The "Out-of-sample" Performance of Long-Run Risk Models

by
Wayne Ferson
Suresh Nallareddy
Biqin Xie

first draft: February 11, 2009
this draft: September 16, 2009

ABSTRACT

This paper studies the ability of long-run risk models, following Bansal and Yaron (2004) and others, to explain out-of-sample asset returns associated with the equity premium puzzle, size and book/market effects, momentum, reversals, and bond returns of different maturity and credit ratings. We examine stationary and nonstationary versions of the models using annual data for 1931-2006. Unlike the in-sample and calibration evidence of previous studies, we find that the models perform no better overall than the simple CAPM. For some effects the long-run risk models deliver smaller average pricing errors, but they have larger error variances than the CAPM, and the mean squared prediction errors are similar. One exception is the Momentum effect, where the long run risk models perform relatively well.

* Ferson is the Ivadelle and Theodore Johnson Chair of Banking and Finance and a Research Associate of the National Bureau of Economic Research, Marshall School of Business, University of Southern California, 3670 Trousdale Parkway Suite 308, Los Angeles, CA. 90089-0804, ph. (213) 740-5615, ferson@marshall.usc.edu, www-rcf.usc.edu/~ferson/. Nallareddy is a Ph.D student at the Marshall School of Business, Leventhal School of Accounting, nallared@usc.edu. Xie is a Ph.D student at the Marshall School of Business, Leventhal School of Accounting, biqinxie@usc.edu. We are grateful to participants in workshops at the University of North Carolina Charlotte, the University of Oregon and the University of Southern California for comments.

The "Out-of-sample" Performance of Long-Run Risk Models

ABSTRACT

This paper studies the ability of long-run risk models, following Bansal and Yaron (2004) and others, to explain out-of-sample asset returns associated with the equity premium puzzle, size and book/market effects, momentum, reversals, and bond returns of different maturity and credit ratings. We examine stationary and nonstationary versions of the models using annual data for 1931-2006. Unlike the in-sample and calibration evidence of previous studies, we find that the models perform no better overall than the simple CAPM. For some effects the long-run risk models deliver smaller average pricing errors, but they have larger error variances than the CAPM, and the mean squared prediction errors are similar. One exception is the Momentum effect, where the long run risk models perform relatively well.

1. Introduction

The long-run risk framework proposed by Bansal and Yaron (2004) has been a phenomenal success in asset pricing. A rapidly expanding literature finds the framework useful for explaining the equity premium puzzle, size and book/market effects, momentum, long-term return reversals in stock returns, risk premiums in bond markets, real exchange rate movements and more (see the review by Bansal, 2007). However, all of the evidence to date is based on calibration exercises or in-sample data fitting. This paper examines the out-of-sample performance of the long-run risk approach.

It is important to examine the ability of long run risk models to fit out-of-sample returns for several reasons. First, asset pricing models are almost always used in practice in an out-of-sample context. Firms want to estimate their costs of capital for future project selection. Risk managers want to model future risk exposures, and academic researchers want to make risk adjustments for as-yet-to-be-discovered empirical phenomena. Second, the evidence for other asset pricing models suggests that out-of-sample performance can be much worse than in-sample fit.[1]

Long-run-risk models feature a small but highly persistent component in consumption that is hard to measure directly in consumption data, but is nevertheless important in asset returns. The persistent component is either modelled as stationary (Bansal and Yaron, 2004) or as cointegrated (Bansal, Dittmar and Kiku, 2007). Either formulation raises potential concerns about the out-of-sample performance. Latent components in stock returns that are stationary but persistent have been shown to exacerbate problems with spurious regression and data snooping (Ferson, Sarkissian and Simin, 2003). Lettau and Ludvigson's (2001) *Cay* variable relies on the estimation of a single cointegrating coefficient, yet its out-of-sample performance in predicting returns is

[1] Early evidence for the Capital Asset Pricing Model (CAPM, Sharpe, 1964) includes Fouse, Jahnke and Rosenberg (1974) and Levy (1974). See Ghysels (1998) and Simin (2008) for evidence on more recent conditional asset pricing models.

degraded markedly when the parameter estimation is restricted to data outside of the forecast evaluation period (see Avramov (2002); also Brennan and Xia, 2005). An out-of-sample analysis is therefore important to understand the performance of long-run risk models.

This paper studies the "out-of-sample" performance of long-run risk models in explaining the equity premium puzzle, size and book/market effects, momentum, reversals, and bond returns of different maturity and credit quality. We examine stationary and nonstationary versions of the models using annual data for 1931-2006. (As explained below, we put "out-of-sample" in quotes to emphasize some qualifying remarks.) Unlike the in-sample and calibration evidence of previous studies, we find that the models perform no better overall than the simple CAPM, but there are some interesting particulars.

In terms of average pricing errors, both versions of the long-run risk model capture much of the equity premium and perform substantially better than the CAPM on the average momentum effect. Both versions of the long-run risk models horribly miss the Credit Premium. The average performance of the cointegrated version of the long run risk model is the better of the two.

The simple consumption beta model is the worst of the models we examine by almost any measure, although its relative performance improves for longer horizon returns and when we attempt to minimize the effects of time aggregation. The classical CAPM captures more than half of the size effect, but has trouble with the Value-Growth effect, and the Momentum effect looks larger after risk adjustment by the CAPM. These results are consistent with previous in-sample evidence in the literature.

In terms of mean absolute deviations (MAD) and mean squared pricing errors (MSE), the classical CAPM turns in the best performance, with the smallest MAD among five models for three to five of seven excess returns, depending on the experiment. In most of the cases where the long-run risk models deliver smaller average pricing errors than the CAPM they also produce larger error variances. These two components of the MSE roughly offset, leaving similar MSE. Statistical tests find few instances where any model performs significantly better than the estimation period sample mean return. The Momentum effect is the main exception, where the long-run risk models perform relatively well.

To evaluate the factors that drive the fit of the long-run risk models we examine models that suppress the consumption growth shocks. We find that these models do not perform as well, indicating that consumption shocks are important factors in the long-run risk models. We examine the robustness of our findings to the method used to estimate the expected risk premiums, to longer holding periods, time aggregation of consumption, to the measures of return, the sample period and to the use of rolling windows versus cross-validation methods.

The rest of the paper is organized as follows. Section 2 summarizes the models and Section 3 our empirical methods. Section 4 describes the data. Section 5 presents our empirical results and Section 6 concludes.

2. The Models

We study stationary and nonstationary versions of long-run risk models. As the long-run risk literature has expanded, papers have offered many different versions of these models. Our goal is not to test every model. Instead, we isolate and characterize the key features that appear in most of the models.

2.1 Stationary Models

Our stationary model specification follows Bansal and Yaron (2004):

$$\Delta c_t = x_{t-1} + \sigma_{t-1}\, \varepsilon_{ct} \qquad (1a)$$

$$x_t = \mu + \rho_x\, x_{t-1} + \varphi\, \sigma_{t-1}\, \varepsilon_{xt} \qquad (1b)$$

$$\sigma_t^2 = \underline{\sigma} + \rho_\sigma (\sigma_{t-1}^2 - \underline{\sigma}) + \varepsilon_{\sigma t}, \qquad (1c)$$

where c_t is the natural logarithm of aggregate consumption expenditures and x_{t-1} is the conditional mean of consumption growth. The conditional mean is the latent, long-run risk variable. It is assumed to be a stationary but persistent stochastic process, with ρ_x less than but close to 1.0. For example, in Bansal and Yaron (2004), $\rho_x = 0.98$; our overall estimate is 0.93. Consumption growth is conditionally heteroskedastic, with conditional volatility σ_{t-1}, given the information at time t-1. The shocks $\{\varepsilon_{ct}, \varepsilon_{xt}, \varepsilon_{\sigma t}\}$ are assumed to be homoskedastic and independent over time, although they may be correlated.

Bansal and Yaron (2004) and Bansal, Kiku and Yaron (2007) show that in this model, assuming Kreps-Porteus (1978) preferences, the innovations in the log of the stochastic discount factor are to good approximation linear in the three-vector of heteroskedastic shocks $u_t = [\sigma_{t-1}\varepsilon_{ct}, \varphi\sigma_{t-1}\varepsilon_{xt}, \varepsilon_{\sigma t}]$. The linear function has constant coefficients. Because the coefficients in the stochastic discount factor are constant, unconditional expected returns are approximately linear functions of the unconditional covariances of return with the heteroskedastic shocks.[2]

We focus on unconditional expected returns as a natural first step, and because many of the anomalies to which the long-run risk framework has been applied are cast in terms of unconditional moments (e.g. the equity premium, the average size and book/market effects). This leaves open the possibility that, like the CAPM in sample (e.g. Jagannathan and Wang (1996), Lettau and Ludvigson (2001), Bali and Engle, 2008),

---

[2] Writing the stochastic discount factor as $m_{t+1} = E_t(m_{t+1}) + u_{t+1}$, the conditional expected excess returns are approximately proportional to $Cov_t\{r_{t+1},m_{t+1}\} = E_t\{r_{t+1},u_{t+1}\}$, where $E_t(.)$ is the conditional expectation and $Cov_t\{.,.\}$ is the conditional covariance. Unconditional expected returns are therefore approximately proportional to $E[E_t\{r_{t+1},u_{t+1}\}]=Cov(r_{t+1},u_{t+1})$. Since u_{t+1} has constant coefficients on the shocks, the coefficients may be brought outside of the covariance operator.

the long-run risk approach works better in a conditional form. We leave that question for future research.

Some formulations of the stationary long-run risk model include an equation for dividends (e.g. Bansal, Dittmar and Lundblad, 2005), which we leave out of the system above because it is not needed for our empirical exercises. The nonstationary model, however, includes a central role for dividends.

