Firm Characteristics, Stock Market Regimes, and the Cross-Section of Expected Returns☆

Chris Kirby[a]

[a] *UNC Charlotte, 9201 University City Blvd., Charlotte, NC 28223; phone: (704) 687-0845; email: ckirby10@uncc.edu*

Abstract

I propose a regime-switching generalization of instrumented principal components analysis (IPCA) that yields new insights about the relation between characteristics, factor loadings, and expected stock returns. Using a two-regime specification, I find evidence of a high-volatility regime in which individual stocks have high conditional expected returns. This contrasts sharply with the pattern of bull and bear regimes that is obtained by analyzing only market returns. Although exact factor pricing can be rejected, characteristics are more strongly related to priced covariances in the high-volatility regime. Furthermore, regime-switching predictability makes a substantial incremental contribution to the out-of-sample explanatory power of IPCA estimates.

Keywords: regime switching, principal components analysis, asset pricing, factor model, price of covariance risk, Fama-MacBeth regressions

---

☆Address correspondence to: Chris Kirby, Belk College of Business, University of North Carolina at Charlotte, 9201 University City Blvd., Charlotte, NC 28223, phone: (704) 687-0845, email: ckirby10@uncc.edu.

Firm Characteristics, Stock Market Regimes, and the Cross-Section of Expected Returns

Abstract

I propose a regime-switching generalization of instrumented principal components analysis (IPCA) that yields new insights about the relation between characteristics, factor loadings, and expected stock returns. Using a two-regime specification, I find evidence of a high-volatility regime in which individual stocks have high conditional expected returns. This contrasts sharply with the pattern of bull and bear regimes that is obtained by analyzing only market returns. Although exact factor pricing can be rejected, characteristics are more strongly related to priced covariances in the high-volatility regime. Furthermore, regime-switching predictability makes a substantial incremental contribution to the out-of-sample explanatory power of IPCA estimates.

Keywords: regime switching, principal components analysis, asset pricing, factor model, price of covariance risk, Fama-MacBeth regressions

1. Introduction

Is the relation between firm characteristics and expected stock returns consistent with the predictions of asset pricing theory? Kelly et al. (2019) use an innovative modeling strategy called instrumented principal components analysis (IPCA) to address this question. In effect, they extend traditional methods of factor analysis to a dynamic setting in which firm characteristics are cross-sectionally correlated with the conditional loadings on one or more latent risk factors. The IPCA estimates are essentially obtained by constructing the linear projections of individual stock returns and their conditional factor loadings on the set of characteristics under consideration. This yields evidence on the extent to which the cross-sectional differences in covariance risk that are captured by the characteristics explain the cross-section of conditional expected stock returns.

The main findings of Kelly et al. (2019) can be summarized succinctly. First, the set of firm characteristics that have incremental explanatory power for the cross-section of expected stock returns is relatively small (around ten in total). Second, a latent factor specification with five factors performs considerably better than existing factor models in asset pricing tests. Because the latent factor specification generally produces pricing errors that are statistically indistinguishable from zero, the evidence suggests that firm characteristics help to explain the cross-section of expected stock returns by capturing cross-sectional differences in priced covariance risk.

However, the IPCA evidence is subject to a potentially noteworthy caveat. Under the maintained assumptions of IPCA, individual stock returns are described by a conditional factor model that has time-varying intercepts and time-varying factor loadings. Although the model seems quite flexible at first glance, it inherently imposes strong restrictions on the set of characteristic-based managed portfolios that determine the nature of the IPCA estimates. Specifically, it implies that the returns for these portfolios are described by a *static* factor model that has constant intercepts and constant loadings. Thus the IPCA estimates fail to account for the possibility that the means, variances, and covariances of the portfolio returns evolve through time. This feature of the IPCA framework has the potential to alter the outcome of the associated asset pricing tests.

To develop insights in this regard, I propose a more flexible version of IPCA that allows the means, variances, and covariances of the managed portfolio returns to differ across economic regimes.

I adopt a regime-switching approach for several reasons. First, the idea of regime changes is both conceptually simple and intuitively appealing. The regimes could be associated with the business cycle, as in Hamilton (1989), or related to other phenomena such as "bull' and "bear" markets. Second, the nonlinear dynamics associated with regime changes can have interesting asset pricing implications. If the changes induce large shifts in the investment opportunity set, then they can give rise to unanticipated effects, such as a non-monotonic relation between risk and expected returns (see, e.g., Whitelaw, 2000). Third, it turns out that the econometric methods of IPCA can easily be generalized to a regime-switching framework.

The first step of the empirical analysis is to specify the characteristics. One possibility would be to use those that have statistically-significant explanatory power in the Kelly et al. (2019) tests. However, their tests emphasize linearity in the relation between firm characteristics and expected stock returns. Because Kirby (2019a) shows that the nonlinear aspects of this relation appear to have important asset pricing implications, I employ a set of 16 characteristics that are chosen based on the results of Freyberger et al. (2019), who apply advanced variable selection techniques to a fairly comprehensive set of 62 firm-specific variables in order to identify those that provide *incremental* information about the cross-section of expected stock returns. Although there are other recent studies, such as Green et al. (2017), that have a similar focus, Freyberger et al. (2019) use nonparametric methods that do not impose strong a priori assumptions regarding functional form.

I begin by considering the results produced by the single-regime version of IPCA. The sample covers May 1967 to December 2018 (620 months of individual stock returns). First I examine the case in which the relation between firm characteristics and expected stock returns is modeled as linear. The results obtained in this case are broadly consistent with those reported by Kelly et al. (2019). The estimated root mean squared pricing error (RMSE) declines as the number of latent factors increases and there are instances in which it is statistically insignificant at the 1% level for specifications with five or six latent risk factors. I then examine the case in which the relation between firm characteristics and expected stock returns is modeled as nonlinear. Although I again find that the magnitude of the pricing errors falls as the number of latent risk factors increases, I do not find any instances in which a specification with six or fewer factors produces a statistically insignificant estimate of the RMSE. Thus the preliminary evidence suggests that it is important to account for nonlinearity in IPCA-based pricing tests.

Next I turn to the regime-switching analysis, which yields a number of new and interesting insights. The analysis centers on simple two-regime models in order to highlight how the results compare to those of prior studies, such as Maheu and McCurdy (2000), that investigate bull and bear market regimes. Like these studies, I find that analyzing market index returns in isolation produces evidence of two persistent regimes: one that has low volatility and a high mean return (the bull regime) and one that has high volatility and a low mean return (the bear regime). However, I also find that analyzing market index returns in conjunction with the returns for characteristic-based managed portfolios produces marked changes in the regime-switching estimates.

First, both regimes have a lower estimated expected duration. The estimate for the high-volatility regime, for example, is only about three months. Second, the time intervals for which the estimated probability of the high-volatility regime is close to one do not correspond to bear market periods, as defined by a common ex-post classification rule. Third, the estimated expected excess return on the market is much higher in the high-volatility regime than in the low-volatility regime. Using stocks of all capitalization levels, for example, the results produced by a six-factor specification imply that the estimated expected excess return on the equally-weighted market index is 0.16% per month in the low-volatility regime and 2.53% per month in the high-volatility regime. Thus the estimates are consistent with a positive relation between the market risk premium and market volatility.

As might be anticipated from these findings, the asset pricing tests indicate that the performance of the conditional factor model differs across regimes. Regardless of the number of risk factors, I find that the estimated RMSE for the high-volatility regime exceeds that for the low-volatility regime by a considerable margin. But it is important to interpret this finding within the broader context of the results. Using zero factors, for example, the estimated RMSE is much larger for the high-volatility regime than for the low-volatility regimes: 0.97% per month versus 0.23% per month using stocks of all capitalization levels. Hence the high-volatility regime clearly poses a pricing challenge. Fitting a six-factor specification of the model reduces the estimated RMSE to 0.23% per month for the high-volatility regime. But this still exceeds the value of 0.09% per month obtained for the low-volatility regime. Using formal regime-specific tests, I find that all of the estimated RMSEs are statically significant at the 1% level for both regimes.

More broadly, the regime-specific pricing tests suggest that the linkages between characteristics, covariances, and expected stock returns are stronger in the high-volatility regime than in the low-volatility regime. This is evidenced in several ways. For example, I find that the risk factors explain a larger fraction of the time-series variation in the managed portfolio returns in the high-volatility regime. I also find that the IPCA analog of the second pass regression for the familiar two-pass methodology generally produces high R-squared statistics for the high-volatility regime, especially for the case in which micro-cap stocks are included in the analysis. Those produced by the six-factor specification, for instance, range from 89% to 93% in this case.

To assess the potential implications of these findings for investors, I investigate the out-of-sample performance of the conditional factor model that underpins regime-switching IPCA. I start by examining the ability of recursively-updated estimates of the regime probabilities to predict the excess returns for the characteristic-based managed portfolios. My approach is straightforward. I regress the excess monthly portfolio returns for month t on the estimated probability of the high-volatility regime for month t, where the regressor is constructed by fitting the regime-switching model to the data observed through month $t-1$. Although a substantial fraction of the managed portfolios display statistically-significant evidence of predictability, the explanatory power of the predictive regressions is fairly low. All of the R-squared statistics are less than 5%.

But low R-squared statistics do not necessarily preclude the construction of profitable dynamic trading strategies. I therefore consider the evidence from portfolio sorts that are designed to isolate the contribution of the predictable variation in the managed portfolio returns to the explanatory power of the conditional factor model for individual stocks. The sorts are conducted using two competing estimates of the vector of excess individual stock returns, both of which are constructed using information available to investors in real time. The first corresponds to the IPCA framework of Kelly et al. (2019) and the second corresponds to my regime-switching version of IPCA.

Because the cross-sectional correlation between the competing estimates of the vector of expected excess returns for a given month is typically fairly high, I use a residual-based sorting scheme to isolate the effect of regime changes. First, I regress the estimated expected excess stock returns produced by the regime-switching model for the month on those produced by the single-regime model of Kelly et al. (2019) and save the resultant vector of residuals. Next, I conduct two independent quintile sorts using the set of available stocks for the month: one in ascending order of the single-regime estimates and one in ascending order of the saved residuals. Finally, I form a set of 25 equally-weighted portfolios by finding the intersections of the quintile groups.

Using the single-regime estimates to sort stocks is quite effective at spreading average returns. Nonetheless, it is clear that the regime-switching estimates provide substantial incremental information about conditional expected stock returns. This is evidenced by the large spread in average returns that is generated by sorting on the regression residuals. Regardless of the level of estimated expected excess returns produced by the single-regime model, the average excess returns obtained by investing in stocks that have high residuals and shorting stocks that have low residuals are positive and statistically significant at the 1% level. The smallest of these high-low spreads is 0.98% per month, which has a *t*-statistic of 3.37. Notably, the findings are robust to excluding micro-cap stocks from the analysis and using value weights to construct the portfolios. This highlights the potential benefits of accounting for regime-switching behavior in investment management.

Overall the analysis lends a good deal of support to the view that firm characteristics display explanatory power in Fama and MacBeth (1973) regressions because they capture cross-sectional variation in priced covariance risk. But the results are less favorable to the hypothesis of exact factor pricing than those of Kelly et al. (2019). Perhaps the most striking finding is that the relation between characteristics and individual stock returns is indicative of the presence of latent regimes that are quite different than those identified by studies that focus only on broad market returns. The evidence points to a relatively short-lived regime in which return volatility is high, expected excess stock returns are high, and there is a strong cross-sectional relation between firm characteristics and priced covariance risk. In addition, it suggests that regime changes are associated with substantial shifts in the investment opportunity set that have a range of important asset-pricing implications. Beyond producing this intriguing set of empirical findings, regime-switching IPCA extends all of the methodological benefits of IPCA to a wider class of estimation problems.

2. Methodology

The IPCA framework of Kelly et al. (2019) generalizes traditional methods of factor analysis to a dynamic setting in which the factor loadings are cross-sectionally correlated with observable firm characteristics (instruments). To illustrate, let $\boldsymbol{r}_t = (r_{1,t}, \ldots, r_{N,t})'$ and $\boldsymbol{f}_t = (f_{1,t}, \ldots, f_{K,t})'$ denote the period t values of an $N \times 1$ vector of individual stock returns that are measured in excess of the risk free rate (excess returns) and a $K \times 1$ vector of latent risk factors.[1] The IPCA methodology assumes that $\boldsymbol{r}_{t+1}$ is described by a conditional linear factor model of the form

$$\boldsymbol{r}_{t+1} = \boldsymbol{a}_t + \sum_{k=1}^{K} \boldsymbol{b}_{k,t} f_{k,t+1} + \boldsymbol{\epsilon}_{t+1}, \tag{1}$$

where $\boldsymbol{a}_t$ is a $N \times 1$ vector of time-varying intercepts, $\boldsymbol{b}_{k,t}$ is an $N \times 1$ vector of time-varying loadings on the kth factor, and $\boldsymbol{\epsilon}_{t+1}$ is an $N \times 1$ vector of errors.

As it stands, the model does not place any restrictions on the dynamics of the intercepts or loadings. Suppose, however, that the values of $\boldsymbol{a}_t$ and $\{\boldsymbol{b}_{k,t}\}_{k=1}^{K}$ are cross-sectionally correlated with one or more firm characteristics whose values are predetermined with respect to the realizations of $\{f_{k,t+1}\}_{k=1}^{K}$ and $\boldsymbol{\epsilon}_{t+1}$. More specifically, suppose that $\boldsymbol{a}_t$ and $\{\boldsymbol{b}_{k,t}\}_{k=1}^{K}$ can be decomposed as

$$\boldsymbol{a}_t = \boldsymbol{C}_t \boldsymbol{\delta} + \boldsymbol{\nu}_t, \tag{2}$$

$$\boldsymbol{b}_{k,t} = \boldsymbol{C}_t \boldsymbol{\gamma}_k + \boldsymbol{\upsilon}_{k,t}, \; k = 1, \ldots, K, \tag{3}$$

where $\boldsymbol{C}_t$ is an observable $N \times J$ matrix whose nth row contains the values of the characteristics for the nth firm. Under the maintained assumptions of IPCA, the elements of $\boldsymbol{\nu}_t$ and $\{\boldsymbol{\upsilon}_{k,t}\}_{k=1}^{K}$ behave like the errors in a cross-sectional regression model. They represent time-varying components of $\boldsymbol{a}_t$ and $\{\boldsymbol{b}_{k,t}\}_{k=1}^{K}$ that are assumed to be cross-sectionally unrelated to the values of the characteristics in the sense that $\boldsymbol{C}_t'\boldsymbol{\nu}_t \xrightarrow{p} \boldsymbol{0}$ and $\boldsymbol{C}_t'\boldsymbol{\upsilon}_{k,t} \xrightarrow{p} \boldsymbol{0}$ as $N \to \infty$ for all t and k.

Kelly et al. (2019) develop an alternating least squares (ALS) algorithm that can be used to fit

[1] The discussion in this section treats N as fixed to avoid unnecessarily complicating the notation. The extension to the case in which the cross-sectional dimension of the data varies through time is obvious. In subsequent sections and the table descriptions, I use N_t to denote the number of stocks for period t.

the IPCA model to individual stock returns and extract estimates of the latent risk factors. First, they combine equations (1), (2) and (3) to obtain

$$\boldsymbol{r}_{t+1} = \boldsymbol{C}_t\boldsymbol{\delta} + \boldsymbol{C}_t\boldsymbol{\Gamma}\boldsymbol{f}_{t+1} + \boldsymbol{\varepsilon}_{t+1}, \tag{4}$$

where $\boldsymbol{\Gamma} = (\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_K)$ is an $J \times K$ matrix of coefficients and

$$\boldsymbol{\varepsilon}_{t+1} = \boldsymbol{\nu}_t + \sum_{k=1}^{K} \boldsymbol{\upsilon}_{k,t} f_{k,t+1} + \boldsymbol{\epsilon}_{t+1} \tag{5}$$

is an $N \times 1$ vector of composite errors. Next, they construct estimators of $\boldsymbol{\delta}$, $\boldsymbol{\Gamma}$, and $\boldsymbol{F} = (\boldsymbol{f}_1, \ldots, \boldsymbol{f}_T)$ by minimizing the average squared composite error across all observations. This entails solving an optimization problem of the form

$$\min_{\boldsymbol{\delta},\boldsymbol{\Gamma},\boldsymbol{F}} \frac{1}{NT} \sum_{t=1}^{T} (\boldsymbol{r}_t - \boldsymbol{C}_{t-1}\boldsymbol{\delta} - \boldsymbol{C}_{t-1}\boldsymbol{\Gamma}\boldsymbol{f}_t)'(\boldsymbol{r}_t - \boldsymbol{C}_{t-1}\boldsymbol{\delta} - \boldsymbol{C}_{t-1}\boldsymbol{\Gamma}\boldsymbol{f}_t) \tag{6}$$

under identification restrictions like those typically employed in traditional principal components analysis (PCA). The details of the ALS algorithm are provided in Appendix A.

2.1. Relation to Fama and MacBeth (1973) regressions

It turns out that the IPCA methodology has close ties to Fama and MacBeth (1973) regressions. To see the nature of these ties, consider the typical case in which $\boldsymbol{C}_t$ is a rank J matrix that has a left pseudoinverse of the form $\boldsymbol{C}_t^+ = (\boldsymbol{C}_t'\boldsymbol{C}_t)^{-1}\boldsymbol{C}_t'$. In this case, the vector of excess stock returns for period $t+1$ has an additive decomposition of the form $\boldsymbol{r}_{t+1} = \boldsymbol{C}_t\boldsymbol{r}_{p,t+1} + \boldsymbol{e}_{t+1}$, where $\boldsymbol{r}_{p,t+1} = \boldsymbol{C}_t^+\boldsymbol{r}_{t+1}$ is a $J \times 1$ vector. As the notation suggests, $\boldsymbol{r}_{p,t+1}$ represents a $J \times 1$ vector of excess returns for a set of managed portfolios whose weights are chosen by regression methods. Specifically, the weights are obtained by fitting a Fama and MacBeth (1973) regression of the excess individual stock returns for period $t+1$ on the firm characteristics for period t.

Because $\boldsymbol{C}_t'\boldsymbol{e}_{t+1} = \boldsymbol{0}$ by construction, it follows by straightforward algebra that the optimization problem in equation (6) can be expressed as

$$\min_{\delta,\Gamma,F} \frac{1}{T}\sum_{t=1}^{T}(\boldsymbol{r}_{p,t}-\boldsymbol{\delta}-\boldsymbol{\Gamma f}_t)'\hat{\boldsymbol{W}}_{t-1}(\boldsymbol{r}_{p,t}-\boldsymbol{\delta}-\boldsymbol{\Gamma f}_t)+\frac{1}{NT}\sum_{t=1}^{T}\boldsymbol{e}_t'\boldsymbol{e}_t, \tag{7}$$

where $\hat{\boldsymbol{W}}_{t-1} = (1/N)\boldsymbol{C}_{t-1}'\boldsymbol{C}_{t-1}$ is the estimated cross-sectional second moment matrix of the characteristics for period $t-1$. Hence, the ALS estimators of $\boldsymbol{\delta}$, $\boldsymbol{\Gamma}$, and $\boldsymbol{F}$ are obtained by fitting a static factor model to $\{\boldsymbol{r}_{p,t}\}_{t=1}^{T}$ via a type of weighted least squares (WLS). Back et al. (2015) refer to the elements of $\boldsymbol{r}_{p,t+1}$ as returns for "pure play portfolios" because $\boldsymbol{C}_t^{+}$, the $J \times N$ matrix of portfolio weights, satisfies $\boldsymbol{C}_t^{+}\boldsymbol{C}_t = \mathbf{I}$. Thus the portfolio associated with the jth characteristic has a unit exposure to this characteristic and zero exposure to the other $J-1$ characteristics. This property makes pure play portfolios an appealing choice for use in asset pricing tests. Following Kelly et al. (2019), I henceforth assume that the leading column of $\boldsymbol{C}_t$ is an $N \times 1$ vector of ones, which is equivalent to assuming the Fama and MacBeth (1973) regression used to construct $\boldsymbol{r}_{p,t+1}$ includes an intercept.

2.2. Asset pricing tests in the IPCA framework

The model in equation (4) implies that $\boldsymbol{\delta} = \mathbf{0}$ if exact factor pricing holds. Following Kelly et al. (2019), I test this hypothesis using an approach that is akin to the two-pass regression methodology of Black et al. (1972). The similarities to two-pass methodology arise from the need to impose identification restrictions in the IPCA framework. Suppose, for example, that $\hat{\boldsymbol{\delta}}$, $\hat{\boldsymbol{\Gamma}}$, and $\hat{\boldsymbol{F}}$ denote the set of estimates obtained by solving the minimization problem in equation (6) for a given dataset. Now suppose that some arbitrary $K \times 1$ vector $\boldsymbol{\mu}$ is used to create a new set of estimates by replacing $\hat{\boldsymbol{\delta}}$ with $\hat{\boldsymbol{\delta}}^{\star} = \hat{\boldsymbol{\delta}} + \hat{\boldsymbol{\Gamma}}\boldsymbol{\mu}$ and replacing $\hat{\boldsymbol{f}}_t$ with $\hat{\boldsymbol{f}}_t^{\star} = \hat{\boldsymbol{f}}_t - \boldsymbol{\mu}$ for all values of t. Because $\hat{\boldsymbol{\delta}}^{\star} + \hat{\boldsymbol{\Gamma}}\hat{\boldsymbol{f}}_t^{\star}$ has the same value as $\hat{\boldsymbol{\delta}} + \hat{\boldsymbol{\Gamma}}\hat{\boldsymbol{f}}_t$, the two sets of estimates produce the same value of the IPCA objective function. Hence, the values of $\boldsymbol{\delta}$ and $\mathrm{E}(\boldsymbol{f}_t)$ are unidentified due to the latent nature of the factors.

The lack of identification can be overcome by imposing a set of $K(K+1)$ restrictions on the parameters of the conditional factor model. I impose restrictions that mirror those typically employed for PCA. Specifically, I restrict the expected value of each element of $\boldsymbol{f}_t$ to be zero, the covariance matrix of $\boldsymbol{f}_t$ to be an identity matrix, and $\boldsymbol{\Gamma}'\boldsymbol{\Gamma}$ to be a diagonal matrix with descending diagonal elements.[2] Estimating $\boldsymbol{\delta}$, $\boldsymbol{\Gamma}$, and $\boldsymbol{F}$ under this set of restrictions is only a first step toward constructing a suitable test statistic. Although the estimate of $\boldsymbol{\Gamma}$ can be used to draw inferences about the explanatory power of the factors for excess stock returns, the estimate of $\boldsymbol{\delta}$ tells us nothing about the pricing errors for the model. This follows because the estimate of $\boldsymbol{\delta}$ embodies the restriction $\mathrm{E}(\boldsymbol{f}_t) = \mathbf{0}$, which implies the compensation for bearing factor risk is zero.

Note that a similar situation arises if the factors are observed but not traded. The factor loadings can be estimated by regressing excess returns on the demeaned factors, but doing so tells us nothing about the pricing errors. As in this analogous case, a simple way forward is to estimate the pricing errors via a second-pass regression. Under the restriction $\mathrm{E}(\boldsymbol{f}_t) = \mathbf{0}$, equation (4) implies that the vector of expected excess returns for the pure play portfolios is given by $\mathrm{E}(\boldsymbol{r}_{p,t}) = \boldsymbol{\delta} + \mathrm{E}(\boldsymbol{\varepsilon}_{p,t})$, where $\boldsymbol{\varepsilon}_{p,t} = \boldsymbol{C}_t^{+}\boldsymbol{\varepsilon}_t$. Recall, however, that $\boldsymbol{\varepsilon}_{p,t} \xrightarrow{p} \mathbf{0}$ as $N \to \infty$ under the maintained assumptions of IPCA. Because this result implies that $\hat{\boldsymbol{\delta}}$ is a consistent estimator of $\mathrm{E}(\boldsymbol{r}_{p,t})$ under large-N asymptotics, it follows that $\hat{\boldsymbol{\mu}} = (\hat{\boldsymbol{\Gamma}}'\hat{\boldsymbol{\Gamma}})^{-1}\hat{\boldsymbol{\Gamma}}'\hat{\boldsymbol{\delta}}$ is a natural regression-based estimator of the vector of factor risk premiums. I therefore use $\hat{\boldsymbol{\alpha}} = \hat{\boldsymbol{\delta}} - \hat{\boldsymbol{\Gamma}}\hat{\boldsymbol{\mu}}$, which is the vector of residuals for the second-pass regression of $\hat{\boldsymbol{\delta}}$ on $\hat{\boldsymbol{\Gamma}}$, as my estimator of the vector of pricing errors for the pure play portfolios.[3]

2.2.1. Test statistics and inference

The statistical significance of estimated pricing errors is most easily assessed using bootstrap methods. One strategy, which exploits the structure of the IPCA problem in equation (7), would be to hold $\{\hat{\boldsymbol{W}}_{t-1}\}_{t=1}^{T}$ fixed and generate bootstrap samples by drawing randomly with replacement from $\{\boldsymbol{r}_{p,t}\}_{t=1}^{T}$. Because equation (4) implies that $\boldsymbol{r}_{p,t} = \hat{\boldsymbol{\delta}} + \hat{\boldsymbol{\Gamma}}\hat{\boldsymbol{f}}_t + \hat{\boldsymbol{\varepsilon}}_{p,t}$ by construction, drawing a value $\boldsymbol{r}^{*}_{p,t}$ at random from $\{\boldsymbol{r}_{p,t}\}_{t=1}^{T}$ is equivalent to setting $\boldsymbol{r}^{*}_{p,t} = \hat{\boldsymbol{\delta}} + \hat{\boldsymbol{\Gamma}}\hat{\boldsymbol{f}}^{*}_t + \hat{\boldsymbol{\varepsilon}}^{*}_{p,t}$, where $\{\hat{\boldsymbol{f}}^{*}_t, \hat{\boldsymbol{\varepsilon}}^{*}_{p,t}\}$ denotes a random draw from $\{\hat{\boldsymbol{f}}_t, \hat{\boldsymbol{\varepsilon}}_{p,t}\}_{t=1}^{T}$. This is akin to bootstrapping pairs in linear regression, a well-established approach for capturing conditional heteorskedasticity (Freedman, 1981). A related strategy would be

[2] Appendix A explains how to impose these restrictions, which identify the loadings and factors up to arbitrary sign changes, on the ALS estimates for a given dataset. To identify the signs, I restrict the estimates of the factor risk premiums used for the asset pricing tests to be positive.

