The "Out-of-sample" Performance of Long-Run Risk Models[@]

Wayne Ferson[a*], Suresh Nallareddy[a], Biqin Xie[a]

[a]*Marshall School of Business, University of Southern California, Los Angeles, CA. 90089 USA*

first draft: February 11, 2009
this draft: January 29, 2012

ABSTRACT

This paper studies the ability of long-run risk models to explain out-of-sample asset returns during 1931-2009. The long-run risk models perform relatively well on the momentum effect. A cointegrated version of the model outperforms the classical, stationary version. Both the long-run and the short run consumption shocks in the models are empirically important for the models' performance. The models' average pricing errors are especially small in the decades from the 1950s to the 1990s. When we restrict the risk premiums to identify structural parameters, this results in larger average pricing errors but often smaller error variances. The mean squared errors are not substantially better than those of the classical CAPM, except for Momentum.

[*]Corresponding author.
email address: ferson@marshall.usc.edu, www-rcf.usc.edu/~ferson/.

[@]We are grateful to an anonymous referee, Ravi Bansal, Jason Beeler, David P. Brown, Dana Kiku, Ayse Imrohoroglu, Selahattin Imrohoroglu, Chris Jones, Sergei Sarkissian, Zhiguang Wang, Jianfeng Yu, Guofu Zhou, and to participants in workshops at the University of British Columbia, Claremont McKenna College, Florida International University, the University of California San Diego, the University of North Carolina Charlotte, University of Notre Dame, Ohio State University, the University of Oregon, the Oxford-Man Institute, Queens University, the University of Southern California, the University of Utah, and the University of Washington for discussions and comments. We are also grateful to participants at the 2010 Duke Asset Pricing Conference, the 2010 First World Finance Conference and the 2010 Conference on Financial Economics and Accounting.

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

ABSTRACT

This paper studies the ability of long-run risk models to explain out-of-sample asset returns during 1931-2009. The long-run risk models perform relatively well on the momentum effect. A cointegrated version of the model outperforms the classical, stationary version. Both the long-run and the short run consumption shocks in the models are empirically important for the models' performance. The models' average pricing errors are especially small in the decades from the 1950s to the 1990s. When we restrict the risk premiums to identify structural parameters, this results in larger average pricing errors but often smaller error variances. The mean squared errors are not substantially better than those of the classical CAPM, except for Momentum.

JEL Classification: G12, E44

Keywords: Long-run risk models, Out-of-sample, Equity premium puzzle, Size effect, Book to market effect, Momentum, Reversals, Term premium, Default premium

1. Introduction

The long-run risk model following Bansal and Yaron (2004) has been a phenomenal success. A rapidly expanding literature finds the model useful for the equity premium puzzle, size and book-to-market effects, momentum, long-term return reversals in stock prices, risk premiums in bond markets, real exchange rate movements and more (see the review by Bansal, 2007). However, the evidence to date is based largely on calibration exercises and some in-sample data fitting. If long-run risk models are to ultimately be useful for practical applications, their performance in an out-of-sample setting is important. This paper provides an empirical analysis of the models from this perspective.

We study the "out-of-sample" performance of long-run risk models for explaining the equity premium puzzle, size and book-to-market effects, momentum, reversals, and bond returns of different maturity and credit quality. Comparing the fit out of sample allows us to evaluate features of model performance that are important for practical applications. (As explained below, we put "out-of-sample" in quotes to emphasize some qualifying remarks.) We examine both stationary and cointegrated versions of the models using annual data for 1931-2009.

Our first main result is that the overall performance of a cointegrated version of the long-run risk model is better than the original, stationary version. Thus, we argue that the out-of-sample performance justifies a more prominent role for cointegrated models in future research.

We find that the long-run risk models perform relatively well in explaining the Momentum Effect. At the same time, they perform poorly in explaining the relative returns to low-grade corporate and long-term government bonds. These results point to the importance of missing factors in the simple, long-run risk models.

To evaluate the risk factors that drive the fit of the long-run risk models we examine models that suppress the consumption-related shocks. We find that these models perform poorly, indicating that the consumption shocks are important risk factors. The long-run risk models also perform better than the simple consumption-based (CCAPM) model. Thus, both the short-run consumption risk factor and the long-run risk factor are important ingredients in the models' performance.

We examine the impact of restricting the risk premiums on the long-run risk models' factors to identify structural parameters of the model. Imposing consistency with economically reasonable values of risk aversion and intertemporal substitution reduces the degrees of freedom, and has the effect of increasing the average out-of-sample pricing errors substantially, while sometimes reducing the pricing error variances. Compared to the classical Capital Asset Pricing Model (CAPM, Sharpe, 1964) the restricted model's overall performance is inferior.

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 returns, dividend yields and interest rates, to the sample period, to outliers in the data, and to biases from lagged stochastic regressors. Holding the parameters fixed, the long-run risk models perform better in the 1951-1990 period than in the decades before or in more recent data, but the cointegrated models dominate in every subperiod.

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 main empirical results and Section 6 examines their robustness. Section 7 concludes.

2. The models

We study stationary and cointegrated versions of long-run risk models. As the long-run risk literature has expanded, papers have offered 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):

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

$$x_t = \mu + \rho_x\, x_{t-1} + \phi\, \sigma_{t-1}\, \varepsilon_{xt} \tag{1b}$$

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

where c_t is the natural logarithm of aggregate consumption expenditures, Δ is the first difference 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 average 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) show that in this model, the innovations in the log of the stochastic discount factor are to good approximation linear in the three-vector of shocks $u_t = [\sigma_{t-1}\varepsilon_{ct},\ \phi\, \sigma_{t-1}\varepsilon_{xt},\ \varepsilon_{\sigma t}]$. 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 returns with the shocks.[1]

We focus on unconditional expected returns, because many of the empirical regularities 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-to-market effects). This leaves open the possibility that, like the CAPM and CCAPM in sample (e.g., Jagannathan and Wang, 1996; Lettau and Ludvigson, 2001), the long-run risk approach works better in a conditional form. We therefore also evaluate simple conditional versions of the models.

2.2. Cointegrated models

The cointegrated model follows Bansal, Dittmar, and Lundblad (2005), Bansal, Gallant, and Tauchen (2007), and Bansal, Dittmar, and Kiku (2009), 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 (2009) allow for time trends in the levels of dividends and consumption, but find that they don't have much effect, and we find a similar

[1] Writing the log 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.

result.

We allow for the possibility that the volatility shock is a priced risk, as in the stationary model. Constantinides and Ghosh (2012) 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 shocks.

3. Empirical methods

The finance literature uses three main approaches to evaluate asset pricing models: calibration, in-sample estimation and out-of-sample fit. These approaches are not mutually exclusive. Long-run risks models have been extensively calibrated with great success. With calibration, values for the parameters of the model are chosen, and the model is solved at these parameter values for the prices of financial assets. The model-generated series of prices and returns are examined to see if their moments match key moments of actual asset prices and returns.

With estimation, the model parameters are chosen to fit a panel of economic series and asset returns, to minimize a statistical criterion. Standard errors for the parameter estimates quantify their precision. Statistical hypothesis tests about the parameters may be conducted and the residuals of the model are examined to assess the fit to the sample. Estimation typically challenges a model in more dimensions at once than calibration. For example, a calibrated parameter value may not be the value that maximizes the likelihood, indicating that more is going on in the data than was captured by the calibration. Estimation sacrifices the clarity of a calibration experiment for this kind of information. Long-run risk models are estimated by Bansal, Kiku, and Yaron (2010) and Constantinides and Ghosh (2012), among others.

The practical utility of an asset pricing model ultimately depends on its ability to fit out-of-sample returns, as most practical applications are, in some sense, out of sample. For example, firms want to estimate costs of capital for future projects, portfolio and risk managers want to know the expected compensation for future risks, and academic researchers will want to make risk adjustments to expected returns in future data. Many of these applications rely on out-of-sample estimates for the required or ex ante expected return, where the model parameters must be chosen on the basis of available data.

This perspective leads naturally to the mean squared pricing error (MSE) criterion for model evaluation. A model that delivers the correct (conditional) mean of the future return minimizes the (conditional) MSE. That is, δ minimizes $E_t\{(r_{t+1} - \delta)^2\}$ when $\delta = E_t(r_{t+1})$, where $E_t(\cdot)$ is the conditional expectation at time t. We use the out-of-sample MSE to compare the models' performance.

We are also interested in the two components of the MSE, the expected pricing error or bias and the error variance. Let $r_{t+1} = E_t(r_{t+1}) + \varepsilon_{t+1}$, where ε_{t+1} is independent of time-t information, and let $\delta(\hat{\phi})$ be the prediction of a model at the estimated parameters $\hat{\phi}$, where ϕ are the true parameters of the model. The out-of-sample pricing error is:

$$\begin{aligned} v_{t+1} &= r_{t+1} - \delta(\hat{\phi}) \\ &= \varepsilon_{t+1} + [E_t(r_{t+1}) - \delta(\phi)] + [\delta(\phi) - \delta(\hat{\phi})] \end{aligned} \quad (3)$$

The expected error $E(v_{t+1})$ reflects a model's average error at its true parameter values (the second term above) plus parameter estimation error (the third term). For the nonlinear models examined here, the third term in Eq. (3) will not average to zero, even if $E(\hat{\phi}) = \phi$, and the average pricing errors are affected by both model error and parameter estimation errors.

The variance of the pricing errors, v_{t+1}, may be used to focus on the relative effects of parameter estimation error when comparing models. Using Eq. (3), the variance of the pricing error is:

$$\begin{aligned} &\text{var}(v_{t+1}) = E\{(v_{t+1} - E(v_{t+1}))^2\} \\ &= \sigma_\varepsilon^2 + \text{var}[E_t(r_{t+1})] + \text{var}[\delta(\hat{\phi})] - 2E\{[E_t(r_{t+1}) - E(r)][\delta(\hat{\phi}) - E(\delta(\hat{\phi}))]\}, \end{aligned} \tag{4}$$

where σ_ε^2 is the variance of the unexpected return, ε_{t+1}. The σ_ε^2 term is likely to be dominant in Eq. (4) when applied to stock returns. Thus, the prediction error variance may be a low power statistic. However, comparing two models on the same return we examine the difference between two Eq. (4)s, and the first two terms drop out of the comparison. Thus, the largest source of noise may be avoided by comparing two models' variances on the same asset.

Our out-of-sample analysis is further motivated by a central feature of long-run-risk models: a small, highly persistent component in consumption and/or dividends – the long-run risk component. This persistent component is modeled either as stationary or as cointegrated. Latent components in stock returns that are stationary but persistent exacerbate problems with spurious regression and data snooping (Ferson, Sarkissian, and Simin, 2003) and may induce lagged stochastic regressor bias (Stambaugh, 1999). These issues may also arise in long-run risk models, in which case the in-sample fit may appear overly optimistic. An out-of-sample analysis can alert us to these issues.

To evaluate out-of-sample fit, we use a traditional "rolling" estimation. We estimate the model parameters over an initial period and predict returns for the next year. The estimation period is rolled forward by one year and the process is repeated. Because the long-run risk models may need long samples to estimate the parameters with precision, we also examine the in-sample fit, using the full sample of data.[2]

3.1. The reliability and power of the methods

It is crucial for the interpretation of the results that the empirical methods are reliable enough to distinguish between models in the face of finite samples and data issues such as time aggregation and noise. We approach this concern from several angles. In the robustness section we examine variations on the approach to address time aggregation, autocorrelation, the length of the asset holding period, finite sample biases, the number of test assets, the method for identifying the stochastic volatility in the model, the measures of the price/dividend ratio and risk-free rate.