2.2 Nonstationary Models

The nonstationary model follows Bansal, Dittmar and Kiku (2007) and Bansal, Gallant and Tauchen (2007), assuming that the natural logarithms of aggregate consumption and dividend levels are cointegrated:

$$d_t = \delta_0 + \delta_1 c_t + \sigma_{t-1} \varepsilon_{dt} \quad (2a)$$

$$\Delta c_t = a + \gamma' s_{t-1} + \varphi_c \sigma_{t-1} \varepsilon_{ct} \equiv x_{t-1} + \varphi_c \sigma_{t-1} \varepsilon_{ct} \quad (2b)$$

$$\sigma_t^2 = \underline{\sigma} + \rho_\sigma (\sigma_{t-1}^2 - \underline{\sigma}) + \varepsilon_{\sigma t}, \quad (2c)$$

where d_t is the natural logarithm of the aggregate dividend level and s_t is a vector of state variables at time t. Bansal, Dittmar and Kiku (2007) allow for time trends in the levels of dividends and consumption, but find that they don't have much effect, so we leave them out.[3]

Assets are priced in Bansal, Dittmar and Kiku (2007) from their consumption betas. These are modelled by representing asset returns (following Campbell and Shiller, 1988) as approximate linear functions of their dividend growth

[3] We examine a specification where we put in time trends and confirm little impact on the out-of-sample results. With seven test assets the largest difference in the average pricing error, with versus without the time trends, is 0.19% per year or about 3% of the average pricing error. The mean absolute deviations of the pricing errors are identical to two decimal places.

rates, dividend/price ratios and the first differences of the log dividend/price ratios. The consumption beta is then decomposed into three corresponding terms. We allow for the possibility that the volatility shock is a priced risk. Constantinides and Ghosh (2008) show that in this version of the model, the innovations in the log stochastic discount factor are approximately linear in the heteroskedastic shocks to consumption growth, the state variables s_t and the cointegrating residual, $\sigma_{t-1}\,\varepsilon_{dt}$. The coefficients in this linear relation are constant over time. Therefore, unconditional expected excess returns are approximately linear in the covariances of return with these four shocks.

3. Empirical Methods

For the stationary model we estimate the following equations:

$$u_{1t} = \Delta c_t - [a_0 + a_1\, rf_{t-1} + a_2 \ln(P/D)_{t-1}] \qquad (3a)$$

$$\equiv \Delta c_t - x_{t-1}$$

$$u_{2t} = u_{1t}^2 - [b_0 + b_1\, rf_{t-1} + b_2 \ln(P/D)_{t-1}] \qquad (3b)$$

$$\equiv u_{1t}^2 - \sigma_{t-1}^2$$

$$u_{3t} = x_t - d - \rho_x\, x_{t-1} \qquad (3c)$$

$$u_{4t} = \sigma_t^2 - f - \rho_\sigma\, \sigma_{t-1}^2 \qquad (3d)$$

$$u_{5t} = r_t - \mu - \beta(u_{1t},u_{3t},u_{4t}) \qquad (3e)$$

$$u_{6t} = \lambda - [\beta'V(r)^{-1}\beta]^{-1}\beta'V(r)^{-1}r_t, \qquad (3f)$$

where r_t is an N-vector of asset returns in excess of a proxy for the risk-free rate and V(r) is the covariance matrix of the excess returns. The system is estimated using the Generalized Method of Moments (GMM, see Hansen, 1982). The moment conditions are $E\{(u_{1t},u_{2t})\,(1, rf_{t-1}, \ln(P/D)_{t-1})\}=0$, $E\{u_{3t}\,(1, x_{t-1})\}=0$, $E\{u_{4t}\,(1, \sigma_{t-1}^2)\}=0$, $E\{u_{5t}$

$(1,u_{1t},u_{3t},u_{4t})\}=0$ and $E\{u_{6t}\}=0$. The system is exactly identified.[4]

The state vector in the model is the risk-free rate and aggregate price/dividend ratio: $s_t = \{rf_t, \ln(P/D)_t\}$. Equations (3a) and (3b) reflect the fact, as shown by Bansal, Kiku and Yaron (2007) and Constantinides and Ghosh (2008), that the conditional mean of consumption growth and the conditional variance can be identified as affine functions of these state variables. Unlike Constantinides and Ghosh, we do not drill down to the underlying structural parameters, but leave the coefficients in reduced form. Constantinides and Ghosh (2008) find that imposing the restrictions implied by the full structure of the model leads to rejections of the model in sample. The reduced forms are all we need to generate fitted expected returns and evaluate the out-of-sample performance.

From the system (3) we identify the priced heteroskedastic shocks that drive the stochastic discount factor. Comparing systems (1) and (3), we have:

$$u_{1t} = \sigma_{t-1}\, \varepsilon_{ct},\;\; u_{3t} = \varphi\, \sigma_{t-1}\, \varepsilon_{xt},\; u_{4t} = \varepsilon_{\sigma t}. \tag{4}$$

In Equation (3e), β is an N x 3 matrix of the betas of the N excess returns in r_t with the priced shocks, $\{u_{1t}, u_{3t}, u_{4t}\}$.

Equation (3f) identifies the three unconditional risk premiums, λ. These are the expected excess returns on the minimum variance, orthogonal, mimicking portfolios for the shocks to the state variables (see Huberman, Kandel and Stambaugh (1987) or Balduzzi and Robotti, 2007). We refer to these risk premium estimates as the GLS risk premiums. We also present results for simpler OLS risk premiums, where we

---

[4] This implies, inter alia, that the point estimates of the fitted expected returns are the same when the covariance matrix of the excess returns, V(r), is estimated, as we do, in a separate step or estimated simultaneously in the system. The standard errors of the coefficients would not be the same, but we don't use them in our analysis.

set V(r) equal to the identity matrix. The OLS risk premiums are not as efficient, asymptotically, as the GLS risk premiums. However, they might be more reliable in small samples. For a given evaluation period the fitted expected excess return is the estimate of βλ, based on the data for a nonoverlapping estimation period as described above.[5]

3.1 Nonstationary Model Estimation

The nonstationary models are estimated using the following system of equations:

$$u_{1t} = d_t - \delta_0 - \delta_1\ c_t \tag{5a}$$

$$u_{2t} = \Delta c_t - [a_0 + a_1\ rf_{t-1} + a_2\ \ln(P/D)_{t-1} + a_3\ u_{1t-1}] \tag{5b}$$

$$\equiv \Delta c_t - x_{t-1}$$

$$u_{3t} = rf_t - [g_0 + g_1\ rf_{t-1} + g_2\ \ln(P/D)_{t-1} + g_3\ u_{1t-1}] \tag{5c}$$

$$u_{4t} = \ln(P/D)_t - [h_0 + h_1\ rf_{t-1} + h_2\ \ln(P/D)_{t-1} + h_3\ u_{1t-1}] \tag{5d}$$

$$u_{5t} = r_t - \mu - \beta(u_{1t}, u_{2t}, u_{3t}, u_{4t}) \tag{5e}$$

$$u_{6t} = \lambda - [\beta' V(r)^{-1} \beta]^{-1} \beta' V(r)^{-1} r_t, \tag{5f}$$

In the nonstationary model there are four priced shocks, as shown by Constantinides and Ghosh (2008), as the cointegrating residual becomes a new state variable.[6] Identification

---

[5] Since we form the mimicking portfolios using only a small set of assets, they are not the maximum correlation portfolios relative to a broader asset universe. This means that the risk premium estimates might "overfit" the test asset returns. This overfitting could hurt the out of sample performance, as issue we plan to explore in a future draft.

[6] We can include a volatility equation in the system (5) as follows:

$$u_{7t} = u_{2t}^2 - [k_0 + k_1\ rf_{t-1} + k_2\ \ln(P/D)_{t-1} + k_3\ u_{1t-1}]$$

$$\equiv u_{2t}^2 - \sigma_{t-1}^2.$$

and estimation are similar to the stationary model.[7]

Both the stationary and nonstationary long-run risk models follow the literature in using a risk-free rate and log price/dividend ratio as state variables. This could be an important aspect of the ability of the models to explain returns. Ferson and Harvey (1999) find that regression coefficients on lagged conditioning variables, including a risk-free rate and dividend yield, are powerful cross-sectional predictors of stock returns. Petkova (2006) argues that innovations in lagged predictors, including a risk-free rate and dividend yield, subsume much of the cross-sectional explanatory power of the size and book/market factors of Fama and French (1993). Campbell (1996) uses the innovations in lagged predictor variables, including a dividend yield and interest rates, as risk factors in an asset pricing model.

We examine the attribution of returns to the priced state variables and conduct experiments to assess the importance of the state variables for the models' explanatory power. We examine versions of the models in which we suppress the consumption-related risk factors. If the risk-free rate and dividend yield are crucial for the ability of the model to fit returns, the models using only these two innovations should perform well. Given the smaller number of parameters to estimate, these "Two-State Variable" models might be expected to perform even better out of sample than the full models.

3.2 Evaluating Model Performance

To evaluate the fit of the models, some long-run risk studies focus on

---

Given (5b) the volatility equation is exactly identified. However, as Constantinides and Ghosh (2008) show, because the volatility is a linear function of the state variables including the cointegrating residual, the volatility shock is not a separately priced factor.

[7] The GMM estimator of the cointegrating parameter δ_1 is superconsistent and has a nonstandard limiting distribution, as shown by Stock (1987). This implies that the shock estimates used in (5b-5e) may be more precise than in a stationary model.

calibration exercises, while others estimate the model parameters in sample. Our focus is the out-of-sample performance. We estimate the reduced-form model parameters using data outside the evaluation period, and then assess the ability of the fitted models to predict returns over an evaluation period of one to three years. Because the long-run risk models may require long data samples to estimate the parameters accurately, we use the entire 1931-2006 period, excepting the forecast evaluation period, for parameter estimation. This approach is similar to the cross-validation method of Stone (1974). It could be called a "step around" analysis, in contrast to the more traditional "step-ahead" analysis.

A disadvantage of our "step-around" approach is that the forecasts could not actually be used in practice because they rely on future data that would not be known at the forecast date. We therefore also conduct a more traditional rolling window analysis, where the forecast evaluation period follows the rolling window estimation period.

3.3 Econometric Issues

Our evidence is subject to a data snooping bias, to the extent that the variables in the model have been identified in the previous literature through an unknown number of searches using essentially the same data. As the number of potential searches is unknown, this bias cannot be quantified. To the extent that such a bias is important, the long-run risk models would be expected to perform worse in actual practice than our evidence suggests. This explains the qualification in the introduction and the title. A "true" out of sample exercise would in our view use fresh data.

The estimation is subject to finite sample biases, as documented by Bansal, Kiku and Yaron (2007). The risk-free rate and dividend yield state variables are modelled as autoregressions and the variables are highly persistent. Following Kendal (1954), Stambaugh (1999) shows that the finite sample bias in regressions using such variables can be substantial. We address finite sample bias by applying a finite sample bias correction to the estimated coefficients on the persistent variables. The correction is developed by Amihud, Hurvich and Wang (2008).

