence interval.

statistical methodology maps this t-statistic to a one-sided p-value of 1.4% under the homoskedasticity assumption. While the t-statistic on *ep* is higher than on our first cross-sectional measure λ^{SRC} (2.81 versus 3.45), the p-value for *ep* is higher than the p-value for λ^{SRC}. This is the motivation for our econometric work in Section 4; the earnings yield is very persistent, and its shocks are strongly negatively correlated with equity-premium shocks, making standard statistical inference misleading.

Using the heteroskedasticity-robust version of our conditional test, which is based on the Eicker-Huber-White *t*-statistics, greatly weakens the case for predictability based on *ep*. When predicting the value-weight CRSP excess return over the entire sample, *ep* is not a statistically significant predictor, with a one-sided *p*-value of 21.5%. The *p*-value for the term yield spread is less affected by the heteroskedasticity adjustment. However, it is also only marginally statistically significant predictor as its *p*-value is 9.5%.

As with most return-prediction exercises, equal-weight index results are a more extreme version of those for the 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 measures (ranging from 3.63 to 4.81), while the term yield spread's is slightly lower (2.45). All OLS *t*-statistics imply rejection of the null at a better than 1% level, even after accounting for the problems caused by persistent and correlated regressors. However, as above, heteroskedasticity adjustment has a dramatic impact on the statistical evidence concerning *ep* (but not for other predictors). While under the homoskedasticity assumption *ep* is significant at better than 1% level, the heteroskedasticity-robust *p*-value is 23%.

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 the importance of heteroskedasticity is to omit this volatile period from estimation. When we estimate the model and *p*-values from the 1946–2002 sample, λ^{SRC} remains statistically significant at a better than 1% level, while the log earnings yield is no longer significant, even at the 10% level.

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 hisweb site, focusing on the top 30% and bottom 30% portfolios sorted on four MSCI 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 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 have low prices (relative to current cash-flow generating ability) and are thus mostly sorted into the value portfolio. Symmetrically, low-beta stocks have relatively high prices and those stocks mostly end up in the growth or high-multiple portfolio. Consequently, a high expected equity premium causes 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 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

Table 5
Predicting the equity premium, country-by-country regressions

Country	Time period	N	$\widehat{\theta}$	OLS t		White t		$\widehat{\rho}$	$\widehat{\gamma}$
Australia	1975:1-2001:12	324	0.0237	1.70	(0.111)	1.60	(0.132)	0.989	−0.229
Austria	1987:1-2001:12	180	−0.0251	−0.61	(0.584)	−0.48	(0.540)	0.935	0.241
Belgium	1975:1-2001:12	324	0.0234	1.36	(0.064)	1.43	(0.056)	0.972	0.196
Denmark	1989:1-2001:12	156	0.0149	0.78	(0.240)	0.78	(0.238)	1.02	0.027
Finland	1988:1-2001:12	168	−0.0150	−0.81	(0.710)	−0.78	(0.701)	1.00	0.151
France	1975:1-2001:12	324	0.0444	2.08	(0.028)	2.08	(0.028)	1.00	0.033
Germany	1975:1-2001:12	324	0.0226	1.32	(0.078)	1.25	(0.089)	0.985	0.018
Hong Kong	1975:1-2001:12	324	0.0200	0.50	(0.278)	0.49	(0.282)	0.977	0.11
Ireland	1991:1-2001:12	132	0.0100	0.39	(0.374)	0.36	(0.386)	0.911	−0.061
Italy	1975:1-2001:12	324	0.0268	0.92	(0.233)	0.92	(0.234)	1.01	0.081
Japan	1975:1-2001:12	324	0.0172	1.40	(0.095)	1.66	(0.058)	0.992	−0.037
Malaysia	1994:1-2001:10	94	0.0418	0.81	(0.089)	0.76	(0.096)	0.918	0.306
Mexico	1982:1-1987:12	72	0.3490	1.41	(0.044)	1.50	(0.036)	0.844	0.311
Netherland	1975:1-2001:12	324	−0.0061	−0.37	(0.596)	−0.30	(0.573)	0.984	0.105
New Zealand	1988:1-2001:12	168	0.0456	1.95	(0.018)	1.98	(0.017)	0.959	0.023
Norway	1975:1-2001:12	324	−0.0053	−0.54	(0.708)	−0.50	(0.697)	0.994	−0.024
Singapore	1975:1-2001:12	324	0.0159	0.76	(0.201)	0.66	(0.229)	0.977	0.094
Spain	1975:1-2001:12	324	0.0366	2.76	(0.004)	2.79	(0.004)	0.986	−0.051
Sweden	1975:1-2001:12	324	0.0177	1.57	(0.046)	1.37	(0.068)	1.01	0.019
Switzerland	1975:1-2001:12	324	−0.0025	−0.15	(0.552)	−0.15	(0.552)	0.974	0.030
United Kingdom	1975:1-2001:12	324	0.0115	0.53	(0.341)	0.44	(0.372)	0.971	−0.153
United States	1926:7-2002:12	918	0.0166	2.41	(0.008)	2.02	(0.019)	0.993	0.089

This table shows results from the model

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

with $\mathrm{Corr}(u_t, v_t) = \gamma$. $x_{t,i} = \lambda^{\mathrm{MSCI}}_{t,i}$ for country i in year t. $\lambda^{\mathrm{MSCI}}_{t,i}$ is constructed by taking the top 30% and bottom 30% portfolios sorted on four Morgan Stanley Capital International 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 ordinary least squares regressions. The OLS t is the homoskedastic t-statistic for testing the null that $\theta = 0$. White t 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 $\theta > 0$. N is the number of observations. The hatted variables are unrestricted ordinary least squares estimates.

precisely. Finally, the parameter estimates for all the countries are similar to those obtained for the United States (with the exception of Mexico with its extremely short sample).

