Sparse Macro Factors

David E. Rapach
Saint Louis University
david.rapach@slu.edu

Guofu Zhou*
Washington University in St. Louis and CAFR
zhou@wustl.edu

October 1, 2018

*Corresponding author. Send correspondence to Guofu Zhou, Olin School of Business, Washington University in St. Louis, St. Louis, MO 63130; e-mail: zhou@wustl.edu; phone: 314-935-6384. We thank seminar participants at Emory University, Georgia State University, and Saint Louis University for insightful comments.

Sparse Macro Factors

Abstract

We use machine learning to estimate sparse principal components (PCs) for 120 monthly macro variables spanning 1960:02 to 2018:06 from the FRED-MD database. For comparison, we also extract the first ten conventional PCs from the macro variables. Each of the conventional PCs is a linear combination of all the underlying macro variables, making them difficult to interpret. In contrast, each of the sparse PCs is a sparse linear combination, whose active weights allow for intuitive economic interpretations of the sparse PCs. The first ten sparse PCs can be interpreted as yields, inflation, production, housing, employment, yield spreads, wages, optimism, money, and credit. Innovations to the conventional (sparse) PCs constitute a set of conventional (sparse) macro factors. Robust tests indicate that only one of the conventional macro factors earns a significant risk premium. In contrast, three of sparse macro factors—corresponding to yields, housing, and optimism—earn significant risk premia. Compared to leading risk factors from the literature, mimicking portfolios for the yields, housing, and optimism factors deliver sizable Sharpe ratios. A four-factor model comprised of the market factor and mimicking portfolio returns for the yields, housing, and optimism factors performs on par with or better than leading multifactor models from the literature in accounting for numerous anomalies in cross-sectional stock returns.

JEL classifications: C38, C58, G12

Key words: Sparse principal component analysis, FRED-MD, Risk premia, Factor mimicking portfolio, Three-pass regression, Multifactor models

1. Introduction

Modern finance theory posits that securities' expected returns are proportional to risk premia, which represent compensation for exposure to systematic risk factors in the macroeconomy. In addition to the return on the market portfolio, the intertemporal capital asset pricing model (ICAPM, Merton 1973) predicts that innovations to state variables that affect the "investment opportunity set" command risk premia for their exposure of investors to reinvestment risk. Because investors hold jobs and own houses and small businesses, risk premia can also originate from shocks to state variables that reflect labor and housing market conditions, as well as the prospects for small businesses (Cochrane 2005, p. 712). In general, innovations to any state variable that affects an investor's marginal utility of consumption potentially constitute a systematic risk factor that requires a risk premium.

While theory admits numerous plausible state variables, the identification of relevant systematic risk factors in security markets is ultimately an empirical issue. In their seminal study, Chen et al. (1986) use the Fama and MacBeth (1973) two-pass regression procedure to test whether a set of common macro state variables earn risk premia. With 20 size-sorted equity portfolios serving as test assets, they find evidence of significant risk premia associated with industrial production, default and term spreads, and inflation. A number of subsequent studies report that a variety of theoretically motivated state variables derived from individual macro series generate significant risk premia in cross-sectional equity returns (e.g., Cochrane 1996; Jagannathan and Wang 1996; Lettau and Ludvigson 2001; Vassalou 2003; Lustig and Nieurwerburgh 2005; Parker and Julliard 2005; Yogo 2006; Malloy et al. 2009). However, the evidence that existing "macro factors" generate significant risk premia does not appear very robust, especially after accounting for potential model misspecification (e.g., Shanken and Weinstein 2006; Kan et al. 2013;

Giglio and Xiu 2018). As the literature now stands, there is still substantial uncertainty surrounding what macro factors are the most relevant for asset pricing.

In this paper, we utilize advances in machine learning to compute *sparse macro factors* from a large macroeconomic database ("big data"). Each of the existing macro factors from the literature is typically constructed using only one or two macro variables. However, a larger number of variables are likely relevant for measuring something like labor or housing market conditions, so that reliance on one or two individual variables potentially ignores pertinent information. To incorporate information from numerous macro variables in diverse categories, we use monthly data for 120 macro variables from the extensive FRED-MD database (McCracken and Ng 2016). The categories for the 120 macro variables include output and income; the labor market; housing; consumption, orders, and inventories; money and credit; yields and exchange rates; and inflation. The 120 variables represent manifold measures of a broad array of potentially relevant risks in the macroeconomy. Employing machine learning tools, we compute sparse principal components (Jolliffe et al. 2003; Zou et al. 2006; d'Aspremont et al. 2008; Shen and Huang 2008; Sigg and Buhmann 2008; Johnstone and Lu 2009; Witten et al. 2009; Journée et al. 2010; Berthet and Rigollet 2013) to capture the systematic risks represented by the entire set of macro variables.

