The Factor-Structure of Time-Series Predictability*†

Patrick C. Kiefer‡ UCLA Anderson

September 6, 2018

Abstract

New light is shed on U.S. stock markets by studying "long memory" portfolios. These portfolios eliminate path dependence in common test asset returns, rendering transition dynamics Markovian. Common variation in returns is identified in terms of a *martingale* component and one or more asymptotically negligible *transitory* components. We find the concentration of return volatility on the martingale component - *the spectral gap* - is countercyclical and predicts annual market returns out-of-sample (o.o.s.) with an R^2 of 10.8%. Value (HML) predictability is concave and front-heavy, peaking at a one-year 14.7% o.o.s. R^2. In contrast, the momentum predictability term structure is convex, insignificant on the short end, but accelerates to 31.4% o.o.s. R^2 at the three-year horizon. We provide evidence the value premium is compensation for exposure to business cycle risks, and that excess returns to size (SMB) are not risk premia. Our findings imply new restrictions on the set of viable parametric asset pricing models.

Keywords: forecasting, time-varying risk, Fama-French factors, Markov process, spectral gap, orthogonal decomposition, fund separation, VAR, state-space models, performance evaluation, filtering

*Current version direct download or website `http://sites.google.com/patrickckiefer/research`

†Author contact: `pkiefer@anderson.ucla.edu`.

‡I thank Mark Grinblatt, Daniel Andrei, Michael Brennan, Mike Chernov, Bernard Herskovic, Jason Hsu, Bob Jarrow, Nick Kiefer, Holger Kraft, Juhani Linnainmaa, Lars Lochstoer, Francis Longstaff, Edward Mehrez, Michael Nowotny, Stavros Panageas, Rob Richmond, Subra Subrahmanyam, Phil Wool and Geoff Zhang. I thank my committee for patience and encouragement. Any errors are mine.

1 Introduction

Linear risk factor models reduce large complex markets to a handful of portfolios, making empirical and theoretical studies of asset markets tractable and efficient. This widely used reduction in dimension is a structural implication of equilibrium asset pricing theory: the so-called *factor replicating portfolios* are sufficient because the distribution of any security's returns can be characterized in terms of its exposure to these portfolios. While this simplification is exploited routinely in the context of unconditional factor modeling, it is rarely used to reduce complexity in problems of predictability.

This is not because equilibrium asset pricing theory is silent on dimension reduction in the time series. Consider a constant relative risk aversion (CRRA) representative agent economy with stochastic volatility - in this model, the conditional volatility of the pricing kernel is simply the risk aversion coefficient times the volatility of aggregate consumption growth. Predictable variation of any asset's excess return can be quantified using its beta, once we have the pricing kernel. The dimension reduction - from a large set of portfolios to the pricing kernel - evident in this model underscores a general implication. If the prices of risk vary through time, their forecasts are sufficient to forecast the excess returns of arbitrary portfolios.

We exploit these implications to parsimoniously characterize conditional mean returns in real time. The method is simple to implement using principal components analysis (PCA) and single-lag vector auto-regressions VAR(1), making it accessible to practitioners and regulators. By appealing to the pricing kernel structure in the time series, we also exploit the theoretical implication linking the dynamics of aggregate risk to the systematic components of predictability. We present several positive implications related to the method's success.

The primary challenge is the estimation of conditional means. Our approach is to construct a richer information set for realized returns with the goal of attaining a simplifying Markovian structure. Along with one-period realized returns, we include rolling returns from portfolios of the same assets that are held "to maturity" without rebalancing. These buy-and-hold portfolios store information about today from several previous periods, and then reveal the news relevant for those periods along with today's realized returns.

The information in these portfolios is redundant in the context of prevailing parametric asset pricing models. A sufficient condition for these portfolios to be redundant is that contemporaneous returns comprise a first-order Markov state vector. Empirically, this is not the case at mid to low frequencies in U.S. stock and fixed income markets[1] With the judicious inclusion of buy and hold portfolios, the Markov condition can be satisfied. From here, a known decomposition of Markovian

[1] A notable case is provided in Cochrane and Piazzesi 2005.

transition dynamics identifies systematic transitory variation - i.e., the factor structure of time-varying expected returns.

A critical conceptual point underpins the preceding description. The problem of estimating conditional expectations for possibly many assets was reduced to estimating parameters of a joint transition distribution for a few common sources of variation. But why might this be an easier task and why might it be justified? After all, if conditional means are prohibitively difficult to estimate, any number of them is prohibitive.

The answer expands on the point made by our CRRA example above. In equilibrium, the risk price is a market-wide variable - the volatility of the pricing kernel. Assuming no arbitrage, we can recover individual expected returns from an estimate of market covariance. Roughly, while the $Q-$ measure lets one avoid calculating drifts in a Black-Scholes option pricing setting, it also facilitates the recovery of individual drifts in a setting where neither the drifts nor the $Q-$ measure are known but the $Q-$ measure is better estimated.

We find evidence this connection is powerful for estimating expected returns. The *spectral gap* measures the fraction of priced risk that is time varying and predicts the market with an out-of-sample R^2 of 10.8% annually. The spectral gap is also used to predict portfolio returns. We forecast the value (HML) and momentum (UMD) premia with annual o.o.s. R^2 's of 14.7% and 6.7% respectively. The forecasts highlight cross-sectional differences in predictability term structures. At the two -year horizon, momentum predictability reaches 14.5%. While market and momentum forecasts exhibit increasing rates of explained variation, the value premium is most predictable at a 1-year horizon.

We investigate the theoretical implications of our findings in two ways. First, for each of the Fama and French risk factors, we replicate the predictive series obtained from the spectral gap within the space of test assets. This allows us to quantify the efficiency gains relative to the underlying factor models. Like the forecasts, the *timing portfolios* are constructed in real time. Unlike the forecasts, the timing portfolios are tradeable. We find returns to the timing portfolios have higher ex-ante Sharpe ratios and higher average realized Sharpe ratios than the underlying factors.

Second, we construct a time series of conditionally unpredictable returns from each of the timing portfolios. We test whether exposures to these shocks can explain cross-sectional variation in returns. In particular, using the underlying factor replicating portfolios as benchmarks, we isolate incremental changes to cross-sectional performance. Gains in efficiency from the market, momentum and value timing portfolios are priced in the cross-section, while gains from size (SMB) timing portfolios are not.

Classically, an empirical risk factor is formed in two steps. Traded assets are sorted into bins according to a characteristic, such as book-to-market equity. Then, a zero-cost portfolio is formed by shorting the lowest bin and buying the highest bin. When the sorting characteristic proxies for exposure to an undiversifiable source of risk, mean returns to the zero-cost portfolio are proportional to the unconditional price of exposure to this risk. The resulting *factor replicating portfolios* form the basis of empirical risk factor models.

However, conditional and unconditional expected returns differ systematically. The dividend yield, *cay*, and other variables forecast market returns, producing estimates of the time-varying expected returns.[2] In equilibrium the expected excess return on the market is equal to the market risk price squared (uniquely up to changes in basis). Similarly, portfolio-level forecasts can be used to estimate time-varying risk prices for value (HML), size (SMB), momentum (UMD) and other non-market factors. Importantly, systematic variation in expected returns contains the time-variation in factor risk prices. When all systematic variation is priced, the two are equivalent.

Evidence suggests the term structure of equity risk is not flat, and that the sign of the slope changes depending on economic conditions.[3] Information in the full term structure may be relevant for forecasting single-period returns, but the exact term structure is not observed and no consensus estimation method exists. To capture this information while remaining agnostic about the structural model, we construct return forecasts of priced risk factors for several horizons separately. Henceforth we refer to these forecasts as *nominal* forecasts.

To estimate time-varying risk prices, we augment asset returns with nominal forecast errors, fitted within each rolling window separately for several horizons. The nominal forecasts we use are fitted values of Fama-French risk-factor returns on standard (lagged) predictors, such as the dividend yield and past returns. We reject two important null hypotheses that characterize nominal forecasts. The first is that nominal forecasts over different horizons are deterministic functions of each other. The second is that every nominal forecast is replicable in the space of (contemporaneous) test assets.

The signals in the nominal forecast errors are a key input for extracting the latent factors and risk price estimates. Consider an event of the sort: "nominal market forecast errors at the 4-year horizon tend to be high when one-month HML returns are low." The information in this event is *not* contained in contemporaneous realized returns if the sequence of nominal 4-year market forecast errors *cannot* be replicated by a portfolio of test assets.[4] Conditioning on these events is valuable empirically. In addition to refining the information set, nominal forecast errors are chosen in exactly the linear combinations of past returns that justify a Markov representation.

[2] The series *cay* is constructed in Lettau and Ludvigson (2001) to capture deviations in the consumption-wealth ratio from its long-run mean.

[3] Binsbergen and Koijen (2016) provide a thorough survey.

[4] Specifically, today's realization of the nominal 4-year market forecast error is the difference between today's one-period realized market return and the prediction of that return made 4-years ago.

Given a Markov representation of returns - including past returns, in the case of U.S. equities - we run a classical principal components analysis for each period over a fixed history length. The PCA maps to a decomposition of a generic vector Euler equation when dynamics are Markovian.[5] The history length is chosen to isolate fluctuations at a particular distance from statistical equilibrium, controlled by the so-called *mixing times*. The representation is updated in each period by repeating this construction.

The construction captures changes in the risk price and composition of transitory factors while keeping the scale of the *average* transitory fluctuation fixed at a constant fraction of total volatility. Technically, for a given threshold, we keep the mean mixing time fixed. Changes in the risk price and composition of transitory factors correspond to *systematic fluctuations in expected returns*. We find that while an average of 86% of quarterly return volatility is concentrated on permanent shocks to market capitalization, the forecasting power comes from incorporating the systematic predictable variation in expected returns - *the expected return factors*.

Writing the Euler equation in terms of decomposed returns shows the *spectral gap* is informative about future expected returns. A naive test of predictability using the lagged spectral gap predicts quarterly, semi-annual and annual market returns with out-of-sample R^2s of 3.8%, 6.5% and 10.8%. More sophisticated statistics and existing predictors are no better than the lagged spectral gap out of sample, with *cay* coming in second with at an annual o.o.s. R^2 of 7.6%. Gap statistics also forecast portfolio returns, predicting value (HML) with an annual out of sample R^2 of 14.7% and Momentum (UMD) with an R^2 of 6.4% (16.1% biannually).

Portfolio-level predictability term structures are significantly heterogeneous. Explained variation increases monotonically with horizon for the market and momentum risk factors. For value, explained variation is hump-shaped, with a maximum R^2 corresponding to the one-year forecasting horizon and decreasing afterward. We find the market loads with a coefficient 0.98 on shocks to the leading component. The leading component's autoregressive coefficient is indistinguishable from zero (the point estimate is 0.089 with s.e. 0.072). In contrast, value returns load significantly on the penultimate factor. The penultimate factor tracks common variation in expected returns - the autoregressive coefficient is 0.649 with an adjusted standard error of 0.092.

In addition to the concave predictability term structure for the value premium, we find the momentum term structure is increasing and convex. Forecast horizons inside of 1 year feature rapidly increasing predictability in the value premium, while momentum is almost unpredictable. Forecast horizons longer than one year feature decreasing predictability of the value premium and simultaneous rapid increases in predictability of the momentum premium. Predictability of market expected returns increase linearly over the same forecasting horizons.

[5] A formal description of this procedure is given in section 5.5.1.

Value-weighted dividend yields play a key role in the construction of productive nominal forecasts. Using the value weighted CRSP index ex-dividend in place of dividend yields when constructing the nominal forecasts significantly restricts the predictive power for each factor other than size. However, predictability for the market picks up at the long end, suggesting capital gains have a small but significant role in positively predicting aggregate expected returns over lower frequencies. The equal-weighted CRSP ex-dividend index input generates predictive power for size but insignificant predictive power for the value premium, and significantly reduced predictive power for the market and momentum premiums.

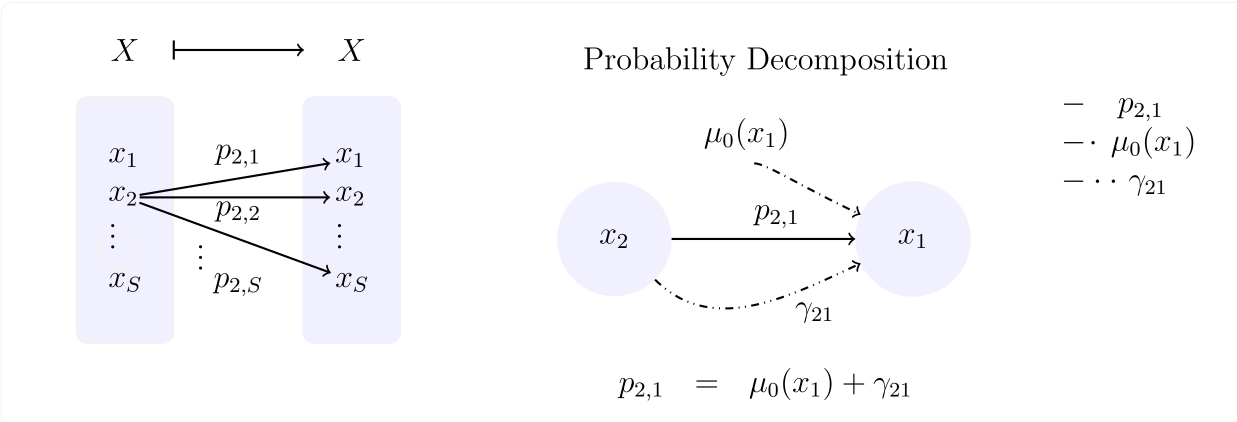
Our findings are consistent with existing evidence on predictability and dividend yields. Nominal forecasts using single lags of dividend yields produce factor return forecasts that are inferior but nonetheless dominate dividend yield forecasting directly. Dividend yields are highly persistent, fluctuating with a half-life of roughly 15 years, and forecast returns. From this standpoint, that the full term structure of forecast errors is informative about expected returns is unsurprising - it is certainly informative about future yields.

We find the spectral gap is a meaningful macroeconomic indicator on its own. Dynamics of the spectral gap measure changes in the concentration of volatility on the leading priced risk factor. In the US stock markets, we find return volatility concentrates countercyclically on permanent, or "long-run" shocks. These findings have implications for parametric stochastic discount factor (SDF) models. Because the spectral gap is an accessible object in any Markov model of asset prices, these findings can help discriminate among competing parametric theories.

The prices of incremental gains from timing portfolios are informative about the risk content of the underlying factor. By projecting the factor forecasts back on to the space of test assets, we limit our analysis to changes in the distribution of risk across portfolios because the total risk is constrained to be the value-weighted excess returns of the test assets. Relative efficiency gains measure the *intensive* margin of factor risk. Marginal efficiency gains arise when marginal and average risk prices diverge. We reject that the marginal prices of exposure to the value and momentum portfolio returns are zero, but we cannot reject that the marginal price of exposure to size is zero.

The study proceeds with a discussion of the literature, followed by a Markov model of asset returns. Section 4 describes the empirical tests derived from the Markov model. The data and empirical results are reported in sections 4.3 and 4.4. Appendices contain technical details and a handful of ancillary charts and tables.

Figure 1: Decomposition over Finite State Space



(a) Predictable fluctuations are identified using a decomposition of transition probabilities. The probability of moving to x_1 given the current state x_2 is $p_{2,1} = \mu_0(x_1) + \gamma_{21}$. The invariant probability $\mu_0(x_1)$ does not depend on x_2, and dominates forecasts asymptotically. The transitory correction γ_{21} becomes negligible in large time, but contributes importantly to local dynamics. Perron-Frobenius and spectral theories provide decompositions of the signed measures $\nu : X \times X \mapsto \mathbb{R}$ rather than the state space X itself. This point is clear for finite-state ergodic processes, pictured here, where it does not make sense to claim subsets of finitely many points are transitory: all points in an ergodic set are visited with probability one in large time.

2 Related Work

Alvarez and Jermann (AJ) (2005), and Hansen and Scheinkman (2011) factorize the pricing kernel into martingale and transitory components. (AJ) find that volatility of the growth rate of the martingale is roughly 90% of the volatility of the stochastic discount factor. Hansen and Scheinkman (2011) use the Perron-Frobenius theory to isolate the asymptotic risk-return tradeoff for an aggregate payoff functional when the underlying dynamics are Markovian. Borovicka, Hansen and Scheinkman (2016) point out the structure and interpretation of the Perron-Frobenius estimate depends heavily on model specification. We use a Perron-Frobenius decomposition to identify predictable fluctuations in latent factors that are negligible asymptotically. These components constitute roughly 15% of the total variation in returns, corroborating findings in (AJ).

Chen, Roll and Ross (1986) find that exposure to innovations in the term spread, credit spread, and industrial productivity help explain cross-sectional variation in average stock returns. Hansen and Jagannathan (1991), (AJ), Ross (2015), and Backus and Chernov (2008) argue that while important work is done using macroeconomic variables to understand asset prices, it is also the case that equilibrium prices reveal information about the macroeconomy. In particular, Hansen and Jagannathan (1991) study restrictions placed by observed prices on the mean and variance of the pricing kernel and argue the pricing kernel must be an order of magnitude more volatile than consumption growth to justify the observed Sharpe ratios. Backus and Chernov (2008) study restrictions on pricing kernel cumulants implied by observed prices and use evidence of higher moments to rule out symmetric dynamics. Ross (2015) argues for recovery of the underlying physical transition dynamics from returns. To our knowledge, identification of persistent dynamics in the pricing kernel from the covariance matrix of returns is a novel contribution.

Koijen, Lustig and Van Nieuwerburgh (2017) propose a three-factor with a separate role for business-cycle fluctuations to explain average returns to stock portfolios sorted on book-to-market equity and maturity-sorted treasury portfolios. Cochrane and Piazzesi (CP) show a single bond factor constructed from a linear combination of forward rates predicts returns on bond portfolios of any maturity and forecasts returns in equity markets. The (CP) factor describes transitory fluctuations in the Koijen, Lustig and Van Nieuwerburgh economy; the market factor captures permanent shocks to cash flow levels, and the third factor updates inflation expectations. Our findings corroborate and extend their classification of factors. We show over 90% of the variation in the market and the momentum (UMD) factor is driven by innovations with no transitory content, while over 70% of the variation in value (HML) is explained by the transitory factors.

Fama and French (1992, 1993) and Asness (1994) document the importance of book-to-market equity for explaining cross-sectional variability in average returns and establish the value (HML) and size (SMB) factors to augment the single-factor (CAPM) model. Jegadeesh and Titman (1993) and Moskowitz and Grinblatt (1999) study momentum by asset and industry, respectively, and find a cross-sectional ranking of past winners and losers incrementally improves dynamic efficiency properties of the Fama-French three factor models. Berk (1995) points out that if size is the expected cash flow level, and two firms have identical "size" but one has lower market cap, it is because it has a higher discount rate. Mechanically, size inversely predicts average returns and will appear to be priced whenever a true factor is missing. We find incremental efficiency gains from timing size are not priced, while gains from timing the market, momentum and value are priced.

Several papers argue for improvements in the prevailing constructions of Fama-French factors. Gerokos and Lihnnainmaa (2012) argue that the HML factor returns can be decomposed into price-driven and book-driven elements, and that only the price-driven component of the HML factor returns can explain cross-sectional variation in returns. Asness and Frazzini (2013) argue that HML contains about 20% momentum, and propose a construction of HML that isolates the "pure value" component. Lihnnainmaa (2015, 2016) finds accrual, investment, and profitability factor constructions that are preferred to the Fama French (2015) constructions for cross-sectional asset pricing. We find evidence the value premium is compensation for shocks to the persistence of expected returns. In contrast, we find momentum is compensation for i.i.d. shocks to realized returns.

Binsbergen, Brandt, and Koijen (2012) synthesize dividend strips at various maturities to analyze the term structure of equity risk. They find that short-run cash flows have higher average returns that the market, implying a downward-sloping term structure of equity premia. Ait-Sahalia, Karaman, and Mancini (2015) provide evidence that the sign of the slope is time-varying and procyclical. Schulz (2016) argues the downward sloping term structure becomes insignificant when tax rates specific to dividend income are considered. Weber (2016) sorts stocks based on a measure of cash flow duration and finds high-duration stocks earn roughly a one-percent premium monthly over low-duration stocks, providing term-structure evidence that does not rely on synthesizing dividend payments. This set of findings motivate us to study nominal forecasts of returns on equity portfolios separately at different horizons.

Bandi and Tamoni (2015) implement a time-scale decomposition of returns and consumption growth by projection on the Haar basis. The result is a representation of the time-series of returns as the sum of moving averages over J non-intersecting intervals of increasing scale 2^j $j \in \{0, 1, ..., J\}$. Severino (2014) shows existence of time-series decompositions based on the sequential application of an isometric operator. The decomposition splits a Hilbert space into an infinite direct sum of rank-one prediction-error subspaces and that sum's (possibly empty) orthogonal complement. The derivation generalizes the Wold representation. Our decomposition is most related methodologically to Bandi and Tamoni (2015) and Severino (2014). The functional basis in our decomposition is identified from a PCA of a generalized covariance matrix of realized returns.

Bansal and Yaron (2005), and Hansen, Heaton and Li (2008) propose and evaluate slow-moving latent growth factors as explanations for unconditional risk premia. Bansal, Kiku and Yaron (2010) sharply characterize the distinction between long run risk - captured by shocks to persistent levels driving growth, and business cycle risk - captured by shocks to persistent growth directly. Hansen and Sargent (2007, 2016) outline the similarities between long-run risk and the implications of robust control policies in financial markets. The authors show the long-run risk model is the model a robust Epstein-Zinn investor would appear to have referenced ex-post. We contribute to this discussion by quantifying slow moving changes in the concentration of return volatility on long-run shocks. Estimates of this quantity predict portfolio returns and contribute to priced factor risk, suggesting a new layer of tests from within parametric long-run-risk or ambiguity aversion models.

Jagannathan and Wang (JW) (1996) take a CAPM model with conditionally dependent parameters and condition down. It is well known that the unconditional version includes correction parameters capturing the correlation between time- varying exposures and time- varying risk prices. JW circumvent the problem of obtaining conditional estimates directly by using the $BAA - AAA$ credit spread as a proxy for the time-varying risk price. They provide GMM estimates that support the conditional CAPM over the implied constant parameter CAPM tested classically. In U.S. stocks, we provide evidence the latent pricing kernel contains the required forecasting variables itself. As a diagnostic, we find including the $BAA - AAA$ credit spread in the calculation of the covariance matrix generates forecasts with o.o.s. R^2 gains of 0% to 2% annually, depending on the portfolio and subsample.

Cochrane (2011) surveys the in-sample evidence suggesting that, across many asset classes, low yields (high prices) today predict low returns in the future - and not cash flow growth. The importance of jointly restricting cash flow and yield parameters in a vector-autoregression (VAR) for evaluating the predictability of discount rates is given by Cochrane (2008). Santa Clara (2015) finds evidence that dividend yields in fact predict cash flows the portfolio-level. Brennan and Taylor (2016) find aggregate and portfolio-level return predictability out of sample. The authors distinguish risk-based sources of predictability from potentially non-equilibrium present-value predictors by comparing variables obtained from the covariance matrix with those obtained from replicating portfolios for present values of cash flow news and discount rate shocks. We base aggregate and portfolio-level o.o.s. forecasts on a small number of common factors - also advocated by Cochrane (2011). To quantify the risk content of our predictors, we test whether the incremental efficiency gains, measured by the difference in returns between a forecast replicating portfolio and the underlying target portfolio, are priced in the cross-section.

Chen (2005) argues that out of sample (o.o.s.) tests of predictability are well suited for studying time series containing one or more discrete structural breaks. In Taiwanese markets, Chen finds that o.o.s. predictive statistics capture discrete structural breaks in savings rates that predict declines in investment rates that followed. Importantly, an in-sample vector-autoregression model misses the break and fails to to reject the null of no-predictability. We provide diagnostics in section (5.5) indicating our o.o.s. predictors react to structural breaks quickly relative to rolling beta and rolling benchmark PCA models.