In addition to the country-by-country regressions, we 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
Predicting the equity premium, pooled international regressions

	No FE		Country FE		Country, time FE	
	All	Excluding US	All	Excluding US	All	Excluding US
$\widehat{\theta}$	0.0102	0.0090	0.0132	0.0123	0.00961	0.00756
t Homoskedastic	3.21	2.53	3.76	3.09	3.32	2.34
p Homoskedastic	(0.0007)	(0.0057)	(0.0001)	(0.0010)	(0.0004)	(0.0010)
t Heteroskedastic	2.69	2.12	3.31	2.73	2.97	2.14
p Heteroskedastic	(0.0036)	(0.0171)	(0.0005)	(0.0032)	(0.0015)	(0.016)
p Clustering by year	2.08	1.65	2.31	1.89	2.57	1.78
p Clustering by year	(0.0189)	(0.0494)	(0.0105)	(0.0295)	(0.0051)	(0.0376)
$\widehat{\rho}$	0.992	0.992	0.990	0.990	0.988	0.987
$\widehat{\gamma}$	0.0519	0.0469	0.0545	0.0497	0.0578	0.0483

This table shows results from the model

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

with $\mathrm{Corr}(u_{t,i}, v_{t,i}) = \gamma$. $x_{t,i} = \lambda^{\mathrm{MSCI}}_{t,i}$ for country i in year t. $\lambda^{\mathrm{MSCI}}_{t,i}$ is constructed by taking the top 30% and bottom 30% portfolios sorted on four Morgan Stanley Capital International 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. t homoskedastic and t heteroskedastic indicate the usual ordinary least squares (OLS) t-statistic and the heteroskedasticity-robust White t-statistic. t clustering by year indicates that we calculate standard errors robust to correlations between firms as well as heteroskedasticity, but assume independence over time. The analogously defined 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.

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 or exclusion of the US data and inclusion or exclusion of country fixed effects in the pooled regression. All p-values are under 5%. Thus we conclude 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 subsumes 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, at least under the homoskedasticity assumption. 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%). Because 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. Allowing for heteroskedasticity changes this conclusion, however. Using the heteroskedasticity-robust test, the p-values are always much larger for ep.

Though the horse race between λ^{SRC} and ep is a draw under the homoskedasticity assumption, many of our alternative cross-sectional measures win 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 forecasts of 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.

5.5. Implications of premia divergence in the 1980s

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 1980s. Curiously, as Fig. 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

Table 7
Multivariate predictors of the excess value-weight market return (R^e_M)

Specification	$\widehat{\theta}_1$	t_1	p_1	p_1^W	$\widehat{\theta}_2$	t_2	p_2	p_2^W	Fp	Fp^W
Prediction equation: $R^e_{M,t} = \theta_0 + \theta_1 \lambda^{\mathrm{SRC}}_{t-1} + \theta_2 ep_{t-1} + u_t$										
1927:5-2002:12	0.012	1.17	0.158	0.191	0.013	2.32	0.108	0.294	0.020	0.409
1927:5-1965:2	0.001	0.045	0.532	0.533	0.031	2.17	0.091	0.366	0.056	0.596
1965:2-2002:12	0.011	0.527	0.285	0.295	0.008	1.36	0.447	0.516	0.592	0.696
1927:5-1946:3	−0.002	−0.030	0.573	0.571	0.042	1.79	0.176	0.342	0.139	0.585
1946:3-1965:2	0.037	2.03	0.038	0.039	0.004	0.225	0.798	0.784	0.086	0.082
1965:2-1984:1	0.027	1.08	0.183	0.209	0.023	2.03	0.172	0.209	0.210	0.283
1984:1-2002:12	−0.034	−0.738	0.778	0.756	0.012	1.39	0.341	0.371	0.527	0.566
Prediction equation: $R^e_{M,t} = \theta_0 + \theta_1 \lambda^{\mathrm{SRC}}_{t-1} + \theta_2 TY_{t-1} + u_t$										
1927:5-2002:12	0.023	2.49	0.006	0.051	0.002	0.564	0.289	0.307	0.017	0.126
1927:5-1965:2	0.038	2.23	0.017	0.095	−0.001	−0.132	0.604	0.582	0.051	0.208
1965:2-2002:12	0.005	0.233	0.357	0.360	0.003	0.791	0.184	0.181	0.670	0.666
1927:5-1946:3	0.069	1.84	0.043	0.162	−0.001	−0.158	0.634	0.614	0.160	0.337
1946:3-1965:2	0.035	2.55	0.006	0.014	0.010	0.912	0.201	0.233	0.005	0.003
1965:2-1984:1	0.010	0.398	0.280	0.283	0.008	1.61	0.038	0.047	0.229	0.223
1984:1-2002:12	−0.012	−0.257	0.524	0.514	−0.002	−0.403	0.633	0.649	0.849	0.849
Full sample results for alternative cross-sectional risk premium measures *Prediction equation*: $R^e_{M,t} = \theta_0 + \theta_1 x_{t-1} + \theta_2 ep_{t-1} + u_t$										
$x_t = \lambda^{\mathrm{REG}}_t$	0.066	2.36	0.015	0.030	0.012	2.12	0.151	0.350	0.003	0.283
$x_t = \lambda^{\mathrm{DP}}_t$	0.0200	1.32	0.107	0.265	0.0143	2.69	0.056	0.175	0.005	0.336
$x_t = \lambda^{\mathrm{DPG}}_t$	0.0076	0.516	0.321	0.409	0.0160	3.02	0.027	0.140	0.010	0.261
$x_t = \lambda^{\mathrm{BM}}_t$	0.0005	1.03	0.174	0.330	0.0140	2.46	0.062	0.101	0.007	0.290
$x_t = \lambda^{\mathrm{BMG}}_t$	0.0009	1.71	0.050	0.257	0.0131	2.44	0.078	0.122	0.003	0.294
$x_t = \lambda^{\mathrm{ER}}_t$	1.766	2.40	0.019	0.029	0.0081	1.57	0.284	0.393	0.011	0.057
Full sample results for alternative cross-sectional risk premium measures *Prediction equation*: $R^e_{M,t} = \theta_0 + \theta_1 x_{t-1} + \theta_2 TY_{t-1} + u_t$										
$x_t = \lambda^{\mathrm{REG}}_t$	0.088	3.35	0.000	0.016	0.001	0.474	0.316	0.333	0.001	0.046
$x_t = \lambda^{\mathrm{DP}}_t$	0.0335	2.38	0.013	0.172	0.0032	1.13	0.136	0.158	0.017	0.248
$x_t = \lambda^{\mathrm{DPG}}_t$	0.0215	1.53	0.066	0.225	0.0032	1.13	0.129	0.153	0.121	0.343
$x_t = \lambda^{\mathrm{BM}}_t$	0.0010	2.39	0.012	0.226	0.0026	0.910	0.184	0.223	0.028	0.275
$x_t = \lambda^{\mathrm{BMG}}_t$	0.0014	2.82	0.003	0.211	0.0029	1.03	0.159	0.174	0.006	0.272
$x_t = \lambda^{\mathrm{ER}}_t$	2.154	3.08	0.001	0.001	0.0005	0.172	0.400	0.411	0.006	0.014