Ferson, Sarkissian and Simin (2003) find that predictive regressions using persistent but stationary regressors are susceptible to spurious regression bias that can be magnified in the presence of data snooping. The spurious regression bias studied by Ferson, Sarkissian and Simin (2003) affects the standard errors and t-ratios of predictive regressions, but not the point estimates of the coefficients. Since our focus is the out-of-sample performance based on the point estimates, spurious regression bias should not affect our estimates of the models' forecasts. However, it could complicate the statistical evaluation of the forecast errors, an issue we address below.

3.4 Statistical Tests

We study the models' pricing errors, presenting their time-series averages, mean squared prediction errors and other summary statistics. To evaluate the statistical significance of the differences between the pricing errors for the various models, we use two approaches. The first approach follows Diebold and Mariano (1995). Let e_t be a vector of forecast errors for period-t returns, stacked up across K-1 models and let e_{0t} be the forecast error of the reference model (K=6 in our examples). Let $d_t = g(e_t) - g(e_{0t})\underline{1}$, where $\underline{1}$ is a K-1 vector of ones and g(.) is a loss function defined on the forecast errors. In our examples, g(.) will be the mean squared error or mean absolute error. Let $\underline{d} = T^{-1}\Sigma_t d_t$ be the sample mean of the difference over the evaluation period. Diebold and Mariano (1995) show that in the scalar case $\sqrt{T}\,(\underline{d} - E(d_t))$ converges in distribution to a normal random variable with mean zero and covariance, Σ, that may be consistently estimated by the usual HAC estimator: $T^{-1}\,\Sigma_t\,\Sigma_\tau\,w(\tau)\,(d_t - \underline{d})(d_{t-\tau} - \underline{d})'$. The appropriate weighting function w(τ) and number of lags depend on the context. We form t-ratios and Wald tests from these estimators.

Because the asymptotic distributions may be inaccurate we present bootstrapped p-values for the tests. Here we resample from the sample values of the mean-centered d_t vector with replacement and construct artificial samples with the same number of observations as the original. Constructing the test statistic on each of 10,000 artificial samples, the empirical p-value is the fraction of the artificial samples that produce a test statistic as large as the one we find in the actual data.

In the presence of a latent stationary but highly persistent variable, Ferson Sarkissian and Simin (2003) show that spurious regression bias produces inconsistent estimates of the standard errors, because the usual diagnostics fail to detect the weak but important autocorrelations in the models errors. Given a large enough sample, this can be addressed by letting the number of lags in the HAC estimator of the standard errors get very large. We conduct some experiments where we let the number of lags be large (in a future draft).

4. Data

We focus on annual data for several reasons. First, much of the evidence on long-run risk models is based on annual consumption data. Second and related, annual consumption data are less affected by problems with seasonality (Ferson and Harvey, 1992) and other measurement errors (Wilcox, 1992).

The annual consumption data are from Robert Shiller's web page and are discussed in Shiller (1989). Annual consumption data are time-aggregated, meaning that the reported figures are the averages of the more frequently-sampled levels. Time aggregation in consumption levels causes (1) a spurious moving average structure in the time-series of the measured consumption growth (Working, 1960); (2) biased estimates of consumption betas (Breeden, Gibbons and Litzenberger, 1989) and (3) biased estimates of the shocks in the long-run risk model (Bansal, Kiku and Yaron, 2007).

The moving average structure is addressed by the choice of weighting matrices in the statistical tests. The bias in consumption betas, as derived by Breeden, Gibbons and Litzenberger, is proportional across assets and therefore our risk premium estimates will be biased in the inverse proportion. For return attribution we examine the products of the betas and the risk premiums. The two effects should cancel out, leaving the predicted expected returns unaffected.

Bansal, Kiku and Yaron (2007) show that under time aggregation of a "true" monthly model, if the risk-free rate and dividend yield state variables are measured at the end of December, for example, regressions like (3a) and (3b) can still identify the true conditional mean, x_{t-1}, and conditional volatility, σ_{t-1}^2. However, the error terms do not reveal the consumption shocks from the true annual pricing kernel. We use quarterly consumption data, sampled annually, in experiments to assess the importance of this issue for our results.

The asset return data in this study are standard. We use the returns of common stocks sorted according to market capitalization and book/market ratios to study the size and value-growth effects. These are the extreme value-weighted decile portfolios provided on Ken French's web site. We also use the extreme value-weighted deciles for momentum winners and losers and for long-term reversal winners and losers, also from French. The market portfolio proxy is the CRSP value-weighted stock return index. The risk-free rate proxy is the one-month return on a three-month Treasury bill from CRSP. All of the returns are continuously-compounded annual returns.

We include bond returns that differ in maturity and credit quality. The short term bond return is the risk-free rate described above. The long-term Government bond splices the Ibbotson Associates 20 year US Government bond return series for 1927-1971, with the CRSP greater than 120 month US Government bond return after 1971. High grade Corporate bonds are the Lehman US AAA Credit Index after 1973, spliced with the Ibbotson Corporate Bond series prior to that date. High-yield bond returns splice the Blume, Keim and Patel (1991) low grade bond index returns for 1927 to

1990, with the Merrill Lynch High Yield US Master Index returns after that date.

Table 1 presents summary statistics of the basic data. We focus on the seven excess returns summarized in the table, which reflect the equity premium, the firm size effect (Small-Big), the book/market effect (Value-Growth), the excess returns of momentum winners over losers (Win-Lose), the excess returns of long-term reversal losers over winners (Reversals), the excess returns of low over high-grade corporate bonds (Credit Premium) and the excess returns of long-term over short-term Treasury bonds (Term Premium).

5. Empirical Results

Table 1 also presents the average, step-around forecast errors generated by the models. Panel A uses the OLS risk premiums and Panel B uses the GLS risk premiums. We compare the performance of the long-run risk models with three alternative models in Table 1. The first is the market-portfolio based CAPM of Sharpe (1964). The second is a simple consumption-beta model following Rubinstein (1976), Breeden, Gibbons and Litzenberger (1989), Parker and Julliard (2005) and Moskowitz and Vissing-Jorgensen (2008). The third model is a "2-State Variable" model in which the consumption-related shocks of the long run risk models are turned off. The two state variables are the shocks to the risk-free rate and log price/dividend ratio.

We also fit the long-run risk models using the whole sample and find that this produces different results than our step-around analysis. The average pricing errors are smaller for about half of the assets and the differences can be substantial. For example, the nonstationary version of the model turns in smaller average pricing errors than in Table 1 for four of the seven excess returns, and the greatest difference is 2.53%. Averaged across the seven assets the stationary model's step-around pricing errors are 1.2% under OLS and 4.8% under GLS. Using the full sample the figures are 0.7% and 4.3%, respectively. Consistent with previous studies that evaluate the models on the basis of in-sample fit, the long run risk models appear to fit the average returns well, fitting better than the CAPM for about half of the assets.[8]

Table 1 Panel A shows that in terms of average step-around pricing errors, the simple consumption beta model is the worst. It does capture some of the term premium (1.7% unadjusted versus an average pricing error of 1.3%) and it reduces the momentum effect from 15% to an average pricing error of 11.9%, but otherwise it performs worse than no risk adjustment at all.

The classical CAPM fits the average equity premium closely. The fit would be perfect except for the fact that the market return mean for the estimation period fluctuates slightly around the overall sample mean. The CAPM captures about 2/3 of the size effect, but has trouble with the Value-Growth effect (unadjusted effect 5.5%, average pricing error 3.5%). This makes sense, given that the book/market effect was identified as a CAPM anomaly. The momentum effect looks larger after risk adjustment by the CAPM, consistent with many earlier studies. The CAPM captures a part of the Term premium, and slightly overshoots the Credit Premium (average pricing error -0.4%).

Both versions of the long-run risk model capture much of the equity premium (7.2% unadjusted, with average pricing errors of 2.2% and 1.8%). Both models perform much better than the CAPM on the average Momentum Effect and Term Premium, and the cointegrated model captures most of the average Reversal Effect (5.6% unadjusted, average pricing error 1.0%). Both of the long-run risk models horribly miss the Credit Premium.

Panel B of Table 1 presents the average pricing errors using the GLS risk premiums. Compared with the OLS case the performance of the consumption beta model

---

[8] Mean Absolute Deviations (MAD) of the pricing errors are reported in Table 2. Using the full sample the long run risk models generate smaller MADs than in Table 2 for five or six of the seven asset returns and beat the CAPM for four to six of the assets, depending on the experiment.

deteriorates, except when explaining momentum and the term premium. The performance of the long-run risk models also generally deteriorates. However, under GLS the models perform better on the credit premium than they do under OLS. This is because the mimicking portfolios under GLS place more weight on the low variance asset returns. Thus, the models do a better job of explaining the low variance assets under GLS.

The CAPM is unaffected by the choice of GLS versus OLS, because the sample mean of the market excess return is the efficient estimator for the market risk premium (e.g. Shanken, 1992). Under GLS the long-run risk models continue to outperform the CAPM in capturing the momentum effect and the term premium.

Overall, the CAPM produces the smallest average pricing errors of all the models for two of the seven returns under OLS, and the nonstationary long run risk model performs best for five of the seven returns. Under GLS the CAPM wins for five of the seven returns and the nonstationary model wins for two of the returns. The average performance of the cointegrated version of the long-run risk model is better of the two.

The right hand column of Table 1 presents the average pricing errors of the 2-State model. The results suggest that the consumption-based risk factors are important for the models' explanatory power. In no case does the model with only the risk-free rate and dividend yield innovations deliver the smallest average pricing error. There is only one case (the credit premium under OLS) where the 2-State model produces a smaller average pricing error than the long-run risk models. These results show that the performance of the long-run risk models does depend on the use of consumption data.[9]

[9] We explore this issue further with a return attribution analysis over the whole sample period. The fitted expected return for each asset is broken down into (beta times risk premium) components for each risk factor. We find that the consumption growth shocks are the least important overall, but they are not negligible and the dividend yield and interest rate shocks are not uniformly dominant. The cointegrating residuals in the nonstationary long-run risk models also contribute substantially to the fitted returns.

Table 1 uses continuously-compounded annual excess returns. We also run the analysis with simple arithmetic excess returns. None of the above conclusions change, except that the relative performance of the nonstationary long-run risk model improves. It now delivers the smallest average pricing error for five of the seven assets.