[3] Another point to note is that $\hat{\boldsymbol{\alpha}} = (\mathbf{I} - \hat{\boldsymbol{\Gamma}}(\hat{\boldsymbol{\Gamma}}'\hat{\boldsymbol{\Gamma}})^{-1}\hat{\boldsymbol{\Gamma}}')\hat{\boldsymbol{\delta}}$ is identical to the estimate of $\boldsymbol{\delta}$ that would result from specifying $\boldsymbol{\Gamma}'\boldsymbol{\delta} = \mathbf{0}$ as part of the identification restrictions. In effect, therefore, the second-pass regression is used to impose K restrictions on the vector $\boldsymbol{\mu} = \mathrm{E}(\boldsymbol{f}_t)$ that maximize the explanatory power of $\boldsymbol{\Gamma}$ from a factor pricing perspective.

to draw randomly from $\{\boldsymbol{r}_{p,t}, \hat{\boldsymbol{W}}_{t-1}\}_{t=1}^{T}$, or equivalently, from $\{\boldsymbol{r}_t, \boldsymbol{C}_{t-1}\}_{t=1}^{T}$, which is the approach that I use for the empirical analysis. It is similar to one used for panel data regressions — drawing entire cross-sections of residuals at random — that is designed to capture the cross-sectional correlation in the errors (see, e.g., Gonalves, 2011).[4]

The tests are conducted using the RMSE for the pure play portfolios. To illustrate, let $\hat{\boldsymbol{\delta}}^*$ and $\hat{\boldsymbol{\Gamma}}^*$ denote estimates of $\boldsymbol{\delta}$ and $\boldsymbol{\Gamma}$ for a given bootstrap sample. Using these estimates, I construct $\hat{\boldsymbol{\alpha}}^* = \hat{\boldsymbol{\delta}}^* - \hat{\boldsymbol{\Gamma}}^* \hat{\boldsymbol{\mu}}^*$, where $\hat{\boldsymbol{\mu}}^* = (\hat{\boldsymbol{\Gamma}}^{*\prime}\hat{\boldsymbol{\Gamma}}^*)^{-1}\hat{\boldsymbol{\Gamma}}^{*\prime}\hat{\boldsymbol{\delta}}^*$. Because the data generating process for the bootstrap samples implies that $\hat{\boldsymbol{\alpha}}$ and $RMSE = (\hat{\boldsymbol{\alpha}}'\hat{\boldsymbol{\alpha}}/J)^{1/2}$ are the population values of the vector of pricing errors and RMSE, I use the centered value of $\hat{\boldsymbol{\alpha}}^*$ and the associated RMSE to test the null hypothesis of exact factor pricing for a given number of factors. Specifically, I use the observed distribution of $RMSE^* = ((\hat{\boldsymbol{\alpha}}^* - \hat{\boldsymbol{\alpha}})'(\hat{\boldsymbol{\alpha}}^* - \hat{\boldsymbol{\alpha}})/J)^{1/2}$ across 500 bootstrap trials to compute 10%, 5%, and 1% critical values for the RMSE under the null hypothesis.

2.3. Recasting IPCA as a likelihood-based procedure

It should be apparent at this point that using IPCA to conduct pricing tests for the cross-section of individual stocks is similar to using PCA to conduct pricing tests for pure play portfolios. I now show that the connection between the IPCA and PCA methodologies is even closer than it might seem. Indeed, one can easily formulate PCA-based estimators of $\boldsymbol{\delta}$ and $\boldsymbol{\Gamma}$ that can be constructed without the use of iterative optimization methods. This strategy plays a key role in the subsequent generalization of IPCA because it implies that the IPCA procedure can be embedded within a likelihood-based framework for estimation and inference.

The first step is to build on an insight that is implicit in a couple of previous arguments. That is, the conditional factor model in equation (4) can be multiplied by $\boldsymbol{C}_t^+$ to obtain

$$\boldsymbol{r}_{p,t+1} = \boldsymbol{\delta} + \boldsymbol{\Gamma}\boldsymbol{f}_{t+1} + \boldsymbol{\varepsilon}_{p,t+1}, \tag{8}$$

which is a *static* linear factor model for the excess returns on the pure play portfolios. One of the

[4] These procedures could be extended to capture serial correlation by sampling blocks of data. But this has little impact empirically because there is little serial correlation in the managed portfolio returns.

immediate implications of equation (8) is that $\boldsymbol{\delta}$ and $\boldsymbol{\Gamma}$ can be estimated via conventional PCA. Specifically, the estimates can be obtained by setting $\hat{\boldsymbol{\delta}} = (1/T)\sum_{t=1}^{T} \boldsymbol{r}_{p,t}$ and setting the columns of $\hat{\boldsymbol{\Gamma}}$ equal to the leading K orthonormal eigenvectors of the sample covariance matrix of $\boldsymbol{r}_{p,t}$. I refer to this strategy for constructing IPCA estimators as the direct solution method for the IPCA problem because it does not rely on any type of iterative estimation scheme.

The direct solution method embodies the same basic assumptions as the ALS algorithm of Kelly et al. (2019). In fact, the arguments that are used to motivate the ALS algorithm are drawn from Kelly et al. (2017), who develop the asymptotic theory for IPCA estimators under circumstances in which $\hat{\boldsymbol{W}}_t = \mathbf{I}$ for all t. There is an obvious link between IPCA and PCA in this case because the IPCA estimators are constructed by applying PCA to the sample covariance matrix of the vector of managed portfolio returns given by $(1/N)\boldsymbol{C}'_{t-1}\boldsymbol{r}_t$. Although Kelly et al. (2017) do not explicitly address the case in which $\hat{\boldsymbol{W}}_t \neq \mathbf{I}$, their arguments can be extended to cover this case by making suitable assumptions about the limiting behavior of $\hat{\boldsymbol{W}}_t$ as $N \to \infty$.[5]

2.3.1. Maximum likelihood estimators

The next step is to link the estimators produced by the direct solution method to the maximum likelihood estimators for a closely related problem. Following Tipping and Bishop (1999), I consider a scenario in which $\boldsymbol{\varepsilon}_{p,t+1} \sim N(\mathbf{0}, \varsigma^2\mathbf{I})$ and $\boldsymbol{f}_{t+1} \sim N(\mathbf{0}, \boldsymbol{\Psi})$, where identification is achieved by assuming that $\boldsymbol{\Gamma}'\boldsymbol{\Gamma}$ is a diagonal matrix with descending diagonal elements and that $\boldsymbol{\Psi} = \mathbf{I}$. Under the static factor model in equation (8), the identification restrictions in conjunction with the

[5]By combining equations (5) and (8), the factor model for the vector of excess returns on the pure play portfolios can be expressed as

$$\boldsymbol{r}_{p,t+1} = \boldsymbol{\delta} + \boldsymbol{\Gamma}\boldsymbol{f}_{t+1} + \hat{\boldsymbol{W}}_t^{-1/2}\left[\frac{1}{N}\boldsymbol{Z}_t'\boldsymbol{\nu}_t + \frac{1}{N}\sum_{k=1}^{K}\boldsymbol{Z}_t'\boldsymbol{\upsilon}_{k,t}f_{k,t+1}\right],$$

where $\hat{\boldsymbol{W}}_t^{-1/2}$ is the symmetric square root of $\hat{\boldsymbol{W}}_t$ and $\boldsymbol{Z}_t = \boldsymbol{C}_t\hat{\boldsymbol{W}}_t^{-1/2}$ satisfies $(1/N)\boldsymbol{Z}_t'\boldsymbol{Z}_t = \mathbf{I}$. Kelly et al. (2017) consider the case in which $\hat{\boldsymbol{W}}_t = \mathbf{I}$ and show that applying PCA to the sample covariance matrix of $\boldsymbol{r}_{p,t+1}$ delivers consistent estimators provided that the term inside the brackets is $O_p(N^{-1/2})$. To extend their asymptotic arguments to cover the more commonly encountered case in which $\hat{\boldsymbol{W}}_t \neq \mathbf{I}$, one could assume, for instance, that $\hat{\boldsymbol{W}}_t$ converges in probability to $\boldsymbol{W}_t$ as $N \to \infty$ for all t, where $\boldsymbol{W}_t$ displays stationary dynamics and has suitable distributional properties.

distributional assumptions for $\boldsymbol{\varepsilon}_{p,t+1}$ and $\boldsymbol{f}_{t+1}$ imply that

$$\boldsymbol{r}_{p,t+1} \sim N(\boldsymbol{\delta}, \boldsymbol{\Gamma}\boldsymbol{\Gamma}' + \varsigma^2\mathbf{I}). \tag{9}$$

It is easy to see the maximum likelihood estimator of $\boldsymbol{\delta}$ is given by $\bar{\boldsymbol{r}}_p = (1/T)\sum_{t=1}^{T} \boldsymbol{r}_{p,t}$. Hence the concentrated log likelihood function can be expressed as

$$\mathcal{L}(\boldsymbol{\Gamma}, \varsigma^2) = -\frac{TJ}{2}\log(2\pi) - \frac{T}{2}\log|\boldsymbol{\Sigma}| - \frac{T}{2}\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{S}), \tag{10}$$

where $\boldsymbol{\Sigma} = \boldsymbol{\Gamma}\boldsymbol{\Gamma}' + \varsigma^2\mathbf{I}$ and $\boldsymbol{S} = (1/T)\sum_{t=1}^{T}(\boldsymbol{r}_{p,t} - \bar{\boldsymbol{r}}_p)(\boldsymbol{r}_{p,t} - \bar{\boldsymbol{r}}_p)'$ denote the population and sample covariance matrix of $\boldsymbol{r}_{p,t}$, respectively.

Now let $\boldsymbol{\Gamma} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}'$ denote the singular value decomposition (SVD) of the loading matrix for the pure play portfolios. The first-order conditions (FOCs) obtained by differentiating $\mathcal{L}(\boldsymbol{\Gamma}, \varsigma^2)$ with respect to $\boldsymbol{\Gamma}$ imply that $\boldsymbol{S}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Gamma} = \boldsymbol{\Gamma}$ (see, e.g., Lawley and Maxwell, 1971). Building on this result, Tipping and Bishop (1999) show that the FOCs for $\boldsymbol{\Gamma}$ imply that

$$\boldsymbol{S}\boldsymbol{u}_k = (d_k^2 + \varsigma^2)\boldsymbol{u}_k, \; k = 1, \ldots, K, \tag{11}$$

where $\boldsymbol{u}_k$ is the kth column of $\boldsymbol{U}$ and d_k is the kth diagonal element of $\boldsymbol{D}$.[6] Equation (11) implies that $\boldsymbol{u}_k$ is an eigenvector of $\boldsymbol{S}$ and that its associated eigenvalue is given by $l_k = d_k^2 + \varsigma^2$. Thus the kth column of the maximum likelihood estimator of $\boldsymbol{\Gamma}$ is given by $\hat{\boldsymbol{\gamma}}_k = \boldsymbol{u}_k(l_k - \hat{\varsigma}^2)^{1/2}$, where l_k denotes the kth largest eigenvalue of $\boldsymbol{S}$ and $\hat{\varsigma}^2$ denotes the maximum likelihood estimator of ς^2, which Tipping and Bishop (1999) show is given by $\hat{\varsigma}^2 = (J - K)^{-1}\sum_{j=K+1}^{J} l_j$.

There are two key takeaways from these results: (i) the maximum likelihood estimator of $\boldsymbol{\delta}$ is identical to the PCA estimator of $\boldsymbol{\delta}$ and (ii) the maximum likelihood of $\boldsymbol{\Gamma}$ has the same column space as the PCA estimator of $\boldsymbol{\Gamma}$. It therefore follows that implementing the asset pricing tests using the

[6] The first step in deriving equation (11) is to note that $(\boldsymbol{\Gamma}\boldsymbol{\Gamma}' + \varsigma^2\mathbf{I})^{-1}\boldsymbol{\Gamma} = \boldsymbol{\Gamma}(\boldsymbol{\Gamma}'\boldsymbol{\Gamma} + \varsigma^2\mathbf{I})^{-1}$. Thus $\boldsymbol{\Sigma}^{-1}\boldsymbol{\Gamma} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}'(\boldsymbol{V}\boldsymbol{D}\boldsymbol{U}'\boldsymbol{U}\boldsymbol{D}\boldsymbol{V}' + \varsigma^2\mathbf{I})^{-1}$. Because $\boldsymbol{\Gamma}'\boldsymbol{\Gamma} = \boldsymbol{V}\boldsymbol{D}^2\boldsymbol{V}'$, it follows that $\boldsymbol{V} = \mathbf{I}$ under the imposed set of identification restrictions. The rest is simple algebra using the properties of orthonormal matrices.

maximum likelihood estimators will yield the same outcome as implementing the tests using the PCA estimators. Although there is no inherent advantage to employing the maximum likelihood estimators in the Kelly et al. (2019) framework, the ability to recast IPCA as a likelihood-based procedure paves the way for generalizing IPCA to other settings.

2.4. Regime-switching IPCA and regime-specific pricing tests

The assumption that $\boldsymbol{\delta}$ and $\boldsymbol{\Gamma}$ are constant is central to all of the foregoing results. As discussed earlier, however, modeling these parameters as constant is equivalent to assuming that the excess returns for the pure play portfolios are described by a static linear factor model. This aspect of IPCA seems overly restrictive in view of the overwhelming evidence that variances and covariances of excess portfolio returns vary through time. It also leaves little room for time-varying risk premiums to play a role in the analysis. To address these issues, I propose a regime-switching generalization of IPCA that allows both the mean vector and covariance matrix of the excess portfolio returns to differ across unobserved economic and/or market regimes.

I favor a regime-switching approach over other potential modeling strategies for several reasons. First, the idea of regime changes is both conceptually simple and intuitively appealing. The regimes could be associated with phases of the business cycle, as in Hamilton (1989), or related to other phenomena such as the familiar notion of "bull' and "bear" markets. Second, the nonlinear dynamics associated with regime changes have the potential to generate interesting asset pricing implications. If the changes induce large shifts in the investment opportunity set, for example, then they can give rise to unanticipated effects, such as a non-monotonic relation between risk and expected returns (see, e.g., Whitelaw, 2000). Third, it turns out that regime-switching models have the advantage of being highly tractable from a computational perspective.

2.4.1. The regime-switching model

I develop my regime-switching generalization of IPCA by building on the regime-switching PCA framework of Kirby (2019b). Let $s_t \in (1, 2, \ldots, M)$ denote an integer-valued state variable that follows a stationary, ergodic, first-order Markov chain. The basic idea behind the proposed approach is to assume that $\boldsymbol{r}_{p,t}|s_t \sim N(\boldsymbol{\delta}_{s_t}, \boldsymbol{\Sigma}_{s_t})$ for all t, where $\boldsymbol{\delta}_i$ and $\boldsymbol{\Sigma}_i = \boldsymbol{\Gamma}_i\boldsymbol{\Gamma}_i' + \varsigma_i^2\mathbf{I}$ denote the mean vector and covariance matrix of the excess portfolio returns in the ith regime. It is well known that a first-order Markov chain can be conveniently described in terms of a vector autoregressive process (see, e.g., Hamilton, 1989). Specifically, the state transitions are governed by a process of the form

$$\boldsymbol{\xi}_{t+1} = \boldsymbol{P}'\boldsymbol{\xi}_t + \boldsymbol{\eta}_{t+1}, \tag{12}$$

where $\boldsymbol{\xi}_t$ is an $M \times 1$ vector whose ith element is given by $1_{(s_t=i)}$ for all i, $\boldsymbol{P}$ is an $M \times M$ transition matrix with typical element $p_{ij} = \Pr(s_{t+1} = j | s_t = i)$, and $\boldsymbol{\eta}_{t+1}$ is a vector martingale difference sequence, i.e., $\mathrm{E}(\boldsymbol{\eta}_{t+1} | \boldsymbol{\xi}_t, \boldsymbol{\xi}_{t-1}, \ldots, \boldsymbol{\xi}_1) = \mathbf{0}$.

Fitting this type of model is straightforward. Let $\boldsymbol{\theta}$ denote a vector that contains all of the unknown parameters. Following Hamilton (1990), one can compute the maximum likelihood estimator of $\boldsymbol{\theta}$ using a simple expectation-maximization (EM) algorithm. A detailed example of the algorithm is contained in Appendix A.[7] The EM algorithm implies that the maximum likelihood estimator of $\boldsymbol{\delta}_i$ is

$$\hat{\boldsymbol{\delta}}_i = \frac{\sum_{t=1}^T \xi_{i,t|T}(\hat{\boldsymbol{\theta}}) \boldsymbol{r}_{p,t}}{\sum_{t=1}^T \xi_{i,t|T}(\hat{\boldsymbol{\theta}})}, \tag{13}$$

where $\xi_{i,t|T}(\hat{\boldsymbol{\theta}})$ is the estimated smoothed probability that $s_t = i$ in period t (i.e., the estimated probability that $s_t = i$ given $\{\boldsymbol{r}_{p,t}\}_{t=1}^T$). It also implies that the maximum likelihood estimator of $\boldsymbol{\Gamma}_i$ satisfies $\boldsymbol{S}_i \hat{\boldsymbol{\Sigma}}_i^{-1} \hat{\boldsymbol{\Gamma}}_i = \hat{\boldsymbol{\Gamma}}_i$, where

$$\boldsymbol{S}_i = \frac{\sum_{t=1}^T \xi_{i,t|T}(\hat{\boldsymbol{\theta}})(\boldsymbol{r}_{p,t} - \hat{\boldsymbol{\delta}}_i)(\boldsymbol{r}_{p,t} - \hat{\boldsymbol{\delta}}_i)'}{\sum_{t=1}^T \xi_{i,t|T}(\hat{\boldsymbol{\theta}})}. \tag{14}$$

Hence, the kth column of the maximum likelihood estimator of $\boldsymbol{\Gamma}_i$ is given by $\hat{\boldsymbol{\gamma}}_{i,k} = \boldsymbol{u}_{i,k}(l_{i,k} - \hat{\varsigma}_i^2)^{1/2}$, where $l_{i,k}$ is the kth largest eigenvalue of $\boldsymbol{S}_i$, $\boldsymbol{u}_{i,k}$ is the associated orthonormal eigenvector, and $\hat{\varsigma}_i^2$ is the maximum likelihood estimator of ς_i^2, which is given by $\hat{\varsigma}_i^2 = (J - K)^{-1} \sum_{j=K+1}^J l_{i,j}$. Because these estimates can be computed without resorting to a secondary layer of iterative optimization, the overall computational demands of the EM algorithm are generally quite low.

[7]The example is for a model that is specified in terms of $\tilde{\boldsymbol{r}}_{p,t} = \bar{\boldsymbol{W}}^{1/2} \boldsymbol{r}_{p,t}$, where $\bar{\boldsymbol{W}} = (1/T) \sum_{t=1}^T \hat{\boldsymbol{W}}_t$. This model is discussed in Section 2.6.

2.4.2. *Regime-specific test statistics and inferences*

The parameter estimates produced by the EM algorithm can be used to conduct regime-specific asset pricing tests. Let $\hat{\boldsymbol{\mu}}_i = (\hat{\boldsymbol{\Gamma}}_i'\hat{\boldsymbol{\Gamma}}_i)^{-1}\hat{\boldsymbol{\Gamma}}_i'\hat{\boldsymbol{\delta}}_i$ denote the $J \times 1$ vector of second-pass coefficients that result from projecting $\hat{\boldsymbol{\delta}}_i$ on $\hat{\boldsymbol{\Gamma}}_i$. The $J \times 1$ vector of estimated pricing errors for regime i is given by $\hat{\boldsymbol{\alpha}}_i = \hat{\boldsymbol{\delta}}_i - \hat{\boldsymbol{\Gamma}}_i\hat{\boldsymbol{\mu}}_i$. Thus the regime-specific estimates of the RMSE are given by $RMSE_i = \sqrt{\hat{\boldsymbol{\alpha}}_i'\hat{\boldsymbol{\alpha}}_i/J}$ for $i = 1, \ldots, M$. Although the limiting distribution of $RMSE_i$ is unavailable, the computational tractability of the EM algorithm makes it easy to implement a parametric bootstrap procedure.

The basic steps for generating a bootstrap sample of size T are as follows. First, generate s_1^* from the M-point discrete distribution implied by the estimated ergodic probabilities (i.e., unconditional probabilities) of the first-order Markov chain. Next, generate $\{s_t^*\}_{t=2}^T$ using the M-point discrete distribution of $s_t|s_{t-1}$ implied by the estimated transition probabilities of the first-order Markov chain. Finally, generate $\boldsymbol{r}_{p,t}^*$ from a multivariate normal distribution with mean vector $\hat{\boldsymbol{\delta}}_{s_t^*}$ and covariance matrix $\hat{\boldsymbol{\Sigma}}_{s_t^*} = \hat{\boldsymbol{\Gamma}}_{s_t^*}\hat{\boldsymbol{\Gamma}}_{s_t^*}' + \hat{\varsigma}_{s_t^*}^2\mathbf{I}$ for $t = 1, 2, \ldots, T$.

The regime-specific pricing tests are conducted in the same fashion as the single-regime pricing tests. Let $\hat{\boldsymbol{\delta}}_i^*$ and $\hat{\boldsymbol{\Gamma}}_i^*$ denote the estimates of $\hat{\boldsymbol{\delta}}_i$ and $\hat{\boldsymbol{\Gamma}}_i$ obtained by fitting the regime-switching model to a given bootstrap sample. The associated vector of estimated pricing errors is given by $\hat{\boldsymbol{\alpha}}_i^* = \hat{\boldsymbol{\delta}}_i^* - \hat{\boldsymbol{\Gamma}}_i^*\hat{\boldsymbol{\mu}}_i^*$, where $\hat{\boldsymbol{\mu}}_i^* = (\hat{\boldsymbol{\Gamma}}_i^{*\prime}\hat{\boldsymbol{\Gamma}}_i^*)^{-1}\hat{\boldsymbol{\Gamma}}_i^{*\prime}\hat{\boldsymbol{\delta}}_i^*$. Thus the observed distribution of $RMSE_i^* = ((\hat{\boldsymbol{\alpha}}_i^* - \hat{\boldsymbol{\alpha}}_i)'(\hat{\boldsymbol{\alpha}}_i^* - \hat{\boldsymbol{\alpha}}_i)/J)^{1/2}$ can be used to obtain critical values for testing the null hypothesis of exact factor pricing in the ith regime. Similarly, the observed distribution of the second-pass R-squared statistic for the ith regime can be used to construct confidence intervals on the regime-specific explanatory power of the model for expected excess returns under the alternative hypothesis.

2.5. *Capturing nonlinear effects*

Recent studies point to nonlinearity as an important feature of the relation between firm characteristics and expected stock returns. For example, Freyberger et al. (2019) combine nonparametric regressions with advanced variable selection techniques to identify characteristics that provide incremental information about the cross-section of expected stock returns. They find that their nonparametric regressions outperforms linear regressions by a substantial margin in out-of-sample applications. Nonlinearity is also important in asset pricing tests. Kirby (2019a), for instance, uses regression-based managed portfolios that capture nonlinearity to test a range of recently-developed asset pricing models. None of the models, including those of Kroencke (2017), He et al. (2017), Stambaugh and Yuan (2017), and Lettau et al. (2019), explain more than a small fraction of the cross-sectional variation in the average portfolio returns.

Fortunately, the IPCA framework can easily be extended to capture nonlinear effects. Although the conditional factor model assumes that the cross-sectional relation between the instrumental variables and factor loadings is linear, it provides a great deal of flexibility in how the instruments are defined. As a consequence, it is compatible with nonparametric methods that rely on basis function expansions to accommodate departures from linearity. I employ an approach that combines the features of additive regression models with those of characteristic-based portfolio sorts.

Let $\{g_{1j,t}, g_{2j,t}, \ldots, g_{Lj,t}\}$ denote a set of L disjoint groups of stocks that are formed by using the jth characteristic to sort stocks for period t and grouping those with similar values of the characteristic together. If N is evenly divisible by L, for example, then $g_{1j,t}$ contains the N/L stocks that have the lowest values of the jth characteristic for period t, $g_{2j,t}$ contains the N/L stocks that have next lowest values of the jth characteristic for period t, and so forth.[8] Next, let $\boldsymbol{I}_{j,t}$ denote an $N \times L$ matrix of indicator variables whose values are defined on the basis of the groups. In particular, let the nth element of the ith column of $\boldsymbol{I}_{j,t}$ be equal to one if the nth firm is in $g_{ij,t}$ and equal to zero otherwise. Finally, let $\boldsymbol{I}_t = (\boldsymbol{I}_{2,t}, \boldsymbol{I}_{3,t}, \ldots, \boldsymbol{I}_{J,t})$ denote an $N \times Q$ matrix that contains the full set of indicator variables for period t, where $Q = L(J-1)$. [9]

Now consider the implications of specifying the elements of $\boldsymbol{I}_t$ as instrumental variables for period t. Under this approach, the associated pure play portfolio returns for period $t+1$ are the estimated coefficients for a Fama and MacBeth (1973) regression of the form

$$\boldsymbol{r}_{t+1} = \boldsymbol{X}_t \boldsymbol{\rho} + \boldsymbol{e}_{t+1}, \tag{15}$$

where $\boldsymbol{X}_t = (\mathbf{1}, \boldsymbol{I}_t)$ and $\boldsymbol{\rho}$ is an $(Q+1)\times 1$ coefficient vector. This is a simple nonparametric regression

[8] If N is not evenly divisible by L, then the groups are constructed such that the stocks are as evenly distributed across groups as possible.

[9] The indicators are indexed starting with $j = 2$ because $j = 1$ corresponds to the constant instrument.

specification: an additive model that employs zero-degree B-splines with quantile-based knots. But it is not identified because $\mathbf{1}$ is in the column space of $\boldsymbol{I}_{j,t}$ for every value of j. One way to achieve identification would be to set the coefficient for one column of $\boldsymbol{I}_{j,t}$ to be zero for all j. Although this is a common way to avoid the "dummy variable trap" in regressions that employ indicator variables, the estimated coefficients are easier to interpret under the following identification strategy.

Let $\boldsymbol{R}$ denote a $(Q+1)\times(J-1)$ matrix used to impose restrictions of the form $\boldsymbol{R}'\boldsymbol{\rho} = \mathbf{0}$. I specify $\boldsymbol{R}$ such that the L elements of $\boldsymbol{\rho}$ that correspond to $\boldsymbol{I}_{j,t}$ are constrained to sum to zero for every value of $j \in (2, \ldots, J)$. As a result, the restricted ordinary least squares (OLS) estimate of the intercept (i.e., the first element of $\hat{\boldsymbol{\rho}}$) is simply the excess return on the equally weighted market index (EWM) for the set of stocks under consideration. In addition, the restricted OLS estimate of the coefficient associated with the ith column of $\boldsymbol{I}_{j,t}$ has a simple interpretation. It is the estimated marginal effect for period $t+1$ of changing the value of the jth characteristic from its average value across all stocks in period t to its average value for the stocks in group $g_{ij,t}$.

Note that $\boldsymbol{R}'\boldsymbol{\rho} = \mathbf{0}$ implies that the restricted OLS estimator must lie in the column space of the orthogonal complement of $\boldsymbol{R}$, which is a $(Q+1)\times(Q-J+2)$ matrix. This property implies that the estimator has a tractable closed-form representation (Faliva and Zoia, 2002). In particular, one can show that $\hat{\boldsymbol{\rho}} = \boldsymbol{R}_{\perp}(\boldsymbol{X}_t\boldsymbol{R}_{\perp})^{+}\boldsymbol{r}_{t+1}$, where $\boldsymbol{R}_{\perp}$ denotes the orthogonal complement of $\boldsymbol{R}$ and $(\boldsymbol{X}_t\boldsymbol{R}_{\perp})^{+} = (\boldsymbol{R}_{\perp}'\boldsymbol{X}_t'\boldsymbol{X}_t\boldsymbol{R}_{\perp})^{-1}\boldsymbol{R}_{\perp}'\boldsymbol{X}_t'$ denotes the left pseudoinverse of $\boldsymbol{X}_t\boldsymbol{R}_{\perp}$. Because the columns of $\boldsymbol{R}_{\perp}$ form a basis for the null space of $\boldsymbol{R}'$, they are given by the orthonormal eigenvectors associated with the non-zero eigenvalues of the symmetric and idempotent matrix $\mathbf{I} - \boldsymbol{R}(\boldsymbol{R}'\boldsymbol{R})^{-1}\boldsymbol{R}'$.