We also use simulations to evaluate the methods. A referee simulates data assuming that the stationary long-run risk model is the true model. We compare the results from simulations that assume that the CAPM is the true model. The distributions of the average pricing errors shift by significant amounts between the two cases, suggesting that our approach has statistical power. For example, when the LRRS model generates the data the average pricing errors we observe for the CAPM in the actual data are far out in the tail of the empirical distribution and would therefore reject the CAPM. In contrast, when the CAPM generates the data the average pricing errors we observe for the CAPM in the actual data are not statistically significant. The pricing error MSE differences between the CAPM and LRRS model present very tight empirical distributions in the simulations, further indicating statistical power.

A formal evaluation of power requires a series of simulations under both the null and

---

[2] In earlier versions of the paper we used an alternative, "step around" approach where we estimate the models' parameters using all of the available data outside of the evaluation period. The results are broadly similar to the full sample estimation reported here, and are available by request to the authors.

alternatives, and is beyond the scope of this paper. However, we conduct simulations where the null hypothesis is that the population value of a t-test for the difference between the prediction errors for two models is zero. An advantage of this approach is that the t-statistic is pivotal, which improves the performance of the bootstrap. We generate the empirical distributions for the difference statistics under the null by resampling from the mean-centered sample statistics. We generate samples with the same sizes as those used in the actual tests. While these simulations do not explicitly model time aggregation or other sources of noise, they implicitly capture the variance that such factors generate in the test statistics.

Using empirical p-values from these simulations, our test statistics have the power to discriminate between the various models. For example, in Table 1 below, we find that the average pricing errors of the stationary long run risk model are significantly different from those of the CAPM in eleven out of the fourteen cases. In Table 3 the MSE pricing errors of the stationary long run risk model and the CAPM are significantly different in two cases. In Table 5, where we examine nested models and the Two-state model is the benchmark, the long-run risk models are significantly different (at the 10% level) from the benchmark model in 27 out of the 42 cases. All of these are examples with rolling estimation, so our approach has power even with the shorter, rolling samples.

3.2. Stationary model estimation

For the stationary model we estimate the following equations:

$$u_{1t} = \Delta c_t - [a_0 + a_1\, rf_{t-1} + a_2 \ln(P/D)_{t-1}] \quad (5a)$$

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

$$u_{2t} = u_{1t}^2 - [b_0 + b_1\, rf_{t-1} + b_2 \ln(P/D)_{t-1}] \quad (5b)$$

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

$$u_{3t} = x_t - d - \rho_x x_{t-1} \tag{5c}$$

$$u_{4t} = \sigma_t^2 - k - \rho_\sigma \sigma_{t-1}^2 \tag{5d}$$

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

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

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 $E\{(u_{1t},u_{2t}) \otimes (1, rf_{t-1}, \ln(P/D)_{t-1})\}=0$, $E\{u_{3t}\,(1, x_{t-1})\}=0$, $E\{u_{4t}\,(1, \sigma_{t-1}^2)\}=0$, $E\{u_{5t}\,(1,u_{1t},u_{3t},u_{4t})\}=0$ and $E\{u_{6t}\}=0$. The system is exactly identified.[3]

The state vector in the model is the risk-free rate and aggregate price/dividend ratio: $s_t = \{rf_t, \ln(P/D)_t\}$. Eqs. (5a) and (5b) reflect the fact, as shown by Bansal, Kiku, and Yaron (2010) and Constantinides and Ghosh (2012), that the conditional mean of consumption growth and the conditional variance can be identified as affine functions of these state variables.

From the system (5) we identify the priced heteroskedastic shocks that drive the stochastic discount factor. Comparing systems (1) and (5), we have:

$$u_{1t} = \sigma_{t-1}\, \varepsilon_{ct},\;\; u_{3t} = \varphi\, \sigma_{t-1}\, \varepsilon_{xt},\;\; u_{4t} = \varepsilon_{\sigma t}. \tag{6}$$

In Eq. (5e), β is an N x 3 matrix of the betas of the N excess returns in r_t with the priced shocks,

[3] 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.

$\{u_{1t}, u_{3t}, u_{4t}\}$.[4]

Eq. (5f) 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, 2008). We refer to these risk premium estimates as the GLS risk premiums. We also present results for simpler OLS risk premiums, where we set V(r) equal to the identity matrix[5]. The OLS risk premiums are not as efficient, asymptotically, as the GLS risk premiums. However, they might be more reliable in small samples[6].

For a given evaluation period the fitted expected excess return is the estimate of β multiplied by the estimate of λ, where the estimates use data for the nonoverlapping estimation period. We denote the model in system (5) as the LRRS model in the tables below.

3.3. A restricted, stationary model

The reduced forms in system (5) are sufficient to generate out-of-sample return predictions. However, the coefficients are functions of the deeper structural parameters of the model. Restricting the coefficients allows us to recover estimates of these parameters, and by reducing the number of parameters to be estimated, may improve the efficiency of estimation and the out-of-sample performance. On the other hand, if the restrictions are inconsistent with the

---

[4] We recover the stochastic volatility from the squared consumption innovations in Eq. (5b). In the model the volatility shocks also drive the volatility of the expected growth shocks, u_{3t}. We also conduct experiments, discussed below, where we recover the volatility from these residuals.

[5] The covariance matrix of the residuals from regressing the excess returns on the factors could alternatively be used here, but the risk premium estimates would be numerically identical to those using the covariance matrix of the returns.

[6] Since we form the mimicking portfolios using only the seven excess returns, they are not the maximum correlation portfolios in a broader asset universe. This means that the risk premium estimates might "overfit" the test asset returns. We explore this issue by varying the number of test assets below.

data they could hurt the out-of-sample performance.

Constantinides and Ghosh (2012) find rejections of long-run risk models when the deeper structure is imposed. They use the unconditional means, variances, autocovariances and higher moments of consumption and dividend growth to identify the structural parameters.

Our restricted models focus on the risk premiums and the predicted required returns. In particular, we leave Eqs. (5a - 5d) unchanged, append one more moment condition to identify κ_1, and replace (5e) and (5f) with:

$$u_{5t} = u_{3t}^2 - \phi^2 u_{1t}^2 \tag{7e}$$

$$u_{6t} = r_t - r_t\{\gamma u_{1t} + \lambda_3 u_{3t} + \lambda_4 u_{4t}\}, \tag{7f}$$

where:

γ is the coefficient of risk aversion,

$$\lambda_4 = (1-\theta)\kappa_1\{\frac{0.5[(\theta - \frac{\theta}{\psi})^2 + (\theta\kappa_1 A_1\phi)^2]}{\theta(1-\kappa_1\rho_\sigma)}\} \equiv (1-\theta)\kappa_1 A_2,$$

$$\kappa_1 = \exp\{E\ln(P/D)\}/[1+\exp\{E\ln(P/D\}],$$

$$\theta = \frac{1-\gamma}{1-\frac{1}{\psi}}, \text{ and}$$

ψ is the elasticity of intertemporal substitution.

The new moment restrictions are $E(u_{5t}) = E(u_{6t}) = 0$. Since the error terms u_{1t}, u_{3t}, and u_{4t} are mean zero, Eq. (7f) specifies the expected risk premiums as $E(r) = \gamma\,\mathrm{cov}(r,u_1) + \lambda_3\,\mathrm{cov}(r,u_3) + \lambda_4\,\mathrm{cov}(r,u_4)$. Note that with N=7 assets in (7f), the moment condition (7e) and the three parameters $\{\phi,\gamma,\psi\}$ the system is now overidentified with 5 degrees of freedom[7]. With these restrictions we do not compare OLS with GLS estimation of the risk premiums. We refer to this restricted version of the model as RLRRS in the tables below.

Fig. 1 presents time-series plots of the estimates of the utility function parameters $\{\gamma, \psi\}$, the risk aversion and the elasticity of intertemporal substitution, from the RLRRS model using rolling estimation. In the estimation we penalize the GMM criterion function to force economically reasonable values of the parameters, that is $\gamma > 0$ and $\psi > 1$. [8] The figure shows that the estimates of the elasticity of substitution stay just above 1.0 and are stable in the rolling estimation. The risk aversion parameter fluctuates more, with values between about one and three, averaging near two. Thus, the model fits returns with economically reasonable parameter values.

3.4. Cointegrated model estimation

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

$$u_{1t} = d_t - \delta_0 - \delta_1 c_t \tag{8a}$$

$$u_{2t} = \Delta c_t - [a_0 + a_1 rf_{t-1} + a_2 \ln(P/D)_{t-1} + a_3 u_{1t-1}] \tag{8b}$$

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

$$u_{3t} = rf_t - [g_0 + g_1 rf_{t-1} + g_2 \ln(P/D)_{t-1} + g_3 u_{1t-1}] \tag{8c}$$

$$u_{4t} = \ln(P/D)_t - [h_0 + h_1 rf_{t-1} + h_2 \ln(P/D)_{t-1} + h_3 u_{1t-1}] \tag{8d}$$

---

[7] It would be possible to go further, restricting the coefficients in (5a – 5d) as in Constantinides and Ghosh (2012). Since this part of the system is exactly identified, we prefer to leave it in reduced form. This way we do not rely to as great an extent on the model-implied predictability and higher moments of consumption and dividend growth, which Constantinides and Ghosh find to be mispecified. Our approach focuses more on information about risk premiums.

[8] The GMM moment conditions may be multiplied by any smooth function of the parameters. We use sums of exponential functions, chosen to get large as γ approaches zero from above or as ψ approaches 1.0 from above. Using iterated GMM with an optimal weighting matrix, the scaling functions cancel out of the final minimized value criterion function. A similar approach is used in Ferson and Constantinides (1991).

$$u_{5t} = r_t - \mu - \beta(u_{1t},u_{2t},u_{3t},u_{4t})' \quad (8e)$$

$$u_{6t} = \lambda - [\beta' V(r)^{-1}\beta]^{-1}\beta' V(r)^{-1} r_t, \quad (8f)$$

In the cointegrated model there are four priced shocks, as shown by Constantinides and Ghosh (2012), as the cointegrating residual becomes a new state variable. Identification and estimation are similar to the stationary model. We refer to this version of the model as LRRC in the tables below.

The GMM estimator of the cointegrating parameter δ_l is superconsistent and has a nonstandard limiting distribution, as shown by Stock (1987). This implies that the u_{1t-1} shock estimates used in (8b-8e) may be more precise than standard estimates. This may contribute to our finding that the cointegrated model performs better than the stationary model. For example, we find that time-series plots of estimated rolling-window cointegration parameters appear smoother than rolling-window estimates of the autoregression coefficient for expected consumption growth.

Both the stationary and cointegrated long-run risk models follow the literature in using a risk-free rate and log price/dividend ratio as state variables. It is natural to ask whether the ability of the long-run risk models to fit returns is driven by the choice of these two state variables.[9] We assess the importance of the state variables for the models' explanatory power by suppressing the consumption-related risk factors. If the risk-free rate and dividend yield are the main drivers of the ability of the model to fit returns, then the models without consumption

[9] Campbell (1996) uses the innovations in lagged predictor variables, including a dividend yield and interest rates, as risk factors in an asset pricing model. 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-to-market factors of Fama and French (1993).

should perform well. Given the smaller number of parameters to estimate, these "2-State-Variable" models might be expected to perform even better out of sample than the full models.

3.5. A restricted cointegrated model

Our restricted version of the cointegrated long-run risk model (denoted RLRRC) follows a similar approach as the RLRRS model. The Eqs. (8a – 8d) are left unchanged. A moment condition to identify κ_1 is included. The variance Eq. (5b) is included in the system but this now includes the lagged cointegrating residual with coefficient b_3, this is used along with (7e) and 5(c) to identify φ in Eq. (6). The coefficients $\{a_1, a_2, a_3\}$ from (8b) and $\{b_1, b_2, b_3\}$ from (5b) are embedded in the restrictions on the risk premiums, as derived by Constantinides and Ghosh (2012). We replace Eqs. (8e) and (8f) with:

$$u_{6t} = r_t - r_t\{\gamma u_{2t} - c_3 u_{3t} - c_4 u_{4t} - c_5 u_{1t}\}, \tag{9}$$

where $c_3 = (\theta - 1)\kappa_1[A_1 a_1 + A_2 b_1]$,

$c_4 = (\theta - 1)\kappa_1[A_1 a_2 + A_2 b_2]$, and

$c_5 = (\theta - 1)\kappa_1[A_1 a_3 + A_2 b_3]$.