5.1 MAD and Mean Square Errors

We are interested in the volatility of the forecast errors as well as their average values. These are summarized in Table 2. We present the mean absolute deviations (MAD) and mean squared forecast errors (MSE). For comparison, the left-hand column reports the MAD and MSE of a Mean Model, where the estimation period sample mean return is the forecast. Once again, the simple consumption beta model performs poorly, producing the largest MAD and MSE's in most of the cases.

In terms of mean absolute deviations, the classical CAPM turns in the best performance, with the smallest MAD for five of the seven excess returns. The historical mean and long run risk models are the best predictors of the momentum effect, and all of the models deliver similar numbers on the Term Premium. The MSEs provide similar impressions. Table A.2 in the appendix presents the results using GLS risk premiums and the impressions are again similar. The classical CAPM delivers the smallest MAD errors for four of the seven asset returns.[10]

The MSEs in Table 2 may be decomposed as the mean pricing error or bias squared plus the error variance. Comparing tables 1 and 2 reveals more about the structure of the pricing errors. The error variance is the dominant component in almost every instance. In most cases where the long-run risk models deliver smaller average pricing errors than the CAPM (Reversals, Term Premium) they also have larger error

---

[10] Using arithmetic returns the results are similar, except that the nonstationary long run risk model's performance is improved. It now has the smallest MAD pricing error for two of the seven returns and appears more similar to the CAPM than in Table 2.

variances, so the MSEs are similar. Momentum is the exception, where the long-run risk models handily trump the CAPM by any measure, but no model beats the Mean Model. When the CAPM confronts the momentum effect the mean error and the error variance contribute nearly equal shares to the MSE. The choice of GLS versus OLS risk premiums has a small effect on the decompositions of the MSEs. Both the average pricing errors and MSEs of the long run risk model increase under GLS.

5.2 Statistical Significance

Table 3 summarizes the Diebold-Mariano (1995) tests for the differences between the pricing errors of the various models. The benchmark model is the estimation period sample mean. Panel A presents tests based on the MADs and Panel B is based on MSEs. Empirical p-values are presented for the two one-sided tests of the null hypothesis that the model in question performs no differently than the Mean Model. These are based on resampling the mean-centered statistics over 10,000 simulation trials, each with the same number of observations as in the sample.

Similar to the findings of Simin (2008), the sample mean model is hard to beat. The CAPM outperforms the mean in predicting the size effect, but otherwise no model beats the mean for the size effect, value-growth, reversals or the term premium.[11] There are a half dozen cases, out of 35 comparisons, where the p-values are less than 5% and the statistics are negative, indicating that the mean model delivers smaller MSEs and MADs.

We also compute Diebold-Mariano tests where the CAPM is the null model. These tests show that the long-run risk models significantly outperform the CAPM on momentum. However, they are significantly worse than the CAPM in

[11] The CAPM is not examined for the equity premium because its forecast of the equity premium is the estimation period sample mean excess return.

explaining the firm size effect and the credit premium.

Clark and West (2007) show that, on the assumption that a given model is correct, a model with more parameters is expected to have larger mean squared prediction errors out of sample, because the noise in estimating the parameters that should really be set to zero will contribute to the error variance. They propose an adjustment to the difference between two models' mean squared prediction errors to account for this affect. Let E_{1t} be the forecasted value from model one which is the more parsimonious model, SSE_1 is the mean squared forecast error, E_{2t} is the forecast from model 2 and SSE_2 is its mean squared forecast error. The proposed test is $SSE_1 - [SSE_2 - \Sigma(E_{1t} - E_{2t})^2/T]$. The third term adjusts for the difference in expected forecast error and a one-sided test is conducted. The null hypothesis is that the two models have the same forecast error if the true parameters were known and the alternative is that the more complex model has smaller forecast error given the true parameters. If we suppress the adjustment term, we obtain the Diebold-Mariano test as a special case.[12]

Table 4 presents the Clark-West tests using OLS in panels A and C and GLS in panels B and D. In Panels A and B the null model is the sample mean. None of the test statistics is significant at conventional levels. Thus, the inability of the long-run risk models, the consumption beta model or the classical CAPM to outperform the sample mean is not an artifact of the greater numbers of parameters in those models.

Panels C and D of Table 4 show that when the 2-State model is the null model, it is significantly worse than the more complex long-run risk models for Momentum, the Value-Growth effect, and under GLS also for the size effect and reversals. This evidence further reinforces the conclusion that the fit of the long-run risk

[12] Following Clark and West (2007), we implement the tests with the adjustment terms as follows. Let: $F_t = (r_t - E_{1t})^2 - [(r_t - E_{2t})^2 - (E_{1t} - E_{2t})^2]$, where r_t is the return being forecast. We construct a t-statistic for the null hypothesis that $E(F_t)=0$ and conduct a one-sided test using bootstrap simulation as described above.

models does depend on the use of the consumption data. It also shows that the test statistics have some power.

5.3 Longer-horizon Returns

Previous studies find that consumption-based asset pricing models do a better job fitting longer-horizon returns (e.g. Daniel and Marshall (1997), Parker and Julliard (2005), Malloy, Moskowitz and Vissing-Jorgensen, 2008). We examine two-year and three-year returns. Table 5 summarizes the average pricing errors for three-year returns, obtained by lengthening the "step around" evaluation period to three years and compounding the returns, resulting in 25 nonoverlapping three year returns. This maintains the previous assumption that the decision interval in the model is one year. The overall conclusions are similar to the previous tables. The CAPM delivers the smallest average pricing error for three or four of the seven assets and the long-run risk models win for one or two each, including the Momentum effect. The 2-State model performs relatively poorly. Two year returns produce similar results.

Table 6 presents the MAD and MSE for the pricing errors under OLS risk premiums. The CAPM performs best for three or four of the seven returns and the stationary long-run risk model wins for one of the seven. (Using GLS risk premiums, not tabulated, the CAPM wins for four to five out of seven and the nonstationary long-run risk model in one case.) The nonstationary version generally outperforms the stationary long-run risk model. The consumption beta model does perform better, in relative terms, on the longer-horizon returns. It delivers smaller MADs than the 2-State model for five of the seven returns and its performance lags that of the other models by smaller margins on the longer-horizon returns.

5.4 The Post-war Period

The "step around" results assume that the reduced-form parameters of the model are essentially constant for the full 1931-2006 period. However, the literature provides evidence that they may not be constant. For example Nelson and Kim (1993), Pastor and Stambaugh (2001) and Lettau and Van Niewerburgh (2007) find structural breaks in stock returns and dividend yields. Constantinides and Ghosh (2008) find that long-run risk models perform better when confined to the post World War II sample period. We therefore examine the performance of the models, restricted to a sample beginning in 1947.

The average pricing errors are recorded in Table A.1 in the appendix. The long-run risk models generally deliver smaller average pricing errors in the post war period than for the full period, consistent with Constantinides and Ghosh (2008). This suggests that the parameters of the model are likely not constant over the longer sample period. The cointegrated version of the model is especially improved. For example, even though the momentum effect is larger after 1947, the cointegrated model captures it well (average pricing error only 0.24%). The performance of the 2-State model deteriorates dramatically, confirming the previous observation that the long-run risk models' fit is not solely driven by the risk-free rate and dividend yield innovations.

During the post-war period the cointegrated long run risk model delivers a smaller average pricing error for the equity premium (0.84%), but the stationary model has a larger pricing error than it does over the full sample. The average size effect is smaller by half in the post-war period but the stationary model has an average pricing error of 4.12%, while the CAPM produces 0.49%. Overall, in terms of average pricing errors, the CAPM performs best for three of the seven assets, the nonstationary model wins for three assets and the stationary model in one case.

We examine (but do not tabulate here) the MAD forecast errors and MSEs during the post-war period. The performance of the long run risk models improves slightly by these measures, relative to the full sample period. Still, the classical CAPM has the smallest MAD error for more of the returns. The stationary long run risk model wins for Value and Momentum, while the cointegrated model wins for Reversals. Statistical tests find few instances where any model outperforms the Mean Model in terms of MAD or MSE.

5.5 Time Aggregation

The consumption data are time aggregated, which as described above leads to a spurious moving average component in consumption growth, biased consumption betas and biased shocks in the consumption-related risk factors. While the available data do not allow us to fully address the issue we can obtain some information on the impact of time aggregation by using higher frequency data, sampled at the end of each year. Jagannathan and Wang (2007) find that consumption-based models work better on data sampled from the last quarter of each year. Monthly US consumption data are available from 1959 in seasonally adjusted form and from 1946 in both seasonally adjusted and unadjusted form. We use the quarterly, not seasonally adjusted data for 1946-2004. The longer historical sample for the quarterly data allows us to compare the results with those in the previous section. Sampling the last quarter of each year we avoid the strong seasonality in consumption expenditures and the smoothing biases in data seasonally adjusted by the Commerce Department's X-11 seasonal adjustment program (see Ferson and Harvey (1992) and Wilcox, 1992). The Appendix tables A.3 and A.4 present the results.

In terms of average pricing errors, the performance of the consumption beta model is improved by the use of the higher-frequency consumption data. For example, comparing Table A.3 with Table A.1, we find smaller average pricing errors for five of seven assets when using the quarterly consumption data. However, no model outperforms the classical CAPM in terms of average pricing errors. The MADs and MSEs reported in Table A.4 show that the consumption beta model actually wins for explaining the value premium. The MAD performance splits fairly evenly across the models in this experiment, and no model is clearly dominant. Momentum is again the exception, where the Mean model and the nonstationary long-run risk model tie for first place.

5.6 Rolling Windows: 1976-2006

One advantage of our "step-around" analysis is that it uses as much of the time series data as possible. A disadvantage is that the forecasts could not actually be used, as they rely on data after the forecast evaluation period. We repeat the analysis using a more traditional rolling estimation and step-ahead forecast scheme. The rolling estimation period is 45 years and the forecast period is the subsequent year.[13]

Table 7 presents the average returns and pricing errors from the rolling step-ahead analysis. Panel A uses the OLS risk premiums and Panel B uses GLS. Over the 1976-2006 period the average Momentum effect, Book-to-market effect and Term premium are all larger than over the sample going back to 1931. The consumption beta model performs worst in the rolling estimation, and is worse than no risk adjustment at all in about 2/3 of the cases. The 2-State model also performs poorly, with the largest average pricing errors in about 1/3 of the cases. The classical CAPM performs worse on momentum and the value premium than it did in the full sample, but it delivers the smallest average pricing error for three of the seven assets. The long run risk models perform worse on the equity premium and size effects than they do in the full sample; but each version of the long run risk model delivers the smallest average pricing error for two of the seven assets. Panel B presents the GLS results and the overall impressions are similar, except that the long-run risk models perform markedly worse under GLS on the equity premium and size effects.