Goyal and Welch (2008) provide evidence that predictors in the literature perform poorly out of sample, and that prediction quality in-sample is often confined to crisis periods. However, although low in absolute terms, Goyal and Welch (2003) and (2008) find the relative forecasting power of the dividend-price ratio is highest using out of sample predictions over the postwar sample ending in 1990. They argue the incremental efficacy of the forecasts arise through the ability of rolling beta estimates to pick up changes in the data generating process (DGP). Similarly to Chen (2005), the authors find a constant coefficient VAR model fails to generate a significantly non-zero R^2. Our predictive efficacy is not limited to crises, suggesting we capture important cases where the underlying model changes smoothly but manifests as a significantly distinct process along sufficiently non-overlapping subsamples.

Kelly and Pruiit (2013) implement a three-pass filter to forecast portfolio returns using a large cross-section of predictor variables. They report test statistics derived to account for the fitting procedure, and measure out-of-sample predictive R^2 by fitting the model to data omitting the target period and predicting the target period. R^2s on annual market predictability reach 13%. We report comparable but lower annual market R^2 of 10.8% without using cross-sectional data and without using future data. Thus, our decomposition is informative in real time. However, the predictive regressions proposed by Kelly and Pruiit (2013) apply to predictive settings where a Markov structure may not be warranted, while our identification relies on this structure.

Bryzgolva (2014) advocates for traded (price-based) proxies for risk factors over macro factors for statistical reasons. A constrained LASSO-style regression penalizes candidate risk factors with poorly measured exposures. Traded factor returns have low levels of idiosyncratic noise, making exposures easier to measure and thus the LASSO procedure penalizes traded proxies less. This helps rationalize the incremental gain in statistical significance from the timing portfolio returns over the extracted latent process. Simultaneously, it underscores the improvements in significance and precision of the timing portfolios relative to the conventional factor replicating portfolios.

Shrinkage estimation procedures - such as the LASSO for the covariance matrix - are a potentially relevant exercise in our setting. We avoid this by first considering the factor returns only, avoiding the rank-deficiency in cross-sections with many test assets. However, the inclusion of nominal forecasts in the generalized covariance matrix estimation introduces rank deficiency. Moreover, one may shrink towards a target horizon Sharpe ratio when the data matrix includes many horizon-specific forecasts. For the purposes of this paper, we manage rank deficiency and sparseness by exploiting the inner-unitary properties of the rotation matrices from a singular value decomposition. Cases of explicit shrinkage are delegated to future work.

We draw on tools from Markov processes, large deviations and random matrix theory. The modern theory of large deviations is due to Donsker and Varadhan (1975a, 1975b) and Gartner (1977), building on Cramer (1938) and Sanov (1957). Exponentially unlikely deviations of Markov processes are described by the Perron-Frobenius theory (Varadhan (1983, 2008), Hansen (2011), Borovicka, Hansen and Scheinkman (2016)). Brownian dynamics for the spectra of random matrices were introduced by Dyson (1962a, 1962b). Erdos and Yao (2017) and Tao (2011) characterize spectral dynamics for sequences of random matrices and nest Dyson Brownian motion as a special case. Knowles, Yao and Yin (2014) provide asymptotics for outlying eigenvalues of covariance matrices when the parameter dimension grows proportionally to sample size.

Interest in Markov-Chain Monte-Carlo (MCMC) methods drove a better understanding of convergence rates for finite-state Markov chains. Estimates have been characterized in terms of the log-Sobolev inequality (Diaconis and Saloff-Coste (1996a)), the Poincare inequality on graph representations (Diaconis and Strook (1991), Tuominen and Tweedie (1994)), and the spectral gap (Diaconis and Saloff-Coste (1996b), Saloff-Coste (2004)). Each of these build on Doob (1959), Nash (1958) and more recently Anderson (1989). Diaconis (2009) provides an excellent discussion of related developments. Fukushima (2010) emphasizes quadratic and bilinear form representations of Markov processes on general state spaces and touches on their spectral content. Chen, Hansen and Scheinkman (2007) make this connection explicit for the Feynman-Kac semigroup.

3 A Markov Model of Returns

We construct an empirical model to use for forecasting. Sections 3.1 and 3.2 describe the process environment and specify asset prices and a risk-neutral measure. Section 3.3 reviews the decomposition. Section 3.4 provides an example to highlight the economics of equilibrium prices and transitory fluctuations. 3.5 - 3.6 present the spectral gap, relates it to a martingale representation, and characterizes the identification of the spectral gap from the empirical covariance matrix.

The lag operator and composition of the lag and Markov operators are defined on the *path* space of the Markov chain, so some technical statements cannot be avoided. Extensive derivations are left for the section (7) appendix, along with the complete set of proofs.

3.1 Market Prices

Undiversifiable risk arises from S - state Markov jump dynamics taking values in a finite ordered set $x_t \in X := \{x^{[1]}, ..., x^{[S]}\} \subset \mathbb{R}^S$. Time is discrete. We take each $x^{[j]} \in \{0,1\}$ so the state at time t, $x_t = x^{[j]}$ is characterized by the index $\{[j] : x^{[j]} = 1\}$. Local dynamics of the Markov chain X_t are described by the *kernel* function $m(x^{[i]}, x^{[j]}) := \Pr(X_t = x^{[j]}|X_{t-1} = x^{[i]})$.

Sequentially traded state-contingent securities $d_n \in F(X) \subset \mathbb{R}^S$ are sufficient to build up rich dynamically complete cross-sections as in Arrow (1953). We specialize asset $n = 0$ to $d_0 = (1, 1, ..., 1)'$.[6] We will extend the marketable security space to include long-lived securities recursively, but we first establish the benchmark asset prices.

Market equilibrium implies a positive pricing kernel exists and can be used in lieu of replication to price arbitrary cash flows (Ross (1976), Harrison and Kreps (1979)). Let $s_{j,t} = s_j(w_j(x_t), t) = \beta_j^t \, \tilde{s}_j(w_j(x_t))$ be the marginal value of wealth for investor j, where $\beta_j \in (0, 1)$ captures time discounting. For any asset n and market prices $p_{n,t}$, individual optimality requires

$$s_{j,t} p_{n,t} = \mathbb{E}_t[s_{j,t+1} d_n(X_{t+1})] \tag{3.1.1}$$

[6]Contingencies d_n are assumed to satisfy $\nu d_n d_n' < \infty$, which in finite dimensions under any positive probability measure ν is equivalent to $d_n \cdot e_i \leq M < \infty$ for every $n, 0 \leq i \leq S$ and some fixed scalar $M \in \mathbb{R}$, where $\{e_i\}, 0 \leq i \leq S$ denotes the standard basis in $\mathbb{R}^S$.

in each period t. In equilibrium, the stochastic discount factor (SDF) $S_{t,t+1} = S(X_{t+1}|x_t, 1)$ encodes market-wide preferences in observed prices, enforcing

$$s_{j,t}\mathbb{E}_t[S_{t,t+1}d_i(X_{t+1})] = \mathbb{E}_t[s_{j,t+1}d_i(X_{t+1})]$$

for any unconstrained investor j, and arbitrary i, t. In particular, 3.1.1 becomes

$$p_{n,t} = \mathbb{E}_t[S_{t,t+1}d_n(X_{t+1})] \tag{3.1.2a}$$

for every n given t and every t. We follow the convention in Alvarez and Jermann (2005) by modeling the SDF as the ratio of the pricing kernels $S_{t,t+k} = s_{t+k}/s_t = \beta^k \tilde{s}_{t+k}/\tilde{s}_t$. Then, the pricing kernel is the particular SDF when the reference period wealth is the numeraire $s_t = S_{0,t} = S(X_t|x_0, t)$, $s_0 \equiv 1$. For the asset $n = 0$, $p_{0,t} = \mathbb{E}_t[S_{t,t+1}] = 1/r_{f,t}$ is the price of a one-period default free bond per unit of face value.

From 3.1.2a, the value of the replicating portfolio for any security n is given by

$$p_{n,t} = \widehat{\mathcal{Q}}_k d_n \tag{3.2.2b}$$

where $\widehat{\mathcal{Q}}_k := 1(x_{t,k})'\widehat{\mathcal{Q}}$, $1(x_{t,k})$ is a column of zeros outside of the indicator at k, and $x_{t,k}$ is shorthand for the conditioning event $x_{t,k} := \{X_t = x^{[k]}\}$. The risk neutral transition dynamics follow by normalizing the entries of $\widehat{\mathcal{Q}}$ so that the rows are probability measures as follows $\mathcal{Q}_k := b_t^{-1}\widehat{\mathcal{Q}}_k$ with $b_t = 1/r_{f,t} = \iota'\mathcal{M}_k s_t$.

3.2 Decomposition

A first-order transition distribution can be decomposed into two orthogonal components.[7] Each transition probability $m(j, k)$ is comprised of a local *transitory* and a non-local *permanent* component. The local components contains state-dependent transitioning information. The non-local component completely determines asymptotic forecasts. Both components are important in finite samples.

Let $\iota = 1_{S\times 1}$; then by our convention, $\mathcal{M}\iota = \iota$. We assume the chain $(X, \mathcal{M})$ is ergodic, which implies a unique invariant $\mu_0' = \mu_0'\mathcal{M}$. The pair (μ_0, ι) are the left and right eigenvectors of $\mathcal{M}$, respectively, normalized so that μ_0 is a probability measure $\mu_0'\iota = 1$. From these assumptions, we

[7]See proposition (7.1) parts $I. - III.$ and corollaries (7.4) – (7.9)).

obtain the representation for the dynamics of distributions over X

$$\mathcal{M}' = \mu_0 \iota' + \mathcal{M}'_\gamma$$

Asset return dynamics inherit this representation

$$\begin{aligned}\mathbb{E}[R_{n,t+1}(X_{t+1})|x_{t,k}] &= 1(x_{t,k})'\mathcal{M}r_n \\ &= \overline{r}_n + 1(x_{t,k})'\mathcal{M}_\gamma r_n\end{aligned}$$

where $\overline{r}_n := r_n \cdot \mu_0 = 1(x_{t,k})'(\mu_0\iota')'r_n$ is the long-run mean return for asset n. The operator $\mathcal{M}'_\gamma$ drives purely transitory variation.[8] In corollary (7.6), we establish the classic Wold representation applied to returns,

$$R_{n,t+1} = \overline{r}_n + r_n \cdot \sum_{s=0}^{\infty} (\mathcal{M}'_\gamma)^s 1(u_{t+1-s})$$

In lemma (7.2), we identify $\nu = \mathcal{M}'1(x_{\cdot,k})$ with probability measures over X given $x_{\cdot,k}$ for any k . Hence, the decomposition states that any transition probability can be written

$$p_{i,j} = \mu_j + \gamma_{i,j} \tag{3.3a}$$

for order pairs $x^{[i]}, x^{[j]}$. 3.3a indicates that conditioning on $x_{t,i}$ we arrive at $x_{t+1,j}$ with probability equal to the long-run occupation rate of the coordinate j plus a correction term $\gamma_{i,j}$. In corollary 7.1.2, we show that $\mathcal{M}_\gamma \iota = 0$, i.e., $\sum_j \gamma_{i,j} = 0$. Hence, if any individual term $\gamma_{i,j}$ is nonzero for a given i, then at least one of the entries is negative for that i.

Transition dynamics in general are distinct from dynamics of conditioning information, but can be characterized tractably. Using proposition (7.1) and corollaries (7.2)-(7.3), expected return dynamics are

$$\begin{aligned}\mathbb{E}_t[R_{t+k}] - \mathbb{E}_{t-1}[R_{t-1+k}] &= \Delta[(\mathcal{M}')^k x_{t-1}] \\ &= \Delta \left(\mathcal{M}'_\gamma\right)^k x_{t-1} + \Delta\widehat{\mathbb{E}}_{t-1,k}\end{aligned}$$

The term $\Delta \left(\mathcal{M}'_\gamma\right)^k$ captures changes in transition probabilities conditionally. Following Dyson

[8] We assume throughout that the columns of $\mathcal{M}'$ are not each identically μ_0, so the decomposition is nontrivial (i.e., $\mathcal{M}'_\gamma \neq \mathbf{0}$).

(1962), the leading order terms for dynamics of $\left(\mathcal{M}'_\gamma\right)^k$ are

$$\Delta[\left(\mathcal{M}'_\gamma\right)^k] = -(1-\lambda_2)^{-k}\gamma\gamma' + o\,(\cdot) \tag{3.1s}$$

The kernel function γ of $\mathcal{M}_\gamma$ can be viewed as the left eigenvectors of $I - \mathcal{M}_\gamma$, even if $\mathcal{M}$ is not reversible. Lower eigenvalues of $\mathcal{M}$ are eigenvalues of $\mathcal{M}_\gamma$. We show this in sections 7.4-7.5. Setting $\langle\gamma, 1\rangle = \boldsymbol{\gamma}$, and $\Delta\widehat{\mathbb{E}}_{t-1,k} = 0$ and conditioning,

$$\mathbb{E}_t[R_{t+1}] - \mathbb{E}_{t-1}[R_t] = -\zeta_t^{-1}\boldsymbol{\gamma}\gamma x_{t-1} \tag{3.2s}$$

describes time-varying mean returns. $\zeta_t = \lambda_1 - \lambda_2 = 1 - \lambda_2$ is the *spectral gap* of $\mathcal{M}$.

The dynamics of spectral data provide a summary for the dynamics of the risk prices because in equilibrium the risk prices are eigenvalues. We consider a fixed known covariance matrix plus mean zero i.i.d. random perturbations.

4 Empirical Implementation

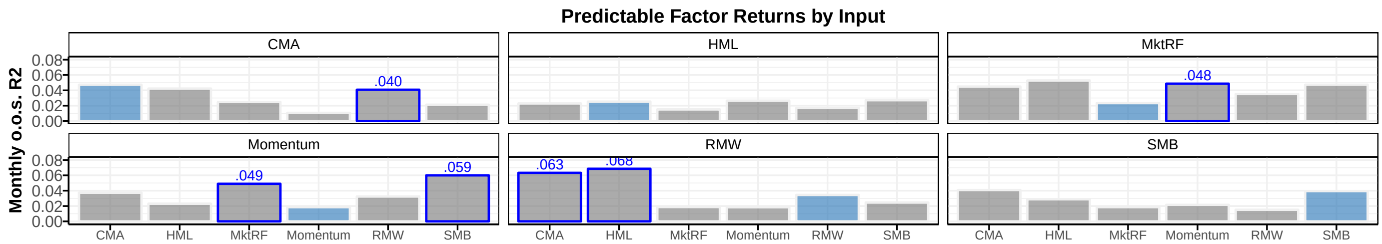
4.1 Nominal Forecasts

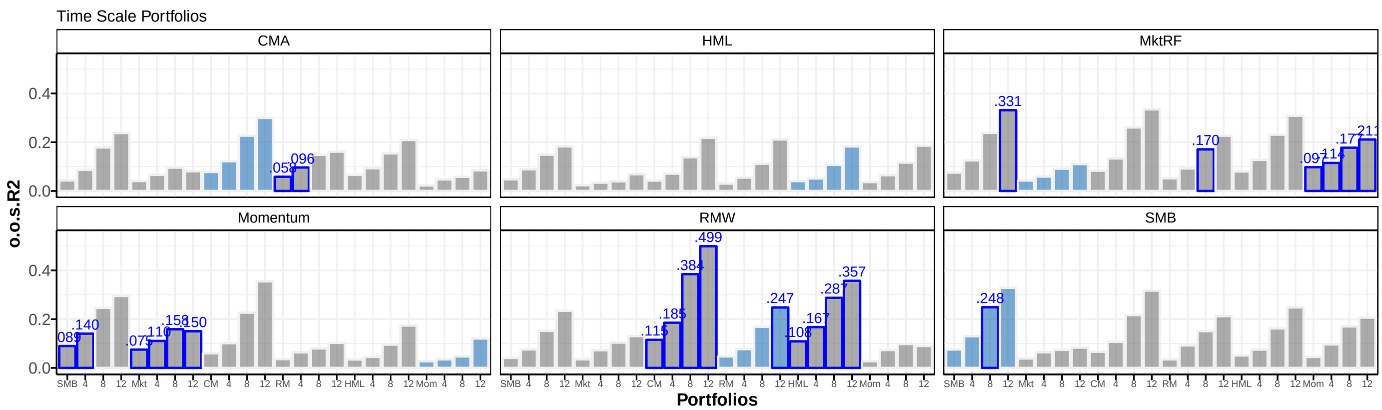
We construct *nominal* risk factor return forecasts within each rolling window for several horizons k separately. Nominal forecasts are fitted values of factor returns on past dividend yields and past returns. The sample window is truncated for each t depending on k, so that the effective windows for nominal forecasts are $\tau_m(t,k) := [t-(T_M-k), t]$. In the main test results reported, we reduce the window for all returns and forecasting series to that of the maximum horizon $\overline{k}$ nominal forecast series, $\tau_m(t,\overline{k})$. More nuanced procedures do not improve the performance significantly.

Nominal forecasts should not be confused with the forecasts made using the decomposition, which are the basis of the o.o.s. tests. The derived forecasts perform significantly better than the nominal forecasts used as inputs because the nominal forecasting procedures exploit none of the Markov structure of equilibrium asset returns.

4.2 Sample Window and Mixing

A distinction is made between the window size T_m and the mixing times $N(\epsilon_0)$ for threshold $\epsilon_0 >$ 0. As a benchmark, the fixed window size T_m proxies for the stationary mixing times $N_t(\epsilon_0) =$





$N(\epsilon_0, R_{t-T_m,t})$. The mixing time at t is defined

$$N_t(\epsilon_0) := \min_s \left\{ \mathbb{R}_+ \ni s \geq t : \max_{x \in X} \parallel \hat{h}_{t,t+s}(x) - \mu_0(x) \parallel \leq \epsilon_0 \right\}$$

In words, the mixing time is the shortest amount of time it can take for the maximum total variation distance between rows of the transition matrix to be within a threshold of size ϵ_0. Using the convergence results from proposition (7.1) part (II), we can bound the mixing time

$$N_t^*(\epsilon_0) = \log(\zeta_{0,t})^{-1}[\log(\epsilon_0) - \log(\chi_t)] \tag{4.2}$$

where $\chi_0 \mu_0 = \hat{h}_{0,0}$ is the initial statistical likelihood ratio of the conditional to the unconditional distributions. Changes in the window size T_m are justified by changes in $N_t(\epsilon_0)$, which for fixed $\epsilon_0 > 0$ varies through estimates of χ_t and $\zeta_{0,t}$. We verify ex-post that we cannot distinguish the sequence of mixing times $N_t(\epsilon_0)$, $t \in [0, T]$ from a covariance stationary process.

It is also possible to choose the window size in each period $T_m = T_m(t)$ to minimize the ℓ^2-distance between the mixing time estimate for that period and a fixed target mixing time $N_0(\epsilon_0)$. Here, the mixing time estimator is formally a transformation of a random sample. This objective corresponds to a well-defined extremum estimator. However, empirically we find this step produces very little movement in the window size $T_m(t)$ over t.

4.3 Data

For priced risk factors, we use monthly returns data on the Fama-French three-factor, Fama French three factor plus momentum (Carhart), and Fama-French 5- factor models. The Fama-French three factors are the Market, Value (HML) and size (SMB), rebalanced annually. Details are in Fama and French (1993). Momentum is constructed from a portfolio long 2-12 month winners and short 2-12 month losers ranked in the cross-section and truncated at the 30% and 70% percentiles. Each reported momentum sorted portfolio is an average of small and large-cap momentum stocks. Momentum is rebalanced monthly.

We consider several cross sections of test assets including the 25 size-BTM portfolios, the 25 size-BTM plus 10 Momentum portfolios, the 25 size -operating profit portfolios and the 32 size-BTM-OP portfolios. The size-BTM portfolios are annually rebalanced and are comprised of the intersection between five market-cap sorted stocks with five book-equity to market-equity (BTM) sorted stocks. The size-BTM plus Momentum add 10 portfolios sorted on 2-12 month returns cross-sectional rankings. The size-OP portfolios sort annually on operating profits (OP): "annual revenues minus COGS, interest expense, and SG&A", normalized by trailing book-equity, and intersect with size. The factor returns data and the test asset returns data are obtained from Ken French's website.[9]

We use predictor variables from Goyal and Welch (2008). Data are available on Amit Goyal's website.[10] We use monthly data for the rolling average 12-month dividends, the rolling average 12-month earnings, and the index level for the $S\&P500$. Data are from 1926-2016. Monthly value-weighted and equal- weighted index total returns and ex-dividend returns over 1926-2016 are from CRSP. 18 portfolios sorted by cash flow to market capitalization and 18 sorted by dividend to market capitalization over the same period are from Ken French. We form three high-low portfolios for cash flow and three for dividends, leveraging the cash flow spread at denary, quinary and tertiary scales. The breakpoints are available on Ken French's website.

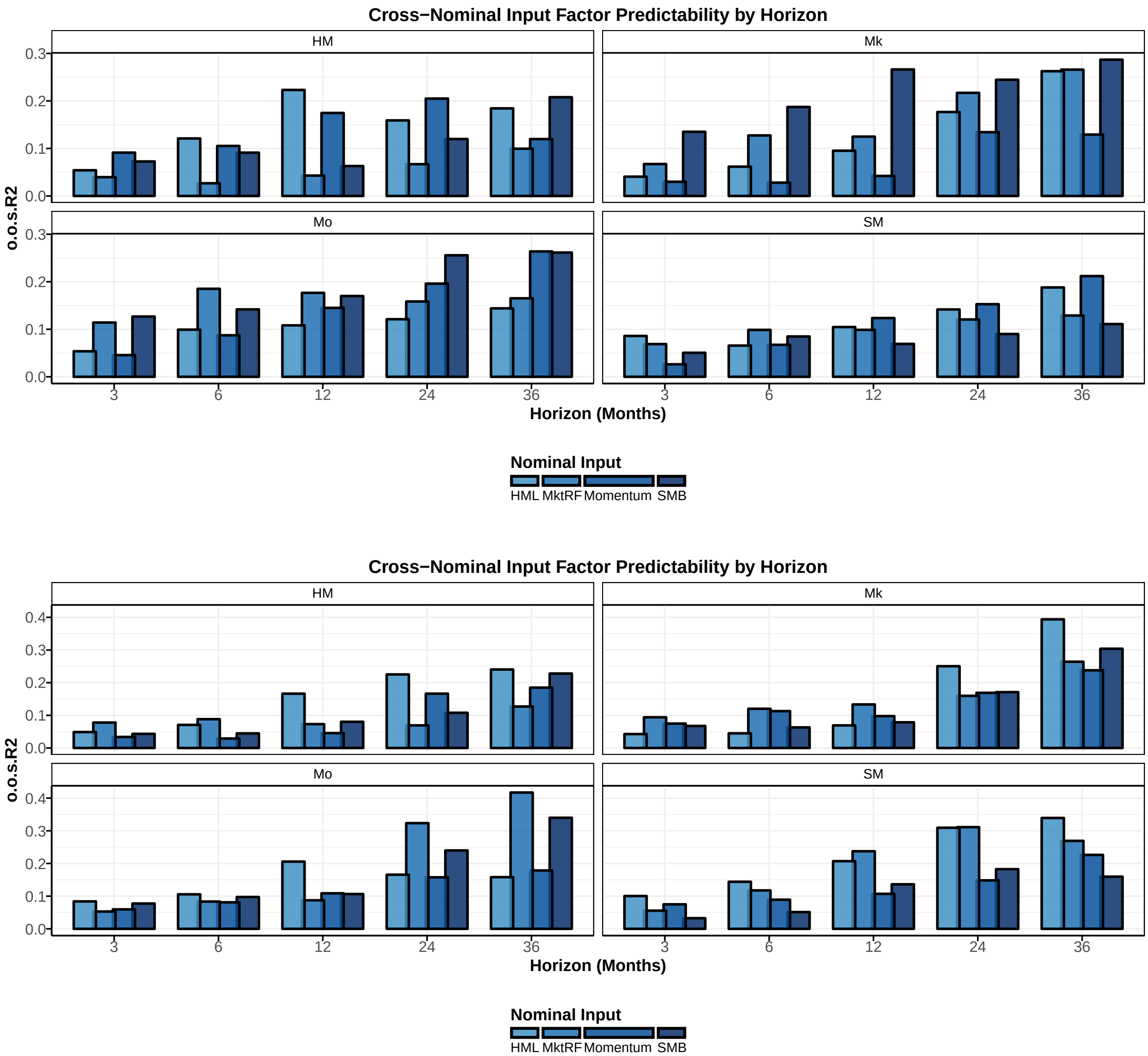
5 Empirical Results

5.1 Latent Factor Dynamics

Persistence estimates for the leading and penultimate latent factors are reported in Table 1. We report coefficient estimates from a first-order autoregression for the demeaned factor processes. The

[9] http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html
[10] http://www.hec.unil.ch/agoyal/

Figure 2: Predictability Term Structures for the Carhart Factors



(a) Out of sample forecast performance for each of the Carhart four factors. Value is forecasted by the volatility of the spectral gap. The spectral gap is calculated up to time t using data from $[T_m - t, t]$ and used to forecast returns for various $t + k$, including returns between periods $t + 1$ and $t + n + 1$ for $n = 1, 2, 4, 8, 12$. Top panel has $T_m = 8$yrs. Lower panel has $T_m = 12$yrs. Carhart Factors are the Fama-French Market, Value and Size factors plus the Momentum factor (UMD). Quarterly data $Q1$ 1967 to $Q3$ 2015 are compounded monthly factor returns from Ken French's website.

leading factor exhibits no predictable deviation from its mean. In contrast, the penultimate factor is highly persistent. Fluctuations around the mean of the penultimate factor are predictable, stationary and statistically significant. GMM standard errors adjust for serial correlation. The sharp contrast in predictability between the leading and penultimate factors is consistent with the predictions of the Markov asset pricing model with nonzero stationary mixing times.

