Risk Price Variation: The Missing Half of the Cross-Section of Expected Returns

Andrew J. Patton and Brian M. Weller*
Duke University

September 14, 2018

Abstract

The Law of One Price is a bedrock of asset pricing theory and empirics. Yet real-world frictions can violate the Law by generating unequal compensation across assets for the same risk exposures. We develop new methods for cross-sectional asset pricing with unobserved heterogeneity in compensation for risk. We extend k-means clustering to group assets by risk prices and introduce a formal test for whether differences in risk premia across market segments are too large to occur by chance. Using portfolios of US stocks, international stocks, and assets from multiple classes, we find significant evidence of cross-sectional variation in risk prices for all 135 combinations of test assets, factor models, and time periods. Variation in risk prices is as important as variation in risk exposures for explaining the cross-section of expected returns.

JEL: G12, G14, C21, C55

Keywords: Risk Premia, Market Segmentation, Clustering, Expectation Maximization

*Email: andrew.patton@duke.edu, brian.weller@duke.edu. We thank Tim Bollerslev, Svetlana Bryzgalova, Anna Cieslak, Stefano Giglio, Cam Harvey, Jia Li, Zhengzi (Sophia) Li, Dong Lou, and seminar participants at Duke Fuqua, the International Symposium on Financial Engineering and Risk Management, the INSEAD Finance Symposium, and Tim Bollerslev's 60th Birthday Conference for valuable comments and discussion. We are grateful to Duke Research Computing and especially Tom Milledge for large-scale computing support.

I. Introduction

Academic models of asset prices often assume away complications like trading frictions and segmented markets, but in practice these simplifying assumptions rarely hold. As the global financial crisis made painfully apparent, arbitrage frictions matter and, even in ordinary times, institutional and informational frictions impede investor participation across markets and make for good deals in the markets' dusty corners (Fama and French (2010), Hou, Xue, and Zhang (2017)). Frictions such as these can break the Law of One Price and generate heterogeneous prices of risk across market segments. Consequently, cross-sectional differences in average returns may derive from differences in both risk exposures and *risk premia.*

In an idealized financial market, cross-sectional differences in risk prices are tamed by two forces. First, risk sharing among market participants typically makes the consumption of a "representative" household the key determinant of risk prices for *all* assets. Second, sophisticated arbitrageurs eliminate differences in compensation for risk that arise on account of short-lived demand pressures. The financial crisis broke both intuitions. In response, recent models such as Brunnermeier and Pedersen (2009) and He and Krishnamurthy (2012) place intermediaries at the heart of the pricing kernel—intermediaries' rather than households' constraints and marginal value of wealth determine equilibrium prices of risk. Likewise, Gârleanu and Pedersen (2011) and Gromb and Vayanos (2018), among others, develop theoretical cross-sectional asset pricing implications of arbitrageur borrowing frictions in exogenously segmented markets. We find empirically that differences in risk prices are so pervasive that such limits to arbitrage must take a central role in cross-sectional asset pricing within and across markets, in crises and in normal times.

Existing research on asset pricing with heterogeneous risk premia proceeds by drawing on *a priori* knowledge to conjecture groups of similarly priced assets (e.g. Fama and French (1993), Foerster and Karolyi (1999), and Griffin (2002) among others). However, only in certain cases do we know how to group assets ex ante, and any conjectured market segments may be incorrect or less informative than other dimensions of variation in risk prices. Moreover, given the sensitivity of estimated risk prices to how assets are grouped together, a skeptical empiricist should impose the same data-snooping hurdle on market segments as on expected return factors.

We propose a new approach to cross-sectional asset pricing with multiple prices of risk. Our methodology extends existing clustering algorithms to "let the data speak" in identifying groups of assets with similar risk prices. In so doing we address the empirical challenge of identifying segmented sets of assets in a wide range of economic settings. We then build on recent work on panel data models, e.g., Lin and Ng (2012) and Bonhomme and Manresa (2015), to propose methods for formally testing whether multiple risk prices are needed in a given data set, or whether frictionless frameworks suffice to explain the cross-section of expected returns.

The core of our estimation technique consists of estimating group assignments and cross-sectional slopes via the expectation-maximization (EM) algorithm. In brief, this algorithm cycles between (1) estimating cross-sectional slopes given conjectured group assignments and (2) reallocating portfolios to groups given estimated cross-sectional slopes. In the first step, our approach generalizes standard cross-sectional techniques such as Fama and MacBeth (1973) regression to settings with multiple risk prices, whereby we estimate cross-sectional slopes period-by-period and group-by-group. The second step then reallocates each portfolio to the group whose cross-sectional slopes best describe its return dynamics. Iterating these steps identifies progressively more important dimensions of heterogeneity in risk prices, and at convergence, the algorithm delivers an optimal set of group assignments and cross-sectional slopes.[1]

The structure of our estimation problem prevents the use of popular, off-the-shelf clustering technologies like k-means. In typical clustering applications, the algorithm groups assets using characteristics that are not a function of other assets or of the group assignment and hence can be taken as fixed. In our setting, the key grouping attributes are cross-sectional slopes, i.e., our estimates of risk premia for each factor and date. It is the fact that these grouping attributes are not fixed for each asset and instead depend on the risk and return characteristics of other cluster members that makes k-means inappropriate in our setting. Our new approach can be interpreted as a generalization of k-means that accommodates this dependence.

To test whether a set of portfolios has segmented risk prices, we need to evaluate whether the incremental explanatory power of multiple clusters warrants the additional parameters we estimate in a multiple-cluster model. A naive test of whether risk prices differ across clusters fails because cluster assignments are *estimated* in our approach, and standard tests for the significance of risk-price differences across clusters severely over-reject in finite samples.[2] We propose a simple alternative to overcome this problem based on subsamples of the data: we estimate the cluster assignments using one subsample, and we test for differences of estimated risk prices in a second subsample using the group assignments from the first subsample. We then construct modified F-statistics for equal risk prices across groups, and we obtain critical values using a permutation method in which we shuffle group assignments to obtain the distribution of the test statistic under the null (see, e.g., Lehmann and Romano (2005)).

The economics of our test differ in important ways from standard tests of asset pricing models such as the GRS test of Gibbons, Ross, and Shanken (1989). The GRS test evaluates whether adding test assets improves Sharpe ratios enough to justify the additional flexibility granted in forming the in-sample mean-variance efficient portfolio. Our test evaluates whether adding *clusters* improves model fit enough to justify the additional flexibility allowed by group-specific slopes to

[1] We pair this iterative approach with multi-start and genetic algorithm global search methods to locate the global-best group assignments and corresponding cross-sectional slopes.

[2] By contrast, no such problems arise when cluster assignments are known *ex ante*, as in Foerster and Karolyi (1999). In the leading example of testing a null of $G = 1$ cluster (i.e, no heterogeneity in risk prices) against $G = 2$ clusters with estimated assignments, we must consider the behavior of a model that allows for two clusters when the true number of clusters is one. Existing work on inference for k-means clustering and related methods that takes into account estimated assignments (e.g. Bonhomme and Manresa (2015)) assumes that clusters are "well separated," making such theory inapplicable to this hypothesis test.

explain the cross-section of returns (we use subsamples with fixed groups to shut down flexibility in group assignments). Rejections in each test signifies that a candidate factor model is incomplete—in not spanning priced risks in the GRS test and in not capturing cross-sectional variation in risk prices in our test.

We apply these tools to analyze risk price heterogeneity in a variety of economic settings, including domestic stock portfolios, international stocks portfolios, and cross-asset class portfolios.[3] To these portfolio sets we apply leading factor models including the CAPM, the Fama and French (1992) three-factor model; the Carhart (1997) four-factor model, the Fama and French (2015) five-factor model, the He, Kelly, and Manela (2017) intermediary-capital factor model, and the Hou, Xue, and Zhang (2015) q-factor model. Finally we split our data samples into subintervals to evaluate segmentation over time within each market-model pair.

Our analysis delivers two new empirical facts on cross-sectional heterogeneity in risk prices. First, we find segmented risk prices in *every* setting examined. We reject unified risk pricing for all 135 combinations of test assets, benchmark factor models, and time periods, with a modal p-value of 0.000. Second, we find variation in risk prices contributes significantly to explaining the cross-section of expected returns. Among US stock portfolios, where segmentation is least pronounced, adding multiple clusters increases cross-sectional dispersion in average returns from 3% to 167%, with 25th and 75th percentiles increases of 16% and 76%, respectively. This increase in cross-sectional explanatory power is comparable to replacing the Fama and French three-factor model with their more recent five-factor model. Ex post Sharpe ratios of global factor models also increase markedly when including factor variants local to each market segment; for domestic equity portfolios, the 25th and 75th percentiles of changes in maximal annual Sharpe ratios are 0.26 and 0.83. The gains from acknowledging multiple market segments are even larger in our analyses of international equity portfolios and portfolios of assets across multiple asset classes. We conclude that risk-price heterogeneity is a pervasive and economically important feature of real-world financial markets.

While we successfully explain substantial additional variation in the cross section of average returns, our paper fundamentally differs from empirical asset pricing papers that introduce new candidate factors to accomplish the same objective. There, an empiricist selects factors from a large and unobserved model space, and the many degrees of freedom in data mining the factor (and its empirical implementation) carry no penalty.[4] Our approach penalizes such fishing expeditions. Rather, once we have selected a conventional factor model, the freedom to choose group assignments and fit additional cross-sectional slopes is explicitly accounted for in our statistical tests of unified versus segmented pricing.

[3]We consider a wide range of portfolio sets in the spirit of Lewellen, Nagel, and Shanken (2010), who emphasize the importance of considering a diverse set of test assets when evaluating asset pricing models.

[4]Harvey (2017) caricatures this search among candidate factors in the motivating example of the 2017 AFA Presidential Address.

One drawback to our approach is that, like principal components analysis and other clustering techniques, it does not assign labels to the segments identified. Nevertheless, labels can often be intuited from estimated segments, as we find in the three representative settings we consider:

1. Domestic equity portfolios cluster by market capitalization into small-cap and large-cap stock portfolios. We contribute to the longstanding debate on whether factors earn differential compensation across stocks of different sizes—e.g., Hong, Lim, and Stein (2000), Grinblatt and Moskowitz (2004), and Israel and Moskowitz (2013)—by identifying market capitalization as the single most important determinant of differences in factor premia among common US stock portfolios.

2. International stock portfolios cluster perfectly by geographic region, but with a twist. Rather than assigning stock portfolios to the four regional groups in our sample, we find only three, with developed North American and Pacific Rim markets integrated in a single cluster. This analysis validates our methodology by recovering international boundaries as a known source of market segmentation (see, for example, Griffin (2002), Hou, Karolyi, and Kho (2011), Fama and French (2012)).

3. Stock, commodity, bond, options, and currency portfolios cleave into (at least) five clusters, each corresponding to one or two asset classes with similar risk prices. Here we contrast with He, Kelly, and Manela (2017) by rejecting the null of equal pricing across asset classes in a model with market and intermediary capital risk factors.

Another challenge we face is distinguishing between omitted clusters and omitted factors. In much of our analysis, we take the factor models as given and complete, and we conduct inference on whether the data support one vs. multiple prices of risk. However, as we show in Section VI.A, omitted factors can manifest as clusters, and vice-versa. Omitting factors creates an omitted variable bias, and groups of assets with greater exposures on omitted factors may appear to have different risk prices from assets with smaller exposures. As a case in point, prior-return sorted portfolios differ from size-value sorted portfolios in earning negative compensation to value-factor exposure in a Fama-French three factor model, whereas they do not in a momentum-augmented Carhart model. Similarly, omitting clusters generates a proliferation of apparent factors that capture the interaction of included factors and group membership dummies; the absence of momentum compensation in Japan can masquerade as a new, Japan-specific factor on which only Japanese stocks load, for example.

We use the model comparison test of Rivers and Vuong (2002) to distinguish between omitted factors and fundamental heterogeneity in risk prices. We compare the "best" cluster-based model, as selected by the Akaike information criterion (AIC), with a "best" extended-factor model containing additional factors extracted via principal components from the residuals of the factor model. The subsample approach used to formally test for multiple clusters serves double duty in allowing us to select comparable models on one partition and evaluate their relative performance on the other partition. Accounting for potential omitted factors leaves our main result intact: we find strong evidence of risk price heterogeneity in almost every setting considered. New dimensions of return variation call for new *clusters* rather than new *factors*.

In this vein, our work suggests a potential reason for the continual discovery of new factors in expected returns. The "factor zoo" described by Cochrane (2011) is too crowded to be plausible, standing at several hundred inhabitants according to Harvey, Liu, and Zhu (2016) and Hou, Xue, and Zhang (2017)'s recent counts. Empirically, we find clusters even in the presence of potential omitted factors, but the reverse may not be true. In general, any factor betas or characteristics that are cross-sectionally correlated with omitted group membership dummies may appear to be new priced factors. Moreover, given the discreteness of clusters in segmented markets, continuous characteristics are unlikely to span group indicators, and several new "factors" may spuriously appear from just a few dimensions of latent market segmentation.

II. Related Literature

Market Segmentation

Prior empirical work on segmentation in asset pricing falls into two broad categories. In the first category, researchers consider whether anomalies are concentrated in particular segments of the US stock market. These studies typically focus on transaction costs and heterogeneous market clientele rather than geography as a source of market frictions. In the second category, researchers evaluate integration across international markets. For both strands of research, two key implications of market integration are (1) equality of risk prices across markets, that is, all assets should earn the same compensation per unit risk, and (2) comparable time series pricing errors for local and cross-segment factor models. Intuitively, perfect global risk sharing makes the addition of local risk factors unnecessary.

To the best of our knowledge, no study systematically analyzes dimensions of segmentation within US stocks. However, existing work proposes several independent dimensions of segmentation among US-traded securities. In a seminal paper, Merton (1987) considers informational frictions in which only some traders are aware of an investment opportunity. The absence of knowledge effectively segments markets. Traders earn abnormal returns on their private information, but at the same time, their concentrated holdings of relatively unknown securities adds to their portfolio risk. Empirically, Kadlec and McConnell (1994) and Foerster and Karolyi (1999) confirm Merton's "investor recognition hypothesis" in illiquid US stocks and in international stocks listed on US exchanges, respectively.

Investor awareness, institutional ownership, and trading frictions vary with market capitalization, and several studies consider whether stocks earn the same risk premia in large- and small-cap segments. Hong, Lim, and Stein (2000) and Grinblatt and Moskowitz (2004) estimate a negative interaction between momentum returns and market capitalization. Israel and Moskowitz (2013) find this conclusion to be an artifact of the particular sample period, and they suggest instead that momentum strategies earn compensation in large and small stocks alike.

Fama and French (1993) consider common factors in stocks and bonds, and they find mixed support for integration of these markets: stock returns load on term-structure factors when other stock factors are included, but bond returns (mostly) do not load on stock factors when other bond factors are included. By contrast, He, Kelly, and Manela (2017) find that an intermediary-based asset pricing model explains the returns on stocks, bonds, options, currencies, commodities, and other asset classes, and with similar prices of risk. We note that with rare exceptions, conclusions on the degree of segmentation within and across markets depend critically on the choice of asset pricing model.

Turning to the international context, evidence is mixed on whether global factor models suffice for pricing the cross-section of international portfolios. Fama and French (1998) argue for a parsimonious, global market and value-factor model to explain value in international markets. Griffin (2002) rejects both local and global factor models using GRS tests of their in-sample mean-variance efficiency (Gibbons, Ross, and Shanken (1989)), and smaller time series alphas from local factor models favor a segmented-markets worldview. Hou, Karolyi, and Kho (2011) support Griffin (2002)'s finding using an expanded cross-section of individual stocks in 49 countries. Less ambiguously, the momentum factor performs well in most countries with the notable exception of Japan (Rouwenhorst (1998), Griffin, Ji, and Martin (2003), and Fama and French (2012)). International variation in the pricing of momentum exposure contradicts global factor models, as does the particularly large Japanese value premium (Asness, Moskowitz, and Pedersen (2013)).

The international finance literature often measures market segmentation by correlation and cointegration rather than by mean-variance spanning or equality of risk prices. For exceptions in which heterogeneous risk prices feature prominently, Errunza and Losq (1985) propose a model of "mildly" segmented markets in which asymmetric barriers to trade break the equality of risk prices across countries. Bekaert and Harvey (1995) estimate a regime-switching model to consider time-variation in the degree of cross-market integration. Regimes are characterized by homogeneous or heterogeneous risk prices depending on the degree of segmentation at a particular date.

Representative of much of the international market segmentation literature, Bekaert, Hodrick, and Zhang (2009) evaluate a large cross-section of country and industry portfolios and find models with both global and regional factors fit the data. The authors emphasize common *factor comovements* rather than common *factor prices*. Pukthuanthong and Roll (2009) critique cross-country correlation measures and focus instead on integration as a high R^2 from global factor models. Our methodology draws on both interpretations of market integration; we recognize groups of assets as stochastically segmented if cross-sectional R^2s are low for cross-group factor models relative to within-group factor models.

Financial Frictions and Asset Prices

Motivated by the global financial crisis, a number of recent studies consider equilibrium asset prices in which frictions play a central role. Gârleanu and Pedersen (2011) show that margin constraints combined with limited capital prevent potential arbitrageurs from eliminating violations of the Law of One Price, much in the spirit of Shleifer and Vishny (1997)'s limits to arbitrage. Gromb and Vayanos (2018) extend Gromb and Vayanos (2002) to consider arbitrage dynamics across segmented markets when asset payoffs are not identical, and they obtain a similar result of arbitrageurs investing to maximize alpha per unit of collateral. Much like our paper, the authors focus on cross-sectional implications of constrained arbitrage, and like us, they assume that market segmentation occurs for exogenous reasons, including "regulation, agency problems, or a lack of specialized knowledge." These rationales apply in all of the settings we consider, in addition to other cross-region frictions considered by the international finance literature and applicable primarily in our analysis of international equities.

Cross-sectional variation in risk prices requires both a segmentation mechanism to break the Law of One Price and heterogeneous agents to generate differences in risk premia. New models focusing on inequality and intermediation provide rationales for such differences. Greenwald, Lettau, and Ludvigson (2016) and Lettau, Ludvigson, and Ma (2018) build on a long tradition of limited-participation models (e.g., Mankiw and Zeldes (1991) and Vissing-Jørgensen (2002)) in which workers and shareholders have different ratios of capital income to labor income and do not share risks. In these setups, the representative agent's consumption does not price assets; rather, capital share-adjusted income enters the stochastic discount factor. He and Krishnamurthy (2012, 2013), Adrian, Erkko, and Muir (2014), and He, Kelly, and Manela (2017) replace the representative household's stochastic discount factor with that of marginal financial intermediaries to derive an intermediary-capital factor. Both classes of models generate imperfect correlation between the marginal value of wealth among households.

Related Statistical Techniques

Our approach combines the estimation of risk premia with the assignment of assets to latent market segments. Correspondingly, we build on the work of others in statistics as well as finance. On the statistics front, we extend seminal algorithms in machine learning and recent work on estimation of group effects in panel models. The core of our estimation technique relies on expectation maximization (EM) to jointly recover group assignments and cross-sectional slopes (Dempster, Laird, and Rubin (1977)). EM is a powerful iterative technique for estimating parameters when full likelihood functions are difficult or impossible to maximize. Algorithmically, our iteration between risk-price estimation and group assignments also parallels Lloyd's algorithm in k-means clustering (MacQueen (1967)). As discussed in the introduction, our methodology generalizes k-means to group assets based on common within-group slopes in addition to within-group intercepts or average values.

The econometric approach of this paper is related to recent work by Lin and Ng (2012), Sarafidis and Weber (2015) and particularly to Bonhomme and Manresa (2015), who study a k-means application in a panel setting with both N and T large. Su, Shi, and Phillips (2016) also consider a panel setting with estimated group assignments, though they employ penalized estimation methods (such as the "lasso") rather than a clustering approach.