$\widehat{\theta}_i$ is the ordinary least squares estimate of θ_i, with $\theta_i = (\theta_1\ \theta_2)'$ in the model $R^e_{M,t} = \mu_1 + \theta' \mathbf{x}_{t-1} + u_t$; $\mathbf{x}_t = \boldsymbol{\mu}_2 + K\mathbf{x}_{t-1} + V_t$. t is the usual t-statistic for testing the null $\theta_i = 0$. The table also reports the p-values for testing the null $\theta_i = 0$ against $\theta_i > 0$, p denoting the p-value computed using the regular t-statistic (t) and p^W using heteroskedasticity–robust White t-statistic. $F\,p$ and $F\,p^W$ denote the p-values for the F-test of the null $\theta_1 = \theta_2 = 0$ with and without imposing the homoskedasticity assumption. All p-values are computed using the conditional bootstrap described in Appendix B.

forecast a low equity premium, while the smoothed earnings yield forecasts a high equity premium.

If *ep* is 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 value-weight 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 in-sample unconditional beta, but our subsequent results are robust to using more elaborate and complex portfolio constructions schemes. 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). Although we do not have a clear prior about the unconditional mean of R^e_{arb}, a natural alternative hypothesis is that the coefficient on λ^{SRC} should be negative while the coefficient on *ep* should be positive. R^e_{arb} is a combination of two bets: (1) buying the market on margin and (2) hedging this equity-premium bet by shorting high-beta stocks and investing the proceeds in low-beta stocks. First, holding the cross-sectional beta premium among stocks constant, a higher

Table 8
Multivariate predictors of the hedged value-weight market return (R^e_{arb})

Specification	$\widehat{\theta}_1$	t_1	p_1	p_1^W	$\widehat{\theta}_2$	t_1	p_1	p_1^W	Fp	Fp^W
Prediction equation: $R^e_{\text{arb},t} = \theta_0 + \theta_1 \lambda^{\text{SRC}}_{t-1} + \theta_2 ep_{t-1} + u_t$										
1927:6-2002:12	−0.030	−2.87	0.002	0.001	0.007	1.13	0.140	0.249	0.013	0.013
1927:6-1965:2	−0.018	−0.965	0.161	0.168	−0.014	−1.10	0.877	0.763	0.018	0.014
1965:2-2002:12	−0.063	−2.29	0.016	0.016	0.010	1.34	0.108	0.196	0.022	0.007
1927:6-1946:3	0.004	0.087	0.522	0.514	−0.018	−0.907	0.821	0.764	0.440	0.728
1946:3-1965:2	−0.016	−1.05	0.150	0.139	0.009	0.600	0.486	0.468	0.605	0.608
1965:2-1984:1	−0.070	−2.75	0.006	0.009	0.007	0.599	0.300	0.308	0.012	0.011
1984:1-2002:12	−0.083	−1.19	0.121	0.152	0.025	1.89	0.028	0.042	0.126	0.108
Full sample results for alternative cross-sectional risk premium measures										
Prediction equation: $R^e_{arb,t} = \theta_0 + \theta_1 x_{t-1} + \theta_2 ep_{t-1} + u_t$										
$x_t = \lambda^{\text{REG}}$	−0.062	−2.32	0.011	0.015	0.001	0.467	0.305	0.384	0.062	0.085
$x_t = \lambda^{\text{DP}}$	−0.072	−4.66	0.000	0.000	0.0073	1.34	0.099	0.189	0.000	0.005
$x_t = \lambda^{\text{DPG}}$	−0.066	−4.36	0.000	0.000	0.0064	1.19	0.130	0.225	0.000	0.006
$x_t = \lambda^{\text{BM}}$	−0.0029	−5.80	0.000	0.000	0.0146	2.53	0.007	0.033	0.000	0.001
$x_t = \lambda^{\text{BMG}}$	−0.0029	−5.31	0.000	0.001	0.0098	1.79	0.041	0.097	0.000	0.004
$x_t = \lambda^{\text{ER}}$	−2.21	−2.61	0.005	0.011	0.0114	1.94	0.036	0.104	0.019	0.063

$\widehat{\theta}_i$ is the ordinary least squares estimate of θ_i, with $\theta_i = (\theta_1\ \theta_2)'$ in the model $R^e_{\text{arb},t} = \mu_1 + \theta' \mathbf{x}_{t-1} + u_t$; $\mathbf{x}_t = \boldsymbol{\mu}_2 + K\mathbf{x}_{t-1} + V_t$. R^e_{arb} is the return on a zero-investment portfolio consisting of 1.21 times the value-weight excess market return, minus the return difference between the highest-beta (10) and lowest-beta (1) deciles. The table reports the *p*-values for testing the null $\theta_i = 0$ against $\theta_i > 0$, *p* denoting the *p*-value computed using the regular *t*-statistic (*t*) and p^W using heteroskedasticity-robust White *t*-statistic. *Fp* and Fp^W denote the *p*-values for the *F*-test of the null $\theta_1 = \theta_2 = 0$ with and without imposing the homoskedasticity assumption. All *p*-values are computed using the conditional bootstrap described in Appendix B.

expected equity premium (as evidenced by a high *ep*) should translate into a high expected return on R^e_{arb}. Second, holding the expected equity premium constant, a higher cross-sectional beta premium (manifest by a high λ^{SRC}) should translate into a low expected return on R^e_{arb}. Thus, one-sided tests are appropriate.