Dynamics of the factors of realized returns are plotted in Figure 2. Figure 2 in combination with Table 1 restate the findings in Alvarez and Jermann (2005) that the bulk of pricing kernel variation comes from the permanent factor. Transitory factors are plotted in Figure 3, which shows the time series of the time-varying expected return factors. Figure 1 charts the empirical densities of conditional risk prices. Figure 4 plots the dynamics of the empirical spectral gap. The spectral gap is strongly countercylical. The interpretation of Figure 4 is that *volatility increases in bad times, but so does the concentration of volatility on the permanent factor*.

5.1.1 Latent and Conventional Factor Composition

The martingale factor captures 84% of the common time series variation in asset returns. The market and the martingale factor are almost identical. The conventional value (HML) factor provides a striking contrast to the market factor. Table 8 breaks down the market and HML loadings on the latent expected return components. Variation in returns to the value factor load significantly on the leading expected returns factor, which drives transitory predictable fluctuations in mean returns. The conventional market factor does not significantly load on the expected return factors.

5.2 Forecasting

We use the common sources of transitory variation to predict factor returns out of sample. Results are reported in Table 2. The spectral gap (lagged for predictive regressions) forecasts the market returns with an o.o.s. R^2 of 3.8% quarterly, 6.5% semiannually, 10.8% annually and 20.3% biannually. Other variants of the spectral gap, including the common negative exponential transform, perform similarly but no better. *cay* performs relatively well, giving an out of sample R^2 of 7.6% on an annual basis, while the dividend yield generates forecasts with o.o.s. R^2 of only 2.9%.

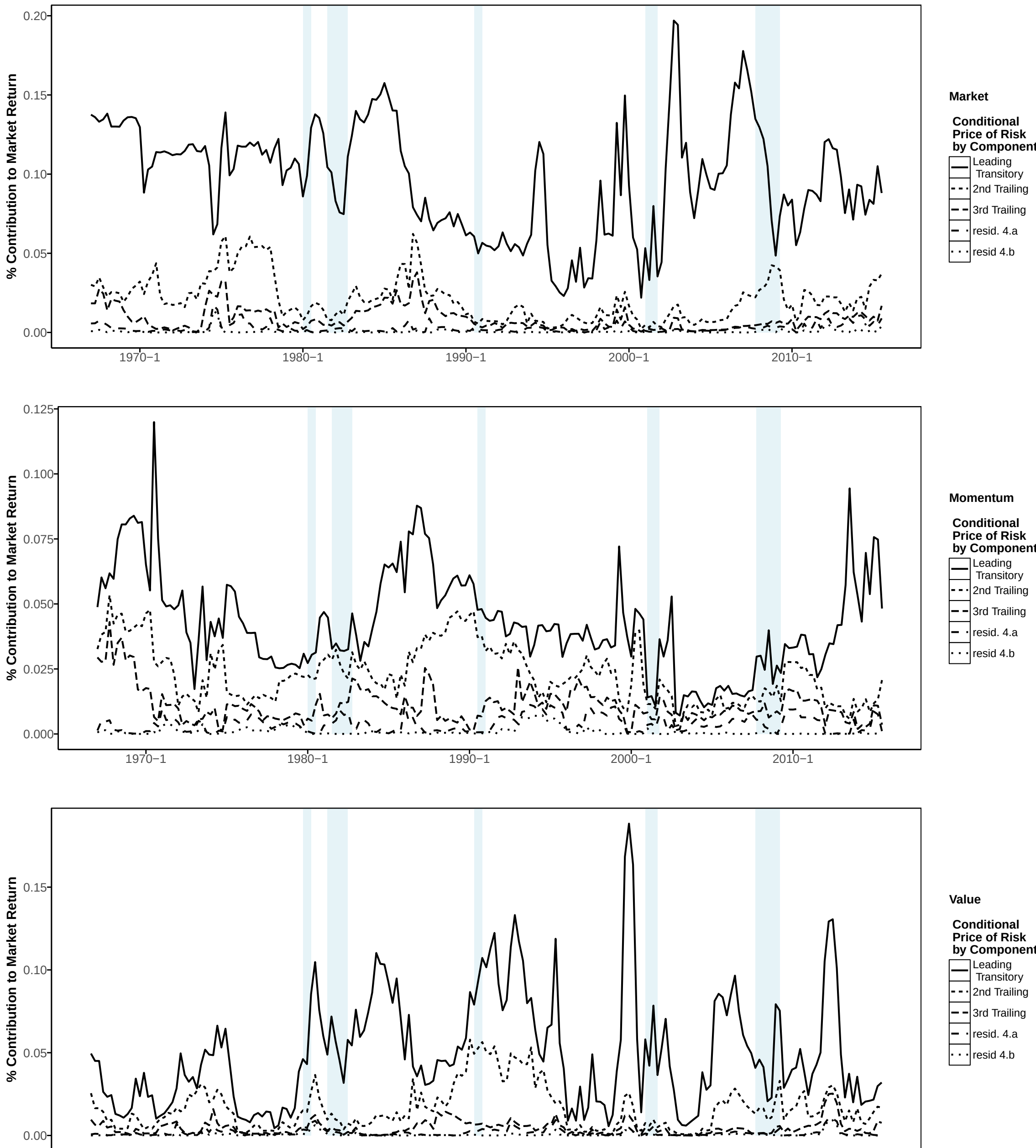
This latter number contrasts with conventional wisdom because it is an out-of-sample estimate. The original predictability studies by Campbell and Shiller (1989) were estimated in sample (see Goyal and Welch, 2008). The roughly 10% R^2 reported by Cochrane (2008) is calculated using the Campbell-Shiller decomposition, which jointly restricts cash flow and discount rates by design.

Table 1: Market Return Predictability by Dividend Yield, *cay* and the Spectral Gap

Panel I shows out of sample predictability of market returns by the spectral gap. The spectral gap is the difference between the conditional volatilities of the permanent and first transitory factors, measured by the second conditional eigenvalue of the empirical decomposition of asset returns. II. shows the out of sample predictability for the dividend yield. Panel III. shows out of sample prediction statistics for *cay*. The spectral gap *excluding* non-market volatility is given in panel IV. The 12-month moving average of monthly dividends, the market index level and *cay* are from Goyal and Welch. Fama and French factor returns quarterly are from Ken French. Data are quarterly from 1967 $Q1$ to 2015 $Q3$.

		Full Sample Estimates		Out of Sample Statistics	
Predictor	k (quarters)	coefficient $\widehat{\phi}_{j,k}$	$t : H_0(0)$	R^2	adj.R^2
I.	1	2.39**	2.770	0.038	0.033
Spectral Gap	2	4.51**	3.650	0.065	0.060
	4	8.01***	4.805	0.108	0.104
	8	14.76***	6.893	0.203	0.199
II.	1	64.066	1.256	0.008	
Dividend	2	143.25*	1.929	0.019	
Yield	4	243.69*	2.385	0.029	
	8	355.27**	2.555	0.034	
III.	1	47.51*	1.944	0.019	
cay	2	101.01**	2.821	0.040	
	4	196.51***	3.974	0.076	
	8	388.12***	5.901	0.157	
IV.	1	1.48*	2.136	0.023	
Market Gap	2	2.90**	2.895	0.042	
	4	5.29**	3.896	0.074	
	8	9.15***	5.095	0.122	

Figure 3: Time Varying Components for the Market, Momentum and Value



(a) Dynamics for the components of common variation for the market, momentum and value expected returns, expressed as a percentage of the total variation of factor returns. Expected return variation comprises roughly 20% of the variation in total returns. Expected return variation comprising market returns contribute an average of less than 14% of variation in total returns, and comprising momentum returns contribute an average of less that 5% of the variation in total returns. Quarterly data $Q1$ 1967 to $Q3$ 2015. Fama French and Carhart factor model returns are from Ken French's website. NBER recessions are in blue.

Table 2: Persistence of first and second components of realized returns.

The first component is not predictable, while the second and trailing components are predictable when expected returns are time-varying. For the j'th component of returns we report the estimated autoregressions

$$x_{t,j} = \phi_{0,j} + \phi_{1,j} x_{t-1,j} + u_{t,j}$$

The components are obtained over rolling $T_m = 15$ year samples of realized returns supplemented with the forecast errors from dividend yields over several horizons. A singular value decomposition of the generalized covariance matrix orders orthogonal components by contribution to variation in realized returns and lower frequency variation from surprises to historical forecasts. We report autoregressions over the full sample for components $j = 1, 2$.

	coefficient	estimate	$s.e$
Leading Factor	$\widehat{\phi}_{0,1}$	0.312	0.624
	$\widehat{\phi}_{1,1}$	0.089	0.072
Trailing Factor	$\widehat{\phi}_{0,2}$	-0.059	0.032
	$\widehat{\phi}_{1,2}$	0.649^{***}	0.055

(a) The penultimate component of the realized returns is the leading component of expected returns. Returns data are quarterly from $Q1$ 1967 to $Q3$ 2015. Fama French and Carhart factor model returns are from Ken French's website.

The forecasts in Cochrane (2008) are necessarily evaluated in sample using fitted values from a vector-autoregression (VAR).

In fact, our evidence corroborates the importance of an aggregate measure of dividend yields for forecasting risk premiums. While the short-horizon performance of the dividend yield is poor out of sample, explained variation grows monotonically, resulting in the stylized fact that low yields today (high prices) are followed by low returns tomorrow. As emphasized by Cochrane (2011), variation in the dividend yield corresponds entirely to variation in mean returns (at the aggregated level), and more so at the long end.

Our benchmark nominal forecast variable is the dividend yield for the CRSP value-weighted index. We construct a rich information space for our decomposition by calculating optimal return forecasts using the dividend yield, separately for lags of 2^k, $k = 0, 1, 2, ..., k^+$. Surprises from different lag lengths are calculated separately. An $AR(1)$ generates signals that are all deterministic functions of each other, eliminating signals relevant over different time scales when they exist.

We also find that constructing nominal forecasts using the CRSP ex-dividend value-weighted index return translates into significantly poorer performance. Value in this case cannot be forecasted. The explained variability in market returns does not exceed 10% out of sample at any horizon in this case. Interestingly, an equally-weighted dividend yield used to generate nominal forecasts also

Table 3: HML Return Predictability

(a) Value return predictability, assessed over the entire sample using a time-series of forecasts made out of sample. The predictive regression is

$$R_{t+k,j} = a_0 + a_1 G_j(\widehat{\zeta}_t) + \varepsilon_t$$

where $G_j(\hat{\zeta}_t \in \{\sigma_1/\overline{\sigma}, \sigma_2\}$. The rolling T_m samples are augmented with information contained in the nominal forecast errors at lags 2^k, $k \in [0,5] \cap \mathbb{N}$. Nominal inputs are listed. Data are from $Q1$ 1967 to $Q3$ 2015 from Ken French, Goyal and Welch (2008) and CRSP.

		Full Sample Estimates			Out of Sample Statistics	
Nominal Input	horizon (quarters)	predictor	estimate	$t: H_0(0)$	R^2	adj.R^2
DP & 10-1 CF Portfolio	1	σ_2	4.97	1.339	0.009	0.004
		$\sigma_1/\overline{\sigma}$	−7.37	−1.272	0.008	
	2	σ_2	24.29 ***	4.505	0.096	0.092
		$\sigma_1/\overline{\sigma}$	−38.41 ***	−4.573	0.099	
	4	σ_2	40.14 ***	5.229	0.127	0.122
		$\sigma_1/\overline{\sigma}$	−67.46 ***	−5.702	0.147	
	8	σ_2	46.41 ***	4.31	0.092	0.087
		$\sigma_1/\overline{\sigma}$	−80.96 ***	−4.887	0.115	
	12	σ_2	46.15 **	3.830	0.075	0.070
		$\sigma_1/\overline{\sigma}$	−84.34 ***	−4.560	0.104	
DP	1	σ_2	28.84	1.501	0.012	0.006
		$\sigma_1/\overline{\sigma}$	−79.50	−1.427	0.010	–
	2	σ_2	125.66 ***	4.508	0.097	0.092
		$\sigma_1/\overline{\sigma}$	−370.56 ***	−4.598	0.100	–
	4	σ_2	191.03 ***	4.768	0.108	0.103
		$\sigma_1/\overline{\sigma}$	−576.42 ***	−4.993	0.117	–
	8	σ_2	214.77 **	3.831	0.074	0.069
		$\sigma_1/\overline{\sigma}$	−666.78 ***	−4.131	0.085	–
	12	σ_2	122.28 *	1.941	0.020	0.015
		$\sigma_1/\overline{\sigma}$	−383.66 **	−2.107	0.024	–

results in diminished efficacy. Unsurprisingly, size is hit least by this distinction. At the three-year horizon, the value weighted dividend yield nominal input predicts 34.7% of the variation in returns to the SMB portfolio. Using the equal -weighted yield generates an R^2 of 30.5%.

Several features of the value forecasts reported in Table 3 are striking. First, at the annual horizon, the variation in value (HML) returns is 14.7% predictable, and semiannually, 9.9% predictable. Because HML is a priced risk factor, these percentages measure the fraction of the variability in value premia that are predictable. Said another way, they capture the time-varying price of risk for exposure to innovations in the HML replicating portfolio. Second, unlike market predictors, which, as emphasized by Cochrane (2008, 2011), increase in explanatory power monotonically with horizon, the value predictors have a half-life of about a year. In every case, the two year prediction is less effective than the one year prediction.

The shape of the term structure of the time-varying value premium is robust to the choice of predictor as long as the predictor works. We report two normalizations of the second moment of the spectral gap that forecast HML returns. Every other priced risk factor we forecast has an upward sloping term structure of predictability, making the value premium unique among priced sources of risk in U.S. equity markets. This finding is also consistent with the findings in Koijen, Lustig and Van Nieuwerburgh, showing value returns load on the transitory variation picked up by the Cochrane-Piazzesi factor.

From Table 4. the momentum factor returns are predictable. The first two years explained variation is low, but accelerates after the second year to meet or exceed market predictability after the 3-year horizon. Momentum predictability jumps from 6.4% at one-year to 16.1% and 31.4% at the two- and three- year horizons, respectively. The curvature of momentum predictability is increasing and convex and in this way stands out from the nearly linear term structure of market predictability. Standard market predictors do poorly with momentum, reported in Table (10). Earnings to price, cay and the dividend yields all predict less than 3% of the variation in momentum return at any horizon.

Figure (1) contrasts the term structures of return predictability for the value (HML), Momentum (UMD), market and size (SMB) factor replicating portfolio returns. Explained variation grows monotonically in forecast horizon for the market, momentum and size. Time-varying expected returns to value are predictable in the short-run, but become negligible in asymptotic forecasts. The market term structure is nearly linear. The momentum term structure is a locally increasing convex function of horizon while value predictability is a concave function with a local maximum at the one-year horizon. Forecasts are given by the lagged spectral gap normalized to match the fit of the realized factor returns on rolling historical $T_m = 15$ year samples.

These forecasts meet or exceed existing predictors in the literature to our knowledge, with the

Table 4: Market Return Predictability

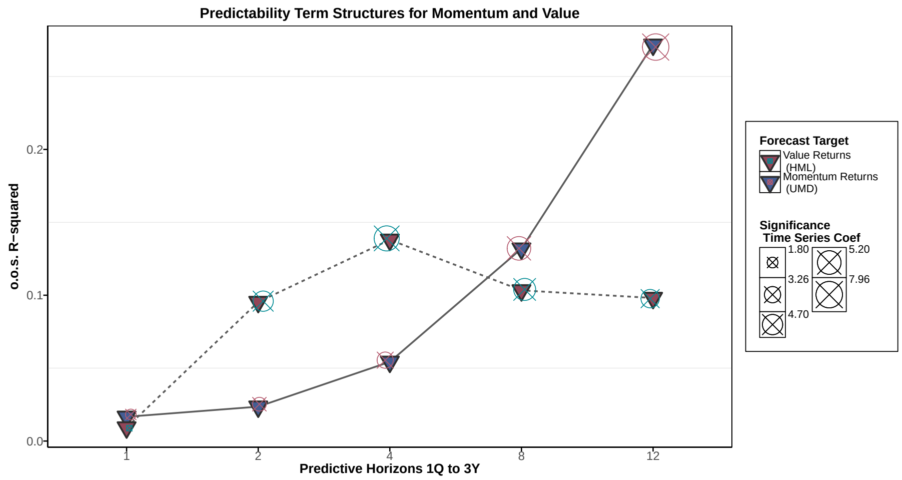
(a) The spectral gap predicts market returns. Predictability regressions are calculated over the entire sample. The predictor series $\hat{\zeta}_t$ is constructed using data up to t for forecasts of returns at $t+k$ or $t+1+k$, $k \in \{1, 2, 4, 8\}$ on rolling windows of T_m =15yr histories. The predictive regression is

$$R_{t+k,j} = a_0 + a_1 G_j(\widehat{\zeta_t}) + \varepsilon_t$$

where $G_j(\hat{\zeta}_t)$ is a functional of the spectral gap time series that can vary only by portfolio, indexed by j. The rolling T_m samples are augmented with information contained in the forecast errors from forecasts made at various lags. The variable $\widehat{\zeta}$ is shorthand for the empirical estimate of the spectral gap. Tr_1 is the (non-normalized) largest contribution to the trace of the covariance matrix. Monthly dividend yields, Fama -French 3-factor and Carhart model returns and *cay* data are from $Q1$ 1967 to $Q3$ 2015 from Ken French, Goyal and Welch (2008) and CRSP.

		Full Sample Estimates			Out of Sample Statistics	
horizon (quarters)	target	predictor	estimate	$t : H_0(0)$	R^2	adj.R^2
1	R_M	$\widehat{\zeta}$	2.546	2.983	0.044	
		$(Tr)_1$	1.354	1.893	0.018	–
		$\widehat{\zeta} + (Tr)_1$	1.169	2.683	0.036	0.031
2	$R_{M,2}$	$\widehat{\zeta}$	3.827	3.065	0.047	
		$(Tr)_1$	2.357	2.257	0.026	
		$\widehat{\zeta} + (Tr)_1$	1.879	2.956	0.044	0.039
4	$R_{M,4}$	$\widehat{\zeta}$	7.708	4.486	0.098	
		$(Tr)_1$	4.682	3.249	0.054	
		$\widehat{\zeta} + (Tr)_1$	3.778	4.320	0.091	0.086
8	$R_{M,8}$	$\widehat{\zeta}$	13.181	5.626	0.151	
		$(Tr)_1$	7.013	3.568	0.067	
		$\widehat{\zeta} + (Tr)_1$	6.167	5.142	0.129	0.124
12	$R_{M,12}$	$\widehat{\zeta}$	18.307	6.480	0.198	
		$(Tr)_1$	9.032	3.791	0.078	
		$\widehat{\zeta} + (Tr)_1$	8.494	5.800	0.165	0.160

Figure 4: Value and Momentum Predictability Term Structures



(a) Forecasts are given by the lagged spectral gap normalized to match the fit of realized factor returns on the rolling historical $T_m = 15$ year samples. The size of circular nodes corresponds to the t -statistic of the time-series coefficient estimates obtained on the full sample. The Fama-French Market, Value and Size factors are quarterly from $Q1$ 1967 to $Q3$ 2015, compounded monthly returns from Ken French.

exception of Kelly and Pruiit (2013) in the case of market returns. Kelly et. al find an o.o.s R^2 of 13% for the market, omitting windows in a neighborhood the forecast target.[11] In the case of momentum returns. Huang (2016) reports a monthly out of sample R^2 of 0.5% for momentum returns using the cross-sectional dispersion of moving average annual returns (the "momentum gap"). This compares roughly to the spectral gap's forecast R^2 of 6.4% when scaled to the one-year horizon.

Using the CRSP value-weighted index ex-dividend as the nominal input produces limited forecasting power but has interesting implications for the market. The market o.o.s. R^2's of 2.8% and 6.9% at the 1 - and 2 -year horizons accelerate to 16.3% at 3-years. At 3 years we report a significant time-series coefficient point estimate 15.54 with t- statistic 5.76%. This suggests capital gains have a non-negligible role forecasting aggregate returns over longer horizons. Using the equally weighted CRSP ex-dividend index to construct nominal forecast inputs unsurprisingly allows our model to forecast size, although not quite as well as with nominal dividend yield inputs. The former and latter 3-year o.o.s. R^2 for size are 30.05% and 34.7% respectively.

We reproduce the analysis shifted forward an extra period to address potential concerns about systematic measurement error. If prices are measured with error in a persistent direction then returns from contiguous intervals are spuriously correlated. This problem is not as likely in liquid stock markets as over the counter or emerging markets contexts. This is confirmed in table 10 for the

[11] Using future observations implies these forecasts cannot be implemented in real time. However, they may be a better measure of the forecasting ability of a theoretical investor within the model. Kelly et. al can also be implemented in non-Markovian settings.

Table 5: Size (SMB) and Momentum (UMD) Out of Sample Predictability

(a) Predictability regressions are calculated over the entire sample. The predictor series $\hat{\zeta}_t$ is constructed using data up to t for forecasts of returns at $t+k$ or $t+1+k$, $k \in \{1, 2, 4, 8, 12\}$. The predictive regression is $R_{t+k,j} = a_0 + a_1 \widehat{\zeta_t} + \varepsilon_t$, where $\widehat{\zeta_t}$ is the spectral gap time series. The rolling $T_m = 15$yr samples are augmented with nominal forecast errors from forecasts made at various lags. $Q1$ 1967 to $Q3$ 2015 data from Ken French, Goyal and Welch (2008) and CRSP.

			Full Sample Estimates		Out of Sample Statistics
Portfolio	quarters k	variable	$\hat{a}_1$	$t : H_0 = 0$	R^2
Size (SMB)	1	$\widehat{\zeta}$	1.067	2.252	0.026
	2	$\widehat{\zeta}$	1.819	2.787	0.039
	4	$\widehat{\zeta}$	4.175	4.507	0.098
	8	$\widehat{\zeta}$	8.495	6.363	0.185
	12	$\widehat{\zeta}$	12.874	7.485	0.248
Momentum (UMD)	1	$\widehat{\zeta}$	−0.544	−1.305	0.009
	2	$\widehat{\zeta}$	−1.207	−2.070	0.022
	4	$\widehat{\zeta}$	−2.939	−3.566	0.064
	8	$\widehat{\zeta}$	−6.209	−5.854	0.161
	12	$\widehat{\zeta}$	−10.395	−8.820	0.314

case of momentum returns, where the annualized point estimate for forecasts from $t+1 \mapsto t+1+k$ is equal to the point estimate obtained from $t \mapsto t+k$ up to one significant digit. We calculate the staggered forecasts for each of the portfolio returns we analyze and find no significant discrepancies.