[13] We tried 30 and 40 year estimation periods, but encountered singularities in the gradients of the GMM criterion in these cases.

Table 8 summarizes the MAD returns and pricing errors for the step-ahead analysis. Panel A uses OLS risk premiums, and once again the consumption beta model fairs poorly. Its MAD forecast error is no better than using a forecasted return of zero for most of the assets, whether based on OLS or GLS risk premiums. The performance of the classical CAPM is similar to its performance over the full sample period, and it delivers the smallest MAD forecast error for two of the seven assets (under OLS or GLS). The nonstationary long-run risk model delivers the smallest MAD error for one return. Under GLS risk premiums in Panel B, the long run risk models perform much worse on the equity premium and value-growth effects than under OLS, while their improvement under GLS on the credit and term premiums is small.

We test for significant differences using the Diebold-Mariano tests and find that few of the models are significantly different from the forecast provided by the rolling sample mean. Overall, the impressions from the rolling, step ahead forecast analyses are similar to those from the full sample, step-around approach.

5.7 The Effects of Bias Correction

In this section we examine versions of the models in which we apply the bias correction factors derived by Amihud, Hurvich and Wang (2008) to the coefficients on the highly persistent variables in the models. The Appendix presents the details of the correction and Table 9 (to appear) summarizes the results.

6. Conclusions

This study examines the ability of long run risk models to fit out-of-sample returns. The question of out-of-sample fit is important because asset pricing models are almost always used in practice in an out-of-sample context and the evidence for other asset pricing models shows that out-of-sample performance can be much worse than in-sample fit. Our results are different in several respects from the conclusions of the previous literature that examines the long-run risk models using calibration and in-sample fit.

We examine stationary and nonstationary versions of long-run risk models using annual data for 1931-2006. Overall, the models perform no better than the simple CAPM, but there are some interesting particulars. The long run risk models perform especially well in capturing the Momentum Effect. In most of the cases where the long-run risk models deliver smaller average pricing errors than the CAPM they also generate larger error variances. These two components offset, leaving similar mean squared pricing errors. Statistical tests find few instances where the MADs or MSEs are significantly better than using the historical mean as the forecast. Our main results are robust over one to three year holding periods, quarterly or annual consumption data, continuous or discrete returns and rolling estimation or cross-validation methods.

The average performance of the cointegrated version of the long run risk model is the better of the two versions. When we restrict to data after 1947, we find that the performance of the long-run risk models improves, consistent with the in-sample tests of Constantinides and Ghosh (2008), and the average performance of the cointegrated version of the model is especially improved. Even in these cases the long-run risk models fail to outperform the simple CAPM, except on Momentum.

Our results are generally not very optimistic about the ability of long-run risk models to provide practically useful forecasts of expected returns, or even better forecasts than the standard CAPM. But they do suggest areas for potential refinements. The subperiod results suggest that a marriage of structural breaks and long-run risk models, as in recent work by Lettau, Ludvigson and Wachter (2008) and Boguth and Keuhn (2009) could prove fruitful. Consumption growth shocks by themselves are not empirically important risk factors, but models that suppress the consumption shocks perform markedly worse, so the macroeconomic data do matter. Sampling not seasonally adjusted consumption data in the last quarter of the year improves the fit of the consumption beta model, consistent with Jagannathan and Wang (2007). The cointegrating residual shocks of the nonstationary long-run risk model contribute substantially to its ability to fit returns. The parameters that determine the cointegrating residual shocks are superconsistent, which theoretically improves the precision with which the shocks can be estimated.[14] Cointegrated versions of the long-run risk models have received less attention in the literature than stationary versions; our results suggest they deserve more attention.

Appendix: Finite Sample Bias Correction

This appendix discusses corrections for finite sample biases due to lagged persistent regressors. Our corrections are a modification of the approach suggested by Amihud, Hurvich and Wang (AHW 2009). They consider a time-series regression of a stock return, y_t on a vector of lagged predictors:

$$y_t = \alpha + \beta' x_{t-1} + u_t, \qquad (A.1)$$

$$x_t = \theta + \Phi x_{t-1} + v_t, \qquad (A.2)$$

$$u_t = \varphi' v_t + e_t, \qquad (A.3)$$

where e_t is independent of v_t and x_{t-1} and the covariance matrix of v_t is Σ_v. They obtain reduced-bias estimates of β through an iterative procedure, exploiting the approximate bias in the OLS estimator of Φ given by Nichols and Pool (1988):

$$E(\Phi^{\wedge}-\Phi) = -(1/T)\Sigma_v \left[(I-\Phi')^{-1} + \Phi'(I-\Phi'\Phi')^{-1} + \Sigma_k \lambda_k(I-\lambda_k\Phi')^{-1}\right]\Sigma_x^{-1} + O(T^{-3/2}), \qquad (A.4)$$

where λ_k is the k-th eigenvalue of Φ' and Σ_x is the covariance matrix of x_t. AHW obtain

---

[14] Empirically, time-series plots of the estimated rolling cointegration parameter appear smoother than estimates of the autoregression coefficient for expected consumption growth.

the reduced-biased estimator of β by iterating between equations (A.4) and (A.2) to obtain a corrected estimator, Φ_c, which is used to generate corrected residuals from (A.2), v_{ct}. With these residuals the following regression delivers the reduced-bias estimate of β:

$$y_t = a + \beta' x_{t-1} + \varphi' v_{ct} + e_t. \qquad (A.5)$$

We modify the AHW procedure as follows. Our goal is to obtain reduced-biased estimates of the innovations ε_t in the projection, $\beta' x_t$, as in the following regression:

$$\beta' x_t = A + B\, \beta' x_{t-1} + \varepsilon_t. \qquad (A.6)$$

For example, our y_t in the stationary long run risk model is a vector consisting of consumption growth and the squared consumption growth innovations. These are both regressed on the same lagged state variables x_{t-1}: the risk free rate and dividend yield. The innovations in the projections are priced risk factors in the model. Since in our case (A.1) is a seemingly unrelated regression, y_t may be a vector of variables with no loss of generality. Our equations (3c) and (3d) regress the projections of the two y_t variables on their lagged values to obtain the shocks, as represented in Equation (A.6). We modify the AHW procedure for Equation (A.6).

In our applications β and B are full rank square matrices. Multiplying the vector x_t by the invertible matrix β,' the autoregressive coefficient $B = \beta' \Phi (\beta')^{-1}$. We also modify Equation (A.5) as follows:

$$y_t = a + \beta' x_{t-1} + \varphi' \varepsilon_{ct} + e_t, \qquad (A.7)$$

where we have replaced v_{ct} with the corrected error from (A.6), ε_{ct}.

For the stationary model we apply the AHW method by starting with Equation (A.1) to get an initial estimate of β and then iterating between equations (A.6) and (A.7). Within this scheme, at each (A.6) step we iterate between (A.6), taking β as fixed, and (A.4) replacing Φ with B and Σ_v with Σ_ε. This produces the reduced bias Φ, ε_{ct}, and β. Applying the reduced bias β to Equation (3a) gives the priced shock, u_{1t}. The shocks u_{3t} and u_{4t} are obtained directly from (A.6).

For the nonstationary model we apply the AHW method on the three-vector consisting of the lagged risk-free rate, dividend yield and the lagged error correction shock from (5a). Since the slope in (5a) is superconsistent, we use its OLS estimate to obtain u_{1t}. We obtain the reduced bias Φ as before, which delivers the shocks u_{3t} and u_{4t} as a subset of the autoregressive system (A.2), which is (5c) and (5d). The reduced bias estimate of β from (A.5) amounts to adding v_{ct} to the regression (5b), which then delivers the shock u_{2t}.

References

Amihud, Yakov, and Clifford M. Hurvich, 2004, Predictive Regressions: A Reduced-Bias Estimation Method, *Journal of Financial and Quantitative Analysis* 39, 2004, 813-841.

Amihud, Yakov, Clifford M. Hurvich and Yi Wang, 2009, Multiple-predictor regressions: Hypothesis testing, *Review of Financial Studies* 22 (1), 413-434.

Avramov, Doron, 2002, Stock Return Predictability and Model Uncertainty, *Journal of Financial Economics* 64, 423-458.

Bali, Turan, and R.F. Engle, 2008, Investigating ICAPM with dynamic conditional correlations, working paper, Baruch and New York Universities.

Bansal, Ravi, 2007, Long Run Risks and Financial Markets, *The Review of the St. Louis Federal Reserve Bank*, summer.

Bansal, Ravi, Robert Dittmar and Dana Kiku, 2009, Cointegration and consumption risks in asset returns, *Review of Financial Studies* 22 (3), 1343-1375.

Bansal, Ravi, R. Gallant and G. Tauchen, 2007, Rational Pessimism, Rational Exuberance and Asset Pricing Models, *Review of Financial Studies* 74 (4), 1005-1033.

Bansal, Ravi, Dana Kiku and Amir Yaron, 2007, Risks for the long run: Estimation and inference, working paper, Duke University.

Bansal, Ravi, and Amir Yaron, 2004, Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles. *Journal of Finance* 59, 1481-1509.

Blume, Marshall E., Donald B. Keim and Sandeep A. Patel, 1991, Returns and Volatility of Low-Grade Bonds 1977-1989, *Journal of Finance* 46, 49-74.

Boguth, Oliver, and Lars-Alexander Kuehn, 2009, Consumption volatility risk, working paper, Carnegie Mellon University.

Breeden, Douglas T., Michael R. Gibbons and Robert H. Litzenberger, 1989, Empirical tests of the consumption CAPM, *Journal of Finance* 44, 231-262.

Brennan, Michael J., and Yihong Xia, 2005, tay's as good as cay, *Finance Research Letters* 2 (1), 1-14.

Campbell, John Y., 1996, Understanding risk and return, *Journal of Political Economy* 104, 298-345.