The equations above (9) in the system are exactly identified, and the final equation provides N=7 moment conditions to identify the parameters $\{\gamma_1, \psi\}$, leaving 5 degrees of freedom.

3.6. Estimating the CAPM

The classical CAPM is compared to the long-run risk models and we estimate it in a similar way for comparability. We use Eq. (5a), where the excess return of the market portfolio replaces the consumption growth rate to define the shocks, u_{1t}. Eq. (5e), with only these shocks, delivers the market betas and a simplified Eq. (5f) delivers the market risk premium as the cross-sectional regression slope of excess returns on market betas. Under GLS, a cross-sectional regression recognizes the market excess return as the efficient risk premium estimator (Shanken, 1992), so we use the average of the market excess return over the estimation period to estimate λ in the GLS case.

3.7. 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 models would be expected to perform worse in actual practice than our evidence suggests. Furthermore, we use revised data that would not be available in real time. These issues explain the qualification in the introduction and the title. An ideal out of sample exercise would in our view use fresh data.

The estimation is subject to finite sample biases. The risk-free rate and dividend yield state variables are modeled as autoregressions and the variables are highly persistent. Stambaugh (1999) shows that the finite sample bias in regressions using such variables can be substantial. We address finite sample bias by applying a correction developed by Amihud, Hurvich, and Wang (2009). These results are discussed below and our implementation is described in an Appendix, available by request to the authors.

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 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' predictions. However, it could complicate the statistical evaluation of the pricing errors, because the standard errors may be inconsistent in the presence of the autocorrelation. We address autocorrelation in the pricing errors using a block bootstrap approach as described below.

Consumption data are time-aggregated, meaning that the reported figures are the averages of the levels during the period. Time aggregation in consumption 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 (2010) 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 (5a) and (5b) can still identify the true conditional mean, x_{t-1}, and conditional volatility, σ_{t-1}. However, the error terms do not reveal the true consumption shocks.

The moving average structure may be addressed by the choice of weighting matrices used in estimation, and the use of block bootstrap methods in evaluating the test statistics. We conduct experiments with quarterly data, as described below, to assess biased consumption shocks. 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.

3.8. Statistical tests

We study the models' pricing errors, focusing on their time-series averages and the mean squared pricing errors (We also have examined the mean absolute deviations, which lead to similar inferences as the MSEs). To evaluate the statistical significance of the differences between the pricing errors for the different models, we use two approaches. The first approach follows Diebold and Mariano (1995). Let e_t be a vector of pricing errors for period-t returns, stacked up across K-1 models and let e_{0t} be the pricing error of the reference model. 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 pricing errors. For example, g(.) can be the mean squared error. Let $\underline{d} = \Sigma_t d_t / T$ be the sample mean of the difference. Diebold and Mariano (1995) show that $\sqrt{T}(\underline{d} - E(d_t))$ 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 $w(\tau)$ and number of lags depend on the context.

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 (or sample a block 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 or larger than the one we find in the actual data.

4. Data

We focus on annual data for several reasons. First, much of the evidence on long-run risk models uses annual consumption data. Second and related, annual consumption data are less affected by problems with seasonality (Ferson and Harvey, 1992) and other measurement errors. The annual consumption data are real nondurables plus real services expenditures from the National Income and Product Accounts, divided by the total U.S. population[10]. We also use not seasonally-adjusted quarterly consumption data, sampled annually, in experiments[11].

The asset return data in this study are standard. We use the returns of common stocks sorted according to market capitalization and book-to-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 losers and winners, also from French. The market portfolio proxy is the CRSP value-weighted stock return index.

When returns are measured in excess of a risk-free rate (the equity and term premium), we use a one-month Treasury bill return as the risk-free rate. We also use a risk-free rate as a state variable. To avoid using the same data on both sides of a regression, we use the lagged three-month Treasury yield as the state variable. In theory, we should use an ex ante real risk-free rate. We conduct experiments, described below, with an alternative proxy that involves inflation data. Inflation is measured using the U.S. Consumer Price Index from NIPA table 2.3.4.

We include bond returns that differ in maturity and credit quality. The short term bond return is the one-month Treasury bill, rolled over each month during the year. The long-term Government bond splices the Ibbotson Associates 20 year US Government bond return series for

---

[10] Nondurable goods expenditures are from NIPA Table 2.3.5, line 8 divided by the price deflator in Table 2.3.4, line 8. Services expenditures are from Table 2.3.5, line 13 divided by Table 2.3.4, line 13. The sum of the two is divided by the population in Table 2.1, line 39.

[11] We obtain not seasonally-adjusted quarterly consumption data from NIPA table 8.2, available during 1947-2004. Nondurable goods expenditures from line 6 of table 8.2 is divided by the price deflator in line 8 of Table 2.3.4. Services expenditures from line 13 of Table 8.2 is divided by the price deflator in line 13 of Table 2.3.4. The sum of the two is divided by the population in line 39 of Table 2.1 to get real per capital consumption.

1931-1971, with the CRSP greater than 120 month US Government bond return after 1971. High grade corporate bond returns are those of 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 1931 to 1990, with the Merrill Lynch High Yield US Master Index returns after that date. All of the returns are continuously-compounded annual returns, although we conduct experiments with arithmetic returns as well.

The second state variable is the log price/dividend ratio. In calculating the price/dividend ratio, the stock price is the real price of the S&P Composite Stock Price Index, and the dividend is the real dividends accruing to the Index. These data are from Robert Shiller's website. Given that stock buybacks represent a significant component of corporate payouts, and that total payout (dividends plus repurchases minus issuances) yield beats dividend yield in explaining returns (Boudoukh, Michaely, Richardson, and Roberts, 2007), we also use an alternative measure of the price/dividend ratio that accounts for share repurchases and equity issuances. We obtain these data from Michael Roberts' website.

Table 1 presents summary statistics of the basic asset return 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-to-market effect (Value-Growth), the excess returns of momentum winners over losers (Win-Lose), the excess returns of long-term reversal losers over winners (Reversal), 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

5.1. Average pricing errors

We use traditional rolling estimation with a 36-year rolling window, and the evaluation period is the subsequent year. (We also examine 24-year rolling windows to check sensitivity to the window length.) We compare the performance of the long-run risk models with three alternative models. The first is the market-portfolio based CAPM of Sharpe (1964). The second is a simple consumption-beta model following Rubinstein (1976) and Breeden, Gibbons, and Litzenberger (1989). 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 risk-free rate and log price/dividend ratio. We also include the restricted versions of the stationary (RLRRS) and cointegrated (RLRRC) long-run risk models described above.

Table 1 presents the average excess returns and average pricing errors from the rolling step-ahead analysis over the 1967-2009 period, and Table 2 presents the full-sample results for 1931-2009. The A panels use OLS risk premiums and the B Panels use GLS (the restricted model are presented in the OLS panels). Tables 1 and 2 include the results of Diebold-Mariano (1995) tests for the differences between the average pricing errors of the various models and those of the CAPM. Empirical p-values are presented (in parentheses) for two-sided tests of the null hypothesis that the model in question performs no differently than the CAPM.

The first important point to make is that the tests for equal average pricing errors have power to discriminate between models. In Table 1 there are 70 cases where a model is compared to the CAPM, and 51 cases produce empirical p-values for the null of equal pricing error that are below 5%. We don’t show the p-values in Table 2, but with the longer sample most of the differences are statistically significant.

The consumption-beta model (CCAPM) performs about as well as the classical CAPM in rolling estimation, with smaller average pricing errors on Momentum, Value-Growth and Reversals, but larger errors on the Size Effect and Credit Premium. However, in the full sample estimation the CCAPM performs much worse than the CAPM for all of the strategies excepting Momentum.

The long-run risk models under rolling estimation produce economically and statistically significantly smaller average pricing errors than the CAPM in most cases. The cointegrated model (LRRC) has the smallest overall errors, improving on the stationary model (LRRS) for the Equity Premium, Size Effect and Momentum. Zurek (2007) finds that a stationary long-run risk model explains about 70% of the momentum effect, with time-varying betas on the expected consumption growth shock. We get smaller percentages here with the unconditional models, but still an impressive performance in explaining the Momentum Premium.

The average pricing error of the LRRC model on the Credit Premium is relatively large. In the full sample both long run risk models overshoot the credit premium by several times its raw magnitude. The fit improves somewhat under GLS estimation, because GLS gives more weight to fitting the low volatility assets when estimating the risk premiums. However, the cost is degraded performance on the Equity Premium and Size Effect under GLS.

It should not be surprising that the long-run-risk models examined here leave large pricing errors for the Credit Premium. The original model of Bansal and Yaron (2004) is designed to capture a few key features of the equity markets. Subsequent work has introduced additional factors for credit risks. Chen (2010) and Chen, Collin-Dufresne, and Goldstein (2008) find that when countercyclical default probabilities (induced by time-varying volatility or asset values) and recovery rates in default are allowed to interact with countercyclical risk aversion, a long-run risk model can be calibrated to match credit spreads. Endogenizing the leverage choice and optimal default boundary, Bharma, Kuehn, and Strebulaev (2010) find that the dynamics of a cross-section of firms is important for calibrating to credit spreads. Perhaps the simple models examined here fail because they are missing these factors.

The poor performance on the Credit Premium and the strong performance on Momentum are an interesting combination. Avramov, Chordia, Jostova, and Phillipov (2007) find that the Momentum Effect is concentrated among firms with low credit ratings. This suggests that additional empirical work on models elaborated for credit risks may prove fruitful.

The long run risk models also fit poorly on the Term Premium under OLS, and similar concerns about the simplicity of the basic models arise here. In particular, inflation risks are likely important for bond pricing. See Eraker (2008), Bansal and Shaliastovich (2010) and Hasseltoft (2008) for long run risk models with stochastic inflation.

The long-run risk models do not fare as well in full-sample estimation as they do in the step-ahead analysis. Under OLS they still display the smallest overall average pricing errors, and the LRRC model does slightly better than the LRRS model. But the low overall averages under OLS include the large negative pricing errors for the Credit Premium, and the CAPM performs significantly better than the LRRS model for four of the seven assets.

It is interesting and perhaps unexpected that the long-run risk models' performance deteriorates under full-sample estimation. Holding the parameters fixed over a long time period could be an important misspecification that rolling estimation avoids. Alternatively, it could be that the long-run risk models perform better in the more recent period that is evaluated under the rolling method. We evaluate these alternatives below.

The 2-State-Variable model is examined in column 7 of tables 1 and 2. Under rolling estimation its average pricing error performance is the worst among all the preceding models, for almost all of the assets, and especially so with GLS risk premiums. In full-sample estimation the 2-State-Variable model performs slightly better, but is still dominated by the long-run risk models. This shows that the strong average pricing error performance of the long-run risk models is not solely driven by the two state variables. The consumption risk factors are important for its performance.

The improved performance of the 2-State-Variable Model in full-sample estimation is interesting in view of the degraded performance of the long-run risk models in the full sample. This suggests that using more data to estimate the part of the model that uncovers the state variable shocks – which involves highly persistent series – may be more of a benefit than allowing those coefficients to drift over time with rolling estimation.