III. Detecting Heterogeneity in Risk Prices

A. Estimation of a Factor Model with Multiple Clusters

Economic Setting

The economy consists of N assets and K asset pricing factors over T dates. Let f_t be a $K \times 1$ vector of asset pricing factors and r_t be an $N \times 1$ vector of asset excess returns at time t. The factors f_t are known and observable (we relax the former assumption in Section VI). Each asset is a member of one of $G \geq 1$ groups, where G is fixed for now. We use $\gamma_i \in \{1, ..., G\}$ to denote the group membership of asset i and define I_i as a $G \times 1$ vector with a one in the γ_i^{th} element and zeros elsewhere. The asset pricing model satisfies

$$\begin{aligned} r_{it} &= \alpha_t I_i + \beta_i \left(f_t + \Phi_t I_i \right) + \epsilon_{it}, \\ 0 &= E\left[\epsilon_t\right] = cov\left(\epsilon_t, f_s\right) = cov\left(\epsilon_t, \Phi_s\right) = cov\left(\Phi_t, f_s\right) = cov\left(\alpha_t, f_s\right), \forall t, s \end{aligned} \tag{1}$$

We set aside conformability of the zero matrices to streamline exposition. For simplicity we assume that errors in (1) are i.i.d. normal with variance σ_i^2. We can relax this assumption to allow for more general return processes so long as the errors have finite fourth moments.

This model differs from a traditional return process in that different groups may have different cross-sectional slopes at each date t and different risk prices on average. Rather than having a scalar average excess return α_t and a $K \times 1$ vector of risk prices λ_t for the entire set of assets, we instead have a $1 \times G$ vector α_t and a $K \times G$ matrix Λ_t, both multiplied by the $G \times 1$ selection vector I_i. Group-specific elements are distinguished by superscripts as $\alpha_t^{(g)}$ for zero-beta rates and as $\lambda_t^{(g)}$ for risk prices. If there is no segmentation, or if all investors can frictionlessly trade all factors, then $\alpha_t = \alpha_t^* \times \mathbf{1}_G'$ and $\Lambda_t = \lambda_t^* \times \mathbf{1}_G'$, and our model nests the typical one in which factor compensation is the same across all assets. Our model and estimation technique also accommodate cases in which groups differ only in their (unconditional) average risk premia $\bar{\Phi}$ or in which dynamics differ only for a single group, for example, momentum behaves differently in Japan, but similarly elsewhere.

In the spirit of Fama and MacBeth (1973) two-stage regressions, the empiricist starts with time series regressions for each asset i,

$$r_{it} = a_i + \beta_i f_t + \eta_{it}. \tag{2}$$

which delivers estimates of risk exposures and idiosyncratic volatility for each portfolio. We make two assumptions to interpret the parameters of this regression. First, we require that the "group-specific" factors Φ are orthogonal to f in the time series; in economic terms, the group-specific distortions are uncorrelated with the factors themselves. This assumption is innocuous in that if the group-specific components are not orthogonal to f, β_i recovers the sensitivity of returns to variation in the common stochastic component. Second, we assume that the time series dimension of our panel is large relative to the cross-sectional dimension, and so the measurement error in the estimated βs and σ_is is negligible, and we can treat them as known (formally, $T/N \to \infty$ as $T, N \to \infty$). This assumption is particularly benign in our empirical estimation given that we use static betas for well-diversified portfolios,[5] and have much larger time series than cross sections.

After computing risk loadings from the time series, the empiricist needs to cross-sectionally estimate the prices of risk for all factors and groups, Λ. Doing so is complicated because we do not know group assignments ex ante. The empiricist must concurrently estimate N group assignments, γ, or equivalently, the $N \times G$ matrix of group assignment indicators, I. We use expectation maximization to address this challenge.

To see why our setting requires a new solution method beyond traditional OLS or maximum-likelihood estimation, consider the error term for asset i under (1) using the time series betas from (2),

$$\hat{\epsilon}_{it} = r_{it} - \alpha_t I_i - \beta_i \Lambda_t I_i = r_{it} - \alpha_t^{(g)} - \beta_i \lambda_t^{(g)}, \text{ for } \gamma_i = g. \tag{3}$$

Up to a constant, the associated log likelihood is

$$\log L(\alpha, \Lambda, I) = -\frac{1}{2} \sum_{g} \sum_{\gamma_i = g} \sum_{t} \frac{1}{\sigma_i^2} \left(r_{it} - \alpha_t^{(g)} - \beta_i \lambda_t^{(g)} \right)^2, \tag{4}$$

and maximizing (4) entails optimizing over $(K+1) \times T \times G$ αs and λs and N group assignments.

Even holding fixed group assignments, attacking this maximization problem directly is difficult because of the number of parameters involved. Direct maximization of (4) becomes impossible computationally when we optimize over the latent group-assignment parameters γ. The objective function in (4) lacks differentiability with respect to discrete group assignments, which complicates or invalidates standard solution techniques for smooth problems.[6] An exhaustive search of group

[5]We use static betas to avoid complicating our approach with estimation error in the generated regressors. An alternative would be to estimate conditional betas and carry the time series estimation errors into the cross-sectional regressions. However this approach adds considerable computational expense because it requires replacing each cross-sectional OLS regression—of which there may be millions in a typical run—with nonlinear regressions that account for this measurement error.

[6]Replacing discrete group assignments with continuous group assignments is an alternative, potentially more tractable modeling choice. However, we are not aware of asset pricing models that deliver partial group memberships, nor is it clear how to incorporate partial memberships into cross-sectional regressions without applying ad hoc

assignments for a "small" problem with 75 portfolios and two groups entails searching over 10^{18} possible groupings just to solve the integer component of the mixed-integer programming problem.

Expectation Maximization Fortunately iterative conditional approaches such as expectation maximization excel in situations in which conditional maximization problems are straightforward but full-problem optimizations—often involving latent parameters—are difficult. In our application the EM algorithm consists of two steps to recover model parameters (α and Λ) and group assignments (γ). First, given the group assignments, we estimate the model parameters ("maximization"). Second, given the model parameters, we reestimate the group assignments ("expectation"). We then iterate between these steps until convergence.[7,8] Both steps are straightforward maximization problems:

1. **Estimate α and Λ given γ.** The first-order conditions of (4) with respect to $\alpha_t^{(g)}$ and $\lambda_t^{(g)}$ are

$$0 = \sum_{\gamma_i = g} \frac{1}{\sigma_i^2} \begin{bmatrix} 1 \\ \beta_i' \end{bmatrix} \left(r_{it} - \alpha_t^{(g)} - \beta_i \lambda_t^{(g)} \right). \tag{5}$$

   Crucially, conditioning on γ delivers separability across time and across clusters. Equation (5) presents the moment conditions of TG cross-sectional regressions with precision weights $1/\sigma_i^2$,

$$r_{it} = \alpha_t^{(g)} + \sum_k \beta_{ik} \lambda_{kt}^{(g)} + \epsilon_{it}, \ \forall i \text{ s.t. } \gamma_i = g, \ \text{for } g = 1, \ldots, G \text{ and } t = 1, \ldots, T. \tag{6}$$

   In short this step reduces to separate Fama-MacBeth regressions for each group g.

2. **Estimate γ given α and Λ.** Fixing the parameters from the conditional MLE, we can focus on a single asset at a time. The maximization problem reduces to finding the group with the

observation weights.

[7] Dempster, Laird, and Rubin (1977) and Wu (1983) prove that this iteration achieves a local solution to the full maximization problem under general conditions. In our setting, the proof of convergence in a finite number of steps is almost immediate. Each step weakly improves the likelihood function, and a (local) maximum exists because the number of possible group assignments is finite (albeit large). Hence the sequence of likelihoods converges to a maximum by the monotone convergence theorem. The algorithm does not "cycle" because each step must weakly improve upon previous steps. Putting these components together, the number of steps is bounded above by G^N. In our applications, the EM algorithm tends to converge far more rapidly, taking between 5 and 15 iterations for the typical starting value and economic setting.

[8] Appendix A details multi-start and genetic algorithm techniques we use to obtain global optima. In practice we find these techniques to be important. For the three leading examples of Section V.C, only in one case would a local optimizer be likely to arrive at the global best group assignments, and in another case, most locally optimal group assignments differ substantially from the global best.

smallest sum of squared errors,

$$\hat{\gamma}_i = \arg\min_g \sum_t \left(r_{it} - \alpha_t^{(g)} - \beta_i \lambda_t^{(g)} \right)^2 .$$

This group assignment is essentially immediate for each security, as it requires only the comparison of a set of G readily computed constants (idiosyncratic variances are fixed and factor out). Note that group memberships are determined by the entire $T \times K$ matrix $\lambda^{(g)}$ rather than by its time-series average $\bar{\lambda}^{(g)}$. Factor realizations at every date are helpful for identifying assets in a common cluster, and groups may be meaningfully distinct in their cross-sectional dynamics regardless of whether their average risk prices are similar.

This procedure has attractive economic properties. The first step generalizes Fama-MacBeth estimation of risk premia to multiple market segments. The cross-sectional slopes $\lambda_t^{(g)}$ are the same as those in standard, single-cluster Fama-MacBeth regression if risk premia are the same across assets. More generally, they are the factor-mimicking portfolio returns obtained using the assets within each cluster—for example, the best approximation of the global momentum factor using North American stocks for one group and Japanese stocks for another. This step accommodates error structures with more complex cross-sectional or time-series dependence with suitable updates of the moment conditions for α and Λ.

The second step assigns assets to the clusters that minimize their pricing errors. Doing so resembles selecting the cluster for each security based on similarity in risk prices, but it is more economically robust because it weights cross-group differences in λs by asset betas; factors that are unimportant for explaining variation in portfolio returns do not determine group membership. This feature guards against grouping assets based on "junk" factors that explain little variation in the panel of realized returns.

From step 1, we see that idiosyncratic volatility serves as observation weights in cross-sectional regressions. Any well-behaved set of weights delivers a consistent estimate of α and Λ as $N \to \infty$, so potential errors in σ_i are immaterial in large samples for this step. In addition, idiosyncratic errors do not enter the second step. Hence the choice of σ_i has no effect asymptotically for parameter estimation or group assignment. However, the way we estimate σ_i matters in practice in two situations. First, in finite samples, σ_i estimated from the time-series residuals maintains efficiency under the null of a single group, for which this choice of weights is optimal. In the worst case for finite N, the estimates in (5) are inefficient under the alternative model of multiple groups, and this inefficiency increases noise in λ estimates, decreases our ability to achieve dispersion among groups, and decreases the probability of rejecting the null. Second, observation weights affect the estimated maximized likelihood and the information-criterion selected number of groups. For this reason, our tests do not rely on a particular choice of G^*(we instead consider a range of values), and we use information criteria only when examining in detail the results from a specific multi-cluster model.

B. Testing for Multiple Clusters

The question at the heart of our paper is whether there exist latent market segments with differing risk prices. In this section we develop a formal methodology to test the null of equal prices of risk in the cross-section (i.e., a single cluster) against the alternative of varying risk prices (i.e., multiple clusters).

We use subsamples of the data to implement our tests for multiple clusters. We partition our data into $\mathcal{R}$ and $\mathcal{P}$ samples, e.g., the first and second halves of the sample, or the odd- and even-dated observations. Let R and P denote the number of observations in each of these samples. For a fixed number of groups G, we estimate our cluster model on the $\mathcal{R}$ sample. This estimation yields parameters for each group $\hat{\alpha}_R^{(g)}$ and $\hat{\lambda}_R^{(g)}$ as well as group assignments $\hat{\gamma}_R$. In estimating these parameters we make standard large-panel assumptions on each group: to estimate cross-sectional slopes consistently we need R, $N \to \infty$, and we also require the number of assets in each cluster, denoted N_g, $g = 1, 2, ..., G$, to be such that $\min_g N_g \to \infty$, i.e. there are no small clusters. Using the group assignments obtained from the first partition, $\hat{\gamma}_R$, we estimate Fama-MacBeth cross-sectional regressions separately for each group on the $\mathcal{P}$ sample, obtaining the parameters $\hat{\alpha}_P^{(g)}$ and $\hat{\lambda}_P^{(g)}$, which we use for testing.

Our tests for multiple clusters take two forms. Our first test evaluates equality of average risk prices across groups for a total of $(G-1)\,K$ restrictions:

$$\begin{aligned} H_0 :&\ \ \bar{\lambda}_k^{(1)} = \bar{\lambda}_k^{(2)} = \ldots = \bar{\lambda}_k^{(G)}\ \forall k \\ vs.\ \ H_1 :&\ \ \bar{\lambda}_k^{(g)} \neq \bar{\lambda}_k^{(g')} \text{ for some } k, g, g'. \end{aligned} \tag{7}$$

where $\bar{\lambda}_k^{(g)} \equiv E[\bar{\lambda}_{kt}^{(g)}]$. Our second test evaluates equality of cross-sectional slopes across groups at each point in time for a total of $(G-1)\,KP$ restrictions:

$$\begin{aligned} H_0 :&\ \ \lambda_{kt}^{(1)} = \lambda_{kt}^{(2)} = \ldots = \lambda_{kt}^{(G)}\ \forall k, t \\ vs.\ \ H_1 :&\ \ \lambda_{kt}^{(g)} \neq \lambda_{kt}^{(g')} \text{ for some } k, g, g', t. \end{aligned} \tag{8}$$

The first test generalizes Fama-MacBeth style t tests to speak to differences in expected returns across market segments. The second test enriches the first by adding the information embedded in the dynamics of cross-sectional slopes to distinguish among groups of assets.[9] Indeed, groups may have different factor dynamics but identical unconditional risk prices. Intuitively, both tests assess whether adding clusters beyond the first generates differences in risk prices beyond what we would

[9] A stronger form of both tests applies if factors are tradeable, namely, under the null hypothesis of no segmentation, the cross-sectional slopes for all segments should equal each other *and* the returns to the factor. We drop this latter condition to accommodate non-tradeable factors such as intermediary capital ratio innovations.

expect by chance. Note that neither test examines the equality of intercepts ($\bar{\alpha}^{(g)}$ or $\alpha_t^{(g)}$) because our focus is on risk price heterogeneity rather than on differences in zero-beta rates. However, both tests can easily be extended to include tests of equality of intercepts, as well.

We define test statistics for (7) and (8) analogously to F statistics for tests of equality of average slopes ("avg") and slope dynamics ("dyn"). Doing so requires some auxiliary quantities. First, let the estimated difference in cross-sectional slopes at date t for clusters g and g' be $\Delta\lambda_t^{(g,g')}$ and the time-series average of this quantity be $\Delta\bar{\lambda}^{(g,g')}$. Second, define $\hat{\Sigma}_{\lambda^{(g)}}$ as the cross-sectional covariance matrix of the parameter estimates at date t for cluster g, and define $\hat{\Sigma}_{\bar{\lambda}^{(g)}}$ as the time-series average of this value. Because cross-sectional slopes are estimated separately group-by-group and date-by-date, the covariance matrices of $\Delta\lambda_t^{(g,g')}$ and $\Delta\bar{\lambda}^{(g,g')}$ are simply $\hat{\Sigma}_{\lambda^{(g)}} + \hat{\Sigma}_{\lambda^{(g')}}$ and $\hat{\Sigma}_{\bar{\lambda}^{(g)}} + \hat{\Sigma}_{\bar{\lambda}^{(g')}}$, respectively. Combining these quantities into test statistics for differences in averages and differences in dynamics between all groups and factors, we obtain

$$F^{Avg} = \frac{1}{(G-1)\,K} \sum_{g=1}^{G-1} \Delta\bar{\lambda}^{(g,g+1)\prime} \left(\hat{\Sigma}_{\bar{\lambda}^{(g)}} + \hat{\Sigma}_{\bar{\lambda}^{(g+1)}}\right)^{-1} \Delta\bar{\lambda}^{(g,g+1)}, \tag{9}$$

$$F^{Dyn} = \frac{1}{(G-1)\,KP} \sum_{g=1}^{G-1} \sum_{t\in\mathcal{P}} \Delta\lambda_t^{(g,g+1)\prime} \left(\hat{\Sigma}_{\lambda_t^{(g)}} + \hat{\Sigma}_{\lambda_t^{(g+1)}}\right)^{-1} \Delta\lambda_t^{(g,g+1)}. \tag{10}$$

Incidentally, both tests downweight between-group differences in factor premia on "junk" factors for which the dispersion in betas is low because the test statistics normalize differences in λs by the precision of λ estimates—which themselves are proportional to the cross-sectional variation in βs—on average or date-by-date.

The test statistics use estimated parameters from the $\mathcal{P}$ sample taking group memberships estimated on the $\mathcal{R}$ sample as fixed. If the dependence between observations in the $\mathcal{R}$ and $\mathcal{P}$ samples is limited, then the over-fitting problem that arises when the same sample is used for both group membership estimation and testing is eliminated.[10] In principle, we can then use standard hypothesis testing methods to implement a test for multiple clusters: for example, the test of equal average risk prices is a simple F test. However in simulation studies with realistic data generating processes, we found poor size control using this approach. We instead adopt a permutation testing approach (see, for example, Lehmann and Romano (2005)) to generate correct critical values for the tests of hypotheses (7) and (8). The permutation tests randomly shuffle group assignments and are particularly well-suited for evaluating whether groups differ in some coefficient(s) of interest, e.g., the set of cross-sectional slopes.

We implement our permutation tests as follows. Given a set of group assignments $\hat{\gamma}_R$ obtained on the $\mathcal{R}$ sample, we draw random group assignments in which portfolios have the same N_g/N

[10] To ensure that the dependence between our subsamples is negligible, we leave a gap of one year between the $\mathcal{R}$ and $\mathcal{P}$ samples. If the number of years in our sample is odd, we use the middle year as the "gap," otherwise we split the sample evenly and then drop the last year of the $\mathcal{R}$ sample.

probabilities of being in group g. We next calculate the statistics in equations (9)–(10) on the $\mathcal{P}$ sample. We repeat this procedure M=5,000 times to obtain a distribution of test statistics. Each permutation acts similar to a bootstrap draw, and generating many permutations fills out the distribution of the test statistic under the null hypothesis.[11] p-values then are computed simply as the proportion of permutation statistics larger than the test statistics. The advantage of this approach is that it adjusts for possible issues arising from departures from our assumptions, such as finite group sizes and sample lengths, while requiring less structure than a traditional bootstrap design.

Finally, the test for multiple clusters requires the researcher to specify the number of clusters under the alternative, G. Rather than imposing a specific value, we instead implement the test for a range of values $G = 2, 3, 4, 5$. We set the largest value under the alternative to five to balance the requirement that each cluster be "large" against the possibility of many market segments in the data. We account for the fact that this method implements four individual tests by using a simple Bonferroni correction. In practice, this means we calculate the overall p-value as the minimum of the individual p-values multiplied by four, and we then compare the overall p-value to desired size of the test (e.g., 5%).

C. Finite Sample Properties of the Test for Multiple Clusters

We analyze the finite-sample properties of the above tests for multiple clusters via an extensive simulation study. Ensuring that the proposed tests have satisfactory finite-sample properties is a necessary condition for interpreting the test results that we present in the next section. We consider a range of values of N, T, K and G, to match the various specifications that we consider in our empirical analysis in Section V.

The DGP for the simulation study is obtained as follows, and it uses data described in detail in the next section. We first estimate factor means, μ_f, and covariance matrices, Σ_f, for two representative asset pricing models, the CAPM ($K = 1$) and Carhart ($K = 4$) factor models, at daily and monthly frequencies. We next estimate time series betas and idiosyncratic volatilities for each of 234 domestic equity portfolios (the portfolio set "P3" described in Section IV) for each factor model at daily and monthly frequencies.

For each simulation s, we draw with replacement N=75 stocks or N=225 stocks, where each draw consists of $(\{\beta_{i1}, \ldots, \beta_{iK}\}, \sigma_{i\epsilon})$ pairs. Asset returns are obtained as $r_{it} = \beta_i f_t + \epsilon_{it}$, with simulated factor returns $f \sim N(\mu_f, \Sigma_f)$ with simulated idiosyncratic returns $\epsilon_i \sim N\left(0, \sigma^2_{i\epsilon}\right)$. We simulate factor returns with two time-series dependency structures: factor returns are either i.i.d. or they exhibit time-series dependence as in a GARCH(1,1) process. We include simulations with volatility dependence to confirm that our procedure is robust to (1) heteroskedasticity and (2) long-

[11] It is possible to consider all possible permutations of the assets when N and G are very small;, however, for the values of N in our applications this approach is infeasible. As noted in Lehmann and Romano (2005), using a randomly drawn subset of all possible permutations, as we do, also controls the level of the test.

range volatility dependence that may introduce dependence between the $\mathcal{R}$ and the $\mathcal{P}$ samples. Our GARCH processes use Zivot (2009)'s parameters of $a = 0.09$ and $b = 0.89$ for daily simulations and $a = 0.18$ and $b = 0.78$ for monthly simulations estimated from S&P 500 data for 1986–2003, and we impose constant conditional correlation among the factors (Bollerslev (1990)), with correlations estimated for the Carhart model on the entire 1963–2016 sample.