Conventional principal component analysis (PCA) owes its popularity to its ability to parsimoniously capture much of the information in a large number of variables. However, this comes at the cost of interpretability, as conventional principal components (PCs) are typically linear combinations of all the underlying variables. Sparse PCA restricts the cardinality of the weight vectors for the PCs, so that the PCs are sparse linear combinations of the underlying variables. By setting many weights to zero, sparse PCA facilitates interpretation of the PCs. The aim of sparse PCA is to improve interpretability via sparsity without unduly sacrificing explanatory ability.

For comparison, we begin by computing the first ten conventional PCs for the 120 macro variables. Each conventional PC is a linear combination of all 120 macro variables, which makes the individual PCs difficult to interpret. We then use the approach of Sigg and Buhmann (2008) to extract the first ten sparse PCs for the macro variables. We impose substantive sparsity in the weight vectors for the sparse PCs: each vector is only allowed to have at most twelve nonzero (or "active") elements. The high degree of sparsity facilitates economic interpretation of the sparse PCs; specifically, we intepret the individual sparse PCs as yields, inflation, production, housing, employment, yield spreads, wages, optimism, money, and credit. Fortunately, the high degree of sparsity comes at relatively little cost: the sparse PCs explain 51% of the total variation in the macro variables, compared to 59% for the conventional PCs. Based on vector autoregressions fitted to the conventional and sparse PCs (in turn), we compute innovations to the conventional and sparse PCs. Innovations to the conventional (sparse) PCs constitute a set of conventional (sparse) macro factors.

We estimate risk premia for each of the conventional and sparse macro factors using the robust three-pass methodology of Giglio and Xiu (2018), which accounts for the omission of relevant risk factors and measurement error. In addition, we use the same set of test assets as Giglio and Xiu (2018), which is comprised of 202 equity portfolios from Kenneth French's Data Library formed on a variety of firm characteristics. Based on conventional PCA, only one of the macro factors earns a significant risk premium. However, based on sparse PCA, three of macro factors—corresponding to yields, housing, and optimism—earn significant risk premia. The Giglio and Xiu (2018) three-pass methodology can also be used to compute mimicking portfolio returns for the sparse macro factors. Mimicking portfolios corresponding to the yields, housing, and optimism factors deliver sizable Sharpe ratios compared to leading risk factors from the literature. A four-factor model comprised of the market factor and mimicking portfolio returns for the yields, housing, and optimism factors also performs on par with or better than leading multifactor models from the literature in explaining 63 anomalies, as well as industry portfolio returns. Given that the sparse macro four-factor model has a straightforward economic interpretation, it represents a valuable benchmark model for asset pricing tests.

Our results indicate that sparse PCA is a valuable machine learning tool for extracting the relevant asset pricing information from a large set of macro variables. In our application, compared to conventional PCA, sparse PCA provides substantial gains in term of economic interpretation and has relatively little cost in terms of explaining the total variation in the macro variables. Moreover, sparse PCA provides important gains with respect to asset pricing: while only one of the conventional macro factors earns a significant risk premium in the cross section of equity returns, three of the sparse macro factors generate statistically and economically significant risk premia, even though the sparse PCs do not directly incorporate information from cross-sectional returns in their construction. In sum, sparse PCA appears better able than conventional PCA to filter the noise in macro variables to more reliably identify the relevant asset pricing signals in macro variables.

The rest of the paper is organized as follows. Section 2 describes sparse PCA and reports the results of conventional and sparse PCA applied to the 120 macro variables from FRED-MD. Section 3 reports risk premia estimates based on the Giglio and Xiu (2018) three-pass methodology. Section 4 compares a four-factor model comprised of the market factor and mimicking portfolio returns for the yields, housing, and optimism factors to leading multifactor models from the literature. Section 5 concludes.

2. Sparse PCA

Conventional PCA is a popular dimension-reduction technique.[1] Suppose that we are interested in reducing the dimension of the $T \times P$ data matrix $\mathbf{X}$, whose columns contain observations for P variables over T periods. We assume that the variables in $\mathbf{X}$ are

[1]See Hastie et al. (2015) for a textbook treatment of conventional and sparse PCA.

standardized, so that each column of $\mathbf{X}$ has zero mean and unit variance. Consider the T-dimensional column vector $\mathbf{X}\mathbf{w}_1$, where $\mathbf{w}_1$ is a P-dimensional column vector of weights. The first PC can be computed in the context of the following optimization problem, which entails maximizing the variance of $\mathbf{X}\mathbf{w}_1$ subject to a unit-norm constraint:

$$\arg\max_{\mathbf{w}_1} \ \mathbf{w}_1^\top \hat{\mathbf{C}} \mathbf{w}_1 \ \text{ subject to } \|\mathbf{w}_1\|_2 = 1, \tag{2.1}$$

where $\hat{\mathbf{C}} = \mathbf{X}^\top \mathbf{X}/T$ and $\|\cdot\|_2$ is the ℓ_2 norm. The first PC is given by $\hat{\mathbf{z}}_1 = \mathbf{X}\hat{\mathbf{w}}_1$, where $\hat{\mathbf{w}}_1$ is the solution to Equation (2.1).