In view of these results, equation (15) implies that the $(Q+1)\times 1$ vector of excess returns for the nonparametric pure play portfolios can be expressed as $\boldsymbol{R}_{\perp}\boldsymbol{r}_{p,t+1}$, where $\boldsymbol{r}_{p,t+1} = (\boldsymbol{X}_t\boldsymbol{R}_{\perp})^{+}\boldsymbol{r}_{t+1}$ is a lower-dimensional vector of excess portfolio returns. Because $\boldsymbol{R}_{\perp}$ is constructed such that $\boldsymbol{R}_{\perp}'\boldsymbol{R}_{\perp} = \mathbf{I}$, it follows that $\boldsymbol{r}_{p,t+1}$ can be used in place of $\boldsymbol{R}_{\perp}\boldsymbol{r}_{p,t+1}$ to conduct the asset pricing tests. If, for example, the vector of estimated pricing errors for the lower-dimensional set of managed portfolios is given by $\hat{\boldsymbol{\alpha}}$, then the vector of estimated pricing errors for the pure play portfolios is given by $\boldsymbol{R}_{\perp}\hat{\boldsymbol{\alpha}}$. Because the value of $\hat{\boldsymbol{\alpha}}'\boldsymbol{R}_{\perp}'\boldsymbol{R}_{\perp}\hat{\boldsymbol{\alpha}}$ is numerically identical to that of $\hat{\boldsymbol{\alpha}}'\hat{\boldsymbol{\alpha}}$, I simply use the RMSE for the lower-dimensional set of managed portfolios to conduct the tests.

2.6. *Additional modeling issues*

Before turning to the empirical analysis, I address a couple of additional modeling issues. One relates to the interpretation of the parameter estimates. Suppose by way of illustration that $\boldsymbol{C}_t$ is replaced with $\boldsymbol{C}_t\boldsymbol{A}$ for all t, where $\boldsymbol{A}$ is a nonsingular $J \times J$ matrix. Because this is a full rank transformation, it will not have any effect on the fitted values for the Fama and MacBeth (1973) regressions. Thus it will not affect the outcome of the asset pricing tests. But it will change the nature of the managed portfolios and the estimates of the factor loadings. As a result, something as innocuous as changing units of measurement will affect the magnitude of the estimated pricing errors. This suggests that some type of standardization would be beneficial.

To address this issue, I standardize every column of $\boldsymbol{C}_t$ except the first to have a mean of zero and variance of one for each time period. Thus the estimated intercept in the Fama and MacBeth (1973) regression for period t is given by $(1/N)\sum_{n=1}^{N} r_{n,t}$, the excess EWM return. Standardizing the columns of $\boldsymbol{C}_t$ has two primary benefits. It implies that all of the characteristics are measured in the same units, thereby promoting direct comparisons across characteristics. In addition, it implies that the cross-sectional distribution of the pure play portfolio weights has a constant mean vector and constant covariance matrix, so the weights should be well behaved for cases in which some of the characteristics display time trends or other forms of nonstationary behavior.[10]

The other issue has to do with the relation between the parameter estimates obtained via the direct solution method and those obtained via the ALS algorithm of Kelly et al. (2019). To see the nature of this relation more clearly, consider a scenario in which $\hat{\boldsymbol{W}}_t$ takes the same value for every time period. In this case, the problem in equation (7) is equivalent to

$$\min_{\boldsymbol{\delta},\boldsymbol{\Gamma},\boldsymbol{F}} \frac{1}{T}\sum_{t=1}^{T}(\boldsymbol{r}_{p,t} - \boldsymbol{\delta} - \boldsymbol{\Gamma}\boldsymbol{f}_t)'\bar{\boldsymbol{W}}(\boldsymbol{r}_{p,t} - \boldsymbol{\delta} - \boldsymbol{\Gamma}\boldsymbol{f}_t) + \frac{1}{NT}\sum_{t=1}^{T}\boldsymbol{e}_t'\boldsymbol{e}_t, \tag{16}$$

where $\bar{\boldsymbol{W}} = (1/T)\sum_{t=1}^{T}\hat{\boldsymbol{W}}_t$. Thus the estimates of $\boldsymbol{\delta}$ and $\boldsymbol{\Gamma}$ can be obtained solving a PCA problem

[10] Using indicator variables to construct nonparametric pure play portfolios delivers the same benefits. But standardization is still useful in this case in order to facilitate comparisons between the parametric and nonparametric results. Hence, I standardize the columns of $\boldsymbol{I}_t$ to have a unit cross-sectional second moment before constructing the managed portfolios used for the asset pricing tests.

of the form

$$\min_{\boldsymbol{\kappa},\boldsymbol{\Lambda},\boldsymbol{F}} \frac{1}{T}\sum_{t=1}^{T}(\tilde{\boldsymbol{r}}_{p,t}-\boldsymbol{\kappa}-\boldsymbol{\Lambda}\boldsymbol{f}_t)'(\tilde{\boldsymbol{r}}_{p,t}-\boldsymbol{\kappa}-\boldsymbol{\Lambda}\boldsymbol{f}_t), \tag{17}$$

where $\tilde{\boldsymbol{r}}_{p,t} = \bar{\boldsymbol{W}}^{1/2}\boldsymbol{r}_{p,t}$, $\boldsymbol{\kappa} = \bar{\boldsymbol{W}}^{1/2}\boldsymbol{\delta}$, and $\boldsymbol{\Lambda} = \bar{\boldsymbol{W}}^{1/2}\boldsymbol{\Gamma}$. Specifically, they can be obtained by setting $\hat{\boldsymbol{\delta}} = \bar{\boldsymbol{W}}^{-1/2}\hat{\boldsymbol{\kappa}}$ and $\hat{\boldsymbol{\Gamma}} = \bar{\boldsymbol{W}}^{-1/2}\hat{\boldsymbol{\Lambda}}$, where $\hat{\boldsymbol{\kappa}}$ and $\hat{\boldsymbol{\Lambda}}$ solve the PCA problem. Note, however, that reproducing the ALS results would require a nonstandard choice of identification restrictions. For example, if the goal is to construct an ALS estimate of $\boldsymbol{\Gamma}$ such that $\hat{\boldsymbol{\Gamma}}'\hat{\boldsymbol{\Gamma}}$ is diagonal with descending diagonal elements, then the PCA estimate of $\boldsymbol{\Lambda}$ would have to be constructed such that $\hat{\boldsymbol{\Lambda}}'\bar{\boldsymbol{W}}^{-1}\hat{\boldsymbol{\Lambda}}$ is diagonal with descending diagonal elements.

Because I want to compare the results produced by the ALS and direct solution methods, I use the following procedure to conduct the econometric analysis for the single-regime version of IPCA. First, I multiply equation (8) by $\bar{\boldsymbol{W}}^{1/2}$ to obtain a static factor model of the form

$$\tilde{\boldsymbol{r}}_{p,t+1} = \boldsymbol{\kappa} + \boldsymbol{\Lambda}\boldsymbol{f}_{t+1} + \tilde{\boldsymbol{\varepsilon}}_{p,t+1}, \tag{18}$$

where $\tilde{\boldsymbol{\varepsilon}}_{p,t+1} = \bar{\boldsymbol{W}}^{1/2}\boldsymbol{\varepsilon}_{p,t+1}$. Next, I construct the PCA estimates of $\boldsymbol{\kappa}$ and $\boldsymbol{\Lambda}$, which should be similar to the estimates of $\boldsymbol{\delta}$ and $\boldsymbol{\Gamma}$ obtained by solving the minimization problem in equation (7) as long as the values of $\hat{\boldsymbol{W}}_t$ are reasonably stable over the sample period. Finally, I multiply the PCA estimates by $\bar{\boldsymbol{W}}^{-1/2}$ to obtain the estimated pricing errors for the pure play portfolios under the direct solution method and implement the asset pricing tests as discussed in Section 2.2.

The approach used for the regime-switching generalization of IPCA is similar. First, I fit a regime-switching model to $\{\tilde{\boldsymbol{r}}_{p,t}\}_{t=1}^{T}$ by assuming that $\tilde{\boldsymbol{r}}_{p,t}|s_t \sim N(\boldsymbol{\kappa}_{s_t}, \boldsymbol{\Omega}_{s_t})$, where $\boldsymbol{\Omega}_{s_t} = \boldsymbol{\Lambda}_{s_t}\boldsymbol{\Lambda}'_{s_t} + \tau^2_{s_t}\mathbf{I}$ has an underlying factor structure (see Appendix A for additional details).[11] Next, I multiply the estimates $\{\boldsymbol{\kappa}_i\}_{i=1}^{M}$ and $\{\boldsymbol{\Lambda}_i\}_{i=1}^{M}$ by $\bar{\boldsymbol{W}}^{-1/2}$ to obtain the estimates of $\{\boldsymbol{\delta}_i\}_{i=1}^{M}$ and $\{\boldsymbol{\Gamma}_i\}_{i=1}^{M}$ for the pure play portfolios. Finally, I carry out the asset pricing tests using regime-specific estimates of the vector of pricing errors for the pure play portfolios. I also conduct tests using the managed portfolios used to fit the the model to assess whether the results are sensitive to the choice of portfolios.

[11] I fit the model to $\{\tilde{\boldsymbol{r}}_{p,t}\}_{t=1}^{T}$ instead of to $\{\boldsymbol{r}_{p,t}\}_{t=1}^{T}$ for the sake of promoting comparisons to the single-regime results produced by the ALS algorithm. The latter approach is an obvious choice more broadly.

3. Data and Preliminary Analysis

Researchers have identified a host of firm-specific variables over the past three decades that appear to have some ability to capture cross-sectional variation in expected stock returns (see, e.g., Harvey et al., 2016). This raises an important question. Which of these characteristics should be employed as instruments in the IPCA framework? I use a relatively small set of characteristics that are chosen based on the results of Freyberger et al. (2019). These authors apply an advanced variable selection procedure to a fairly comprehensive set of 62 characteristics in order to identify those that provide incremental information about the cross-section of expected returns. Their analysis is distinguished from earlier efforts along these lines, such as that of Green et al. (2017), by their use of nonparametric methods that do not impose strong functional-form assumptions.

Freyberger et al. (2019) report that only 13 of the 62 characteristics survive the selection procedure using data for 1965 to 2014. I use all of these characteristics and include three others that are identified in various subperiod tests.[12] Specifically, I use the log of market equity (LME), the log of the book-to-market equity ratio (LBM), an eleven-month stock return (MOM), a six-month stock return (IMOM), a one-month stock return (REV), the log of realized monthly volatility (LVOL), the log of monthly share turnover (LTO), standardized unexpected monthly volume (SUV), the annual growth rate of total assets (INV), the ratio of annual sales to market equity (SM), the ratio of annual net operating assets to total assets (NOA), the stock price relative to its 52 week high price (P52H), the annual industry-adjusted profit margin (APM), the annual return on cash (ROC), a fiscal-year-based measure of annual share issuance (NSFY), and a measure of annual share issuance for the most recent 12 months (NSTY). Variable definitions are given in Appendix B, which also describes how I match characteristics that depend on balance sheet items with stock returns.

The data are drawn from standard sources. I obtain monthly stock returns for the Center for Research in Security Prices (CRSP), annual balance sheet items from the Compustat Annual Industrial file, and the monthly risk free rate from the Ken French data library. Firms are required

[12]There are a few cases in which I employ a transformed version of the characteristic used by Freyberger et al. (2019). For example, I use the logarithm of realized volatility rather than the volatility itself because the log transformation reduces the influence of values in the long right tail of this characteristic's cross-sectional distribution.

to have at lest two years of data on Compustat to alleviate backfilling bias. The sample period is April 1967 to December 2018.[13] Table 1 reports descriptive statistics for the firm characteristics. These statistics are for the sample used to conduct the baseline empirical tests.[14] To assess whether my findings are sensitive to sample composition, I also employ a secondary sample (the "no-micro-caps" sample) that excludes stocks whose market equity is smaller than the 20th percentile of the NYSE market-equity distribution. Stocks that have a missing return or a missing value for any of the 16 characteristics for a given month are excluded from the analysis for that month.

3.1. Fama and MacBeth (1973) regressions

Table 2 summarizes the results obtained by fitting monthly cross-sectional regessions of excess stock returns on the characteristics and on characteristic-based indicator variables for the $L = 4$ case (i.e., quartile indicators). The findings are in line with those of prior research. In particular, the regression evidence provides broad support for the view that the characteristics provide incremental information about the cross-section of expected returns. Using the characteristics as regressors, for example, all but three of the Fama and MacBeth (1973) t-statistics for the average estimated slope coefficients are statistically significant at the 5% level. The exceptions are the average estimated slopes for the IMOM, LTO, and APM characteristics. Although this finding suggests that these characteristics have negligible incremental explanatory power, it is not dispositive given that the regressions in question impose a strong a priori assumption on the functional form of the relation between the characteristics and expected stock returns.

Indeed, the results for the additive regressions indicate that all 16 characteristics provide incremental information about the cross-section of expected returns. There are certainly instances in which one or more of the indicators for a given characteristic is statistically insignificant at the 5% level. But at least one of the indicators is statistically significant for every characteristic. Although the indicators for the first and fourth quartiles tend to have the largest absolute t-statistics,

[13]Because the nonparametric regressions have a large number of regressors, I require the no-micro-caps sample to contain at least 500 stocks for every month considered. This screen results in an April 1967 start date.

[14]The statistics are similar to those reported in prior studies that examine one or more of the characteristics. Exact matches are unlikely for a variety of reasons, such as differences in sample periods, winsorization or trimming at different levels, minor differences in variable definitions, and differences in the treatment of missing values.

there are exceptions. For example, the t-statistics for the log realized volatility are 0.45, 6.77, 3.42, and -4.98, which suggests that the cross-sectional relation between realized volatility and expected stock returns is not monotonic. On the whole, the results of the nonparametric regressions appear to be broadly consistent with those reported by Freyberger et al. (2019).

3.2. Baseline asset pricing tests

Table 3 examines the pricing performance of various specifications of the single-regime version of the IPCA factor model. This provides a baseline set of results that can be compared with those produced by the regime-switching model. Panels A and B report estimates of the RMSE for two sets of managed portfolios. The estimates in the first seven columns are for the 17 pure play portfolios obtained via linear Fama-MacBeth regressions (the "linear pure play portfolios"). Those in the remaining columns are for the 49 portfolios obtained via additive regressions that use quartile indicator variables and restrict the four coefficients associated with each characteristic to sum to zero (the "additive pure play portfolios").

First I discuss the results for the linear pure play portfolios. Panel A shows the results obtained using the ALS algorithm of Kelly et al. (2019). The $K = 0$ column reports the estimated RMSE with zero factors in the model (i.e., the value of $(\hat{\delta}'\hat{\delta}/17)^{1/2}$). It is 0.36% per month using the sample that includes all stocks and 0.22% per month using the sample that excludes micro-cap stocks. The next six columns show the estimated RMSE for values of K that range from one to six.

The estimated RMSE for the one-factor model is 0.29% per month, which suggests that the first factor does relatively little to explain the cross-section of average portfolio returns. Adding a second factor to the model reduces the estimated RMSE to 0.27% per month. Although there is no additional reduction in the estimated RMSE for the $K = 3$ or $K = 4$ specifications, moving to a five-factor model causes the estimated RMSE to fall to 0.15% per month. With the addition of a sixth factor, it falls again to 0.11% per month. The same general pattern is evident for the no-micro-caps sample. The estimated RMSE for the one-factor specification is 0.17% per month, which again suggests that the first factor does relatively little to explain the cross-section of average returns. The estimated RMSEs for the two-, three-, and four-factor specifications are slightly lower, but larger reductions occur with the addition of the fifth and sixth factor to the model.

Panel B replicates the analysis of panel A using the direct solution method. The differences in the estimated RMSEs across the two panels are very minor, which is the expected finding in light of the discussion in Sections 2.3 and 2.6. Using all stocks, for example, the estimated RMSE produced by the direct solution method is 0.29% per month for the one-factor model and 0.12% per month for the six-factor model. Because the estimated RMSEs in panel B closely mirror those in panel A, it appears that the direct solution method represents a promising alternative to ALS for estimating parameters and conducting IPCA-based asset pricing tests.

The critical values in panel C lend additional context to these findings. The critical values are based on the bootstrap distribution of the estimated RMSE under the null hypothesis of exact factor pricing. As might be anticipated, the results indicate that the one-factor model is sounded rejected. The critical values for the 1% significance level are only 0.06% (ALS method) and 0.05% (direct solution method) per month. However, the critical values rise as more factors are added, which causes the evidence against exact factor pricing to become weaker as K increases. Indeed, using all stocks yields 1% critical values for the six-factor model that exceed the estimated RMSEs for the model. Of course some might question how well a six-factor model aligns with the conventional view that the number of sources of priced risk should be fairly small. Nonetheless, the general thrust of the evidence is reasonably favorable to the hypothesis that the 16 firm characteristics serve as proxies for priced covariance risk.

Turning to the results for the additive pure play portfolios, I find that the $K = 0$ values of the estimated RMSE are a good deal smaller than those for the linear-regression portfolios: 0.19% per month using all stocks and 0.13% per month using the no-micro-caps sample. This finding appears to be driven by the portfolios associated with the indicator variables for the middle two quartiles. Because these portfolios generally have average excess returns that are closer to zero than the portfolios associated with the indicator variables for the top and bottom quartiles, they help to bring down the average value of the squared deviations of the elements of $\hat{\boldsymbol{\delta}}$ from zero.

The general pattern of the estimated RMSEs with $K > 0$ is similar to that observed for the linear pure play portfolios. The ALS method yields an estimated RMSE for the one-factor model of 0.15% per month using all stocks and 0.10% per month using the no-micro-caps sample. In comparison, the corresponding values under the direct solution method are 0.15% and 0.09% per month. But all of the estimated RMSEs fall as K increases. With six factors, for example, the estimated RMSEs falls to 0.10% (all stocks) and 0.07% (no micro-caps) using the ALS method, whereas the direct solution method yields values of 0.10% (all stocks) and 0.06% (no micro-caps).

More broadly, however, there is some divergence between the inferences for the additive pure play portfolios and those for the linear pure play portfolios. Using the additive pure play portfolios, the estimated RMSEs for the five- and six-factor parameterizations exceed the 1% critical values in almost every case. Although this finding suggests that the conditional factor model is ultimately unsuccessful in explaining the nonlinear effects captured by the additive regressions, the results still point to substantial linkages between firm charactersitics and priced covariance risk.

3.3. An initial look at the regime-switching evidence

Table 4 reports selected results for a two-regime version of the regime-switching model. I fit the model to $\{\tilde{\boldsymbol{r}}_{p,t}\}_{t=1}^{T}$ using values of K from one to six as described in Section 2.4.1. However, my discussion focuses on the estimates of the regime-specific means and volatilities for a single portfolio: the equally-weighted market index. These estimates are of particular interest because the literature contains a number of prior studies that use two-state Markov switching models to characterize bull and bear market regimes (see, e.g., Gordon and St-Amour, 2000; Maheu and McCurdy, 2000; Pagan and Sossounov, 2003; Guidolin and Timmermann, 2005; Kole and van Dijk, 2017). I can therefore use the results of these studies as a frame of reference for assessing my findings.

The one-factor parameterization of the model clearly yields evidence of two distinct market regimes. The mean and volatility of the excess market returns are estimated to be 0.11% and 4.92% for the first regime and 2.77% and 7.66% for the second regime. These figures translate into annualized values of 1.3% and 17.0% for the first regime and 33.2% and 26.5% for the second regime. The 95% confidence intervals for the regime-specific means indicate that the estimates are reasonably precise. The upper confidence limit for regime one is 0.53%, which is well below the lower confidence limit of 1.48% for regime two. Note, however, that the spread between the estimated means is considerably smaller for the no-micro-caps sample, which suggests that market capitalization may play a meaningful role in the subsequent asset pricing tests.

Moving to a two-factor parameterization causes the estimated mean for regime one to rise and that for regime two to fall. The estimates are 0.50% and 1.88% for the sample that includes all stocks and 0.61% and 0.93% using the no-micro-caps sample. Moving to a three-factor parameterization leads to a further narrowing of the spread between the estimated means. But the spread widens with the addition of the fourth, fifth, and sixth factors. The six-factor parameterization therefore produces estimated means that are similar to those produced by the one-factor parameterization.

Regardless of the choice of K, the estimated mean and estimated volatility for the first regime are lower than the estimated mean and estimated volatility for the second regime. This finding contrasts sharply with those of prior studies that fit univariate two-regime specifications to market index returns. Their model fitting results imply that one of the regimes displays a high mean and low volatility (the bull regime) and the other displays a low mean and high volatility (the bear regimes). For example, Maheu and McCurdy (2000, p. 104) report that "the bull-market label refers to the high-return, low-volatility state, whereas the bear-market label refers to the low-return, high-volatility state of the stock market." However, I find that the picture changes once the relation between firm characteristics, expected returns, and volatility is taken into account.

To investigate further, I consider the estimated smoothed probabilities of the high-volatility regime that are obtained by fitting a univariate two-regime specification to the excess monthly returns on the equally-weighted market index. These probabilities are plotted in the top panel of Figure 1. The rectangular shaded areas on the plot show ex-post bear market periods under a common classification rule that relies strictly on observed market returns (Lunde and Timmermann, 2004). The rule is to declare the start of a bear market once the month-end value of the market index falls by at least 15% from its most recent high and to declare the start of a bull market once the month-end value of the market index rises by at least 20% from its most recent low.

The pattern of the estimates produced by the univariate fit is consistent with that reported by prior studies. The mean and volatility of excess market returns are estimated to be 1.16% and 3.96% in regime one and 0.09% and 8.37% in regime two for the sample that includes all stocks. In addition, the periods during which the estimated smoothed probability of the high volatility regime is close to one line up well with the ex post bear market periods. This finding is also consistent with the results reported in prior studies (see, e.g., Kole and van Dijk, 2017).

The remaining two panels of Figure 1 show the estimated smoothed probabilities of the high-volatility regime that correspond to the Table 4 results. The middle panel is for the $K = 1$ case and the bottom panel is for the $K = 6$ case. These plots look quite different than the plot produced by the univariate fit to monthly excess market returns. Not only do the estimated regimes appear to be considerably less persistent than those identified by the univariate fit, there also appears to be a much weaker correlation between the estimated smoothed probabilities of the high volatility regime and the bear market periods identified by the ex-post classification rule.

Figure 2 helps to explain the contrast between the model fitting results for the univarate and multivariate specifications. It plots the regime-specific estimates of the expected excess returns for the linear pure play portfolios side by side to highlight the magnitude of the differences across regimes. In almost all cases, the estimated expected excess return for the high-volatility regime is larger in magnitude than that for the low-volatility regime. The estimates suggest that market portfolio undergoes the largest shift in expected excess returns across regimes. However, the evidence points to substantial shifts for the LME, REV, SUV, LTO, INV, and P52H portfolios as well.

The main takeaways from the preliminary analysis of the regime-switching evidence can be summarized succinctly. The model fitting results point to the presence of a regime in which the linear-regression portfolios have both high volatility and high expected excess returns. In addition, they indicates that this regime is relatively short lived. Using all stocks, for example, the estimated expected regime durations in Table 4 range from a low of 2.3 months for $K = 1$ to a high of 3.2 months for $K = 5$. The estimates are slighted higher for the no-micro-caps sample. Still, the upper end of the 95% confidence interval for expected duration never exceeds 5.1 months.

4. Main Empirical Results

The analysis thus far has yielded several interesting insights. First, there are stretches of time during which individual stocks have both high volatility and high expected excess returns. Because this insight emerges from the dynamics of returns for managed portfolios that have characteristic-dependent weights, it seems to be broadly consistent with the view that the characteristics proxy for factor loadings, thereby capturing cross-sectional differences in priced covariance risk. Second, any link between the explanatory power of characteristics and the familiar concept of bull and bear markets is relatively weak. The high-volatility regime has a relatively short expected duration and fails to display any of the features typically associated with bear markets. Third, regime changes are associated with substantial shifts in the investment opportunity set.

I now look more closely at the asset pricing implications of the regime-switching evidence by conducting regime-specific analogs of the baseline asset pricing tests. Table 5 reports the results of the regime-specific pricing tests for the managed portfolios constructed via linear cross-sectional regressions. Panel A is for the 17 portfolios used to fit the regime-switching model and panel B is for the 17 linear pure play portfolios. The first two columns report the average value of the R-squared statistics that capture the regime-specific explanatory power of the factors for the excess portfolio returns.[15] The remaining columns report regime-specific values of the estimated RMSEs, the associated 1% critical values for testing the hypothesis of exact factor pricing, and regime-specific 95% confidence intervals for the second-pass R^2 statistic.

I begin by discussing the results for the managed portfolios used to fit the model (panel A). Several aspects of the results obtained using all stocks stand out immediately. First, the estimated RMSE with zero factors in the model, which measures the distance of the estimated expected portfolio returns from the origin, is much larger in the high-volatility regime than in the low-volatility regime: 0.97% versus 0.23% per month. Second, the one-factor parameterization of the model explains a substantial fraction of the cross-sectional variation in the estimated expected excess portfolio returns, but its explanatory power is confined to the high-volatility regime. The estimated RMSE for this regime is 0.48% per month, which is about half as large as that for the $K = 0$ case, whereas the estimated RMSE for the low-volatility regime is the same as that obtained with $K = 0$. Third, the second-pass R-squared statistics are considerably higher for the high-volatility regime than for the low-volatility regime, which points to a stronger cross-sectional relation between the characteristics

[15]For example, the jth diagonal element of $\hat{\boldsymbol{\Gamma}}_i'\hat{\boldsymbol{\Gamma}}_i$ represents the estimated variance of the explained component of the excess return for the jth pure play portfolio in the ith regime. This estimated explained variance can be divided by the jth diagonal element of $\boldsymbol{S}_i(\hat{\boldsymbol{\theta}})$ to obtain an R-squared statistic for the jth portfolio in the ith regime. The values reported in the tables are obtained by averaging the regime-specific R-squared statistics computed in this manner.

and estimated expected excess returns in the high-volatility regime.

The estimated RMSEs produced by the two-, three-, and four-factor parameterizations of the model do not vary much for a given regime. The values range from 0.22% to 0.23% per month for the low-volatility regime and from 0.40% to 0.42% per month for the high-volatility regime. But the estimated RMSEs fall by a substantial amount when the fifth and sixth factors are included in the model. The six-factor parameterization, for instance, produces an estimated RMSE of 0.09% per month for the low-volatility regime and 0.23% per month for the high-volatility regime.

The critical values reported next to the estimated RMSEs indicate that the hypothesis of exact factor pricing is rejected at the 1% significance level for both regimes, regardless of the choice of K. This is an initial indication that the regime-specific tests produce more definitive evidence against this hypothesis than the single-regime tests. Despite this finding, the results still point to a fairly strong linkage between firm characteristics and covariance risk in each of the regimes. It is true that the six-factor parameterization of the model produces an estimated RMSE for the high-volatility regime that is more than double that for the low-volatility regime. In both absolute and proportional terms, however, moving from zero to six factors leads to a more pronounced reduction in the estimated RMSE for the high-volatility regime than for the low-volatility regime.