The variables are jointly significant for the full period as well as for both subperiods. However, because 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.

These results on the predictabilty of R^e_{arb} are relatively insensitive to using tests that are more robust to heteroskedasticity. While the volatility of the realized equity premium is systematically related to the earnings yield, the volatility of the beta-hedged market return is much less so. Therefore, the relative insensitivity of these tests to heteroskedasticity adjustments makes sense.

Even a cursory examination of the fitted values suggests that these predictability results are also economically significant. In the beginning of year 1982, the predicted value for R^e_{arb} is over 20% annualized in the regression that uses λ^{SRC} and *ep* as forecasting variables. (For reference, this conditional mean is more than three standard errors from zero.) Because the unconditional volatility of R^e_{arb} 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 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, the 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 the 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.

Because 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, b; 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 highly correlated with 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. Our claim is not that one should use the CAPM and the Fed model for relative valuation of stocks. We merely show 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 has asked 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, 1988a, b; Fama and French, 1988, 1989; Hodrick, 1992). Although the usual asymptotic p-values indicate 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 homoskedastic Gaussian errors.

We use a novel method to solve for the exact small-sample p-values in the case of homoskedastic Gaussian errors. The method is based on the Jansson and Moreira (2003) idea of first 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) with homoskedastic Gaussian errors, we provide the finance community with a function that enables an applied researcher to implement a correctly sized test of predictability in seconds.

Appendix A. λ^{SRC} identification requires cross-sectional variation in growth rates

This appendix explains how identifying expected-equity premium variation with the measure λ^{SRC} requires cross-sectional variation in expected growth rates. Let D/P_i be the dividend yield, g_i the expected growth rate (in excess of the risk-free rate), and β_i the beta for stock i. Let k^* be the typical expected excess return on the market and k_e the deviation of the expected excess return on the market from k^*. The Gordon model and the CAPM state that $D/P_i = \beta_i(k^* + k_e) - g_i$. We calibrate the model such that when the equity premium is at k^*, the cross-sectional variance of g_i is c times that of $\beta_i k^*$. We make two

further simplifying assumptions that (1) expected growth rates and discount rates are uncorrelated and that (2) both variables are uniformly distributed. (Uniform distributions are convenient, because in this case simple correlations are equal to rank correlations in large samples.)

Simple algebra shows that the (rank) correlation (Γ) between dividend yields and betas is equal to

$$\Gamma = \frac{k^* + k_e}{\sqrt{(k^* + k_e)^2 + ck^{*2}}}. \tag{19}$$

The change in this correlation in response to a small change in k_e at $k_e = 0$ is

$$\left.\frac{\partial \Gamma}{\partial k_e}\right|_{k_e=0} = \frac{ck^*}{[k^{*2} + ck^{*2}]^{3/2}}. \tag{20}$$

As long as c is positive and growth rates vary across stocks, the (rank) correlation of β_i with D/P_i will vary with the equity premium. Furthermore, as one can see from these formulas, increasing the cross-sectional variation in expected growth rates makes many of our ordinal equity-premium measures more sensitive to changes in the equity premium, at least as long as betas and expected growth rates are not correlated in the cross-section.

We calibrate the above equations using an estimate of c from Cohen et al. (2003, 2005b). Those authors estimate that approximately 75% of the cross-sectional variation in valuation multiples is caused by expected growth rates and only 25% by discount rates, giving a variance ratio of $c = 75\%/25\% = 3$. The following exhibit illustrates how even a very modest level of cross-sectional spread in growth rates allows changes in the (rank) correlation between beta and dividend yield to be strongly related to changes in the equity premium.

c	0	0.5	1	2	**3**	4	5
$k^* =$	0.07	0.07	0.07	0.07	**0.07**	0.07	0.07
Γ	1.00	0.82	0.71	0.58	**0.50**	0.45	0.41
$\frac{\partial \Gamma}{\partial k_e}$	0.00	55.54	72.15	78.55	**76.53**	73.01	69.43
$\frac{\partial \Gamma}{\partial k_e}/100$	0.00	0.56	0.72	0.79	**0.77**	0.73	0.69
$\frac{\partial \Gamma}{\partial k_e}/\Gamma$	0.00	68.03	102.04	136.05	**153.06**	163.27	170.07
$\frac{\partial \Gamma}{\partial k_e}/(100\Gamma)$	0.00	0.68	1.02	1.36	**1.53**	1.63	1.70

Appendix B. Statistical appendix

B.1. Algorithm for computing q_α^{nn}

In this section, we describe the algorithm for computing q_α^{nn}, the neural network approximation to the critical values defined in Eq. (10). We choose the parameters of the neural net to minimize the sum of squared size distortions over a grid of

(ρ, γ, T) values:

$$(\widehat{\psi}, \widehat{\xi}) = \text{argmin}_{(\psi,\xi)} \sum_{j=1}^{J} \sum_{i=1}^{N} \left(\alpha_j - B^{-1} \sum_{b=1}^{B} 1_h[\widehat{t}_{b,i} > q_{\alpha_j}^{nn}(X_{b,i}, \psi, \xi)] \right)^2. \tag{21}$$

$i = 1, \ldots, N$ indexes a grid of (ρ_i, γ_i, T_i) triplets. For each i, we simulate B data sets from the null model, with $\theta = 0$, i.i.d. 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 b^{th} simulated sample generated from (ρ_i, γ_i, T_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:

$$\Pr[\widehat{t} > q_\alpha^{nn}(X, \psi, \xi); \rho_i, \gamma_i, T_i] \approx B^{-1} \sum_{b=1}^{B} 1[\widehat{t}_{b,i} > q_\alpha^{nn}(X_{b,i}, \psi, \xi)]. \tag{22}$$

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

$$1_h(x) = (1 + \mathrm{e}^{-x/h})^{-1}. \tag{23}$$

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

The neural net critical values used in this paper were computed setting $B = 10,000$ and $h = 0.01$. Reparameterizing $\rho = 1 + c/T$, the grid points were all possible combinations of

$$\begin{aligned} -c &\in (T, 75, 50, 30, 20, 15, 12, 10, 8, 6, 4, 2, 0)/T, \\ -\gamma &\in (0, 0.2, 0.4, 0.5, 0.6, 0.7, 0.8, 0.85, 0.86, 0.88, 0.90, 0.92, 0.94, 0.96, 0.98), \\ T &\in (60, 120, 240, 480, 840), \\ \alpha &\in (0.01, 0.025, 0.05, 0.10, 0.50, 0.90, 0.95, 0.975, 0.99). \end{aligned} \tag{24}$$

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.