For the "daily" design, we set T=10,000, and for the "monthly" design we set T=300. We estimate the multi-cluster factor models using the methods described in Section III.A, using data from our $\mathcal{R}$ sample (the first half minus a year at the end), and we implement tests on the $\mathcal{P}$ sample (the second half). We report rejection frequencies for nominal 0.05 level tests,[12] for a single alternative $G \in \{2, 3, 4, 5\}$, and for a multiple comparison using $G = 2, 3, 4, 5$ and a Bonferroni correction. We simulate under each design S=500 times. Due to computation constraints, we reduce the number of permutations within each simulation from M=5,000 to M=500. A test of correct size should reject 5% of the time, up to simulation variability.

Table I shows that the proposed testing procedure generally has rejection rates that are comparable to the nominal test sizes. For the "daily" simulation design, in the left panel, almost all rejection frequencies are between 0.03 and 0.08. For the "monthly" simulation design, in the right panel, rejection frequencies are generally close to the nominal 0.05 level in the i.i.d. case, though they are sometimes larger in the GARCH case in the lower-right, particularly when N=225. In this case our assumption of T/N being "large" is less plausible. Rejection rates do not appear sensitive to the choice of factor model (CAPM or Carhart) with the exception of the large N and small T design. Overall, we conclude that the test has satisfactory size control in finite samples, though the simulations suggest caution in interpreting borderline results in the monthly samples with larger factor models.

IV. Data

Our data consist of risk factors and well-diversified portfolios common throughout the empirical asset pricing literature. To ensure that our conclusions on market segmentation are robust to choices of a particular factor model, we include several leading factor models in our analysis. These models include the CAPM; the Fama and French (1992) three-factor model ("FF3F"); the Carhart (1997) four-factor model ("Carhart"); the Fama and French (2015) five-factor model ("FF5F"); the He, Kelly, and Manela (2017) intermediary-capital factor model ("HKM");[13] and the Hou, Xue, and Zhang (2015) q-factor model ("HXZQ"). Ken French's website provides the Fama and French

[12]The size control for significance levels of 0.01 and 0.10 are very similar to the 0.05 results presented here.

[13]The intermediary capital factor is available monthly from January 1970 and daily from January 2000. For sample periods in which daily data are available for portfolio returns but not for the intermediary capital factor, we use monthly intermediary capital factors and monthly portfolio returns to estimate betas and idiosyncratic volatilities, and we convert monthly volatility estimates to daily estimates by dividing by $\sqrt{21}$. In periods for which both data frequencies are available, the cross-sectional correlation in betas and idiosyncratic volatilities estimated using daily and monthly intermediary capital factors is about 90%, with slight variation depending on the portfolio set considered.

Table I: Finite Sample Rejection Rates

This table reports the proportion of simulations in which we reject a single cluster in favor of multiple clusters using the two F tests described in Section III.B when the data comes from a single cluster model. We consider 16 simulation designs: N=75 and N=225 simulated portfolios; CAPM (K=1) and Carhart (K=4) factor models; i.i.d. and GARCH(1,1) factor realizations; and T=10,000 days (left columns) and T=300 months (right columns). We use S=500 simulations of each design. GARCH processes for the factors use parameters $a = 0.09$ and $b = 0.89$ for daily simulations and $a = 0.18$ and $b = 0.78$ for monthly simulations (both from Zivot (2009)), and factors have constant conditional correlation (Bollerslev (1990)). All tests are at the 0.05 nominal level. We report rejection frequencies for a single alternative $G \in \{2, 3, 4, 5\}$, and for a multiple comparison using G=2,3,4,5 and a Bonferroni correction.

Design			**T = 10,000 Days**					**T = 300 Months**				
N	K	**GARCH**	**G=2**	**=3**	**=4**	**=5**	**∈2–5**	**G=2**	**=3**	**=4**	**=5**	**∈2–5**
			Panel A: Tests for Equality of $\bar{\lambda}_k$ Across Groups									
75	**1**	**N**	0.05	0.07	0.07	0.05	0.06	0.04	0.04	0.07	0.07	0.07
75	**4**	**N**	0.06	0.06	0.04	0.05	0.07	0.07	0.06	0.06	0.05	0.07
225	**1**	**N**	0.04	0.06	0.06	0.07	0.07	0.05	0.05	0.05	0.06	0.06
225	**4**	**N**	0.04	0.03	0.07	0.07	0.06	0.08	0.08	0.11	0.08	0.10
75	**1**	**Y**	0.06	0.08	0.07	0.05	0.07	0.04	0.05	0.07	0.07	0.07
75	**4**	**Y**	0.05	0.05	0.04	0.05	0.07	0.06	0.06	0.06	0.04	0.07
225	**1**	**Y**	0.04	0.06	0.07	0.06	0.05	0.04	0.05	0.05	0.03	0.05
225	**4**	**Y**	0.06	0.06	0.06	0.06	0.08	0.07	0.09	0.09	0.08	0.11
			Panel B: Tests for Equality of λ_{kt} Across Groups									
75	**1**	**N**	0.08	0.04	0.03	0.08	0.08	0.03	0.03	0.03	0.10	0.06
75	**4**	**N**	0.04	0.01	0.04	0.08	0.07	0.06	0.04	0.03	0.05	0.05
225	**1**	**N**	0.05	0.11	0.09	0.09	0.08	0.04	0.05	0.07	0.06	0.06
225	**4**	**N**	0.08	0.07	0.06	0.03	0.07	0.10	0.12	0.09	0.06	0.12
75	**1**	**Y**	0.09	0.04	0.01	0.08	0.07	0.05	0.05	0.05	0.13	0.08
75	**4**	**Y**	0.04	0.01	0.03	0.07	0.04	0.06	0.03	0.02	0.05	0.04
225	**1**	**Y**	0.07	0.08	0.09	0.08	0.08	0.06	0.06	0.05	0.06	0.06
225	**4**	**Y**	0.07	0.08	0.07	0.04	0.06	0.13	0.13	0.12	0.07	0.17

factors and the momentum factor both in domestic and global versions (as in Fama and French (2012)). Asaf Manela's website provides the intermediary capital factor. Finally, Lu Zhang shares his investment and return-on-equity factors for the q-factor model. We adjust the time series length and sampling frequency of the factors to match the corresponding portfolio sets described below.

Throughout our analysis we use value-weighted portfolios as our test assets.[14] We obtain characteristic-sorted equity portfolio data from Ken French's website. Our domestic equity data are daily for 1963 to 2016. The domestic equity portfolios are formed using a standard sorting characteristics including market capitalization, book-to-market ratios, prior returns, investment rates, operating profitability, market beta, and industry group. We also use double-sorted portfolios based on market capitalization and market beta, book-to-market ratio, prior return, investment rate, and operating profitability. The international stock portfolio data are also daily and run from 1991 to 2016. They consist of double-sorted size-book-to-market and size-prior return portfolios for developed markets in North America, Europe, Japan, and Asia Pacific (excluding Japan) regions. Details on the underlying data and security set are provided in Fama and French (2012). In our analysis we pair these portfolios with global versions of the Carhart factors, and we substitute the global market factor for the US market factor in the intermediary-capital factor model. Because return on equity is constructed using accounting data that varies across countries and is not readily accessible, we drop the q-factor model in studying risk price heterogeneity for these portfolios.

The cross-asset class sample is mainly courtesy of Asaf Manela, and our data is monthly for 1986 to 2010. In addition to domestic equities, these data include commodities, US Treasuries and corporate bonds, sovereign bonds, options, and currencies.[15] Unlike the domestic and international stock portfolios, the diverse He, Kelly, and Manela (2017) portfolios necessarily start and end at different dates as new asset classes come into being and database availability changes. While Manela's data include some asset classes from 1970, to maintain a near-balanced panel, we initialize the sample at the start of the year that commodity and options data begin (1986) and end the sample at the end of the year that foreign exchange data end (2010). He, Kelly, and Manela (2017) describes portfolio construction and primary sources for these data in greater detail.

Table II summarizes the seven collections of portfolios. The first component of each subtable marks the constituents of each portfolio set. Our smallest and largest portfolio sets consist of 75 (P1) and 234 (P3) domestic equity portfolios, respectively. All portfolio sets sort on at least three variables (e.g., region, size, and book-to-market ratio) to ensure coverage of several potential dimensions of market segmentation. The second component of each subtable reports summary

[14] We use portfolios rather than individual securities for the usual reasons: (1) to increase the stability of security risk characteristics over time; (2) to decrease the measurement error in betas through diversification of idiosyncratic risk; and (3) to reduce the sparsity of the matrix of realized returns. Importantly, risk premia estimates obtained using portfolio returns do not generally apply to portfolio constituents because comovements between securities strongly influence portfolio dynamics.

[15] We exclude credit default swaps and sovereign bonds from our analysis because data for these asset classes are available only in the second half of our sample period (2001–2012 and 1995–2011, respectively).

Table II: Summary Statistics for Collections of Test Portfolios

This table reports the composition and summary statistics for eight main collections of test portfolios. The first panel consists of domestic test portfolios. The top component indicates whether a set of single- or double-sorted portfolios is included in portfolio collections P1–P3. “Mom,” “Inv,” and “Prof” abbreviate prior return, investment, and operating profitability sorted portfolios, respectively, and “Ind” is the Fama-French 49 industry portfolios. We download daily value-weighted portfolios from the data library on Ken French’s website (“KF”). The middle component reports the average of the mean daily returns (ave $\bar{r}$) and standard deviation of daily returns (ave σ_r) within a portfolio set, as well as the standard deviation of these quantities σ ($\bar{r}$) and σ (σ_r), all in annualized percentage terms. The second panel consists of international test portfolios P4–P5. These portfolios parallel the value-weighted daily size-value and size-momentum portfolios for our four global regions. The third panel consists of diversified domestic and international equity and non-equity assets used in He, Kelly, and Manela (2017) (“HKM”). We download the corresponding monthly portfolio returns for non-equity assets from Asaf Manela’s website. Where data is not compiled by the authors, we report primary sources including the Commodities Research Bureau (“CRB”), the CRSP Fama Bond Portfolios (“CRSP”), Nozawa (2017) (“N”), Constantinides, Jackwerth, and Savov (2013) (“CJS”), Lettau, Maggiori, and Weber (2014) (“LMM”), and Menkhoff, Sarno, Schmeling, and Schrimpf (2012) (“MSSS”). We add start and end years for data availability in this panel because they differ across asset classes.

(a) Domestic Equity Portfolios, Daily Data, 1963–2016

#	**Size-β_{mkt}**	**Size-B/M**	**Size-Mom**	β_{mkt}	**B/M**	**Size**	**Mom**	**Ind**	**Inv**	**Prof**	**Size-Inv**	**Size-Prof**
P1	✓	✓	✓									
P2	✓	✓	✓	✓	✓	✓	✓					
P3	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Ave $\bar{r}$	12.71	13.10	12.91	11.13	12.45	12.53	10.90	11.88	11.71	10.84	13.04	12.65
Ave σ_r	17.36	17.18	17.78	17.77	16.71	16.47	18.06	21.57	16.69	16.93	16.98	17.08
σ ($\bar{r}$)	1.76	2.63	4.01	0.67	1.84	0.90	3.73	1.90	1.60	1.21	2.26	1.95
σ (σ_r)	4.58	1.72	2.94	4.93	1.39	0.98	3.32	4.88	1.57	1.45	1.71	1.56
N	25	25	25	10	10	10	10	49	10	10	25	25
Source	KF	KF	KF	KF	KF	KF	KF	KF	KF	KF	KF	KF

Table II: Summary Statistics for Collections of Test Portfolios (Continued)

(b) International Equity Portfolios, Daily Data, 1991–2016

	Size-Value 25				**Size-Momentum 25**			
#	**North America**	**Asia-Pacific**	**Europe**	**Japan**	**North America**	**Asia-Pacific**	**Europe**	**Japan**
P4	✓	✓	✓	✓				
P5	✓	✓	✓	✓	✓	✓	✓	✓
Ave $\bar{r}$	11.67	9.63	8.39	4.48	11.70	8.09	6.61	4.72
Ave σ_r	19.32	17.69	16.71	21.74	19.60	18.36	17.41	22.21
$\sigma(\bar{r})$	2.30	4.00	2.16	2.18	4.43	6.38	5.14	2.29
$\sigma(\sigma_r)$	2.13	1.90	2.85	2.06	2.84	2.94	3.51	2.34
N	25	25	25	25	25	25	25	25
Source	KF	KF	KF	KF	KF	KF	KF	KF

(c) Cross-Asset Class Portfolios, Monthly Data, 1986–2010

#	**Size-B/M**	**Size-β_{mkt}**	**Size-Mom**	**Commod**	**US Bonds**	**Options**	**FX**
P6	✓			✓	✓	✓	✓
P7	✓	✓	✓	✓	✓	✓	✓
Ave $\bar{r}$	12.61	12.92	12.40	5.61	7.11	8.23	3.20
Ave σ_r	19.83	19.26	20.88	25.87	3.78	16.07	8.21
$\sigma(\bar{r})$	2.64	1.49	3.59	6.22	1.58	5.79	3.41
$\sigma(\sigma_r)$	3.13	4.93	4.68	6.24	1.70	2.03	1.41
N	25	25	25	23	20	18	12
Source	KF	KF	KF	HKM/CRB	HKM/CRSP/N	HKM/CJS	HKM/LMM/MSSS
HKM Start Year	1970	N/A	N/A	1986	1974	1986	1976
HKM End Year	2012	N/A	N/A	2012	2012	2012	2010

statistics for each set of sorted portfolios or asset class. We report the first and second moments of average returns and return volatility for each grouping. Dispersion in average returns and volatility within each portfolio set is on the order of several percent per year, indicating considerable variation in factor exposures or risk prices within each group. These quantities also vary across sorts, indicating that different sorting variables capture different dimensions of heterogeneity.

For each factor model-portfolio set combination, we repeat our analysis using both the full time series and shorter subsamples. Just as the risk characteristics of portfolios may change over time, so too may the market frictions that separate portfolios into segments with different risk prices. We split the data into two halves for international equity and cross-asset class analyses, and we increase the number of splits to three for domestic equity analyses because equity data are available for a longer time period. Splitting the sample allows our methodology to accommodate time-varying risk characteristics and segmentation without incurring the extreme computational costs of rolling estimation of group assignments. At the same time it allows us to evaluate stability in these characteristics, as we do in Appendix B.

V. Segmentation Everywhere

A. Testing for Market Segmentation

Table III presents our main empirical result: we find evidence of segmentation in *all* combinations of test assets, benchmark factor models, and time periods.

Focusing first on the left columns of Table III, we find evidence of differing *average* prices of risk across clusters for most domestic equity portfolio sets and factor models, and for almost all international equity and multi-asset class portfolio sets and factor models. The main exceptions to this strong evidence of multiple clusters are portfolio set P1, which is comprised of the most common double-sorted domestic equity portfolios, and the CAPM factor model. P1 is not sufficiently diverse to feature meaningful heterogeneity in average risk prices, and the CAPM is too poor a model of cross-sectional variation in expected returns for differences in average risk prices to be detected. Differences in unconditional risk premia are important for almost all other environments with richer cross-sections of assets or better factor models for returns.

Tests of equal risk *dynamics* in the right columns make much stronger statements about segmentation. We find that *every test except one* (out of 69 in total) rejects the null hypothesis of a single cluster at the 5% significance level, and all p-values are less than 0.1% for portfolio sets P2–P7. Our simulation study suggests that these tests somewhat over-reject for monthly data with larger factor models, but the p-values reported here are far from borderline cases that warrant statistical caution. From these strong rejections we conclude that cross-sectional variation in risk prices is ubiquitous. In addition to the international and cross-asset class contexts, where we may have anticipated segmentation a priori, even one of the world's most-developed and liquid markets

Table III: Bonferroni Adjusted p-Values for Tests of Multiple Clusters Against a Single Clusters

This table reports p-values from F tests described in Section III.B for comparing a multiple-clusters model to a single-cluster model. We perform pairwise comparisons of G=2,3,4,5 clusters against one cluster and obtain a Bonferroni-adjusted p-value by taking the minimum individual test p-value and multiplying it by four. The left columns report tests of equality of average risk prices across clusters for all factors ($\bar{\lambda}_k$), and the right columns report tests of equality of cross-sectional slopes across clusters for all factors and dates (λ_{kt}). Portfolio sets and factor models are described in Section IV. For the He, Kelly, and Manela (2017) factor model, we use daily data for the most recent time period and monthly data for earlier time periods. We do not have sufficient coverage for their intermediary capital factor for 1963–1980 to include it in the domestic equity portfolio analysis. The q-factor (HXZQ) model is excluded from the international portfolio analysis because we do not have global return-on-equity factor data.

(a) Domestic Equity Portfolios

		Equal Average Risk Prices $\bar{\lambda}_k$				**Equal Risk Prices** λ_{kt}			
Portfolios	Model	1963–2016	1963–1980	1981–1998	1999–2016	1963–2016	1963–1980	1981–1998	1999–2016
P1	CAPM	0.312	0.003	0.025	0.561	0.026	0.026	0.012	0.006
	FF3F	0.057	0.112	0.166	0.011	0.000	0.000	0.000	0.000
	Carhart	0.050	0.279	0.287	0.102	0.000	0.000	0.000	0.000
	FF5F	0.078	0.054	0.009	0.375	0.000	0.074	0.000	0.000
	HKM	0.236		0.134	0.057	0.037		0.010	0.001
	HXZQ	0.000	0.896	0.000	0.671	0.000	0.000	0.000	0.000
P2	CAPM	0.001	0.050	0.014	0.219	0.000	0.000	0.000	0.000
	FF3F	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Carhart	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000
	FF5F	0.001	0.002	0.000	0.100	0.000	0.000	0.000	0.000
	HKM	0.345		0.004	0.038	0.000		0.000	0.000
	HXZQ	0.000	0.009	0.000	0.000	0.000	0.000	0.000	0.000
P3	CAPM	0.052	0.001	0.000	0.006	0.000	0.000	0.000	0.000
	FF3F	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000
	Carhart	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000
	FF5F	0.000	0.018	0.002	0.000	0.000	0.000	0.000	0.000
	HKM	0.543		0.000	0.007	0.000		0.000	0.000
	HXZQ	0.000	0.007	0.000	0.000	0.000	0.000	0.000	0.000

Table III: Tests of Multiple Groups Against a Single Group (Continued)

(b) International Equity Portfolios

		Equal Average Risk Prices $\bar{\lambda}_k$			**Equal Risk Prices λ_{kt}**		
Portfolios	Model	1991–2016	1991–2003	2004–2016	1991–2016	1991–2003	2004–2016
P4	CAPM	0.000	0.000	0.026	0.000	0.000	0.000
	FF3F	0.002	0.000	0.001	0.000	0.000	0.000
	Carhart	0.000	0.000	0.000	0.000	0.000	0.000
	FF5F	0.018	0.000	0.000	0.000	0.000	0.000
	HKM	0.000	0.000	0.000	0.000	0.000	0.000
	HXZQ						
P5	CAPM	0.000	0.282	0.000	0.000	0.000	0.000
	FF3F	0.000	0.000	0.000	0.000	0.000	0.000
	Carhart	0.000	0.000	0.000	0.000	0.000	0.000
	FF5F	0.000	0.002	0.000	0.000	0.000	0.000
	HKM	0.000	0.000	0.000	0.000	0.000	0.000
	HXZQ						

(c) Cross-Asset Class Portfolios

		Equal Average Risk Prices $\bar{\lambda}_k$			**Equal Risk Prices λ_{kt}**		
Portfolios	Model	1986–2010	1986–1997	1998–2010	1986–2010	1986–1997	1998–2010
P6	CAPM	0.000	0.000	0.000	0.000	0.000	0.000
	FF3F	0.000	0.000	0.000	0.000	0.000	0.000
	Carhart	0.000	0.000	0.000	0.000	0.000	0.000
	FF5F	0.000	0.010	0.000	0.000	0.000	0.000
	HKM	0.000	0.000	0.000	0.000	0.000	0.000
	HXZQ	0.000	0.000	0.000	0.000	0.000	0.000
P7	CAPM	0.000	0.000	0.000	0.000	0.000	0.000
	FF3F	0.000	0.016	0.000	0.000	0.000	0.000
	Carhart	0.000	0.000	0.000	0.000	0.000	0.000
	FF5F	0.000	0.053	0.000	0.000	0.000	0.000
	HKM	0.000	0.000	0.000	0.000	0.000	0.000
	HXZQ	0.000	0.000	0.000	0.000	0.000	0.000

such as US stocks exhibits significant variation in compensation to factor exposure. Given the large size of our typical panels, these strong rejections of the null of unified risk prices may not correspond with *economically* meaningful differences in cross-sectional dispersion of realized returns and expected returns. The next subsection confirms that they do.