As anticipated, the estimated RMSEs are generally lower for the sample that excludes micro-cap stocks. The second-pass R-squared statistics also show an increase in most cases for the low-volatility regime and a small decrease for the high-volatility regime. But there is nothing in the results to indicate that micro-cap stocks drive the outcome of the pricing tests. For example, the estimated RMSE for $K = 0$ is 0.19% per month for the low-volatility regime and 0.48% per month for the high-volatility regime. Using a one-factor parameterization, the values fall to 0.16% and 0.28% per month, respectively. So the explanatory power of the first factor is largely confined to the high-volatility regime, as is the case using all stocks.

Similarly, the estimated RMSEs decline as the second, third, fourth, and fifth factors are added to the model. The main deviation from the pattern observed using all stocks is that adding the fourth factor produces a larger decline in the estimated RMSE for the low-volatility regime and adding the fifth factor produces a smaller decline in the estimated RMSE for the high-volatility regime. Using six factors, the estimated RMSE is 0.06% per month for the low-volatility regime and 0.18% per month for the high-volatility regime. These figures are statistically significant at the 1% level, which is the same finding obtained with six factors using all stocks.

Now consider the results for the linear pure play portfolios (panel B). In general, they look similar to those for the managed portfolios used to fit the model. The estimated RMSE with $K = 0$ is much larger in the high-volatility regime than in the low-volatility regime: 0.93% versus 0.23% per month. In addition, the explanatory power of the one-factor parameterization is confined to the high-volatility regime, where it again produces an estimated RMSE that is about half as large as that for the $K = 0$ case. The two-, three-, and four-factor parameterizations perform only slightly better than the one-factor parameterization, but the estimated RMSEs fall by a substantial amount when the fifth and sixth factors are included in the model. For example, the six-factor parameterization produces an estimated RMSE of 0.09% per month for the low-volatility regime and 0.26% per month for the high-volatility regime. Both estimates are statistically significant at the 1% level.

Although there are some minor differences in the results across panels, none of them are particularly noteworthy. The values of the average R-squared statistics reported in the first two columns are a little lower in panel B than in panel A, as are the values of the second-pass R-squared statistics. But this finding is not surprising given that the results in panel A are for the portfolios whose excess returns are used to fit the model. In short, the tests for the linear pure play portfolios convey the same message as those for the portfolios used to fit the model.

Table 6 reports the results of the regime-specific pricing tests for the managed portfolios constructed via additive cross-sectional regressions using quartile-based indicator variables. The results in panel A are for the 49 portfolios used to fit the regime-switching model. The estimated RMSE with $K = 0$ is 0.13% per month for the low-volatility regime and 0.48% per month for the high-volatility regime. These values are about half as large as the corresponding values in panel A of Table 5, which mirrors the finding produced by the single-regime model (see Table 3).

Increasing the value of K causes the estimated RMSEs to fall in a now familiar manner: slowly with the addition of the second, third, and fourth factors and then somewhat more rapidly with the addition of the fifth and sixth factors. The estimated RMSEs for the six-factor parameterization are

0.08% per month for the low-volatility regime and 0.14% per month for the high-volatility regime. These figures are statistically significant at the 1% level. The results for the additive pure play portfolios in panel B are similar, both in terms of the values of the estimated RMSEs and the overall pattern of the estimates. Thus all of the results are broadly consistent with those in Table 5.

But further comparisons reveal a couple of signs that the portfolios constructed via additive regressions may pose a more difficult pricing challenge than those constructed via linear regressions. Take the results obtained for the six-factor parameterization using all stocks. The second-pass regression for the low-volatility regime produces R-squared statistics of 86% (panel A) and 84% (panel B) in Table 5, whereas the corresponding statistics in Table 6 are 59% and 45%. In addition, the ratio of the estimated RMSE to the 1% critical value is higher for additive regression portfolios than for the linear regression portfolios. Although the differences are not dramatic, they are another indication that it may be important to account for nonlinear effects in assets pricing tests.

In the end, two main conclusions emerge from the regime-specific pricing tests that are summarized in Tables 5 and 6. The first is that the findings are less favorable to the hypothesis of exact factor pricing than those of Kelly et al. (2019). This hypothesis is rejected at the 1% significance level for any regime-switching parameterization of the conditional factor model that includes six or fewer factors. The second is that despite these rejections, the evidence indicates that a substantial fraction of the explanatory power of firm characteristics for expected stock returns stems from their ability to capture cross-sectional variation in exposures to latent risk factors. The tests also suggest that the relation between the characteristics and priced covariance risk is stronger in the high-volatility regime, especially for the case in which micro-cap stocks are included in the analysis.

4.1. A closer look at the factors

One of the features of the regime-switching model is that it allows the factors themselves to differ across regimes. Thus it should be possible to lend further context to the results by constructing regime-specific estimates of the latent factor realizations. Using the linear pure play portfolios, for example, the OLS estimator of $\boldsymbol{f}_t$ for the ith regime can be expressed as $\hat{\boldsymbol{\Gamma}}_i^{+}\boldsymbol{r}_{p,t}$, where $\hat{\boldsymbol{\Gamma}}_i^{+} = (\hat{\boldsymbol{\Gamma}}_i'\hat{\boldsymbol{\Gamma}}_i)^{-1}\hat{\boldsymbol{\Gamma}}_i'$ denotes the left pseudoinverse of the matrix of loadings for the ith regime. This is just a regime-specific analog of the vector of estimated factor realizations that would be obtained by applying PCA to the sample covariance matrix of $\boldsymbol{r}_{p,t}$. I examine the estimates implied by the identification restriction $\hat{\boldsymbol{\Gamma}}_i'\hat{\boldsymbol{\Gamma}}_i = \mathbf{I}$. Hence, the jth row of $\hat{\boldsymbol{\Gamma}}_i^+$, which contains the weights placed on the various portfolios to estimate the realization of the jth factor for the ith regime, has unit length for all j.

Figure 3 contains six panels that show the weights used to compute the estimated factor realizations for the low-volatility regime (i.e, each panel plots a row of $\hat{\boldsymbol{\Gamma}}_1^+$). Several of the plots lend themselves to straightforward interpretation. For example, it is apparent that the first factor is essentially just the excess return on the EWM. This finding is consistent with the evidence from prior PCA studies, which typically report that market-wide movements are a dominant factor in explaining common variation in individual stock returns (see, e.g., Ait-Sahalia and Xiu, 2019). In addition, the interpretation of the weights for the second, third, and fourth factors is reasonably clear. It appears that the second factor is closely associated with firm size (market capitalization), the third factor is closely associated with momentum, and the fourth factor is is closely associated with the closeness of the stock price to its 52 week high price.

In comparison, the weights for the fifth and sixth factors are much less transparent. Those for the fifth factor imply relatively large long positions in the LBM and MOM portfolios that are counterbalanced by relatively large short positions in the REV and P52H portfolios. This combination does not seem particularly intuitive. The weights for the sixth factor also entail substantial positions in a range of portfolios. There are relatively large short positions in the REV, LVOL, LTO, SM, and NOA portfolios that are counterbalanced by long positions in the SUV and INV portfolios. It appears, therefore, that the fifth and sixth factors, which play a crucial role in boosting the pricing performance of the model in the low-volatility regime, defy easy interpretation.

Figure 4 shows the analogous plots for the high-volatility regime. Once again it appears that the first factor is essentially just the excess return on the EWM. For the second factor, however, the weights look quite different than those for the low-volatility regime. Almost all of the weight is placed on the linear pure play portfolio for the P52H characteristic. Because the estimated expected excess return for this portfolio is positive in the high-volatility regime (see Figure 2), it appears that the second factor for this regime is tied to some type of price momentum that is distinct from that captured by the linear pure play portfolio for the MOM characteristic.

The weights for the third through sixth factors are also notably different than in the low-volatility regime. But this finding is not particularly surprising. Suppose for the sake of illustration that the true unobserved factors were the same in each regime. Because the identification restriction $\hat{\mathbf{\Gamma}}_i'\hat{\mathbf{\Gamma}}_i = \mathbf{I}$ implies that the estimated factors are ordered according to their estimated variance (from high to low), there still might not be a direct correspondence between the estimated factors across regimes. If sorting the true factors by their regime-specific variances produces a different ordering in each regime, then this alone could result in a reordering of the estimated factors across regimes.

The fifth and sixth factors are again of particular interest because of the role that they play in boosting the pricing performance of the model. Unlike in the low-volatility regime, it is arguable that the weights for these factors have a reasonably clear interpretation. They suggest that fifth factor is tied to return reversals and that the sixth factor is tied to price momentum as it is typically measured in the asset pricing literature. Therefore, three out of the six factors for the high volatility regime appear to be primarily associated with past price behavior.

4.2. Firm-level correlations between characteristics and subsequent returns

Up to now the analysis has focused on the cross-sectional relation between firm characteristics and expected stock returns. But what are the implications of the evidence with respect to the explanatory power of the characteristics in the *time-series* dimension? To lay the foundation for addressing this question, I examine the estimated firm-level correlations between the characteristics and subsequent excess stock returns. Let $t \in t_n$ indicate that t corresponds to a month for which the nth firm has a non-missing excess return. Under the regime-switching model, the estimated regime i correlation between the firm's jth characteristic and its subsequent excess stock return is given by

$$\hat{\varrho}_{nj,i} = \frac{\sum_{t\in t_n} \xi_{i,t|T}(\hat{\boldsymbol{\theta}})(c_{nj,t-1} - \bar{c}_{nj,i})(r_{n,t} - \bar{r}_{n,i})}{(\sum_{t\in t_n} \xi_{i,t|T}(\hat{\boldsymbol{\theta}})(c_{nj,t-1} - \bar{c}_{nj,i})^2)^{1/2}(\sum_{t\in t_n} \xi_{i,t|T}(\hat{\boldsymbol{\theta}})(r_{n,t} - \bar{r}_{n,i})^2)^{1/2}}, \tag{19}$$

where $c_{nj,t-1}$ is the value of the jth characteristic for month $t-1$, $r_{n,t}$ is the excess stock return for month t, $\bar{c}_{nj,i} = \sum_{t\in t_n} \xi_{i,t|T}(\hat{\boldsymbol{\theta}})c_{nj,t-1}/\sum_{t\in t_n} \xi_{i,t|T}(\hat{\boldsymbol{\theta}})$, and $\bar{r}_{n,i} = \sum_{t\in t_n} \xi_{i,t|T}(\hat{\boldsymbol{\theta}})r_{n,t}/\sum_{t\in t_n} \xi_{i,t|T}(\hat{\boldsymbol{\theta}})$.

Panel A of Table 7 reports selected percentiles of the cross-sectional distribution of $\hat{\varrho}_{nj,i}$. The first set of results is for the low-volatility regime and the second is for the high-volatility regime. I use all available stocks to estimate $\boldsymbol{\theta}$, but exclude those that have fewer than 60 non-missing excess returns when estimating the correlations to ensure that they display reasonable precision.[16] The median value of the estimated correlations is close to zero for most of the characteristics. It ranges from -0.05 to 0.05 for the low-volatility regime and from -0.08 to 0.08 for the high-volatility regime. However, the cross-sectional dispersion in the estimated correlations is also fairly large. The spread between the 5th and 95th percentiles of the cross-sectional distribution ranges from 0.29 to 0.41 for the low-volatility regime and from 0.50 to 0.64 for the high-volatility regime. This finding hints at substantial return predictability for a sizable fraction of individual stocks.

Panel B of the table provides additional insights on the predictability issue. It examines the regime-specific distributions of the R-squared statistic produced by time-series regressions of excess stock returns on the full set of firm characteristics. These statistics are constructed as follows for stocks with at least 60 non-missing excess returns. Let $\boldsymbol{c}_{n,t-1}$ denote a row vector that contains the month $t-1$ values of the characteristics for the nth firm. First, I solve the WLS regression problem

$$\min_{\boldsymbol{\beta}_{n,i}} \sum_{t \in t_n} (r_{n,t} - \boldsymbol{c}_{n,t-1}\boldsymbol{\beta}_{n,i})^2 \xi_{i,t|T}(\hat{\boldsymbol{\theta}}) \tag{20}$$

for $i = 1, 2$ to obtain $\hat{\boldsymbol{\beta}}_{n,1}$ and $\hat{\boldsymbol{\beta}}_{n,2}$, the regime-specific estimates of the coefficients.[17] Next, I compute

$$\hat{\sigma}^2_{e,i} = \sum_{t \in t_n} (r_{n,t} - \boldsymbol{c}_{n,t-1}\hat{\boldsymbol{\beta}}_{n,i})^2 \xi_{i,t|T}(\hat{\boldsymbol{\theta}}) \quad \text{and} \quad \hat{\sigma}^2_{r,i} = \sum_{t \in t_n} (r_{n,t} - \bar{r}_{n,i})^2 \xi_{i,t|T}(\hat{\boldsymbol{\theta}}) \tag{21}$$

for $i = 1, 2$, where $\bar{r}_{n,i} = \sum_{t \in t_n} r_{n,t}\xi_{i,t|T}(\hat{\boldsymbol{\theta}})$. Finally, I compute the R-squared statistic for the ith regime as $R^2_i = 1 - (\hat{\sigma}^2_{e,i}/\hat{\sigma}^2_{r,i})$.

---

[16] More specifically, I estimate $\boldsymbol{\theta}$ using $\tilde{\boldsymbol{r}}_{p,t} = \bar{\boldsymbol{W}}_c^{1/2}\boldsymbol{C}^+_{t-1}\boldsymbol{r}_t$, where $\boldsymbol{r}_t$ is the $N_t \times 1$ vector of excess stock returns for month t, $\boldsymbol{C}^+_{t-1}$ is the left pseudoinverse of an $N_t \times 17$ matrix $\boldsymbol{C}_{t-1}$ whose nth row contains a 1 along with the values of 16 predetermined characteristics for the nth firm, and $\bar{\boldsymbol{W}}_c = (\sum_{t=1}^T N_t)^{-1} \sum_{t=1}^T \boldsymbol{C}'_{t-1}\boldsymbol{C}_{t-1}$.

[17] This strategy is inspired by standard results for Markov-switching regressions. A Markov-switching regression with Gaussian errors for each regime would yield FOCs for $\boldsymbol{\beta}_{n,i}$ of the form $\sum_{t \in t_n}(r_{n,t} - \boldsymbol{c}_{n,t-1}\boldsymbol{\beta}_{n,i})\xi_{i,t|T}(\cdot) = \boldsymbol{0}$, where $\xi_{i,t|T}(\cdot)$ denotes the smoothed probability of regime i for period t implied by the full set of model parameters. The FOCs for the WLS problem take the same form except that the smoothed probability of regime i for period t is replaced by the estimated probability produced by the regime-switching IPCA model.

The median R-squared is 44% for the low-volatility regime and 63% for the high-volatility regime. These values may seem implausibly high at first glance. But it is important to bear in mind that the R-squared statistic is an upward-biased estimate of the fraction of the return variance that is explained by the characteristics. For example, if the error term for a regression is serially uncorrelated and conditionally homoskedastic, then TR^2 — the sample size times the R-squared statistic — has a well-known limiting distribution (see, e.g., Kirby, 1997). For the case in which population R-squared is equal to zero, it converges to a chi square random variable whose degrees of freedom equals the number of regressors (excluding the intercept). Thus the R-squared statistic for a regression with an intercept, 16 characteristics, and the minimum allowed number of observations has an expected value of around $16/60 = 27\%$ under the null hypothesis that excess returns are unpredictable. Furthermore, the 95th percentile of its distribution is around $26.3/60 = 48\%$.

Although the cross-sectional distribution of the R-squared statistic in the current setting is unknown, the influence of the upward bias is likely to be quite substantial. To get a better idea of the extent to which the characteristics capture time-series predictability, I examine how the distributional properties of the R-squared statistic change as I increase the lower bound on number of observations used to fit the regressions. Figure 5 uses histograms to illustrate the findings. I construct the histograms by requiring that stocks have a minimum of either 120 or 240 monthly observations to be included in the analysis. The top row of plots is for the low-volatility regime and the bottom row is for the high volatility regime. In each case, I show the results obtained using all stocks and those obtained using the no-micro-caps sample.

As expected, increasing the sample size for the predictive regressions shifts the cross-sectional distribution of the R-squared statistic to the left. Consider the results for the low-volatility regime. The average R-squared using a minimum of 120 observations is 15.3% whereas that using a minimum of 240 observations is 9.6%. Excluding micro-cap stocks leads to a further reduction in the average R-squared, but the changes are relatively minor. It falls to 13.1% using a minimum of 120 observations and 8.4% using a minimum of 240 observations. Under the null of no predictability, the expected R-squared for a regression with 240 observations is around 6.7% if the errors are serially uncorrelated and conditionally homoskedastic. The evidence therefore suggests that the predictable component of returns for a typical stock is reasonably small in the low-volatility regime.

But the results for the high-volatility regime are more striking. The average R-squared is 32.7% using a minimum 120 observations and 21.0% using a minimum 240 observations. Thus the evidence points to a surprisingly high level of predictability. Notably, this finding is not driven by micro-cap stocks. The average R-squared for the predictive regressions that exclude these stocks is 34.8% using a minimum 120 observations and 21.9% using a minimum 240 observations. Thus the results point to a strong relation between firm characteristics and conditional expected returns in the high-volatility regime for stocks of all capitalization levels.

Despite the intriguing nature of these findings, their implications for investors are as yet unclear. One obvious caveat is that the regime-specific estimates of the predictable variation in returns correspond to unobserved regimes. Hence there is no way to construct a feasible trading strategy that fully captures regime-specific levels of predictability. In addition, it is unclear whether the regime-specific R-squared statistics are at all indicative of the *out-of-sample* performance of any type of implementable trading strategy. This is an empirical question that can only be answered by conducting out-of-sample tests. Fortunately, the tractability of the regime-switching estimators allows for a straightforward assessment of the out-of-sample evidence.

4.3. Predictive regressions for pure play portfolios

As a first step in assessing the evidence, I examine the ability of recursively-updated estimates of the regime probabilities to predict the excess returns for the linear pure play portfolios. The estimates are constructed as follows. Let $\{\tilde{\boldsymbol{r}}_{p,t}\}_{t=1}^{h}$ denote the sequence of observations that constitutes the initial holdout sample (i.e., $1 < h < T$). First, I estimate $\boldsymbol{\theta}$ by fitting the regime-switching model to $\{\tilde{\boldsymbol{r}}_{p,t}\}_{t=1}^{h}$ and denote the resultant estimate by $\hat{\boldsymbol{\theta}}_1$. Second, I compute $\xi_{2,h+1|h}(\hat{\boldsymbol{\theta}}_1)$, the estimated probability of the high-volatility regime in period $h + 1$ conditional on observing $\{\tilde{\boldsymbol{r}}_{p,t}\}_{t=1}^{h}$. I then repeat these steps for each remaining observation by, for example, redefining the holdout sample to be $\{\tilde{\boldsymbol{r}}_{p,t}\}_{t=1}^{h+1}$ and refitting the model to obtain $\xi_{2,h+2|h+1}(\hat{\boldsymbol{\theta}}_2)$.

After computing the time series of recursively-updated probability estimates, I use OLS to fit predictive regressions of the form

$$r_{pj,t} = \mu_1 + \Delta_{12}\xi_{2,t|t-1}(\hat{\boldsymbol{\theta}}_{t-h}) + u_t,\ t = h+1, \ldots, T, \tag{22}$$

where $r_{pj,t}$ is the excess return for the jth linear pure play portfolio. The results for $h = 120$ are summarized in Table 8. It reports the OLS estimates of the coefficients, their t-statistics, and the regression R-squared. The t-statistics are based on heteroskedasticity-robust standard errors.

The estimate of μ_1 for a given portfolio represents its estimated expected excess return for the case in which $\xi_{2,t|t-1}(\hat{\boldsymbol{\theta}}_{t-h}) = 0$. Thirteen of the estimates are statistically significant at the 10% significance level (as are twelve at the 5% level), which is consistent with the evidence that the characteristics help to explain the cross-sectional of expected returns in the low-volatility regime. More notably, however, seven of the estimates of Δ_{12} are statistically significant at the 5% level. So almost half of the linear pure play portfolios display evidence of out-of-sample predictability.

On the other hand, the R-squared statistics for the portfolios that produce statistically significant estimates of Δ_{12} are fairly small. They range from 1.4% for the EWM and APM portfolios to 4.9% for the P52H portfolio. In addition, the evidence of predictability is weaker for the case in which micro-cap stocks are excluded from the analysis. Six of the estimates of Δ_{12} are statistically significant at the 5% significance level, and the associated R-squared statistics range from 1.3% to 3.7%. These findings make it tempting to conclude that little can be gained by exploiting the regime-switching forecasts of the excess portfolio returns. But prior research suggests that low R-squared statistics do not necessarily preclude the construction of profitable trading strategies. I investigate this possibility by considering the evidence from portfolio sorts.

4.4. Evidence from portfolios sorts

Let $\boldsymbol{m}_t$ denote the conditional mean of the vector of fitted values produced by the Fama and MacBeth (1973) regression for month $t+1$. That is, $\boldsymbol{m}_t = \mathrm{E}(\hat{\boldsymbol{r}}_{t+1}|\mathcal{I}_t)$, where $\hat{\boldsymbol{r}}_{t+1} = \boldsymbol{C}_t\boldsymbol{r}_{p,t+1}$ and $\mathcal{I}_t$ denotes the month t information set. Because $\boldsymbol{C}_t \in \mathcal{I}_t$, it follows that $\boldsymbol{m}_t = \boldsymbol{C}_t\boldsymbol{m}_{p,t}$, where $\boldsymbol{m}_{p,t} = \mathrm{E}(\boldsymbol{r}_{p,t+1}|\mathcal{I}_t)$. The results reported in Table 8 point to predictable variation in $\boldsymbol{r}_{p,t+1}$ and hence a time-varying value of $\boldsymbol{m}_{p,t}$. The portfolio sorts are designed to isolate the contribution of this predictability to the explanatory power of the conditional factor model that underpins IPCA.

I start by constructing recursively-updated estimates of $\{\boldsymbol{m}_{p,t}\}_{t=h}^{T}$ for two separate cases. The first corresponds to the IPCA framework of Kelly et al. (2019). Because their version of IPCA treats $\boldsymbol{m}_{p,t}$ as constant, the associated estimate of $\boldsymbol{m}_{p,h}$ is simply $\bar{\boldsymbol{r}}_{p(h)} = (1/h)\sum_{t=1}^{h} \boldsymbol{r}_{p,t}$. The second corresponds to the regime-switching version of IPCA. Under the two-regime model, the estimate of $\boldsymbol{m}_{p,h}$ is computed as $\hat{\boldsymbol{m}}_{p,h} = \sum_{j=1}^{2} \xi_{j,h+1|h}(\hat{\boldsymbol{\theta}}_1)\hat{\boldsymbol{\delta}}_{j,1}$, where $\hat{\boldsymbol{\theta}}_1$ and $\hat{\boldsymbol{\delta}}_{j,1}$ denote the estimates of $\boldsymbol{\theta}$ and $\boldsymbol{\delta}_j$ that result from fitting the regime-switching model to $\{\tilde{\boldsymbol{r}}_{p,t}\}_{t=1}^{h}$.

This procedure, which uses only the information available to investors in real time, delivers two competing estimates of $\boldsymbol{m}_t$ for each value of $t \geq h$. Because the cross-sectional correlation between the two estimates is likely to be fairly high, I employ residual-based portfolio sorts to isolate the effect of regime changes. To see the basic idea, suppose that $\hat{\boldsymbol{m}}_h = \boldsymbol{C}_t\hat{\boldsymbol{m}}_{p,h}$ is the estimate of $\boldsymbol{m}_h$ produced by the regime-switching model. First, I regress $\hat{\boldsymbol{m}}_h$ on $\boldsymbol{C}_t\bar{\boldsymbol{r}}_{p(h)}$ and save the resultant vector of residuals. Next, I conduct two independent quintile sorts using the set of available stocks for month h: one in ascending order of the elements of $\boldsymbol{C}_t\bar{\boldsymbol{r}}_{p(h)}$ and one in ascending order of the saved residuals. Finally, I form 25 portfolios by finding the intersections of the quintile groups. These portfolios are formed at the end of month h and held until the end of month $h + 1$.

Table 9 examines the performance of portfolios formed in this fashion. The results in panel A are for equally-weighted portfolios that include all stocks. Those in panel B are for value-weighted portfolios that exclude micro-cap stocks. In each case, the initial group of columns reports the results obtained by regressing the regime-switching estimate of $\boldsymbol{m}_t$ on that implied by the single-regime model of Kelly et al. (2019), and the remaining group of columns reports the results obtained by regressing the single-regime estimate of $\boldsymbol{m}_t$ on that implied by the regime-switching model.

Panel A tells a clear-cut story. First consider the results obtained by regressing the regime-switching estimate of $\boldsymbol{m}_t$ on that implied by the single-regime model. The average excess returns for the high and low portfolios in each column are markedly different, and the high-minus-low portfolio has an average excess return that is positive and statistically significant at the 1% level in every case. The smallest of the column-wise high-low differences is 1.77%, which has a t-statistic of 9.03. So using the single-regime estimates to sort stocks is very effective at spreading average returns. At the same time, however, the high and low portfolios in each row also produce markedly different average excess returns. The smallest of these row-wise high-low differences is 0.98%, which has a *t*-statistic of 3.37. Thus the evidence indicates that the regime-switching estimates provide substantial *incremental* information about conditional expected stock returns.

The key point is that the row-wise differences in average returns are driven by differences between the regime-switching and single-regime estimates of $\boldsymbol{m}_{p,t}$. There is no reason to expect the differences in average returns to be positive and statistically significant unless the regime-switching estimates capture predictable variation in $\boldsymbol{r}_{p,t+1}$. Although the evidence suggests that predictable component of $\boldsymbol{r}_{p,t+1}$ is relatively small, it is sufficient to generate substantial gains in the context of portfolio sorts. This finding highlights the potential relevance of regime switching for investment management applications that exploit the explanatory power of firm characteristics.

The same conclusions emerge from the results obtained by regressing the single-regime estimate of $\boldsymbol{m}_t$ on that implied by the regime-switching model. The high and low portfolios in each column produce markedly different average excess returns, and the high-minus-low portfolio has an average excess return that is positive and statistically significant at the 1% level in every case. The smallest of the column-wise high-low differences is 1.96%, which has a *t*-statistic of 8.99. Hence, using the regime-switching estimates to sort stocks is very effective at spreading average returns.

Note, however, that the residual sorts produce a very different pattern than that seen previously. First, the differences in average excess returns between the high and low portfolios in each row are not as pronounced. Second, the high-minus-low portfolios have *negative* average excess returns in every case. To understand these findings, consider a stock that has a high estimated expected excess return according to the regime-switching model. If the single-regime estimates are less informative than the regime-switching estimates about conditional expected stock returns, then this stock will be more likely to have a negative residual than a positive residual. Consequently, the residuals will tend to display a negative cross-sectional correlation with conditional expected stock returns.

Panel B shows that these findings are robust to excluding micro-cap stocks and using value weights instead of equal weights to form the portfolios. Consider the results obtained by regressing the regime-switching estimate of $\boldsymbol{m}_t$ on that implied by the single-regime model. Although the sorts produce lower spreads between the average excess portfolio returns those that use all available stocks, I still find that the average excess returns for the high-minus-low portfolios are statistically significant at the 5% level in every case. In addition, the smallest of the row-wise high-low differences is 0.58%, which has a t-statistic of 2.05. Thus the residuals still provide substantial incremental information about conditional expected stock returns. This bolsters the prior findings regarding the likely relevance of regime switching in investment management applications.