The weight vector $\hat{\mathbf{w}}_2$ for second PC is found by maximizing the variance of $\mathbf{X}\mathbf{w}_2$ subject to $\|\mathbf{w}_2\|_2 = 1$ and $\mathbf{w}_2$ orthogonal to $\mathbf{w}_1$, while the second PC itself given by $\hat{\mathbf{z}}_2 = \mathbf{X}\hat{\mathbf{w}}_2$. Continuing in this manner, the first $\hat{p} \ll P$ PCs are given by $\hat{\mathbf{z}}_p = \mathbf{X}\hat{\mathbf{w}}_p$ for $p = 1, \ldots, \hat{p}$, where $\|\hat{\mathbf{w}}_p\|_2 = 1$ for $p = 1, \ldots, \hat{p}$ and $\hat{\mathbf{w}}_j^\top \hat{\mathbf{w}}_k = 0$ for $j \neq k$. The PCs themselves are also uncorrelated with each other. Intuitively, PCA uses the first $\hat{p}$ "dominant" PCs to reduce the dimension of the data from P to $\hat{p}$, while still capturing much of the variation in the data.

Because the elements of $\hat{\mathbf{w}}_p$ for $p = 1, \ldots, \hat{p}$ are all typically nonzero, the individual PCs can be difficult to interpret.[2] Sparse PCA harnesses machine learning tools to induce sparsity in the weight vectors. The aim is to facilitate the interpretability of the PCs without unduly sacrificing the ability of the PCs to capture the variation in the data.

We implement sparse PCA via the approach of Sigg and Buhmann (2008), which induces sparsity by directly imposing the cardinality restriction $\|\mathbf{w}_p\|_0 \leq K$ on each weight vector. Sigg and Buhmann (2008) develop a version of the expectation-maximization (EM) algorithm (Dempster et al. 1977) based on a probabilistic expression of PCA to compute

[2] Of course, we can rotate the PCs using any full rank $\hat{p} \times \hat{p}$ matrix $\mathbf{H}$, and the rotated PCs and weights will explain the same variation in $\mathbf{X}$ as the original PCs and weights. However, the individual rotated PCs can still be difficult to interpret (e.g. Jolliffe 1995).

sparse weight vectors and corresponding sparse PCs, which we denote by $\tilde{\mathbf{w}}_p$ and $\tilde{\mathbf{z}}_p$, respectively, for $p = 1, \ldots, \hat{p}$.[3]

We compute sparse PCs for 120 macro variables from the FRED-MD database (McCracken and Ng 2016). FRED-MD is a comprehensive macro database of 134 monthly U.S. variables compiled from the popular FRED database hosted by the Federal Reserve Bank of St. Louis.[4] We use 120 variables from the July 2018 vintage of FRED-MD that are available continuously starting in 1960. The variables cover a wide array of categories (output and income; labor market; housing; consumption, orders, and inventories; money and credit; interest and exchange rates; prices); as such, they capture much of the macro information available to investors.

Table 1 lists the 120 macro variables as defined by their FRED tickers and provides variable descriptions based on the Updated Appendix for the FRED-MD database.[5] Before conducting sparse PCA, we make two adjustments to the variables. First, where necessary, we transform the variables to render them stationary, as indicated in the second column of Table 1. Second, we adjust the variables for any lags in their reporting. Interest rates, exchange rates, interest rate spreads, and oil prices are reported without delay, so that no timing adjustment is needed for these variables. Nearly all of the remaining variables are reported with a one-month delay. In this case, we lag each observation by one month to account for the reporting delay. A few variables are reported with a two-month delay; accordingly, we lag each observation by two months for these variables. The timing adjustments better reflect the flow of information to investors. After making the adjustments, our sample spans 1960:02 to 2018:06.

[3] Analogously to conventional PCA, Sigg and Buhmann (2008) compute successive sparse weight vectors via iterative deflation, whereby $\tilde{\mathbf{w}}_p$ is computed after projecting the data onto the orthogonal subspace defined by the first $p-1$ sparse PCs. Unlike conventional PCs, the individual sparse PCs are not necessarily uncorrelated.

[4] FRED is available at `https://fred.stlouisfed.org/`; FRED-MD is available at Michael McCracken's webpage at `https://research.stlouisfed.org/econ/mccracken/fred-databases/`.

[5] The S&P 500 dividend yield (DIVYLD) is based on data from Robert Shiller's webpage at `http://www.econ.yale.edu/ shiller/data.htm`.