Despite the strong rejections of integrated markets, our results are conservative in several respects. First, our ability to identify market segments depends on whether the selected factors have heterogeneous risk prices. In this respect our paper joins most others in the segmentation literature in depending on the choice of factor model. We address this issue in part by evaluating market segmentation with a battery of leading factor models and diverse portfolio sets. Second, the Bonferroni adjustment for penalizing multiple tests is too severe if tests are correlated, as our tests likely are. Third, to the extent that group assignments are time-varying, our subsample approach for testing for multiple clusters will have lower power, as using the "wrong" clusters for the $\mathcal{P}$ sample decreases the improvement in model fit with multiple clusters relative to a single cluster. Evidence of segmented markets comes through strongly in Table III despite these features.

B. The Economic Importance of Market Segmentation

To assess the economic contribution of heterogeneous risk-price models, we first define two measures of the dispersion in expected returns explained by a particular model. Perhaps of greatest academic interest is the estimated explanatory power for the cross-section of expected returns. We select the number of clusters $G^* \in \{2, 3, 4, 5\}$ as the value that minimizes the AIC. We define our first measure for the G^*-cluster model as the variance of unconditional model-implied average returns across test assets,

$$\sigma^2_{G^*}(\bar{r}) \equiv var_i\left(\frac{1}{T}\sum_{t=1}^{T}\left(\alpha_t^{(g)} + \beta_i\lambda_t^{(g)}\right)\right), \tag{11}$$

We calculate the variance in expected returns analogously for the one-cluster model using standard, full-sample Fama-MacBeth estimates for α_t and λ_t, and we report the ratio of $\sigma^2_{G^*}(\bar{r})/\sigma^2_1(\bar{r})$.

Our second measure quantifies Sharpe ratio improvements from using factors constructed from multiple clusters rather than from one. This measure is the difference in maximal in-sample Sharpe ratios achievable using factor-mimicking portfolios from all clusters and from a single combined cluster,

$$\Delta SR_{G^*} \equiv \sqrt{\mu'_\Lambda \Sigma_\Lambda^{-1}\mu_\Lambda} - \sqrt{\mu'_\lambda \Sigma_\lambda^{-1}\mu_\lambda}. \tag{12}$$

If the factors are tradeable, $\sqrt{\mu'_\lambda \Sigma_\lambda^{-1}\mu_\lambda}$ is simply the maximal in-sample Sharpe ratio achievable using the factors alone, $\sqrt{\mu'_f \Sigma_f^{-1}\mu_f}$.

The second measure takes the perspective of an arbitrageur able to frictionlessly invest in all portfolio sets. To the extent that risk prices vary in the cross-section, investing across clusters improves the maximal in-sample Sharpe ratio by enabling (1) tilts toward "local" factors with particularly high compensation per unit risk and (2) diversification of risk across imperfectly correlated mimicking portfolios. ΔSR_{G^*} answers how much better our hypothetical unconstrained investor can do in mean-variance terms by recognizing risk price heterogeneity. Equivalently, from a no-arbitrage perspective, it represents the magnitude of segmentation frictions in the size of gains remaining *despite* the activities of sophisticated potential arbitrageurs. In this respect, ΔSR_{G^*} measures cross-sectional limits of arbitrage described by Gromb and Vayanos (2018).

Table IV reports these measures for all combinations of test assets, benchmark factor models, and time periods. Despite not being the optimization objective, dispersion in average returns increases almost everywhere, and often considerably. Starting with the domestic portfolios and setting aside outlier CAPM and He, Kelly, and Manela (2017) (HKM) rows, increases in cross-sectional dispersion in average returns range from 3% to 167%, with with 25th and 75th percentiles increases of 16% and 76%, respectively. These improvements are on par with adding additional factors to standard asset pricing models; for example, augmenting the Fama-French three-factor model with investment and profitability factors, as in the Fama-French five-factor model, increases cross-sectional dispersion in average returns by 49% for P1 portfolios, 16% for P2 portfolios, and 28% for P3 portfolios. Performance gains are much larger for the CAPM and HKM models because (1) both models perform quite badly in explaining cross-sectional variation in average returns, and (2) additional clusters likely pick up omitted factors, a possibility we address in Section VI. Conversely, while our statistical tests reject the null of a single cluster for all portfolio sets, the improvement in likelihood from adding additional clusters is sometimes small. The AIC selects a single cluster in several instances for portfolio set P1. This portfolio set is sufficiently homogeneous that the four- and five-factor models we consider capture most of the economically interesting cross-sectional variation in returns, and additional clusters do not add much in terms of explained expected-return variation and increased ex post Sharpe ratios.

Turning attention to the second and third panels, we see larger improvements in cross-sectional dispersion in model-implied average returns. The smallest increase in the international setting is 11%, and the largest increases are of several hundred percent, again excluding the global CAPM from consideration because of clearly omitted factors. As we discuss in a detailed example in the next section, international markets have region-specific variants of their factors that make global models problematic. Intriguingly, by this metric, inter-regional segmentation does not diminish over time for these markets, suggesting that barriers to international arbitrage remain important. Finally, in the cross-asset class context, we see comparable gains in performance for explaining the cross-section of average returns, even though with the exception of HKM we limit ourselves to models designed with only domestic equities in mind. However, in this panel we also observe one reduction in cross-sectional explanatory power when adding clusters to the Carhart model. This situation can occur if portfolios with relatively extreme betas belong to groups with low

Table IV: Contribution of Clusters to Expected Return Variation and Sharpe Ratio Improvements

Table reports the ratio of cross-section variance in expected returns explained by multiple-cluster models to single-cluster models and the difference in maximal in-sample Sharpe ratios between multiple-cluster models and single-cluster models. We construct these measures as follows. First we estimate group assignments, risk prices, and likelihoods for models with $G = \{1,2,3,4,5\}$ groups using the entirety of each sample period. Next, we select the number of groups using the AIC. '*' indicate instances in which the AIC selects a single cluster. For the one-cluster and G^* cluster models, we then calculate the cross-sectional variance in average returns ($\text{var}(\bar{r})$) as well as the maximal in-sample Sharpe ratio attainable using the mimicking portfolios ($\sqrt{\mu_\Lambda' \Sigma_\Lambda^{-1} \mu_\Lambda}$). We tabulate the ratio of cross-sectional variances $\text{var}(\bar{r})^{(G^*)}/\text{var}(\bar{r})^{(1)}$ as "$\text{Var}(\bar{r})$" and increases in annualized Sharpe ratios $\sqrt{\mu_\Lambda' \Sigma_\Lambda^{-1} \mu_\Lambda} - \sqrt{\mu_\lambda' \Sigma_\lambda^{-1} \mu_\lambda}$ as "ΔSR." We repeat this procedure for all combinations of portfolio sets, risk models, and sample periods. Portfolios and models are described in the text. For the He, Kelly, and Manela (2017) factor model, we use daily data for the most recent time period and monthly data for earlier time periods. We do not have sufficient coverage for their intermediary capital factor for 1963–1980 to include it in the domestic equity portfolio analysis. The q-factor (HXZQ) model is excluded from the international portfolio analysis because we do not have global return-on-equity factor data.

(a) Domestic Equity Portfolios

		1963–2016		**1963–1980**		**1981–1998**		**1999–2016**	
Portfolios	Model	$\text{Var}(\bar{r})$	ΔSR	$\text{Var}(\bar{r})$	ΔSR	$\text{Var}(\bar{r})$	ΔSR	$\text{Var}(\bar{r})$	ΔSR
P1	CAPM	3.77	0.26	55.78	1.04	2.31	0.77	161.88	0.15
	FF3F	1.81	0.74	1.19	0.26	1.44	0.70	1.80	-0.09
	Carhart	1.03	0.10	*	*	1.03	0.13	1.38	0.16
	FF5F	*	*	*	*	1.95	1.41	1.10	0.15
	HKM	8.25	0.33			34.85	1.18	5.30	0.38
	HXZQ	1.16	0.67	*	*	*	*	2.67	0.29
P2	CAPM	6.48	0.55	80.02	1.61	1.02	0.20	52.70	0.32
	FF3F	2.15	0.69	1.40	0.40	2.67	1.93	2.48	0.74
	Carhart	1.05	0.07	1.17	0.29	1.17	0.95	1.30	0.61
	FF5F	1.61	0.53	1.15	0.27	1.71	0.79	1.05	0.24
	HKM	2.60	0.17			7.72	0.68	4.16	0.35
	HXZQ	1.23	0.47	0.97	0.25	1.76	2.11	2.31	0.68
P3	CAPM	2.75	0.17	22.66	0.84	1.22	0.46	6.16	0.48
	FF3F	1.67	0.82	1.41	0.96	1.67	1.57	1.75	0.47
	Carhart	1.41	0.86	1.16	0.35	1.07	0.49	1.55	0.85
	FF5F	1.49	0.54	1.29	0.84	2.57	1.79	1.22	0.14
	HKM	6.25	0.08			55.58	1.49	10.42	0.72
	HXZQ	1.51	0.69	1.03	0.11	1.20	1.90	2.39	0.69

Table IV: Contribution of Clusters to Expected Return Variation and Sharpe Ratio Improvements (Continued)

(c) International Equity Portfolios

		1991–2016		**1991–2003**		**2004–2016**	
Portfolios	Model	Var($\bar{r}$)	ΔSR	Var($\bar{r}$)	ΔSR	Var($\bar{r}$)	ΔSR
P4	CAPM	17.79	0.46	761.91	0.84	19.80	0.35
	FF3F	3.35	0.35	6.61	0.73	16.26	0.34
	Carhart	3.03	0.35	4.02	0.86	6.63	0.02
	FF5F	2.26	0.34	1.97	0.79	1.33	0.44
	HKM	2.63	0.43	7.92	0.98	2.57	0.51
	HXZQ						
P5	CAPM	1126.70	0.57	116.83	0.47	200.27	0.89
	FF3F	2.33	0.32	5.23	1.12	18.59	0.82
	Carhart	3.80	1.04	2.61	1.31	11.28	1.22
	FF5F	2.54	0.51	1.31	0.44	1.11	0.24
	HKM	7.17	0.80	5.46	0.47	5.30	0.93
	HXZQ						

(d) Cross-Asset Class Portfolios

		1986–2010		**1986–1997**		**1998–2010**	
Portfolios	Model	Var($\bar{r}$)	ΔSR	Var($\bar{r}$)	ΔSR	Var($\bar{r}$)	ΔSR
P6	CAPM	11.97	1.03	1.12	0.48	69.22	1.70
	FF3F	1.33	0.55	1.05	0.19	1.84	0.87
	Carhart	0.81	0.61	0.97	0.11	1.13	0.68
	FF5F	1.15	0.42	1.22	0.24	2.15	0.93
	HKM	7.48	0.79	1.03	0.33	19.69	1.40
	HXZQ	1.05	0.44	1.33	0.25	2.69	1.06
P7	CAPM	4.93	1.11	1.23	1.00	5.95	0.68
	FF3F	1.27	0.80	1.03	0.48	1.34	1.26
	Carhart	1.08	0.87	1.10	0.46	2.48	1.80
	FF5F	1.23	0.95	1.18	0.39	2.80	1.93
	HKM	4.47	0.87	0.95	0.50	2.40	1.02
	HXZQ	1.21	0.90	1.28	0.47	1.56	1.41

compensation for risk, as in Frazzini and Pedersen (2014), who find that high-market beta stocks have particularly low compensation for risk.

Perhaps of greater practical interest are increases in Sharpe ratios that can be attained by unconstrained arbitrageurs who use "local" versions of the factors. As with the dispersion in expected returns metric, using multiple clusters is as important for expanding the mean-variance frontier as using cutting-edge factor models. Returning to the domestic portfolios of the first panel of Table IV, typical improvements in annual Sharpe ratios are comparable to the market's Sharpe ratio of approximately 0.4, and some are as large as the maximal Carhart model Sharpe ratio of 1.3. Moving to the second and third panels, Sharpe ratio improvements of the multiple-cluster models are comparable with the first panel. Large improvements to attainable Sharpe ratios for international portfolios reinforce Asness, Moskowitz, and Pedersen (2013)'s point that cross-region, multi-factor strategies have highly desirable risk-return characteristics, even when compared with similar strategies within a single region.

Taken together, Tables III and IV reveal segmentation everywhere. Cross-sectional differences in factor compensation represent an important new dimension of heterogeneity that is absent from standard asset pricing models. This dimension is as important as differences among risk models, even in the low-friction setting of US equity markets, and especially so in more challenging international and cross-asset class settings. Of course, evidence for segmentation everywhere begs the question of from whence it comes.

C. Risk Price Heterogeneity in Common Economic Settings: Detailed Examples

In this section we investigate the dimensions of cross-sectional heterogeneity in risk prices in three representative settings: domestic portfolios with the Carhart four factors; international equity portfolios with the global Carhart factors; and cross-asset class portfolios with the market and He, Kelly, and Manela (2017) intermediary capital factors. The portfolios correspond to P3, P5 and P7, respectively, in Table II.

Up to this point, our statistical tests and economic interpretation have focused on whether one or multiple sets of risk prices obtain in the data. Here we are more interested in interpreting levels and cross-group differences in average cross-sectional slopes, and a different underlying theory is needed. Specifically, to interpret results from Fama-MacBeth tests for differences in average slopes from zero and from each other, we must account for the impact of estimated group memberships on the cross-sectional slopes.

Bonhomme and Manresa (2015) provide general conditions for problems related to ours under which the parameter estimates based on estimated group memberships have the same limiting distribution as the (infeasible) parameter estimates based on *true* group memberships. A critical condition for the asymptotic negligibility of the error in estimated group memberships is that the clusters are "well separated," that is, that there are indeed multiple clusters in the data. Table III's strong rejection of a single cluster for these examples indicates that this separation condition is met. In our application, this equivalence of the limiting distributions implies that we can estimate the group memberships (γ) and Fama-MacBeth parameters (α and Λ) using the EM algorithm described in Section III.A and then conduct inference on the Fama-MacBeth parameters using standard methods (e.g., t- and F-tests on the time series of the estimated coefficients).

Domestic Equity Portfolios

Table V reports Fama-MacBeth regression estimates from one- and two-cluster models. The top panel reports p-values associated with tests of G clusters against one cluster (the null hypothesis). Tests of equality of average risk prices rejects the null against alternatives of $G = 2$, $G = 5$, and, more generally, $G > 1$ clusters with Bonferroni-adjusted p-values.[16] Tests of equality of all cross-sectional slopes strongly reject the null for all $G \in \{2, 3, 4, 5\}$. The top panel also reports the log likelihoods and AICs associated with G-cluster models in the full sample. We select two clusters using the AIC on this full sample.

The bottom panel presents standard Fama-MacBeth regression estimates in the leftmost column ("All"), as well as group-specific Fama-MacBeth estimates in columns for each group ("Group 1" and "Group 2"). ρ rows report the time-series correlation of factor-mimicking portfolio returns (cross-sectional slopes) and factor realizations, and the last column reports standard F tests in the style of Fama-MacBeth for equality of average risk prices across groups.[17] We also report average within- and across-cluster R^2s for the one- and G^*-cluster models.

The clusters have unequal sizes of 148 portfolios in Group 1 and 86 portfolios in Group 2. Comparing the columns for Groups 1 and 2, Group 2 has larger factor premia for the market, value, and momentum factors, and a smaller factor premium for the size factor. Between-group differences in risk premia for non-market factors are economically large, at about 5%–6%, and these differences are highly significant statistically as judged by F tests for equality of average premia. Both clusters approximate the time series factors for the market, value, and momentum reasonably well, with factor correlations ranging from 75% to 96% for these factors. Group 2's SMB-mimicking portfolio only achieves a correlation of 59% with the SMB factor; by comparison, Group 1's SMB-mimicking portfolio achieves a correlation of 78%.

From the portfolio counts for the two groups reported at the bottom of Table V, we see that Group 2 contains all 81 single- and double-sorted portfolios in which market equity is below the 60th percentile of NYSE stocks, as well as five industry portfolios. Group 1 contains all double-

[16]Intriguingly our test does not reject equality of means for $G = 3$ and $G = 4$ clusters. In these cases the optimal group assignments consist of clusters with smaller dispersion in average factor premia, at least for the typical cluster pair. Maximizing (4) in the three- and four-cluster cases generates groups that instead differ more in their factor dynamics than in their average risk premia.

[17]Namely, the F statistic is constructed as the sum of squared between-cluster differences in average λs normalized by the inverse covariance matrix of the time series of λs. This F statistic equals the square of the usual Fama-MacBeth t-statistic in the case of a single factor and pair of clusters.

Table V: Domestic Equity Portfolios (P3) Example: Domestic Carhart, 1963–2016

The top table reports p-values from F tests described in Section III.B for comparing a multiple-clusters model to a single-cluster model. The underlying factor model is the Carhart four-factor model, and our sample period is 1963–2016. The bottom table reports full-sample Fama-MacBeth estimates of average cross-sectional slopes and associated t-statistics for each group in a model with one cluster ("All") and in a model with $G^* = 2$ clusters selected by the full-sample AIC. Standard errors are Newey-West with 252 daily lags. $\rho_{f,\lambda^{(g)}}$ are the correlations of the factor return and the factor-mimicking portfolio return for cluster g. $p_F\left(\bar{\lambda} =\right)$ is the p-value associated with equality of factor means for the particular factor assuming fixed group memberships. Average cross-sectional R^2s are reported both within each cluster (R^2_G) and across clusters ($R^2_{Combined}$).

Sample	**# Clusters**	**1**	**2**	**3**	**4**	**5**
$\mathcal{P}$	$p_F\left(\bar{\lambda}_k\right)$: 1 vs. G	–	**0.000**	0.487	0.490	0.000
	$p_F\left(\lambda_{kt}\right)$: 1 vs. G	–	**0.000**	0.000	0.000	0.000
Full	LL ($\times 10^{-6}$)	6.438	**6.508**	6.525	6.538	6.553
	AIC ($\times 10^{-6}$)	-12.813	**-12.887**	-12.858	-12.820	-12.785

	Single-Cluster Model	**Two-Cluster Model**		
	All	**Group 1**	**Group 2**	$p_F\left(\bar{\lambda} =\right)$
$\bar{\alpha}$	7.79	7.24	6.71	0.76
t-stat	(3.53)	(2.64)	(2.79)	
$\bar{\lambda}_{MKT}$	-1.13	-0.52	2.33	0.20
t-stat	(-0.44)	(-0.16)	(1.02)	
$\rho_{f,\lambda^{(g)}}$	[0.83]	[0.75]	[0.77]	
$\bar{\lambda}_{HML}$	3.79	2.12	8.12	0.00
t-stat	(2.26)	(1.35)	(3.66)	
$\rho_{f,\lambda^{(g)}}$	[0.95]	[0.92]	[0.82]	
$\bar{\lambda}_{SMB}$	1.60	2.21	-2.54	0.03
t-stat	(0.98)	(1.21)	(-0.86)	
$\rho_{f,\lambda^{(g)}}$	[0.98]	[0.78]	[0.59]	
$\bar{\lambda}_{UMD}$	7.11	5.57	10.43	0.00
t-stat	(3.46)	(2.91)	(4.01)	
$\rho_{f,\lambda^{(g)}}$	[0.99]	[0.96]	[0.92]	
R^2_G	0.91	0.90	0.94	
$R^2_{Combined}$	0.91	0.92		
ME 1-3	81	0	81	
ME 4-5	54	54	0	
Industry	49	44	5	
Other	50	50	0	
N_G	234	148	86	
T	13469	13469	13469	

sorted portfolios in the largest and second-largest market-capitalization groups as well as single-sorted market-capitalization decile portfolios 7–10 and 44 of the 49 industry portfolios. Because our portfolios are value-weighted, the single characteristic-sorted portfolios behave similarly to the highest market capitalization stocks, and they generate similar premia to the high market capitalization portfolios of Group 1. In short, the most important dimension of heterogeneity within our collection of US stock portfolios is that of market capitalization, whereby small stocks earn greater risk premia than large stocks on all but the size factor.