Even though we simulated only over negative correlations, q_α^{nn} is valid over the entire range, $-1 < \gamma < 1$. To understand why, first consider the case in which γ is negative. Our rejection rule is

$$\text{if } \widehat{\gamma} < 0 \quad \text{reject when } \widehat{t} > -\mu(X) + \sigma(X)\Phi^{-1}(.95). \tag{25}$$

Next consider the case in which γ is positive. We replace x_t by $-x_t$, thus reversing the sign of the correlation and making our approximate quantile function valid. This transformation also reverses the sign of θ, so instead of testing the null $\theta = 0$ against the positive alternative $\theta > 0$, we test the null $\theta = 0$ against the negative alternative $\theta < 0$. Instead of rejecting when the t-statistic $\widehat{t}$ (computed from x_t) is greater than the 95% quantile of the conditional null distribution we reject when the transformed t-statistic $-\widehat{t}$ (computed from $-x_t$) is less than the 5% quantile of the conditional null. Because X is invariant to replacing x_t with $-x_t$, we reject when

$$-\widehat{t} < -\mu(X) + \sigma(X)\Phi^{-1}(.05). \tag{26}$$

Because $\Phi^{-1}(\alpha) = -\Phi^{-1}(1-\alpha)$, the rejection rule becomes

$$\text{if } \widehat{\gamma} > 0, \quad \text{reject when } \widehat{t} > \mu(X) + \sigma(X)\Phi^{-1}(.95). \tag{27}$$

This leads to our general rejection rule $\widehat{t} > \text{sign}(\widehat{\gamma})\mu(X) + \sigma(X)\Phi^{-1}(0.95)$, valid whether $\widehat{\gamma}$ is positive or negative.

The simulated sample sizes range from 60 to 840 observations. The quantile function perhaps is not accurate for samples smaller than 60 observations, but unreported Monte Carlo simulations indicate that it is accurate for any sample size greater than 60 observations, including samples larger than 840 observations. There are sound theoretical reasons to believe that the function works for samples larger than 840 observations. As T becomes large, asymptotic approximations become accurate. If x_t is stationary, the input vector X converges to $(1, 0, 0, 1, \gamma)'$ and the critical function returns the usual standard normal approximation. If ρ is modeled as a unit root (or local to unity) as in Jansson and Moreira (2003), their asymptotic approximations imply that the conditional critical function $q(S, \alpha)$ converges to a function that does not depend on T and delivers the correct test size in any large sample. So our critical function returns asymptotically sensible critical values whether we have a unit root or not.

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) 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, p. 262), we randomly draw each element of ψ and ξ from an independent Normal$(0, 1/5)$ distribution. We then iterate from the starting values for ψ and ξ using the Broyden-Fletcher-Goldfarb-Shanno optimization algorithm. All computations were done using Ox 3.00, a programming language described in Doornik (2001).

We repeated this algorithm for many different randomly drawn starting values for ψ and ξ. Some of the starting values led to solutions with minimal size distortions; the rejection frequencies were visually similar to those in Fig. 1. Some starting values converged at parameters that did not lead to accurate solutions.

B.2. Constructing confidence intervals for the univariate case

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 $\mathscr{P} = (\mu_1, \mu_2, \theta, \rho, \sigma_u^2, \sigma_v^2, \gamma)'$ 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:

$$\inf_{\mathscr{P}} \Pr[\theta \in C; \mathscr{P}] \geqslant \alpha \quad \text{for all } \mathscr{P}. \tag{28}$$

C is a random interval, because it is a function of the data, and $\Pr[\theta \in C; \mathscr{P}]$ denotes the probability that θ is in C given the parameters $\mathscr{P}$. 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:

$$\begin{aligned}\Pr[\overline{\theta} \in C; \mathscr{P}] &\equiv \Pr[\text{fail to reject null } \theta = \overline{\theta} \text{ when null is true}; \mathscr{P}] \\ &= 1 - \alpha \text{ for all } \mathscr{P}.\end{aligned} \tag{29}$$

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.[8]

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

$$\widetilde{y}_t = \mu_1 + \widetilde{\theta} x_{t-1} + u_t, \tag{30}$$

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

B.3. Conditional bootstrap algorithm for the multivariate model

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 $\widehat{\Sigma}^{1/2}(\widehat{\Sigma}^{1/2})' = \widehat{\Sigma}$. 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

$$X = (\mathrm{vec}(\widehat{K}_R)' \, \mathrm{se}(x_1) \, \mathrm{se}(x_2) \, \widehat{\mathrm{Corr}}(x_1, x_2))', \tag{31}$$

where $\mathbf{x}_t = (x_{1,t}, x_{2,t})$, $[\mathrm{se}(x_1)]^2 = 1/(T-1) \sum (x_{1,t-1} - \overline{x}_1)^2$ is the estimated variance of x_1, $[\mathrm{se}(x_2)]^2$ is the estimated variance of x_2, and $\widehat{\mathrm{Corr}}(x_1, x_2)$ 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 $d_b = \max_i |(X_i - X_{b,i})/s_i|$, 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 $d_{(1)} \leqslant d_{(2)} \leqslant \cdots \leqslant d_{(B)}$. Let $\mathscr{D}$ denote the set of $\widehat{t}_b$ where the corresponding X_b is among the N that are nearest to the actual sufficient statistics:

$$\widehat{t}_b \in \mathscr{D} \text{ iff } \; d_{(b)} \leqslant d_{(N)}. \tag{32}$$

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

(5) The set of draws $\mathscr{D}$ are treated as draws from the conditional distribution of $\widehat{t}$ given S. We estimate the 100αth quantile of the conditional distribution with the 100αth empirical quantile of the sample of draws $\mathscr{D}$.