Campbell, John Y., and Robert J. Shiller, 1988, Stock Prices, Earnings, and Expected Dividends, *Journal of Finance* 43, 661-676.

Chen, Long, Ralitsa Petkova and Lu Zhang, 2008, The Expected Value Premium, *Journal of Financial Economics* 87, 269-280.

Clark, Todd E., and Kenneth D. West, 2007, Approximately normal tests for equal predictive accuracy in nested models, *Journal of Econometrics* 138, 291-311.

Constantinides, George M., and Anisha Ghosh, 2008, Asset pricing tests with long-run risks in consumption growth, working paper, University of Chicago.

Daniel, Kent, and David Marshall, 1997, The equity premiums puzzle and the risk-free rate puzzle at long horizons, *Macroeconomic Dynamics* 1(2), 452-484.

Diebold, Francis X., and Roberto S. Mariano, 1995, Comparing predictive Accuracy, *Journal of Business and Economic Statistics* 13,253-265.

Epstein, Lawrence, and Stanley Zin, 1989, Substitution, risk aversion and the temporal behavior of consumption and asset returns: A theoretical framework. *Econometrica* 57, 937-969.

Ferson, Wayne, and Campbell Harvey, 1992, Seasonality and Consumption-Based Asset Pricing, *Journal of Finance* 47, 511-552.

Ferson, Wayne, and Campbell Harvey, 1999, Conditioning Variables and Cross-section of Stock Returns, *Journal of Finance* 54, 1325-1360.

Ferson, Wayne, Timothy Simin and Sergei Sarkissian, 2003, Spurious Regressions in Financial Economics? *Journal of Finance* 58, 1393-1414.

Fouse, William L., William W. Jahnke and Barr Rosenberg, 1974, Is Beta Phlogiston? *Financial Analysts Journal* 30, no. 1 pp. 70-80.

Ghysels, Eric, 1998, On Stable Factor Structures in the Pricing of Risk: Do Time-Varying Betas Help or Hurt? *Journal of Finance* 53, 549-573.

Jagannathan, Ravi, and Yong Wang, 2007, Lazy Investors, discretionary consumption and the cross-section of equity return, *Journal of Finance*, (August).

Jagannathan, Ravi, and Zhenyu Wang, 1996, The conditional CAPM and the cross-section of expected returns, *Journal of Finance* 51, 3-53.

Hansen, Lars Peter, 1982, Large Sample Properties of Generalized Method of Moments Estimators, *Econometrica* 50, 1029-1054.

Kendal, M. G. 1954, Note on bias in the estimation of autocorrelation. *Biometrika* 41, 403-404.

Kreps, David, and Evan Porteus, 1978, Temporal resolution of uncertainty and dynamic choice theory, *Econometrica* 46, 185-200.

Lettau, Martin, Sydney Ludvigson and Jessica Wachter, 2008, The declining equity premium: What role does macroeconomic risk play? *Review of Financial Studies* 21, 1653-1687.

Lettau, Martin, and Syndey Ludvigson, 2001a, Consumption, aggregate wealth and expected stock returns, *The Journal of Finance* 56, 815-849.

Lettau, Martin, and Syndey Ludvigson, 2001b, Resurrecting the (C)CAPM: A cross-sectional test when risk premia are time-varying, *Journal of Political Economy* 109, 1238-1287.

Lettau, Martin, and Stijn Van Nieuwerburgh, 2008, Reconciling the Return Predictability Evidence, *Review of Financial Studies* 21, 1607-1652.

Levy, Robert A., 1974, Beta coefficients as predictors of return, *Financial Analysts Journal* 30 (1), 61-69.

Malloy, Christopher J., Tobias J. Moskowitz, and Annette Vissing-Jorgensen, 2008. Long-Run Stockholder Consumption Risk and Asset Returns, Northwestern University, Working Paper.

Nelson, Charles R., and Myung J. Kim, 1993, Predictable Stock Returns: The Role of Small Sample Bias, *Journal of Finance* 48, 641-661.

Newey, Whitney K., and Kenneth D. West, 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703-708.

Parker, Jonathan A., and Christian Julliard, 2005, Consumption Risk and the Cross Section of Expected Returns, *Journal of Political Economy* 113, 185-222.

Pastor, Lubos, and Robert F. Stambaugh, 2001, The Equity Premium and Structural Breaks, *Journal of Finance* 56, 1207-1239.

Petkova, Ralista, 2006, Do the Fama-French factors proxy for innovations in predictive variables? *Journal of Finance* 61, 581-612.

Rubinstein, Mark, 1976, The valuation of uncertain income streams and the pricing of options, *Bell Journal of Economics and Management Science* 7, 407-425.

Shiller, Robert, 1989, *Market Volatility*.

Simin, Timothy, 2008, The (Poor) Predictive Performance of Asset Pricing Models, *Journal of Financial and Quantitative Analysis* 43, 355-380.

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

Stock, James H., 1987, Asymptotic Properties of least squares estimators of cointegrating vectors, *Econometrica* 55, 1035-1056.

Stone, M., 1974, Cross-validatory choice and assessment of statistical predictions, *Journal of the Royal Statistical Society*, Series B 36, 111-147.

Wilcox, David W., 1992, The Construction of U.S. Consumption Data: Some Facts and Their Implications for Empirical Work, *American Economic Review* 82, 922-941.

Working, Holbrook, 1960, Note on the Correlation of First Differences of Averages in a Random Chain, *Econometrica* 28, 916-918.

Table 1

Mean excess returns and pricing errors (actual minus forecast) during model evaluation years, 1932-2006. The number of observations is 75. The continuously-compounded returns are in annual percent. CCAPM is the simple consumption beta model. CAPM is the Capital Asset Pricing Model. LRRN is the Nonstationary Long Run Risk Model and LRRS is the Stationary Long Run Risk Model. 2-STATE is the model with only shocks to the two state variables (risk free rate and log price/dividend ratio).

Premium:	Mean Excess	Model Pricing Errors: CCAPM	CAPM	LRRN	LRRS	2-STATE
Panel A: OLS Risk Premiums						
Equity Premium	7.18	8.60	0.89	2.14	1.75	2.55
Small-Big	5.10	6.23	1.97	0.00	-1.07	3.06
Value-Growth	5.48	6.79	3.54	3.51	4.74	7.74
Win-Lose	15.04	11.85	17.19	2.15	4.44	14.36
Reversal	5.55	7.53	4.92	1.00	4.15	4.71
Credit Premium	1.87	2.10	-0.40	-5.84	-6.08	1.63
Term Premium	1.74	1.30	1.37	0.32	0.36	-1.94
Panel B: GLS Risk Premiums						
Equity Premium	7.18	10.54	0.89	3.62	2.67	3.46
Small-Big	5.10	7.76	1.97	4.95	7.26	9.82
Value-Growth	5.48	8.59	3.54	8.65	8.85	10.16
Win-Lose	15.04	7.48	17.19	5.20	6.67	10.22
Reversal	5.55	10.23	4.92	6.74	9.41	10.75
Credit Premium	1.87	2.43	-0.40	-1.27	-0.85	1.77
Term Premium	1.74	0.69	1.37	0.17	-0.43	-0.43

Table 2

Mean absolute excess returns (MAD), mean squared excess returns (MSE), MAD and MSE pricing errors during model evaluation years, 1932-2006. The number of observations is 75. The returns are continuously-compounded annual returns and the statistics are stated as annual decimal fractions. CCAPM is the simple consumption beta model. CAPM is the Capital Asset Pricing Model. LRRN is the Nonstationary Long Run Risk Model and LRRS is the Stationary Long Run Risk Model. These results are based on OLS risk premium estimates. 2-STATE is the model with only shocks to the two state variables (risk free rate and log price/dividend ratio). MEAN is the estimation period sample mean.

Premium:	Excess	Model Pricing Errors:					
	Returns	CCAPM	CAPM	LRRN	LRRS	2-STATE	MEAN
Panel A: MAD Returns and Pricing Errors (OLS)							
Equity Premium	0.159	0.164	0.142	0.144	0.143	0.144	0.142
Small-Big	0.188	0.191	0.187	0.193	0.195	0.189	0.192
Value-Growth	0.181	0.184	0.179	0.180	0.180	0.186	0.181
Win-Lose	0.202	0.184	0.215	0.140	0.145	0.196	0.137
Reversal	0.178	0.182	0.177	0.178	0.180	0.179	0.179
Credit Premium	0.072	0.072	0.069	0.085	0.087	0.071	0.070
Term Premium	0.065	0.065	0.065	0.064	0.064	0.068	0.065
Panel B: MSE Returns and Pricing Errors (OLS)							
Equity Premium	0.036	0.038	0.032	0.032	0.031	0.032	0.032
Small-Big	0.060	0.062	0.058	0.060	0.060	0.060	0.060
Value-Growth	0.048	0.050	0.046	0.046	0.047	0.051	0.046
Win-Lose	0.057	0.050	0.064	0.036	0.037	0.055	0.036
Reversal	0.052	0.055	0.051	0.051	0.052	0.052	0.050
Credit Premium	0.009	0.010	0.009	0.012	0.013	0.009	0.009
Term Premium	0.007	0.007	0.007	0.007	0.007	0.007	0.007

Table 3

Diebold-Mariano (1995) Test Statistic for mean absolute excess returns (MAD) and mean squared pricing errors (MSE) during model evaluation years, 1932-2006. The number of observations is 75. The benchmark model is the Sample Mean Model. The returns are continuously-compounded annual returns. CCAPM is the simple consumption beta model. CAPM is the Capital Asset Pricing Model. LRRN is the Nonstationary Long Run Risk Model and LRRS is the Stationary Long Run Risk Model. 2-STATE is the model with only shocks to the two state variables (risk free rate and log price/dividend ratio). These results are based on OLS risk premium estimates. The t-stats are constructed using OLS assumptions, for the two-sided test of the null hypothesis that a model performs as well as the mean. The empirical p values for the t-stats are based on 10,000 simulation trials in each case.