For comparison, we also conduct conventional PCA for the 120 macro variables. Table 2 reports weights for the first ten conventional PCs, while Figure 1 depicts the PCs themselves.[6] The first ten conventional PCs collectively explain 59% of the total variation in the macro variables. The weights in Table 2 illustrate the difficulty in interpreting conventionalPCs. All of the weights in Table 2 are nonzero, and the weights for individual PCs are often sizable for variables across a variety of categories; for example, the first PC evinces relatively large weights (in magnitude) for variables relating to employment, housing, yields, yield spreads, and prices. Similarly, the conventional PCs in Figure 1 reflect influences from an amalgam of variables. In sum, conventional PCs extracted from the 120 macro variables are difficult to interpret economically. Of course, conventional PCA maximizes the total variation in the underlying variables that is explained by the PCs, so that it is not designed to facilitate economic interpretation.

Table 3 and Figure 2 report sparse weights and PCs, respectively, computed using the Sigg and Buhmann (2008) EM algorithm.[7] We set $K = 12$, so that each weight vector only contains up to 10% of the 120 macro variables. The substantive sparsity in each weight vector enables us to intuitively interpret the ten sparse PCs as follows:

1. *Yields*. The second column of Table 3 shows that the first sparse PC is predominantly a linear combination of the nominal interest rates included in FRED-MD, so that we interpret the first sparse PC as "yields." In accord with this interpretation, Panel A of Figure 2 shows that the first sparse PC follows well-known fluctuations in nominal interest rates over the postwar era.

2. *Inflation*. From the third column of Table 3, we see that the second sparse PC is essentially a linear combination of various producer and consumer price indices and personal consumption expenditure deflators. We thus label the second sparse PC as "inflation." Panel B of Figure 2 indicates that the second sparse PC displays the

[6]Considering a maximum of ten, the Bai and Ng (2002) PC_{p_2} modified information criterion indicates ten significant PCs ($\hat{p} = 10$).

[7]We estimate the sparse weights using the `R` package `nsprcomp` (Sigg 2018).

well-known "Great Inflation" of the 1970s and early 1980s and subsequent "Great Disinflation."

3. *Production*. The third sparse PC is primarily a linear combination of the industrial production measures in FRED-MD, as well as manufacturing employment (see the fourth column of Table 3). We label the thrid sparse PC as "production." It exhibits downward spikes during business-cycle recessions in Panel C of Figure 2.

4. *Housing*. According to the fifth column of Table 3, the fourth sparse PC is predominantly a linear combination of the housing start and new private housing permit variables in FRED-MD, as well as real estate loans. We thus call this sparse PC "housing." The fourth sparse PC clearly depicts the housing market cycle in Panel D of Figure 2, including the long bull market from the early 1990s to the mid 2000s, followed by the housing market collapse corresponding to the Global Financial Crisis.

5. *Employment*. As shown in the sixth column of Table 3, the weights for the fifth sparse PC are concentrated in the unemployment and employment variables appearing in FRED-MD. The fifth sparse PC, which we label as "employment," exhibits clear procyclical behavior in Panel E of Figure 2.

6. *Yield spreads*. The sixth sparse PC is predominantly a linear combination of the interest rate spreads included in FRED-MD (see the seventh column of Table 3), so that we label this sparse PC as "yield spreads." The sixth sparse PC in Panel F of Figure 2 displays the distinct countercyclical fluctuations that are known to characterize yield spreads.

7. *Wages*. We refer to the seventh sparse PC as "wages," as the active elements of its weight vector include the various measures of average hourly earnings found in FRED-MD (see the eighth column of Table 3). More generally, the seventh sparse

PC appears to reflect drivers of costs for firms. This interpretation is reflected in Panel G of Figure 2, where the seventh sparse PC evinces a secular increase from the late 1960s to early 1980s, in line with the sharp increases in costs experienced by firms during this period.

8. *Optimism*. From the ninth column of Table 3, it is evident that the eighth sparse PC is a linear combination of variables from a variety of categories, including real personal income and consumption, retail sales, help wanted, overtime, and new orders for durable goods. A common theme among these variables is that they reflect how optimistic households and firms are about future economic conditions.[8] We thus label the eighth sparse PC as "optimism." The eighth sparse PC experiences sharp declines during recessions in Panel H of Figure 2, especially for the relatively deep recessions of the mid 1970s and early 1990s, as well as the recent Great Recession.

9. *Money*. As can be seen from the tenth column of Table 3, the active weights for the ninth sparse PC include all of the money stock measures in FRED-MD. Accordingly, we label this sparse PC as "money." The ninth sparse PC exhibits declines in the late 1970s and early 1980s in Panel I of Figure 2, corresponding to the "Volcker disinflation," as well as sharp increases more recently during the Global Financial Crisis, reflecting "quantitative easing" by the Fed.

10. *Credit*. The final sparse PC has relatively large loadings on credit variables from FRED-MD, including commercial and industrial loans, total nonrevolving credit, the credit-to-income ratio, consumer motor vehicle loans outstanding, and total consumer loans and leases outstanding (see the eleventh column of Table 3). As shown in Panel J of Figure 2, the tenth sparse PC tends to fall during recessions.