This bootstrap procedure computes a nonparametric nearest neighbor estimate of the conditional quantile of $\widehat{t}$ given X. Chaudhuri (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, because X is a high-dimensional vector the curse of dimensionality requires B to be extraordinarily large, possibly in the billions. Thus if we take the Chaudhuri (1991) theory literally, it is not computationally feasible to precisely estimate the conditional quantile. However, the Monte Carlo results reported below suggest that the conditional bootstrap accomplishes the more modest goal of improving on the parametric bootstrap.

We simulate five thousand 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^e_M, 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 test the null $\theta_1 = 0$ against the one-sided alternative $\theta_1 > 0$. We compute critical values using the parametric bootstrap and the conditional bootstrap. For the parametric bootstrap we simulate 20 thousand 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 is computed taking $B = 20,000$ and $N = 1,000$.

The above experiment yields the following results. When θ_1 is the coefficient on λ_t^{SRC}, the parametric bootstrap rejects the null 5.48% of the time and the conditional bootstrap rejects 3.84% of the time. When θ_1 is the coefficient on ep_t the rejection frequencies are 11.46% and 4.78%, respectively. We then simulate 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 rejects 6.36% of the time and the conditional bootstrap rejects 3.32% of the time. When θ_1 is the coefficient on ep_t the rejection frequencies are 15.64% and 7.80%. These Monte Carlo results suggests that conditional inference yields a significant improvement even in the computationally more challenging multivariate problem.

We choose the N and B used for Tables 7 and 8 as follows. For the p-values that test the nulls $\theta_1 = 0$, $\theta_2 = 0$, and $\theta_1 = \theta_2 = 0$, we set $N = 10,000$ and $B = 200,000$.

B.4. Discussion of test power

The Monte Carlo experiments in Section 4, as well as the theory underlying conditional inference, demonstrate that the size of our conditional test is correct. The test is also powerful. Jansson and Moreira (2003) show that the conditional t test is most powerful among the class of similar tests. Similar tests deliver correct size for all values of the nuisance parameters ρ, μ_1, and μ_2. Therefore, a test can only be more powerful than ours if it either under-rejects or over-rejects for some values of the nuisance parameters. One way to add power to a test is to make strong a priori assumptions about the values of ρ, μ_1, and μ_2. For example, if one knows the value of ρ, then it is straightforward to construct a test that is more powerful than ours. If one's belief in ρ is incorrect, the test has power or size problems or both.

In particular, our procedure has good power relative to alternative procedures proposed by Lewellen (2004), Torous et al. (2005), and Campbell and Yogo (2006). Lewellen (2004) provides a good example of a clever test that improves power by making strong

assumptions about nuisance parameters. Lewellen (2004) derives an one-sided test (alternative hypothesis is $\theta > 0$) assuming that $\rho = 1$ and $\gamma < 0$. If we know for sure that ρ equals one, then Lewellen's test has a power advantage over our test. However, Lewellen (2004) shows that the power of his test declines dramatically as ρ declines. To address the power issue, Lewellen (2004) also describes a Bonferroni procedure to improve power for ρ below one. (For some of our predictive variables $\rho < 1$ and $\gamma > 0$, thus Lewellen's tests would over-reject the null.)

Torous et al. (2005) use a Bonferroni procedure first proposed by Cavanaugh et al. (1995). They form a confidence interval for ρ, then construct the optimal test of $\theta = 0$ at all values of ρ in the interval. If none of the tests rejects for any ρ in the interval, they do not reject the null that $\theta = 0$. Their test is also designed to have high power if ρ is close to one.

Campbell and Yogo (2006) make stronger assumptions about the nuisance parameters than we do, and if their assumptions are correct, their test could therefore be more powerful. Campbell and Yogo (2006) and Torous et al. (2005) both use Bonferroni tests that rely on a first-stage confidence interval for ρ. Campbell and Yogo (2006) differ in that they form the interval using a newer, more powerful test: the Dickey-Fuller generalized least squares (DF-GLS) test of Elliot et al. (1996). The DF-GLS test is potentially more powerful than traditional methods but also makes stronger assumptions about nuisance parameters.

To better understand Campbell and Yogo's approach in practice, suppose that the intercept in the second equation, μ_2, is known to be zero and does not need to be estimated. Incorporating this prior information into a test should improve the power of the test. In practical applications μ_2 is probably not exactly zero. However, suppose that μ_2 is by some metric small, and further suppose that ρ is close to one. Then, over time, x_t varies so much that it dominates any small value of μ_2 (in other words, μ_2 is small relative to the standard deviation of x_t,) and the data behave approximately as if μ_2 were zero. Therefore, one can construct tests under the assumption that μ_2 is zero, and in a large sample the assumption does not lead to size problems. This approximation is the basis for the DF-GLS test used by Campbell and Yogo. In summary, Campbell and Yogo are thus implicitly making the assumption that ρ is close to one and μ_2 is close to zero.

Monte Carlo experiments confirm the above logic. The tests by Lewellen (2004), Torous et al. (2005), and Campbell and Yogo (2006) have better power than our test if ρ is close to one and γ is close to negative one. As ρ decreases and γ increases, our test becomes more powerful. In the simulations, we focus on a long (1926–2002) monthly time series of log dividend yields and excess returns, available from Motohiro Yogo's website. We use this time series for the following reasons. First, this series is the most persistent of the commonly used predictor variables. Second, shocks to dividend yield have much stronger negative correlation with returns than shocks to any other of the commonly used predictor variables. Third, this series has a long time span, and thus many of the asymptotic approximations are likely be accurate. Fourth, a series that starts in 1926 has a relatively low intial value, which is relevant for some of the tests we examine.