In summary, the picture that emerges from the out-of-sample tests is unambiguous. Not only does the evidence show that the regime-switching estimates capture time variation in the conditional expected excess returns for the linear pure play portfolios, it also reveals that the entire cross-section of expected excess stock returns displays statistically and economically significant out-of-sample predictability. Importantly, these findings are robust to the exclusion of micro-cap stocks from the analysis. In view of the evidence from the portfolio sorts, further study of the asset pricing implications of time-varying risk premia for pure play portfolios would be a worthwhile endeavor.

5. Conclusions

The recent study of Kelly et al. (2019) finds that using a small set of firm characteristics to instrument for the loadings on latent risk factors delivers a conditional factor model that largely explains the cross-sectional relation between characteristics and conditional expected stock returns. However, the managed portfolios that play a crucial role in their IPCA framework are implicitly assumed to follow a static factor model. I relax this strong assumption by introducing regime-switching IPCA, which allows both the factors and loadings to differ across economic regimes. Implementing this generalized version of IPCA yields a number of noteworthy findings.

First, the managed portfolio returns display clear evidence of regime-switching dynamics. But the properties of the estimated regimes are very different from those obtained by fitting regime-switching models to market returns. Instead of the usual pattern of bull and bear regimes reported in prior studies, the estimates indicate that are stretches of time with a typical duration of around three months during which individual stocks have both high volatility and high expected excess returns. This is suggestive of a scenario in which the magnitudes of the factor loadings are large in the high-volatility regime, the prices of risk are high in the high-volatility regime, and the firm characteristics capture cross-sectional differences in priced covariance risk.

Second, the results of regime-specific asset pricing tests are less favorable to the hypothesis of exact factor pricing than those of the single-regime tests implemented by Kelly et al. (2019). I find that the hypothesis is rejected at the 1% significance level for any regime-switching parameterization of the conditional factor model that includes six or fewer factors. But the decline in the magnitude of the estimated pricing errors as more factors are added to the model is substantial. Thus the evidence indicates that much of the explanatory power of firm characteristics for expected stock returns stems from their ability to capture cross-sectional variation in exposures to latent risk factors.

Third, the regime-switching estimates display an impressive degree of predictive ability in out-of-sample tests. Portfolio sorts that isolate the incremental contribution of regime-switching dynamics to the explanatory power of out-of-sample estimates of conditional expected stock returns produce economically and statistically-significant spreads in average returns. The spread in average returns declines if I use value weighting and exclude micro-cap stocks from the analysis. Nonetheless, it remains both economically and statistically significant. This finding highlights the potential benefits of accounting for regime switching in investment management applications.

Beyond these findings, regime-switching IPCA extends all of the methodological benefits of IPCA to a more general setting. For example, Kelly et al. (2019, p. 523) note that IPCA "allows investors and managers to easily assess a firm's cost of capital without relying on the obviously misspecified CAPM beta or other factor loadings that may be infeasible to estimate with time series regression." The same is true of regime-switching IPCA, which produces cost of capital estimates that capture the predictable variation in expected returns associated with predictable variation in latent economic regimes. It is easy to envision many other uses of the methodology as well.

Appendix A. Estimation Algorithms for IPCA and Regime-Switching IPCA

Much of the econometric analysis relies on iterative methods for solving optimization problems. In particular, I use an ALS algorithm to jointly estimate the parameters and factors in the IPCA framework, and an EM algorithm to estimate the parameters of the Markov switching process that governs the dynamics of the mean vector and covariance matrix of the excess managed portfolio returns under my regime-switching generalization of IPCA. These algorithms are presented below.

A.1. ALS algorithm for IPCA

Following Kelly et al. (2019), the IPCA estimators of $\boldsymbol{\delta}$, $\boldsymbol{\Gamma}$, and $\boldsymbol{F}$ are obtained by minimizing $(NT)^{-1}\sum_{t=1}^{T}\boldsymbol{\varepsilon}_t'\boldsymbol{\varepsilon}_t$. The FOCs imply that the estimators satisfy

$$\begin{pmatrix} \hat{\boldsymbol{\delta}} \\ \hat{\boldsymbol{\gamma}} \end{pmatrix} = \left(\sum_{t=1}^{T} \begin{pmatrix} 1 & \hat{\boldsymbol{f}}_t' \\ \hat{\boldsymbol{f}}_t & \hat{\boldsymbol{f}}_t\hat{\boldsymbol{f}}_t' \end{pmatrix} \otimes (\boldsymbol{C}_{t-1}'\boldsymbol{C}_{t-1}) \right)^{-1} \sum_{t=1}^{T} \begin{pmatrix} (\boldsymbol{C}_{t-1}'\boldsymbol{C}_{t-1})\boldsymbol{r}_{p,t} \\ (\hat{\boldsymbol{f}}_t \otimes (\boldsymbol{C}_{t-1}'\boldsymbol{C}_{t-1}))\boldsymbol{r}_{p,t} \end{pmatrix} \tag{A1}$$

and

$$\hat{\boldsymbol{f}}_t = (\hat{\boldsymbol{\Gamma}}'\boldsymbol{C}_{t-1}'\boldsymbol{C}_{t-1}\hat{\boldsymbol{\Gamma}})^{-1}(\hat{\boldsymbol{\Gamma}}'\boldsymbol{C}_{t-1}'\boldsymbol{C}_{t-1})(\boldsymbol{r}_{p,t} - \hat{\boldsymbol{\delta}}),\ t = 1,\dots,T, \tag{A2}$$

where $\hat{\boldsymbol{\gamma}} = \mathrm{vec}(\hat{\boldsymbol{\Gamma}})$ and $\otimes$ denotes the Kronecker product. Under the maintained assumptions of the IPCA model, the estimators of $\boldsymbol{\delta}$ and $\boldsymbol{\Gamma}$ are consistent provided that $\boldsymbol{C}_{t-1}'\boldsymbol{\varepsilon}_t \xrightarrow{p} \mathbf{0}$ as $N \to \infty$ for all t (see Kelly et al., 2017, for a discussion of regularity conditions).

The ALS algorithm delivers estimates of $\boldsymbol{\delta}$, $\boldsymbol{\Gamma}$, and $\boldsymbol{F}$ that satisfy equations (A1) and (A2) for a given dataset. Let $\hat{\boldsymbol{F}}^{(0)}$ denote an initial estimate (guess) for the matrix of factor realizations. Substituting $\hat{\boldsymbol{F}}^{(0)}$ into equation (A1) yields $\hat{\boldsymbol{\delta}}^{(1)}$ and $\hat{\boldsymbol{\gamma}}^{(1)}$, the estimates of the parameter vectors for iteration one. Similarly, substituting $\hat{\boldsymbol{\delta}}^{(1)}$ and $\hat{\boldsymbol{\gamma}}^{(1)}$ into equation (A2) yields $\hat{\boldsymbol{F}}^{(1)}$, an updated estimate of the matrix of factor realizations. This completes the first iteration. The iterations continue in the same fashion until the reduction in the value of the objective function from one iteration to the next falls below a specified tolerance, such as machine precision.

A.1.1. Identification restrictions

Suppose that $\hat{\boldsymbol{\delta}}$, $\hat{\boldsymbol{\Gamma}}$, and $\hat{\boldsymbol{F}}$ satisfy equations (A1) and (A2) for a given dataset. If these estimates are obtained via ALS, then they will not, in general, satisfy any particular set of identification restrictions. But this is easily remedied. Let $\bar{\boldsymbol{f}} = (1/T)\sum_{t=1}^{T}\hat{\boldsymbol{f}}_t$, $\boldsymbol{H} = (1/T)\sum_{t=1}^{T}(\hat{\boldsymbol{f}}_t - \bar{\boldsymbol{f}})(\hat{\boldsymbol{f}}_t - \bar{\boldsymbol{f}})'$, and $\boldsymbol{G} = \hat{\boldsymbol{\Gamma}}\boldsymbol{H}^{1/2}$. Further, let $\boldsymbol{G} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}'$ denote the singular value decomposition of $\boldsymbol{G}$. Identification can be achieved by replacing $\hat{\boldsymbol{\delta}}$, $\hat{\boldsymbol{\Gamma}}$, and $\hat{\boldsymbol{F}}$ with alternative estimates of the form $\hat{\boldsymbol{\delta}}^{\star} = \hat{\boldsymbol{\delta}} + \hat{\boldsymbol{\Gamma}}\bar{\boldsymbol{f}}$, $\hat{\boldsymbol{\Gamma}}^{\star} = \boldsymbol{U}\boldsymbol{D}$, and $\hat{\boldsymbol{f}}_t^{\star} = \boldsymbol{V}'\boldsymbol{H}^{-1/2}(\hat{\boldsymbol{f}}_t - \bar{\boldsymbol{f}})$ for $t = 1, 2, \ldots, T$. It is easy to see that doing so leaves the sum of squares that defines the IPCA estimators unchanged, and it is easy to verify that the alternative estimates satisfy a set of identification restrictions that mirror those typically imposed in PCA. Specifically, the sample mean of $\hat{\boldsymbol{f}}_t^{\star}$ is zero, the sample covariance matrix of $\hat{\boldsymbol{f}}_t^{\star}$ is an identity matrix, and $\hat{\boldsymbol{\Gamma}}^{\star\prime}\hat{\boldsymbol{\Gamma}}^{\star}$ is diagonal with descending diagonal elements.

A.2. EM algorithm for the Markov switching process

The EM algorithm for the Markov switching process is adapted from Kirby (2019b). Let $\mathcal{F}_t = \{\tilde{\boldsymbol{r}}_{p,1}, \tilde{\boldsymbol{r}}_{p,2}, \ldots, \tilde{\boldsymbol{r}}_{p,t}\}$ contain the excess managed portfolio returns observed though period t. Four sets of conditional probabilities must be computed in order to implement the algorithm. The first and second set are $\boldsymbol{\xi}_{t|t}(\boldsymbol{\theta}) = \mathrm{E}(\boldsymbol{\xi}_t|\mathcal{F}_t;\boldsymbol{\theta})$ and $\boldsymbol{\xi}_{t+1|t}(\boldsymbol{\theta}) = \mathrm{E}(\boldsymbol{\xi}_{t+1}|\mathcal{F}_t;\boldsymbol{\theta})$ for $t = 1, 2, \ldots, T$, where $\boldsymbol{\theta}$ denotes a vector that contains the parameters of the model. These are computed using a pair of simple recursions (Hamilton, 1989, 1990). Specifically, $\boldsymbol{\xi}_{t+1|t}(\boldsymbol{\theta}) = \boldsymbol{P}'\boldsymbol{\xi}_{t|t}(\boldsymbol{\theta})$ and

$$\boldsymbol{\xi}_{t|t}(\boldsymbol{\theta}) = \frac{\boldsymbol{\xi}_{t|t-1}(\boldsymbol{\theta}) \odot \boldsymbol{\phi}_t(\boldsymbol{\vartheta})}{\mathbf{1}'(\boldsymbol{\xi}_{t|t-1}(\boldsymbol{\theta}) \odot \boldsymbol{\phi}_t(\boldsymbol{\vartheta}))}, \tag{A3}$$

where $\odot$ denotes the Hadamard product, $\mathbf{1}$ is an $M \times 1$ vector of ones, $\boldsymbol{\phi}_t(\boldsymbol{\vartheta}) = (\phi_{1t}(\boldsymbol{\vartheta}), \ldots, \phi_{Mt}(\boldsymbol{\vartheta}))'$ is an $M \times 1$ vector of Gaussian densities whose ith element is given by

$$\phi_{it}(\boldsymbol{\vartheta}) = \frac{1}{(2\pi)^{J/2}|\boldsymbol{\Omega}_i|^{1/2}} \exp\left(-\frac{1}{2}(\tilde{\boldsymbol{r}}_{p,t} - \boldsymbol{\kappa}_i)'\boldsymbol{\Omega}_i^{-1}(\tilde{\boldsymbol{r}}_{p,t} - \boldsymbol{\kappa}_i)\right) \tag{A4}$$

with $\boldsymbol{\Omega}_i = \boldsymbol{\Lambda}_i\boldsymbol{\Lambda}_i' + \tau_i^2\mathbf{I}$, and $\boldsymbol{\vartheta}$ denotes a column vector that contains the density function parameters (i.e., all elements of $\boldsymbol{\kappa}_i$, $\boldsymbol{\Lambda}_i$, and τ_i^2 for $i = 1, \ldots, M$). The third and and fourth sets are $\boldsymbol{\xi}_{t|T}(\boldsymbol{\theta}) =$

$\mathrm{E}(\boldsymbol{\xi}_t|\mathcal{F}_T;\boldsymbol{\theta})$ for $t = 1, 2, \ldots, T$ and $\boldsymbol{\zeta}_{t|T}(\boldsymbol{\theta}) = \mathrm{E}(\boldsymbol{\xi}_{t-1}\boldsymbol{\xi}_t'|\mathcal{F}_T;\boldsymbol{\theta})$ for $t = 2, \ldots, T$. Kim (1994) shows that

$$\boldsymbol{\xi}_{t|T}(\boldsymbol{\theta}) = \boldsymbol{\xi}_{t|t}(\boldsymbol{\theta}) \odot (\boldsymbol{P}'(\boldsymbol{\xi}_{t+1|T}(\boldsymbol{\theta}) \oslash \boldsymbol{\xi}_{t+1|t}(\boldsymbol{\theta}))), \tag{A5}$$

and

$$\boldsymbol{\zeta}_{t|T}(\boldsymbol{\theta}) = (\boldsymbol{\xi}_{t-1|t-1}(\boldsymbol{\theta})(\boldsymbol{\xi}_{t|T}(\boldsymbol{\theta}) \oslash \boldsymbol{\xi}_{t|t-1}(\boldsymbol{\theta}))') \odot \boldsymbol{P}, \tag{A6}$$

where $\oslash$ denotes element-by-element division.

Because the only restriction placed on the transition probabilities is that they sum to one, the equations that determine the maximum likelihood estimators for the model take a relatively simple form (Hamilton, 1990). For example, the maximum likelihood estimator of $\boldsymbol{\vartheta}$ satisfies

$$\sum_{t=1}^{T} \left(\left. \frac{\partial \boldsymbol{\varphi}_t(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}'} \right|_{\boldsymbol{\vartheta}=\hat{\boldsymbol{\vartheta}}} \right)' \boldsymbol{\xi}_{t|T}(\hat{\boldsymbol{\theta}}) = \mathbf{0}, \tag{A7}$$

where $\boldsymbol{\varphi}_t(\boldsymbol{\vartheta}) = (\log \phi_{1t}(\boldsymbol{\vartheta}), \ldots, \log \phi_{Mt}(\boldsymbol{\vartheta}))'$, and

$$\hat{p}_{ij} = \frac{\sum_{t=2}^{T} \zeta_{ij,t|T}(\hat{\boldsymbol{\theta}})}{\sum_{t=2}^{T} \xi_{i,t-1|T}(\hat{\boldsymbol{\theta}})}, \tag{A8}$$

where $\zeta_{ij,t|T}(\hat{\boldsymbol{\theta}})$ is the ijth element of $\boldsymbol{\zeta}_{t|T}(\hat{\boldsymbol{\theta}})$ and $\xi_{i,t-1|T}(\hat{\boldsymbol{\theta}})$ is the ith element of $\boldsymbol{\xi}_{t-1|T}(\hat{\boldsymbol{\theta}})$. The EM algorithm finds the maximum likelihood estimators via fixed-point iteration. It is initiated with a guess at the solution, say $\hat{\boldsymbol{\theta}}^{(0)}$, and the four sets of conditional probabilities implied by this guess are used to compute an updated estimate of $\boldsymbol{\theta}$ that increases the value of the log likelihood function (see Hamilton, 1990, for further elaboration).

Let $\hat{\boldsymbol{\theta}}^{(n)}$ denote the estimate of $\boldsymbol{\theta}$ at the end of the nth iteration. It is easy to show that equation (A7) implies that the updated estimate of $\boldsymbol{\kappa}_i$ is given by

$$\hat{\boldsymbol{\kappa}}_i^{(n+1)} = \frac{\sum_{t=1}^{T} \xi_{i,t|T}(\hat{\boldsymbol{\theta}}^{(n)})\tilde{\boldsymbol{r}}_{p,t}}{\sum_{t=1}^{T} \xi_{i,t|T}(\hat{\boldsymbol{\theta}}^{(n)})}. \tag{A9}$$

The updated estimates of $\boldsymbol{\Lambda}_i$ and τ_i^2 follow from equation (A7) by noting that the maximum likelihood estimators of these parameters satisfy probability-weighted versions of the FOCs for the static factor model considered by Tipping and Bishop (1999).[18] Specifically, the updated estimates are obtained from the spectral decomposition of the probability-weighted covariance matrix

$$\boldsymbol{S}_i(\hat{\boldsymbol{\theta}}^{(n)}) = \frac{\sum_{t=1}^{T} \xi_{i,t|T}(\hat{\boldsymbol{\theta}}^{(n)})(\tilde{\boldsymbol{r}}_{p,t} - \hat{\boldsymbol{\kappa}}_i^{(n)})(\tilde{\boldsymbol{r}}_{p,t} - \hat{\boldsymbol{\kappa}}_i^{(n)})'}{\sum_{t=1}^{T} \xi_{i,t|T}(\hat{\boldsymbol{\theta}}^{(n)})}. \tag{A10}$$

Let $l_{i,j}^{(n)}$ denote the jth largest eigenvalue of $\boldsymbol{S}_i(\hat{\boldsymbol{\theta}}^{(n)})$ and let $\boldsymbol{u}_{i,j}^{(n)}$ denote the corresponding orthonormal eigenvector. The jth column of $\hat{\boldsymbol{\Lambda}}_i^{(n+1)}$ is given by $\hat{\boldsymbol{\lambda}}_{i,j}^{(n+1)} = \boldsymbol{u}_{i,j}^{(n)}(l_{i,j}^{(n)} - \hat{\tau}_i^{2(n+1)})^{1/2}$, where $\hat{\tau}_i^{2(n+1)} = (J-K)^{-1}\sum_{j=K+1}^{J} l_{i,j}^{(n)}$. Finally, the updated estimate of p_{ij} is given by

$$\hat{p}_{ij}^{(n+1)} = \frac{\sum_{t=2}^{T} \hat{\zeta}_{ij,t|T}(\hat{\boldsymbol{\theta}}^{(n)})}{\sum_{t=2}^{T} \hat{\xi}_{i,t-1|T}(\hat{\boldsymbol{\theta}}^{(n)})}. \tag{A11}$$

Stacking the full set of updated parameter estimates into a vector yields $\hat{\boldsymbol{\theta}}^{(n+1)}$.

This version of the algorithm assumes that $\boldsymbol{\xi}_{1|0}$ is set equal to a fixed value that is unrelated to $\boldsymbol{\theta}$. If $\boldsymbol{\xi}_{1|0}$ is regarded as a separate vector of unknowns to be estimated by maximum likelihood, then $\hat{\boldsymbol{\xi}}_{1|0}^{(n+1)} = \boldsymbol{\xi}_{1|T}(\hat{\boldsymbol{\theta}}^{(n)})$ is added to the updating equations. The algorithm terminates when either $\|\hat{\boldsymbol{\theta}}^{(n+1)} - \hat{\boldsymbol{\theta}}^{(n)}\|$ or the increase in the value of the log likelihood from one iteration to the next falls below a specified tolerance. Because the log likelihood is given by $\sum_{t=1}^{T} \log \mathbf{1}'(\boldsymbol{\xi}_{t|t-1}(\boldsymbol{\theta}) \odot \boldsymbol{\phi}_t(\boldsymbol{\vartheta}))$, its value can easily be computed at each iteration.

Appendix B. Variable Definitions

All of the data items used to construct the firm characteristics are drawn from CRSP and Compustat. I exclude firms with less than two years of Compustat data to mitigate the impact of backfilling and I winsorize each characteristic on a month-by-month basis at the 0.5th and 99.5th percentiles of its cross-sectional distribution. To account for the delay between the end of a firm's fiscal year and the release of its annual report, I match balance sheet information for firms whose fiscal years end in

[18] A detailed treatment of the first-order conditions for static factor models can be found in Lawley and Maxwell (1971).

month $t-16$ with stock returns for months $t-11$ to t. Hence, all annual balance sheet items are lagged by a minimum of four months with respect to the interval that is covered by the returns.

The definitions of the LME, LBM, MOM, IMOM, REV, LVOL, LTO, SUV, INV, SM, NOA, P52H, APM, ROC, NSFY, and NSTY characteristics are provided in the enumerated list below. CRSP items names and Compustat mnemonics are shown in roman capital letters. The function lag($\cdot$) denotes the first annual lag of the argument. Because of the timing convention used to match annual data items with stock returns, a t subscript for a Compustat-based variable indicates that it is constructed using data for fiscal years that end in month $t-5$ or earlier.

1. $LME_{i,t}$ denotes the logarithm of the market equity of firm i in month t. Market equity is $|\mathrm{PRC}| \times (\mathrm{SHROUT}/1000)$, where PRC is the stock price from the monthly stock file. *LME* is updated monthly.
2. $LBM_{i,t}$ denotes the logarithm of the book-to-market equity ratio for stock i in month t. Market equity is $|\mathrm{PRC}| \times (\mathrm{SHROUT}/1000)$. Book equity is defined as in Fama and French (1992): shareholders equity (SEQ), plus balance-sheet deferred taxes and investment tax credit (TXDITC), if available, minus the book value of preferred stock, which is either its redemption value (PSTKRV), liquidation value (PSTKL), or par value (PSTK), in this order of preference. If SEQ is missing, then I substitute common equity plus preferred stock (CEQ plus PSTK), if available, or assets minus liabilities (AT minus LT), if available, in this order of preference. Market equity is updated monthly. Book equity is updated annually.
3. $MOM_{i,t}$ denotes the momentum for stock i in month t, which is measured using the stock return over the first 11 months of the 12-month interval. That is, $MOM_{i,t} = (\prod_{n=1}^{11}(1 + R_{i,t-n})) - 1$, where $R_{i,t}$ is the return for stock i in month t (RET from the monthly stock file). *MOM* is updated monthly.
4. $IMOM_{i,t}$ denotes the intermediate-range momentum for stock i in month t, which is measured using the stock return over the first 6 months of the prior 7-month interval. That is, $IMOM_{i,t} = (\prod_{n=1}^{6}(1 + R_{i,t-n})) - 1$, where $R_{i,t}$ is the return for stock i in month t (RET from the monthly stock file). *IMOM* is updated monthly.
5. $REV_{i,t}$ denotes the return for stock i in month t that is used to capture the short-term reversal

effect (i.e., the return that is skipped when computing $MOM_{i,t}$ and $IMOM_{i,t}$). It is RET from the monthly stock file. *REV* is updated monthly.

6. $LVOL_{i,t}$ denotes the logarithm of realized volatility for stock i in month t. I compute this variable as

$$LVOL_{i,t} = \log\left(\sum_{d=1}^{D_{i,t}} (R_{i,t_d} - \hat{m}_t(R_{i,t_d}))^2\right)^{1/2},$$

where R_{i,t_d} is the stock return for day d of month t (RET from the daily stock file), $D_{i,t}$ is the number of days with non-missing daily returns for stock i in month t, and $\hat{m}_t(R_{i,t_d}) = (1/D_{i,t}) \sum_{d=1}^{D_{i,t}} R_{i,t_d}$. I treat $LVOL_{i,t}$ as missing if $D_{i,t} < 13$. *LVOL* is updated monthly.

7. $LTO_{i,t}$ denotes the logarithm of turnover for stock i in month t. Turnover is VOL/(10 $\times$ SHROUT), where VOL and SHROUT are trading volume and shares outstanding from the monthly stock file. *LTO* is updated monthly.

8. $SUV_{i,t}$ denotes the standardized unexpected volume for stock i in month t. Let V_{i,t_d} and R_{i,t_d} denote the trading volume and stock return for day d of month t (VOL and RET from the daily stock file). To construct this variable for month t, I fit a linear regression model of the form

$$\log(1 + V_{i,t_d}) = \psi_0 + \psi_1 \max(0, R_{i,t_d}) + \psi_2 \min(0, R_{i,t_d}) + v_{i,t_d}$$

using a 66-day window that ends on the last trading day of month t. Using the residuals from this regression, I compute

$$SUV_{i,t} = \sum_{d=1}^{D_{i,t}} \hat{v}_{i,t_d} / \sqrt{D_{i,t}\hat{\sigma}_v^2},$$

where $\hat{v}_{i,t_d}$ is the residual for day t_d, $\hat{\sigma}_v^2$ is the sample variance of the regression residuals for the 66-day estimation window, and $D_{i,t}$ is the number of days with non-missing residuals for month t. The SUV is treated as missing if either $D_{i,t} < 13$ or the estimation window contains less than 39 non-missing daily observations. *SUV* is updated monthly.

9. $INV_{i,t}$ denotes investment for firm i in month t as measured by annual asset growth. It is the continuously-compounded growth rate of total firm assets over the prior fiscal year (log(AT/lag(AT))). *INV* is updated annually.

10. $SM_{i,t}$ denotes the ratio of annual sales (SALE) to market equity for firm i in month t. Market equity is |PRC| × (SHROUT/1000). Market equity is updated monthly. Sales is updated annually.

11. $NOA_{i,t}$ denotes the net operating assets to total assets ratio for firm i in month t. It is the ratio of operating assets minus operating liabilities for the prior fiscal year to beginning-of-year total assets. Operating assets are assets (AT) minus cash and short-term investments (CHE). Operating liabilities are assets (AT), minus debt in current liabilities (DLC), minus long-term debt (DLTT), minus minority interest (MIB), minus par value of preferred stock (PSTK), minus common equity (CEQ). Missing values of MIB and PSTK are set to zero. *NOA* is updated annually.

12. $P52H_{i,t}$ denotes the ratio of the stock price (|PRC|) for stock i in month t to its high price over the previous 52 weeks. The 52-week high price is computed from the daily stock file (the maximum value of |PRC| for the 52-week period). *P52H* is updated monthly.

13. $APM_{i,t}$ denotes the industry-adjusted profit margin for firm i in month t. Profit margin is operating income after depreciation (OIADP) divided by sales (SALE). Industry-adjusted profit margin is obtained by subtracting the average profit margin at the 48 Fama-French industry level (Fama and French, 1997). The industries are defined using standard industrial classification codes from Compustat (SICH). If the Compustat code is missing, I substitute the CRSP code (SICCD), if available. *APM* is updated annually.

14. $ROC_{i,t}$ denotes the return on cash for firm i in month t. *ROC* is the ratio of market equity plus long-term debt (DLTT) minus total assets (AT) to cash (CHE), where market equity is |PRC| × (SHROUT/1000). *ROC* is updated annually.

15. $NSFY_{i,t}$ denotes a fiscal-year-based measure of new share issues for firm i in month t. *NSFY* = log(*SO*/lag(*SO*)), where *SO* is the value of split-adjusted shares outstanding from Compustat (AJEX times CSHO). *NSFY* is updated annually.

16. $NSTY_{i,t}$ denotes a measure of new share issues over the prior 12 months for firm i in month t. *NSTY* = log(*SO*/lag(*SO*)), where *SO* is the value of split-adjusted shares outstanding using the shares outstanding (SHROUT) and the factor to adjust shares (FACSHR) from the CRSP

monthly stock file. *NSTY* is updated monthly.

References

Ait-Sahalia, Y., and D. Xiu (2019), Principal component analysis of high-frequency data, *Journal of the American Statistical Association* 114, 287–303.