The results for the restricted versions of the long-run risk models are in the far right columns of tables 1 and 2. In terms of average pricing errors the restricted models perform markedly worse than their reduced-form counterparts, and worse than the CAPM except on Momentum. Their performance is similar in the full sample and rolling estimation. Requiring the prices of covariance risk to match economically reasonable values of the utility function parameters pulls the risk premiums away from values that fit the average excess returns.[12] Of course, the poor performance of the restricted models may not be entirely surprising. Given data issues such as time aggregation and the imperfect proxies for objects in the model, like the risk-free rate and dividend-price ratio, forcing the restrictions on the risk premiums may be expected to hurt the fitted models' performance. We assess the effects on pricing error variances in the

[12] Tables 1 and 2 use continuously-compounded annual excess returns. We also run the analysis with simple arithmetic excess returns. None of the conclusions change, and the relative performance of the cointegrated long-run risk model slightly improves.

next section.

5.2. Mean squared pricing errors

We are interested in the precision of the models' predictions as well as the average errors. Tables 3 and 4 present the differences in the mean squared pricing errors (MSE) from those of the CAPM. The MSE errors for each model are presented as a percentage of the MSE errors of the CAPM, whose MSEs are shown in the left-most columns of the tables. The empirical p-value in parentheses tests the hypothesis that the MSE differences are zero. These are based on two-sided tests of the null hypothesis.

As described above, the MSE may be decomposed as the mean pricing error squared plus the error variance. We find that the error variance is the dominant component in almost every instance, excepting Momentum. When the CAPM confronts Momentum the mean error and the error variance contribute nearly equal shares to the MSE. Thus, with this exception we interpret the MSE results as indicating mainly the effects of parameter estimation error on the pricing accuracy.

Table 3 presents results using rolling estimation. In contrast to the tests in Table 1 there are fewer significant differences between the MSEs of the CAPM and the other models, but there are some significant differences. The long-run risk models outperform the CAPM in explaining Momentum on the basis of MSEs. All of the models, except the 2-State-Variable model under OLS, significantly outperform the CAPM on the Momentum effect. The 2-State-Variable model is significantly worse than the CAPM in explaining the Value-Growth premium and, under GLS, the Size Effect, Reversals and the Term Premium.

The fact that there are fewer significant differences in the MSEs, compared with the average pricing errors, is not simply a lack of statistical power. The differences in the MSEs are economically small. For example, the MSE differences for Momentum become statistically significant when they are about 20% of the CAPM value, which corresponds to an MSE difference of about 2% per year. For the Value-Growth Premium, an MSE difference of only 0.2% per year is statistically significant.

In the full sample estimation of Table 4 more of the MSE differences are statistically significant. All of the models beat the CAPM on the Momentum Effect, but the CAPM significantly outperforms the long-run risk models in explaining the Value-Growth effect and Reversals, on an MSE basis. In these cases, while the MSE differences are statistically significant they are economically small.

The right-hand columns of tables 3 and 4 present the MSE differences for the restricted long-run risk models. With a few exceptions the results are similar to, but slightly worse on average, than those for the models without the parameter restrictions. This reflects the fact the while the restrictions increase the bias and the average pricing errors, they improve estimation efficiency by reducing the number of parameters, and the error variance benefit shows up in the MSEs. In the case of Momentum the bias effect exerts itself again in the MSEs, and the restricted versions of the long-run risk models perform substantially worse than their unrestricted counterparts on Momentum.

It makes sense that the long-run risk models don't substantially outperform the CAPM on a MSE basis, with the exception of Momentum. With that exception, the MSEs are driven by the effects of the parameter estimation error variance. With more parameters to estimate than the CAPM, the long-run risk models' MSE performance is more impacted by estimation error. We address this issue with formal tests in the next section.

5.3. Testing nested models

The MSE results are relatively more favorable to the restricted long-run risk models and to simpler models like the CAPM and the CCAPM than are the average pricing errors. However, Clark and West (2007) show that a model with more parameters than a "true" model is expected to have larger MSEs out of sample because the noise in estimating the parameters that should really be set to zero will contribute to the error variance. Thus, the mean squared errors of the long-run risk models could be inflated relative to the simpler models they nest. Clark and West propose an adjustment to the difference between two models' mean squared pricing errors to account for this effect. Let E_{1t} be the predicted value from model 1 which is the more parsimonious model and MSE_1 is its mean squared pricing error. E_{2t} is the prediction from the more complex model 2 and MSE_2 is its mean squared pricing error. The test examines MSE_1 - [MSE_2 - $\Sigma_t(E_{1t} - E_{2t})^2/T$]. The third term adjusts for the difference in expected MSE under the null and a one-sided test is conducted. The null hypothesis is that the two models have the same pricing error and the alternative is that the more complex model has smaller pricing error. If we suppress the adjustment term, we obtain the Diebold-Mariano test as a special case.[13]

Table 5 presents Clark-West tests for the long-run risk models using rolling estimation, where the benchmark models are the CCAPM and the 2-State-Variable Model. (We don't use the restricted long-run risk models as benchmarks, because Table 3 shows that even without the adjustment they perform slightly worse than their unrestricted long-run risk counterparts.) When the t-statistics are negative it means that the more complex model performs worse than the

[13] 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 predicted. 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.

simpler model, even after the adjustment. When the 2-State-Variable model is the benchmark model (the C and D panels) the t-statistics are almost always positive, and the long-run risk models significantly outperform the 2-State-Variable model. The unrestricted models perform significantly better than the 2-State model on four to six of the seven assets and the restricted long-run risk models outperform on five of the seven assets. This is not a surprising result given our previous findings on the poor performance of the 2-State model, but it demonstrates that the Clark-West tests have statistical power.

In Table 5 panels A and B, the CCAPM is the benchmark model and the unrestricted long-run risk models never significantly outperform the CCAPM on an adjusted MSE basis. However, the restricted models perform substantially better than the CCAPM for four of the seven assets. Before the adjustment for the larger number of parameters the restricted LRR models do not seem to perform better than the CCAPM (Table 3). Thus, we can reject the simpler CCAPM in favor of the long-run risk models when we restrict the risk premiums to be consistent with reasonable utility function parameters and adjust for the larger number of parameters.

If we think of the long-run risk models as essentially combining elements of the CCAPM and the 2-State model, then Table 5 says that both aspects are significant for the MSE performance. The consumption side of the models in particular, is crucial for fitting the asset returns, as the long-run risk models substantially outperform the 2-State model that leaves out the consumption risk factors. The restricted versions of the LRR models outperform the CCAPM, showing that the long-run risk factors are also important.

5.4. Comparison of subperiods

In the previous sections we find differences in the results using full-sample versus rolling estimation. The differences could be related to in-sample versus step-ahead fit, or to the different evaluation periods. We run the rolling analysis again, removing data from the pre-World War II period and find that the long-run risk models perform even better. Thus, some of the improved performance of the long-run risk models in the rolling, step-ahead analysis likely reflects the different sample periods. But we would like to know if this is due to parameter estimation or if it reflects a better fit in different historical periods at the same parameter values. In this section we hold the model parameters fixed and vary the evaluation period. We use estimation over the full sample to obtain the parameters. This allows the models to base the parameter estimates on a long time series; but of course, the predictions would not be feasible in practice. Table 6 presents the results.

The format of Table 6 is similar to the previous tables, except that we compare four, approximately 20 year evaluation periods and show results for an equally-weighted portfolio of the test assets. All of the models perform at their worst in the first subperiod, 1931-50, by any measure. All of the models perform best during the two middle periods, 1951-70 and 1971-90, and their performance deteriorates again during 1991-2009. Thus, poor performance in the pre-World War II period is not unique to the long-run risk models, and all the models are challenged by recent economic times.

Confirming the importance of the subperiods, we conduct another experiment where we hold the evaluation period fixed, using the 1967-2009 period as in Table 1 and 3, and change the parameter values. We consider the full-sample parameter estimates, estimates using 1967-2009 data, and the rolling estimates of table 1 and 3. We find much less variation in the models' performance with the different parameter estimates than across the subperiods in Table 6 (Tables of these results are available by request).

Under OLS the average pricing error performance of the unrestricted long-run risk model in Panel A is dramatically better during the two middle periods. While the CAPM also performs better during these periods, the differences are much less dramatic. This suggests that the improved relative average pricing error performance of the long-run risk models under rolling estimation is likely attributable to the different evaluation periods.

The restricted long-run risk models are the worst performing models in each subperiod and experience a relatively small improvement in average pricing errors during the middle periods. We saw above that their relative performance in rolling estimation was not improved, compared with full-sample estimation. Imposing the parameter restrictions destroys the long-run risk models' ability to fit the average returns so well during the 1951-1990 period.

Panel B summarizes the MSEs under OLS estimation. The MSEs of the CAPM are dramatically larger during the first subperiod, 1931-50. During this period all of the models outperform the CAPM on an MSE basis, except for the restricted long-run risk models. This seems different from the previous tables, where except for Momentum the CAPM performs about as well as the LRR models on a MSE basis. While the average pricing error performance of all the models deteriorates during the most recent twenty year period, the MSEs present a different pattern. The CAPM MSEs increase slightly compared with the prior 40 year period, but the long-run risk models' MSEs drop to about 60% of those of the CAPM, from larger fractions in the previous periods.

The poor MSE performance of the CAPM in Table 6 reflects the impact of Momentum on the equally-weighted portfolio. Momentum is a high volatility return, and the average pricing error of the CAPM on momentum is very large, as we observed above. The combined effects have a large influence on the MSE of the portfolio.

In Panels C and D of Table 6 we summarize the results across the assets using a "GLS portfolio." GLS estimation transforms the TxN data matrix of excess returns, r, using $rV(r)^{-1/2}$, where $V(r)$ is the full sample covariance matrix. This is scaled to a set of seven portfolio weights that sum to 1.0.[14] The GLS portfolio gives a weight of only 12.3% to the Momentum strategy and the results are very different.

With the relatively small weight that the GLS portfolio applies to Momentum, and the larger weights on the Value-Growth, Credit and Term premiums, the average pricing error results appear more favorable to the CAPM. On an average pricing error basis the CAPM is consistently the second best model, underforming only the unrestricted cointegrated model. The relative MSE performance across the models is now fairly stable over the subperiods.

5.5 Outliers

The poor performance of the models in the first twenty-year subperiod of Table 6 suggests the influence of outliers during the Great Depression of the 1930s. We examine the absolute pricing errors that exceed three standard errors, for each asset-model combination. The year 1931 is an outlier for the Equity Premium under four of the models. The year 1933 is an outlier for Reversals under all seven models. Momentum and the Credit Premium generate outliers in those years and in the years 1937, 1973, 2008, and 2009. The year 1982 is an outlier for the Term Premium under the CCAPM. Removing these outlier years results in a larger

[14] The N-vector of weights is $rV(r)^{-1/2}(\underline{1}/rV(r)^{-1/2}\underline{1})$, where $\underline{1}$ is an N-vector of ones and / indicates element-by-element division. Listed in the same order as the assets in Table 1, the weight vector takes the values (0.014, 0.002, 0.207, 0.123, 0.269, 0.255, 0.129)'.

average pricing error for all the models during the first subperiod, where the outliers are negative. The MSE for the CAPM is reduced from 2.46% to 1.71%. The long-run risk model MSEs improve, changing from 84-86% of the CAPM's MSEs to 70-72% when the outliers are removed. The restricted versions of the models perform slightly worse than before.

There are no outliers in the second subperiod. In the third subperiod removing 1973 and 1982 lowers the average pricing errors by 30-60 basis points for all of the models, and the CAPM's MSE drops slightly to 0.59%, but the relative MSE performance of the models is unchanged. A similar effect is observed when 2008 and 2009 are removed from the fourth subperiod.

5.6. Conditional models

Conditional models arguably better reflect the economic structure of long-run risk models. For example in the models market yields, expected consumption, consumption variances and other quantities are affine functions of the state variables and thus vary over time. In our estimation using equation systems (5) and (7), everything above the last two equations is conditioned on the lagged risk-free rate and dividend yield. In this section we explore conditional versions of the models. A priori, conditional models may empirically perform better or worse than unconditional models. While conditional models use more information to generate required returns, Ghysels (1998) and Simin (2008) find that conditional beta pricing models do not outperform simpler models in an out-of-sample context.