The patterns of predictability extend to the full sample of available U.S. equity data, beginning in 1926 when the NYSE emerged as a dominant national exchange, among other reasons.[12]. The term structure of momentum is convex and increasing beyond the 1-2 year horizon. Value is concave and decreasing after the 1-year horizon, both as in the primary sample. The long-end of the value term structure decreases more slowly and from a slightly lower rate of predictability, maximized at an out of sample R^2 of 11.2%.

Table 6: Average and standard deviations of estimated conditional means for the Market. Annualized returns.

Variable	Standard Deviation	Unconditional Mean	Forecast Sharpe Ratio
Conditional Means	9.924	7.084	0.714
	9.742	7.084	0.727
	9.410	7.084	0.753
	9.064	7.084	0.782

5.3 Carhart Model Factor Replication

The timing portfolio returns for a particular factor are not identical to the sequence of conditional forecasts obtained for that factor. This is because the conditional forecasts are constructed by exploiting information in the errors from past nominal forecasts that are not replicable in the space of the factors' test assets. Nonetheless, a projection of the conditional forecasts onto the test asset space improves the efficiency properties of the factor replicating portfolios.

The timing portfolio for HML returns can be split into two components. The raw HML Sharpe ratio over the full sample is 0.177, while the first HML timing component achieves 0.230 and the second component achieves 0.283. Results are stated in Table 5. The hybrid timing component replicates HML returns but with lower volatility, producing a Sharpe ratio of 0.301 over the full sample. HML does better prior to the 2007 financial crisis with a Sharpe ratio of 0.190. The timing portfolio components pre-crisis generate Sharpe ratios of 0.259 and 0.223 for the penultimate and trailing components, respectively. The overall timing portfolio does less well excluding the crisis, but we still detect a significant improvement. The trailing component - both the trailing factor and the weight of the HML replication on that factor - drives performance of the HML timing portfolio

[12]Brown, Mulherin, and Weidenmier (2006) discuss the pre-1926 history of the stock exchange industry in the U.S.

during the financial crises.

When several components are predictable, a simple diversification argument suggests the hybrid returns will be more efficient than either of the individual component replicating returns, which is part of what we see here. This intuition is delicate: theory tells us that investors are willing to sacrifice some mean-variance efficiency for inter-temporal hedging opportunities represented by the HML factor. We should expect to see a diversification benefit from combining the components within the HML factor, as we do, but it is not obvious what to expect combining components from different factors with the market.

The momentum factor timing portfolio is concentrated on a single expected return factor. The Sharpe ratio rises from 28% to 32.8%. The Sharpe ratios, means and volatility are significantly estimated. Moreover, we reject the null hypothesis that the difference in the Sharpe ratios is zero. Results are reported in Table 5.

The SMB timing portfolio represents an improvement in mean-variance efficiency of 63%. Statistics are reported in Table 5, along with the standard factor replicating portfolio statistics for reference. Interestingly, the incremental gains in efficiency do not arise from more precise estimates of the conditional price of risk for the size factor. This and related findings are the subject of the proceeding section.

5.4 Gains in Dimension

Using the full BTM 25 xsection, we compare the dimension of the representation of 97.5% variation threshold we find the inclusion of the nominal forecasts increases the dimension between one and three degrees. Heuristically, the dimensions can be thought of as state variables.

The fact that forecasts improve for horizons inside of four years as we bring the window T_m down from $T_m = 15$ to $T_m = 12$ and $T_m = 8$ is not surprising if the filtering procedure is working as expected. These choices represent an increase in the mixing time holding a threshold fixed. As a result, higher frequency fluctuations are emphasized at the expense of lower frequency transitory variation.

5.5 Summary of Empirical Method

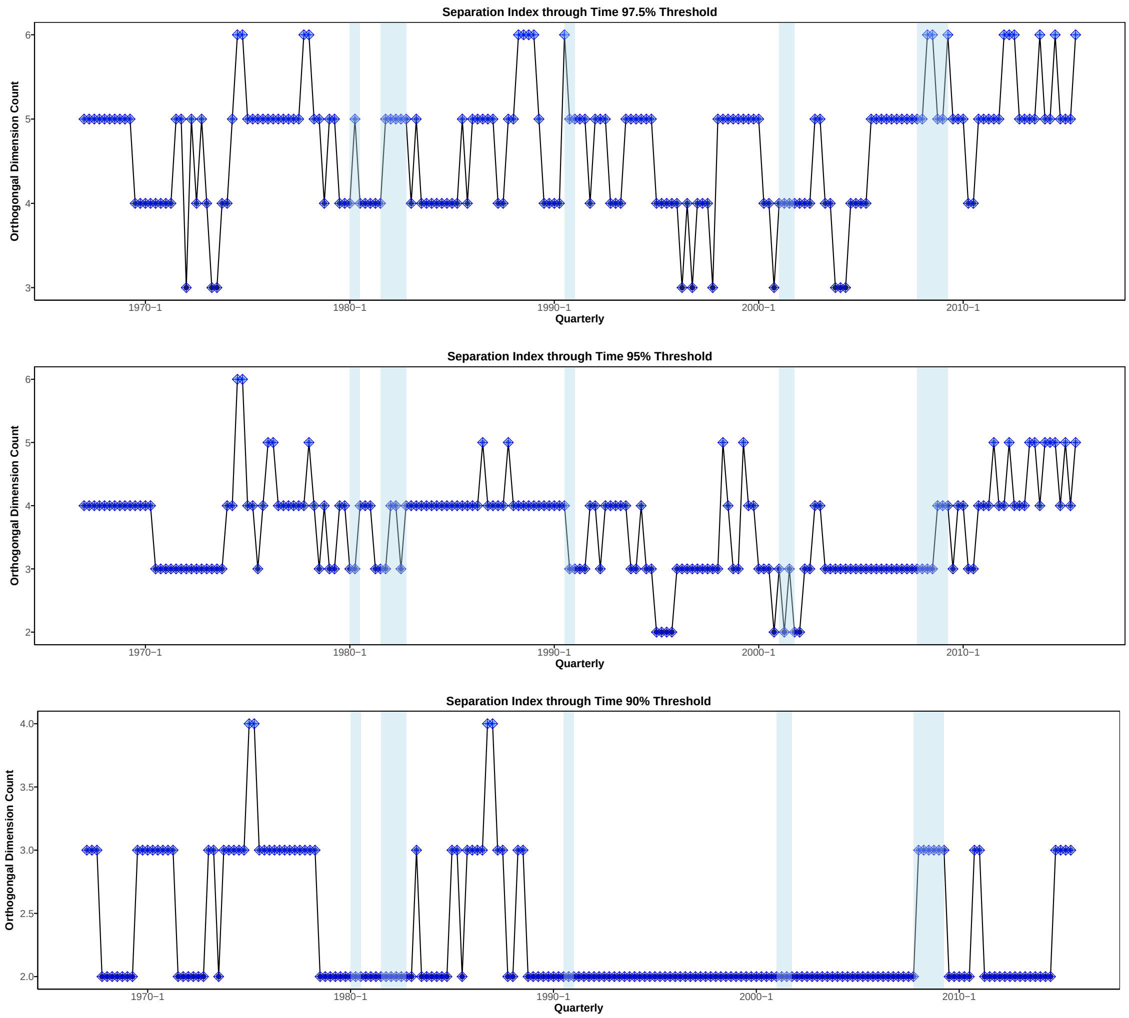
To summarize, pick a threshold $\epsilon_0 > 0$ for the mixing times $N_t(\epsilon_0)$ and target $N_0(\epsilon_0)$. This choice implies some sample window size T_m such that the mixing time evaluated on the window $\tau_m(t)$ is near $N_0(\epsilon_0)$. Within each sample window $\tau_m(t)$, nominal forecasts are fitted for each of the risk

Table 7: Sharpe Ratio Comparisons
Value, Momentum, and Size Timing Portfolios

Top: Full sample and pre- 2007 financial crisis Sharpe ratios and t- statistics for HML and HML timing portfolio by component. Statistics correspond to pre-crisis estimates. Mid: Sharpe ratios for momentum, the first two components of the expected return factors weighted corresponding to their contribution to momentum, and the momentum timing strategy. Lower panel: size, size timing portfolio and market Sharpe ratios. Test assets are the Fama-French FF25 Size/BTM plus 10 momentum portfolios. Factor data are the FF3 plus Momentum factor returns. Data are quarterly from 1927 $Q1$ to 2015 $Q3$.

Strategy	Value	Penultimate	Trailing	Value Timing
Monthly SR				
Full Sample	0.177	0.230	0.283	0.301
Pre-2007	0.190	0.259	0.223	0.243
		(3.403)	(2.920)	(3.191)
Strategy	Momentum			Momen. Timing
Monthly SR				
Full Sample	0.2801	0.3283	0.3009	0.3284
				(4.055)
Strategy	Size			Size Timing
Monthly SR				
Full Sample	0.1352			0.2138
				(2.941)

Figure 5: Orthogonal Dimensions Count 85% – 97.5% -thresholds



(a) The time series of the number of orthogonal dimensions needed to explain each threshold percentage of the variation from that date on a rolling historical 15-year window. Quarterly Fama -French 3-factor and Carhart model returns data are from 1967 $Q1$ to 2016 $Q4$. NBER recessions are in blue.

factor returns and combined with the contemporaneous realized returns to create an augmented panel. Calculate a singular value decomposition (SVD) of the series in this panel and extract the significant dimensions. Transform the diagonal elements using $\Lambda \mapsto C(T_m)(1-\Lambda)^{-1}$. Forecast the component dimensions $t \to t+1$ by linear autoregressions (do not include the first component in general). Regress the target portfolio returns on the components in window $[T_m - t, t]$ and keep the coefficient estimates. The latent component forecasts give forecasts for the target portfolios by weighting the component forecasts with the portfolio's coefficient estimates.

The specialized case where the best predictor is the spectral gap follows from equilibrium asset pricing theory, as described in the modeling sections. However, statistics contained in the trailing solutions to the eigenvalue problems are myriad. The above described predictive routine applies generally, as long as a first-order transition representation is justified.

6 Conclusion

We extract predictable components from priced risk factors and show these components can be used to improve allocation efficiency in real time. Latent, transitory components of factor risk prices contain valuable information about near and medium term evolution of the state of the economy. Novel evidence connecting time-series predictability and time-varying risk prices for common factors in equity markets is provided.

The *spectral gap* measures the fraction of volatility concentrated on long-run shocks, varies through time and predicts market returns. We use these factors to predict the market as well as several portfolios including value, size and momentum. For the market and value, we find out-of-sample R^2's of 10.8% and 14.7% respectively, for annual returns. For size, annual returns are 13.7% predictable out of sample. Momentum predictability is low at the short end, but reaches nearly 30% at the three year horizon. More generally, we document a heterogeneous term structure of predictability across types of portfolios. Most strikingly, the value premium predictability is concave and most predictable at the one-year horizon, while momentum predictability is convex. This finding contributes an interesting wrinkle to the value and momentum "puzzle" regarding the high average returns but significant negative correlations of these series.

The discrepancy between the variation of the permanent and transitory components is a well studied object that emerges in a range of contexts in nature. Because the general representation for the concentration of volatility on the leading factor is measured by (one-minus) the difference between the first two eigenvalues of the Laplacian of a dynamical system, it is called the *spectral gap*. In asset markets, the spectral gap and related statistics are found by specializing a known decomposition of Markov transition dynamics. This decomposition identifies a factor structure of expected returns from the permanent and transitory fluctuations. The *common* sources of transitory fluctuations comprise the expected return *factors*. Naturally, the expected return factors are also the common sources of time-series predictability.

We begin with priced risk factors so we know the benchmark portfolios are priced, and find that incremental efficiency gains associated with identifying the expected return factors are also priced. Incremental gains from timing HML are compensation for exposure to business - cycle news, while timing Momentum compensates for permanent wealth shocks. Timing gains are not priced for size. We isolate the component of market returns that are priced but find a residual process with significant variation that has no cross-sectional pricing power.

Our findings suggest investors are able to allocate capital with more precision using past returns data alone than previously indicated. This distinction is most relevant in economies where we distinguish between revisions in state variables given a known, fixed model and economies where dynamical features require shocks to the model. For example, it is not only the case that the evolution of technology is uncertain and risky, but also the case that the way markets equilibrate is uncertain, particularly in the context of the given technological uncertainty. Because this problem immediately becomes high dimensional, the most natural device for managing it is high dimensional random variables in the form of random local transition dynamics. This modeling choice is simplified in practice because the distribution of the eigenvalues (and eigenfunctions) of the system are well studied and tractably summarize high dimensional information.

Our characterization of the latent components has implications for economic modeling. Rational agents making allocations optimally in an asset pricing model must be endowed with the spectral data. In models where data are initialized at a deterministic invariant distribution this distinction is immaterial. In more general models, incorporation of spectral data into the decision makers' information set can impact model implications.

The predictable component in market returns is related to slow moving cash-flow yields, placing restrictions on parametrizations of stochastic discount factor models. Two natural theoretical benchmarks are i.) persistent level shocks to growth under representative Epstein Zin preferences, and ii.) i.i.d. growth (i.e., output is a random walk) under representative ambiguity aversion. Hansen (2011) points out that ambiguity averse investors ex-post look like rational expectations investors if the equilibrium data generating process is the worst-case model. For plausibly indistinguishable models, the worst case model is the long-run risk model. The findings and methods in this paper suggest an opportunity to distinguish these stories in finite samples by looking directly at dynamics in the spectral data (after making some assumptions regarding the underlying equilibrium and its data generator). Given the factor structure of time-varying discount rates, surprise revisions to persistent premia are priced cross-sectionally (ex-post in finite samples) if the long run risk model is the true model, but not necessarily when the representative investor is ambiguity averse.

Our analysis revealed the importance of the countercyclical concentration of return volatility on permanent shocks - both as an economic concept and a convenient modeling object. This suggests a similar analysis can be fruitful in a variety of other asset markets. Because these analyses are cast in terms of objects that are common to many asset pricing models, new empirical evidence implies tractable restrictions to the set of plausible parametrizations of stochastic discount factor models.

References

[1] Fernando Alvarez and Urban J. Jermann, *Using Asset Prices to Measure the Persistence of the Marginal Value of Wealth*, 2005, Econometrica, Vol. 73, No. 6

[2] Theodore W. Anderson, *Second Order Moments of a Stationary Markov Chain with Applications*, Technical Report No. 22 U.S. Army Research Office and Stanford, February 1989

[3] Clifford S. Asness, Andrea Frazzini, Ronen Israel and Tobias J. Moskowitz, *Fact Fiction and Value Investing*, Journal of Portfolio Management, Fall 2015

[4] Clifford S. Asness, Andrea Frazzini, Ronen Israel and Tobias J. Moskowitz, *Fact Fiction and Momentum Investing*, Journal of Portfolio Management, Fall 2014

[5] Federico M. Bandi and Andrea Tamoni, *Business Cycle Consumption Risk and Asset Prices*, 2015, Working Paper

[6] Ravi Bansal and Bruce N Lehman, *Growth Optimal Portfolio Restrictions on Asset Pricing Models*, Macroeconomic Dynamics, 1, pp. 333-354, 1997

[7] Ravi Bansal and Amir Yaron, *Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles*, The Journal of Finance, Vol. 59, No. 4 (Aug.), 2004, pp. 1481-1509

[8] Alex Bloemendal Antti Knowles Horng-Tzer Yau Jun Yin, *On the principal components of sample covariance matrices*, Working Paper, Harvard Mathematics Department, February 14, 2015

[9] Michael Brennan and Alex Taylor, *Expected Returns and Risk in the Stock Market*, UCLA Anderson School Working Paper, 2016

[10] Jaraslov Borovicka, Lars Peter Hansen and Jose A Scheinkman, *Misspecified Recovery*, Journal of Finance, Vol.71, No.6, 2016

[11] Svetlana Bryzgalova, *Spurious Factors in Linear Asset Pricing Models*, LSE Working Paper 2014

[12] Svetlana Bryzgalova and Christian Julliard, *The Consumption Risk of Bonds and Stocks*, Working Paper, September 15, 2015

[13] Craig Burnside, *Identification and Inference in Linear Stochastic Discount Factor Models with Excess Returns*, Working Paper, June 2015

[14] Carhart, M. M., *On Persistence in Mutual Fund Performance.*, 1997, Journal of Finance, no 52 pg 57-82

[15] Nai-Fu Chen, Richard Roll and Stephen A. Ross, *Economic Forces and the Stock Market*, Journal of Business, Vol.59 No. 3, 1986

[16] Shiu-Sheng Chen, *A Note on In-Sample and Out-of-Sample Tests for Granger Causality*, Journal of Forecasting, No. 24, 453-464, (2005)

[17] John Cochrane, *Discount Rates*, Presidential Address, American Finance Association, 2011

[18] John Cochrane and Monika Piazzesi, *Bond Risk Premia*, 2005, American Economic Review, v 95 n 1

[19] John Cochrane, *The Dog That Did Not Bark: A Defense of Return Predictability*, 2008, The Review of Financial Studies, v 21 n 4

[20] Persi Diaconis, *The Markov Chain Monte-Carlo Revolution*, Bulletin of the American Mathematical Society, Vol. 46 No. 2 April 2009, pp. 179-205

[21] Persi Diaconis and Laurent Saloff-Coste, *Comparison Theorems for Reversible Markov Chains* The Annals of Applied Probability, Vol.3 No. 8 1993

[22] J.L. Doob, *Discrete Potential Theory and Boundaries*, Journal of Mathematics and Mechanics, Vol.8 No.3, 1959, pp. 433-458

[23] M.D. Donsker and S.R.S. Varadhan, *Asymptotic Evaluation of Certain Markov Process Expectations in Large Time, I*,II*,III**, Communications in Pure and Applied Mathematics, 1975 - 1976, Vol. 28 - 32

[24] Pierre Collin-Dufresne, Michael Johannes, and Lars A. Lochstoer, *Parameter Learning in General Equilibrium: The Asset Pricing Implications*, American Economic Review, 2016

[25] D.Duffie, J. Geanakoplos, A. Mas-Colell, A. McLennan *Stationary Markov Equilibria*, Econometrica, Vol. 62, No. 4. (Jul.), 1994, pp. 745-781.

[26] Lazlo Erdos, Horng-Tzer Yau, *Dynamical approach to random matrix theory*, Institute of Science and Technology, Austria. Manuscript, 2017

[27] Lazlo Erdos, *Notes on the Matrix Dyson Equation*, Lecture Notes, Institute of Advanced Studies, Princeton University, 2016

[28] Eugene Fama and Ken French, *Common Risk Factors in Stocks and Bonds*, Journal of Financial Economics, 33 (1993) 3-56

[29] Eugene F. Fama and Kenneth R. French, *A Five-Factor Asset Pricing Model* September 2014, Working Paper

[30] Gibbons, Ross and Shanken, *A Test of the Efficiency of a Given Portfolio*, Econometrica, Vol. 57, No. 5 (September), 1989, 1121 to 1152

[31] Joseph Gerakos Juhani T. Linnainmaa, *Decomposing Value*, Chicago Working Paper, September 19, 2012

[32] Valentin Haddad, Serhiy Kozak, Shri Santosh, *Predicting Relative Returns*, Working Paper, 2017

[33] Lars Hansen, *Dynamic Valuation Decomposition with Stochastic Economies* NBER and Chicago Working Paper, 2011; Koopmnan and Fisher-Schulz Lecture series, 2008

[34] Lars Peter Hansen and Thomas J. Sargent, *Fragile beliefs and the price of uncertainty*, Journal of Quantitative Economics 1, 2010

[35] Lars Peter Hansen and Jose A. Scheinkman, *Long Term Risk: An Operator Approach*, Econometrica, Vol. 77, No. 1 (January, 2009), 177-234

[36] J. Michael Harrison and David M. Kreps, *Martingales and Arbitrage in Multiperiod Securities Markets*, Journal of Economic Theory, 20, 381-408, 1979

[37] Campbell R. Harvey, Yan Liu, Heqing Zhu, *. . . and the Cross-Section of Expected Returns*, Review of Financial Studies, October 2015

[38] Simon Huang, *The Momentum Gap and Return Predictability*, Working paper, 2016

[39] Ravi Jagannathan; Zhenyu Wang, *The Conditional CAPM and the Cross-Section of Expected Returns* The Journal of Finance, Vol. 51, No. 1. (Mar.), 1996, pp. 3-53.

[40] Brian Kelly, Seth Pruitt, Yinan Su, *Instrumented Principal Components Analysis*, 2017, Working Paper

[41] Brian Kelly and Seth Pruitt, *Market Expectations and the Cross-section of Present Values*, 2013, Journal of Finance, Vol. 68, No. 5

[42] Bryan Kelly, Seth Pruitt and Yinan Su, *Some Characteristics Are Risk Exposures and the Rest Are Irrelevant*, Working Paper, 2017

[43] Peter Lax, *Functional Analysis*, Pure and Applied Mathematics, 2002

[44] Lettau, M., Ludvigson, S.C., *Consumption, aggregate wealth and expected stock returns*, Journal of Finance 56 (3), 2001, 815-849

[45] Thomas M. Liggett, *Continuous Time Markov Processes*, American Mathematical Society, Graduate Texts in Mathematics, Vol. 133, 2010

[46] Andrew W. Lo, *The Statistics of Sharpe Ratios*, Financial Analysts Journal, July/August 2002

[47] David G. Luenberger, *Optimization by Vector Space Methods*, Wiley Professional Paperback Series, 1969

[48] Paulo Maio and Pedro Santa-Clara, *Dividend Yields, Dividend Growth, and Return Predictability in the Cross Section of Stocks*, Journal of Financial and Quantitative Analysis, Vol. 50, Nos. 1/2, Feb./Apr. 2015

[49] Markowitz, Harry, *Portfolio Selection*, Journal of Finance 7, 1952, 77 to 91.

[50] Chistopher J. Malloy, Tobias J. Moskowitz, and Annette Vissing-Jorgensen, *Long-Run Stockholder Consumption Risk and Asset Returns*, Journal of Finance, December 2009

[51] Merton, Robert C., *Optimal Consumption and Portfolio Rules in a Continuous Time Model*, Journal of Economic Theory 3, 1971, 373 to 413.

[52] Merton, Robert C., *An Intertemporal Capital Asset Pricing Model*, Econometrica 41, 1973, 867 to 887.

[53] Michaud, Robert O., *The Markowitz optimization Enigma: Is Optimized Optimal?*, Financial Analysts Journal, Jan/Feb 1989

[54] Alan Moreira and Tyler Muir, *Volatility-Managed Portfolios*, Journal of Finance, May 2017

[55] Tobias J. Moskowitz and Mark Grinblatt *Do Industries Explain Momentum?*, The Journal of Finance, Vol. 54, No. 4, (Aug.), 1999

[56] John Nash, *Continuity of Solutions of Parabolic and Elliptic Differential Equations*, American Journal of Mathematics, Vol.80, No.4, 1958, pp. 931-954

[57] Newey, W.K., West, K.D., 1987. *A simple, positive semidefinite, heteroskedasticity and autocorrelation consistent covariance matrix*, Econometrica 55, 703-708.

[58] Fulvio Ortu, Andrea Tamoni and Claudio Tebaldi, *Long-Run Risk and the Persistence of Consumption Shocks*, The Review of Financial Studies, Vol. 26, No. 11 (November), 2013, pp. 2876-2915

[59] Jonathan A. Parker and Christine Julliard, *Consumption Risk and the Cross Section of Expected Returns*, Journal of Political Economy, 2005, vol. 113, no. 1

[60] Ross, Stephen A. *The Arbitrage Theory of Capital Asset Pricing*, Journal of Economic Theory 13, 1976, 341 to 360.