Figure I illustrates our group-specific Fama-MacBeth regressions graphically. As these regressions have multiple (correlated) regressors, we cannot simply plot returns against betas. Instead, we plot the time series averages of residual returns against residual betas, where residual returns are the error term in $r_{it} = \eta_{0t} + \sum_{j \neq k} \eta_{jt}\beta_{ij} + \tilde{r}_{it}$ and residual betas are the error term in $\beta_{ik} = \delta_0 + \sum_{j \neq k} \delta_j \beta_{ij} + \tilde{\beta}_{ik}$. Both regressions are estimated group-by-group as in Section III.A, and the line of best fit for each group has slope $E\left[\lambda_k^{(g)}\right]$.[18] This plot highlights the key economic differences in average risk premia between small- and large-cap stock portfolio clusters.

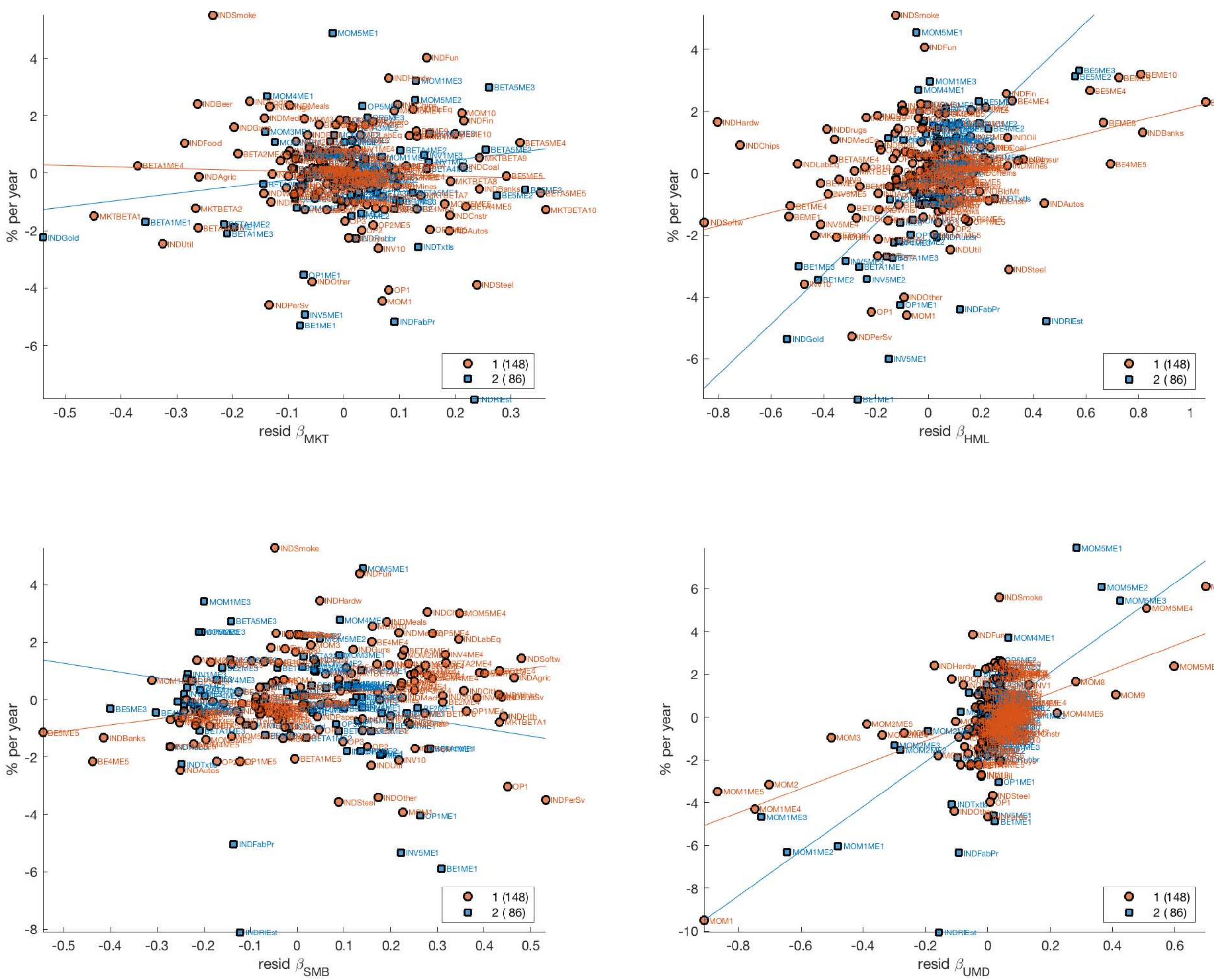
Our findings agree with Hong, Lim, and Stein (2000), Grinblatt and Moskowitz (2004), and Israel and Moskowitz (2013) in identifying market capitalization as an important determinant of cross-sectional differences in risk prices. Unlike these studies, we let the data inform us that size matters for anomaly compensation, and the split between the small and large size quintiles we identify is exactly that determined by previous studies. Notably, these size effects are not sample dependent, as Israel and Moskowitz (2013) caution. While other time periods may see more or less heterogeneity in risk prices along the size dimension, we nonetheless reject equal risk prices across all sets of domestic equity portfolios and across all time periods. Moreover, with the possible exception of the earliest sample period, the P3/Carhart group assignments are stable over time, as discussed in Appendix B; 90% of the group assignments agree between the second and third subsamples, while the agreement between the first/second and first/third subsamples is around two-thirds.

While Group 1 and Group 2 portfolios differ in their risk premia for all four factors, the momentum factor provides the strongest evidence of between-group segmentation. Both groups' mimicking portfolios nearly span the dynamics of the momentum factor, but the average compensation differential between portfolios is nearly 6% per year. Other factor-mimicking portfolios covary less across groups and have smaller compensation differentials per unit β, and hence going long one group and short the other has a worse risk-return trade-off. If arbitrageurs could frictionlessly trade both portfolio sets or a portfolio set and the factor itself, these differences would be driven to

[18]In the Fama-MacBeth procedure, we run multivariate regressions for each date t, $r_{it} = \alpha_t + \sum_k \beta_{ik}\lambda_{kt} + \epsilon_{it}$. By the Frisch–Waugh–Lovell theorem, the coefficient λ_{kt} in each multivariate regression is identical to the coefficient in the univariate regression, $\tilde{r}_{it} = \tilde{\beta}_{ik}\lambda_{kt} + \epsilon_{it}$, where $\tilde{r}_{it}$ and $\tilde{\beta}_{ik}$ are residuals from cross-sectional regressions, $\beta_{ik} = \delta_0 + \sum_{j \neq k} \delta_j \beta_{ij} + \tilde{\beta}_{ik}$ and $r_{it} = \eta_{0t} + \sum_{j \neq k} \eta_{jt}\beta_{ij} + \tilde{r}_{it}$. Averaging across dates, and letting $T \to \infty$, $E[\tilde{r}_i] = E\left[\tilde{\beta}_{ik}\lambda_k\right] = \tilde{\beta}_{ik}E[\lambda_k]$. Plotting average *residual* returns against *residual* betas delivers the same average cross-sectional slope $E[\lambda_k]$ for the line of best fit as we obtain using Fama-MacBeth in the multivariate regressions.

Figure I: Domestic Equity Portfolios (P3) Example: Domestic Carhart, 1963–2016

Figure presents cross-sectional slopes of returns on factor exposure by cluster. The underlying factor model is the Carhart four-factor model, and our sample period is 1963–2016. We select the number of groups as $G^{*} = 2$ using the minimum AIC criterion, where AICs are tabulated in Table V. We vary residual betas on the x-axis because slopes in the multivariate cross-sectional regressions are identical to the slopes on residual betas. Portfolios are color coded by segment and identified with abbreviated portfolio names next to each point. Cluster sizes are in parentheses.



zero. Either real-world market participants suffer large implementation costs in replicating the momentum factor, as Novy-Marx and Velikov (2016) and Patton and Weller (2018) argue, or trading long-short momentum factor portfolios is especially costly in certain parts of the domestic equity universe (as Lesmond, Schill, and Zhou (2004) and others suggest).

We caveat these results with a drawback to our approach: more than half of our portfolios are sorted on the dimension of market capitalization, so our methodology has high resolution to detect differences in compensation among differently-sized stocks. While we use a relatively expansive portfolio set, we are equipped only to detect heterogeneity in pricing along the sorting dimensions analyzed by other researchers, as we take as test assets only portfolios that have previously appeared in the literature. Hence we can only conclude that market capitalization is the most important determinant of cross-sectional variation in risk prices among our stock portfolios. The study of other domestic equity test assets may find dimensions along which risk premia vary even more strongly.

International Equity Portfolios

Table VI and Figure II present Fama-MacBeth regression estimates for the global Carhart model and 200 size-value and size-momentum sorted portfolios. The top panel of Table VI indicates that we require (only) three sets of risk prices to fit the data well, although the AIC is almost identical for four sets of risk prices. From the rightmost column, the greatest differences in risk prices are found for the market and momentum factors. In particular, the third cluster has a strongly negative equity premium and no momentum premium, whereas the first cluster has large and positive equity and momentum risk premia. By contrast with the previous example, between-group factor dynamics are quite different among portfolios: for example, the Group 2's mimicking portfolio returns are more than 60% correlated with the global factors, whereas Group 3's mimicking portfolio returns are only 30%–40% with these factors. Allowing for heterogeneous factor dynamics also contributes to substantial improvements in average cross-sectional R^2s, suggesting that each cluster has strong within-cluster or local factor structures.

The estimated assignments tabulated in Table VI and depicted in Figure II clusters assets perfectly by geographic region: North America and Asia-Pacific excluding Japan (Group 1), Europe (Group 2), and Japan (Group 3). There are no portfolio clusters with portfolios outside of these regional splits.[19] In the bottom-right plot, we see the well-documented failure of momentum in Japan (e.g., Rouwenhorst (1998) and Griffin, Ji, and Martin (2003)) set against the large momentum premium of the other three regions. Likewise, we see a greater value premium in Japan than any other region, consistent with Asness, Moskowitz, and Pedersen (2013)'s argument that a combined value-momentum strategy performs well across all major international regions. Intriguingly only

[19] The four-cluster model, with results summarized in the top panel of Table VI, separates the portfolios exactly into the four geographic regions, again with no exceptions.

Table VI: International Equity Portfolios (P5) Example: Global Carhart, 1991–2016

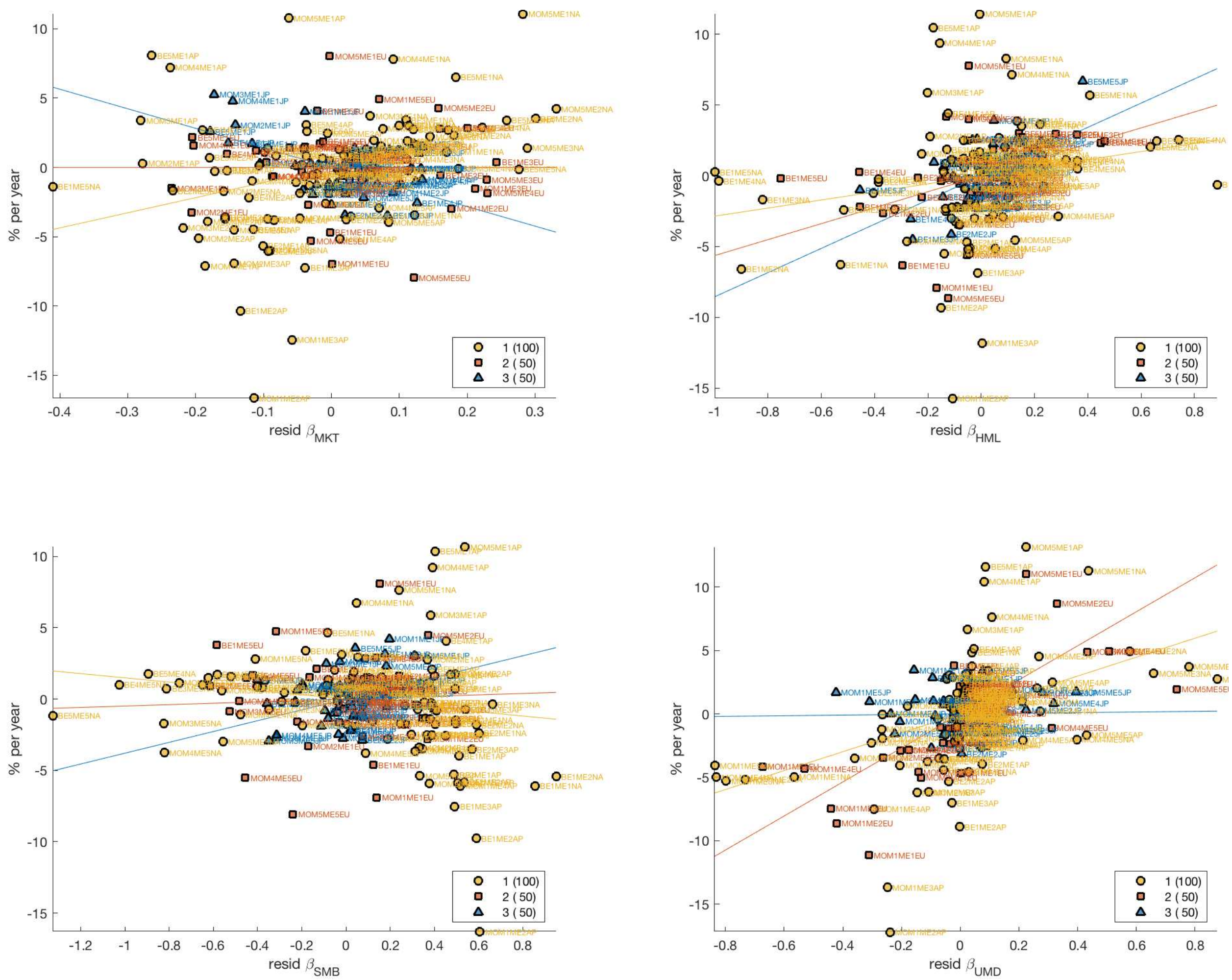
The top table reports p-values from F tests described in Section III.B for comparing a multiple-clusters model to a single-cluster model. The underlying factor model is the Carhart four-factor model with global factors described in Fama and French (2012), and our sample period is 1991–2016. The bottom table reports full-sample Fama-MacBeth estimates of average cross-sectional slopes and associated t-statistics for each group in a model with one cluster ("All") and in a model with $G^* = 3$ clusters selected by the full-sample AIC. Standard errors are Newey-West with 252 daily lags. $\rho_{f,\lambda^{(g)}}$ are the correlations of the factor return and the factor-mimicking portfolio return for cluster g. $p_F\left(\bar{\lambda} =\right)$ is the p-value associated with equality of factor means for the particular factor assuming fixed group memberships. Average cross-sectional R^2s are reported both within each cluster (R^2_G) and across clusters ($R^2_{Combined}$).

Sample	**# Clusters**	**1**	**2**	**3**	**4**	**5**
$\mathcal{P}$	$p_F\left(\bar{\lambda}_k\right)$: 1 vs. G	–	0.000	**0.000**	0.000	0.000
	$p_F\left(\lambda_{kt}\right)$: 1 vs. G	–	0.000	**0.000**	0.000	0.000
Full	LL ($\times 10^{-6}$)	2.223	2.312	**2.339**	2.355	2.355
	AIC ($\times 10^{-6}$)	-4.415	-4.562	**-4.585**	-4.584	-4.554

	Single-Cluster Model	**Three-Cluster Model**			
	All	**Group 1**	**Group 2**	**Group 3**	$p_F\left(\bar{\lambda} =\right)$
$\bar{\alpha}$	3.25	-1.49	2.05	5.35	0.56
t-stat	(0.75)	(-0.23)	(0.60)	(1.24)	
$\bar{\lambda}_{MKT}$	4.78	10.91	0.01	-14.11	0.06
t-stat	(1.32)	(1.79)	(0.00)	(-1.89)	
$\rho_{f,\lambda^{(g)}}$	[0.63]	[0.43]	[0.69]	[0.34]	
$\bar{\lambda}_{HML}$	-1.22	2.85	5.64	8.55	0.13
t-stat	(-0.54)	(1.51)	(1.87)	(1.99)	
$\rho_{f,\lambda^{(g)}}$	[0.46]	[0.83]	[0.62]	[0.33]	
$\bar{\lambda}_{SMB}$	-1.63	-1.49	0.49	3.79	0.49
t-stat	(-0.72)	(-0.60)	(0.24)	(0.82)	
$\rho_{f,\lambda^{(g)}}$	[0.67]	[0.59]	[0.64]	[0.32]	
$\bar{\lambda}_{UMD}$	8.30	7.46	13.42	0.25	0.00
t-stat	(2.81)	(2.64)	(3.69)	(0.04)	
$\rho_{f,\lambda^{(g)}}$	[0.94]	[0.89]	[0.77]	[0.40]	
R^2_G	0.70	0.83	0.94	0.93	
$R^2_{Combined}$	0.70		0.83		
NA	50	50	0	0	
AP	50	50	0	0	
EU	50	0	50	0	
JP	50	0	0	50	
N_G	200	100	50	50	
T	6783	6783	6783	6783	

Figure II: International Equity Portfolios (P5) Example: Global Carhart, 1991–2016

Figure presents cross-sectional slopes of returns on factor exposure by cluster. The underlying factor model is the Carhart four-factor model with global factors described in Fama and French (2012), and our sample period is 1991–2016. We select the number of groups as $G^* = 3$ using the minimum AIC criterion, where AICs are tabulated in Table VI. We vary residual betas on the x-axis because slopes in the multivariate cross-sectional regressions are identical to the slopes on residual betas. Portfolios are color coded by segment and identified with abbreviated portfolio names next to each point. Cluster sizes are in parentheses.



Group 1 securities earn meaningfully positive compensation to the global market factor.

This analysis inverts the standard asset pricing paradigm of selecting regions and comparing risk premia or mean-variance efficient portfolios, as in Griffin (2002), Hou, Karolyi, and Kho (2011), Fama and French (2012). Instead, we confirm that the cross-section of portfolio returns itself suffices to identify the region to which assets belong, although our analysis introduces a surprising twist in that we find developed Pacific Rim markets to be integrated with the US and Canada. That our estimated group boundaries coincide with geographic ones, taken by others previously as given because of institutional barriers to arbitrage, serves as reassurance that our methodology will detect important sources of segmentation, and encourages us to apply it in settings for which the critical dimensions of segmentation are not known ex ante.

Cross-Asset Class Portfolios

Table VII and Figure III report results from our clustering methodology applied to the He, Kelly, and Manela (2017) market-intermediary capital factor model with 148 cross-asset class portfolios (P7). In conducting this exercise we use the authors' monthly data and a similar set of test assets. However, rather than finding support for unified factor pricing, we instead find (at least) five clusters in the data, as selected by the AIC. We stop at five clusters because the "no small clusters" assumption is already strained by groups of size 20.

Estimates of the equity premium range from -2.4% per year in cluster Group 4 to 45.9% per year in cluster Group 1, with a full-sample average equity premium of 7.14% per year. The intermediary capital factor price varies from -48.4% per year in Group 1 to 22.8% per year in Group 2. This large variation is not due to the clusters being too small to estimate λs well—our smallest group has 20 portfolios, and our most extreme risk prices come from clusters of 20 and 30 portfolios (by comparison five of the eight asset classes analyzed individually in He, Kelly, and Manela (2017) have 20 or fewer portfolios). An test for equality of average risk premia across clusters rejects equality for both factors with p-values of 0.00 and 0.06.

Table VII also reveals important differences in the dynamics of risk premia across clusters, and these differences in dynamics are missed by the test of differences in average risk prices. While the market factor is highly correlated with market factor-mimicking portfolios for each cluster other than the first (similar to the international equities portfolio example), the intermediary capital factor looks very different from its mimicking portfolios. Three of the five clusters' local variants are less than 50% correlated with the global factor. This finding reinforces the main point of Haddad and Muir (2018), who discover important cross-asset class differences in the time-variation in risk premia as a function of barriers to direct participation by households (equivalently in their model, the degree of intermediation).

As Figure III illustrates, our algorithm suggests that the degree of reliance on financial intermediaries may not be the most important determinant of risk-premium dynamics across asset classes.