Comparing the results in Tables 2 and 3 and Figures 1 and 2, the substantive sparsity imposed on the weight vectors greatly facilitates economic interpretation of the sparse

[8] For example, in line with consumption-smoothing logic, corresponding increases in income and consumption indicate that households expect a persistent increase in income.

vis-à-vis the conventional PCs. Fortunately, the increased interpretability of the sparse PCs comes at relatively little cost in terms of explanatory ability: despite the high degree of sparsity, the first ten sparse PCs still explain 51% of the total variation in the 120 macro variables (compared to 59% for the first ten conventional PCs).

3. Risk Premia

As a first step in estimating risk premia, we fit first-order vector autoregressions to the set of conventional and sparse PCs (in turn), and we use the fitted processes to compute innovations to the conventional and sparse PCs. The innovations to the conventional (sparse) PCs constitute our set of conventional (sparse) macro factors. Table 4 reports correlations for the conventional and sparse macro factors. Although the conventional PCs are uncorrelated by construction, the innovations to the conventional PCs in Panel A are correlated. However, most of the correlations are relatively small in magnitude. Similarly, the innovations to the sparse PCs are correlated in Panel B, but many of the correlations are limited in magnitude. Of course, asset pricing tests do not require the macro factors to be uncorrelated.

We estimate risk premia for the macro factors using the recently developed three-pass methodology of Giglio and Xiu (2018). Their methodology has a number of attractive features. Importantly, it can recover the risk premium for any observable factor, regardless of whether the model includes all of the relevant risk factors. The methodology is also robust to measurement error. Furthermore, Giglio and Xiu (2018) show that their methodology has an "ideal" mimicking portfolio interpretation. Indeed, as a byproduct of their methodology, we can readily compute a mimicking portfolio return series for any nontradable factor.

The first step of the three-pass methodology applies conventional PCA to the sample covariance matrix for the test asset excess returns. We use the 202 portfolios from Giglio and Xiu (2018) as test assets, with data updated through 2018:06. The portfolios are all available from Kenneth French's Data Library,[9] and they are formed by sorting on a variety of firm characteristics, including size, book-to-market ratio, industry classification, operating profitability, investment, variance, net issuance, accruals, betas, and momentum. To help ensure that we span the relevant excess return space, we compute the first ten PCs and corresponding loadings for the 202 test asset excess returns.[10]

The second step of the three-pass procedure estimates risk premia for the PCs from the first step via a cross-sectional regression that relates the average excess returns for the test assets to the loadings for the PCs from the first step. The R_f^2 statistic of 67.72% in Table 5 indicates that the loadings for the first ten PCs account for much of the cross-sectional variation in expected excess returns for the test assets.

The final step of the three-pass methodology entails the estimation of a time-series regression that relates an observed factor to the set of PCs estimated in the first step. The risk premium for the observed factor is then estimated as the sum of the product of each of the estimated slope coefficients in the third-step time-series regression and their corresponding estimated risk premia in the second-step cross-sectional regression. In addition, the fitted values for the time-series regression in the third step provide a mimicking portfolio return series for the observed factor. Giglio and Xiu (2018) supply expressions for the asymptotic variance of the estimated risk premium for the observed factor. Simulations in Giglio and Xiu (2018) indicate that their three-pass methodology makes reliable inferences in finite samples of sizes similar to ours.

The first five columns of Table 5 report three-pass regression results for the conventional macro factors. The R_g^2 statistic in the fourth column is the goodness-of-fit measure for the time-series regression in the third step, while the fifth column reports the annualized Sharpe ratio for the factor mimicking portfolio. According to the t-statistics in the

[9] Available at `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`. The sample begins in 1963:07.

[10] Considering a maximum of ten, the Bai and Ng (2002) PC_{p_2} modified information criterion indicates ten significant PCs for the 202 test asset excess returns.

third column of Table 5, only one of the conventional macro factors, corresponding to the eighth conventional PC, earns a significant risk premium at the 5% level. Based on the ninth column of Table 2 and Panel H of Figure 1, it is difficult to provide a straightforward economic interpretation of this factor.

Three-pass regression results for the sparse macro factors are reported in the last five columns of Table 5. There is more widespread evidence of statistically and economically significant risk premia for the sparse macro factors. Specifically, the risk premia for the yields and housing factors are significant at the 1% level in Table 5, while that for the optimism factor is significant at the 5% level.[11] The yields, housing, and optimism factors deliver annualized Sharpe ratios ranging in magnitude from 0.53 to 1.02. These Sharpe ratios are sizable relative to those for popular risk factors from the literature (e.g., market, size, value, and momentum factors).

It is economically sensible that the yields, housing, and optimism factors earn significant risk premia in Table 5. Merton (1973) identifies the interest rate as a candidate for a state variable in the ICAPM, as changes in the interest rate affect the investment opportunity set. The significant risk premium earned by the yields factor in our analysis accords with Merton (1973) and further indicates that common fluctuations in multiple interest rates are relevant for asset pricing. Because housing is the primary form of wealth for many households, the significant risk premium earned by the housing factor in Table 5 is highly plausible. The optimism factor, which includes consumption growth, likely captures key components of an investor's marginal utility of consumption, so that its significant risk premium in Table 5 is economically reasonable.