MAD	**Equity Premium**		**Small-Big**		**Value-Grow**		**Win-Lose**		**Reversal**		**Credit Premium**		**Term Premium**	
MODEL	**t-stat**	**p value**	**t-stat**	**p value**	**t-stat**	**p value**	**t-stat**	**p value**	**t-stat**	**p value**	**t-stat**	**p value**	**t-stat**	**p value**
CAPM	n.a.	n.a.	2.64	0.01	0.43	0.66	**-4.79**	**0.00**	0.23	0.82	1.33	0.19	-0.19	0.85
CCAPM	**-2.71**	**0.01**	0.26	0.80	-0.46	0.64	**-3.87**	**0.00**	-0.37	0.71	-1.44	0.15	-0.09	0.93
LRRS	-0.38	0.71	-1.46	0.15	0.09	0.93	-1.35	0.17	-0.24	0.81	**-2.62**	**0.01**	0.32	0.75
LRRN	-0.81	0.42	-1.16	0.25	0.17	0.86	-0.96	0.34	0.39	0.70	**-2.47**	**0.02**	0.66	0.52
2-STATE	-0.94	0.35	1.03	0.30	-0.66	0.51	**-4.18**	**0.00**	-0.01	0.99	-1.17	0.24	-1.25	0.21

MSE	**Equity Premium**		**Small-Big**		**Value-Grow**		**Win-Lose**		**Reversal**		**Credit Premium**		**Term Premium**	
MODEL	**t-stat**	**p value**	**t-stat**	**p value**	**t-stat**	**p value**	**t-stat**	**p value**	**t-stat**	**p value**	**t-stat**	**p value**	**t-stat**	**p value**
CAPM	n.a.	n.a.	1.78	0.08	0.16	0.87	**-3.68**	**0.00**	-0.37	0.71	1.08	0.28	-0.14	0.90
CCAPM	**-2.10**	**0.04**	-0.65	0.52	-1.18	0.24	**-2.48**	**0.02**	-1.25	0.22	-0.43	0.67	0.01	1.00
LRRS	1.02	0.32	-0.82	0.41	-0.23	0.82	-0.38	0.71	-0.88	0.38	**-2.31**	**0.02**	0.35	0.74
LRRN	-0.15	0.88	-0.19	0.86	-0.09	0.93	-0.19	0.85	-0.35	0.73	**-2.16**	**0.03**	0.93	0.38
2-STATE	-0.26	0.80	-0.18	0.86	-1.28	0.20	**-2.87**	**0.01**	-0.59	0.55	-0.27	0.79	-1.24	0.22

Table 4

Clark-West (2007) Test Statistic for mean squared pricing errors (MSEs) during model evaluation years, 1932-2006. The number of observations is 75. The returns are continuously-compounded annual returns. LRRN is the Nonstationary Long Run Risk Model and LRRS is the Stationary Long Run Risk Model. 2-STATE is the model with only shocks to the two state variables (risk free rate and log price/dividend ratio). Panel A and Panel C results are based on OLS risk premium estimates. Panel B and Panel D results are based on GLS risk premium estimates. The t-stats are constructed using OLS assumptions, for the one-sided test of the null hypothesis that a model performs no better than the mean. The empirical p values for the t-stats are based on 10,000 simulation trials in each case.

Panel A: Clark-West (2007) Test Statistic: Benchmark Model Sample Mean (OLS)

Mean	Equity Premium		Small-Big		Value-Grow		Win-Lose		Reversal		Credit Premium		Term Premium	
Model	**t-stat**	**p value**	**t-stat**	**p value**	**t-stat**	**p value**	**t-stat**	**p value**	**t-stat**	**p value**	**t-stat**	**p value**	**t-stat**	**p value**
LRRS	1.30	0.07	-0.51	0.69	0.66	0.25	0.74	0.21	-0.05	0.53	0.83	0.21	0.82	0.19
LRRN	0.19	0.42	-0.06	0.51	0.54	0.29	0.45	0.32	-0.08	0.54	0.81	0.22	1.23	0.06

Panel B: Clark-West (2007) Test Statistic: Benchmark Model Sample Mean (GLS)

Mean	Equity Premium		Small-Big		Value-Grow		Win-Lose		Reversal		Credit Premium		Term Premium	
Model	**t-stat**	**p value**	**t-stat**	**p value**	**t-stat**	**p value**	**t-stat**	**p value**	**t-stat**	**p value**	**t-stat**	**p value**	**t-stat**	**p value**
LRRS	1.04	0.12	0.35	0.36	0.20	0.42	0.55	0.28	0.37	0.36	0.99	0.16	0.22	0.41
LRRN	0.50	0.31	0.44	0.33	0.26	0.39	0.44	0.33	0.58	0.28	1.23	0.11	0.49	0.29

Panel C: Clark-West (2007) Test Statistic: Benchmark Model 2-State Model (OLS)

2-State	Equity Premium		Small-Big		Value-Grow		Win-Lose		Reversal		Credit Premium		Term Premium	
Model	**t-stat**	**p value**	**t-stat**	**p value**	**t-stat**	**p value**	**t-stat**	**p value**	**t-stat**	**p value**	**t-stat**	**p value**	**t-stat**	**p value**
LRRS	1.31	0.06	0.64	0.26	**2.74**	**0.00**	**6.56**	**0.00**	-0.37	0.64	1.75	0.05	**2.58**	**0.01**
LRRN	0.50	0.31	0.77	0.21	**2.97**	**0.00**	**6.50**	**0.00**	1.33	0.09	1.71	0.05	**2.84**	**0.00**

Panel D: Clark-West (2007) Test Statistic: Benchmark Model 2-State Model (GLS)

2-State	Equity Premium		Small-Big		Value-Grow		Win-Lose		Reversal		Credit Premium		Term Premium	
Model	**t-stat**	**p value**	**t-stat**	**p value**	**t-stat**	**p value**	**t-stat**	**p value**	**t-stat**	**p value**	**t-stat**	**p value**	**t-stat**	**p value**
LRRS	1.29	0.06	**2.78**	**0.00**	**2.64**	**0.00**	**3.92**	**0.00**	**2.38**	**0.00**	1.72	0.05	-0.63	0.72
LRRN	0.03	0.47	**3.04**	**0.00**	**3.73**	**0.00**	**4.00**	**0.00**	**3.30**	**0.00**	1.90	0.03	0.40	0.34

Table 5

Mean excess returns and pricing errors (actual minus forecast) during model evaluation years 1934-2006, using three-year returns obtained by lengthening the "step around" evaluation period to three years and compounding the annual returns. The number of observations is 25. CCAPM is the simple consumption beta model. CAPM is the Capital Asset Pricing Model. LRRN is the Nonstationary Long Run Risk Model and LRRS is the Stationary Long Run Risk Model. 2-STATE is the model with only shocks to the two state variables (risk free rate and log price/dividend ratio).

Premium:	Mean Excess	Model Pricing Errors: CCAPM	CAPM	LRRN	LRRS	2-STATE
Panel A: OLS Risk Premiums						
Equity Premium	20.12	16.38	0.01	8.88	11.49	19.52
Small-Big	11.91	12.49	3.97	-8.80	-10.77	10.24
Value-Growth	13.41	14.81	12.29	1.45	16.94	20.61
Win-Lose	47.53	45.43	50.30	13.68	19.00	42.27
Reversal	12.34	8.33	20.04	-7.82	-5.69	9.26
Credit Premium	6.11	9.04	-3.04	7.00	6.76	6.22
Term Premium	4.66	6.17	3.71	-4.85	2.50	-3.62
Panel B: GLS Risk Premiums						
Equity Premium	20.12	18.60	0.01	7.37	14.28	28.26
Small-Big	11.91	12.12	3.97	-15.34	-9.16	18.23
Value-Growth	13.41	13.91	12.29	-12.98	17.40	18.04
Win-Lose	47.53	46.99	50.30	16.64	22.11	38.66
Reversal	12.34	10.93	20.04	-10.12	-4.38	9.36
Credit Premium	6.11	7.45	-3.04	1.79	7.58	13.08
Term Premium	4.66	5.20	3.71	-4.06	4.68	4.34

Table 6

Mean absolute excess returns (MAD), mean squared excess returns (MSE), MAD and MSE pricing errors during model evaluation years 1934-2006,using three-year returns obtained by lengthening the "step around" evaluation period to three years and compounding the annual returns. The number of observations is 25. CCAPM is the simple consumption beta model. CAPM is the Capital Asset Pricing Model. LRRN is the Nonstationary Long Run Risk Model and LRRS is the Stationary Long Run Risk Model. These results are based on OLS risk premium estimates. 2-STATE is the model with only shocks to the two state variables (risk free rate and log price/dividend ratio). MEAN is the estimation period sample mean.

Premium:	Excess	Model Pricing Errors:					
	Returns	CCAPM	CAPM	LRRN	LRRS	2-STATE	MEAN
Panel A: MAD Returns and Pricing Errors (OLS)							
Equity Premium	0.298	0.260	0.242	0.246	0.256	0.309	0.242
Small-Big	0.407	0.404	0.399	0.460	0.464	0.400	0.430
Value-Growth	0.336	0.345	0.323	0.328	0.382	0.359	0.328
Win-Lose	0.504	0.491	0.531	0.290	0.309	0.463	0.252
Reversal	0.318	0.320	0.333	0.303	0.312	0.326	0.318
Credit Premium	0.160	0.186	0.165	0.171	0.160	0.170	0.157
Term Premium	0.124	0.133	0.120	0.134	0.138	0.138	0.125
Panel B: MSE Returns and Pricing Errors (OLS)							
Equity Premium	0.135	0.118	0.100	0.114	0.112	0.145	0.100
Small-Big	0.272	0.272	0.254	0.296	0.294	0.271	0.277
Value-Growth	0.178	0.189	0.172	0.168	0.216	0.200	0.175
Win-Lose	0.315	0.298	0.344	0.131	0.134	0.271	0.099
Reversal	0.141	0.144	0.165	0.133	0.129	0.146	0.137
Credit Premium	0.046	0.053	0.046	0.048	0.044	0.052	0.045
Term Premium	0.026	0.030	0.025	0.029	0.030	0.028	0.027

Table 7

Mean excess returns and pricing errors (actual minus forecast) during model evaluation years, 1976-2006, using a 45-year rolling estimation window. The number of observations is 31. The continuously-compounded returns are in annual percent. CCAPM is the simple consumption beta model. CAPM is the Capital Asset Pricing Model. LRRN is the Nonstationary Long Run Risk Model and LRRS is the Stationary Long Run Risk Model. 2-STATE is the model with only shocks to the two state variables (risk free rate and log price/dividend ratio).