Table VII: Cross-Asset Class Portfolios (P7) Example: He, Kelly, and Manela (2017) Factors, 1986–2010

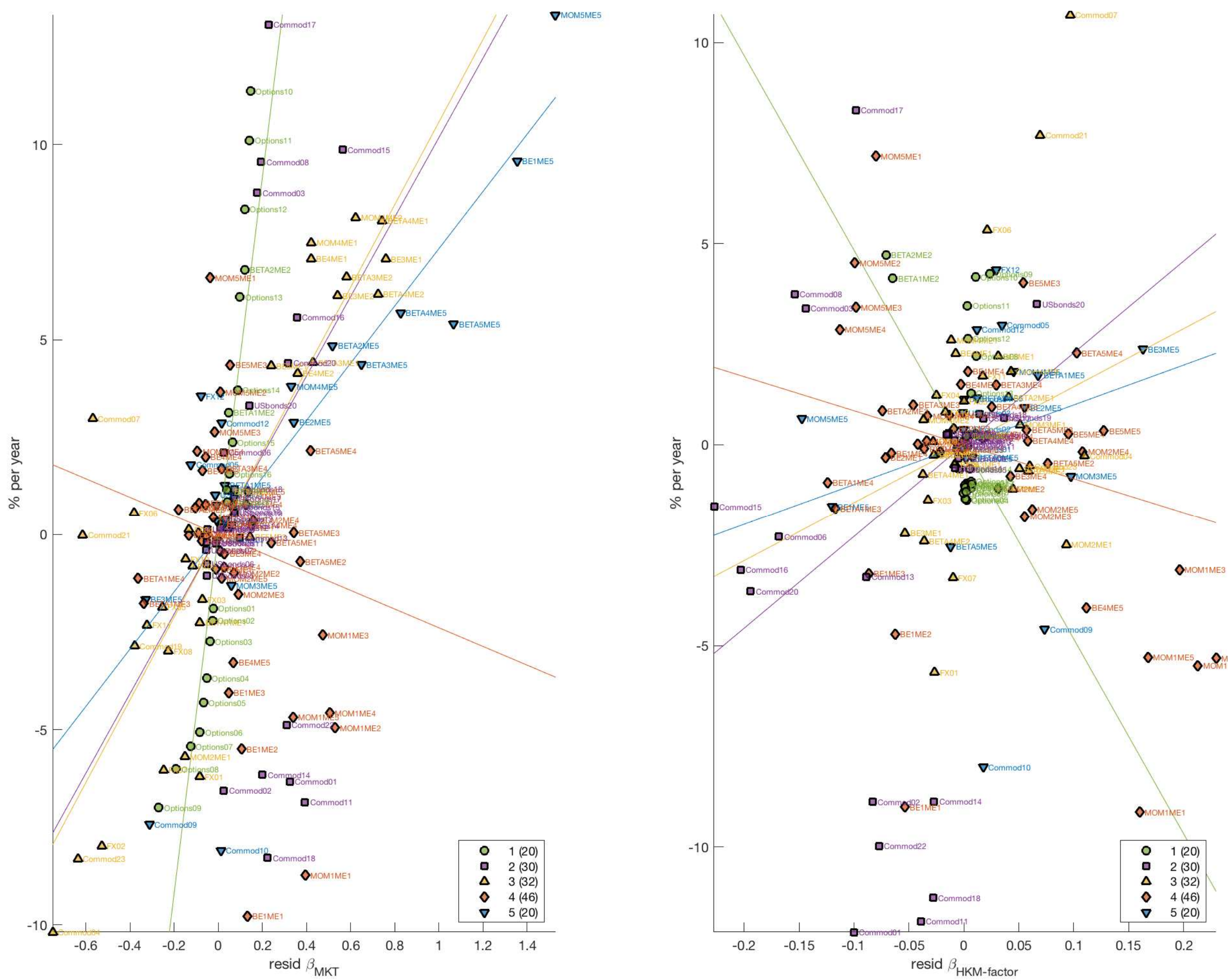
The top table reports p-values from F tests described in Section III.B for comparing a multiple-clusters model to a single-cluster model. The underlying factor model is the He-Kelly-Manela two-factor model with the intermediary capital factor from He, Kelly, and Manela (2017), and our sample period is 1986–2010. The bottom table reports full-sample Fama-MacBeth estimates of average cross-sectional slopes and associated t-statistics for each group in a model with one cluster ("All") and in a model with $G^* = 5$ clusters selected by the full-sample AIC. Standard errors are Newey-West with 12 monthly lags. $\rho_{f,\lambda^{(g)}}$ are the correlations of the factor return and the factor-mimicking portfolio return for cluster g. $p_F\left(\bar{\lambda} =\right)$ is the p-value associated with equality of factor means for the particular factor assuming fixed group memberships. Average cross-sectional R^2s are reported both within each cluster (R^2_G) and across clusters ($R^2_{Combined}$).

Sample	**# clusters**	**1**	**2**	**3**	**4**	**5**
$\mathcal{P}$	$p_F\left(\bar{\lambda}_k\right)$: 1 vs. G	–	0.000	0.000	0.000	**0.000**
	$p_F\left(\lambda_{kt}\right)$: 1 vs. G	–	0.000	0.000	0.000	**0.000**
Full	LL ($\times 10^{-4}$)	5.194	5.463	5.546	5.614	**5.661**
	AIC ($\times 10^{-4}$)	-10.302	-10.754	-10.834	-10.881	**-10.893**

	Single-Cluster Model	**Five-Cluster Model**					
	All	**Grp 1**	**Grp 2**	**Grp 3**	**Grp 4**	**Grp 5**	$p_F\left(\bar{\lambda} =\right)$
$\bar{\alpha}$	0.62	-31.05	2.67	-0.13	11.42	0.54	0.00
t-stat	(6.41)	(-4.53)	(4.02)	(-0.05)	(2.76)	(6.81)	
$\bar{\lambda}_{MKT}$	7.14	45.85	10.18	10.57	-2.39	7.33	0.00
t-stat	(2.22)	(4.77)	(1.30)	(2.53)	(-0.48)	(2.10)	
$\rho_{f,\lambda^{(g)}}$	[0.98]	[0.33]	[0.55]	[0.75]	[0.66]	[0.98]	
$\bar{\lambda}_{HKM}$	9.30	-48.38	22.84	14.43	-8.47	9.91	0.06
t-stat	(1.18)	(-1.34)	(1.75)	(1.15)	(-0.87)	(1.14)	
$\rho_{f,\lambda^{(g)}}$	[0.62]	[0.16]	[0.36]	[0.45]	[0.51]	[0.56]	
R^2_G	0.74	0.98	0.58	0.85	0.91	0.84	
$R^2_{Combined}$	0.74			0.88			
Options	18	18	0	0	0	0	
Commod.	23	0	14	5	0	4	
US Bonds	20	0	16	0	0	4	
FX	12	0	0	11	0	1	
Stocks	75	2	0	16	46	11	
N_G	148	20	30	32	46	20	
T	300	300	300	300	300	300	

Figure III: Cross-Asset Class Portfolios (P7) Example: He, Kelly, and Manela (2017) Factors, 1986–2010

Figure presents cross-sectional slopes of returns on factor exposure by cluster. The underlying factor model is the He-Kelly-Manela two-factor model with the intermediary capital factor from He, Kelly, and Manela (2017), and our sample period is 1986–2010. We select the number of groups as $G^{*} = 5$ using the minimum AIC criterion, where AICs are tabulated in Table VII. We vary residual betas on the x-axis because slopes in the multivariate cross-sectional regressions are identical to the slopes on residual betas. Portfolios are color coded by segment and identified with abbreviated portfolio names next to each point. Cluster sizes are in parentheses.



Although the groups do not cleave quite as cleanly as in the international equity portfolios example, the green cluster (Group 1) consists almost exclusively of US options; the purple cluster (Group 2) is evenly split between commodities and US bonds; the yellow cluster (Group 3) includes all but one foreign exchange portfolio, 16 small-cap stock portfolios, and five commodities; the red cluster (Group 4) is comprised exclusively of US stock portfolios; and the blue cluster (Group 5) has 11 large-cap stock portfolios, and the remaining commodity, US bond, and foreign exchange portfolios. Some of these splits are readily conjectured ex ante, for example, options and stocks look different from bonds and commodities, and some of these splits are not, for example, the mixing of commodities and US bonds. Our approach is uniquely positioned to find such unconventional partitions of the data by risk prices. That the assets are estimated to separate mostly back into asset class groups suggests that either the HKM measure does not capture intermediaries' pricing kernel, or that intermediary asset pricing fails to unify risk prices.

VI. Omitted Factors or Fundamental Heterogeneity?

A. Clusters as Factors and Factors as Clusters

In this section we consider the possibility of omitted factors manifesting as differences in risk prices and vice-versa.[20] To start, consider two simple models:

$$\text{Model 1: } r_{it} = \alpha_t^{(1)} \mathbf{1}(\gamma_i = 1) + \alpha_t^{(2)} \mathbf{1}(\gamma_i = 2) + \epsilon_{it}, \tag{13}$$

$$\text{Model 2: } r_{it} = \tilde{\alpha}_t + \beta_i \eta_t + \tilde{\epsilon}_{it}. \tag{14}$$

The first model consists of two clusters, with cluster memberships determined by $\gamma_i \in \{1, 2\}$, each with time-varying average returns, $\alpha_t^{(g)}$, within the cluster. The second model consists of a single cluster with time-varying average returns and factor realizations as well as heterogeneous risk exposures β_i. Note that β_i can be a time series beta or a characteristic. To complete our notation, let N_i be the number of assets in cluster i, and define $\Delta\alpha_t \equiv \alpha_t^{(1)} - \alpha_t^{(2)}$.

Clusters as Factors Suppose that the true data-generating process (DGP) is (13), but we instead estimate (14). To simplify this case, assume $\epsilon_{it} \perp \beta_i$ at each date. Within each cross-section, the OLS estimate for η_t is

$$\begin{aligned} \hat{\eta}_t = \frac{cov(r_{it}, \beta_i)}{var(\beta_i)} &= \frac{\left(\alpha_t^{(1)} - \alpha_t^{(2)}\right) cov(\mathbf{1}(\gamma_i = 1), \beta_i) + cov(\epsilon_{it}, \beta_i)}{var(\beta_i)} \\ &= \Delta\alpha_t \frac{var(\mathbf{1}(\gamma_i = 1))}{var(\beta_i)} \left(E[\beta_i | \gamma_i = 1] - E[\beta_i | \gamma_i = 2]\right). \end{aligned} \tag{15}$$

[20] An in-depth analysis of clusters as a source of the proliferation of factors in the cross-section of expected returns is the subject of ongoing research.

The time series average of $\hat{\eta}_t$ replaces $\Delta\alpha_t$ with $\Delta\bar{\alpha}$. If average returns in each cluster are different, i.e. $\bar{\alpha}^{(1)} \neq \bar{\alpha}^{(2)}$, then $\bar{\eta} \neq 0$, and *any* characteristic that varies on average across groups will appear to be priced. This expression is readily extended to a multivariate context by invoking the Frisch–Waugh–Lovell theorem, and it extends to more than two clusters by replacing (15) with a $G-1$ set of indicators corresponding to membership in each cluster other than the last one.

Factors as Clusters Now suppose that the true DGP is (14), but we instead estimate (13). We assume $E\left[\tilde{\epsilon}_{it}|i \in G_1\right] = 0$ to keep the exposition as simple as possible. Within each cross-section, the OLS estimate for $\Delta\alpha_t$ is

$$\begin{aligned}\widehat{\Delta\alpha_t} &= \frac{cov\left(r_{it}, \mathbf{1}\left(\gamma_i = 1\right)\right)}{var\left(\mathbf{1}\left(\gamma_i = 1\right)\right)} = \frac{cov\left(\beta_i \eta_t, \mathbf{1}\left(\gamma_i = 1\right)\right) + cov\left(\tilde{\epsilon}_{it}, \mathbf{1}\left(\gamma_i = 1\right)\right)}{var\left(\mathbf{1}\left(\gamma_i = 1\right)\right)} \\ &= \eta_t\left(E\left[\beta_i|i \in G_1\right] - E\left[\beta_i|i \in G_2\right]\right). \end{aligned} \tag{16}$$

Hence, so long as (a) the factor exposure or characteristic β_i has different conditional means across clusters, and (b) the factor realization is not precisely zero, the difference in cross-sectional means $\widehat{\Delta\alpha_t}$ is also nonzero. Moreover, if the factor is priced with $\bar{\eta} \neq 0$, then the difference in average returns across clusters, $\overline{\Delta\alpha}$, will appear to be nonzero. By parallel with the "clusters as factors" case, this expression can also be extended to a multifactor model with one or more omitted factors.

B. Economic Restrictions of a Segmented-Markets Model

To link the stylized example in the previous subsection to our empirical analysis below, we now show that a K-factor asset pricing model with G clusters is a special case of a larger, single-cluster, asset pricing model with an additional $(G-1)\times K$ factors, with factor loadings that are constrained to be zero in specific cases. To show this equivalence, we combine groups in (1) as

$$\begin{aligned} r_{it} &= \left(\bar{\alpha}_t + \alpha_t^{(1)}\right) + \sum_k \beta_{ik}\left(f_{kt} + \phi_{kt}^{(1)}\right) \\ &+ \sum_{g>1}\left(\alpha_t^{(g)} - \left(\bar{\alpha}_t + \alpha_t^{(1)}\right)\right) 1\left(\gamma_i = g\right) + \sum_{g>1}\sum_k \beta_{ik} 1\left(\gamma_i = g\right)\phi_{kt}^{(g)} + \epsilon_{it}. \end{aligned} \tag{17}$$

The first line of (17) resembles a standard realized return model for assets in an integrated market. Market segmentation adds group-specific zero-beta rates $\alpha_t^{(g)}$ as well as group-specific factors $\phi_t^{(g)}$. The coefficients on the group-specific factors take one of two values. For assets in segment g, the group-specific factor loadings β_{ik} are the same as those on the corresponding global factors. For assets in other segments, these loadings are zero. This structure on local factors also affects the output from our cross-sectional regressions. We cannot estimate a $G \times K$-factor version of (17) if markets are segmented because f and Φ together have rank $(G+1)\times K$ whereas the betas have rank $G \times K$.

The factor-mimicking portfolio interpretation of Fama-MacBeth cross-sectional slopes helps to clarify differences between segmented-market and extended-factor models.[21] Equation (5) delivers factor-mimicking portfolio returns group by group, that is, the cross-sectional slopes are each segment's approximation of the global factor return given the assets in that segment. In the international example of Section V.C, these factor-mimicking portfolios are the approximations to global value, momentum, etc., using each region's size-value and size-momentum portfolios. Imposing the redundant-or-zero structure on betas in (17) maintains this feature. By contrast, in global models with unrestricted betas on all factors, the cross-sectional slopes take on a different interpretation. The second-stage slopes on β_{ik} represent the mimicking portfolio return using all assets, regardless of market segment, and zeroing out the local components.

C. Empirically Distinguishing Between Factors and Clusters

As the preceding examples make clear, it is challenging to distinguish between omitted factors and multiple clusters without imposing structure on what omitted factors might look like and how numerous they might be. Indeed, as shown in the previous section, factor models of arbitrary length nest cluster-based models as a special case. For this reason, we compare omitted-factor and cluster-based models of comparable size (defined in a variety of ways).

Rivers and Vuong (2002) provide our framework for model comparison. Specifically, for two non-nested models that minimize in-sample (weighted) squared errors, e.g., using linear regression, Rivers and Vuong (2002) demonstrate that the models can be compared using their (weighted) cross-sectional mean squared errors (MSE) date-by-date. Under the null that the models M_1 and M_2 are equally accurate, they show that,

$$\frac{1}{\sqrt{T}} \sum_t \left(MSE_t^{(M_1)} - MSE_t^{(M_2)} \right) \sim N\left(0, V\right), \tag{18}$$

where V is the asymptotic variance of the difference in MSEs, which is easily computed using a HAC estimator, e.g. Newey and West (1987). Hence given a particular choice of cluster and factor model, we can use a simple t test on the difference in (weighted) MSEs to evaluate the null of equal fit against the alternative of unequal fit. As in our estimation, we use idiosyncratic variance as weights in the MSE.

The challenge is how to choose the cluster and factor models to compare. For this purpose, we retain the $\mathcal{R}$ and $\mathcal{P}$ partitions described in Section III.B. We first use the in-sample AIC on the $\mathcal{R}$ subsample to select the "best" cluster model and fix group assignments. We then estimate this model on the $\mathcal{P}$ subsample, using the group assignments from the $\mathcal{R}$ sample. Next, we retain the time-series regression estimates of factor betas and residual variation, and we extract additional factors from these residuals on $\mathcal{R}$ using principal components analysis. These additional factors

[21] We thank Stefano Giglio for this insight.

maximize explanatory power of the variation in returns not spanned by the included factors, and they serve as our candidate omitted factors; in this respect our methodology mimics Giglio and Xiu (2017)'s approach for extracting omitted factors that may bias second-stage estimates of risk premia. We retain factor loadings (βs) estimated on the $\mathcal{R}$ subsample to fix factor identities and employ them in the cross-sectional regressions on the $\mathcal{P}$ subsample.

Because we wish to speak to a large range of possible omitted-factor models, we consider several choices for the "best" omitted-factor model rather than a single one. We enumerate these models on the basis on the number of additional included principal components, K_1^*, K_2^* or K_3^*. In the K_1^* models, we consider factor models augmented with the first three principal components of the panel of residual returns on the $\mathcal{R}$ partition. Our choice of three PCs is ad hoc and intended only to provide a uniform omitted-factor benchmark. In the K_2^* models, we consider $G^* - 1$ additional factors, where G^* is the number of clusters selected by the AIC on $\mathcal{R}$. We make this choice of the number of additional factors to give both cluster and factor models similar flexibility in fitting the data. A model with G^* clusters adds $G^* - 1$ partitions of the data on which to fit K-factor models, and a model with a single cluster and $K + K_2^*$ factors also has an $G^* - 1$ additional degrees of freedom to fit each cross section. The K_3^* models give maximum flexibility to extended-factor models to fit the data; just as we choose the number of clusters using the in-sample AIC, so too do we let the AIC on the $\mathcal{R}$ partition dictate the number of factors. We impose one restriction to make the selection process comparable, that is, much as we do not allow the AIC for the cluster models to select $G^* > 5$, we do not let the number of additional factors selected exceed $(G^* - 1)(K + 1) - 1$. This upper bound is one less than the number of additional variables that allow a factor model to perfectly replicate a cluster model. In settings with large N and large T, factor models with so much additional flexibility cannot "lose" to cluster models, making the model comparison uninformative and violating the non-nestedness condition of Rivers and Vuong (2002).

Equipped with models chosen on the $\mathcal{R}$ partition, Table VIII reports discretized t statistics for the hypothesis of equal mean squared errors on $\mathcal{P}$ for factor models augmented with additional clusters and our three choices of additional factors. We code p-values below 0.1, 0.05, and 0.01 with one, two, or three – or +, respectively. Positive entries signify that the cluster model performs better, and negative entries signify that the extended-factor model performs better. We use Newey-West standard errors with 252 daily or 12 monthly lags to allow for serial dependence in squared errors.

Focusing first on the domestic equity portfolios, we obtain mixed results on the importance of clusters versus factors depending on the breadth of the portfolio set considered. For the narrowest portfolio set, P1, model comparisons are roughly split between favoring multiple-cluster models and multiple-factor models. This result parallels Table IV in that we again find that this portfolio set is too limited in variety for risk price heterogeneity to be the dominant consideration for explaining variation in returns, and this feature is reinforced by the fact that the $\mathcal{R}$-partition AIC in some

Table VIII: Comparison of Multiple-Cluster and Omitted-Factor Models

Table reports discretized t-statistics for the average time-series differences in mean-squared errors between omitted-factor and multiple-cluster models. We code p-values below 0.1, 0.05, and 0.01 with one, two, or three − or +, respectively. Positive (negative) values indicate support for the multiple-cluster (omitted-factor) models. To estimate these quantities, we first estimate group assignments, risk prices, and likelihoods for models with $G = 1, \ldots, 5$ groups on the $\mathcal{R}$ partition (the first half of the sample). We then select the number of groups G^* using the AIC on the $\mathcal{R}$ sample. '*' indicate instances in which the AIC selects a single cluster. We compare this multiple-cluster model with extended-factor models with additional factors extracted using principal components on the $N \times R$ panel of time-series residuals from the conjectured factor model on the $\mathcal{R}$ partition. These models augment the conjectured model with $K_1^* = 3$, $K_2^* = G^* - 1$, and $K_3^* = \min\left(\arg\min_K AIC_K, (G^* - 1)(K+1) - 1\right)$ principal components. By parallel with the selection of the number of clusters, we fix the extracted principal component loadings and select models using the AIC on the $\mathcal{R}$ partition. t statistics compare models across dates in the $\mathcal{P}$ partition, and standard errors are Newey-West with 252 daily or 12 monthly lags. We then repeat this procedure for all combinations of portfolio sets, risk models, and sample periods. Portfolios and models are described in the text. For the He, Kelly, and Manela (2017) factor model, we use daily data for the most recent time period and monthly data for earlier time periods. We do not have sufficient coverage for their intermediary capital factor for 1963–1980 to include it in the domestic equity portfolio analysis. The q-factor (HXZQ) model is excluded from the international portfolio analysis because we do not have global return-on-equity factor data.