In sum, sparse PCA offers two key advantages over conventional PCA in our application. First, by imposing substantive sparsity in the weight vectors, sparse PCA facilitates the economic interpretation of the macro factors. Second, sparse PCA better captures the information in macro variables that is relevant to investors. We find it interesting that

[11] The risk premium for the yield spreads factor is significant at the 10% level.

sparse PCA proves superior to conventional PCA for identifying systematic risk factors in the macroeconomy, as the construction of the sparse weight vectors does not directly incorporate information from cross-sectional equity returns. Our results indicate that sparsity is a powerful machine learning tool for identifying the relevant asset pricing signals in macro variables.

4. Comparisons with Leading Multifactor Models

In this section, we analyze the performance of a sparse macro four-factor model comprised of the market factor and mimicking portfolio returns for the yields, housing, and optimism factors. Similarly to Hou et al. (2015), we examine the ability of the sparse macro four-factor model to account for a plethora of anomalies in cross-section equity returns relative to leading multifactor models from the literature. The three models from the literature are the Carhart (1997) four-factor, Fama and French (2015) five-factor, and Hou et al. (2015) q-factor models. The first model augments the market, size, and value factors from Fama and French (1993) with a momentum factor (Jegadeesh and Titman 1993), while the second model adds operating profitability and investment factors to the three Fama and French (1993) factors. Finally, the q-factor model includes market, size, investment, and profitability factors.[12] We estimate pricing errors for each multifactor model for two sets of test assets using the Gibbons et al. (1989) framework.

The first set of test assets is comprised of 63 spread portfolio returns for 1974:01 to 2016:12 from Huang et al. (2018) that represent numerous anomalies from the literature. For each of the multifactor models, Panel A of Table 6 reports the average of the absolute values of the alphas across the 63 portfolios, as well as the Gibbons et al. (1989) W_u statistics for testing the null hypothesis that the alphas are jointly zero. The average magnitude of the alphas is largest for the Carhart (1997) four-factor model (0.37%), followed

[12] Factor data for the Carhart (1997) four-factor and Fama and French (2015) five-factor models are from Kenneth French's Data Library. We thank Kewei Hou for providing data for the Hou et al. (2015) q factors.

by the Fama and French (2015) five-factor model (0.28%). The average magnitude of the alphas is lowest for the Hou et al. (2015) q-factor model (0.23%), while that for the sparse macro four-factor model is nearly the same (0.24%). The null hypothesis that the alphas are jointly zero is rejected at the 1% level for all four multifactor models, with the Carhart (1997) four-factor and Fama and French (2015) five-factor models displaying the largest test statistic values.

Following the recommendation of Lewellen et al. (2010), we also examine the ability of the multifactor models to explain industry portfolio excess returns in Panel B of Table 6. The test assets are 30 industry portfolios from Kenneth French's Data Library, and the sample period covers 1967:01 to 2016:12.[13] The sparse macro four-factor model generates the smallest average magnitude of the alphas (0.17%), followed closely by the Carhart (1997) four-factor model (0.18%). The Hou et al. (2015) q-factor and Fama and French (2015) five-factor models produce average alpha magnitudes of 0.20% and 0.24%, respectively. As in Panel A, the W_u statistic is significant at the 1% level for all four models. The test statistic takes the smallest value for the Hou et al. (2015) q-factor model, followed by the sparse macro four-factor and Carhart (1997) four-factor models.

Overall, the sparse macro four-factor models performs on par with or better than leading multifactor models from the literature with respect to explaining challenging aspects of cross-sectional equity returns. The sparse macro four-factor model has the added benefit that its factors have a straightforward economic interpretation, so that the model constitutes an informative benchmark for asset pricing tests.

5. Conclusion

In this paper, we extract macro risk factors from a comprehensive data set of 120 monthly macro variables covering 1960:02 to 2018:06 from the FRED-MD database. We apply, for

[13] The sample period is based on data availability for the q factors.

the first time in finance, sparse PCA to obtain macro factors as sparse linear combinations of the underlying macro variables. Sparse PCA has two advantages relative to conventional PCA. First, it allows for natural economic interpretation of the factors. Second, it reduces the noise in irrelevant variables. Using a set of 202 portfolios formed from a variety of firm characteristics as test assets, we find three major macro risk factors—corresponding to yields, housing, and optimism—in the cross section of US stock returns. The macro factors earn significant risk premia based the state-of-the-art three-pass methodology of Giglio and Xiu (2018). Mimicking portfolios for the three macro factors deliver large annualized Sharpe ratios (in magnitude). In addition, a four-factor model comprised of the market factor and mimicking portfolio returns for the yields, housing, and optimism factors generally performs as well as or better than leading multifactor models from the literature in accounting for anomalies in cross-sectional equity returns.