[61] Steve Ross, *The Recovery Theorem*, Journal of Finance, Vol. LXX, No. 2, 2015

[62] Walter Rudin, *Functional Analysis*, Second Edition, McGraw-Hill Inc, 1991

[63] Federico Severino, *Isometric Operators on Hilbert spaces and the Classical Wold Decomposition for Stationary Time Series*, Boccini University Working Paper, September, 2014

[64] Daniel M. Stroock, *Introduction to Markov Processes*, Springer Graduate Texts in Mathematics, Vol. 230, Second Edition, 2014

[65] Terrance Tao and Van Vu, *Random Matrices: Universality of Local Eigenvalue Statistics up the the Edge*, UCLA Mathematics Working Paper, 2009

[66] S. R. S. Varadhan, *Large Deviations*, The Annals of Probability, Vol. 36, No. 2, 397-419, 2008

[67] Ivo Welch and Amit Goyal, *A Comprehensive Look at The Empirical Performance of Equity Premium Prediction*, Review of Financial Studies, v.21 n.4 2008

[68] Ivo Welch and Amit Goyal, *Predicting the Equity Premium with Dividend Ratios*, Management Science, Vol. 49, No. 5, 2003, pp. 639-654

Technical Appendices

7 Appendix: Markov Asset Pricing Details

7.1 Technical Environment

Invariant Measure on X The *kernel* of an homogeneous Markov chain is a time-invariant function $m(x^{[i]}, x^{[j]}) : X \times 2^X \longrightarrow [0, 1]$ containing the transition probabilities for X_t, $\Pr(X_t = x^{[j]} | X_{t-1} = x^{[i]}) =: m(x^{[i]}, x^{[j]})$ for every ordered pair of states. Let $S = \#X$; then $\sum_{k=1}^{S} m(x^{[i]}, x^{[k]}) = 1$ for each i. The transition matrix induced by the kernel has entries $(\mathcal{M})_{i,j} = m(x^{[i]}, x^{[j]})$.

We assume the chain is irreducible, i.e., $m(x^{[k]}, x^{[j]}) > 0$ for every $0 \le k, j \le S$. We also assume $\mathcal{M}$ is aperiodic: $\gcd\{n : m(x^{[j]}, x^{[j]})^n > 0\} = 1$ for each $x^{[j]} \in X$ ("gcd" finds the greatest common divisor). In finite states irreducibility and aperiodicity reduce to irreducibility for every $n \in \mathbb{N}$, $m(x^{[k]}, x^{[j]})^n > 0$, where $j = k$ captures the aperiodicity condition (cf., Hairer, 2006, E3.9 p.14).

Define the space of probability measures over X

$$F_\mu := \left\{ q = \begin{bmatrix} q_1 \\ q_2 \\ \vdots \\ q_S \end{bmatrix} : \sum_{i=1}^{S} q_i = 1,\ q_i \in [0,1] \right\} \subset \mathbb{R}^S$$

Let $\{e_i\,,\ 1 \le i \le S\}$ be standard basis vectors for $\mathbb{R}^S$.

Proposition 7.1 *There is a unique invariant measure* $\mu_0 \in F_\mu$ *satisfying* $\mu_0' = \mu_0' \mathcal{M}^N$ *for every* $N \ge 1$ *and* $e_i \cdot \mu_0 > 0$ *for every* $0 \le i \le S$.

Proof Levin, Peres and Wilmer (2008), p. 12 Proposition 1.14

□

Consider a sequence of elements $h_n \in F_\mu$ generated by $\mathcal{M}'$ and denote the k'th outcome h_k. It is clear that when $h_0 = \mu_0$, $(\mathcal{M}')^K h_0 =: h_K = \mu_0$ for every $K \in \mathbb{N}$.

The Path Space Fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Define the product space $Z := \{X\}^{\mathbb{N}} = X \times X \times ... = \prod_{n \in \mathbb{N}} X_n$. Z contains all infinite sequences of elements of X. A path $z_s\ s \in \mathbb{N}$ is an arithmetic function $z : \mathbb{N} \to \mathbb{R}$. A *sample path* is a finite sequence of elements in X, $z_{s \le N} \in \{X\}^N$ on the truncated domain $\{0, 1, 2, ..., N\}$. Events $\omega \in \Omega$ realize as paths $\omega \mapsto Z$, which we occasionally emphasize by writing $\{z\}(\omega) : \Omega \longrightarrow Z \subset \mathbb{R}^{S\mathbb{N}}$.

We use the discrete σ- field for X, written 2^X. A sequence of refinements to $(\Omega, \oslash) =: \mathcal{F}_0$, $\mathcal{F}_N \subset \mathcal{F}_{N+1}$, is *generated* by the sample-path events $\{\omega : \{z\}_N(\omega) \in \sigma(\{X\}^N)\} \in \mathcal{F}_N$ for every N. Each $\mathcal{F}_0 \subset \mathcal{F}_n \subset \mathcal{F}$ for $n < \infty$. We write $\sigma(\{X\}^N) = (2^X)^N$ for finite N.

Lemma 7.2 $\sigma(\prod_{n \in \mathbb{N}} X_n) := \sigma(X^{\mathbb{N}}) = \sigma(X)^{\mathbb{N}}$

Proof The countable product of finite sets, 2^X, in this case given by $(2^X)^{\mathbb{N}} = \sigma(X)^{\mathbb{N}}$, is countable. Hence, using Theorem 4.44, Aliprantis and Border (2006), p.149, for countable σ- fields, $\sigma(\prod_{n \in \mathbb{N}} X_n) = \prod_{n \in \mathbb{N}} \sigma_n(X)$.

□

In words, the σ -field generated by the countable product space is equivalent to the countable product of the σ -fields generated by the state space X (which is not true in general). The implication is that any event in the sequence space can be written as the countable product of elements in 2^X, or equivalently, some countable product of elements of X.

Denote the finite-dimensional distributions of the Markov chain $\mathbb{P}_{\mu,N}(b)$, $b \in (2^X)^N$ for each initial distribution over X, $\mu(x)$. The Dirac mass δ_x on points in X is a particular initial distribution, in which case we write $\mathbb{P}_{x,\cdot}$. The probability of a particular sample path $(2^X)^N \ni z_N : x_0 \rightsquigarrow x_N$ is written

$$h_0(\{z_N\}) = \mu(x_0)m(x_0, x_1)m(x_1, x_2)...m(x_{N-1}, x_N)$$

for initial distribution μ.

The probability of any event $b_N \in (2^X)^N$ can be written as the sum of probabilities of paths $\{z_N\}$ where it is true, $\mathbb{P}_{\mu,N}(b_N) = \sum_{\{z_N\}\in b_N} h_0(\{z_N\})$, which themselves can be enumerated. To track the paths where b_N is true, each ordered configuration j in the set of configurations $J(b_N)$ is written as a path index, $x_{1(j)}, x_{2(j)}, ..., x_{N(j)}$, where the notation is shorthand for $x_{t(j)} = x_t^{k(j)}$. The set $J(b_N)$ is finite for $N < \infty$, and at most countable for $N = \infty$. The probability of any sample event is a sum of finite products of the kernel

$$\mathbb{P}_{\mu,N}(b_N) = \mu(x) \sum_{j\in J(b_N)} \prod_{n(j)=1}^{N} m(x_{n(j)-1}, x_{n(j)})$$

The preceding argument ensures the sample distributions are *consistent*,

$$\begin{aligned}
\mathbb{P}_{\mu,N+1}(b_1, b_2, ..., b_N, X) &= \mu(x) \left[\sum_{j\in J(\boldsymbol{b})} \prod_{n(j)=1}^{N} m(x_{n(j)-1}, x_{n(j)}) \right] \sum_{i=1}^{S} m(x_N, x_{N+1}^{[i]}) \\
&= \mathbb{P}_{\mu,N}(b_1, b_2, ..., b_N) \qquad (7.1a)
\end{aligned}$$

for $b_i \in \sigma(\{X\}^i)$, every finite N and for $\boldsymbol{b} := (b_1, b_2, ..., b_N) \in (2^X) \times (2^X)^2 \times ... \times (2^X)^N$. This definition of consistency is standard (cf. Durrett (2010, p. 366)). The product measure of the chain is given by the pushforward,

$$\mathbb{P}_\mu(b) := \mathbb{P} \circ \{z\}^{-1}(b) = \mathbb{P}(\omega : \{z\}(\omega) \in b)$$

for every initial distribution μ on X.

Proposition 7.3 *The product measure $\mathbb{P}_\mu$ exists and is uniquely characterized by the finite-dimensional distributions.*

Proof By application of the Kolmogorov extension theorem, letting $N \to \infty$ in 7.1a (Durrett (2010), p. 366, Theorem A.3.1)

□

Definition: Define the *tail* σ -field $\mathscr{T} := \bigcap_{n\geq 1}\{\sigma(\bigcup_{m\geq n}\mathcal{F}_m)\}$. An ergodic set

$$\left\{B \subset (2^X)^{\mathbb{N}} : \{\omega : \{z\}(\omega) \in B\} \subset \mathscr{T}\right\}$$

is a subset of a state space that has $\mathbb{P}_x(b) := \mathbb{P} \circ \{z\}^{-1}(b) \in \{0,1\}$ for every $b \in B$ and each initial $x \in X$. A Markov process is ergodic if its state space is an ergodic set.

Corollary 7.4 *An irreducible aperiodioc finite-state Markov chain is ergodic*

Proof Tuominen and Tweedie (1994), pp. 779-780, Theorem 2.3.

□

7.2 Decomposition and Wold Representation

Remark: A *reversible* Markov operator admits a positive real point spectral decomposition. Important departures from reversibility are of economic interest, so we avoid the assumption of reversibility. See Hansen and Scheinkman (1995) for an early discussion of irreversible Markov models in asset pricing.

Definition: The *dual* of $F(X)$ written $F(X^*)$ is the space of linear functionals η_f on $F(X)$, $\eta_f : F(X) \longrightarrow \mathbb{R}$.

Proposition 7.5 (Decomposition)
I.
The image of X^* under $\mathcal{M}'$ is the direct sum of an invariant subspace $\mathscr{I}_0 = \mathscr{I}(\mu_0)$ and its orthogonal complement $\mathscr{I}_0^{\perp}$,

$$\mathscr{R}(\mathcal{M}') = \mathscr{I}_0 \bigoplus \mathscr{I}_0^{\perp}$$

Moreover, for any integer $k \geq 0$, the image space $\mathscr{R}([\mathcal{M}']^k)$ admits a decomposition into $\mathscr{I}_0$ and a space that depends on k

$$\mathscr{R}([\mathcal{M}']^k) = \mathscr{I}_0 \bigoplus (\mathscr{I}_0^{\perp})_k$$

In particular, for $\mathcal{M}'_\gamma :$ Span$(\mathcal{M}'_\gamma) = \mathscr{I}_0^{\perp} \subset \mathscr{R}(\mathcal{M}')$ and $1 \leq k \in \mathbb{N}$,

$$(\mathscr{I}_0^{\perp})_k = \mathit{Span}\left([\mathcal{M}'_\gamma]^k\right)$$

II.
The columns of $(\mathcal{M}'_\gamma)^k, k \in \mathbb{N}$ converge to the origin in the operator norm topology

$$((\mathcal{M}'_\gamma)^k)_{,j} \longrightarrow_{k\longrightarrow\infty} \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$

Each column $(\mathcal{M}')_{,j}$ *of* $\mathcal{M}'$ *converges*

$$(\mathcal{M}')^k_{,j} \longrightarrow \mu_0$$

in total variation to a distribution μ_0 *that is independent of* j.

We first state a lemma we will need. A general proof is provided in section 7.3.2.

Lemma 7.6 *Recall the dual space* X^* *contains the linear functionals* η *on the range of* X *under* $\mathcal{M}$. *Then,*

- *We can identify* X^* *with the space of probability measures* $F_\mu(X)$ *over* X, $F_\mu(X) \equiv X^*$
- *The adjoint* $\mathcal{M}'$ *is a continuous automorphism on* $F_\nu(X \times X)$. *In particular, each row of* $\mathcal{M}'$ *takes* $F_\mu(X) \longrightarrow F_\mu(X)$

Remark: In the matrix case, this point is easy to illustrate. Consider a stochastic matrix $\mathcal{S}$ and a probability distribution ν on $\mathbb{R}^N$, with N equal to the column count of $\mathcal{S}$. Then for $1_{N\times 1} =: \mathbf{1}$, $\mathcal{S}\mathbf{1} = \mathbf{1}$ and $\nu'\mathbf{1} = 1$. Consider $\hat{\nu} = \mathcal{S}'\nu$. Then

$$\hat{\nu}'\mathbf{1} = (\mathcal{S}'\nu)'\mathbf{1} = \nu'\mathcal{S}\mathbf{1} = \nu'\mathbf{1} = 1$$

so the rows of $\mathcal{S}'$ take probability measures to probability measures.

Proof of part I. By the Perron-Frobenius theorem, $\mathcal{M}$ has largest eigenvalue $\lambda_0 = 1$ with corresponding left and right eigenvectors μ_0' and $\iota = 1_{S\times 1}$, respectively.

Eigenvectors exhibit a scale symmetry $\mathcal{M}\iota = \iota \Leftrightarrow c\mathcal{M}\iota = c\iota$, $0 \neq c \in \mathbb{R}$, and equivalently $c^*\mu_0' = c^*\mu_0'\mathcal{M}$ for any $0 \neq c^* \in \mathbb{R}$. Without loss of generality, we pick the the unitary scale normalization $c^* = c^{-1}$ so that $\iota'\mu_0 = 1$ so μ_0 is a probability distribution. Because the stochastic matrix restricts $c \equiv 1$, this choice of scale pins down the *absolute* scale $c^* = 1$.

Write the algebraic multiplicity of eigenvalue j as $\chi(\{j\})$, and the geometric multiplicity $g(\{j\})$. Denoting the largest eigenvalue $j = 0$, another appeal to the Perron-Frobenius theorem gives $\chi(\{0\}) = g(\{0\}) = 1$. We can now claim the following.

Lemma 7.7 $\mu_0\iota'$ *is a rank-one projection.*

Proof of lemma 7.7 The linear operation $\mu_0\iota'$ is idempotent

$$\mu_0\iota'(\mu_0\iota') = \mu_0(\iota'\mu_0)\iota' = \mu_0\iota'$$

which immediately implies that $I - \mu_0\iota'$ is orthogonal to $\mu_0\iota'$,

$$(I - \mu_0\iota')\mu_0\iota' = \mu_0\iota' - \mu_0\iota' = \mathbf{0}$$

so $\mu_0\iota'$ and $I - \mu_0\iota'$ are projections. That $g(\{0\}) = 1$ implies that the subspace

$$E_0 := \{\hat{\mu} : (\mathcal{M}' - \lambda_0 I)\hat{\mu} = \mathbf{0}\}$$

has dimension one, hence $\text{rank}(\mu_0\iota') = 1$ and $\mu_0\iota'$ is a rank-one projection.

□

To prove part **I.** of proposition (7.1), note that if $\text{rank}(\mathcal{M}') = 1$, the result is trivial with $\mathscr{I}_0^{\perp}$ empty. Suppose $\text{rank}(\mathcal{M}') > 1$. Put

$$\mathcal{M}'_\gamma = \mathcal{M}' - \mu_0\iota'$$

Denote $\text{Span}(\mu_0\iota') =: \mathscr{I}_0$. We intend to characterize $\mathscr{I}_0^{\perp} \subset \mathscr{R}(\mathcal{M}')$ in terms of $\mathcal{M}'_\gamma$.

We can see that

$$\mathcal{M}'_\gamma\mu_0\iota' = (\mathcal{M}' - \mu_0\iota')\mu_0\iota' = \mathcal{M}'(I_S - \mu_0\iota')\mu_0\iota' = \mathbf{0} \tag{6.a}$$

The second equality has used the right-eigenvector of the adjoint $\mathcal{M}'\mu_0 = \mu_0$, and the final equality follows given $\mu_0\iota'$ is a rank-one projection.

We verify that, in addition,

$$\mu_0\iota'\mathcal{M}'_\gamma = (\mathcal{M}_\gamma\iota\mu_0')' = ((\mathcal{M} - \iota\mu_0')\iota\mu_0')' = (\mathcal{M}(I_S - \iota\mu_0')\iota\mu_0')' = \mathbf{0} \tag{6.b}$$

using the right-eigenvector $\iota = \mathcal{M}\iota$, and the facts that $\text{rank}(\mu_0\iota') = \text{rank}(\iota\mu_0')$ and $\mu_0\iota'$ is a rank-one projection.

The image of $\mathcal{M}'$ takes the form $\mathscr{R}(\mathcal{M}') = \mathscr{I}_0 \bigoplus \text{Span}(\mathcal{M}'_\gamma)$, because in (6.a)-(6.b), we have shown $\mathscr{I}_0^{\perp} = \text{Span}(\mathcal{M}'_\gamma)$.

Now define $\mathcal{M}'_{\gamma,k} = (\mathcal{M}')^k - \mu_0\iota'$. By definition of the invariant measure μ_0, we have the right-eigenvector arguments for integer powers k,

$$\begin{aligned}(\mathcal{M}')^k\mu_0 &= \mu_0 \\ (\mathcal{M})^k\iota &= \iota\end{aligned} \tag{6.c}$$

Combining the first identity in (6.c) with the arguments (6.a), we obtain

$$\mathcal{M}'_{\gamma,k}\mu_0\iota' = [(\mathcal{M}')^k - \mu_0\iota']\mu_0\iota' = (\mathcal{M}')^k(I_S - \mu_0\iota')\mu_0\iota' = \mathbf{0}$$

Repeating the arguments in (6.b) using the second identity in (6.c), and the and the elementary fact $((\mathcal{M})^k)' = (\mathcal{M}')^k$,

$$\mu_0\iota'\mathcal{M}'_{\gamma,k} = ([(\mathcal{M}')^k - \mu_0\iota']'\iota\mu_0')' = (\mathcal{M}^k(I_S - \iota\mu_0')\iota\mu_0')' = \mathbf{0}$$

Finally, for $0 < k \in \mathbb{N}$, it is clear that

$$\begin{aligned}(\mathcal{M}'_\gamma)^k &= (\mathcal{M}')^k(I_S - \mu_0\iota')^k \\ &= (\mathcal{M}')^k(I_S - \mu_0\iota') \\ &= (\mathcal{M}')^k - \mu_0\iota' \\ &= \mathcal{M}'_{\gamma,k}\end{aligned}$$

We conclude that for any integer $k \geq 1$, $(\mathscr{I}_0^\perp)_k = \text{Span}(\mathcal{M}'_{\gamma,k})$ with $\mathcal{M}'_{\gamma,k} = (\mathcal{M}'_\gamma)^k$ and $\mathcal{M}'_\gamma = \mathcal{M}' - \mu_0\iota'$. In particular, $\mathscr{R}((\mathcal{M}')^k) = \mathscr{I}_0 \bigoplus \text{Span}((\mathcal{M}'_\gamma)^k)$.

□

Before proving part II., we state some useful corollaries.

Corollary 7.8 $\mu_0\iota'$ *and* $\mathcal{M}'_\gamma$ *commute.*

Corollary 7.9 (Chapman-Kolmogorov) *The operators* $\{(\mathcal{M}')^n, n \in \mathbb{Z}^+\}$ *form an abelian semigroup under (matrix) multiplication. Every k- decomposition is contained in the semigroup. An identical statement is true for the operators* $\{(\mathcal{M})^n, n \in \mathbb{Z}^+\}$*.*

Proof Pick finite positive integers n_1, n_2 and denote $N = n_1 + n_2$. Then,

$$\begin{aligned}(\mathcal{M}')^{n_1}(\mathcal{M}')^{n_2} &= (\mu_0\iota' + \mathcal{M}'_\gamma)^N \\ &= \sum_{n=0}^{N} \binom{N}{n} (\mu_0\iota')^{N-n}(\mathcal{M}'_\gamma)^n = \mu_0\iota' + (\mathcal{M}'_\gamma)^N\end{aligned}$$

The first line is trivial, and shows the semigroup property $(\mathcal{M}')^{n_1}(\mathcal{M}')^{n_2} = (\mathcal{M}')^N$ and hence commutativity $(\mathcal{M}')^N = (\mathcal{M}')^{n_2}(\mathcal{M}')^{n_1}$. The second line applies the results from proposition (7.1) finitely many times to extend the semigroup property to the decomposition.

□

Corollary 7.10 (γ- Semigroup) *The operators* $\{(\mathcal{M}'_\gamma)^n, n \in \mathbb{Z}^+\}$ *form an abelian semigroup under matrix multiplication.*

Corollary 7.11 (Wold Time Series Representation) *Under the assumptions needed to decompose* $(\mathcal{M}')^k = \mu_0\iota' + (\mathcal{M}'_\gamma)^k$*, the classic Wold representation is justified.*

Proof First, recall each $\mathbb{R}^S \ni 1(x_{t,\cdot})$ is a degenerate probability distribution with mass on the coordinate of the realized element $x_{t,\cdot} \in X$. Recall the definitions $R_{n,t+1} = r_n \cdot X_{t+1}$, and hence $x_{t+1} = \iota'\mathcal{M}'1(x_{t,\cdot}) + u_{t+1}$. Equivalently,

$$1(x_t) = \mathcal{M}'1(x_{t-1}) + 1(u_t)$$

If $\mathcal{M}' \equiv \mu_0\iota'$, the decomposition is $r_n \cdot X_{t+1} = r_n \cdot \mu_0 + u_{t+1}$ for every t. Consider Rank$(\mathcal{M}') > 1$. Then,

$$\begin{aligned}
r_n \cdot X_{t+1} &= r_n \cdot (\mathcal{M}' 1(x_t) + 1(u_{t+1})) \\
&= r_n \cdot \big([\mu_0\iota' + \mathcal{M}'_\gamma] 1(x_t) + 1(u_{t+1})\big) \\
&= r_n \cdot \mu_0 + r_n \cdot \big(\mathcal{M}'_\gamma 1(x_t) + 1(u_{t+1})\big) \\
&= r_n \cdot \mu_0 + r_n \cdot \big(\mathcal{M}'_\gamma\,(\mathcal{M}' 1(x_{t-1}) + 1(u_t)) + 1(u_{t+1})\big) \\
&= r_n \cdot \mu_0 + r_n \cdot \big(\mathcal{M}'_\gamma([\mu_0\iota' + \mathcal{M}'_\gamma] 1(x_{t-1}) + 1(u_t)) + 1(u_{t+1})\big) \\
&= r_n \cdot \mu_0 + r_n \cdot \big(\mathcal{M}'_\gamma(\mathcal{M}'_\gamma 1(x_{t-1}) + 1(u_t)) + 1(u_{t+1})\big) \\
&= r_n \cdot \mu_0 + r_n \cdot \big(\mathcal{M}'_\gamma(\mathcal{M}'_\gamma\,(\mathcal{M}' 1(x_{t-2}) + 1(u_{t-1})) + 1(u_t)) + 1(u_{t+1})\big) \\
&\vdots \\
&= r_n \cdot \mu_0 + r_n \cdot \sum_{s=0}^{\infty} (\mathcal{M}'_\gamma)^s 1(u_{t+1-s})
\end{aligned}$$

□

Remark: The Wold representation expresses the path $\{z\}(\omega)$ in terms of its martingale-difference (i.e., white-noise) basis. A finite sample version is obtained as a special case. For generalized Wold representations of isometric operators, see Severino (2014).

Remark: The $\mu_0\iota'$ shocks are *permanent* in the sense that $\mu_0\iota' 1(u_t) = (\mathcal{M}')^N \mu_0\iota' 1(u_t)$ for every horizon N.

Corollary 7.12 (Preservation of Probabilities) *The rows of $\mathcal{M}_\gamma$ sum to zero. When Rank$(\mathcal{M}) > 1$, there are nontrivial payoff vectors $d_n \in F(K)$ such that*

$$|e_i \cdot \mathcal{M}'_\gamma d_n| > 0$$

for some $i \in \{1, ..., S\}$

Proof Put $1 = 1_{S\times 1}\,(= \iota)$. Recall the right eigenvector of $\mathcal{M}\iota = \iota$, so that

$$1 = \mathcal{M}\iota = (\iota\mu_0' + \mathcal{M}_\gamma)\iota = \iota\mu_0'\iota + \mathcal{M}_\gamma\iota = 1 + \mathcal{M}_\gamma\iota$$

Clearly, $\mathcal{M}_\gamma\iota = \mathbf{0}$.