The conditional models assume that the expected risk premiums are affine functions of the conditioning variables, but the betas are held fixed, consistent with Gibbons and Ferson (1985) and Campbell (1987). Constant betas imply zero covariances between betas and risk premiums, and that could be important for capturing a cross section of average portfolio returns sorted according to the Size Effect (e.g. Chan and Chen, 1988) or the Value-Growth Effect (e.g. Santos and Veronesi, 2006). However, Ferson and Harvey (1991) show that time-varying risk premiums are much more important than time-varying betas for capturing variation in expected returns over time. In rolling estimation the betas are allowed to vary over time in an unstructured way.

To estimate the conditional models we replace the specification for λ in Eqs. (5f) and (7f) with affine functions of the state variables: the lagged dividend yield and risk-free rate in (5f) and including the lagged cointegrating residual in (7f). The error terms are orthogonal to a constant and the lagged predictors. The fitted values of the risk premiums are evaluated at the most recently available values of the lagged predictors when forming the fitted expected excess returns. The same procedure is used for the CCAPM, the 2-State model and the CAPM.

Recall that in the CAPM the GLS risk premium estimate is simply the expected market portfolio excess return. In the GLS case we use the predicted value from a regression of the market excess return on the lagged risk-free rate and dividend yield as the market risk premium estimator. For the conditional versions of the restricted long-run risk models we impose that the error terms u_{6t} in (7f) and (9) are orthogonal to the same lagged variables. This implicitly allows time-varying conditional covariances without restricting their functional forms.

The average pricing error results for the conditional models are summarized in Table 7. The CCAPM performs better in its conditional form on most of the asset returns and by almost 50 basis points overall. For the other models the comparison is mixed across the assets and the overall results are similar to the models with unconditional risk premiums. Conditioning affects the pricing errors of the restricted cointegrated long-run risk model more than the stationary model, but the impact differs across the assets with no clear pattern.

We also examine the MSE pricing errors of the conditional models. The CAPM performs better on Momentum in its conditional form. The comparisons between the other models and the CAPM are broadly similar to those for the unconditional models. It should not be surprising that our main results are robust to the conditional model extensions. Rolling estimation of the "unconditional" models in Table 1 allows ad-hoc time variation in the parameters, and we find that the rolling and conditional model risk premium estimates are highly correlated. For example, the correlation between the rolling estimates in Table 1 with those in Table 7 for the short-run and long-run consumption risk premiums are 0.89 and 0.79, respectively.

In summary, many of our findings are robust to the use of conditional models. The long-run risk models perform better than the CAPM in terms of average pricing errors but not in terms of the MSEs. The cointegrated long-run risk model outperforms the stationary version and does especially well on Momentum. GLS estimation improves the fit to the low volatility returns at the expense of the higher volatility returns. The 2-State-Variable model performs significantly worse than the long-run risk models, indicating that the long-run risk factors are important ingredients for the models' out-of-sample fit.

5.7. Return Decompositions

To gain more insight into the economics behind the model evaluations, we examine the results of the rolling estimation in more detail. In particular, we study the rolling betas and risk prices associated with the various factors in the models. Previous studies provide similar analyses based on calibrated models and in-sample estimation.[15] Our main goal is to study the out-of-sample performance, so we do not present the details of these exercises, but summarize here the

[15] See, for example, Bansal, Dittmar, and Lundblad (2005) and Bansal, Dittmar, and Kiku (2009).

main observations (tables of these results are available by request).

In general, we find that both the short and long-run consumption factors carry positive risk prices. The equity-based portfolios all feature positive consumption betas, and Momentum has a relatively large consumption beta. Both short and long-run consumption risks contribute substantial positive amounts to the winner-loser portfolio in the LRRS model, and the sum of the consumption-related shocks determines more than 50% of the fitted momentum return in the LRRC model. The size effect and Value-Growth portfolios feature positive short and long-run consumption betas in the LRRS model and positive short-run consumption risk betas in the LRRC model.

Cointegration typically appears to improve the model specification, leading to more economically appealing decompositions, consistent with Bansal, Dittmar, and Kiku (2009) and our out-of-sample results. Comparing the LRRS and LRRC models also leads to an interesting characterization of the Small-Big, Value-Growth and Reversal strategies. These are all in some sense, “contrarian” strategies. In the LRRS model, volatility risk appears important for their determination. However, introducing the cointegrating residual factor in the LRRC model reveals that these strategies feature a negative exposure to shocks that raise dividends above their cointegrating trend. Such shocks are “good news”, as indicated by a positive estimated risk premium.

6. Robustness

This section describes a number of experiments to evaluate the robustness of our main results. Additional tables are available in an unpublished appendix, by request to the authors.

6.1. Autocorrelation

We conduct experiments to assess the impact of autocorrelation in the pricing errors on the statistical tests. The first-order autocorrelations of the pricing errors are similar across the models and usually smaller than 10%, except in a couple of cases. With a sample of 79, the approximate standard error of the autocorrelations is about 11%. Autocorrelated pricing errors reflect some persistent misspecification of the models. Time aggregation of the consumption data is one likely source of such misspecification. In the case of the size effect, some of the autocorrelation (0.41, the largest example) is likely spurious, resulting from stale pricing of the smallest firms.

We repeat some of the tests in tables 1-4 using a block bootstrap approach with block lengths of 2 and 4. A block size of two accommodates an MA(1) error structure, as would be implied by time aggregation. We find little effect on the empirical p-values in tables 2 and 4. The effect is greater in tables 1 and 3 with the shorter, rolling samples. The p-values for the average pricing error differences get larger, and five cases that were significant at the 5% level in Panel A (and the same number in Panel B) of Table 1 are no longer significant at the 10% level. The restricted cointegrated model seems to be the most affected. In Table 3 the p-values get smaller when the block bootstrap is used. Five MSE differences that were not significant at the 10% level in this table become significant under the block bootstrap with 4 lags.

6.2. Earnings and price momentum

Chordia and Shivakumar (2006) attribute much of the *Price* Momentum Effect that we examine above (PMOM) to earnings momentum, which is significantly related to various economic risk variables. This raises the possibility that the explanatory power of the long-run risk models for PMOM works through earnings momentum (EMOM). Following their approach we sort stocks into EMOM portfolios on the basis of the most recently-reported standardized unexpected earnings (SUEs). They use SUE deciles but don't cross-sort on PMOM; we cross-sort on PMOM and EMOM.

Like Chordia and Shivakumar (2006) we find that the average EMOM premium is larger than the PMOM premium. EMOM largely subsumes PMOM, but EMOM is not subsumed by PMOM in the two-way sorts, which also reveal that EMOM is stronger among momentum winners than among momentum losers.

If a model works primarily through EMOM, we would expect that it would perform well on EMOM, but not as well on any remaining PMOM effect when EMOM is controlled for. We focus on the LRRC model and find that this does not appear to be the case. The average pricing errors are larger fractions of the original effects for PMOM, within an EMOM sort, than they are for EMOM. Thus, we find no evidence that the explanatory power of the LRRC model for Momentum is driven by earnings momentum.

6.3. Time aggregation

The consumption data are time aggregated, which leads to a spurious moving average component in consumption growth, biased consumption betas and biased shocks in the consumption-related risk factors. The latter issue is the most difficult to address. We 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. We use quarterly, not seasonally adjusted data for 1948-2004. By sampling the not seasonally adjusted data in the last quarter of each year we avoid the strong seasonality in consumption expenditures and the smoothing biases in data that are seasonally adjusted by the Commerce Department (see Ferson and Harvey, 1992).

In terms of average pricing errors, the performance of the consumption beta model is improved by the use of the higher-frequency consumption data. The relative performance of the classical CAPM and the long-run risk models is similar to what we observed before, in terms of average pricing errors. The MSEs show the CAPM and the cointegrated long-run risk model delivering the best and similar performances. Thus, the main findings are robust to the alternative consumption data.

6.4. 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; Malloy, Moskowitz, and Vissing-Jorgensen, 2009). We examine two-year and three-year returns, compounding the returns but maintaining the assumption that the decision interval in the model is one year. The CCAPM and 2-State-Variable model deliver the largest pricing errors, similar in size to the raw returns. The cointegrated long-run risk model delivers the smallest errors for two of the seven assets and in no case does it perform the worst. The advantage of the CAPM under GLS is not apparent here. Two-year returns produce similar results.

We also examine the MSEs. Under OLS risk premiums, the consumption beta model performs better, in relative terms, on the longer-horizon returns. The cointegrated model generally outperforms the stationary long-run risk model. Using GLS risk premiums, the CAPM wins for four to five out of seven assets on a MSE basis. Thus, overall the conclusions are similar to those from the previous tables.

6.5. The effects of bias correction

We examine models in which we apply versions of the bias correction procedure developed by Amihud, Hurvich, and Wang (2009) to the coefficients on the highly persistent variables. In the stationary version of the long-run risk model the bias corrections have a trivial impact, and the average pricing errors are typically identical to one basis point per year. In the cointegrated model the bias corrections have a larger impact. For example, under OLS the average pricing errors are smaller by as much as 0.4-0.5% per year for four of the seven assets. Thus, bias-corrected estimation appears to be useful, and we recommend that future work estimating cointegrated long-run risk models apply bias corrections.

6.6. Sensitivity to the number of test assets

The risk premiums are estimated using the seven test assets, but the maximum correlation portfolio in a larger set of assets delivers different risk premiums. With fewer assets we likely "overfit" the risk premiums. (Intuitively, with the same number of assets as risk premiums the unrestricted models reduce to using the sample mean of each asset to predict the expected return.) We increase the number of assets to 31 by adding Size deciles 2 to 9, Book-to-Market deciles 2 to 9, and Momentum deciles 2 to 9. As expected, the pricing errors on the original seven assets get larger for all the models except the CAPM, although only slightly larger for the consumption CAPM, but the differences are not large enough to change the previous impressions.

6.7. Alternative price/dividend measures

The "true" price/dividend ratio envisioned by the long-run risk models is not observable, and the standard measures use trailing corporate dividends to smooth out seasonality. The use of annual data helps mitigate some of the concerns about seasonality, but the "true" aggregate dividend is hard to measure. Given that share repurchases represent an important component of payouts (Boudoukh, Michaely, Richardson, and Roberts, 2007), we examine an alternative price/dividend measure that accounts for share repurchases and equity issuances. Using this alternative price/dividend measure, the models' pricing errors on average get larger with the exception of the 2-State-Variable model, but the changes do not substantially alter the previous impressions. The improved performance of the 2-State-Variable model likely reflects the improved predictive ability of the alternative price/dividend measure.

6.8. Alternative risk-free rate measures

Our main results use a nominal, three-month Treasury rate as the risk-free rate state variable. This is not the real, ex ante yield envisioned in the models. Measuring returns in excess of the risk-free rate reduces or eliminates concerns about real versus nominal returns, but not about the risk-free rate used as a state variable. We conduct experiments to evaluate the sensitivity of our findings to the state variable. Following Bansal, Kiku, and Yaron (2009), a real risk-free rate is estimated as the annualized yield on the 3-month treasury bill minus the trailing 12-month realized annual inflation. Overall, the results using this proxy are similar to the results using the three-month treasury yield as the state variable.

6.9. Alternative variance identification

An alternative version of the LRRS model identifies the stochastic volatility from the squared residuals of the expected consumption, or fitted x_t process instead of the unexpected consumption growth. Under OLS the alternative variance model performs slightly worse on the Size and Value-Growth effects and on the Term Premium, but slightly better on the rest and the overall average pricing error is identical. The MSE's are virtually identical to the previous version of the model. Under GLS the alternative variance model performs worse on the Equity Premium, Momentum, the Credit and Term Premium but much better on the Size Effect. The MSEs of the alternative variance model are slightly smaller overall under GLS estimation.