For our Monte Carlo experiments, we estimate the following model:

$$\begin{aligned} y_t &= \mu_1 + \theta x_{t-1} + u_t, \\ x_t &= \widetilde{\mu}_2 + z_t, \quad \text{and} \\ z_t &= \rho z_{t-1} + v_t, \end{aligned} \tag{33}$$

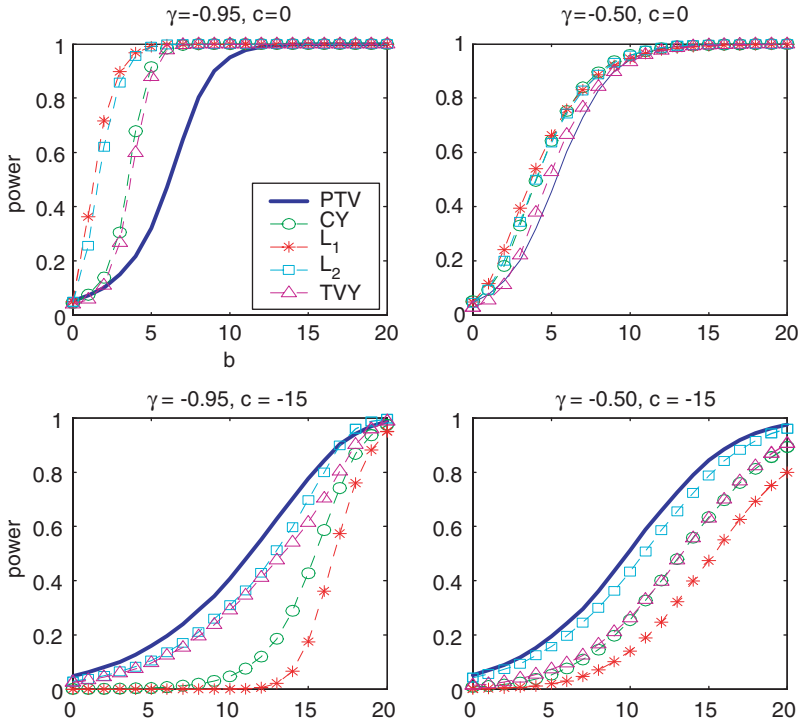

Fig. 3. Power of alternative tests. The plots compare small-sample rejection probabilities for various tests. The labels are Polk-Thompson-Vuolteenaho (PVT), Campbell-Yogo (CY), Lewellen's test (L_1), Lewellen's Bonferroni test (L_2), and Torous, Valkanov, and Yan's Bonferroni test (TVY). We provide results for various values of $c = T(\rho - 1)$, $b = \theta/T$, and γ. There are ten thousand Monte Carlo simulations.

with $\rho = 1 + c/T$ and $\theta = b/T$. This specification appears in the current version of Campbell and Yogo's working paper. It is equivalent to the two-equation system in Eq. (6) that we use in the rest of the paper, with $\mu_2 = \widetilde{\mu}_2(1-\rho)$. The point estimates are

$$\begin{aligned}
&\mu_1 = 0.0319, \quad \widetilde{\mu}_2 = -3.3621, \quad x_0 = -3.0015, \quad z_0 = 0.3606, \quad \text{and} \\
&\sigma_1 = \sqrt{\text{Var}(u_t)} = 0.0545, \quad \sigma_2 = \sqrt{\text{Var}(v_t)} = 0.0565, \\
&\gamma = \text{Corr}(e_t, v_t) = -0.9543.
\end{aligned} \tag{34}$$

The confidence interval for $c \equiv T(\rho - 1)$, based on inverting the Dickey-Fuller statistic, ranges approximately from $c = -16$ to $c > 0$.

Fig. 3 compares the small-sample rejection probabilities of five alternative tests: our test (PTV), the Campbell-Yogo test (CY), Lewellen's test (L_1), Lewellen's Bonferroni test (L_2), and the Bonferroni test of Torous, Valkanov, and Yan (TVY). The Monte Carlo results use the parameter estimates and starting values given above, with a few variations. We set $c = 0$, i.e., a unit root, and $c = -15$, both of which are in the confidence interval. The value $c = -15$ calibrates to $\rho = 1 + c/T = 0.9836$ for 913 observations. We set the correlation $\gamma = -0.95$, the correlation in Campbell and Yogo's data, and $\gamma = -0.50$, which is closer to the correlations we see for many other predictor series.

We calculate the CY test by following the instructions in their Appendix C, except that we do not implement any of the corrections for serial correlation in the errors e_t or v_t. To implement the CY test we electronically store the critical-value tables from the most recent version of their working paper.

The Lewellen test (L_1) is optimal for $\rho = 1$. As Lewellen has shown, it has poor power when ρ is below one. Lewellen proposes but did not empirically implement a Bonferroni version of the test (L_2) to improve power for small ρ. Our implementation of the (L_2) test is partly based on suggestions and clarifications by Jonathan Lewellen in private correspondence. The p-value for Lewellen's Bonferroni test is

$$p_{\text{bonferroni}} = \min[2P, P + D], \tag{35}$$

where

$$P = \min[p_{\text{lewellen}}, p_{\text{stambaugh}}], \tag{36}$$

and D is the p-value for a unit root test of $\rho = 1$, based on the sampling distribution of $\widehat{\rho}$. p_{lewellen} is the p-value for the Lewellen test (L_1), and $p_{\text{stambaugh}}$ is the p-value for the Stambaugh test. We calculate $p_{\text{stambaugh}}$ assuming that the Stambaugh bias-corrected t-statistic is normally distributed. We also ran the simulations using a bootstrap procedure to calculate the p-values. The results are qualitatively similar.

When c equals zero and the correlation γ is very close to negative one, the competing procedures (CY, L_1, L_2, TVY) have more power than our test. The competing procedures also do not over-reject in this case. This is not surprising, as the alternative tests are developed with this particular situation in mind. Our test is superior in terms of power for smaller values of c and larger values of γ. When $c = -15$, many of the alternative procedures have size distortions, causing them to under-reject the null and have low power. The CY and L_1 tests have particularly low power for $c = -15$ and $\gamma = -0.50$. In particular, for the predictor variables and sample periods we test in our paper, c point estimates range approximately from -1 to -75 and γ point estimates from -0.7 to $+0.4$. As the above experiments show, these ranges include many values of γ for which our test is superior.

References

Adrian, T., Franzoni, F., 2002. Learning about beta: an explanation of the value premium. Unpublished working paper. Massachusetts Institute of Technology, Cambridge, MA.

Ang, A., Bekaert, G., 2001. Stock return predictability: is it there? Unpublished working paper. Columbia University Graduate School of Business, New York.