Back, K., N. Kapadia, and B. Ostdiek (2015), Testing factor models on characteristic and covariance pure plays, Working Paper, Rice University.

Black, F., M. Jensen, and M. Scholes (1972), The capital asset pricing model: some empirical tests. In Jensen, M. (Ed.), Studies in the Theory of Capital Markets. Praeger, New York.

Faliva, M., and M. Zoia (2002), On a partitioned inversion formula having useful applications in econometrics, *Econometric Theory* 18, 525–530.

Fama, E., and K. French (1992), The cross-section of expected stock returns, *Journal of Finance* 47, 427–465.

Fama, E., and K. French (1997), Industry costs of equity, *Journal of Financial Economics* 43, 153–193.

Fama, E., and J. MacBeth (1973), Risk, return, and equilibrium: Empirical tests, *Journal of Political Economy* 81, 607–636.

Freedman, D. (1981), Bootstrapping regression models, *Annals of Statistics* 9, 1218–1228.

Freyberger, J., A. Neuhierl, and M. Weber (2019), Dissecting characteristics nonparametrically, forthcoming in *Review of Financial Studies*.

Gonalves, S. (2011), The moving blocks bootstrap for panel linear regression models with individual fixed effects, *Econometric Theory* 27, 1048–1082.

Gordon, S., and P. St-Amour (2000), A preference regime model of bull and bear markets, *American Economic Review* 90, 1019–1033.

Green, J., J. Hand, and X. F. Zhang (2017), The characteristics that provide independent information about average U.S. monthly stock returns, *Review of Financial Studies* 30, 4389–436.

Guidolin, M., and A. Timmermann (2005), Economic implications of bull and bear regimes in UK stock and bond returns, *The Economic Journal* 115, 111–143.

Hamilton, J. (1989), A new approach to the economic analysis of nonstationary time series and the business cycle, *Econometrica* 57, 357–384.

Hamilton, J. (1990), Analysis of time series subject to change in regime, *Journal of Econometrics* 45, 39–70.

Harvey, C., Y. Liu, and H. Zhu (2016), … and the cross-section of expected returns, *Review of Financial Studies* 29, 5–68.

He, Z., B. Kelly, and A. Manela (2017), Intermediary asset pricing: New evidence from many asset classes, *Journal of Financial Economics* 126, 1–35.

Kelly, B., S. Pruitt, and Y. Su (2017), Instrumented principal component analysis, Working paper, University of Chicago.

Kelly, B., S. Pruitt, and Y. Su (2019), Characteristics are covariances: A unified model of risk and return, *Journal of Financial Economics* 134, 501–524.

Kim, C. (1994), Dynamic linear models with Markov-switching, *Journal of Econometrics* 60, 1–22.

Kirby, C. (1997), Measuring the predictable variation in stock and bond returns, *Review of Financial Studies* 10, 579–630.

Kirby, C. (2019a), Firm characteristics, cross-sectional regression estimates, and asset pricing tests, forthcoming, *Review of Asset Pricing Studies*.

Kirby, C. (2019b), Regime-switching PCA and industry costs of equity, Working paper, University of North Carolina at Charlotte

Kole, E., and D. van Dijk (2017), How to identify and forecast bull and bear markets? *Journal of Applied Econometrics* 32, 120–139.

Kroencke, T. (2017), Asset pricing without garbage, *Journal of Finance* 72, 47–98.

Lettau, M., S. Ludvigson, and S. Ma (2019), Capital share risk in U.S. asset pricing, *Journal of Finance* Advance Access published March 30, 2019, 10.1111/jofi.12772.

Lawley, D., and A. Maxwell (1971), *Factor analysis as a statistical method*, 2nd ed. Elsevier, New York.

Lunde, A. and A. Timmermann, (2004). Duration dependence in stock prices: An analysis of bull and bear markets, *Journal of Business and Economic Statistics* 22, 253–273.

Maheu, J., and H. McCurdy (2000), Identifying bull and bear markets in stock returns, *Journal of Business and Economic Statistics* 18, 100–112.

Pagan, A., and K. Sossounov (2003), A simple framework for analysing bull and bear markets, *Journal of Applied Econometrics* 18, 23–46.

Stambaugh, R., and Y. Yuan, (2017), Mispricing factors, *Review of Financial Studies* 30, 1270–315.

Tipping, M., and C. Bishop (1999), Probabilistic principal component analysis, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 61, 611–622.

Whitelaw, R. (2000), Stock market risk and return: an equilibrium approach, *Review of Financial Studies* 13, 521–548.

Table 1
Summary Statistics for Firm Characteristics

	Mean	Vol	Percentiles						
			1st	10th	25th	50th	75th	90th	99th
LME	5.04	2.21	0.58	2.25	3.42	4.92	6.55	7.99	10.47
LBM	−0.54	0.96	−3.34	−1.75	−1.08	−0.46	0.07	0.56	1.65
MOM	0.13	0.60	−0.80	−0.45	−0.20	0.05	0.32	0.70	2.31
IMOM	0.07	0.41	−0.67	−0.34	−0.15	0.03	0.22	0.47	1.50
REV	0.01	0.15	−0.37	−0.15	−0.06	0.00	0.07	0.17	0.53
LVOL	−2.16	0.72	−3.71	−2.97	−2.60	−2.17	−1.71	−1.28	−0.49
LTO	−3.40	1.33	−6.95	−5.09	−4.24	−3.33	−2.47	−1.76	−0.67
SUV	−0.00	2.62	−6.36	−3.24	−1.64	−0.02	1.60	3.21	6.79
INV	0.13	0.34	−0.59	−0.13	−0.01	0.08	0.19	0.42	1.50
SM	2.41	4.54	0.01	0.20	0.46	1.08	2.51	5.44	22.17
NOA	0.65	0.50	−0.15	0.16	0.41	0.66	0.82	0.99	2.19
P52H	1.60	1.51	1.00	1.02	1.08	1.24	1.60	2.33	6.73
APM	−0.63	8.32	−11.93	−0.20	−0.05	0.00	0.06	0.14	0.80
ROC	1.16	103.86	−222.4	−28.19	−9.38	−0.79	5.20	21.82	263.88
NSFY	0.07	0.27	−0.16	−0.02	0.00	0.01	0.04	0.20	1.26
NSTY	0.04	0.15	−0.17	−0.03	0.00	0.00	0.03	0.16	0.71

The table reports the mean (Mean), volatility (Vol), and selected percentiles of the values of the firm characteristics. The characteristics are the log of market equity (LME), the log of the book-to-market equity ratio (LBM), an eleven-month stock return (MOM), a six-month stock return (IMOM), a one-month stock return (REV), the log of realized monthly volatility (LVOL), the log of monthly share turnover (LTO), standardized unexpected monthly volume (SUV), the annual growth rate of total assets (INV), the ratio of annual sales to market equity (SM), the ratio of annual net operating assets to total assets (NOA), the stock price relative to its 52 week high price (P52H), the annual industry-adjusted profit margin (APM), the annual return on cash (ROC), a fiscal-year-based measure of annual share issuance (NSFY), and a measure of annual share issuance for the most recent 12 months (NSTY). Appendix B contains detailed variable definitions. I compute the statistics for the sample that is used to fit the Fama and MacBeth (1973) regressions whose results are summarized in Table 2. Stocks that have a missing value of the regressand or any of the regressors for a given month are excluded from the regression for that month. The sample period is May 1967 to December 2018.

Table 2
Average Coefficient Estimates and *t*-Statistics for Linear and Nonlinear Fama-MacBeth Regressions

	Linear model		Nonlinear model that uses quartile indicators as regressors: Quartile 1		Quartile 2		Quartile 3		Quartile 4	
Characteristic	Slope	*t*-stat	Coeff	*t*-stat	Coeff	*t*-stat	Coeff	*t*-stat	Coeff	*t*-stat
LME	−0.25	−4.39	0.41	5.06	−0.08	−1.88	−0.07	−1.55	−0.27	−4.02
LBM	0.18	4.38	−0.25	−6.00	−0.13	−5.40	0.01	0.49	0.37	9.11
MOM	0.46	8.22	−0.26	−4.28	−0.10	−3.94	0.02	0.79	0.34	6.10
IMOM	0.02	0.65	−0.13	−3.16	−0.00	−0.05	0.01	0.29	0.12	3.29
REV	−0.74	−16.46	1.02	17.72	0.15	7.33	−0.24	−10.43	−0.94	−18.11
LVOL	−0.24	−5.11	0.02	0.45	0.19	6.77	0.09	3.42	−0.30	−4.98
LTO	−0.01	−0.13	−0.12	−2.03	0.13	4.69	0.11	3.79	−0.11	−1.82
SUV	0.52	21.44	−0.58	−20.56	−0.15	−8.00	0.15	8.09	0.58	19.01
INV	−0.19	−6.02	0.20	5.45	−0.00	−0.16	−0.02	−1.24	−0.17	−5.21
SM	0.09	2.84	−0.23	−4.37	−0.07	−2.97	0.04	1.57	0.26	5.34
NOA	−0.16	−5.87	0.29	8.01	0.06	2.46	−0.04	−1.84	−0.31	−9.11
P52H	0.48	6.20	0.23	4.30	0.04	1.27	−0.24	−8.43	−0.03	−0.38
APM	0.04	1.61	−0.12	−3.06	−0.10	−4.35	0.01	0.46	0.21	7.59
ROC	0.03	2.08	−0.17	−4.13	0.03	1.13	0.06	1.76	0.09	2.42
NSFY	−0.07	−3.95	−0.01	−0.38	0.01	0.60	0.09	4.38	−0.09	−3.18
NSTY	−0.15	−8.88	0.10	4.07	0.01	0.25	0.02	0.93	−0.12	−4.46

The table reports the average estimated slope coefficients and Fama and MacBeth (1973) *t*-statistics for monthly cross-sectional regressions of monthly excess percentage stock returns on two different sets of predetermined regressors. All regressions include an intercept. The explanatory variables for the linear model are the log of market equity (LME), the log of the book-to-market equity ratio (LBM), an eleven-month stock return (MOM), a six-month stock return (IMOM), a one-month stock return (REV), the log of realized monthly volatility (LVOL), the log of monthly share turnover (LTO), standardized unexpected monthly volume (SUV), the annual growth rate of total assets (INV), the ratio of annual sales to market equity (SM), the ratio of annual net operating assets to total assets (NOA), the stock price relative to its 52 week high price (P52H), the annual industry-adjusted profit margin (APM), the annual return on cash (ROC), a fiscal-year-based measure of annual share issuance (NSFY), and a measure of annual share issuance for the most recent 12 months (NSTY). I standardized each variable to have a cross-sectional mean of zero and a cross-sectional variance of one for every month in the sample period. Appendix B contains detailed variable definitions. The explanatory variables for the nonlinear model are indicator variables that are constructed by the sorting stocks monthly into quartiles on the basis of each characteristic. In particular, the ith indicator variable for the jth characteristic takes a value of one if the stock falls into the ith quartile of the jth characteristic's cross-sectional distribution for the month and zero otherwise. The coefficients on the four indicator variables for a given characteristic are constrained to sum to zero. I *do not* standardize the indicator variables for the regressions whose results are reported in the table (they are standardized for the purposes of the asset pricing tests). Hence, the estimated coefficient for the ith indicator captures the marginal effect of changing the value of the jth characteristic from its average value across all stocks to its average value for the ith quartile. Stocks that have a missing value of the regressand or any of the regressors for a given month are excluded from the regression for that month. The sample period is May 1967 to December 2018 (620 monthly regressions).

Table 3
Asset Pricing Tests for Regression-Based Managed Portfolios Under the Single-Regime Model

A. Results using the ALS algorithm

	RMSE for 17 portfolios with $\boldsymbol{r}_{p,t} = \boldsymbol{C}_{t-1}^{+}\boldsymbol{r}_t$							RMSE for 49 portfolios with $\boldsymbol{r}_{p,t} = (\boldsymbol{X}_{t-1}\boldsymbol{R}_{\perp})^{+}\boldsymbol{r}_t$						
	K=0	K=1	K=2	K=3	K=4	K=5	K=6	K=0	K=1	K=2	K=3	K=4	K=5	K=6
All stocks	0.36	0.29	0.27	0.27	0.27	0.15	0.11	0.19	0.15	0.15	0.14	0.14	0.10	0.10
No micro-caps	0.22	0.17	0.16	0.15	0.15	0.12	0.10	0.13	0.10	0.09	0.09	0.09	0.08	0.07

B. Results using the direct solution method

	RMSE for 17 portfolios with $\boldsymbol{r}_{p,t} = \boldsymbol{C}_{t-1}^{+}\boldsymbol{r}_t$							RMSE for 49 portfolios with $\boldsymbol{r}_{p,t} = (\boldsymbol{X}_{t-1}\boldsymbol{R}_{\perp})^{+}\boldsymbol{r}_t$						
	K=0	K=1	K=2	K=3	K=4	K=5	K=6	K=0	K=1	K=2	K=3	K=4	K=5	K=6
All stocks	0.36	0.29	0.26	0.26	0.26	0.15	0.12	0.19	0.15	0.15	0.14	0.14	0.10	0.10
No micro-caps	0.22	0.16	0.16	0.15	0.15	0.11	0.09	0.13	0.09	0.09	0.09	0.09	0.07	0.06

C. Bootstrap critical values

	Resampling from $\{\boldsymbol{r}_t, \boldsymbol{C}_{t-1}\}_{t=1}^{T}$						Resampling from $\{\boldsymbol{r}_t, \boldsymbol{X}_{t-1}\}_{t=1}^{T}$					
	ALS algorithm			Direct solution			ALS algorithm			Direct solution		
K	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
1	0.05	0.06	0.06	0.04	0.05	0.05	0.03	0.03	0.03	0.02	0.03	0.03
2	0.07	0.07	0.08	0.05	0.05	0.06	0.04	0.04	0.05	0.02	0.03	0.03
3	0.06	0.07	0.08	0.05	0.06	0.07	0.04	0.04	0.05	0.03	0.04	0.04
4	0.06	0.07	0.08	0.05	0.06	0.07	0.03	0.04	0.04	0.03	0.03	0.04
5	0.11	0.13	0.15	0.11	0.12	0.15	0.06	0.06	0.07	0.05	0.06	0.07
6	0.11	0.12	0.15	0.09	0.10	0.13	0.07	0.07	0.08	0.06	0.07	0.08
No micro-caps												
1	0.05	0.05	0.06	0.04	0.04	0.05	0.03	0.03	0.03	0.02	0.02	0.03
2	0.04	0.05	0.05	0.04	0.04	0.05	0.03	0.03	0.03	0.02	0.03	0.03
3	0.05	0.05	0.06	0.04	0.04	0.05	0.03	0.03	0.03	0.02	0.03	0.03
4	0.07	0.08	0.09	0.06	0.06	0.09	0.03	0.03	0.04	0.03	0.03	0.04
5	0.10	0.11	0.12	0.11	0.12	0.13	0.05	0.06	0.07	0.06	0.06	0.07
6	0.07	0.07	0.09	0.06	0.07	0.08	0.05	0.05	0.06	0.04	0.05	0.06

The table reports estimates of the root mean squared pricing error (RMSE) in % per month for two sets of regression-based managed portfolios. The estimates are obtained via instrumented principal components analysis (IPCA). The vector of excess returns for the first set of portfolios is given by $\boldsymbol{r}_{p,t} = \boldsymbol{C}_{t-1}^{+}\boldsymbol{r}_t$, where $\boldsymbol{r}_t$ is the $N_t \times 1$ vector of excess stock returns for month t and $\boldsymbol{C}_{t-1}^{+}$ is the left pseudoinverse of $\boldsymbol{C}_{t-1}$, an $N_t \times 17$ matrix whose nth row contains a 1 along with the values of 16 predetermined characteristics for the nth firm (see Table 1). I standardize each column of $\boldsymbol{C}_{t-1}$ except the first to have a mean of zero and a variance of one for every month. The vector of excess returns for the second set of portfolios is given by $\boldsymbol{r}_{p,t} = (\boldsymbol{X}_{t-1}\boldsymbol{R}_{\perp})^{+}\boldsymbol{r}_t$, where $(\boldsymbol{X}_{t-1}\boldsymbol{R}_{\perp})^{+}$ is the left pseudoinverse of an $N_t \times 49$ matrix, $\boldsymbol{X}_t = (\boldsymbol{1}, \boldsymbol{I}_{2,t}, \ldots, \boldsymbol{I}_{17,t})$ is a $N_t \times 65$ matrix that is formed by conducting monthly sorts of stocks into quartiles using each of 16 firm characteristics (see Table 1), and $\boldsymbol{R}_{\perp}$ is the orthogonal complement of a 65×16 matrix $\boldsymbol{R}$ that is used to restrict the coefficients on the four indicator variables associated with each characteristic to sum to zero. The nth element of the ith column of the $N_t \times 4$ matrix $\boldsymbol{I}_{j,t}$ is equal to one if the nth firm is in the ith quartile of the sorted values of the jth characteristic for month t and equal to zero otherwise. I standardize each column of $\boldsymbol{I}_{j,t}$ to have a unit cross-sectional second moment for every value of $j \in (2, \ldots, 17)$. Panel A summarizes the results produced by the alternating least squares (ALS) algorithm of Kelly et al. (2019). Panels B summarizes the results produced by the direct solution method using conventional PCA (see Section 2 for further elaboration). Panel C reports critical values for testing the hypothesis of exact factor pricing based on 500 bootstrap trials. The bootstrap is implemented by drawing randomly with replacement from either $\{\boldsymbol{r}_t, \boldsymbol{C}_{t-1}\}_{t=1}^{T}$ or $\{\boldsymbol{r}_t, \boldsymbol{X}_{t-1}\}_{t=1}^{T}$. The no-micro-caps sample excludes stocks that are smaller than the 20th percentile of the NYSE market-equity distribution. The sample period is May 1967 to December 2018.

Table 4
Estimated Market Index Parameters and Expected Regime Durations for the Two-Regime Model

A. Using $\tilde{\boldsymbol{r}}_{p,t} = \bar{\boldsymbol{W}}^{1/2}\boldsymbol{C}_{t-1}^{+}\boldsymbol{r}_t$ to fit the model

	95% CI for the mean						95% CI for volatility						95% CI for expected duration					
	Regime 1			Regime 2			Regime 1			Regime 2			Regime 1			Regime 2		
K	LCL	Est	UCL	LCL	Est	UCL	LCL	Est	UCL	LCL	Est	UCL	LCL	Est	UCL	LCL	Est	UCL
1	−0.34	0.11	0.53	1.48	2.77	3.87	4.57	4.92	5.21	6.65	7.66	8.30	5.49	6.77	8.68	1.94	2.33	2.81
2	0.10	0.50	0.86	0.27	1.88	3.53	4.16	4.46	4.74	8.04	9.29	10.27	8.27	10.87	14.32	2.33	2.98	3.78
3	0.22	0.63	0.99	−0.09	1.41	3.00	4.25	4.56	4.85	7.91	9.15	10.11	8.11	10.63	14.12	2.29	2.92	3.72
4	−0.13	0.25	0.62	0.96	2.24	3.58	3.97	4.27	4.55	7.65	8.64	9.51	6.68	8.55	11.19	2.52	3.16	3.91
5	−0.22	0.18	0.54	1.08	2.38	3.65	3.91	4.19	4.48	7.58	8.60	9.46	6.47	8.26	10.74	2.56	3.19	3.93
6	−0.25	0.16	0.55	1.22	2.53	3.76	4.01	4.32	4.60	7.54	8.55	9.44	6.70	8.56	11.14	2.48	3.10	3.91
No micro-caps																		
1	0.06	0.45	0.81	0.13	1.53	3.17	4.16	4.46	4.73	6.88	8.04	8.98	7.94	10.47	13.61	2.08	2.62	3.37
2	0.19	0.61	0.97	−0.44	0.93	2.54	4.15	4.43	4.71	7.12	8.36	9.23	10.02	13.57	19.45	2.35	3.13	4.07
3	0.21	0.62	0.97	−0.53	0.86	2.43	4.00	4.27	4.55	7.21	8.31	9.15	9.46	13.08	18.69	2.63	3.53	4.73
4	0.18	0.55	0.90	−0.28	1.05	2.44	3.83	4.11	4.37	7.20	8.24	9.06	8.24	11.43	15.68	2.73	3.57	4.60
5	0.12	0.50	0.84	−0.09	1.26	2.71	3.92	4.19	4.46	7.20	8.28	9.18	9.59	13.23	18.96	2.76	3.79	5.04
6	0.07	0.47	0.82	0.09	1.36	2.81	3.94	4.21	4.48	7.11	8.19	9.06	8.90	12.00	16.45	2.68	3.52	4.61

B. Using $\tilde{\boldsymbol{r}}_{p,t} = \bar{\boldsymbol{W}}^{1/2}(\boldsymbol{X}_{t-1}\boldsymbol{R}_{\perp})^{+}\boldsymbol{r}_t$ to fit the model

K	LCL	Est	UCL	LCL	Est	UCL	LCL	Est	UCL	LCL	Est	UCL	LCL	Est	UCL	LCL	Est	UCL
1	−0.30	0.15	0.66	1.33	2.61	3.66	4.51	4.92	5.24	6.87	7.66	8.40	5.19	6.43	7.99	1.89	2.29	2.71
2	−0.02	0.39	0.84	0.70	2.36	3.85	4.19	4.53	4.83	8.00	9.19	10.25	9.02	11.85	15.76	2.38	3.11	4.08
3	−0.15	0.24	0.62	1.24	2.67	4.05	4.04	4.38	4.65	7.96	9.03	10.02	6.72	8.47	10.68	2.02	2.51	3.10
4	−0.15	0.27	0.67	1.05	2.41	3.65	4.20	4.54	4.84	7.56	8.55	9.50	6.23	7.77	9.76	2.01	2.52	3.09
5	−0.17	0.22	0.63	1.31	2.66	3.95	4.11	4.44	4.71	7.76	8.79	9.75	7.22	9.15	11.73	2.21	2.85	3.56
6	−0.26	0.14	0.53	1.23	2.66	3.88	4.10	4.42	4.72	7.53	8.49	9.45	6.36	7.95	10.05	2.22	2.76	3.40
No micro-caps																		
1	0.05	0.48	0.88	−0.01	1.38	2.48	4.19	4.51	4.79	6.73	7.75	8.70	7.19	9.01	11.43	1.98	2.48	3.05
2	0.16	0.54	0.93	−0.28	1.11	2.19	4.11	4.41	4.68	6.93	7.86	8.75	7.50	9.54	12.11	2.17	2.73	3.40
3	0.33	0.70	1.11	−0.86	0.55	1.74	3.99	4.30	4.55	7.24	8.25	9.22	7.89	10.09	12.68	2.17	2.74	3.42
4	0.21	0.60	1.02	−0.43	0.89	2.00	3.85	4.15	4.41	7.27	8.24	9.12	7.20	9.13	11.64	2.17	2.78	3.49
5	0.02	0.42	0.87	0.13	1.49	2.59	3.93	4.23	4.49	7.10	8.06	9.01	7.88	10.08	12.81	2.38	2.99	3.78
6	0.06	0.45	0.88	0.09	1.30	2.34	3.81	4.08	4.34	7.09	8.06	8.96	7.58	9.69	12.44	2.55	3.29	4.16

The table reports estimates for the market index that result from fitting a two-state Markov switching model in which $\tilde{\boldsymbol{r}}_{p,t}|s_t \sim N(\boldsymbol{\kappa}_{s_t}, \boldsymbol{\Lambda}_{s_t}\boldsymbol{\Lambda}'_{s_t} + \tau^2_{s_t}\mathbf{I})$, where $s_t \in (1,2)$ follows a first-order Markov chain. In panel A, $\tilde{\boldsymbol{r}}_{p,t} = \bar{\boldsymbol{W}}^{1/2}\boldsymbol{C}_{t-1}^{+}\boldsymbol{r}_t$, where $\boldsymbol{r}_t$ is the $N_t \times 1$ vector of excess stock returns for month t, $\boldsymbol{C}_{t-1}^{+}$ is the left pseudoinverse of an $N_t \times 17$ matrix $\boldsymbol{C}_{t-1}$ whose nth row contains a 1 along with the values of 16 characteristics for the nth firm (see Table 1), and $\bar{\boldsymbol{W}} = (\sum_{t=1}^{T} N_t)^{-1}\sum_{t=1}^{T}\boldsymbol{C}'_{t-1}\boldsymbol{C}_{t-1}$. I standardize each column of $\boldsymbol{C}_{t-1}$ except the first to have a mean of zero and variance of one for every month. In panel B, $\tilde{\boldsymbol{r}}_{p,t} = \bar{\boldsymbol{W}}^{1/2}(\boldsymbol{X}_{t-1}\boldsymbol{R}_{\perp})^{+}\boldsymbol{r}_t$, where $(\boldsymbol{X}_{t-1}\boldsymbol{R}_{\perp})^{+}$ is the left pseudoinverse of an $N_t \times 49$ matrix, $\boldsymbol{X}_t = (\mathbf{1}, \boldsymbol{I}_{2,t}, \ldots, \boldsymbol{I}_{17,t})$ is a $N_t \times 65$ matrix that is formed by conducting monthly sorts of stocks into quartiles using each of 16 firm characteristics, $\boldsymbol{R}_{\perp}$ is the orthogonal complement of a 65×16 matrix $\boldsymbol{R}$ that is used to restrict the coefficients on the four indicator variables associated with each characteristic to sum to zero, and $\bar{\boldsymbol{W}} = (\sum_{t=1}^{T} N_t)^{-1}\sum_{t=1}^{T}\boldsymbol{R}'_{\perp}\boldsymbol{X}'_{t-1}\boldsymbol{X}_{t-1}\boldsymbol{R}_{\perp}$. The nth element of the ith column of the $N_t \times 4$ matrix $\boldsymbol{I}_{j,t}$ is equal to one if the nth firm is in the ith quartile of the sorted values of the jth characteristic for month t and equal to zero otherwise. I standardize each column of $\boldsymbol{I}_{j,t}$ to have a unit cross-sectional second moment for every value of $j \in (2, \ldots, 17)$. The confidence intervals are computed via a parametric bootstrap using 500 trials. The no-micro-caps sample excludes stocks that are smaller than the 20th percentile of the NYSE market-equity distribution. The sample period is May 1967 to December 2018.