7. Conclusions

This study examines the ability of long-run risk models to fit out-of-sample returns. The question of the out-of-sample fit is important because asset pricing models are almost always used in practice in an out-of-sample context. We examine stationary and cointegrated versions of long-run risk models using annual data for 1931-2009.

The long-run risk models perform well in capturing the Momentum Effect. The performance of the cointegrated version of the long-run risk model is markedly better than the stationary version. While the long-run risk models often deliver smaller average pricing errors than the CAPM, the mean squared pricing errors (MSEs) are not significantly better in economic terms, except in explaining momentum. However, the long-run risk models do perform significantly better than models that suppress their consumption-related shocks. Holding the parameters fixed, the models perform better during 1951-1990 than in the decades before or after. Sampling not seasonally adjusted consumption data in the last quarter of the year improves the fit, consistent with Jagannathan and Wang (2007). Bias corrected regressions improve the performance of cointegrated versions of the models but have little effect on stationary versions. Restricting the models with economically reasonable preference parameters increases the average pricing errors, but often improves the error variances.

Our results suggest several areas for future research. Cointegrated versions of the long-run risk models have received less attention in the literature than stationary versions. Our results suggest that cointegrated models deserve more attention. We examine only the most basic versions of the long-run risk models, and they perform relatively poorly in explaining credit spread returns and excess long-term bond returns. Recent papers have refined the models, incorporating inflation and credit risk dynamics to achieve success in calibrating to term premiums and credit spreads. More empirical work on these refined models is clearly warranted. For example, the simple long-run risk models turn in a strong performance in explaining Momentum. Avramov, Chordia, Jostova, and Philipov (2007) find that momentum is concentrated in low credit risk stocks. This suggests that refined versions of long run risk models with credit risks should perform well in empirical tests. In our view, the practical importance of long-run risk models is yet to be fully established, and will depend on their performance in future empirical estimation and tests.

References

Amihud, Y., Hurvich, C., Wang, Y., 2009. Multiple-predictor regressions: hypothesis testing. Review of Financial Studies 22, 413-434.

Avramov, D., Chordia, T., Jostova, G., Philipov, A., 2007. Momentum and credit rating. Journal of Finance 62, 2503-2520.

Balduzzi, P., Robotti, C., 2008. Mimicking portfolios, economic risk premia, and tests of multi-beta models. Journal of Business and Economic Statistics 26, 354-368.

Bansal, R., 2007. Long run risks and financial markets. The Review of the St. Louis Federal Reserve Bank 89, 1-17.

Bansal, R., Dittmar, R., Kiku, D., 2009. Cointegration and consumption risks in asset returns. Review of Financial Studies 22, 1343-1375.

Bansal, R., Dittmar, R., Lundblad, C., 2005. Consumption, dividends, and the cross section of equity returns. Journal of Finance 60, 1639-1672.

Bansal, R., Gallant, R., Tauchen, G., 2007. Rational pessimism, rational exuberance and asset pricing models. Review of Financial Studies 74, 1005-1033.

Bansal, R., Shaliastovich, I., 2010. A long-run risks explanation of predictability puzzles in bond and currency markets. Unpublished working paper. Duke University.

Bansal, R., Kiku, D., Yaron, A., 2009. An empirical evaluation of the long-run risks models for asset prices. NBER working paper series.

Bansal, R., Kiku, D., Yaron, A., 2010. Risks for the long run: Estimation and inference. Unpublished working paper. Duke University.

Bansal, R., Yaron, A., 2004. Risks for the long run: A potential resolution of asset pricing puzzles. Journal of Finance 59, 1481-1509.

Bharma, H., Kuehn, L., Stebulaev, I., 2010. The levered equity risk premium and credit spreads: a unified framework. Review of Financial Studies 23, 645-703.

Blume, M., Keim, D., Patel, S., 1991. Returns and volatility of low-grade bonds 1977-1989. Journal of Finance 46, 49-74.

Boudoukh, J., Michaely, R., Richardson, M. P., Roberts, M. R., 2007. On the importance of measuring payout yield: implications for empirical asset pricing. Journal of Finance 62, 877-915.

Breeden, D., Gibbons, M., Litzenberger, R., 1989. Empirical tests of the consumption CAPM. Journal of Finance 44, 231-262.

Campbell, J., 1987. Stock returns and the term structure. Journal of Financial Economics 18, 373- 399.

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

Chan, K. C., Chen, N., 1988. An unconditional asset-pricing test and the role of firm size as an instrumental variable for risk. Journal of Finance 43, 309-325.

Chen, H., 2010. Macroeconomic conditions and the puzzles of credit spreads and capital structure. Journal of Finance 65, 2171-2212.

Chen, L., Collin-Dufresne, P., Goldstein, R., 2009. On the relation between the credit spread puzzle and the equity premium puzzle. Review of Financial Studies 22, 3367-3409.

Chordia, T., Shivakumar, L., 2006. Earnings and price momentum. Journal of Financial Economic 80, 627-656.

Clark, T., West, K., 2007. Approximately normal tests for equal predictive accuracy in nested models. Journal of Econometrics 138, 291-311.

Constantinides, G., Ghosh, A., 2012. Asset pricing tests with long-run risks in consumption growth. Review of Asset Pricing Studies 1, 96-136.

Daniel, K., Marshall, D., 1997. The equity premiums puzzle and the risk-free rate puzzle at long horizons. Macroeconomic Dynamics 1, 452-484.

Diebold, F., Mariano, R., 1995. Comparing predictive accuracy. Journal of Business and Economic Statistics 13, 253-265.

Eraker, B., 2008. Affine general equilibrium models. Management Science 54, 2068-2080.

Fama, E., French, K., 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33, 3-56.

Ferson, W., Constantinides, G., 1991. Habit persistence and durability in aggregate consumption: empirical tests. Journal of Financial Economics 29, 199-240.

Ferson, W., Harvey, C., 1991. The variation of economic risk premiums. Journal of Political Economy 99, 385-415.

Ferson, W., Harvey, C., 1992. Seasonality and consumption-based asset pricing. Journal of Finance 47, 511-552.

Ferson, W., Harvey, C., 1999. Conditioning variables and cross-section of stock returns. Journal of Finance 54, 1325-1360.

Ferson, W., Sarkissian, S., Simin,T., 2003. Spurious regressions in financial economics? Journal of Finance 58, 1393-1414.

Gibbons, M., Ferson, W., 1985. Testing asset pricing models with changing expectations and an unobservable market portfolio. Journal of Financial Economics 14, 216-236.

Ghysels, E., 1998. On stable factor structures in the pricing of risk: do time-varying betas help or hurt? Journal of Finance 53, 549-573.

Jagannathan, R., Wang, Y., 2007. Lazy investors, discretionary consumption and the cross-section of equity return. Journal of Finance 62, 1623-1661.

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

Hasseltoft, H., 2011. Stocks, bonds and long-run consumption risks. Journal of Financial and Quantitative Analysis, forthcoming.

Hansen, L., 1982. Large sample properties of generalized method of moments estimators. Econometrica 50, 1029-1054.

Huberman, G., Kandel, S., Stambaugh, R., 1987. Mimicking portfolios and exact arbitrage pricing. Journal of Finance 42, 1-9.

Lettau, M., Ludvigson, S., 2001. Consumption, aggregate wealth and expected stock returns. Journal of Finance 56, 815-849.

Malloy, C., Moskowitz, T., Vissing-Jorgensen, A., 2009. Long-run stockholder consumption risk and asset returns. Journal of Finance 64, 2427-2479.

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

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

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

Santos, T., Veronesi, P., 2006. Labor income and predictable stock returns. Review of Financial Studies 19, 1-44.

Shanken, J., 1992. On the estimation of beta-pricing models. Review of Financial Studies 5, 1-33.

Sharpe, W., 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. Journal of Finance 19, 425-442.

Simin, T., 2008. The (poor) predictive performance of asset pricing models. Journal of Financial and Quantitative Analysis 43, 355-380.

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

Stock, J., 1987. Asymptotic properties of least squares estimators of cointegrating vectors. Econometrica 55, 1035-1056.

Working, H., 1960. Note on the correlation of first differences of averages in a random chain. Econometrica 28, 916-918.

Zurek, P., 2007. Momentum and long-run risks. Unpublished working paper. Wharton.

Table 1: Average Out-of-Sample Pricing Errors

Average excess returns and out-of-sample pricing errors (actual minus prediction) are shown based on rolling estimation during 1967-2009, using a window length of 36 years. The continuously-compounded returns are in annual percent, and measured in excess of a one-month Treasury bill return. CCAPM is the simple consumption beta model. CAPM is the Capital Asset Pricing Model. LRRC is the cointegrated long-run risk model and LRRS is the stationary long-run risk model. 2-State-Variable is the model with only shocks to the two state variables (the risk-free rate and log price/dividend ratio). RLRRC is restricted version of the cointegrated long-run risk model and RLRRS is restricted version of the stationary long-run risk model. The p-values in parentheses are for two-sided Diebold-Mariano (1995) test of the null hypothesis that the model in question performs no differently than the classical CAPM. The empirical p-values are based on 10,000 simulation trials. p-values of 10% or less are in bold face.

Premium:	Mean Excess	Average Pricing Errors						
	Returns	CCAPM	CAPM	LRRC	LRRS	2-State-Variable	RLRRC	RLRRS
Panel A: OLS Risk Premiums								
Equity Premium	3.88	2.45	2.48	0.76	0.88	3.22	4.01	12.20
		(0.97)		**(0.00)**	**(0.01)**	(0.21)	**(0.00)**	**(0.08)**
Small-Big	2.50	2.07	1.21	0.22	2.06	2.64	2.65	3.22
		(0.26)		**(0.07)**	(0.18)	**(0.00)**	**(0.00)**	**(0.03)**
Value-Growth	6.47	1.64	5.51	3.31	1.65	5.95	6.35	5.13
		(0.00)		**(0.00)**	**(0.00)**	**(0.04)**	**(0.00)**	(0.67)
Win-Lose	17.95	12.10	18.30	7.66	9.58	16.97	17.80	12.20
		(0.00)		**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.08)**
Reversal	5.39	0.92	5.14	-1.46	-0.02	4.68	5.39	4.66
		(0.00)		**(0.00)**	**(0.00)**	(0.14)	**(0.08)**	(0.56)
Credit Premium	0.58	0.81	0.09	2.43	0.80	0.16	0.59	1.45
		(0.03)		**(0.00)**	(0.12)	(0.70)	**(0.00)**	**(0.08)**
Term Premium	1.82	1.22	1.86	0.02	-0.01	1.50	1.81	4.55
		(0.00)		**(0.00)**	**(0.00)**	(0.15)	**(0.05)**	(0.62)
Average	5.51	3.03	4.94	1.85	2.13	5.02	5.51	6.20
Panel B: GLS Risk Premiums								
Equity Premium	3.88	2.12	-2.14	1.52	1.71	4.40	n.a.	n.a.
		(0.00)		**(0.00)**	**(0.00)**	**(0.00)**		
Small-Big	2.50	1.55	-1.05	1.24	3.11	7.18		
		(0.00)		**(0.01)**	**(0.00)**	**(0.00)**		
Value-Growth	6.47	1.40	5.16	3.00	2.04	8.48		
		(0.00)		**(0.01)**	**(0.00)**	**(0.00)**		
Win-Lose	17.95	12.10	19.80	7.52	10.11	13.40		
		(0.00)		**(0.00)**	**(0.00)**	**(0.00)**		
Reversal	5.39	0.04	5.39	-0.95	0.22	8.24		
		(0.00)		**(0.00)**	**(0.00)**	**(0.00)**		
Credit Premium	0.58	0.58	-0.69	0.79	0.53	0.87		
		(0.00)		**(0.00)**	**(0.00)**	**(0.00)**		
Term Premium	1.82	1.32	1.36	0.73	1.11	1.83		
		(0.85)		**(0.00)**	(0.36)	**(0.01)**		
Average	5.51	2.73	3.98	1.98	2.69	6.34		

Table 2: Full Sample Average Pricing Errors, 1931-2009

Average excess returns and out-of-sample pricing errors (actual minus prediction) are shown based full sample estimation during 1931-2009. The number of observations is 79. The continuously-compounded returns are in annual percent, and measured in excess of a one-month Treasury bill return. CCAPM is the simple consumption beta model. CAPM is the Capital Asset Pricing Model. LRRC is the cointegrated long-run risk model and LRRS is the stationary long-run risk model. 2-State-Variable is the model with only shocks to the two state variables (the risk-free rate and log price/dividend ratio). RLRRC is restricted version of the cointegrated long-run risk model and RLRRS is restricted version of the stationary long-run risk model.