With respect to future research, sparse PCA can be applied—potentially as widely as conventional PCA—to assess macro risk factors not only for stocks, but also for bonds, currencies, and other assets. In addition, the impact of other types of economic variables, such as investor sentiment and news, can be analyzed using sparse PCA. Indeed, the information in any type of "big data" in finance can be conveniently summarized by sparse PCA. Our results suggest that sparse PCA is promising strategy for identifying the most relevant information for investors in large data sets.

References

Bai, J. and S. Ng (2002). Determining the Number of Factors in Approximate Factor Models. *Econometrica* 70:1, 191–221.

Berthet, Q. and P. Rigollet (2013). Optimal Detection of Sparse Principal Components in High Dimension. *Annals of Statistics* 41:4, 1780–1815.

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

Chen, N.-F., R. Roll, and S. A. Ross (1986). Economic Forces and the Stock Market. *Journal of Business* 59:3, 383–403.

Cochrane, J. H. (1996). A Cross-Sectional Test of an Investment-Based Asset Pricing Model. *Journal of Political Economy* 104:3, 572–621.

Cochrane, J. H. (2005). *Asset Pricing*. Revised Edition. Princeton, NJ: Princeton University Press.

d'Aspremont, A., F. Bach, and L. E. Ghaoui (2008). Optimal Solutions for Sparse Principal Component Analysis. *Journal of Machine Learning Research* 9:1, 1269–1294.

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

Fama, E. F. and K. R. French (1993). Common Risk Factors in the Returns on Stocks and Bonds. *Journal of Financial Economics* 33:1, 3–56.

Fama, E. F. and K. R. French (2015). A Five-Factor Asset Pricing Model. *Journal of Financial Economics* 116:1, 1–22.

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

Gibbons, M. R., S. A. Ross, and J. Shanken (1989). A Test of the Efficiency of a Given Portfolio. *Econometrica* 57:5, 1121–1152.

Giglio, S. and D. Xiu (2018). Asset Pricing with Omitted Factors. Manuscript.

Hastie, T., R. Tibshirani, and M. Wainwright (2015). *Statistical Learning with Sparsity: The Lasso and Generalizations*. London: CRC Press.

Hou, K., C. Xue, and L. Zhang (2015). Digesting Anomalies: An Investment Approach. *Review of Financial Studies* 28:3, 650–703.

Huang, D., J. Li, and G. Zhou (2018). Shrinking Factor Dimension: A Reduced-Rank Approach. Manuscript.

Jagannathan, R. and Z. Wang (1996). The Conditional CAPM and the Cross-Section of Expected Returns. *Journal of Finance* 51:1, 3–53.

Jegadeesh, N. and S. Titman (1993). Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency. *Journal of Finance* 48:1, 65–91.

Johnstone, I. M. and A. Y. Lu (2009). On Consistency and Sparsity for Principal Components Analysis in High Dimensions. *Journal of the American Statistical Association* 104:486, 682–703.

Jolliffe, I. T. (1995). Rotation of Principal Components: Choice of Normalization Constraints. *Journal of Applied Statistics* 22:1, 29–35.

Jolliffe, I. T., N. T. Trendafilov, and M. Uddin (2003). A Modified Principal Component Technique Based on the LASSO. *Journal of Computational and Graphical Statistics* 12:3, 531–537.

Journée, M., Y. Nesterov, P. Richtárik, and R. Sepulchre (2010). Generalized Power Method for Sparse Principal Component Analysis. *Journal of Machine Learning Research* 11:1, 517–533.

Kan, R., C. Robotti, and J. Shanken (2013). Pricing Model Performance and the Two-Pass Cross-Sectional Regression Methodology. *Journal of Finance* 68:6, 2617–2649.

Lettau, M. and S. Ludvigson (2001). Resurrecting the (C)CAPM: A Cross-Sectional Test When Risk Premia Are Time-Varying. *Journal of Political Economy* 109:6, 1238–1287.

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

Lustig, H. and S. G. V. Nieurwerburgh (2005). Housing Collateral, Consumption Insurance and Risk Premia: An Empirical Perspective. *Journal of Finance* 60:3, 1167–1219.

Malloy, C. J., T. J. Moskowitz, and A. Vissing-Jørgensen (2009). Long-Run Stockholder Consumption Risk and Asset Returns. *Journal of Finance* 64:6, 2427–2479.

McCracken, M. W. and S. Ng (2016). FRED-MD: A Monthly Database for Macroeconomic Research. *Journal of Business and Economic Statistics* 34:4, 574–589.

Merton, R. C. (1973). An Intertemporal Capital Asset Pricing Model. *Econometrica* 41:5, 867–887.

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

Parker, J. A. and C. Julliard (2005). Consumption Risk and the Cross Section of Expected Returns. *Journal of Political Economy* 113:1, 185–222.