Given the invariant $\mu_0' = \mu_0'\mathcal{M}$, the matrix $\mathcal{M}_\gamma$ is identically zero when $\mathcal{M}$ is identically $(\mu_0\iota')'$. Given μ_0, a necessary and sufficient condition for Rank$(\mathcal{M}) = 1$ is that the rows of $\mathcal{M}$ are constant multiples of μ_0'.

Take $\mathcal{M}'$ such that Rank$(\mathcal{M}') > 1$, and consider d_n s.t. $|d_n \cdot e_i| > 0$ for some $i \in \{1, ..., S\}$. Then $\mathcal{M}'d_n = \mu_0\iota' d_n + \mathcal{M}'_\gamma d_n$ and $0 \leq |e_i \cdot (\mathcal{M}' - \mu_0\iota')d_n| = |e_i \cdot \mathcal{M}'_\gamma d_n|$ for every $1 \leq i \leq S$. Suppose for a contradiction that $e_i \cdot (\mathcal{M}' - \mu_0\iota')d_n = 0$ for every i. In other words, $\mathcal{M}'d_n = \mu_0\iota'd_n$ for every nonzero $d_n \in F_d(K)$. Because $\mathcal{M}'\mu_0 = \mu_0$, these together imply

$$\begin{aligned} \mathcal{M}'(\mu_0 - d_n) &= \mu_0(1 - \iota'd_n) \\ &= \mu_0\iota'(\tfrac{1}{S}\iota - d_n) \end{aligned} \tag{1.a}$$

The condition $\text{Rank}(\mathcal{M}') > 1$ ensures $\mathcal{M}'$ is not identically $\mu_0\iota'$. In particular, in $\mathscr{R}(\mathcal{M}') = \mathscr{I}_0 \bigoplus \mathscr{I}_0^{\perp}$, the $\mathscr{I}_0^{\perp}$ is nonempty. So, (1.a) implies $d_n = \mu_0 = \frac{1}{S}\iota$, the uniform distribution over Z_S. If μ_0 is not uniform, we are done. If μ_0 is uniform, then $\mathcal{M}'\iota = \iota$, so $\mathcal{M}'$ is doubly stochastic. As a result, $\mathcal{M}'d_n = \mu_0\iota'd_n = \frac{1}{S}\iota\iota'd_n$, which implies each row of $\mathcal{M}'$ is identically the uniform distribution. Hence, contrary to our assumption, $\text{Rank}(\mathcal{M}') = 1$. Thus, for some $i \in \{1, ..., S\}$, $0 < |e_i \cdot (\mathcal{M}' - \mu_0\iota')d_n| = |e_i \cdot \mathcal{M}'_\gamma d_n|$.

□

Remark: $\mathcal{M}'_\gamma$ redistributes probability within the transition dynamics. Relative to dynamics under $\mu_0\iota'$, $\mathcal{M}'_\gamma$ generates non-negligible distortions to asset payoffs locally in time.

It turns out these distortions are limited to risky assets.

Corollary 7.13 (State Dependent Payoffs) *Every nontrivial constant payoff is in the image space of $\mu_0\iota'$. In particular, the uniform distribution is never the range of $\mathcal{M}'_\gamma$.*

Proof Take $\text{Rank}(\mathcal{M}') > 1$ so $\text{Rank}(\mathcal{M}'_\gamma) \geq 1$. Put

$$Ker(\mathcal{M}_\gamma) := \{v \in C_K(X) : \mathcal{M}_\gamma v = 0_{S\times 1}\}$$

Recall $1_{S\times 1} = \iota$ and $c\mathcal{M}_\gamma\iota = 0$ for any scalar $c \in \mathbb{R}$. It follows that

$$b_0 := \{v \in C_K(X) : v = c\,\iota\,,\ 0 \neq c \in \mathbb{R}\} \subset Ker(\mathcal{M}_\gamma)$$

Now, using[13]

$$\mathscr{R}(\mathcal{M}'_\gamma) = Ker(\mathcal{M}_\gamma)^{\perp}$$

we conclude that $b_0 \not\subseteq \mathscr{R}(\mathcal{M}'_\gamma)$. In particular, $b_0 \subset \mathscr{I}_0$.

□

7.3 Convergence Rates

To study convergence, we review a few definitions and known results. A handful of purely analytical definitions are relegated to section (7.6.1.).

[13] e.g., Luendberger, sec 6.6 pp 155, theorem 1.

Definition: The *total variation* of a signed measure is

$$\| \eta \|_{TV} := \sup_{A \in \mathcal{B}(X)} |\eta(A)|$$

Definition: The total variation *distance* between two probability measures μ, ν is

$$d(\mu,\nu)_{TV} = \| \mu - \nu \|_{TV} = \sup_{b \in (2^X)^{\mathbb{N}}} |\mu(b) - \nu(b)|$$

which itself takes values in $[0,1]$.

Lemma 7.14 *The set of probability measures F_μ is metrized by total variation $d = d_{TV}$. In particular (F_μ, d_{TV}) inherits the topology induced by total variation. Now write $\sigma(F_\mu)$ for the Borel σ -field on F_μ with open sets generated by the total variation distance d_{TV}. Then $(F_\mu, \sigma(F_\mu))$ is a measurable space.*

Proof Den Hollander (2011), p.13 II.1

□

Lemma 7.15 *On countable state space, total variation distance is equivalent to*

$$\sup_{b \in (2^X)^{\mathbb{N}}} |\mu(b) - \nu(b)| = \frac{1}{2} \sum_{a \in X} |\mu(a) - \nu(a)|$$

Proof Levin, Peres and Wilmer, 2008, Proposition 4.2.

□

An immediate corollary of lemma (7.7) is that for finite X, total variation is an ℓ^1 *norm*

$$\begin{aligned} \sum_{a \in 2^X} |h_0(a) - \mu_0(a)| &= \sum_{a \in 2^X} |\mathsf{h}_0(a) - 1| \mu_0(a) \\ &= \| \mathsf{h}_0 - \iota \|_{\ell^1(\mu_0)} \end{aligned}$$

under μ_0, where the elementwise quotient $\mathsf{h}_0 = h_0/\mu_0$ is the likelihood of h_0 with respect to μ_0.

Definition: The operator norm $\| \cdot \|_{Op}$ for operator $\mathbb{T}$ is

$$\| \mathbb{T} \|_{Op} \ := \sup_{\{v \in \mathcal{D}(\mathbb{T}),\, \|v\|_{\ell^1} = 1\}} \| \mathbb{T} v \|_{\ell^1}$$

Lemma 7.16 *For continuous linear operators $\mathbb{T}, \mathbb{S}$*

$$\| \mathbb{T}'\mathbb{T} \|_{Op} = \| \mathbb{T} \|^2_{Op} \qquad \text{(a.)}$$

$$\| \mathbb{S}\mathbb{T} \|_{Op} \leq \| \mathbb{S} \|_{Op} \| \mathbb{T} \|_{Op} \qquad \text{(b.)}$$

Proof
(a.) Lax (2002), Theorem 14., section 19.7, p. 222
(b.) Lax (2002) Theorem 8., section 15.4, p. 168

□

Proof of Prop. 7.1 part II. We need only consider the nontrivial case Rank$(\mathcal{M}') > 1$. Pick an initial probability over X given by $\hat{h}_0$. Then $\hat{h}'_{0,k} = \hat{h}'_0 \mathcal{M}^k$, equivalently $(\mathcal{M}')^k \hat{h}_0 = \hat{h}_{0,k}$. Define the likelihood $\mathsf{h}_0 = \hat{h}/\mu_0$ with the quotient elementwise as before. Consider first the case $\hat{h}_0 \neq \mu_0$. We are interested in $\| (\mathcal{M}')^k \hat{h}_0 - \mu_0 \|_{TV}$. From part I.,

$$\begin{aligned} (\mathcal{M}')^k \hat{h}_0 - \mu_0 &= \mu_0 \iota' \hat{h}_0 + (\mathcal{M}'_\gamma)^k \hat{h}_0 - \mu_0 \\ &= (\mathcal{M}'_\gamma)^k \hat{h}_0 + \mu_0(\iota' \hat{h}_0 - 1) \\ &= (\mathcal{M}'_\gamma)^k \hat{h}_0 \end{aligned}$$

The third equality is simply $\iota'\hat{h}_0 = 1$ because $\hat{h}_0$ is a probability measure. We can now write

$$\begin{aligned} 2 \parallel (\mathcal{M}')^k \hat{h}_0 - \mu_0 \parallel_{TV}^2 &= 2 \parallel (\mathcal{M}'_{\boldsymbol{\gamma}})^k \hat{h}_0 \parallel_{TV}^2 \\ &= \parallel (\mathcal{M}'_{\boldsymbol{\gamma}})^k \mathsf{h}_0 \parallel_{\ell^1(\mu_0)}^2 \\ &\leq \parallel (\mathcal{M}'_{\boldsymbol{\gamma}})^{2k} \mathbf{1} \parallel_{\ell^1(\mu_0)} \parallel \mathsf{h}_0^2 \parallel_{\ell^1(\mu_0)} \end{aligned}$$

using Cauchy-Schwarz.

We showed in corollary 7.13 that the condition $\mathcal{M}_{\boldsymbol{\gamma}}\iota = \mathcal{M}'_{\boldsymbol{\gamma}}\iota$ imposes that $\mathcal{M}$ is doubly stochastic and therefore that μ_0 is uniform $\mu_0 = \frac{1}{S}\iota$ and have ruled this case out. Because $\mathbf{1}(c) := \{v : v = c\mathbf{1}\, 0 \neq c \in \mathbb{R}\} \subset \text{Ker}(\mathcal{M}_{\boldsymbol{\gamma}})$, we conclude $\mathcal{M}'_{\boldsymbol{\gamma}}\mathbf{1}$ is not trivial. Proceeding

$$\begin{aligned} \parallel (\mathcal{M}'_{\boldsymbol{\gamma}})^{2k} \mathbf{1} \parallel_{\ell^1(\mu_0)} \parallel \mathsf{h}_0^2 \parallel_{\ell^1(\mu_0)} &= 2 \parallel (\mathcal{M}'_{\boldsymbol{\gamma}})^{2k} \mathbf{1} \parallel_{\ell^1(\mu_0)} \parallel \hat{h}_0^2 \parallel_{TV} \\ &\leq 2 \parallel (\mathcal{M}'_{\boldsymbol{\gamma}})^{2k} \mathbf{1} \parallel_{\ell^1(\mu_0)} \end{aligned} \tag{6.1.a}$$

because the total variation distance is no greater than one.

The operator $\mathcal{M}'_{\boldsymbol{\gamma}}$ is a contraction for every $k \in \mathbb{N}$ and is therefore uniformly bounded. Because it is also linear, by the open mapping theorem it is continuous in the operator norm topology (e.g., Lax (2002), Theorem 12. p. 170). Now define $\mathbf{M}_\gamma := \mathcal{M}_{\boldsymbol{\gamma}}\mathcal{M}'_{\boldsymbol{\gamma}}$. The operator $\mathbf{M}_\gamma$ is linear, uniformly bounded in k and symmetric. We have

$$\begin{aligned} 2 \parallel (\mathcal{M}'_{\boldsymbol{\gamma}})^{2k} \mathbf{1} \parallel_{\ell^1(\mu_0)} &\leq 2 \max_{||v||=1} \parallel (\mathcal{M}'_{\boldsymbol{\gamma}})^{2k} v \parallel_{\ell^1(\mu_0)} \\ &= 2 \parallel (\mathcal{M}'_{\boldsymbol{\gamma}})^{2k} \parallel_{Op} \\ &\leq 2 \parallel \mathcal{M}'_{\boldsymbol{\gamma}} \parallel_{Op}^{2k} \\ &= 2 \parallel \mathcal{M}_{\boldsymbol{\gamma}}\mathcal{M}'_{\boldsymbol{\gamma}} \parallel_{Op}^{k} \\ &= 2 \parallel \mathbf{M}_\gamma \parallel_{Op}^{k} \\ &\leq 2\rho(\mathbf{M}_\gamma)^k \end{aligned} \tag{6.1.b}$$

where $\rho(\mathbf{M}_\gamma)$ is the spectral radius of $\mathbf{M}_\gamma$, and we have used submultiplicativity and part (a.) in lemma 7.18.

When $\text{rank}(\mathcal{M}') > 1$, the symmetric positive definite operator $\mathcal{M}\mathcal{M}'$ has at least two eigenvalues. In particular, $\rho(\mathbf{M}_\gamma)$ is bounded above by the second largest eigenvalue of $\mathcal{M}\mathcal{M}'$. Because $\mathcal{M}\iota = \iota$, the largest eigenvalue of $\mathcal{M}\mathcal{M}'$ is one. The second largest eigenvalue ζ of $\mathbf{M}$ is strictly less than one. Moreover, if $\mathcal{M}\psi = \zeta^{1/2}\psi$ for $\mathfrak{Re}(\zeta^{1/2}) < 1$, then

$$\mathcal{M}\psi = \iota\mu_0'\psi + \mathcal{M}_{\boldsymbol{\gamma}}\psi = \mathcal{M}_{\boldsymbol{\gamma}}\psi = \zeta^{1/2}\psi$$

Consider $\varphi'\mathcal{M} = \varphi'\iota\mu_0' + \varphi'\mathcal{M}_{\boldsymbol{\gamma}}$ and $\varphi'\psi = \zeta^{1/2}$, so that

$$\parallel \mathcal{M}'\varphi \parallel = \parallel \mu_0\iota'\varphi + \mathcal{M}'_{\boldsymbol{\gamma}}\varphi \parallel = \parallel \mathcal{M}'_{\boldsymbol{\gamma}}\varphi \parallel \leq |\zeta^{1/2}||\varphi'\psi| = \zeta$$

We conclude

$$\rho(\mathbf{M}_\gamma) \leq \zeta < 1 \tag{6.1.c}$$

Combining (6.1.a) - (6.1.c), we have

$$\| (\mathcal{M}')^k \hat{h}_0 - \mu_0 \|_{TV} \leq \zeta^{k/2} \longrightarrow 0 \quad k \to \infty$$

Now, the case not yet considered is when $h_0 = \mu_0$ but $\mathcal{M}$ is not identically $\iota\mu_0'$. We are interested in the rate at which

$$\| \iota' \mathcal{M}^k f - \mu_0' f \|_{TV} \longrightarrow 0 \qquad k \to \infty$$

However, it is straightforward to bound this deviation using the same radius $\rho(\mathbf{M}_\gamma)$. Expediting the argument using details in (6.1.a)-(6.1.c),

$$\begin{aligned} 2 \| \iota' \mathcal{M}^k f - \mu_0' f \|_{TV}^2 &= 2 \| (\iota' \mathcal{M}^k - \mu_0') f \|_{TV}^2 \\ &= \| (\iota' \mathcal{M}^k - \mu_0') f \|_{\ell^1}^2 \\ &\leq \| \mathcal{M}_{\boldsymbol{\gamma}}^{2k} \|_{\ell^1} \| f^2 \|_{\ell^1} \\ &\leq 2 \| \mathcal{M}_{\boldsymbol{\gamma}} \|_{Op}^{2k} \\ &= 2 \| \mathbf{M}_\gamma \|_{Op}^{k} \\ &\leq 2\rho(\mathbf{M}_\gamma)^k \end{aligned}$$

Hence,

$$\| \iota' \mathcal{M}^k f - \mu_0' f \|_{TV} \leq \zeta^{k/2} \longrightarrow 0 \quad k \to \infty$$

□

Remark: A square root is natural when ζ is viewed as a singular value of $\mathcal{M}'$, in which case the corresponding eigenvalue of $\mathcal{M}'$ is $\zeta^{1/2}$.

Remark: The generator $\mathcal{L}$ inherits a decomposition,

$$(\mathcal{L}'h)(x) = (\mathcal{M}_{\boldsymbol{\gamma}}' h)(x) + ((\mu_0 \iota' - I)h)(x)$$

from proposition 7.1. A function g is γ - *harmonic* when

$$\begin{aligned} 0 &= ((\mathcal{M}_{\boldsymbol{\gamma}}' - I)g)(x) \\ &= (\mathcal{L}'g)(x) - (\mu_0 \iota' g)(x) \end{aligned}$$

emphasizing that such a g has conditionally mean-zero transitory contributions.

Definition: ν - harmonics A function h such that $(\mathcal{L}h)(x) = 0$ for every x is *harmonic* with respect the measure $\nu_{\mathcal{M}}$ on X' when in addition, $\langle \mathcal{L}h, 1\rangle_{\nu(\mathcal{M})} = \langle h, \mathcal{L}^* \rangle_{\nu(\mathcal{M})}$.

Proposition 7.17 *Then,*

- *Every harmonic function h with respect to $\nu_{\mathcal{M}}$ is a martingale under $\mathbb{P}_{\nu_{\mathcal{M}}}$ as a function of the initial conditions*
- *Every martingale is contained in the kernel of $\mathcal{M}'_\gamma$ - i.e., every $\nu_{\mathcal{M}}$ - harmonic function is γ -harmonic when $\mathcal{M}_\gamma$ decomposes $\mathcal{M}$ and $\nu'_{\mathcal{M}} = \nu'_{\mathcal{M}}\mathcal{M}$*

Proof Part i) is given by Doob (1959). Part ii) follows from lemma (7.8) and $\text{Ker}(\mathcal{A}) = \mathscr{R}(\mathcal{A}')^{\perp}$.

□

7.4 Covariance Matrix of Returns

Define

$$\mathscr{E}(\phi) = \parallel \phi \parallel^2_{\ell^2(\mu_0)} - \parallel \mathcal{M}'\phi \parallel^2_{\ell^2(\mu_0)}$$

We will make use of a variational representation of the spectral gap given by Diaconis and Strook 1991,

$$\zeta := 1 - \lambda_1 = \inf_{\phi} \left\{ \frac{\mathscr{E}(\phi)}{\parallel \phi \parallel^2_{\ell^2(\mu_0)}} \text{ s.t. } \parallel \phi \parallel^2_{\ell^2(\mu_0)} > 0 \right\}$$

We now unpack the contents of $\mathscr{E}(\phi)$ in terms of realized returns data.

Inner Product To compare any two sequences a and b in Z, we consider the natural inner product $\langle \{z_a\}, \{z_b\} \rangle$ on ℓ^2 and the normalized inner product $\langle \{z_a\}, \{z_b\} \rangle_{\mu_0}$.

Lemma 7.18 *The path space appended with either of these inner products is a Hilbert space.*

Proof Strook, 2014, p. 139

□

We occasionally write ϕ for paths of the Markov chain. Recall $\mathscr{E}_M(\phi) = \mathscr{E}_M(\phi, \phi) := \langle \phi, (I - \mathcal{M}\mathcal{M}')\phi \rangle_\mu$, and that for real operators B, the adjoint operator is B', $\langle B\phi, \phi \rangle_\mu = \langle \phi, B'\phi \rangle_\mu$. $\mathscr{E}_M(\phi)$ rearranges,

$$\begin{aligned} \mathscr{E}_M(\phi, \phi) &= \langle \phi, (I - \mathcal{M}\mathcal{M}')\phi \rangle_\mu \\ &= \langle \phi, \phi \rangle_\mu - \langle \phi, \mathcal{M}\mathcal{M}'\phi \rangle_\mu \\ &= \mathbb{E}[\phi\phi'] - \langle \mathcal{M}'\phi, \mathcal{M}'\phi \rangle_\mu \\ &= \mathbb{E}[\phi\phi'] - \mathbb{V}(\mathcal{M}'\phi) \end{aligned}$$

Furthermore, $\mathbb{E}[\phi\phi'] = \langle \phi, \phi \rangle_\mu = \mathbb{V}(\phi) + \mu_0\mu_0'$, where $\mathbb{V}(\phi)$ is the variance of ϕ, and $\mathbb{V}(\mathcal{M}'\phi)$ is the variance of the conditional mean of ϕ.

Define the lag operator $L\phi_{t+1} = \phi_t$. Each path realization is an ℓ^2 sequence of the form $\phi = \{z\}(\omega) = (a_0u_0), (a_1u_1), (a_2u_2), ... =: A_0W_0$. The u_j are serially uncorrelated $\langle u_j, u_k \rangle = 0$ for $j \neq k$. The lag operator L maps $Z \to Z$ via the shift $LA_0W_0 \mapsto A_0W_{-1}$. The adjoint of the lag operator L^* maps $L^*A_0W_0 \mapsto A_1W_0$. Heuristically,

$$\langle LA_tW_t, A_tW_t \rangle = \langle A_tW_{t-1}, A_tW_t \rangle = A'_{t+1}A_t$$
$$\langle A_tW_t, L^*A_tW_t \rangle = \langle A_tW_t, A_{t+1}W_t \rangle = A'_tA_{t+1}$$

as required.

Lemma 7.19 (Lag Operator Isometry) *With this construction, $L : Z \to Z$ is an isometry on $(Z, \mu_0, || \cdot ||_{\ell^2})$.*

Proof We sketch a proof here for intuition. A detailed proof and discussion is given in Severino (2014). An isometric operator leaves norms and inner products unchanged when applied symmetrically. Notice

$$\begin{aligned}
\langle LA_tW_t, LA_tW_t \rangle &= \langle A_tW_{t-1}, A_tW_{t-1} \rangle \\
&= A'_tA_t \\
&= \langle A_tW_t, A_tW_t \rangle
\end{aligned}$$

The same argument shifting coefficients A_t in place of W_t shows the adjoint L^* is also an isometry. □

Proposition 7.20 *The form $\mathscr{E}_M(\phi)$ measures the unconditional variance of the forecast errors generated by $\mathcal{M}'$.*

Proof Recall

$$u_{t+1} = \phi_{t+1} - \mathcal{M}'\phi_t = (I - \mathcal{M}'L)\phi_{t+1}$$

The forecast errors are conditionally mean-zero by construction. The unconditional mean is also zero,

$$\begin{aligned}
\mathbb{E}_0[u_{t+1}] &= \mu_0\iota'\phi_{t+1} - \mu_0\iota'\mathcal{M}'L\phi_{t+1} \\
&= \mu_0\iota'\phi_{t+1} - \mu_0\iota'L\phi_{t+1} - \mu_0\iota'\mathcal{M}'_\gamma L\phi_{t+1} \\
&= \mu_0\iota'(\phi_{t+1} - \phi_t) \\
&= 0 \qquad (7.10.a)
\end{aligned}$$

because the span of $\mathcal{M}'_\gamma$ is orthogonal to $\mathscr{I}_0$ and the lag operator is identity on $\mathscr{I}_0$.