Premium:	Mean Excess Returns	Average Pricing Errors CCAPM	CAPM	LRRC	LRRS	2-State-Variable	RLRRC	RLRRS
Panel A: OLS Risk Premiums								
Equity Premium	5.87	2.91	3.73	-3.15	-2.31	1.86	6.03	5.73
Small-Big	4.65	5.93	3.67	0.83	0.03	3.94	5.00	4.65
Value-Growth	4.51	5.33	3.84	4.95	5.92	8.66	4.87	4.54
Win-Lose	15.69	7.71	16.60	7.10	6.89	11.80	15.10	15.40
Reversal	5.28	10.20	5.16	6.08	6.35	4.13	5.79	5.43
Credit Premium	1.14	-2.14	0.27	-4.22	-3.93	2.92	1.18	1.00
Term Premium	1.69	2.12	1.61	-1.76	-1.90	-3.70	1.82	1.69
Average	5.55	4.58	4.98	1.40	1.58	4.23	5.68	5.49
Panel B: GLS Risk Premiums								
Equity Premium	5.87	2.84	0.00	-0.46	1.36	4.70	n.a.	n.a.
Small-Big	4.65	5.96	1.95	6.37	6.49	10.90		
Value-Growth	4.51	5.35	2.67	5.53	7.81	9.86		
Win-Lose	15.69	7.52	18.20	6.17	4.86	6.54		
Reversal	5.28	10.30	4.95	9.27	10.40	11.00		
Credit Premium	1.14	-2.22	-1.24	-1.85	-0.98	3.43		
Term Premium	1.69	2.13	1.47	-0.77	-0.45	-0.67		
Average	5.55	4.55	4.00	3.47	4.21	6.54		

Table 3: Out-of-Sample MSEs Relative to the CAPM (%)

MSEs relative to the CAPM (%) are presented based on rolling estimation during 1967-2009, using a window length of 36 years. The relative MSE pricing error is calculated as 100(1+ ln(Model MSE /CAPM MSE)). The continuously-compounded returns are in annual percent, and measured in excess of a one-month Treasury bill return. CCAPM is the simple consumption beta model. CAPM is the Capital Asset Pricing Model. LRRC is the cointegrated long-run risk model and LRRS is the stationary long-run risk model. 2-State-Variable is the model with only shocks to the two state variables (the risk-free rate and log price/dividend ratio). RLRRC is restricted version of the cointegrated long-run risk model and RLRRS is restricted version of the stationary long-run risk model. The p-values in parentheses are for two-sided Diebold-Mariano (1995) tests of the null hypothesis that the model in question performs no differently than the classical CAPM. The empirical p-values are based on 10,000 simulation trials. p-values of 10% or less are in bold face.

Premium:		MSE relative to CAPM (%)					
	CAPM MSEs	CCAPM	LRRC	LRRS	2-State-Variable	RLRRC	RLRRS
Panel A: MSE relative to CAPM (%), OLS							
Equity Premium	3.88	91.8	104.0	95.0	99.5	97.8	308.0
		(0.30)	(0.68)	(0.54)	(0.95)	(0.64)	(0.48)
Small-Big	5.21	102.0	100.0	100.0	102.0	100.0	106.0
		(0.75)	(0.92)	(0.98)	(0.61)	(1.00)	(0.42)
Value-Growth	4.13	87.1	99.4	97.5	105.0	100.0	101.0
		(0.16)	(0.90)	(0.75)	**(0.01)**	(0.95)	(0.80)
Win-Lose	9.78	78.9	76.7	82.5	97.5	96.9	194.0
		(0.00)	**(0.04)**	**(0.04)**	(0.32)	**(0.00)**	(0.63)
Reversal	4.43	105.0	104.0	106.0	105.0	101.0	98.9
		(0.59)	(0.67)	(0.48)	**(0.06)**	(0.46)	(0.70)
Credit Premium	1.34	93.0	105.0	107.0	103.0	97.0	127.0
		(0.31)	(0.64)	(0.44)	(0.11)	(0.18)	(0.58)
Term Premium	1.02	93.7	97.0	103.0	103.0	99.6	229.0
		(0.14)	(0.66)	(0.72)	(0.55)	(0.39)	(0.62)
Panel B: MSE relative to CAPM (%), GLS							
Equity Premium	3.88	92.9	107.0	92.4	101.0	n.a.	n.a.
		(0.44)	(0.44)	(0.44)	(0.91)		
Small-Big	5.28	93.4	106.0	94.1	117.0		
		(0.40)	(0.50)	(0.63)	**(0.10)**		
Value-Growth	4.09	85.1	108.0	103.0	121.0		
		(0.17)	(0.37)	(0.76)	**(0.01)**		
Win-Lose	10.20	78.0	70.3	77.4	84.3		
		(0.00)	**(0.02)**	**(0.01)**	**(0.03)**		
Reversal	4.46	104.0	115.0	110.0	120.0		
		(0.73)	(0.22)	(0.37)	**(0.00)**		
Credit Premium	1.33	94.0	103.0	104.0	105.0		
		(0.31)	(0.71)	(0.58)	(0.19)		
Term Premium	0.99	98.1	101.0	105.0	108.0		
		(0.61)	(0.86)	(0.23)	**(0.01)**		

Table 4: Full Sample MSEs Relative to the CAPM (%), 1931-2009

MSEs relative to the CAPM (%) are based on full sample estimation during 1931-2009. The relative MSEs for a model is calculated as 100(1+ln(Model MSE/MSE CAPM)). The number of observations is 79. The continuously-compounded returns are in annual percent, and measured in excess of a one-month Treasury bill return. CCAPM is the simple consumption beta model. CAPM is the Capital Asset Pricing Model. LRRC is the cointegrated long-run risk model and LRRS is the stationary long-run risk model. 2-State-Variable is the model with only shocks to the two state variables (the risk-free rate and log price/dividend ratio). RLRRC is restricted version of the cointegrated long-run risk model and RLRRS is restricted version of the stationary long-run risk model. The p-values in parentheses are for two-sided Diebold-Mariano (1995) tests of the null hypothesis that the model performs no differently than the classical CAPM. The empirical p-values are based on 10,000 simulation trials. p-values of 10% or less are in bold face.

Premium:		MSE relative to CAPM (%)					
	CAPM MSEs	CCAPM	LRRC	LRRS	2-State-Variable	RLRRC	RLRRS
Panel A: MSE relative to CAPM (%), OLS							
Equity Premium	4.30	98.7	99.1	98.0	97.6	105.0	104.0
		(0.15)	(0.90)	(0.76)	(0.23)	**(0.03)**	**(0.04)**
Small-Big	5.41	104.0	97.6	97.5	100.0	102.0	102.0
		(0.07)	(0.39)	(0.48)	(0.13)	**(0.09)**	(0.11)
Value-Growth	4.38	103.0	102.0	105.0	113.0	102.0	101.0
		(0.04)	**(0.06)**	**(0.04)**	**(0.01)**	**(0.06)**	**(0.07)**
Win-Lose	7.94	68.2	66.6	66.1	81.1	93.8	95.1
		(0.00)	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**
Reversal	5.13	114.0	102.0	103.0	98.1	101.0	101.0
		(0.00)	**(0.03)**	**(0.02)**	**(0.06)**	**(0.03)**	**(0.03)**
Credit Premium	1.42	103.0	112.0	110.0	106.0	101.0	101.0
		(0.48)	(0.13)	(0.18)	(0.23)	(0.59)	(0.63)
Term Premium	0.69	103.0	101.0	101.0	115.0	101.0	100.0
		(0.04)	(0.93)	(0.88)	(0.25)	**(0.06)**	**(0.07)**
Panel B: MSE relative to CAPM (%), GLS							
Equity Premium	4.16	102.0	100.0	100.0	105.0	n.a.	n.a.
		(0.53)	(0.91)	(0.77)	(0.31)		
Small-Big	5.31	106.0	107.0	107.0	120.0		
		(0.13)	(0.11)	**(0.10)**	**(0.01)**		
Value-Growth	4.30	105.0	105.0	112.0	119.0		
		(0.08)	**(0.07)**	**(0.02)**	**(0.01)**		
Win-Lose	8.49	60.9	57.7	55.0	58.5		
		(0.00)	**(0.00)**	**(0.00)**	**(0.00)**		
Reversal	5.11	115.0	111.0	115.0	117.0		
		(0.00)	**(0.00)**	**(0.00)**	**(0.00)**		
Credit Premium	1.44	102.0	101.0	99.6	107.0		
		(0.20)	(0.24)	(0.41)	(0.42)		
Term Premium	0.68	103.0	97.7	97.1	97.5		
		(0.05)	(0.71)	(0.58)	(0.67)		

Table 5: Clark-West Tests on Out-of-Sample MSEs

Clark-West (2007) tests on out-of-sample MSEs are presented based on rolling estimation during 1967-2009. The number of observations in the evaluation period is 43. In panels A and B the benchmark model is the CCAPM. In panels C and D the benchmark model is the 2-State-Variable Model. The t-statistics examine the one-sided tests of the null hypothesis that the two models have the same mean squared pricing error. The alternative hypothesis is that the more complex model has smaller error. The empirical p-values are based on 10,000 simulation trials. p-values of 10% or less are in bold face. The continuously-compounded returns are in annual percent, and measured in excess of a one-month Treasury bill return. LRRC is the cointegrated long-run risk model and LRRS is the stationary long-run risk model. 2-State-Variable is the model with only shocks to the two state variables (the risk-free rate and log price/dividend ratio). RLRRC is restricted version of the cointegrated long-run risk model, and RLRRS is restricted version of the stationary long-run risk model. Panel A and C results are based on OLS risk premium estimates. Panel B and D results are based on GLS risk premium estimates.

Panel A: OLS Estimation. The Benchmark Model is the CCAPM.

	Equity Premium		**Small-Big**		**Value-Grow**		**Win-Lose**		**Reversal**		**Credit Premium**		**Term Premium**	
	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**
LRRC	-0.99	0.87	0.63	0.27	-1.38	0.93	0.69	0.33	1.00	0.15	-0.68	0.77	0.21	0.40
LRRS	-0.21	0.59	0.72	0.23	-1.39	0.94	-0.34	0.67	0.38	0.36	-1.36	0.97	-0.09	0.53
RLRRC	-0.82	0.78	1.41	**0.08**	-0.70	0.75	-2.56	0.99	1.32	**0.09**	0.09	0.51	-0.92	0.81
RLRRS	0.97	**0.08**	1.59	**0.06**	0.53	0.26	-1.48	0.99	1.63	**0.04**	-0.66	0.79	0.95	**0.08**

Panel B: GLS Estimation. The Benchmark Model is the CCAPM.

	Equity Premium		**Small-Big**		**Value-Grow**		**Win-Lose**		**Reversal**		**Credit Premium**		**Term Premium**	
	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**
LRRC	-0.78	0.79	-0.74	0.72	-1.49	0.94	1.28	0.16	-0.21	0.57	-0.86	0.86	0.00	0.48
LRRS	0.64	0.25	0.37	0.34	-2.84	1.00	0.43	0.38	-0.28	0.58	-1.43	0.97	-0.77	0.78

Panel C: OLS Estimation. The Benchmark Model is 2-State-Variable Model.