Table 5
Regime-Specific Asset Pricing Tests for Managed Portfolios Constructed via Linear Regression

A. Results for $\tilde{\boldsymbol{r}}_{p,t} = \bar{\boldsymbol{W}}^{1/2}\boldsymbol{C}_{t-1}^{+}\boldsymbol{r}_t$

			RMSE & 1% crit val				95% CI for second pass R^2					
	Avg ret R^2		1st regime		2nd regime		1st regime			2nd regime		
K	1st	2nd	RMSE	CV	RMSE	CV	LCL	R^2	UCL	LCL	R^2	UCL
0	–	–	0.23	–	0.97	–	–	–	–	–	–	–
1	0.17	0.22	0.23	0.04	0.48	0.16	0.00	0.00	0.21	0.50	0.75	0.84
2	0.26	0.33	0.23	0.04	0.41	0.16	0.03	0.21	0.44	0.35	0.72	0.87
3	0.32	0.43	0.23	0.03	0.42	0.13	0.10	0.33	0.55	0.32	0.66	0.85
4	0.37	0.47	0.22	0.03	0.40	0.10	0.04	0.10	0.32	0.54	0.77	0.88
5	0.40	0.51	0.09	0.06	0.30	0.12	0.73	0.86	0.90	0.74	0.88	0.94
6	0.43	0.53	0.09	0.06	0.23	0.13	0.76	0.86	0.90	0.83	0.93	0.97
No micro-caps												
0	–	–	0.19	–	0.48	–	–	–	–	–	–	–
1	0.18	0.18	0.16	0.04	0.28	0.17	0.00	0.27	0.57	0.03	0.67	0.87
2	0.25	0.31	0.16	0.03	0.26	0.15	0.06	0.42	0.67	0.04	0.53	0.86
3	0.29	0.40	0.14	0.03	0.25	0.11	0.26	0.57	0.75	0.09	0.51	0.85
4	0.33	0.45	0.10	0.04	0.24	0.10	0.50	0.71	0.83	0.12	0.59	0.86
5	0.37	0.48	0.10	0.03	0.19	0.14	0.49	0.69	0.83	0.40	0.79	0.93
6	0.40	0.50	0.06	0.04	0.18	0.11	0.78	0.90	0.94	0.49	0.83	0.94
B. Results for $\boldsymbol{r}_{p,t} = \boldsymbol{C}_{t-1}^{+}\boldsymbol{r}_t$												
0	–	–	0.23	–	0.93	–	–	–	–	–	–	–
1	0.15	0.16	0.23	0.05	0.52	0.21	0.00	0.00	0.22	0.41	0.69	0.80
2	0.22	0.22	0.23	0.05	0.43	0.23	0.08	0.26	0.47	0.36	0.70	0.85
3	0.28	0.33	0.23	0.05	0.45	0.19	0.13	0.35	0.55	0.32	0.62	0.81
4	0.31	0.38	0.22	0.04	0.42	0.14	0.06	0.13	0.34	0.48	0.73	0.85
5	0.33	0.42	0.10	0.08	0.35	0.16	0.64	0.80	0.87	0.65	0.82	0.91
6	0.36	0.46	0.09	0.07	0.26	0.16	0.71	0.84	0.89	0.81	0.91	0.95
No micro-caps												
0	–	–	0.18	–	0.47	–	–	–	–	–	–	–
1	0.15	0.14	0.15	0.05	0.27	0.23	0.01	0.32	0.61	0.03	0.67	0.86
2	0.19	0.25	0.15	0.04	0.29	0.19	0.08	0.47	0.70	0.01	0.43	0.83
3	0.22	0.31	0.14	0.04	0.27	0.15	0.16	0.54	0.74	0.05	0.42	0.83
4	0.26	0.36	0.11	0.05	0.26	0.12	0.43	0.69	0.82	0.07	0.52	0.85
5	0.30	0.40	0.11	0.04	0.21	0.17	0.41	0.65	0.80	0.33	0.77	0.92
6	0.34	0.44	0.07	0.05	0.18	0.14	0.68	0.85	0.92	0.49	0.83	0.94

The table reports regime-specific estimates of the root mean squared pricing error (RMSE) for two sets of regression-based managed portfolios. Regime 1 is the low-volatility regime and regime 2 is the high-volatility regime. Panel A is for a set of 17 portfolios whose vector of excess returns for period t is given by $\tilde{\boldsymbol{r}}_{p,t} = \bar{\boldsymbol{W}}^{1/2}\boldsymbol{C}_{t-1}^{+}\boldsymbol{r}_t$, where $\boldsymbol{r}_t$ is the $N_t \times 1$ vector of excess stock returns for month t, $\boldsymbol{C}_{t-1}^{+}$ is the left pseudoinverse of an $N_t \times 17$ matrix $\boldsymbol{C}_{t-1}$ whose nth row contains a 1 along with the values of 16 predetermined characteristics for the nth firm (see Table 1), and $\bar{\boldsymbol{W}} = (\sum_{t=1}^{T} N_t)^{-1} \sum_{t=1}^{T} \boldsymbol{C}_{t-1}'\boldsymbol{C}_{t-1}$. I standardize each column of $\boldsymbol{C}_{t-1}$ except the first to have a mean of zero and variance of one for every month. Panel B is for a set of 17 portfolios whose vector of excess returns for period t is given by $\boldsymbol{r}_{p,t} = \boldsymbol{C}_{t-1}^{+}\boldsymbol{r}_t$. Hence, each of these portfolios is a pure play on the associated firm characteristic. The regime-specific parameters are estimated by fitting a regime-switching model to $\{\tilde{\boldsymbol{r}}_{p,1}, \tilde{\boldsymbol{r}}_{p,2}, \ldots, \tilde{\boldsymbol{r}}_{p,T}\}$. Under the model, $\tilde{\boldsymbol{r}}_{p,t}|s_t \sim N(\boldsymbol{\kappa}_{s_t}, \boldsymbol{\Lambda}_{s_t}\boldsymbol{\Lambda}_{s_t}' + \tau_{s_t}^2\mathbf{I})$, where $s_t \in (1,2)$ is a discrete state variable that follows a first-order Markov chain. The first two columns report the average estimated fraction of the variation of the excess portfolio returns that is explained by the latent factors (Avg ret R^2). The next six columns report the estimated RMSEs for the first and second regimes along with the 1% critical values (CVs) for testing the hypothesis of exact factor pricing. The final six columns report 95% confidence intervals for the R^2 statistics obtained by regressing the estimated vector of excess portfolio returns on the estimated matrix of factor loadings (the second-pass regressions). I use a parametric bootstrap procedure to compute the critical values and confidence intervals. The no-micro-caps sample excludes stocks that are smaller than the 20th percentile of the NYSE market-equity distribution. The sample period is May 1967 to Dec 2018.

Table 6
Regime-Specific Asset Pricing Tests for Managed Portfolios Constructed via Nonlinear Regression

A. Results for $\tilde{\boldsymbol{r}}_{p,t} = \bar{\boldsymbol{W}}^{1/2}(\boldsymbol{X}_{t-1}\boldsymbol{R}_{\perp})^{+}\boldsymbol{r}_t$

	Avg ret R^2		RMSE & 1% crit val				95% CI for second pass R^2					
			1st regime		2nd regime		1st regime			2nd regime		
K	1st	2nd	RMSE	CV	RMSE	CV	LCL	R^2	UCL	LCL	R^2	UCL
0	–	–	0.13	–	0.48	–	–	–	–	–	–	–
1	0.15	0.18	0.13	0.02	0.22	0.08	0.00	0.00	0.27	0.51	0.79	0.87
2	0.25	0.27	0.13	0.02	0.21	0.08	0.00	0.09	0.40	0.30	0.77	0.89
3	0.32	0.39	0.12	0.02	0.20	0.06	0.17	0.23	0.40	0.51	0.82	0.90
4	0.35	0.45	0.11	0.02	0.20	0.06	0.18	0.25	0.45	0.48	0.79	0.89
5	0.38	0.49	0.09	0.02	0.17	0.07	0.40	0.48	0.60	0.69	0.88	0.93
6	0.41	0.51	0.08	0.03	0.14	0.06	0.52	0.59	0.68	0.76	0.91	0.95
Excluding micro-caps												
0	–	–	0.11	–	0.25	–	–	–	–	–	–	
1	0.13	0.14	0.09	0.02	0.15	0.08	0.00	0.30	0.63	0.00	0.67	0.85
2	0.20	0.31	0.09	0.02	0.14	0.07	0.05	0.37	0.66	0.08	0.61	0.84
3	0.25	0.37	0.07	0.02	0.15	0.06	0.41	0.68	0.83	0.03	0.28	0.74
4	0.29	0.43	0.06	0.02	0.15	0.05	0.44	0.69	0.84	0.08	0.48	0.80
5	0.33	0.46	0.06	0.02	0.12	0.06	0.48	0.65	0.81	0.36	0.80	0.91
6	0.36	0.49	0.05	0.02	0.11	0.05	0.58	0.73	0.86	0.44	0.81	0.92

B. Results for $\boldsymbol{r}_{p,t} = (\boldsymbol{X}_{t-1}\boldsymbol{R}_{\perp})^{+}\boldsymbol{r}_t$

K	Avg ret R^2 1st	Avg ret R^2 2nd	1st regime RMSE	1st regime CV	2nd regime RMSE	2nd regime CV	1st regime LCL	1st regime R^2	1st regime UCL	2nd regime LCL	2nd regime R^2	2nd regime UCL
0	–	–	0.13	–	0.46	–	–	–	–	–	–	–
1	0.09	0.10	0.13	0.03	0.22	0.10	0.00	0.01	0.30	0.44	0.76	0.85
2	0.14	0.16	0.13	0.02	0.22	0.10	0.02	0.13	0.43	0.27	0.73	0.86
3	0.17	0.24	0.13	0.03	0.21	0.08	0.12	0.18	0.36	0.47	0.79	0.88
4	0.20	0.27	0.12	0.02	0.21	0.07	0.05	0.12	0.37	0.42	0.76	0.87
5	0.23	0.31	0.11	0.03	0.18	0.08	0.26	0.35	0.51	0.65	0.86	0.92
6	0.26	0.33	0.10	0.03	0.16	0.08	0.36	0.45	0.58	0.70	0.89	0.93
Excluding micro-caps												
0	–	–	0.11	–	0.25	–	–	–	–	–	–	
1	0.08	0.08	0.09	0.02	0.15	0.10	0.01	0.36	0.65	0.01	0.65	0.84
2	0.12	0.18	0.09	0.02	0.15	0.08	0.07	0.43	0.70	0.05	0.56	0.82
3	0.14	0.21	0.08	0.02	0.16	0.08	0.29	0.64	0.81	0.02	0.23	0.70
4	0.17	0.26	0.07	0.02	0.15	0.07	0.35	0.65	0.82	0.05	0.44	0.78
5	0.19	0.29	0.07	0.02	0.13	0.07	0.35	0.58	0.78	0.30	0.77	0.89
6	0.24	0.31	0.06	0.02	0.11	0.07	0.47	0.67	0.83	0.40	0.79	0.90

The table reports regime-specific estimates of the root mean squared pricing error (RMSE) for two sets of regression-based managed portfolios. Regime 1 is the low-volatility regime and regime 2 is the high-volatility regime. Panel A is for a set of 49 portfolios whose vector of excess returns for period t is given by $\tilde{\boldsymbol{r}}_{p,t} = \bar{\boldsymbol{W}}^{1/2}(\boldsymbol{X}_{t-1}\boldsymbol{R}_{\perp})^{+}\boldsymbol{r}_t$, where $\boldsymbol{r}_t$ is the $N_t \times 1$ vector of excess stock returns for month t, $(\boldsymbol{X}_{t-1}\boldsymbol{R}_{\perp})^{+}$ is the left pseudoinverse of an $N_t \times 49$ matrix, $\boldsymbol{X}_t = (\mathbf{1}, \boldsymbol{I}_{2,t}, \ldots, \boldsymbol{I}_{17,t})$ is a $N_t \times 65$ matrix that is formed by conducting monthly sorts of stocks into quartiles using each of 16 firm characteristics (see Table 1), $\bar{\boldsymbol{W}} = (\sum_{t=1}^{T} N_t)^{-1} \sum_{t=1}^{T} \boldsymbol{R}'_{\perp}\boldsymbol{X}'_{t-1}\boldsymbol{X}_{t-1}\boldsymbol{R}_{\perp}$, and $\boldsymbol{R}_{\perp}$ is the orthogonal complement of a 65×16 matrix $\boldsymbol{R}$ that is used to restrict the coefficients on the four indicator variables associated with each characteristic to sum to zero. The nth element of the ith column of the $N_t \times 4$ matrix $\boldsymbol{I}_{j,t}$ is equal to one if the nth firm is contained in ith quartile of the sorted values of the jth characteristic for month t and equal to zero otherwise. I standardize each column of indicators to have a unit cross-sectional second moment for each month in the sample period. Panel B is for a set of 49 portfolios whose vector of excess returns for period t is given by $\boldsymbol{r}_{p,t} = (\boldsymbol{X}_{t-1}\boldsymbol{R}_{\perp})^{+}\boldsymbol{r}_t$. The regime-specific parameters are estimated by fitting a regime-switching model to $\{\tilde{\boldsymbol{r}}_{p,1}, \tilde{\boldsymbol{r}}_{p,2}, \ldots, \tilde{\boldsymbol{r}}_{p,T}\}$. Under the model, $\tilde{\boldsymbol{r}}_{p,t}|s_t \sim N(\boldsymbol{\kappa}_{s_t}, \boldsymbol{\Lambda}_{s_t}\boldsymbol{\Lambda}'_{s_t} + \tau^2_{s_t}\mathbf{I})$, where $s_t \in (1,2)$ is discrete state variable that follows a first-order Markov chain. The first two columns report the average estimated fraction of the variation of the excess portfolio returns that is explained by the latent factors (Avg ret R^2). The next six columns report the estimated RMSEs for the first and second regimes along with the 1% critical values (CVs) for testing the hypothesis of exact factor pricing. The final six columns report 95% confidence intervals for the R^2 statistics obtained by regressing the estimated vector of excess portfolio returns on the estimated matrix of factor loadings (the second-pass regressions). I use a parametric bootstrap procedure to compute the critical values and confidence intervals. The no-micro-caps sample excludes stocks that are smaller than the 20th percentile of the NYSE market-equity distribution. The sample period is May 1967 to December 2018.

Table 7
Estimated Regime-Specific Correlations between Characteristics and Subsequent Excess Stock Returns

A. Percentiles of estimated firm-level correlations

	Low volatility regime					High volatility regime				
Characteristic	5th	10th	50th	90th	95th	5th	10th	50th	90th	95th
LME	−0.21	−0.17	−0.05	0.07	0.11	−0.35	−0.29	−0.08	0.14	0.21
LBM	−0.14	−0.09	0.05	0.18	0.23	−0.21	−0.14	0.08	0.28	0.35
MOM	−0.16	−0.12	0.01	0.14	0.19	−0.29	−0.22	−0.02	0.18	0.26
IMOM	−0.15	−0.11	0.02	0.15	0.19	−0.26	−0.19	−0.00	0.19	0.27
REV	−0.23	−0.18	−0.04	0.10	0.15	−0.36	−0.29	−0.07	0.14	0.21
LVOL	−0.19	−0.14	−0.01	0.11	0.16	−0.25	−0.18	0.02	0.24	0.31
LTO	−0.18	−0.14	−0.01	0.11	0.16	−0.25	−0.18	0.02	0.20	0.27
SUM	−0.14	−0.10	0.02	0.15	0.19	−0.24	−0.16	0.04	0.24	0.30
INV	−0.20	−0.16	−0.02	0.11	0.16	−0.29	−0.22	−0.03	0.17	0.25
SM	−0.15	−0.10	0.03	0.19	0.24	−0.27	−0.20	−0.00	0.26	0.35
NOA	−0.19	−0.15	−0.02	0.10	0.14	−0.28	−0.21	−0.02	0.18	0.25
P52H	−0.18	−0.13	0.01	0.17	0.23	−0.24	−0.17	0.04	0.31	0.40
APM	−0.14	−0.10	0.01	0.12	0.15	−0.24	−0.17	0.05	0.21	0.27
ROC	−0.19	−0.15	−0.02	0.12	0.17	−0.27	−0.20	−0.00	0.20	0.29
NSFY	−0.19	−0.14	−0.02	0.12	0.16	−0.27	−0.21	−0.04	0.15	0.23
NSTY	−0.18	−0.14	−0.02	0.11	0.15	−0.28	−0.22	−0.04	0.15	0.22

B. Percentiles of R-squared for firm-level regressions

	0.24	0.27	0.44	0.66	0.73	0.36	0.41	0.63	0.91	0.96

Panel A of the table reports selected percentiles of the cross-sectional distribution of the estimated regime-specific correlations between excess individual stock returns and 16 predetermined firm characteristics. To estimate the correlation between the excess returns of the nth stock and the jth characteristic in regime $i \in (1,2)$, I compute the sample correlation between the weighted excess returns and the weighted characteristic values using the full time series of available excess returns for the stock, where the weight for month t is the square root of the estimated smoothed probability of regime i for month t. I estimate the smoothed probabilities by fitting a two-sate Markov-switching model to the excess returns for a set of 17 regression-based managed portfolios. The vector of excess portfolio returns for month t is given by $\tilde{\boldsymbol{r}}_{p,t} = \bar{\boldsymbol{W}}^{1/2}\boldsymbol{C}^{+}_{t-1}\boldsymbol{r}_t$, where $\boldsymbol{r}_t$ is the $N_t \times 1$ vector of excess stock returns for month t, $\boldsymbol{C}^{+}_{t-1}$ is the left pseudoinverse of an $N_t \times 17$ matrix $\boldsymbol{C}_{t-1}$ whose nth row contains a 1 along with the values of 16 predetermined characteristics for the nth firm (see Table 1), and $\bar{\boldsymbol{W}} = (\sum_{t=1}^{T} N_t)^{-1} \sum_{t=1}^{T} \boldsymbol{C}'_{t-1}\boldsymbol{C}_{t-1}$. I standardize each column of $\boldsymbol{C}_{t-1}$ except the first to have a mean of zero and variance of one for every month. Under the Markov switching model, $\tilde{\boldsymbol{r}}_{p,t}|s_t \sim N(\boldsymbol{\kappa}_{s_t}, \boldsymbol{\Lambda}_{s_t}\boldsymbol{\Lambda}'_{s_t} + \tau^2_{s_t}\mathbf{I})$, where $s_t \in (1,2)$ is a discrete state variable that follows a first-order Markov chain. Panel B reports selected percentiles of the cross-sectional distribution of the regime-specific R-squared statistics obtained by fitting time-series regressions of weighted excess individual stock returns on the weighted values of the 16 characteristics using the full time series of available excess returns for each stock. The estimated correlations and R-squared statistics for stocks that have less than 60 monthly observations are excluded from the analysis. The sample period is May 1967 to December 2018.

Table 8
Predictive Power of Recursively-Updated Regime Forecasts for Pure Play Portfolio Returns

$$r_{pj,t} = \mu_1 + \Delta_{12}\xi_{2,t|t-1}(\hat{\boldsymbol{\theta}}_{t-h}) + u_t,\ t = h+1,\ldots,T$$

	All stocks					No micro-caps				
	Estimates & t-statistics					Estimates & t-statistics				
Portfolio	$\hat{\mu}_1$	$t(\hat{\mu}_1)$	$\hat{\Delta}_{12}$	$t(\hat{\Delta}_{12})$	R^2	$\hat{\mu}_1$	$t(\hat{\mu}_1)$	$\hat{\Delta}_{12}$	$t(\hat{\Delta}_{12})$	R^2
EWM	0.47	1.77	1.39	2.39	0.014	0.60	2.33	0.68	1.21	0.003
LME	−0.09	−1.49	−0.48	−2.94	0.025	−0.03	−0.70	−0.43	−3.57	0.037
LBM	0.21	4.89	−0.15	−1.25	0.005	0.03	0.75	0.00	0.03	0.000
MOM	0.44	7.40	−0.02	−0.12	0.000	0.25	4.00	−0.20	−0.94	0.003
IMOM	0.03	0.67	−0.02	−0.16	0.000	0.04	0.85	0.15	1.11	0.003
REV	−0.49	−11.47	−0.47	−3.53	0.038	−0.24	−6.04	−0.31	−2.16	0.017
LVOL	−0.24	−4.62	−0.04	−0.29	0.000	−0.18	−3.71	0.03	0.21	0.000
LTO	−0.01	−0.24	0.14	1.00	0.003	−0.05	−1.25	0.22	2.10	0.013
SUV	0.42	14.06	0.26	4.22	0.039	0.19	8.14	0.15	2.82	0.019
INV	−0.11	−2.93	−0.27	−2.92	0.023	−0.02	−0.54	−0.22	−2.36	0.017
SM	0.05	1.55	0.15	1.61	0.008	0.01	0.21	0.12	1.49	0.006
NOA	−0.19	−6.02	−0.01	−0.13	0.000	−0.14	−4.07	0.01	0.13	0.000
P52H	0.25	3.45	0.97	3.93	0.049	−0.11	−2.22	0.38	2.18	0.017
APM	0.08	2.94	−0.14	−2.34	0.014	0.08	2.95	−0.07	−1.00	0.003
ROC	0.04	2.63	−0.04	−0.97	0.003	0.02	0.89	0.00	0.03	0.000
NSFY	−0.09	−3.91	0.06	1.16	0.003	0.01	0.58	−0.10	−1.89	0.010
NSTY	−0.16	−7.79	−0.05	−1.19	0.003	−0.07	−4.08	−0.07	−1.51	0.007

The table reports the results of predictive regressions for a set of 17 regression-based managed portfolios (linear pure play portfolios) that are constructed using 17 predetermined regressors (an intercept and the 16 firm characteristics in Table 1). The dependent variable for month t is the excess percentage portfolio return for the month ($r_{pj,t}$ for $j \in (1,\ldots,17)$). The explanatory variables for month t are 1 and $\xi_{2,t|t-1}(\hat{\boldsymbol{\theta}}_{t-h})$, an estimate of the conditional probability that the Markov-switching process for the portfolio returns is in the high-volatility regime in month t. The conditional probability for month t is obtained by estimating $\boldsymbol{\theta}$, the parameter of the Markov switching model, using the data observed through month $t-1$. Hence, the regressors for the predictive regressions are in the month $t-1$ information set of investors. The recursively-updated estimates of $\boldsymbol{\theta}$ are denoted by $\{\hat{\boldsymbol{\theta}}_i\}_{i=1}^{T-h}$, where $h = 120$ is the number of observations in the initial holdout sample. The vector of portfolio returns for month t is given by $\boldsymbol{r}_{p,t} = \boldsymbol{C}_{t-1}^{+}\boldsymbol{r}_t$, where $\boldsymbol{r}_t$ is the $N_t \times 1$ vector of excess stock returns for month t and $\boldsymbol{C}_{t-1}^{+}$ is the left pseudoinverse of an $N_t \times 17$ matrix $\boldsymbol{C}_{t-1}$ whose jth row contains a 1 along with the values of 16 other characteristics for the jth firm. I standardize each column of $\boldsymbol{C}_{t-1}$ except the first to have a mean of zero and variance of one for every month. The no-micro-caps sample excludes stocks that are smaller than the 20th percentile of the NYSE market-equity distribution. The sample period is May 1967 to December 2018.

Table 9
Portfolios Formed Using One-Month-Ahead Forecasts of Individual Stock Returns

	Residual from a regression of regime-switching on single-regime forecasts						Residual from a regression of single-regime on regime-switching forecasts					
Regressor	L	2	3	4	H	H−L	L	2	3	4	H	H−L
Panel A: All stocks using equal weights												
Low (L)	−1.00	0.03	0.21	0.23	−0.01	0.98	−0.11	0.18	0.15	−0.14	−1.07	−0.97
	(−3.61)	(0.13)	(0.92)	(0.95)	(−0.05)	(3.37)	(−0.39)	(0.79)	(0.67)	(−0.55)	(−3.81)	(−3.42)
2	−0.39	0.39	0.70	0.75	0.61	0.99	0.43	0.62	0.68	0.38	−0.16	−0.59
	(−1.41)	(1.61)	(3.07)	(3.06)	(2.04)	(3.40)	(1.51)	(2.62)	(3.00)	(1.55)	(−0.58)	(−2.06)
3	−0.04	0.75	1.04	1.12	1.11	1.15	0.75	1.03	1.07	0.85	0.31	−0.44
	(−0.13)	(3.00)	(4.45)	(4.45)	(3.52)	(3.74)	(2.53)	(4.02)	(4.58)	(3.45)	(1.12)	(−1.51)
4	0.41	1.05	1.46	1.46	1.52	1.11	1.02	1.45	1.42	1.25	0.81	−0.21
	(1.42)	(4.07)	(5.75)	(5.50)	(4.52)	(3.38)	(3.17)	(5.44)	(5.59)	(4.82)	(2.73)	(−0.67)
High (H)	1.68	1.80	2.06	2.17	3.14	1.46	2.73	2.24	2.11	2.04	2.23	−0.50
	(4.98)	(5.85)	(6.93)	(6.81)	(6.65)	(3.26)	(5.81)	(6.41)	(6.65)	(6.68)	(6.33)	(−1.23)
H−L	2.68	1.77	1.85	1.94	3.15		2.84	2.06	1.96	2.17	3.30	
	(12.69)	(9.03)	(9.46)	(9.60)	(9.40)		(8.19)	(8.24)	(8.99)	(10.68)	(14.20)	
Panel B: No micro-caps using value weights												
Low (L)	−0.24	−0.05	0.39	0.58	0.52	0.76	0.36	0.27	0.28	0.01	−0.43	−0.78
	(−0.80)	(−0.20)	(1.43)	(2.17)	(1.71)	(2.64)	(1.27)	(1.03)	(1.13)	(0.03)	(−1.40)	(−2.80)
2	−0.06	0.71	0.69	0.96	0.84	0.90	0.81	0.86	0.69	0.79	0.08	−0.72
	(−0.21)	(2.98)	(2.89)	(3.76)	(2.85)	(2.97)	(2.94)	(3.52)	(2.96)	(3.27)	(0.30)	(−2.50)
3	0.14	0.77	0.88	0.90	1.11	0.98	0.93	1.00	0.86	0.81	0.43	−0.49
	(0.49)	(3.23)	(3.77)	(3.78)	(3.64)	(3.05)	(3.27)	(4.18)	(3.72)	(3.42)	(1.51)	(−1.64)
4	0.55	0.89	1.17	1.11	1.22	0.67	1.05	1.08	1.17	1.03	0.65	−0.40
	(2.00)	(3.90)	(5.11)	(4.74)	(4.02)	(2.18)	(3.33)	(4.42)	(5.08)	(4.48)	(2.37)	(−1.33)
High (H)	0.74	1.16	1.31	1.44	1.32	0.58	1.23	1.31	1.43	1.41	1.17	−0.06
	(2.54)	(4.65)	(5.38)	(5.75)	(4.43)	(2.05)	(3.56)	(4.87)	(5.61)	(5.36)	(4.03)	(−0.20)
H−L	0.98	1.22	0.93	0.86	0.80		0.88	1.04	1.15	1.40	1.60	
	(5.14)	(6.44)	(4.64)	(4.11)	(3.53)		(3.96)	(4.74)	(5.78)	(6.47)	(7.70)	

Each panel of the table reports average excess returns for two sets of 25 portfolios. The first set of portfolios is obtained by finding the intersection of independent sorts that group stocks into quintiles using (i) single-regime forecasts of excess returns and (ii) residuals from a cross-sectional regression of regime-switching forecasts on the single regime forecasts (i.e., the component of the regime-switching forecasts that is cross-sectionally uncorrelated with the single-regime forecasts). The second set of portfolios is obtained by finding the intersection of independent sorts that group stocks into quintiles using (i) regime-switching forecasts of excess returns and (ii) residuals from a cross-sectional regression of single-regime forecasts on the regime-switching forecasts. The portfolio returns for month $t+1$ are for portfolios that are formed in month t. I compute the single-regime forecasts for month $t+1$ as $\hat{\boldsymbol{r}}_{t+1|t} = \boldsymbol{C}_t\hat{\boldsymbol{\mu}}_{p,t}$, where $\boldsymbol{C}_t$ is an $N_t \times 17$ matrix whose jth row contains a 1 along with the values of the 16 characteristics for the jth firm and $\hat{\boldsymbol{\mu}}_{p,t}$ is the average value of $\boldsymbol{r}_{p,t} = \boldsymbol{C}^{+}_{t-1}\boldsymbol{r}_t$ over the first t months of the sample period. I compute the regime-switching forecast for month $t+1$ as $\hat{\boldsymbol{r}}_{t+1|t} = \boldsymbol{C}_t\hat{\boldsymbol{m}}_{p,t}$, where $\hat{\boldsymbol{m}}_{p,t}$ is the estimate of $\mathrm{E}(\boldsymbol{r}_{p,t+1}|\boldsymbol{r}_{p,1}, \ldots, \boldsymbol{r}_{p,t}; \boldsymbol{\theta})$ obtained by fitting a two-state Markov switching model to the collection of managed portfolio returns observed through month t. Panel A reports results for equally-weighted portfolios that are formed using all stocks. Panel B reports results for value-weighted portfolios that are formed using a sample that excludes stocks whose market equity is smaller than the 20th percentile of the NYSE market-equity distribution. The sample period is May 1967 to December 2018.