(a) Domestic Equity Portfolios

		1963–2016			**1963–1980**			**1981–1998**			**1999–2016**		
Portfolios	Model	K_1^*	K_2^*	K_3^*	K_1^*	K_2^*	K_3^*	K_1^*	K_2^*	K_3^*	K_1^*	K_2^*	K_3^*
P1	CAPM	−−	0	+++	−−−	−−−	−−−	−	+++	+++	−−	0	−−
	FF3F	−−−	0	0	0	+++	+++	0	+++	+++	−−	0	−−
	Carhart	+++	+++	+++	*	*	*	+++	+++	+++	++	+++	+++
	FF5F	*	*	*	*	*	*	+++	+++	+++	−	+++	0
	HKM	−−−	0	0				−−−	0	0	++	+++	+++
	HXZQ	+++	+++	+++	*	*	*	*	*	*	*	*	*
P2	CAPM	−−−	−−−	0	−−−	−−−	−	−−−	−−−	−−−	−−	−−	−−−
	FF3F	−−−	0	−−−	0	+++	+++	0	+++	+++	0	+++	−−
	Carhart	++	+++	+++	+++	+++	+++	0	+++	0	−−−	+++	−−−
	FF5F	−	+++	+++	+++	+++	+++	0	+++	0	−−	+++	0
	HKM	−−−	0	0				0	+++	+++	0	++	−−−
	HXZQ	0	+++	+++	+++	+++	+++	+++	+++	+++	0	+++	0
P3	CAPM	+++	+++	+++	+++	+++	−−−	+++	+++	+++	0	−−	0
	FF3F	+++	+++	+++	+++	+++	+++	+++	+++	+++	0	+++	−−−
	Carhart	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	−−
	FF5F	+++	+++	0	+++	+++	+++	+++	+++	+++	0	+++	−−
	HKM	0	+++	+++				+++	+++	+++	+++	+++	−−−
	HXZQ	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	0

Table VIII: Comparison of Multiple-Cluster and Omitted-Factor Models (Continued)

(c) International Equity Portfolios

		1991–2016			1991–2003			2004–2016		
Portfolios	Model	K_1^*	K_2^*	K_3^*	K_1^*	K_2^*	K_3^*	K_1^*	K_2^*	K_3^*
P4	CAPM	+++	+++	+++	+++	+++	+++	0	+++	0
	FF3F	+++	+++	+++	+	+++	+++	−−−	+++	−−−
	Carhart	−−−	+++	−−−	+++	+++	+++	−−−	+++	−−−
	FF5F	+++	+++	+++	+++	+++	+++	+++	+++	+++
	HKM	+++	+++	+++	+++	+++	+++	+++	+++	+++
	HXZQ									
P5	CAPM	+++	+++	+++	+++	+++	+++	+++	+++	+++
	FF3F	+++	+++	0	+++	+++	+++	+++	+++	+++
	Carhart	+++	+++	+++	+++	+++	+++	+++	+++	+++
	FF5F	0	+++	0	+++	+++	+++	+++	+++	+++
	HKM	+++	+++	+++	+++	+++	0	+++	+++	+++
	HXZQ									

(d) Cross–Asset Class Portfolios

		1986–2010			1986–1997			1998–2010		
Portfolios	Model	K_1^*	K_2^*	K_3^*	K_1^*	K_2^*	K_3^*	K_1^*	K_2^*	K_3^*
P6	CAPM	+++	+++	+++	++	++	++	+++	+++	+++
	FF3F	+++	+++	+++	++	++	++	+++	+++	+++
	Carhart	+++	+++	+++	+	+	+	+++	+++	+++
	FF5F	+++	+++	+++	+	++	0	+++	+++	++
	HKM	+++	+++	+++	++	++	++	+++	+++	+++
	HXZQ	+++	+++	+++	+	++	+	+++	+++	+++
P7	CAPM	+++	+++	++	++	+	0	+++	+++	0
	FF3F	+++	+++	0	+	++	0	+++	+++	+++
	Carhart	+++	+++	+++	++	++	0	+++	+++	+++
	FF5F	+++	+++	+++	++	++	+	+++	+++	0
	HKM	+++	+++	+++	++	++	+	+++	+++	+++
	HXZQ	+++	+++	+++	++	++	+	+++	+++	+++

cases only chooses a single cluster. Progressing to P2 and P3, for which domestic equity portfolios become more diverse, so too does the importance of clusters relative to omitted factors increase: both portfolio sets point strongly to clusters as more important than missing factors for explaining variation in returns. There are two exceptions to this finding. First, for the CAPM rows, we see in both cases that omitted factors are still more important than risk price heterogeneity for P2. This result is unsurprising given the CAPM's limited ability to explain cross-sectional variation in expected or realized returns—clearly other factors are missing, and risk price heterogeneity on the market is less vital than for other factors. Second, our tests favor omitted-factor models relative to multiple-cluster models with K_3^* additional factors in the most recent time period. We interpret this finding as evidence of increased capital-market integration across US stocks, perhaps associated with the sharp decline in transactions costs during this time. Nevertheless, risk price heterogeneity still dominates omitted-factor explanations with comparably sized models (K_2^*) and for models with only a small number of omitted factors (K_1^*).

In contrast with the domestic equity portfolios, the international equity and cross-asset class settings unambiguously favor heterogeneous risk prices over richer factor models. This feature echoes the strong rejections of equal risk prices for P4–P7 documented in Table III. We conclude that any analysis of the cross-section of expected returns in such settings requires consideration of segmentation across asset classes, notwithstanding recent evidence of integrated intermediation across markets or reduced barriers to trade across global regions.

One notable feature shared with the domestic equity portfolios is that more diverse portfolios more strongly favor risk price heterogeneity over missing factors. Nothing in our analysis mechanically generates this result, and broader portfolio sets could well have required additional factors to explain other cuts of the investible universe (consider, for example, the additional factors and sorted portfolios of Fama and French (2015)). Rather, more exotic corners of the market or nonstandard portfolio formations illuminate pricing discrepancies even among the factors included in relatively parsimonious models. As empirical asset pricing continues to examine increasingly diverse portfolio sets in response to greater data availability and data-mining concerns, we anticipate heterogeneous risk pricing models to become commensurately more important.

VII. Conclusion

We present new methods for detecting and estimating heterogeneous risk prices in a cross-section of assets. Our approach marries traditional asset pricing methods for risk price estimation and machine learning methods for clustering data. Using this methodology, we find statistically significant and economically important evidence of market segmentation across all portfolios, factor pricing models, and time periods. Arbitrage frictions matter universally, not just in highly specialized assets or during crisis periods: even within low-friction, high-participation markets, we find that compensation per unit risk varies significantly across assets.

Segmented risk prices challenge leading models of risk and return. At best, segmentation implies that these factor models are incomplete and miss important cross-sectional variation in expected returns. However, our contribution is not simply another attack on commonly used factor models in finance. Rather, our findings give fresh motivation to consider limits to arbitrage in security markets. We offer a structured alternative for how, not just whether, frictionless factor models might be improved upon in empirical applications. In doing so we suggest a promising new direction for the study of the cross-section of expected returns.

Our findings have important practical implications. Given the ability to invest across groups of assets, sophisticated investors should direct their capital to markets with the highest compensation per unit risk. We offer concrete guidance for identifying these groups of assets. Likewise, potential arbitrageurs across segments can earn the difference between risk prices net of implementation costs. While such long-short strategies are not true arbitrage opportunities—there are too few market segments to be well-diversified, and local factors are imperfectly correlated across segments—they nonetheless represent "good deals" in the sense of Cochrane and Saá-Requejo (2000) and contribute to substantial improvements in ex post Sharpe ratios.

Risk price heterogeneity also provides a novel explanation for the "factor zoo" of Cochrane (2011) and Harvey, Liu, and Zhu (2016). The examples of Section VI.A indicate that any factor whose loadings align with different segments may appear to be priced. Our analysis suggests that missing clusters are more important than omitted factors for most combinations of factor models and portfolio sets, with the possible exception of US equities in the most recent period. The natural follow-up question, and the subject of ongoing work, is the extent to which so-called expected return factors in US stocks are instead proxies for membership in market segments with different risk prices.

References

Adrian, Tobias, Etula Erkko, and Tyler Muir. 2014. "Financial Intermediaries and the Cross-Section of Asset Returns." *The Journal of Finance* 69 (6):2557–2596.

Arthur, David and Sergei Vassilvitskii. 2007. "K-means++: The Advantages of Careful Seeding." In *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '07. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 1027–1035.

Asness, Clifford S., Tobias J. Moskowitz, and Lasse Heje Pedersen. 2013. "Value and Momentum Everywhere." *Journal of Finance* 68 (3):929–985.

Bekaert, Geert and Campbell R. Harvey. 1995. "Time-Varying World Market Integration." *Journal of Finance* 50 (2):403–444.

Bekaert, Geert, Robert J. Hodrick, and Xiaoyan Zhang. 2009. "International Stock Return Comovements." *Journal of Finance* 64 (6):2591–2626.

Bollerslev, Tim. 1990. "Modelling the Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized Arch Model." *The Review of Economics and Statistics* 72 (3):498–505.

Bonhomme, Stéphane and Elena Manresa. 2015. "Grouped Patterns of Heterogeneity in Panel Data." *Econometrica* 83 (3):1147–1184.

Brunnermeier, Markus K. and Lasse Heje Pedersen. 2009. "Market Liquidity and Funding Liquidity." *The Review of Financial Studies* 22 (6):2201–2238.

Carhart, Mark M. 1997. "On Persistence in Mutual Fund Performance." *Journal of Finance* 52 (1):57–82.

Cochrane, John H. 2011. "Presidential Address: Discount Rates." *Journal of Finance* 66 (4):1047–1108.

Cochrane, John H. and Jesús Saá-Requejo. 2000. "Beyond Arbitrage: Good-Deal Asset Price Bounds in Incomplete Markets." *Journal of Political Economy* 108 (1):79–119.

Constantinides, George M., Jens Carsten Jackwerth, and Alexi Savov. 2013. "The Puzzle of Index Option Returns." *Review of Asset Pricing Studies* 3 (2):229–257.

Deb, Kalyanmoy. 2000. "An Efficient Constraint Handling Method for Genetic Algorithms." *Computer Methods in Applied Mechanics and Engineering* 186 (2):311 – 338.

Deep, Kusum, Krishna Pratap Singh, M.L. Kansal, and C. Mohan. 2009. "A Real Coded Genetic Algorithm for Solving Integer and Mixed Integer Optimization Problems." *Applied Mathematics and Computation* 212 (2):505 – 518.

Dempster, A. P., N. M. Laird, and D. B. Rubin. 1977. “Maximum Likelihood from Incomplete Data via the EM Algorithm.” *Journal of the Royal Statistical Society. Series B (Methodological)* 39 (1):1–38.

Errunza, Vihang and Etienne Losq. 1985. “International Asset Pricing under Mild Segmentation: Theory and Test.” *Journal of Finance* 40 (1):105–124.

Fama, Eugene F. and K. R. French. 2010. “Luck Versus Skill in the Cross-Section of Mutual Fund Returns.” *Journal of Finance* 65 (5):1915–1947.

Fama, Eugene F. and Kenneth R. French. 1992. “The Cross-Section of Expected Stock Returns.” *Journal of Finance* 47 (2):427–465.

———. 1993. “Common Risk Factors in the Returns on Stocks and Bonds.” *Journal of Financial Economics* 33 (1):3 – 56.

———. 1998. “Value Versus Growth: The International Evidence.” *Journal of Finance* 53 (6):1975–1999.

———. 2012. “Size, Value, and Momentum in International Stock Returns.” *Journal of Financial Economics* 105 (3):457 – 472.

———. 2015. “A Five-Factor Asset Pricing Model.” *Journal of Financial Economics* 116 (1):1 – 22.

Fama, Eugene F. and James D. MacBeth. 1973. “Risk, Return, and Equilibrium: Empirical Tests.” *Journal of Political Economy* 81 (3):607–636.

Foerster, Stephen R. and G. Andrew Karolyi. 1999. “The Effects of Market Segmentation and Investor Recognition on Asset Prices: Evidence from Foreign Stocks Listing in the United States.” *Journal of Finance* 54 (3):981–1013.

Frazzini, Andrea and Lasse Heje Pedersen. 2014. “Betting Against Beta.” *Journal of Financial Economics* 111 (1):1 – 25.

Gârleanu, Nicolae and Lasse Heje Pedersen. 2011. “Margin-based Asset Pricing and Deviations from the Law of One Price.” *The Review of Financial Studies* 24 (6):1980–2022.

Gibbons, Michael R., Stephen A. Ross, and Jay Shanken. 1989. “A Test of the Efficiency of a Given Portfolio.” *Econometrica* 57 (5):1121–1152.

Giglio, Stefano and Dacheng Xiu. 2017. “Inference on Risk Premia in the Presence of Omitted Factors.” Working paper.

Greenwald, Daniel L., Martin Lettau, and Sydney C. Ludvigson. 2016. "Origins of Stock Market Fluctuations." Working paper.

Griffin, John M. 2002. "Are the Fama and French Factors Global or Country Specific?" *Review of Financial Studies* 15 (3):783–803.

Griffin, John M., Xiuqing Ji, and J. Spencer Martin. 2003. "Momentum Investing and Business Cycle Risk: Evidence from Pole to Pole." *Journal of Finance* 58 (6):2515–2547.

Grinblatt, Mark and Tobias J. Moskowitz. 2004. "Predicting Stock Price Movements from Past Returns: The Role of Consistency and Tax-Loss Selling." *Journal of Financial Economics* 71 (3):541 – 579.

Gromb, Denis and Dimitri Vayanos. 2002. "Equilibrium and welfare in markets with financially constrained arbitrageurs." *Journal of Financial Economics* 66 (2):361 – 407.

———. 2018. "The Dynamics of Financially Constrained Arbitrage." *The Journal of Finance* .

Haddad, Valentin and Tyler Muir. 2018. "Do Intermediaries Matter for Aggregate Asset Prices?" Working paper.

Harvey, Campbell R. 2017. "Presidential Address: The Scientific Outlook in Financial Economics." *Journal of Finance* 72 (4):1399–1440.

Harvey, Campbell R., Yan Liu, and Heqing Zhu. 2016. "... and the Cross-Section of Expected Returns." *Review of Financial Studies* 29 (1):5–68.

He, Zhiguo, Bryan Kelly, and Asaf Manela. 2017. "Intermediary Asset Pricing: New Evidence from Many Asset Classes." *Journal of Financial Economics* 126 (1):1 – 35.

He, Zhiguo and Arvind Krishnamurthy. 2012. "A Model of Capital and Crises." *The Review of Economic Studies* 79 (2):735–777.

———. 2013. "Intermediary Asset Pricing." *American Economic Review* 103 (2):732–70.

Hong, Harrison, Terence Lim, and Jeremy C. Stein. 2000. "Bad News Travels Slowly: Size, Analyst Coverage, and the Profitability of Momentum Strategies." *Journal of Finance* 55 (1):265–295.

Hou, Kewei, G. Andrew Karolyi, and Bong-Chan Kho. 2011. "What Factors Drive Global Stock Returns?" *Review of Financial Studies* 24 (8):2527–2574.

Hou, Kewei, Chen Xue, and Lu Zhang. 2015. "Digesting Anomalies: An Investment Approach." *Review of Financial Studies* 28 (3):650–705.

———. 2017. "Replicating Anomalies." Working paper.

Israel, Ronen and Tobias J. Moskowitz. 2013. "The Role of Shorting, Firm Size, and Time on Market anomalies." *Journal of Financial Economics* 108 (2):275 – 301.

Kadlec, Gregory B. and John J. McConnell. 1994. "The Effect of Market Segmentation and Illiquidity on Asset Prices: Evidence from Exchange Listings." *Journal of Finance* 49 (2):611–636.

Lehmann, E.L. and J.P. Romano. 2005. *Testing Statistical Hypotheses, Third Edition*. New York: Springer.

Lesmond, David A., Michael J. Schill, and Chunsheng Zhou. 2004. "The Illusory Nature of Momentum Profits." *Journal of Financial Economics* 71 (2):349 – 380.

Lettau, Martin, Sydney C. Ludvigson, and Sai Ma. 2018. "Capital Share Risk in U.S. Asset Pricing." Working paper.

Lettau, Martin, Matteo Maggiori, and Michael Weber. 2014. "Conditional Risk Premia in Currency Markets and Other Asset Classes." *Journal of Financial Economics* 114 (2):197 – 225.

Lewellen, Jonathan, Stefan Nagel, and Jay Shanken. 2010. "A Skeptical Appraisal of Asset Pricing Tests." *Journal of Financial Economics* 96 (2):175 – 194.

Lin, Chang-Ching and Serena Ng. 2012. "Estimation of Panel Data Models with Parameter Heterogeneity when Group Membership is Unknown." *Journal of Econometric Methods* 1 (1):42–55.

MacQueen, J. 1967. "Some methods for classification and analysis of multivariate observations." In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*. Berkeley, Calif.: University of California Press, 281–297.

Mankiw, N.Gregory and Stephen P. Zeldes. 1991. "The Consumption of Stockholders and Nonstockholders." *Journal of Financial Economics* 29 (1):97 – 112.

Menkhoff, Lukas, Lucio Sarno, Maik Schmeling, and Andreas Schrimpf. 2012. "Carry Trades and Global Foreign Exchange Volatility." *The Journal of Finance* 67 (2):681–718.

Merton, Robert C. 1987. "A Simple Model of Capital Market Equilibrium with Incomplete Information." *Journal of Finance* 42 (3):483–510.

Newey, Whitney K. and Kenneth D. West. 1987. "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica* 55 (3):703–708.

Novy-Marx, Robert and Mihail Velikov. 2016. "A Taxonomy of Anomalies and Their Trading Costs." *Review of Financial Studies* 29 (1):104–147.

Nozawa, Yoshio. 2017. "What Drives the Cross-Section of Credit Spreads?: A Variance Decomposition Approach." *Journal of Finance* 72 (5):2045–2072.

Patton, Andrew J. and Brian M. Weller. 2018. "What You See is Not What You Get: The Costs of Trading Market Anomalies." Working paper.

Pukthuanthong, Kuntara and Richard Roll. 2009. "Global Market Integration: An Alternative Measure and Its Application." *Journal of Financial Economics* 94 (2):214 – 232.

Rivers, Douglas and Quang Vuong. 2002. "Model Selection Tests for Nonlinear Dynamic Models." *Econometrics Journal* 5 (1):1–39.

Rouwenhorst, K. Geert. 1998. "International Momentum Strategies." *Journal of Finance* 53 (1):267–284.

Sarafidis, Vasilis and Neville Weber. 2015. "A Partially Heterogeneous Framework for Analyzing Panel Data." *Oxford Bulletin of Economics and Statistics* 2 (77):274 – 296.

Shleifer, Andrei and Robert W. Vishny. 1997. "The Limits of Arbitrage." *Journal of Finance* 52 (1):35–55.

Su, Liangjun, Zhentao Shi, and Peter C.B. Phillips. 2016. "Identifying Latent Structures in Panel Data." *Econometrica* 6 (84):2215 – 2264.

Vissing-Jørgensen, Annette. 2002. "Limited Asset Market Participation and the Elasticity of Intertemporal Substitution." *Journal of Political Economy* 110 (4):825–853.

Wu, C. F. Jeff. 1983. "On the Convergence Properties of the EM Algorithm." *Annals of Statistics* 11 (1):95–103.

Zivot, Eric. 2009. "Practical Issues in the Analysis of Univariate GARCH Models." In *Handbook of Financial Time Series*, edited by Thomas Mikosch, Jens-Peter Kreiß, Richard A. Davis, and Torben Gustav Andersen. Berlin, Heidelberg: Springer Berlin Heidelberg, 113–155.