Using (7.10.a), the unconditional variance is of the form

$$\begin{aligned}\mathbb{V}(u) = \langle u, u\rangle_\mu &= \langle (I - \mathcal{M}'L)\phi, (I - \mathcal{M}'L)\phi\rangle_\mu \\ &= \langle \phi, (I - \mathcal{M}'L)'(I - \mathcal{M}'L)\phi\rangle_\mu \\ &= \langle \phi, (I - I\mathcal{M}'L - L'\mathcal{M}I + L'\mathcal{M}\mathcal{M}'L)\phi\rangle_\mu \\ &= \mathbb{E}[\phi\phi'] + \langle \mathcal{M}'L\phi, \mathcal{M}'L\phi\rangle_\mu - \langle \phi, \mathcal{M}'L\phi\rangle_\mu - \langle \phi, L'\mathcal{M}\phi\rangle_\mu \\ &= \mathbb{E}[\phi\phi'] + \mathbb{V}(\mathcal{M}'\phi) - \langle \phi, \mathcal{M}'L\phi\rangle_\mu - \langle \phi, L'\mathcal{M}\phi\rangle_\mu \\ &= \mathbb{E}[\phi\phi'] + \mathbb{V}(\mathcal{M}'\phi) - \langle \phi, \mathcal{M}'L\phi\rangle_\mu - \langle \mathcal{M}'L\phi, \phi\rangle_\mu\end{aligned}$$

where we have used the isometry of the lag operator L to get from the fourth to the fifth line, and the adjoint $(\mathcal{M}'L)' = L'\mathcal{M}$. Now

$$\begin{aligned}\langle \mathcal{M}'L\phi, \phi\rangle_\mu &= \langle \mathcal{M}'\phi_t, \phi_{t+1}\rangle_\mu \\ &= \langle \phi_{t+1} - u_{t+1}, \phi_{t+1}\rangle_\mu \\ &= \langle \phi_{t+1}, \phi_{t+1}\rangle_\mu - \langle u_{t+1}, \phi_{t+1}\rangle_\mu\end{aligned}$$

Using the Wold representation for the second term in the last equality,

$$\begin{aligned}\langle u_{t+1}, \phi_{t+1}\rangle_\mu &= \langle u_{t+1}, \mu_0\iota' + \sum_{s=0}^{\infty} (\mathcal{M}_\gamma)^s\, u_{t+1-s}\rangle_\mu \\ &= \langle u_{t+1}, u_{t+1}\rangle_\mu\end{aligned}$$

so

$$\langle \mathcal{M}'\phi_t, \phi_{t+1}\rangle_\mu = \langle \phi_{t+1}, \phi_{t+1}\rangle_\mu - \langle u_{t+1}, u_{t+1}\rangle_\mu$$

Applying a symmetric argument to $\langle \phi, \mathcal{M}'L\phi\rangle_\mu$, we have

$$\langle \phi_t, \mathcal{M}'\phi_{t+1}\rangle_\mu = \langle \phi_{t+1}, \phi_{t+1}\rangle_\mu - \langle u_{t+1}, u_{t+1}\rangle_\mu$$

Consolidating terms gives

$$\begin{aligned}\langle u, u\rangle_\mu &= \mathbb{E}[\phi\phi'] + \mathbb{V}(\mathcal{M}'\phi) - \langle \phi, \mathcal{M}'\phi\rangle_\mu - \langle \mathcal{M}'\phi, \phi\rangle_\mu \\ &= \mathbb{E}[\phi\phi'] + \mathbb{V}(\mathcal{M}'\phi) - 2\langle \phi, \phi\rangle_\mu + 2\langle u, u\rangle_\mu \\ &\Longrightarrow \\ -\langle u, u\rangle_\mu &= \mathbb{V}(\mathcal{M}'\phi) - \mathbb{E}[\phi\phi'] \\ &= -\mathscr{E}_M(\phi, \phi)\end{aligned}$$

□

Remark: Explicit time indices indicate material distinctions. For example, we have not shown $\langle \mathcal{M}'\phi, \phi\rangle = \langle \phi, \mathcal{M}'\phi\rangle$, rather $\langle \mathcal{M}'\phi_t, \phi_{t+1}\rangle = \langle \phi_{t+1}, \mathcal{M}'\phi_t\rangle$.

7.4.1 Wold Representations for Second Moments

For $\epsilon > 0$, each $\varsigma \in (0, 1+\epsilon]$ defines an operator

$$(\mathcal{R}_{\boldsymbol{\gamma}}(\varsigma)\phi)(x) = \varsigma^{-1} \sum_{n=0}^{\infty} \left(\varsigma^{-1}(\mathbf{M}_{\gamma}\phi)(x)\right)^n$$

We will also consider the *resolvent* as a *function* of ς,

$$\mathcal{R}_{\boldsymbol{\gamma}}(\varsigma) = \varsigma^{-1} \sum_{n=0}^{\infty} \left(\varsigma^{-1}\mathbf{M}_{\gamma}\right)^n$$

Because $\rho(\mathbf{M}_{\gamma}) < 1$, we have $\mathcal{R}_{\boldsymbol{\gamma}}(\varsigma) = (\varsigma I - \mathbf{M}_{\gamma})^{-1}$ for $\varsigma \in (\rho(\mathbf{M}_{\gamma}), 1+\epsilon]$, and

$$\mathcal{R}_{\boldsymbol{\gamma}}(1)(I - \mathbf{M}_{\gamma}) = I$$

A proof for the case of Neumann series is given in Lax ((2010), Theorem 3 p. 195).

7.5 Identification of Spectral Gap from Realized Returns

Proposition 7.21 (Identification) *Consider the Markov environment above with rank($\mathcal{M}$) > 1. The singular value decomposition of a panel of realized returns can be expressed in terms of the Markov transition*

$$\Lambda_{PCA} \doteqdot \mathsf{V}' D_{1-\lambda} \Sigma \mathsf{V}$$

where $\doteqdot$ reads "unitarily equivalent" and where

$$\mathcal{R}_{\gamma}(1) = \mathsf{U} D_{1-\lambda} \mathsf{U}'$$

$$(D_{1-\lambda})_{i,j} = \begin{cases} \frac{1}{1-\lambda_j} & i = j \\ 0 & i \neq j \end{cases}$$

Proof Recall that for any integer k, $\mu_0 \iota' \mathcal{M}_{\boldsymbol{\gamma}}^k = \mathbf{0}$, and for any $s > 0$, $\langle u_t, u_{t-s} \rangle = 0$. In particular, we have

$$\begin{aligned}
\mathbb{V}(R_t) &= \left\langle R_t - \overline{R}, R_t - \overline{R} \right\rangle \\
&= \langle \sum_{s=0}^{\infty} (\mathcal{M}'_\gamma)^s u_{t-s}, \sum_{s=0}^{\infty} (\mathcal{M}'_\gamma)^s u_{t-s} \rangle \\
&= \langle u_t, u_t \rangle + \langle \sum_{s=1}^{\infty} (\mathcal{M}'_\gamma)^s u_{t-s}, \sum_{s=1}^{\infty} (\mathcal{M}'_\gamma)^s u_{t-s} \rangle \\
&= \langle u_t, u_t \rangle + \left\langle \mathcal{M}'_\gamma u_{t-1}, \mathcal{M}'_\gamma u_{t-1} \right\rangle + \langle \sum_{s=2}^{\infty} (\mathcal{M}'_\gamma)^s u_{t-s}, \sum_{s=2}^{\infty} (\mathcal{M}'_\gamma)^s u_{t-s} \rangle
\end{aligned}$$

Now,

$$\begin{aligned}
\left\langle \mathcal{M}'_\gamma u_{t-1}, \mathcal{M}'_\gamma u_{t-1} \right\rangle &= \left\langle \mathcal{M}'_\gamma L u_t, \mathcal{M}'_\gamma L u_t \right\rangle \\
&= \left\langle \mathcal{M}'_\gamma u_t, \mathcal{M}'_\gamma u_t \right\rangle \\
&= \left\langle \mathcal{M}_\gamma \mathcal{M}'_\gamma u_t, u_t \right\rangle \\
&= \langle \mathbf{M}_\gamma u_t, u_t \rangle
\end{aligned}$$

using the lag operator isometry and the adjoint operation. Then,

$$\mathbb{V}(R_t) = \langle u_t, u_t \rangle + \langle \mathbf{M}_\gamma u_t, u_t \rangle + \left\langle (\mathcal{M}'_\gamma)^2 u_{t-2}, (\mathcal{M}'_\gamma)^2 u_{t-2} \right\rangle + ...$$

Similarly for the second order case,

$$\begin{aligned}
\left\langle (\mathcal{M}'_\gamma)^2 u_{t-2}, (\mathcal{M}'_\gamma)^2 u_{t-2} \right\rangle &= \left\langle (\mathcal{M}'_\gamma)^2 L L u_t, (\mathcal{M}'_\gamma)^2 L L u_t \right\rangle \\
&= \left\langle (\mathcal{M}'_\gamma)^2 L u_t, (\mathcal{M}'_\gamma)^2 L u_t \right\rangle \\
&= \left\langle (\mathcal{M}'_\gamma)^2 u_t, (\mathcal{M}'_\gamma)^2 u_t \right\rangle \\
&= \left\langle \mathcal{M}_\gamma (\mathcal{M}'_\gamma)^2 u_t, \mathcal{M}'_\gamma u_t \right\rangle \\
&= \left\langle (\mathcal{M}_\gamma)^2 (\mathcal{M}'_\gamma)^2 u_t, u_t \right\rangle \\
&= \left\langle (\mathbf{M}_\gamma)^2 u_t, u_t \right\rangle
\end{aligned}$$

using associativity. Continuing,

$$\begin{aligned}
\mathbb{V}(R_t) &= \langle u_t, u_t \rangle + \langle \mathbf{M}_\gamma u_t, u_t \rangle + \left\langle (\mathbf{M}_\gamma)^2 u_t, u_t \right\rangle + \left\langle (\mathcal{M}'_\gamma)^3 u_{t-3}, (\mathcal{M}'_\gamma)^3 u_{t-3} \right\rangle + ... \\
&\vdots \\
&= \lim_{N \to \infty} \sum_{s=0}^{N} \langle (\mathbf{M}_\gamma)^s u_t, u_t \rangle
\end{aligned}$$

This sum is absolutely convergent because the eigenvalues are bounded inside the unit circle uniformly in parameter $k \in \mathbb{N}$.

Write $\sum_{n=0}^{\infty} m_n = \int_{\mathbb{R}} m(n)\nu(dn)$ for the counting measure $\hat{\nu}(n) = n\nu(n)$. Now invoke Fubini's theorem to change the order of integration. This justifies applying bilinearity of the inner product countably many times

$$\begin{aligned}\sum_{s=0}^{\infty}\langle(\mathbf{M}_{\gamma})^{s}u_{t},u_{t}\rangle &= \left\langle\sum_{s=0}^{\infty}(\mathbf{M}_{\gamma})^{s}u_{t},u_{t}\right\rangle \\ &= \langle\mathcal{R}_{\gamma}(1)u_{t},u_{t}\rangle\end{aligned}$$

Because the series is convergent, the operator $\mathcal{R}_{\boldsymbol{\gamma}}(1)$ can be written concisely in $\mathbf{M}_{\gamma}$:

$$\mathcal{R}_{\boldsymbol{\gamma}}(1) = (I - \mathbf{M}_{\gamma})^{-1}$$

We require the following

Lemma 7.22 *The operator $\boldsymbol{M}_{\gamma}$ admits a (weak) positive real point spectral decomposition $\boldsymbol{M}_{\gamma} = \mathsf{U}\Lambda_{\gamma}\mathsf{U}'$ such that every eigenvalue is bounded on unit interval $1 > m \geq (\Lambda_{\gamma})_{j,j} \geq 0$. Furthermore, at least one eigenvalue is strictly positive.*

Proof The operator $\mathcal{M}'_{\boldsymbol{\gamma}} = \mathcal{M}' - \mu_0\iota'$ has a zero eigenvalue. Moreover, any nonzero eigenvalue ϕ_j has real part strictly bounded inside of the unit interval $[-1+\epsilon_-, 1-\epsilon_+]$, for $1 > \epsilon > 0$, by the Perron-Frobenius theorem. Hence by orthogonality any eigenvalue of $\mathcal{M}_{\boldsymbol{\gamma}}$ is bounded inside the same interval. To see this, note that λ_j, ϕ_j such that $\mathcal{M}\phi_j = \lambda_j$ and $j \neq 0$,

$$\mathcal{M}\phi_j = \iota\mu'_0\phi_j + \mathcal{M}_{\boldsymbol{\gamma}}\phi_j = \mathcal{M}_{\boldsymbol{\gamma}}\phi_j = \lambda_j\phi_j$$

because the dual bases of different eigenspaces are orthogonal by construction (for every $j \neq i$, $\phi'_i\phi_j = 0$; in this case, the long run expected value of ϕ_j is zero, written $\mu'_0\phi_j = 0$). Because $\text{rank}(\mathcal{M}) > 1$, there is at least one nonzero eigenvalue of $\mathcal{M}'_{\boldsymbol{\gamma}}$, λ^*_j, with eigenvectors ψ_j,

$$\mathcal{M}'_{\boldsymbol{\gamma}}\psi_j = \lambda^*_j\psi_j \quad \lambda^*_j \in \mathbb{C} \tag{8.j}$$

Clearly, any ψ_j satisfying 8.j gives

$$\begin{aligned}[\mathcal{M}'_{\boldsymbol{\gamma}}\psi_j]'\mathcal{M}'_{\boldsymbol{\gamma}}\psi_j &= \psi'_j\mathcal{M}_{\boldsymbol{\gamma}}\mathcal{M}'_{\boldsymbol{\gamma}}\psi_j \\ &= \lambda^{*2}_j\psi'_j\psi_j > 0\end{aligned}$$

We conclude that $0 < \lambda^{*2}_j < 1$. Recall $\mathbf{M}_{\gamma} := \mathcal{M}_{\boldsymbol{\gamma}}\mathcal{M}'_{\boldsymbol{\gamma}}$ and write the above condition

$$\langle\psi'_j, \psi'_j\mathbf{M}'\rangle = \lambda^{*2}_j\langle\psi'_j, \psi'_j\rangle$$

That it is an eigenvalue of $\mathbf{M}_{\gamma}$ follows from

$$\lambda^{*2}_j = \frac{\parallel \mathcal{M}'_{\boldsymbol{\gamma}}\psi_j \parallel}{\parallel \psi_j \parallel} \geq \frac{\parallel \mathcal{M}'_{\boldsymbol{\gamma}}\psi \parallel}{\parallel \psi \parallel}$$

for any $\psi \neq \psi_j$ since ψ_j is an eigenvalue of $\mathcal{M}'_{\boldsymbol{\gamma}}$.

□

Now, using $\mathcal{R}_\gamma(1)(I - \mathsf{U}\Lambda_\gamma\mathsf{U}') = I$, and the unitary identity $\mathsf{U}' = \mathsf{U}^{-1}$, we have

$$\begin{aligned}(I - \mathbf{M}_\gamma)^{-1} &= (I - \mathsf{U}\Lambda_\gamma\mathsf{U}')^{-1} \\ &= (\mathsf{U}\mathsf{U}' - \mathsf{U}\Lambda_\gamma\mathsf{U}')^{-1} \\ &= \left(\mathsf{U}[\mathsf{U}' - \Lambda_\gamma\mathsf{U}']\right)^{-1} \\ &= \left(\mathsf{U}[I - \Lambda_\gamma]\mathsf{U}'\right)^{-1} \\ &= \mathsf{U}'\left[I - \Lambda_\gamma\right]^{-1}\mathsf{U}\end{aligned}$$

where we have used $\mathsf{U}' = \mathsf{U}^{-1}$ in the second and fifth equality and that $I - \Lambda_\gamma$ is diagonal. Define

$$(D_{1-\lambda})_{i,j} := \begin{cases}\frac{1}{1-\lambda_j} & i = j \\ 0 & i \neq j\end{cases}$$

We have shown

$$(I - \mathbf{M}_\gamma)^{-1} = \mathsf{U}'D_{1-\lambda}\mathsf{U}$$

Returning to the resolvent decomposition,

$$\begin{aligned}\mathbb{V}(\phi_t) &= \langle \mathsf{U}D_{1-\lambda}\mathsf{U}'u, u\rangle \\ &= \langle D_{1-\lambda}^{1/2}\mathsf{U}'u, D_{1-\lambda}^{1/2}\mathsf{U}'u\rangle \\ &= \langle D_{1-\lambda}^{1/2}u, D_{1-\lambda}^{1/2}u\rangle \\ &= D_{1-\lambda}^{1/2}\langle u, u\rangle D_{1-\lambda}^{1/2} \\ &= D_{1-\lambda}\Sigma\end{aligned}$$

Write $R \cdot \mu_0 = R_0$ and consider the SVD of the sample fluctuations of realized returns around their mean: $R_{t-T,t} - \overline{R}_T = \mathsf{V}D^{1/2}\mathsf{W}'$. The covariance matrix is simply

$$\begin{aligned}\mathbb{V}(R_t) &:= \langle R_t, R_t\rangle - R_0R_0 \\ &= \mathsf{V}D^{1/2}\mathsf{W}'\mathsf{W}D^{1/2}\mathsf{V}' \\ &= \mathsf{V}D\mathsf{V}'\end{aligned}$$

Of course, $D = \Lambda_{PCA}$. Hence,

$$\begin{aligned}\mathsf{V}\Lambda_{PCA}\mathsf{V}' &= D_{1-\lambda}CC' \\ \langle \mathsf{V}\Lambda_{PCA}^{1/2}, \mathsf{V}\Lambda_{PCA}^{1/2}\rangle &= \langle \mathsf{U}D_{1-\lambda}^{1/2}u, \mathsf{U}D_{1-\lambda}^{1/2}u\rangle\end{aligned}$$

□

Remark: From the proof of proposition (7.2.1), we can see the unitarily equivalent bundles specified by $\doteqdot$ can be summarized by the equalities

$$\begin{aligned} \mathsf{V}\Lambda_{PCA}\mathsf{V}' &= D_{1-\lambda}\mathsf{U}\Sigma\mathsf{U}' \\ &= D_{1-\lambda}^{1/2}\Sigma D_{1-\lambda}^{1/2} \\ &= \mathsf{U}'D_{1-\lambda}\mathsf{U}\Sigma \end{aligned}$$

7.5.1 Martingale Representation

Markov dynamics define a serially uncorrelated mean-zero process $u_{t+1} = r(x_{t+1}) - (\mathcal{M}r)(x_t)$ with boundary $u_0 = 0$. We construct a martingale u_t^* recursively $u_t^* = u_t + u_{t-1}^*$ with boundary $u_0^* = r(x_0)$. Using the definition of u_{t+1} and the operator $(\mathcal{L}r)(x_t) = (\mathcal{M}r)(x_t) - r(x_t)$, the martingale takes the form

$$u_{t+1}^* = r(x_{t+1}) - \sum_{s=0}^{t}(\mathcal{L}r)(x_{t-s}) \tag{3.2a}$$

The *generator* $\mathcal{L} = \mathcal{M} - I$ takes martingales of $\mathcal{M}$ to its kernel. Because $\mathcal{M}$ has eigenvalues in the closed unit circle, $-\mathcal{L}$ has eigenvalues in $[0, 1)$. For values near zero and $\mathcal{M}\psi_j = \lambda_j(x)\psi_j$,

$$\mathcal{L}\psi_j = -e^{-i\lambda_j(x)}\psi_j$$

is a good first-order approximation. Hence, the spectral gap is equivalently the smallest nonzero eigenvalue of the (negative) generator.

Write $\hat{\zeta}_t^{-1} := \zeta_t^{-1}\boldsymbol{\gamma}\gamma x_{t-1}$. From the construction of the martingale u_{t+1}^* in 3.2a, we see that the cumulative process associated with the changes in expected returns is, unsurprisingly, the expected return itself. Using 3.2a with $\Delta\hat{\mathbb{E}}_{t-1,1} \equiv 0$, we can express the expected return process

$$\mathbb{E}_t[R_{t+1}] = \hat{\zeta}_t^{-1}(x_t) - \sum_{n=0}^{t-1}(\mathcal{L}\hat{\zeta}_{t-1-n}^{-1})(x_{t-1-n}) \tag{3.2b}$$

in terms of first differences of the inverse spectral gap.

7.5.2 Covariance Matrix with Non-Centered SVD

The noncentered SVD puts $\phi = A\Lambda_1 B'$. Then,

$$E[\phi - E\phi][\phi - E\phi]' = E\left[A\Lambda_1 B'(I - \tfrac{1}{T}11')\left(A\Lambda_1 B'(I - \tfrac{1}{T}11')\right)'\right] = E[A\Lambda_1 B' I_1 I_1' B\Lambda A']$$

The matrix $I_1 := I - \frac{1}{T}11'$ is symmetric, hence $I_1 I_1' = (I_1)^2$. Element-wise,

$$(I_1)^2_{i,i} = (1 - T^{-1})^2 + T^{-2}(T-1) = 1 - T^{-1} = (I_1)_{i,i}, \text{and,}$$
$$(I_1)^2_{i,j} = -2(1 - T^{-1})T^{-1} + T^{-2}(T-2) = -T^{-1} = (I_1)_{i,j}$$

so I_1 is a projection. Hence $E_T[A\Lambda_1 B' I_1 I_1' B \Lambda A'] = T^{-1} A\Lambda_1 B' I^{[T]} B\Lambda A'$ where $I^{[T]} = T I_{T\times T} - 1_T 1_T' = T I_1$.

Under mild restrictions,

$$T^{-1} A\Lambda_1 B' I^{[T]} B\Lambda A' \longrightarrow_{T\to\infty} \mathbb{V}(\phi)$$

where $\mathbb{V}(\phi)$ is the population covariance of ϕ.

7.6 Changes of Measure using Deviations from μ_0

The process has long run mean μ_0. We calculate the empirical distribution $\widehat{h}_t = \frac{1}{T} h_t$ where $h_t = \sum_{s=1}^{T} \delta_x(s)$ which is different than μ_0 in almost every finite sample. Precisely, for $a_\epsilon := \mu_0 + \epsilon$,

$$\begin{aligned}\mathbb{P}(h_t \geq T a_\epsilon) \; &\leq \; e^{-kTa_\epsilon} \mathbb{E}[e^{k\, h_t}] \\ &= e^{-TI(a_\epsilon) - o(T^{-1})}\end{aligned}$$

where $I(a_\epsilon) = \sup_k \{k a_\epsilon - \log(\mathbb{E}[e^{k\, h_t}])\}$.

Remark: In the standard normal case, optimality puts $k = \epsilon$. Hence $I(x) = \frac{1}{2}x^2$ and $I(0) = I'(0) = 0$. $I(0) = 0$ expresses the law of large numbers, while $I'(0) = 0$ captures the inflection point of $I(x)$, which locally measures the rate at which the large-time outcomes can deviate from their limiting behavior. $I(x)$ is dubbed the *rate function*; see Veradhan (1979, 2008).

7.7 Proofs of Mathematical Lemmas (Online Appendix)

Proof of Lemma 7.2 For each $f \in F(X)$, the Riesz representation theorem identifies the linear functional

$$\eta(f) = \int f \eta_f = \langle f, \eta_f \rangle$$

as the inner product of f against a unique function $\eta_f \in F(X)$. We can take $\eta(f) = (\mathcal{M}f)(1)$ for consistency, but all linear functionals take the form needed.

First consider the simple functions $f = \sum_i 1_{E_i}$, for disjoint E_i with $\bigvee_i E_i = (2^X)^{\mathbb{N}}$. Using the countable additivity of the Stieltjes integral, clearly

$$(\eta(f))(E_i) = \int 1_{E_i} \eta_f = \mathbb{P}_{\eta(f)}(E_i)$$

is a countably additive set function. Because the path space is countable, the integral is finite on every subset of $(2^X)^{\mathbb{N}}$, which ensures the set function is inner and outer regular (see Tao 2010, p. 152, 1.10.12). We conclude $\mathbb{P}_{\eta(f)}(E_i)$ is a Radon measure (Tao (2010), Theorem 1.10.11.). Moreover, $\eta(f)(Z) = \sum_i \mathbb{P}(E_i) = 1$ for any disjoint partition of $(2^X)^{\mathbb{N}}$ so $\mathbb{P}_{\eta(f)}$ is a probability measure on $(X^{\mathbb{N}}, (2^X)^{\mathbb{N}})$.

We have shown every simple function $f \in F(1_X) \subset F(X)$ corresponds to a unique $\eta_f \in F(1_X)^*$ which itself corresponds to a probability measure $\mathbb{P}_{\eta(f)}$. This correspondence is unique up to functions f_1, f_2 that agree a.e. in the sense that if for every g, $\langle f_1 - f_2, g\rangle = 0$, then $\mathbb{P}_{\eta(f_1)}(E_j) = \mathbb{P}_{\eta(f_2)}(E_j)$ for every measurable $E_j \subset (2^X)^{\mathbb{N}}$.

Recall the definition of the adjoint map $\mathcal{M}'$ given $\mathcal{M}$,

$$\langle \mathcal{M}f, g\rangle = \langle f, \mathcal{M}'g\rangle$$

with $g = \eta_{\mathcal{M}f} \in F(X')^*$ and hence $\mathcal{M}'g = \mathcal{M}'\eta_{\mathcal{M}f} \in F(X)^*$. The adjoint exists by the Riesz representation theorem (Tao (2010), p. 54, 1.4.15) and (if necessary) extends to all of X^* by the Hahn-Banach theorem. In particular, the adjoint map can be constructed directly for each $\eta_f \in F(X')^*$ via the composition

$$\mathcal{M}'\eta_f = \eta_f \circ \mathcal{M}$$

(Rudin 1991, p. 98 Theorem 4.10). We conclude that for simple functions $f \in F(1_X)$, the adjoint operator $\mathcal{M}'$ maps the dual of the range of $\mathcal{M}$ to the dual of the range of $\mathcal{M}^{-1}\mathcal{M}$, ($= I$ on Hilbert space), and we can identify each element of $F(1'_X)^*$ with a probability measure over $\mathscr{R}(\mathcal{M})$.