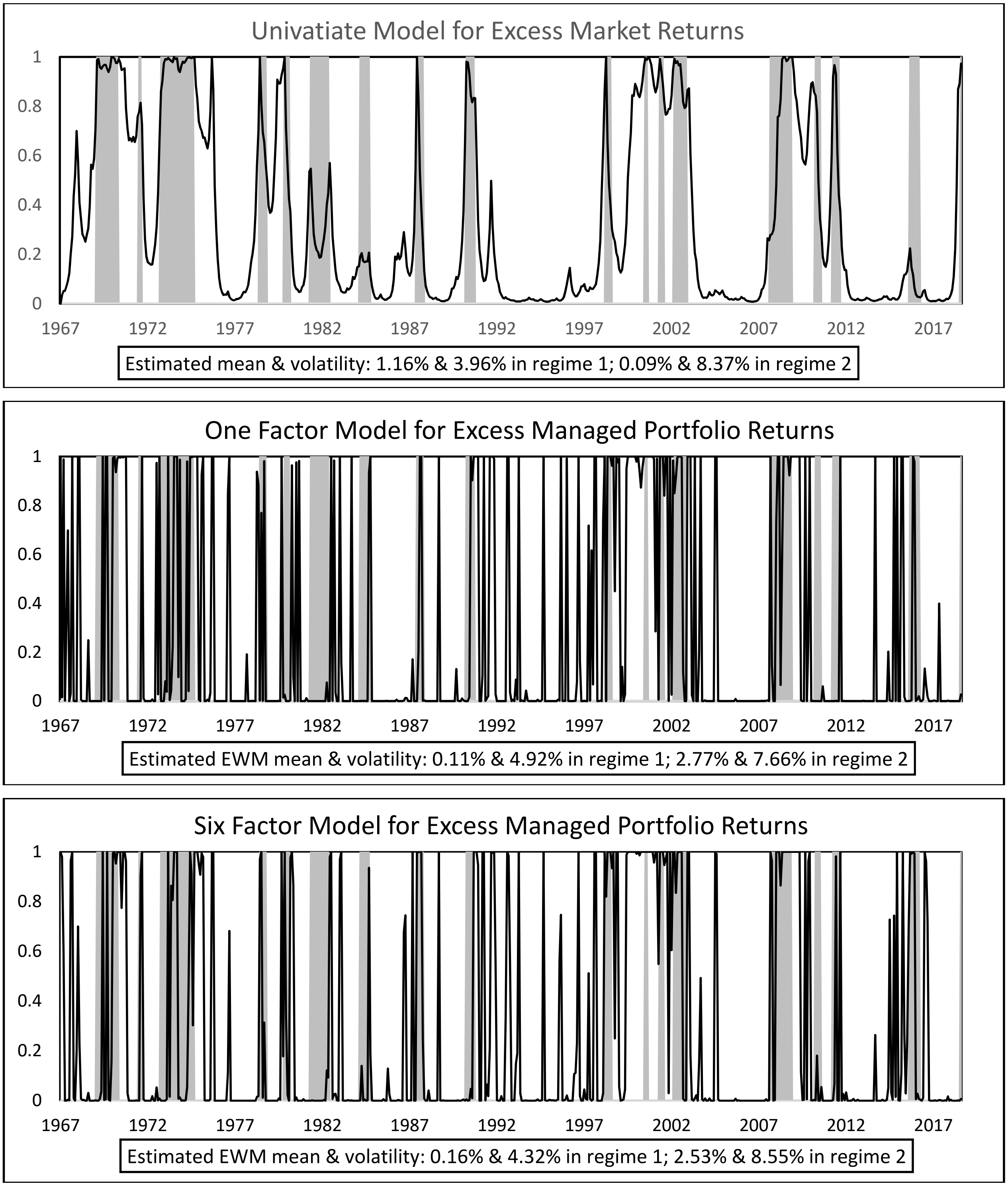

Figure 1. Fitted Time Series of Smoothed Probabilities of the High-Volatility Regime

The plots show estimated smoothed probabilities of the high-volatility regime produced by a two-state Markov switching model. The top plot is obtained by fitting a univariate version of the model to monthly excess returns on an equally-weighted market index. The remaining two plots are obtained by fitting the model to monthly excess returns for a set of 17 managed portfolios and imposing either a one- or six-factor structure on the covariance matrix for each regime. The vector of excess portfolio returns for period t is given by $\tilde{\boldsymbol{r}}_{p,t} = \bar{\boldsymbol{W}}^{1/2}\boldsymbol{C}^{+}_{t-1}\boldsymbol{r}_t$, where $\boldsymbol{r}_t$ is the $N_t \times 1$ vector of excess stock returns for month t, $\boldsymbol{C}^{+}_{t-1}$ is the left pseudoinverse of an $N_t \times 17$ matrix $\boldsymbol{C}_{t-1}$ whose nth row contains a 1 along with the values of 16 predetermined characteristics for the nth firm, and $\bar{\boldsymbol{W}} = (\sum_{t=1}^{T} N_t)^{-1} \sum_{t=1}^{T} \boldsymbol{C}'_{t-1}\boldsymbol{C}_{t-1}$. The shaded areas show ex-post bear market periods under a commonly-used demarcation rule (15% loss ⇒ start of bear market; 20% gain ⇒ start of bull market). The sample period is May 1967 to December 2018.

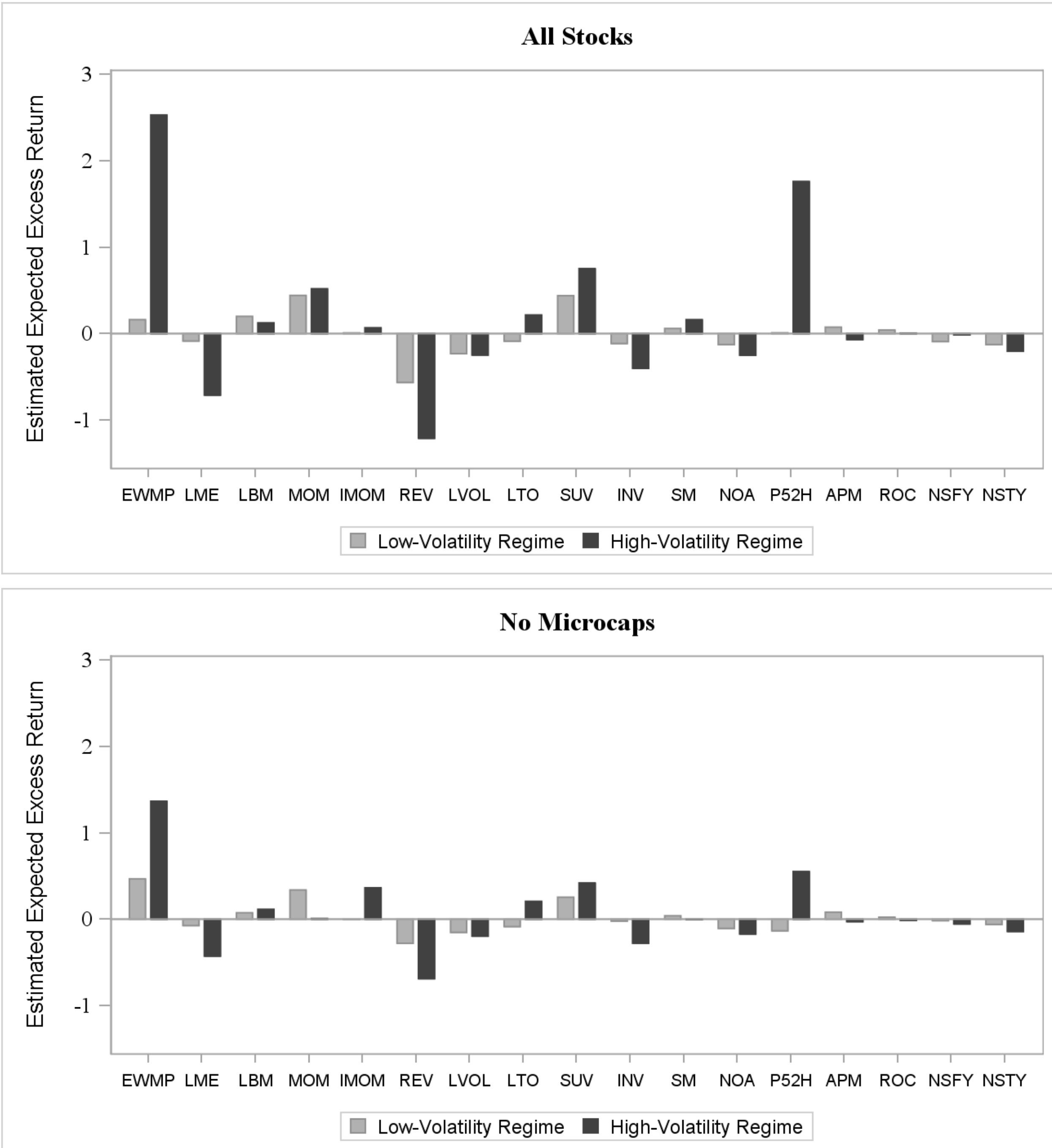

Figure 2. Estimated Expected Excess Portfolio Returns Under the Six-Factor Model
The plot shows regime-specific estimates of the expected excess returns for a set of 17 linear pure play portfolios. The estimates for a given portfolio are plotted side by side to highlight the differences across regimes. The vector of excess portfolio returns for period t is given by $\boldsymbol{r}_{p,t} = \boldsymbol{C}_{t-1}^{+}\boldsymbol{r}_t$, where $\boldsymbol{r}_t$ is the $N_t \times 1$ vector of excess stock returns for month t and $\boldsymbol{C}_{t-1}^{+}$ is the left pseudoinverse of an $N_t \times 17$ matrix $\boldsymbol{C}_{t-1}$ whose nth row contains a 1 along with the values of 16 predetermined characteristics for the nth firm. I obtain the regime-specific estimates by fitting a two-state Markov switching model to the monthly excess returns for a related set of 17 regression-based managed portfolios whose vector of excess returns for period t is given by $\tilde{\boldsymbol{r}}_{p,t} = \bar{\boldsymbol{W}}^{1/2}\boldsymbol{C}_{t-1}^{+}\boldsymbol{r}_t$, where $\bar{\boldsymbol{W}} = (\sum_{t=1}^{T} N_t)^{-1} \sum_{t=1}^{T} \boldsymbol{C}_{t-1}'\boldsymbol{C}_{t-1}$. The sample period is May 1967 to December 2018.

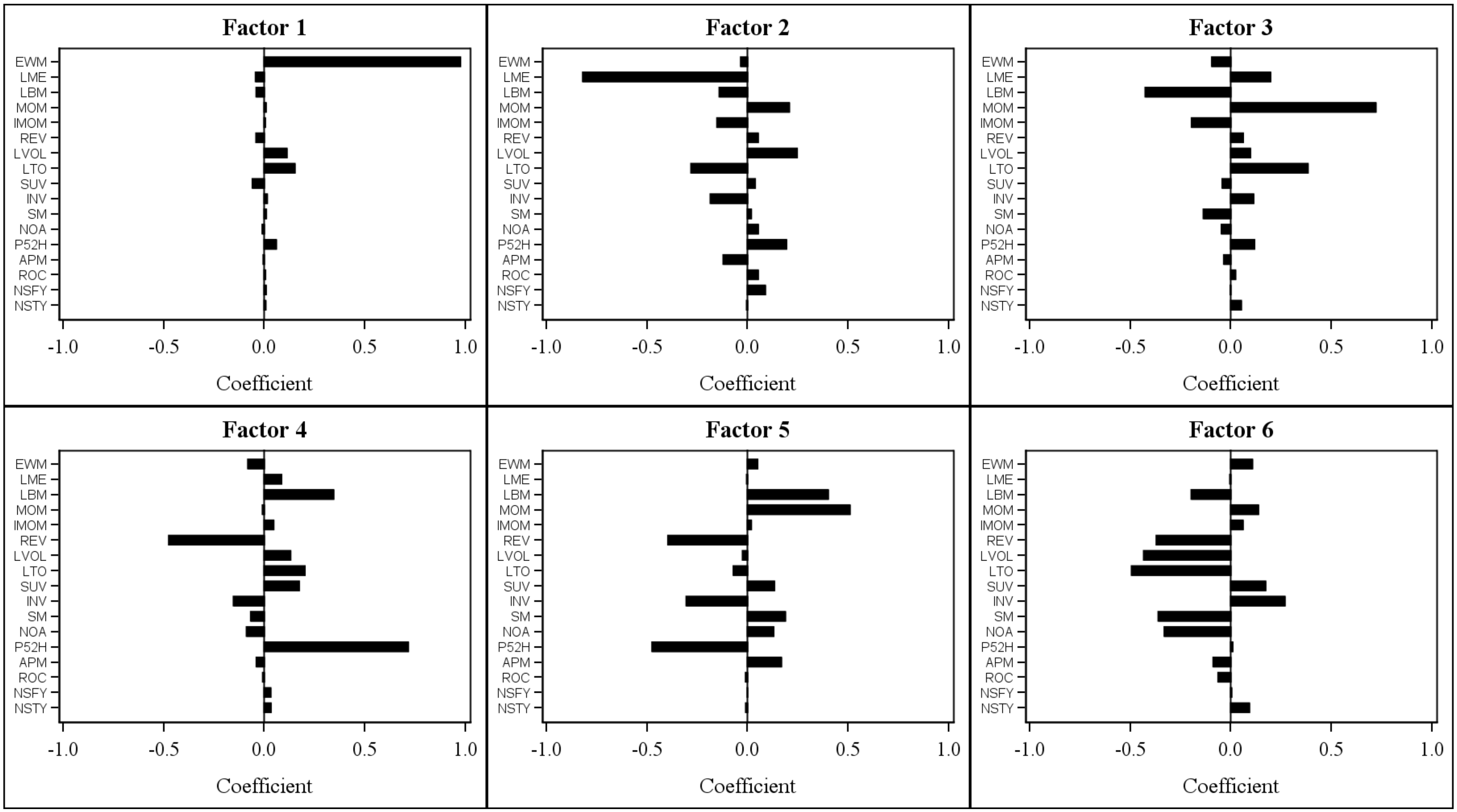

Figure 3. Weights Used to Form the Estimated Regime-Switching IPCA Factors for the Low-Volatility Regime

The plots show the coefficients (weights) applied to the excess returns for 17 linear pure play portfolios to obtain estimates of the latent factors for regime one (the low-volatility regime). The vector of excess portfolio returns for period t is given by $\boldsymbol{r}_{p,t} = \boldsymbol{C}_{t-1}^{+}\boldsymbol{r}_t$, where $\boldsymbol{r}_t$ is the $N_t \times 1$ vector of excess stock returns for month t and $\boldsymbol{C}_{t-1}^{+}$ is the left pseudoinverse of an $N_t \times 17$ matrix $\boldsymbol{C}_{t-1}$ whose nth row contains a 1 along with the values of 16 predetermined characteristics for the nth firm. The estimated vector of factor realizations for month t given that the process is in regime one is given by $\hat{\boldsymbol{\Gamma}}_1^{+}\boldsymbol{r}_{p,t}$, where $\hat{\boldsymbol{\Gamma}}_1^{+} = (\hat{\boldsymbol{\Gamma}}_1'\hat{\boldsymbol{\Gamma}}_1)^{-1}\hat{\boldsymbol{\Gamma}}_1'$ denotes the left pseudoinverse of the matrix of estimated factor loadings for regime one. I construct the estimates under the identification restriction $\hat{\boldsymbol{\Gamma}}_1'\hat{\boldsymbol{\Gamma}}_1 = \mathbf{I}$ so that the jth row of $\hat{\boldsymbol{\Gamma}}_1^{+}$, which contains the weights placed on the various portfolios to compute the realization of the jth factor for regime one, has unit length for all j. The sample period is May 1967 to December 2018.

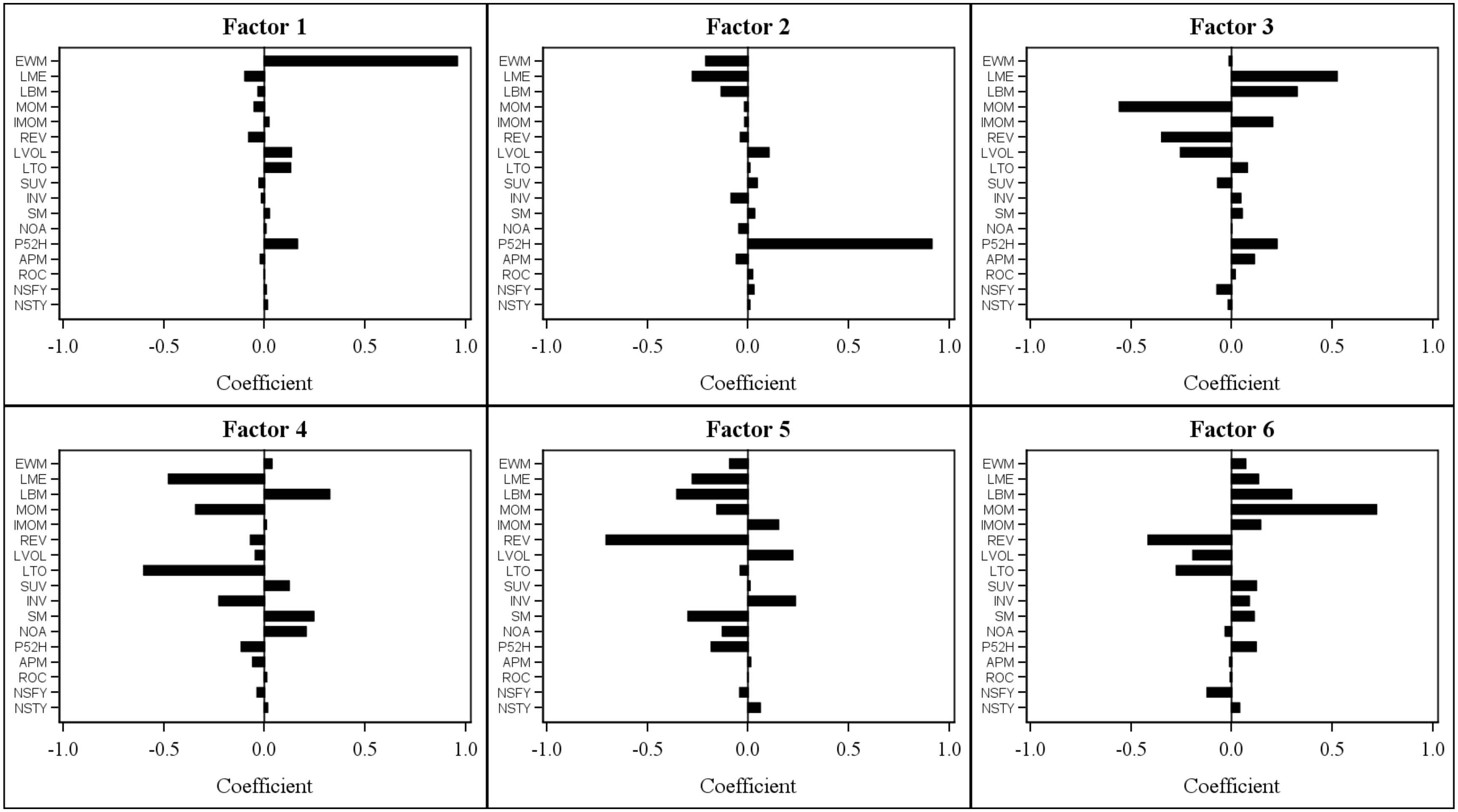

Figure 4. Weights Used to Form the Estimated Regime-Switching IPCA Factors for the High-Volatility Regime
The plots show the coefficients (weights) applied to the excess returns for 17 linear pure play portfolios to obtain estimates of the latent factors for regime two (the high-volatility regime). The vector of excess portfolio returns for period t is given by $\boldsymbol{r}_{p,t} = \boldsymbol{C}_{t-1}^{+}\boldsymbol{r}_t$, where $\boldsymbol{r}_t$ is the $N_t \times 1$ vector of excess stock returns for month t and $\boldsymbol{C}_{t-1}^{+}$ is the left pseudoinverse of an $N_t \times 17$ matrix $\boldsymbol{C}_{t-1}$ whose nth row contains a 1 along with the values of 16 predetermined characteristics for the nth firm. The estimated vector of factor realizations for month t given that the process is in regime two is given by $\hat{\boldsymbol{\Gamma}}_2^{+}\boldsymbol{r}_{p,t}$, where $\hat{\boldsymbol{\Gamma}}_2^{+} = (\hat{\boldsymbol{\Gamma}}_2'\hat{\boldsymbol{\Gamma}}_2)^{-1}\hat{\boldsymbol{\Gamma}}_2'$ denotes the left pseudoinverse of the matrix of estimated factor loadings for regime two. I construct the estimates under the identification restriction $\hat{\boldsymbol{\Gamma}}_2'\hat{\boldsymbol{\Gamma}}_2 = \mathbf{I}$ so that the jth row of $\hat{\boldsymbol{\Gamma}}_2^{+}$, which contains the weights placed on the various portfolios to compute the realization of the jth factor for regime two, has unit length for all j. The sample period is May 1967 to December 2018.

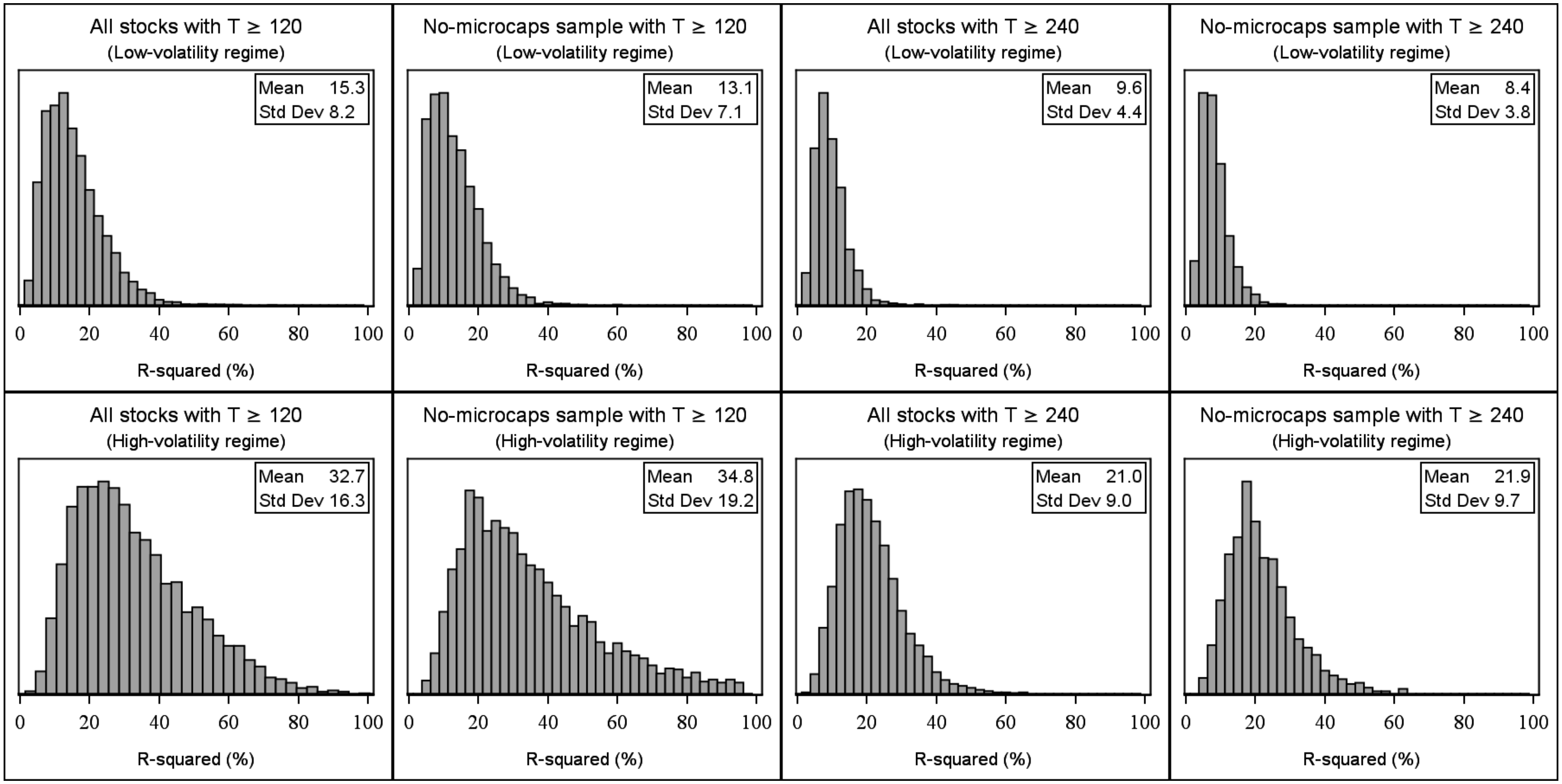

Figure 5. Distributional Properties of the *R*-squared Statistic for Regressions of Excess Stock Returns on Firm Characteristics

The plots show histograms that illustrate the regime-specific distributions of the *R*-squared statistic produced by time-series regressions of excess individual stock returns on the full set of firm characteristics used for the empirical analysis. I require that a stock have at least 60 non-missing excess returns to be included in the plots. The regime-specific *R*-squared statistics are constructed via weighted least squares regressions that employ the square root of the estimated smoothed regime probabilities as weights. The probabilities are estimated by fitting a two-state Markov switching model to the vector of excess returns for a set of 17 regression-based managed portfolios. The vector of excess portfolio returns for period t is given by $\tilde{\boldsymbol{r}}_{p,t} = \bar{\boldsymbol{W}}^{1/2}\boldsymbol{C}^{+}_{t-1}\boldsymbol{r}_t$, where $\boldsymbol{r}_t$ is the $N_t \times 1$ vector of excess stock returns for month t, $\boldsymbol{C}^{+}_{t-1}$ is the left pseudoinverse of an $N_t \times 17$ matrix $\boldsymbol{C}_{t-1}$ whose nth row contains a 1 along with the values of 16 predetermined characteristics for the nth firm, and $\bar{\boldsymbol{W}} = (\sum_{t=1}^{T} N_t)^{-1} \sum_{t=1}^{T} \boldsymbol{C}'_{t-1}\boldsymbol{C}_{t-1}$. The no-micro-caps sample excludes stocks whose market equity is smaller than the 20th percentile of the NYSE market-equity distribution. The sample period is May 1967 to December 2018.