A. Finding Global Optima

A. Numerical Issues and Equivalence with Generalized k-Means

We face three related optimization challenges in implementing the EM algorithm: selecting starting values; achieving global rather than local optima; and avoiding empty clusters. The choice of starting values for γ matters because expectation maximization finds local solutions, and final group assignments may depend heavily on initial group assignments. Likewise, we need a procedure to escape local basins of attraction on the likelihood surface in order to achieve the global maximum likelihood. Finally, clusters may depopulate to fewer than K elements as portfolios are reshuffled after λs are set. However, such collections of group assignments cannot be global optima because repopulating these clusters with (at least) K elements increases the likelihood function unless all portfolios are perfectly fit.

While we apply EM as the solution method to our maximization problem (4), our procedure closely resembles Lloyd's algorithm in k-means clustering. Like our algorithm, Lloyd's algorithm consists of update and assignment steps. Typically the update step consists of generating "centroids" by averaging characteristics within a group. This step is equivalent to minimizing squared errors within each group using a model with only a constant term. Linear regression also minimizes squared errors within each group, but it accommodates multi-factor models. Hence the first step of our methodology is a straightforward extension of k-means in which group "characteristics" are slopes for each date t and factor k. The assignment step of our methodology—choosing the cluster that minimizes errors given slopes—is exactly the same as in standard k-means in the sense of selecting the group that minimizes squared errors. However, in contrast with standard k-means, the importance of each characteristic for group assignment varies across observations in proportion with betas; the larger β_{ik} in absolute value, the more important λ_ks are for determining asset i's cluster.

Because of the similarity of our approach to k-means, we borrow and extend a common initialization method, known as "k-means++" (Arthur and Vassilvitskii (2007)). k-means++ is an algorithm designed to choose cluster centers such that Lloyd's algorithm achieves a clustering solution that is competitive with the global optimum. We discuss our extension of k-means++ in Appendix A.C. While proving the desirable properties of this algorithm is beyond the scope of our paper, we do find significant reductions in squared errors at the solution relative to initialization by choosing cluster memberships at random, suggesting that our variant inherits some of k-means++'s desirable properties.

We take two approaches to find global solutions. First, we initialize our version of k-means++ at $2N$ starting group assignments. We then run the EM algorithm from each assignment to find local optima. If our initializations cover the most promising basins of attraction, this step alone will suffice to locate the global maximum likelihood and corresponding group assignments. However, searching over γ is a high-dimensional problem requiring at least N group assignments to G groups, and $2N$ starting points may be insufficient to find global solutions. For this reason, we couple our multi-start approach with an explicit global optimizer able to accommodate integer problems.

In particular we use MATLAB's mixed-integer programming implementation of the genetic algorithm based on Deb (2000) and Deep et al. (2009). We initialize the population of the genetic algorithm with the $2N$ solutions of the EM algorithm as well as with N non-optimized initializations of our variant of k-means++. In so doing we cover a large number of local optima while allowing the algorithm to search new combinations and mutations toward a global solution. Note that we only need to search over γs, because the group assignments imply αs and Λs and likelihoods, and at the global best choice of groups, no groups need to be reassigned once αs and Λs are estimated. Once we have the final population from the genetic algorithm, we take the $2N$ highest likelihood-values from the local and global procedures and apply the EM algorithm to each to ensure that near-optima from the genetic algorithm are (at least) local optima. We select the highest-likelihood solution from this procedure and retain the corresponding group assignments and cross-sectional slopes.

In addition to using global optimization techniques directly, we also make a minor modification to the standard EM algorithm to avoid suboptimal assignments in which at least one cross-section is too small to obtain slopes. In particular, if after reassignment a cluster would have fewer than $K + 1$ elements, we introduce a likelihood hurdle for moving portfolios. Elements can only be reassigned if the improvement in likelihood is greater than c, and we choose the smallest c such that no cluster after reassignment would have fewer than $K + 1$ elements. Only then do we reassign portfolios to groups. Such a c always exists because in the worst case we can set c equal to the maximal change in likelihoods across observations to ensure no changes in group assignments occur. Once a group has $K + 1$ elements it will not depopulate in the subsequent iteration because all portfolios are perfectly fit, and c may return to 0. This step avoids convergence to dominated local optima with vanishing groups.

B. Comparisons of Local and Global Solutions

Figure A.I compares local solutions to the global best solution to (4) using the EM–genetic algorithm–EM procedure described in the preceding section. To illustrate possible optimization outcomes, we analyze the detailed examples described in V.C: domestic equity portfolios with a Carhart four-factor model; international equity portfolios with a global Carhart four-factor model; and cross-asset class portfolios with a two-factor intermediary capital factor model.

As a preliminary step, we define two measures for summarizing the distance between two optimizations. The first distance uses group assignments. Let the group assignment of the global best solution be γ^*, and the group assignment of a candidate alternative solution be γ'. Because groups are only known up to a permutation of their labels—there's no difference between 1-2 and 2-1 in a

Figure A.I: Comparison of Local and Global Solutions

Figure plots the distribution of optimizer solutions at the conclusion of the EM–genetic algorithm–EM procedure described in Section III.A. We analyze the detailed examples described in V.C: domestic equity portfolios (P3) with a Carhart four-factor model (top left); international equity portfolios (P5) with a global Carhart four-factor model (top right); and cross-asset class portfolios (P7) with two-factor intermediary capital factor model (bottom). In each case we use the AIC to select the number of groups, $G^* = \arg\min_G AIC_G$. Each subfigure displays two plots. The top plots are histograms of distances of the final $2N$ points from the likelihood-maximizing group assignment γ^*. Given cluster assignments γ^* and γ', we iterate over all possible permutations of the group labels for γ' and retain the permutation with the maximum number of groups in common. The histogram reports the distribution of this one minus proportion to obtain a "distance." The bottom of each subfigure plots z-scores of the $2N$ log-likelihoods against these distances. We mark the best optimization—on the y-axis with a γ^* distance of 0—with a red x.

(a) Domestic Equity Portfolios: Domestic Carhart, 1963-2016

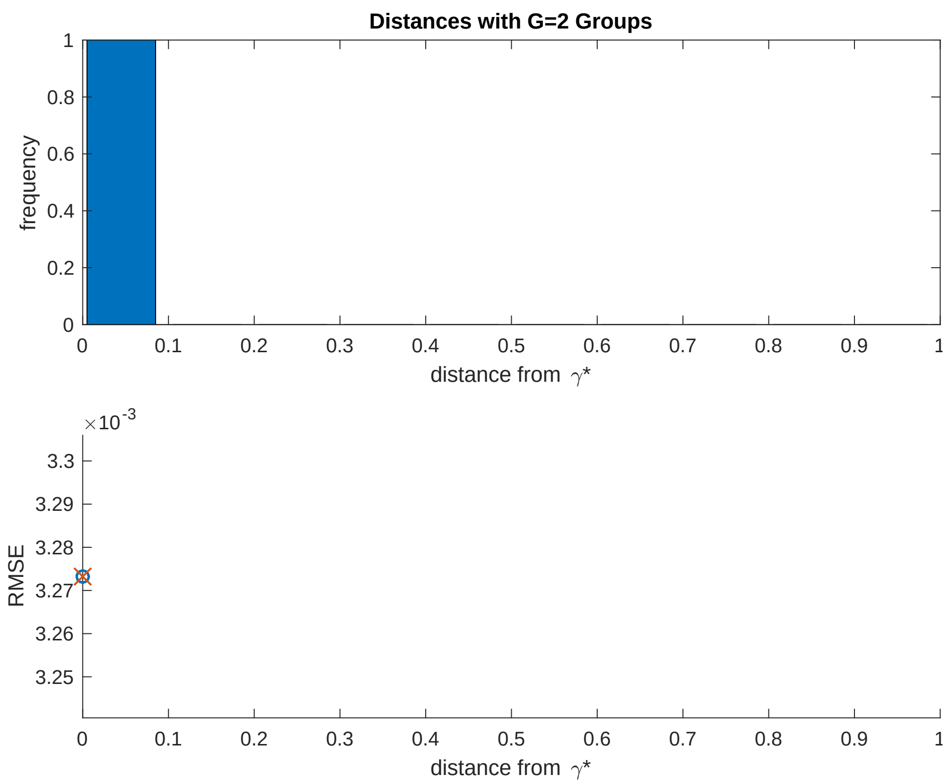
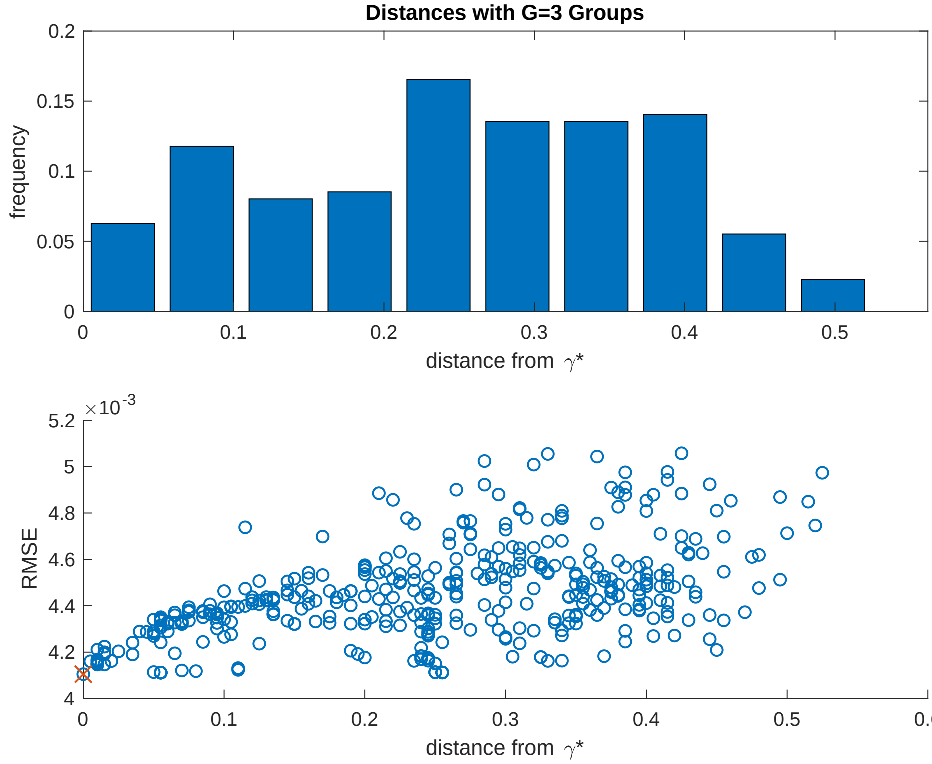
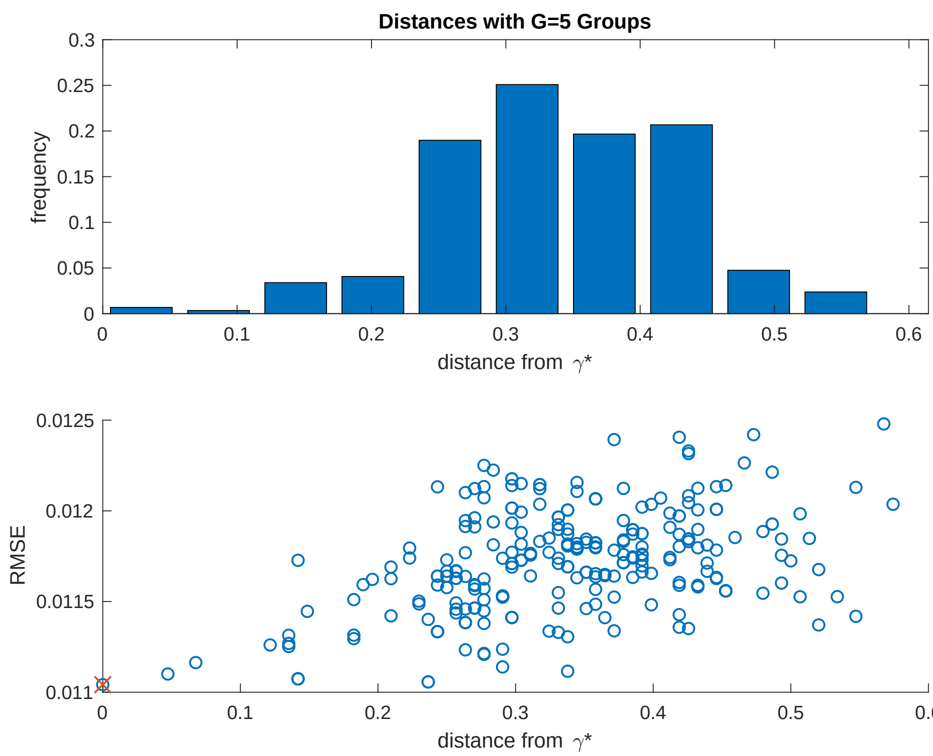


Figure A.I: Comparison of Local and Global Solutions (Continued)

(b) International Equity Portfolios: Global Carhart, 1991-2016



(c) Cross-Asset Class Portfolios: He, Kelly, and Manela (2017) Factors, 1986–2010



two-asset economy, for example—we need our distance measure to be robust to relabelings. To enforce this robustness, given cluster assignments γ^* and γ', we iterate over all possible permutations of the group labels for γ' and retain the permutation with the maximum number of groups in common. We define our first distance as one minus this proportion in common. The second distance uses the root mean squared error (RMSE), a standard measure of model quality. This metric is especially appropriate in our setting because the MSE is a linear function of (hard-to-interpret) log likelihoods, the optimizer maximand. We compare RMSEs to assess whether the optimizer reaches solutions of similar quality, even though the group assignments associated with those solutions may be quite different.

The first figure depicts a case in which all initial values converge to the same solution. Distances are zero for both metrics for all of the $2N$ final values of our procedure. This outcome suggests that our first setting is so well-behaved that EM converges to the global best solution from a wide range of starting values. The second figure indicates a large number of local maxima. Each point away from the global best represents group assignments different from γ^* and a higher associated RMSE. In this case the local solutions appear to populate a continuum in which distances gradually increase in group-space and RMSE-space. Here a local optimizer has a low chance of finding the global best solution, but many local solutions are "close" to the global solution. The third figure similarly suggests many local maxima. However, in this case, we see no clustering around the best solution. From an optimization standpoint, most runs fall into other basins of attraction with disparate group assignments, and a local optimizer would be highly unlikely to stumble upon group assignments near the global best. This said, despite the moderate-to-large differences in assignment distances, other partitions capture risk price heterogeneity almost as well as the segments described in V.C (as judged by the RMSE). Multiple dimensions of risk price heterogeneity are important in the cross-asset class setting, and the null hypothesis of unified risk pricing is likely to be rejected along more than one of them.

C. Extension of k-means++ to Cross-Sectional Slopes

The k-means++ algorithm of Arthur and Vassilvitskii (2007) proceeds as follows:

1. Choose the first cluster center at random among the existing data points.
2. For the c centers that have already been chosen, calculate distances of all data points to all cluster centers. Define $D(x)$ as the distance of data point x to the nearest center.
3. Choose a new center from the data points with probabilities proportional to squared distance.
4. Repeat steps (2) and (3) until the desired number of clusters have been chosen.

This initialization ensures that all clusters are well spaced, and this spacing alone dramatically reduces squared errors (and often, the run time) of the k-means algorithm.

Adapting k-means++ requires only a change in how we define our data points. Because our characteristics are cross-sectional slopes, we require at least $K+1$ assets to define each "data point." We add two preliminary steps A1–A2 to the algorithm to accommodate this difference:

A1. Draw $H = NM$ groups of size $M = \lfloor 1.5\,(K+1) \rfloor$ from the N assets.

A2. For each group $h = 1, \ldots, H$, estimate $\alpha_t^{(h)}$ and $\lambda_{kt}^{(h)}$ for all k and t.

Note that if the only factor is a constant ($K = 0$) we are back to the k-means++ case up to using random draws of portfolios rather than the set of portfolios itself.

The first step (A1) lets the number of potential cluster locations grow with the number of portfolios and factors. To precisely parallel k-means++ would require picking all $\begin{pmatrix} N \\ M \end{pmatrix}$ combinations of groups for candidate cluster centers, but this number of groups grows exponentially with K and is too large to be implementable. The second step (A2) obtains the characteristics (α and Λ) that determine cluster centers and cluster distances. Armed with these quantities we can proceed with k-means++ as before using α and Λ for each group in place of the underlying return data. In a zero-factor model, α *is* the underlying return data, and our algorithm again reduces to k-means++.

We also add a final step (B1) to k-means++ to prevent situations in which we obtain outlier (very small) or empty initial clusters:

B1. Drop the first selected cluster, and assign portfolios to the remaining clusters to obtain initial γs. Go back to step A1 if the size of the smallest cluster is too small relative to the size of the largest cluster, $N_{max}/N_{min} > 6$.

This modification ensures that we satisfy the theoretical requirement that the smallest cluster be large enough to use large-N asymptotics for the cross-sectional step. The first part makes a dominant first cluster less likely. The second part restarts the modified k-means++ algorithm from scratch when some clusters are too small.

B. Stability of Group Assignments

Our testing procedure assumes that group assignments are fixed over our sample period. However, just as the risk characteristics of portfolios may change over time, so too may the market frictions that separate portfolios into segments with different risk prices. We evaluate the stability of group assignments by comparing assignments across different sub-windows for the same factor models and portfolio sets. Our stability measure is the proportion of groups assignments in common, where we take the highest proportion of common assignments over all permutations of group labels (as in our distance measure of Appendix A.B). We use the number of clusters with the smallest AIC from the full sample.

Table A.I: Stability of Group Assignments Over Time

Table reports the proportion of stable group assignments across time periods. We first use the AIC to select the number of groups indicated in the full sample, $G^* = \arg\min_G AIC_G$. Then, for each date set, we estimate group assignments with G^* clusters. Given cluster assignments for two date sets, we report the proportion of group assignments that agree, taking into account that group labels are arbitrary. We repeat this procedure for all combinations of portfolio sets, risk models, and sample periods. Portfolios and models are described in the text. Bolded values indicate proportions for the three examples of Section V.C. For the He, Kelly, and Manela (2017) factor model, we use daily data for the most recent time period and monthly data for 1963–2016 and 1981–1998; we do not have sufficient coverage for their intermediary capital factor for 1963–1980 to include it. The q-factor (HXZQ) model is excluded from the international portfolio analysis because we do not have a global return-on-equity factor.

(a) Domestic Equity Portfolios

	Model	P1	P2	P3		Model	P1	P2	P3		Model	P1	P2	P3
Period 1	CAPM	0.81	0.67	0.56	**Period 1**	CAPM	0.84	0.57	0.55	**Period 2**	CAPM	0.89	0.53	0.61
1963–1980	FF3F	0.71	0.52	0.51	1963–1980	FF3F	0.75	0.39	0.56	1981–1998	FF3F	0.96	0.45	0.63
	Carhart	0.76	0.59	**0.68**		Carhart	0.71	0.63	**0.69**		Carhart	0.95	0.84	**0.90**
Period 2	FF5F	0.56	0.74	0.53	**Period 3**	FF5F	0.56	0.65	0.53	**Period 3**	FF5F	0.89	0.84	0.91
1981–1998	HKM				1999–2016	HKM				1999–2016	HKM	0.83	0.77	0.69
	HXZQ	0.81	0.69	0.56		HXZQ	0.77	0.85	0.57		HXZQ	0.96	0.61	0.50

(b) International Equity Portfolios

	Model	P4	P5
Period 1	CAPM	1.00	1.00
1991–2003	FF3F	0.68	0.96
	Carhart	0.73	**0.95**
Period 2	FF5F	0.98	1.00
2004–2016	HKM	1.00	0.77
	HXZQ		

(c) Cross-Asset Class Portfolios

	Model	P6	P7
Period 1	CAPM	0.70	0.61
1986–1997	FF3F	0.51	0.56
	Carhart	0.67	0.53
Period 2	FF5F	0.72	0.77
1998–2010	HKM	0.55	**0.53**
	HXZQ	0.49	0.59

Table A.I reports the proportions in common for all pairs of sub-windows for all sets of factor models and portfolio sets for which multiple sub-windows are available. Generally stability is high throughout: the average stability value exceeds 0.7, and about 30% of values exceed 0.8. In rare cases, stability is quite low for the domestic equity portfolios, such as for the Fama-French five-factor model applied to the widest domestic equity set. This limited stability suggests that segmentation indeed changes over time, at least for some risk factors and economic settings. Overall we interpret Table A.I as indicating that full-sample analysis should be complemented by sub-window analyses to capture the dimensions of risk-price heterogeneity most relevant during any particular time period.