Ang, A., Chen, J., 2004. CAPM over the long-run: 1926–2001. Unpublished working paper. University of Southern California, Los Angeles, CA.

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

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

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

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

Bishop, C.M., 1995. Neural Networks for Pattern Recognition. Oxford University Press, New York.

Black, F., 1972. Capital market equilibrium with restricted borrowing. Journal of Business 45, 444–454.
Campbell, J.Y., 1987. Stock returns and the term structure. Journal of Financial Economics 18, 373–399.
Campbell, J.Y., Cochrane, J.H., 1999. Force of habit: a consumption-based explanation of aggregate stock market behavior. Journal of Political Economy 107, 205–251.
Campbell, J.Y., Shiller, R.J., 1988a. Stock prices, earnings, and expected dividends. Journal of Finance 43, 661–676.
Campbell, J.Y., Shiller, R.J., 1988b. The dividend–price ratio and expectations of future dividends and discount factors. Review of Financial Studies 1, 195–228.
Campbell, J.Y., Shiller, R.J., 1998. Valuation ratios and the long-run stock market outlook. Journal of Portfolio Management 24 (2), 11–26.
Campbell, J.Y., Vuolteenaho, T., 2003. Bad beta, good beta. Unpublished working paper. Harvard University, Cambridge, MA.
Campbell, J.Y., Yogo, M., 2006. Efficient tests of stock return predictability. Journal of Financial Economics, forthcoming.
Cavanaugh, C., Elliot, G., Stock, J., 1995. Inference in models with nearly nonstationary regressors. Econometric Theory 11, 1131–1147.
Chaudhuri, P., 1991. Nonparametric quantile regression. Annals of Statistics 19, 760–777.
Chen, X., White, H., 1999. Improved rates and asymptotic normality for nonparametric neural network estimators. IEEE Transactions in Information Theory 45, 682–691.
Cohen, R., Polk, C., Vuolteenaho, T., 2003. The value spread. Journal of Finance 58, 609–641.
Cohen, R., Polk, C., Vuolteenaho, T., 2005a. Money illusion in the stock market: the Modigliani-Cohn hypothesis. Quarterly Journal of Economics 120, 639–668.
Cohen, R., Polk, C., Vuolteenaho, T., 2005b. The price is (almost) right. Unpublished working paper, Northwestern University and Harvard University, Evanston, IL, Cambridge, MA.
Davis, J.L., Fama, E.F., French, K.R., 2000. Characteristics, covariances, and average returns: 1929 to 1997. Journal of Finance 55, 389–406.
Doornik, J.A., 2001. Object-Oriented Matrix Programming using Ox 3.0. Timberlake Consultants Ltd. Oxford, London, England, www.nuff.ox.ac.uk/Users/Doornik.
Elliot, G., Rothenberg, T., Stock, J., 1996. Efficient tests for an autoregressive unit root. Econometrica 64, 813–836.
Fama, E.F., 1998. Determining the number of priced state variables in the ICAPM. Journal of Financial and Quantitative Analysis 33, 217–231.
Fama, E.F., French, K.R., 1988. Dividend yields and expected stock returns. Journal of Financial Economics 22, 3–27.
Fama, E.F., French, K.R., 1989. Business conditions and expected returns on stocks and bonds. Journal of Financial Economics 25, 23–49.
Fama, E.F., French, K.R., 1992. The cross-section of expected stock returns. Journal of Finance 47, 427–465.
Fama, E.F., French, K.R., 1999. Forecasting profitability and earnings. Journal of Business 73, 161–176.
Fama, E.F., French, K.R., 2002. The equity premium. Journal of Finance 57, 637–659.
Franzoni, F., 2002. Where is beta going? The riskiness of value and small stocks. Unpublished working paper. Massachusetts Institute of Technology, Cambridge, MA.
Gordon, M., 1962. The Investment, Financing, and Valuation of the Corporation. Irwin, Homewood, IL.
Graham, B., Dodd, D.L., 1934. Security Analysis, first ed. McGraw-Hill, New York.
Hodrick, R.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, M., Moreira, M., 2003. Conditional inference in models with nearly nonstationary regressors. Unpublished working paper. Harvard University, Cambridge, MA.
Keim, D., Stambaugh, R., 1986. Predicting returns in the stock and bond markets. Journal of Financial Economics 17, 357–390.
Kothari, S.P., Shanken, J., Sloan, R.G., 1995. Another look at the cross-section of expected stock returns. Journal of Finance 50, 185–244.
Lakonishok, J., Shleifer, A., Vishny, R.W., 1994. Contrarian investment, extrapolation, and risk. Journal of Finance 49, 1541–1578.
Lewellen, J., 2004. Predicting returns with financial ratios. Journal of Financial Economics 74, 209–235.

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

Lintner, J., 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, T., 1993. Practical Neural Network Recipes in C++. Academic Press, Boston, MA.

Merton, R.C., 1973. An intertemporal capital asset pricing model. Econometrica 41, 867–887.

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

Miller, M., Modigliani, F., 1961. Dividend policy, growth, and the valuation of shares. Journal of Business 34, 411–433.

Modigliani, F., Cohn, R.A., 1979. Inflation, rational valuation, and the market. Financial Analysts Journal (March–April), pp. 24–44.

Nelson, C.R., Kim, M.J., 1993. Predictable stock returns: the role of small sample bias. Journal of Finance 48, 641–661.

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

Ritter, J.R., Warr, R., 2002. The decline of inflation and the bull market of 1982–1999. Journal of Financial and Quantitative Analysis 37, 29–61.

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

Rosenberg, B., Reid, K., Lanstein, R., 1985. Persuasive evidence of market inefficiency. Journal of Portfolio Management 11, 9–17.

Ross, S.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, W., 1964. Capital asset prices: a theory of market equilibrium under conditions of risk. Journal of Finance 19, 425–442.

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

Shiller, R.J., 2000. Irrational Exuberance. Princeton University Press, Princeton, NJ.

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

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

Torous, W., Valkanov, R., Yan, S., 2005. On predicting stock returns with nearly integrated explanatory variables. Journal of Business 77, 937–966.

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