The collection of simple functions $F(1_X)$ is dense in $F(X)$. By the Stone-Weierstrass theorem, $F(X)$ can be obtained from $F(1_X)$ by including the limit points of $F(1_X)$. Now, applying the linear functionals η_f to the limit points of $F(1_X)$, it is clear that under the weak* topology the collection of probability measures corresponding to simple functions is dense in the space of probability measures over $\mathscr{R}(\mathcal{M})$. The corresponding limit points are obtained from the weak* limits of functionals of simple functions

$$\eta(f_n) \longrightarrow \eta(f_\infty)$$

The result follows by application of the Stone-Weierstrass theorem to the dual $F(X')^*$ given $F(1'_X)^*$. □

Following Tao (2010, p.14), the Reisz functional representation $\eta(f) = \int f \mathrm{d}(\eta_f)$ defines a measure m such that $\int \mathrm{d}\eta_f = \int f \mathrm{d}m$ and for any g, $\int g\eta_f = \int gf \mathrm{d}m$.

Proof of lemma 7.7b Recall that $g(\{0\}) = 1$ implies that the subspace

$$E_0 := \{\hat{\mu} : (\mathcal{M}' - \lambda_0 I)\hat{\mu} = \mathbf{0}\}$$

has dimension one. Any $\hat{\mu}$ such that $\mathcal{M}'\hat{\mu} = \hat{\mu}$ has the form $\hat{\mu} = c\mu_0$ for non-zero $c \in \mathbb{R}/\{0\}$. Hence $\mathrm{rank}(\mu_0\iota') = 1$, and in this case, $c \equiv 1$ for every row of $\mathcal{M}$. Moreover, $\dim \mathrm{Span}(I_S) = S$. By the rank-nullity theorem, the kernel of $\mu_0\iota'$ has dimension $S - 1$. Because $\mu_0\iota' \perp I - \mu_0\iota'$, the range of $I - \mu_0\iota'$ coincides with the nullspace of $\mu_0\iota'$. Hence $\mathrm{rank}(I - \mu_0\iota') + \mathrm{rank}(\mu_0\iota') = S$. In finite dimensions, all bases are isometrically isomorphic, so if the rank of two operators are equal, their span is equivalent (up to unitary maps). We conclude

$$\{v : v = \mathcal{M}'u\,,\, u \in \mathbb{R}^S\} = \left\{v = v_0 + v_1 : v_0 = \mu_0\iota' u\,,\, v_1 = (I - \mu_0\iota')u\,,\, u \in \mathbb{R}^S\right\}$$

Together, $\mu_0\iota', (I - \mu_0\iota')$ span the image of X under $\mathcal{M}'$.

□

Definition: A σ - finite signed measure is a measure[14] $\eta : \mathcal{X} \to [-\infty, \infty]$ on a σ -field $\mathcal{X}$ such that for some countable partition $\{E_j\}_{j\in\mathbb{N}}$ with $\bigvee_j E_j = \mathcal{X}$, $\eta(\mathcal{X}) = \sum_{j\in\mathbb{N}} \eta(E_j)$ with each $\eta_j(E_j) < \infty$.

[14] A measure is a countably additive set function.

Lemma 7.23 *Every σ -finite signed measure η is the sum of two unsigned measures $\eta = \eta_+ + \eta_-$ that are mutually singular.*[15]

Proof Hahn-Jordan decomposition theorem (Tao (2010) p.17 Theorem 1.2.2; p.19 E1.2.5) □

Remark: The characteristic polynomial of $\mathcal{M}' - I$ can be expressed

$$\det(\mathcal{M}' - \lambda I) = (\lambda - \lambda_0)^{\chi(0)} \prod_{j=1}^{N_0-1} (\lambda - \lambda_j)^{\chi(j)}$$

with the algebraic multiplicity $\chi(0) = 1$. The characteristic polynomial can also be written

$$\det(\mathcal{M}' - \lambda I) = \left(\lambda - \iota'\mathcal{M}'\mu_0\right)^{\chi(0)} \prod_{j=1}^{N_0-1} \left(\lambda - v_j'\mathcal{M}'\nu_j\right)^{\chi(j)}$$

Hence for rank$(\mathcal{M}) > 1$, the Perron-Frobenius theorem implies that for some $\lambda_1 < 1$,

$$v_1'\mathcal{M}'\nu_1 = \lambda_1$$

to ensure the additional singularity outside a neighborhood of $\lambda_0 = 1$.

8 Appendix: Additional Charts and Tables

Table 8: Market and Value Decomposition by Latent Component (Unconditional)

Series	term	estimate	statistic
Market	(Intercept)	−0.587***	−7.588
	Time Effect	1.273***	28.351
	Component_2	0.014	1.467
	Component_3	0.005	0.241
Value (HML)	(Intercept)	0.206***	5.538
	Time Effect	0.883***	32.476
	Component_2	0.024**	3.699

(a) Quarterly Fama -French 3-factor and Carhart model returns data are from 1967 $Q1$ to 2016 $Q4$ available on Ken French's website. Dividends, earnings and *cay* data are from Goyal and Welch 2008, available on Amit Goyal's website.

$$R_{t+k,j} = a_0 + a_1 G_j(\widehat{\zeta}_t) + \varepsilon_t$$

[15] A measure ν_a is singular with respect to ν_b if for some $b \in \sigma(\text{supp}(\nu_b))$ such that $\nu_b(b) = 0$, $\nu_a(b) > 0$.

Table 9: Momentum Return Predictability

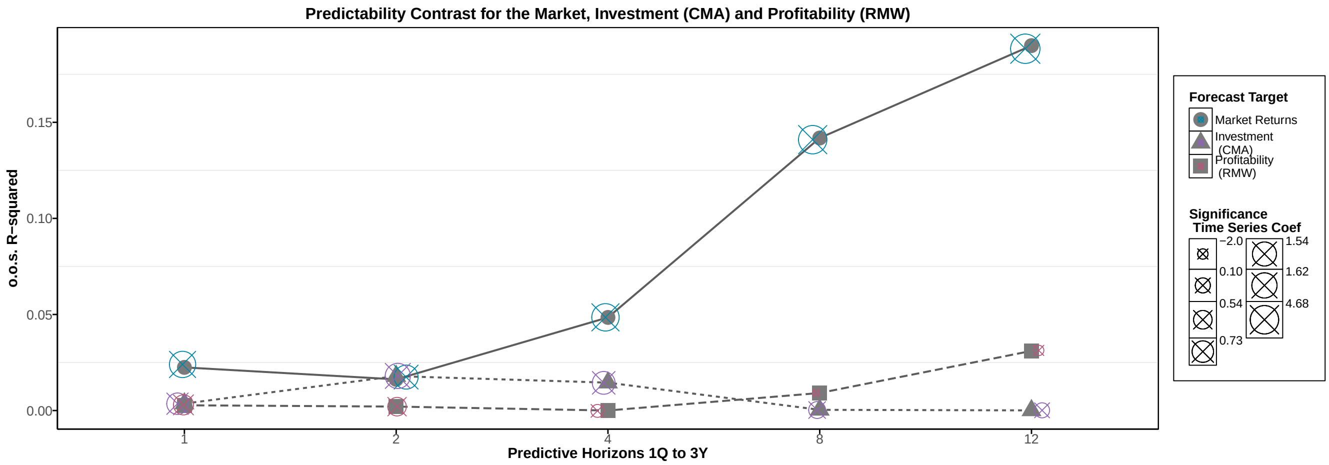
(a) Predictability regressions are calculated over the entire sample. The predictor series $\hat{\zeta}_t$ is constructed using data up to t for forecasts of returns at $t+k$ or $t+1+k$, $k \in \{1, 2, 4, 8\}$ on rolling windows of T_m =15yr histories. The predictive regression is

$$R_{t+k,j} = a_0 + a_1 G_j(\widehat{\zeta_t}) + \varepsilon_t$$

where $G_j(\hat{\zeta}_t)$ is a functional of the spectral gap time series that can vary only by portfolio, indexed by j. The rolling T_m samples are augmented with information contained in the forecast errors from forecasts made at various lags. The variable $\widehat{\zeta}$ is shorthand for the empirical estimate of the spectral gap. Tr_1 is the (non-normalized) largest contribution to the trace of the covariance matrix. Data are from $Q1$ 1967 to $Q3$ 2015 from Ken French, Goyal and Welch (2008) and CRSP.

Momentum (UMD)						
Nominal Input	horizon (quarters)	predictor	estimate.	$t : H_0 = 0$	R^2	adj.R^2
Lagged DP	1	$\widehat{\zeta}$	-0.190	-1.570	0.013	
		$\widehat{\zeta}_1$+Tr_1	-0.094	-1.569	0.013	0.008
	2	$\widehat{\zeta}$	-0.330	-1.939	0.019	
		$\widehat{\zeta}_1$+Tr_1	-0.164	-1.944	0.020	0.014
	4	$\widehat{\zeta}$	-0.723	-2.977	0.045	
		$\widehat{\zeta}_1$+Tr_1	-0.361	-3.001	0.046	0.041
	8	$\widehat{\zeta}$	-1.532	-4.748	0.112	
		$\widehat{\zeta}_1$+Tr_1	-0.770	-4.817	0.115	0.110
	12	$\widehat{\zeta}$	-2.715	-7.117	0.230	
		$\widehat{\zeta}_1$+Tr_1	-1.365	-7.239	0.236	0.231
Lagged DP	1	$\widehat{\zeta}$	-0.255	-1.802	0.017	
+		$\widehat{\zeta}_1$+Tr_1	-0.113	-1.716	0.015	0.010
CRSP Hi-Lo	2	$\widehat{\zeta}$	-0.426	-2.142	0.024	
Cash Flow		$\widehat{\zeta}_1$+Tr_1	-0.193	-2.083	0.022	0.017
	4	$\widehat{\zeta}$	-0.921	-3.269	0.054	
		$\widehat{\zeta}_1$+Tr_1	-0.421	-3.198	0.052	0.047
	8	$\widehat{\zeta}$	-1.922	-5.202	0.132	
		$\widehat{\zeta}_1$+Tr_1	-0.886	-5.117	0.128	0.123
	12	$\widehat{\zeta}$	-3.390	-7.966	0.272	
		$\widehat{\zeta}_1$+Tr_1	-1.561	-7.771	0.262	0.258
Lagged DP	1	Com.var	-0.267	-1.511	0.012	0.007
	2	Com.var	-0.511	-2.063	0.022	0.017
	4	Com.var	-1.180	-3.358	0.057	0.052
	8	Com.var	-2.527	-5.498	0.145	0.140
	12	Com.var	-4.387	-8.368	0.292	0.288

Figure 6: Predictability Term Structures



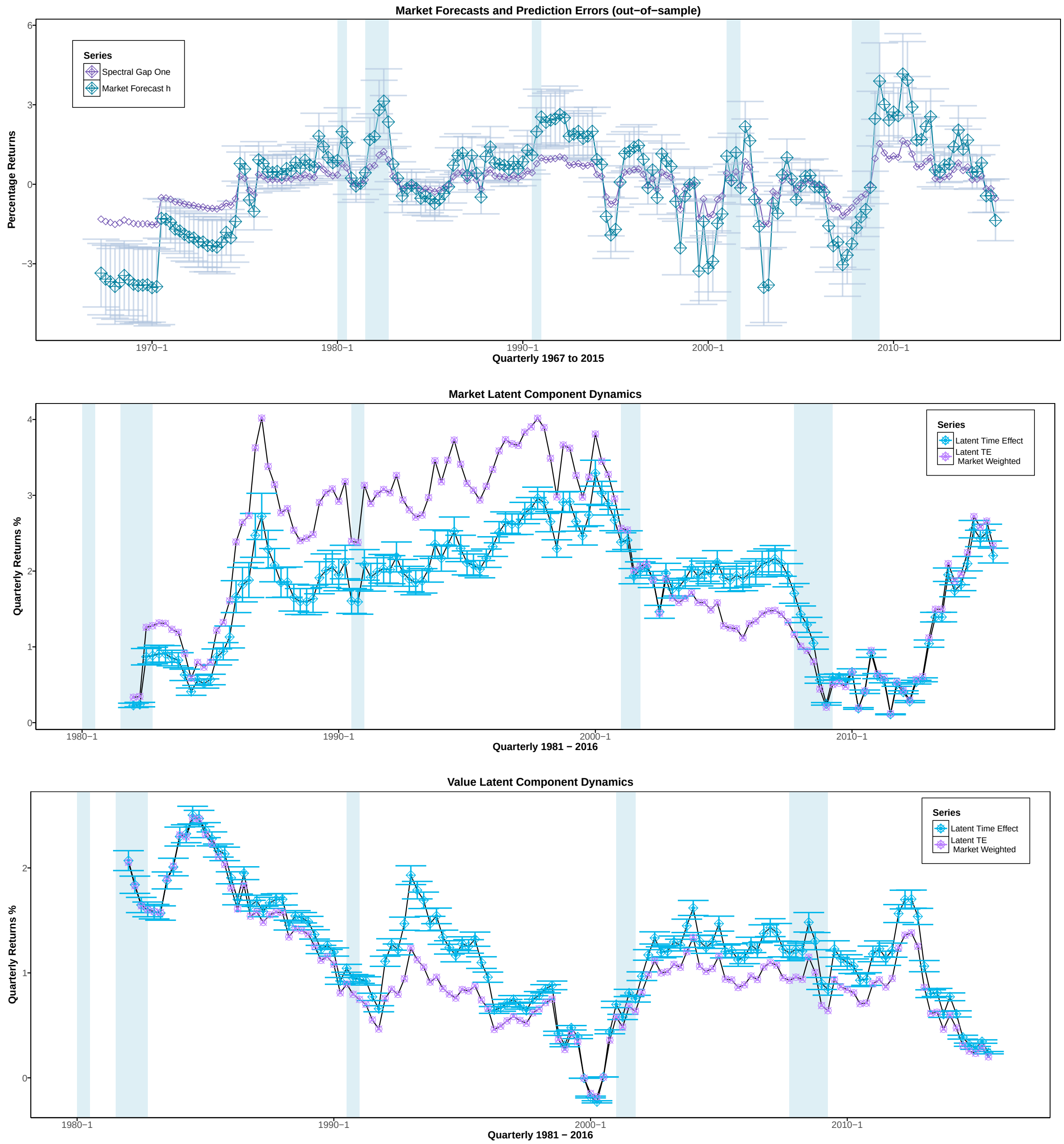
(a) Profitability and Investment factors are not predictable by levels of the spectral gap. Forecasts are given by the lagged spectral gap normalized to match the fit of realized factor returns on the rolling historical $T_m = 15$ year samples. The spectral gap is calculated up to time t on a rolling 15 year history of returns and used to forecast returns for various t+k, including returns between periods t+1 and various t+n+1. The size of circular nodes corresponds to the absolute value of the test statistic against the null of zero (t -statistics) of the time-series coefficient estimates obtained on the full sample. The Fama-French Market, Value, Size, Profitability and Investment factors are quarterly from $Q1$ 1967 to $Q3$ 2015, compounded from monthly factor returns from Ken French's website.

Table 10: Nominal Forecast Inputs from CRSP Value Weighted Return ex-Dividend

Capital gains do not contain the same information in the nominal forecast errors as dividend yields. Forecasting with capital gains for the value and momentum factors have negligible explanatory power. Capital gains based forecasts have some power for the market, presented below. The variable $\widehat{\zeta}$ is shorthand for the empirical estimate of the spectral gap. Tr_1 is the (non-normalized) largest contribution to the trace of the covariance matrix.

Portfolio	horizon (quarters)	predictor	estimate.	$t : H_0 = 0$	R^2	adj.R^2
Momentum	1	$\widehat{\zeta}_1$	1.349	1.848	0.017	0.012
(UMD)	2	$\widehat{\zeta}_1$	2.131	2.000	0.021	0.015
	4	$\widehat{\zeta}_1$	4.981	3.339	0.057	0.051
	8	$\widehat{\zeta}_1$	8.165	3.871	0.078	0.072
	12	$\widehat{\zeta}_1$	11.724	4.330	0.099	0.094

Figure 7: Time-Varying Expected Market and Value Latent Component Dynamics



(a) Time variation in mean returns is represented by the time-series of conditional forecasts. One year forecasts of the Market by the spectral gap. The spectral gap is the ratio of the first and second eigenfunctions of the Markov operator with range equal to the span of observed returns. Dynamics of the leading component in the decomposition of the market and value risk factors. Error bars are given on the series weighted by the conditional exposure to the leading component. The conditional exposure is estimated on a rolling $T_m = 15$ year window for consistency with the decomposition. Newey-West standard errors are shown. Data are quarterly from 1967 $Q1$ to 2016 $Q4$. Fama French and Carhart factor model monthly returns are from Ken French's website. NBER recessions are in blue.

Table 11: Momentum Return Predictability: Staggered Forecast Intervals

(a) Predictability regressions are calculated over the entire sample. The predictor series $\hat{\zeta}_t$ is constructed using data up to t for forecasts of returns at $t+k$ or $t+1+k$, $k \in \{1,2,4,8\}$ on rolling windows of T_m =15yr histories.The variable $\widehat{\zeta}$ is shorthand for the empirical estimate of the spectral gap. Tr_1 is the (non-normalized) largest contribution to the trace of the covariance matrix. The predictive regression is

$$R_{t+k,j} = a_0 + a_1 G_j(\widehat{\zeta}_t) + \varepsilon_t$$

where $G_j(\hat{\zeta}_t)$ is a functional of the spectral gap time series that can vary only by portfolio, indexed by j. The rolling T_m samples are augmented with information contained in the forecast errors from forecasts made at various lags. Data are from $Q1$ 1967 to $Q3$ 2015 from Ken French, Goyal and Welch (2008) and CRSP.

	Momentum (UMD)				
horizon (quarters)	predictor	estimate.	$t : H_0 = 0$	R^2	adj.R^2
$1 \mapsto 2$	$\widehat{\zeta}_1$	-0.573	-1.741	0.016	0.010
$1 \mapsto 2$	$\overline{\zeta}$	-0.458	-1.566	0.013	0.007
$1 \mapsto 2$	DP	-11.199	-0.245	0.0003	-0.005
$1 \mapsto 2$	EP	25.985	1.340	0.009	0.004
$1 \mapsto 2$	cay	-8.645	-0.397	0.001	-0.004
$1 \mapsto 3$	$\widehat{\zeta}_1$	-0.815	-1.762	0.016	0.011
$1 \mapsto 3$	$\overline{\zeta}$	-0.794	-1.932	0.019	0.014
$1 \mapsto 3$	DP	20.669	0.321	0.001	-0.005
$1 \mapsto 3$	EP	52.926	1.952	0.019	0.014
$1 \mapsto 3$	cay	-22.927	-0.732	0.003	-0.002
$1 \mapsto 5$	$\widehat{\zeta}_1$	-1.758	-2.718	0.038	0.033
$1 \mapsto 5$	$\overline{\zeta}$	-1.736	-3.022	0.046	0.041
$1 \mapsto 5$	DP	63.123	0.693	0.003	-0.003
$1 \mapsto 5$	EP	86.587	2.269	0.026	0.021
$1 \mapsto 5$	cay	-19.263	-0.426	0.001	-0.004
$1 \mapsto 9$	$\widehat{\zeta}_1$	-3.694	-4.491	0.099	0.094
$1 \mapsto 9$	$\overline{\zeta}$	-3.415	-4.612	0.104	0.099
$1 \mapsto 9$	DP	220.372	1.857	0.018	0.013
$1 \mapsto 9$	EP	130.607	2.637	0.036	0.031
$1 \mapsto 9$	cay	38.784	0.641	0.002	-0.003

Table 12: Market Return Predictability DP and $CF-10$ Nominal Inputs

(a) The spectral gap is taken from the decomposition using returns of portfolios sorted on cash flows as nominal inputs. The predictive regression is

$$R_{t+k,j} = a_0 + a_1 G_j(\widehat{\zeta_t}) + \varepsilon_t$$

where $G_j(\hat{\zeta}_t)$ is a functional of the spectral gap time series that can vary only by portfolio, indexed by j. The rolling T_m samples are augmented with information contained in the forecast errors from forecasts made at various lags. Predictability regressions are calculated over the entire sample. The variable $\widehat{\zeta}$ is shorthand for the empirical estimate of the spectral gap. Tr_1 is the (non-normalized) largest contribution to the trace of the covariance matrix. The predictor series $\hat{\zeta}_t$ is constructed using data up to t for forecasts of returns at $t+k$ or $t+1+k$, $k \in \{1, 2, 4, 8\}$ on rolling windows of T_m =15yr histories. Data are from $Q1$ 1967 to $Q3$ 2015 from Ken French, Goyal and Welch (2008). The nominal incputs are obtained from a cross-section of stocks sorted into deciles by earnings to market cap and using the high-low portfolio return as a nominal signal. Data are from CRSP.

		Full Sample Estimates		Out of Sample	
h (quarters)	predictor	$\widehat{a}_1$	$t\ H_0(0)$	R^2	adj.R^2
1	$\widehat{\zeta}$	0.851	2.630	0.035	
	$\widehat{\zeta}_1$+Tr_1	0.388	2.460	0.031	0.025
2	$\widehat{\zeta}$	1.383	2.925	0.043	
	$\widehat{\zeta}_1$+Tr_1	0.639	2.775	0.039	0.034
4	$\widehat{\zeta}$	2.728	4.125	0.084	
	$\widehat{\zeta}_1$+Tr_1	1.267	3.947	0.077	0.072
8	$\widehat{\zeta}$	4.311	4.573	0.105	
	$\widehat{\zeta}_1$+Tr_1	1.971	4.362	0.097	0.091
12	$\widehat{\zeta}$	6.054	5.051	0.131	
	$\widehat{\zeta}_1$+Tr_1	2.721	4.806	0.120	0.114

Table 13: Size Factor Predictability: Spectral Gap

(a) The predictor series $\hat{\zeta}_t$ is constructed using data up to t for forecasts of returns at $t+k$ or $t+1+k$, $k \in \{1, 2, 4, 8\}$ on rolling windows of T_m =15yr histories. The variable $\widehat{\zeta}$ is shorthand for the empirical estimate of the spectral gap. Tr_1 is the (non-normalized) largest contribution to the trace of the covariance matrix. The predictive regression is

$$R_{t+k,j} = a_0 + a_1 G_j(\widehat{\zeta}_t) + \varepsilon_t$$

where $G_j(\hat{\zeta}_t)$ is a functional of the spectral gap time series that can vary only by portfolio, indexed by j. The rolling T_m samples are augmented with information contained in the forecast errors from forecasts made at various lags. Predictability regressions are calculated over the entire sample. Data are from $Q1$ 1967 to $Q3$ 2015 from Ken French, Goyal and Welch (2008) and CRSP.

		Full Sample Estimates		Out of Sample	
h (quarters)	predictor	$\widehat{a}_1$	$t\ H_0(0)$	R^2	adj.R^2
1	$\widehat{\zeta}$	0.455	2.587	0.034	
	Tr_1	0.411	2.604	0.034	
	$\widehat{\zeta}_1$+Tr_1	0.218	2.608	0.034	0.029
2	$\widehat{\zeta}$	0.849	3.508	0.061	
	Tr_1	0.790	3.635	0.065	
	$\widehat{\zeta}_1$+Tr_1	0.413	3.593	0.064	0.059
4	$\widehat{\zeta}$	1.862	5.439	0.137	
	Tr_1	1.695	5.514	0.140	
	$\widehat{\zeta}_1$+Tr_1	0.896	5.509	0.140	0.136
8	$\widehat{\zeta}$	3.706	7.535	0.242	
	Tr_1	3.296	7.428	0.237	
	$\widehat{\zeta}_1$+Tr_1	1.762	7.527	0.241	0.237
12	$\widehat{\zeta}$	5.206	8.048	0.276	
	Tr_1	4.482	7.601	0.254	
	$\widehat{\zeta}_1$+Tr_1	2.432	7.860	0.267	0.262