	Equity Premium		**Small-Big**		**Value-Grow**		**Win-Lose**		**Reversal**		**Credit Premium**		**Term Premium**	
	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**
LRRC	-0.14	0.55	0.97	0.17	1.97	**0.04**	3.51	**0.00**	1.20	0.12	0.87	0.16	1.43	**0.06**
LRRS	1.59	**0.06**	0.94	0.18	1.78	**0.04**	3.47	**0.00**	0.74	0.23	0.49	0.31	1.25	**0.10**
RLRRC	0.63	0.23	0.75	0.28	1.26	**0.04**	0.32	0.34	1.67	**0.04**	2.67	**0.00**	1.06	**0.10**
RLRRS	1.07	**0.07**	0.35	0.39	1.31	**0.03**	-0.96	0.92	1.54	**0.01**	-0.15	0.62	1.12	**0.07**

Panel D: GLS Estimation. The Benchmark Model is 2-State-Variable Model.

	Equity Premium		**Small-Big**		**Value-Grow**		**Win-Lose**		**Reversal**		**Credit Premium**		**Term Premium**	
	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**
LRRC	-0.15	0.57	1.92	**0.02**	2.19	**0.02**	2.92	**0.00**	1.95	**0.02**	1.52	**0.08**	2.37	**0.01**
LRRS	2.41	**0.01**	2.74	**0.00**	2.54	**0.00**	2.10	**0.02**	2.14	**0.02**	1.53	**0.07**	1.21	0.12

Table 6: Comparison of Subperiods

Pricing errors for portfolios of the seven test assets are compared, estimating the model parameters using the full sample and evaluating the pricing errors over subperiods. The returns are continuously-compounded annual percents. CCAPM is the simple consumption beta model. CAPM is the Capital Asset Pricing Model. LRRC is the cointegrated long-run risk model and LRRS is the stationary long-run risk model. 2-State-Variable is the model with only shocks to the two state variables (the risk-free rate and log price/dividend ratio). RLRRC is restricted version of the cointegrated long-run risk model and RLRRS is restricted version of the stationary long-run risk model. The MSE relative to the CAPM is calculated as 100x[1+ln(model MSE/CAPM MSE)].

Premium:	Subperiods			
	1931-1950	1951-1970	1971-1990	1991-2009
Panel A: OLS				
Average Pricing Errors of Equally-weighted Portfolio				
CCAPM	6.13	3.17	3.75	5.19
CAPM	6.81	3.53	4.10	5.54
LRRC	3.16	-0.11	0.40	2.06
LRRS	3.38	0.12	0.69	2.23
2-State-Variable	5.97	2.74	3.29	5.01
RLRRC	7.39	4.23	4.84	6.29
RLRRS	7.20	4.04	4.63	6.11
Panel B: OLS				
CAPM MSE	2.46	0.52	0.71	0.77
MSEs of Equally-Weighted Portfolio Relative to the CAPM (%)				
CCAPM	94.1	95.3	95.8	92.7
LRRC	85.8	72.7	77.4	60.7
LRRS	84.3	72.9	79.0	58.9
2-State	96.8	89.0	93.7	94.3
RLRRC	102.1	109.0	108.0	110.0
RLRRS	101.0	106.0	105.0	107.0

Premium:	Subperiods			
	1931-1950	1951-1970	1971-1990	1991-2009
Panel C: GLS				
Average Pricing Errors of GLS Portfolio				
CCAPM	5.70	2.48	4.39	5.58
CAPM	5.45	1.88	3.76	4.58
LRRC	5.02	1.53	3.49	4.21
LRRS	5.77	2.61	4.54	4.84
2-State-Variable	7.90	4.46	6.39	7.14
RLRRC	6.41	2.93	4.86	6.29
RLRRS	6.23	2.74	4.65	6.11
Panel D: GLS				
CAPM MSE	2.13	0.47	0.88	0.85
MSEs of GLS Portfolio Relative to the CAPM (%)				
CCAPM	100.0	106.0	106.0	107.0
LRRC	99.2	95.7	99.3	104.0
LRRS	101.0	107.0	108.0	110.0
2-State	114.0	130.0	127.0	136.0
RLRRC	105.0	110.0	110.0	113.0
RLRRS	104.0	108.0	108.0	111.0

Table 7: Average Pricing Errors and MSEs of the Conditional Models, 1967-2009

Average excess returns, out-of-sample pricing errors (actual minus prediction), and MSEs relative to the CAPM (%) are shown based on rolling estimation of the conditional models during 1967-2009, using a window length of 36 years. The continuously-compounded returns are in annual percent, and measured in excess of a one-month Treasury bill return. CCAPM is the simple consumption beta model. CAPM is the conditional Capital Asset Pricing Model. LRRC is the cointegrated long-run risk model and LRRS is the stationary long-run risk model. 2-State-Variable is the model with only shocks to the two state variables (the risk-free rate and log price/dividend ratio). The p-values in parentheses are for two-sided Diebold-Mariano (1995) test of the null hypothesis that the model in question performs no differently than the classical, unconditional CAPM. The empirical p-values are based on 10,000 simulation trials. p-values of 10% or less are in bold face.

Premium:		Average Pricing Errors						
	Unconditional CAPM	CCAPM	CAPM	LRRC	LRRS	2-State-Variable	RLRRC	RLRRS
Panel A: OLS Risk Premiums								
Equity Premium	2.48	2.20	3.84	2.03	1.72	4.62	4.04	19.70
		(0.72)	(0.14)	(0.62)	(0.44)	**(0.00)**	**(0.00)**	**(0.08)**
Small-Big	1.21	1.71	1.60	-0.62	2.43	2.28	2.70	3.87
		(0.64)	(0.43)	(0.20)	(0.39)	(0.27)	**(0.00)**	**(0.07)**
Value-Growth	5.51	0.88	5.22	2.18	2.08	5.11	6.29	3.96
		(0.00)	(0.24)	**(0.01)**	**(0.00)**	(0.41)	**(0.00)**	(0.63)
Win-Lose	18.30	10.90	18.10	4.58	8.47	17.70	17.80	7.21
		(0.00)	(0.52)	**(0.00)**	**(0.00)**	(0.40)	**(0.00)**	**(0.08)**
Reversal	5.14	-0.17	4.97	-4.55	-1.79	3.58	5.43	4.02
		(0.00)	**(0.10)**	**(0.00)**	**(0.00)**	**(0.06)**	**(0.04)**	(0.57)
Credit Premium	0.09	0.77	0.29	3.29	0.73	-0.03	0.58	2.23
		(0.13)	(0.24)	**(0.00)**	(0.25)	(0.58)	**(0.00)**	**(0.07)**
Term Premium	1.86	1.07	2.21	0.98	-0.51	2.19	1.81	7.01
		(0.01)	**(0.00)**	**(0.07)**	**(0.00)**	(0.33)	**(0.04)**	(0.62)
Average	4.94	2.48	5.18	1.13	1.88	5.06	5.52	6.86
Panel B: GLS Risk Premiums								
Equity Premium	-2.14	2.27	-0.03	2.82	3.02	5.49	n.a.	n.a.
		(0.00)	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**		
Small-Big	-1.05	1.47	-0.15	0.35	4.57	5.22		
		(0.00)	**(0.04)**	(0.25)	**(0.00)**	**(0.00)**		
Value-Growth	5.16	1.53	5.14	3.70	3.18	8.39		
		(0.00)	(0.92)	(0.32)	(0.12)	**(0.00)**		
Win-Lose	19.80	12.20	19.20	5.57	8.89	14.30		
		(0.00)	**(0.04)**	**(0.00)**	**(0.00)**	**(0.00)**		
Reversal	5.39	0.03	5.24	-3.74	-1.04	6.33		
		(0.00)	**(0.07)**	**(0.00)**	**(0.00)**	(0.19)		
Credit Premium	-0.69	0.59	-0.32	1.66	1.47	1.22		
		(0.00)	**(0.01)**	**(0.00)**	**(0.00)**	**(0.00)**		
Term Premium	1.36	1.28	1.68	0.91	0.62	1.86		
		(0.72)	**(0.00)**	(0.23)	(0.20)	**(0.05)**		
Average	3.98	2.77	4.39	1.61	2.96	6.12		

Premium:	Unconditional CAPM MSEs	MSE relative to unconditional CAPM (%)						
		CCAPM	CAPM	LRRC	LRRS	2-State-Variable	RLRRC	RLRRS
Panel C: MSE relative to unconditional CAPM (%), OLS								
Equity Premium	3.88	92.7	117.0	123.0	110.0	108.0	97.9	431.0
		(0.41)	**(0.10)**	**(0.04)**	(0.39)	(0.45)	(0.65)	(0.49)
Small-Big	5.21	107.0	106.0	117.0	123.0	108.0	100.0	119.0
		(0.35)	(0.19)	(0.16)	**(0.10)**	(0.36)	(0.97)	(0.36)
Value-Growth	4.13	89.3	100.0	105.0	109.0	101.0	100.0	127.0
		(0.33)	(0.97)	(0.63)	(0.46)	(0.79)	(0.95)	(0.33)
Win-Lose	9.78	76.1	101.0	88.3	81.8	101.0	96.7	284.0
		(0.00)	(0.72)	(0.67)	**(0.05)**	(0.85)	**(0.00)**	(0.63)
Reversal	4.43	114.0	98.7	122.0	120.0	111.0	101.0	104.0
		(0.24)	(0.20)	(0.13)	(0.11)	(0.14)	(0.43)	(0.13)
Credit Premium	1.34	95.6	102.0	133.0	116.0	105.0	97.1	171.0
		(0.52)	(0.53)	(0.18)	(0.20)	(0.12)	(0.19)	(0.58)
Term Premium	1.02	96.9	103.0	120.0	112.0	112.0	99.5	343.0
		(0.69)	(0.13)	**(0.04)**	(0.32)	**(0.06)**	(0.24)	(0.61)
Panel D: MSE relative to unconditional CAPM (%), GLS								
Equity Premium	3.88	92.8	112.0	123.0	110.0	104.0	97.4	456.0
		(0.46)	(0.24)	**(0.10)**	(0.50)	(0.79)	(0.78)	(0.36)
Small-Big	5.28	95.9	103.0	101.0	120.0	99.7	98.4	114.0
		(0.59)	(0.43)	(0.89)	(0.26)	(0.97)	(0.81)	(0.41)
Value-Growth	4.09	87.0	100.0	118.0	118.0	117.0	102.0	145.0
		(0.24)	(0.99)	(0.20)	(0.17)	**(0.08)**	(0.73)	(0.29)
Win-Lose	10.20	78.2	99.6	79.6	78.0	86.5	93.0	296.0
		(0.00)	(0.79)	(0.42)	**(0.03)**	**(0.02)**	**(0.00)**	(0.63)
Reversal	4.46	108.0	98.7	125.0	119.0	109.0	101.0	113.0
		(0.49)	(0.16)	(0.13)	(0.13)	(0.23)	(0.55)	(0.19)
Credit Premium	1.33	95.6	101.0	120.0	110.0	105.0	98.2	186.0
		(0.41)	(0.73)	(0.33)	(0.41)	(0.20)	(0.51)	(0.51)
Term Premium	0.99	98.7	103.0	116.0	114.0	114.0	103.0	372.0
		(0.76)	**(0.09)**	(0.11)	(0.16)	**(0.03)**	(0.12)	**(0.10)**

Figure 1: Time-Series Plots of $\{\gamma, \psi\}$ from Rolling Estimation of the RLRRS Model

Time-series plots of the utility function parameters $\{\gamma, \psi\}$, the risk aversion and the intertemporal elasticity of substitution (IES) from rolling estimation of the RLRRS model. The estimation period is 1967-2009 and the window length is 36 years.