Premium:	Mean Excess	Model Pricing Errors: CCAPM	CAPM	LRRN	LRRS	2-STATE
Panel A: OLS Risk Premiums						
Equity Premium	6.37	8.39	0.10	4.27	4.52	7.95
Small-Big	3.64	5.24	0.14	2.22	4.94	4.31
Value-Growth	8.60	10.49	7.58	5.35	3.89	8.09
Win-Lose	17.16	12.62	19.30	5.06	8.71	15.75
Reversal	4.34	7.15	4.76	-3.40	-0.69	4.31
Credit Premium	1.61	1.94	0.18	1.92	0.85	1.39
Term Premium	2.93	2.30	2.43	-0.94	-1.26	3.76
Panel B: GLS Risk Premiums						
Equity Premium	6.37	10.47	0.10	7.74	7.88	8.07
Small-Big	3.54	6.89	0.14	4.77	9.56	9.06
Value-Growth	8.60	12.42	7.58	5.13	4.75	10.41
Win-Lose	17.16	7.95	19.30	2.67	4.48	12.99
Reversal	4.35	10.05	4.84	-2.09	2.76	8.10
Credit Premium	1.61	2.28	0.18	2.51	2.23	1.92
Term Premium	2.93	1.65	2.43	0.64	0.43	3.73

Table 8

Mean absolute excess returns (MAD) and MAD pricing errors during model evaluation years based on a 45 year rolling estimation period and step-ahead forecasts. The forecast evaluation period is 1976-2006. The number of observations is 31. The returns are continuously-compounded annual returns and the statistics are stated as annual decimal fractions. CCAPM is the simple consumption beta model. CAPM is the Capital Asset Pricing Model. LRRN is the Nonstationary Long Run Risk Model and LRRS is the Stationary Long Run Risk Model. 2-STATE is the model with only shocks to the two state variables (risk free rate and log price/dividend ratio). MEAN is the estimation period sample mean.

Premium:	Excess	Model Pricing Errors:					
	Returns	CCAPM	CAPM	LRRN	LRRS	2-STATE	MEAN
Panel A: MAD Returns and Pricing Errors (OLS)							
Equity Premium	0.149	0.145	0.116	0.136	0.129	0.143	0.116
Small-Big	0.158	0.176	0.165	0.163	0.171	0.183	0.166
Value-Growth	0.164	0.190	0.181	0.174	0.173	0.187	0.180
Win-Lose	0.211	0.201	0.240	0.174	0.183	0.219	0.159
Reversal	0.163	0.185	0.177	0.180	0.174	0.187	0.171
Credit Premium	0.058	0.070	0.066	0.073	0.075	0.069	0.067
Term Premium	0.070	0.090	0.091	0.095	0.089	0.095	0.092
Panel B: MAD Returns and Pricing Errors (GLS)							
Equity Premium	0.149	0.153	0.116	0.152	0.147	0.144	0.116
Small-Big	0.158	0.183	0.165	0.181	0.188	0.207	0.166
Value-Growth	0.164	0.202	0.181	0.190	0.208	0.203	0.180
Win-Lose	0.211	0.177	0.240	0.165	0.170	0.212	0.159
Reversal	0.163	0.193	0.177	0.195	0.199	0.197	0.171
Credit Premium	0.058	0.070	0.068	0.071	0.068	0.071	0.067
Term Premium	0.070	0.090	0.091	0.092	0.088	0.095	0.092

Table A.1

Mean excess returns and pricing errors (actual minus forecast) during model evaluation years, 1947-2006. The number of observations is 60. The continuously-compounded returns are in annual percent. CCAPM is the simple consumption beta model. CAPM is the Capital Asset Pricing Model. LRRN is the Nonstationary Long Run Risk Model and LRRS is the Stationary Long Run Risk Model. These results use GLS risk premiums. 2-STATE is the model with only shocks to the two state variables (risk free rate and log price/dividend ratio).

Premium:	Mean	Model Pricing Errors:				
	Excess	CCAPM	CAPM	LRRN	LRRS	2-STATE
Equity Premium	6.46	9.76	0.06	0.84	4.03	1.93
Small-Big	2.39	5.01	0.49	-1.09	4.12	12.52
Value-Growth	5.80	8.86	5.78	-0.28	-4.20	11.15
Win-Lose	17.29	9.87	19.81	0.24	1.33	9.37
Reversal	3.19	7.78	4.74	0.77	-0.56	12.78
Credit Premium	1.58	2.13	0.06	0.06	1.58	1.02
Term Premium	1.08	0.05	0.73	-0.24	-2.50	-0.95

Table A.2

Mean absolute excess returns (MAD), mean squared excess returns (MSE), MAD and MSE pricing errors during model evaluation years, 1932-2006. The number of observations is 75. The returns are continuously-compounded annual returns and the statistics are stated as annual decimal fractions. CCAPM is the simple consumption beta model. CAPM is the Capital Asset Pricing Model. LRRN is the Nonstationary Long Run Risk Model and LRRS is the Stationary Long Run Risk Model. These results are based on GLS risk premium estimates. 2-STATE is the model with only shocks to the two state variables (risk free rate and log price/dividend ratio). MEAN is the estimation period sample mean.

Premium:	Excess Returns	Model Pricing Errors: CCAPM	CAPM	LRRN	LRRS	2-STATE	MEAN
Panel A: MAD Returns and Pricing Errors (GLS)							
Equity Premium	0.159	0.173	0.142	0.145	0.143	0.145	0.142
Small-Big	0.188	0.193	0.187	0.189	0.192	0.197	0.192
Value-Growth	0.181	0.188	0.179	0.188	0.190	0.194	0.181
Win-Lose	0.202	0.161	0.215	0.150	0.156	0.173	0.137
Reversal	0.178	0.187	0.177	0.181	0.185	0.188	0.179
Credit Premium	0.072	0.073	0.069	0.069	0.068	0.072	0.070
Term Premium	0.065	0.064	0.065	0.065	0.065	0.065	0.065
Panel B: MSE Returns and Pricing Errors (GLS)							
Equity Premium	0.036	0.042	0.032	0.032	0.031	0.032	0.032
Small-Big	0.060	0.063	0.058	0.061	0.064	0.067	0.060
Value-Growth	0.048	0.053	0.046	0.052	0.053	0.056	0.046
Win-Lose	0.057	0.041	0.064	0.038	0.040	0.045	0.036
Reversal	0.052	0.060	0.051	0.053	0.057	0.060	0.050
Credit Premium	0.009	0.010	0.009	0.009	0.009	0.009	0.009
Term Premium	0.007	0.007	0.007	0.007	0.007	0.007	0.007

Table A.3

Mean excess returns and pricing errors (actual minus forecast) during model evaluation years, 1948-2004. The number of observations is 57. The consumption data are quarterly and not seasonally adjusted. The continuously-compounded returns are in annual percent. CCAPM is the simple consumption beta model. CAPM is the Capital Asset Pricing Model. LRRN is the Nonstationary Long Run Risk Model and LRRS is the Stationary Long Run Risk Model. These results use GLS risk premiums. 2-STATE is the model with only shocks to the two state variables (risk free rate and log price/dividend ratio).

Premium:	Mean			Model Pricing Errors:		
	Excess	CCAPM	CAPM	LRRN	LRRS	2-STATE
Panel A: OLS Risk Premiums						
Equity Premium	6.44	4.00	0.00	0.08	0.23	4.78
Small-Big	2.42	3.32	0.53	-1.30	3.33	4.02
Value-Growth	5.66	2.11	5.67	-0.31	-4.33	6.65
Win-Lose	17.91	14.91	20.39	1.43	5.01	17.08
Reversal	3.34	4.96	4.91	3.10	2.53	4.61
Credit Premium	1.44	-0.08	-0.09	-1.48	-0.97	1.27
Term Premium	1.12	7.45	0.76	1.61	-1.63	0.29
Panel B: GLS Risk Premiums						
Equity Premium	6.44	4.63	0.00	-1.06	0.48	2.83
Small-Big	2.42	3.11	0.53	-1.31	3.12	13.33
Value-Growth	5.66	2.99	5.67	1.63	-4.74	10.49
Win-Lose	17.91	15.66	20.39	2.20	4.91	12.08
Reversal	3.34	4.56	4.91	3.36	2.46	11.15
Credit Premium	1.44	0.22	-0.09	-1.40	-0.99	1.58
Term Premium	1.12	5.88	0.76	1.14	-1.27	-0.12

Table A.4

Mean absolute excess returns (MAD), mean squared excess returns (MSE), MAD and MSE pricing errors during model evaluation years, 1948-2004. The number of observations is 57. The consumption data are quarterly and not seasonally adjusted. The returns are continuously-compounded annual returns and the statistics are stated as annual decimal fractions. CCAPM is the simple consumption beta model. CAPM is the Capital Asset Pricing Model. LRRN is the Nonstationary Long Run Risk Model and LRRS is the Stationary Long Run Risk Model. These results are based on GLS risk premium estimates. 2-STATE is the model with only shocks to the two state variables (risk free rate and log price/dividend ratio). MEAN is the estimation period sample mean.

Premium:	Excess Returns	Model Pricing Errors: CCAPM	CAPM	LRRN	LRRS	2-STATE	MEAN
Panel A: MAD Returns and Pricing Errors (GLS)							
Equity Premium	0.152	0.142	0.134	0.133	0.134	0.149	0.134
Small-Big	0.163	0.164	0.162	0.169	0.166	0.167	0.167
Value-Growth	0.166	0.165	0.166	0.167	0.166	0.168	0.166
Win-Lose	0.213	0.193	0.230	0.133	0.139	0.207	0.132
Reversal	0.165	0.168	0.168	0.169	0.170	0.167	0.168
Credit Premium	0.058	0.056	0.056	0.058	0.055	0.058	0.057
Term Premium	0.072	0.093	0.072	0.070	0.074	0.073	0.073
Panel B: MSE Returns and Pricing Errors (GLS)							
Equity Premium	0.031	0.028	0.028	0.027	0.027	0.030	0.028
Small-Big	0.046	0.046	0.045	0.048	0.048	0.048	0.048
Value-Growth	0.041	0.038	0.040	0.039	0.041	0.042	0.039
Win-Lose	0.062	0.053	0.072	0.031	0.034	0.060	0.032
Reversal	0.039	0.041	0.041	0.040	0.041	0.041	0.040
Credit Premium	0.006	0.005	0.006	0.006	0.005	0.006	0.006
Term Premium	0.008	0.014	0.008	0.008	0.008	0.008	0.008