Shanken, J. and M. I. Weinstein (2006). Economic Forces and the Stock Market Revisited. *Journal of Empirical Finance* 13:2, 129–144.

Shen, H. and J. Z. Huang (2008). Sparse Principal Component Analysis Via Regularized Low Rank Matrix Approximation. *Journal of Multivariate Analysis* 99:6, 1015–1034.

Sigg, C. D. (2018). `nsprcomp`. `R` Package Version 0.5.1–2.

Sigg, C. D. and J. M. Buhmann (2008). "Expectation-Maximization for Sparse and Non-Negative PCA". In *Proceedings of the 25th International Conference on Machine Learning*, pp. 960–967.

Vassalou, M. (2003). News Related to Future GDP Growth as a Risk Factor in Equity Returns. *Journal of Financial Economics* 68:1, 47–73.

Witten, D. M., R. Tibshirani, and T. Hastie (2009). A Penalized Matrix Decomposition, with Applications to Sparse Principal Components and Canonical Correlation Analysis. *Biostatistics* 10:3, 515–534.

Yogo, M. (2006). A Consumption-Based Explanation of Expected Stock Returns. *Journal of Finance* 61:2, 539–580.

Zou, H., T. Hastie, and R. Tibshirani (2006). Sparse Principal Component Analysis. *Journal of Computational and Graphical Statistics* 15:2, 265–286.

Table 1
Macro variable descriptions

(1) FRED ticker	(2) Transformation	(3) Description
RPI	Log growth	Real Personal Income
W875RX1	Log growth	Real Personal Income Excluding Transfer Receipts
DPCERA3M086SBEA	Log growth	Real Personal Consumption Expenditures
CMRMTSPLx	Log growth	Real Manufacturing and Trade Industries Sales
RETAILx	Log growth	Retail and Food Services Sales
INDPRO	Log growth	Industrial Production Index
IPFPNSS	Log growth	Industrial Production Index: Final Products and Nonindustrial Supplies
IPFINAL	Log growth	Industrial Production Index: Final Products (Market Group)
IPCONGD	Log growth	Industrial Production Index: Consumer Goods
IPDCONGD	Log growth	Industrial Production Index: Durable Consumer Goods
IPNCONGD	Log growth	Industrial Production Index: Nondurable Consumer Goods
IPBUSEQ	Log growth	Industrial Production Index: Business Equipment
IPMAT	Log growth	Industrial Production Index: Materials
IPDMAT	Log growth	Industrial Production Index: Durable Materials
IPNMAT	Log growth	Industrial Production Index: Nondurable Materials
IPMANSICS	Log growth	Industrial Production Index: Manufacturing (SIC)
IPB51222S	Log growth	Industrial Production Index: Residential Utilities
IPFUELS	Log growth	Industrial Production Index: Fuels
CUMFNS	Difference	Capacity Utilization: Manufacturing
HWI	Difference	Help-Wanted Index for United States
HWIURATIO	Difference	Ratio of Help Wanted to Number Unemployed
CLF16OV	Log growth	Civilian Labor Force
CE16OV	Log growth	Civilian Employment
UNRATE	Difference	Civilian Unemployment Rate
UEMPMEAN	Difference	Average Duration of Unemployment (Weeks)
UEMPLT5	Log growth	Civilians Unemployed—Less Than 5 Weeks
UEMP5TO14	Log growth	Civilians Unemployed for 5–14 Weeks
UEMP15OV	Log growth	Civilians Unemployed—15 Weeks and Over
UEMP15T26	Log growth	Civilians Unemployed for 15–26 Weeks
UEMP27OV	Log growth	Civilians Unemployed for 27 Weeks and Over

The table provides descriptions of 120 macro variables from the FRED-MD database. The first column identifies the variable based on its FRED ticker. (DIVYLD is the S&P 500 dividend yield based on data from Robert Shiller's webpage.) The second column reports how the variable is transformed before computing the principal components; - indicates that the variable is not transformed. The description in the third column is based on the FRED-MD Updated Appendix.

Table 1 (continued)

(1) FRED ticker	(2) Transformation	(3) Description
CLAIMSx	Log growth	Initial Claims
PAYEMS	Log growth	All Employees: Total Nonfarm
USGOOD	Log growth	All Employees: Goods-Producing Industries
CES1021000001	Log growth	All Employees: Mining and Logging: Mining
USCONS	Log growth	All Employees: Construction
MANEMP	Log growth	All Employees: Manufacturing
DMANEMP	Log growth	All Employees: Durable Goods
NDMANEMP	Log growth	All Employees: Nondurable Goods
SRVPRD	Log growth	All Employees: Service-Providing Industries
USTPU	Log growth	All Employees: Trade, Transportation, and Utilities
USWTRADE	Log growth	All Employees: Wholesale Trade
USTRADE	Log growth	All Employees: Retail Trade
USFIRE	Log growth	All Employees: Financial Activities
USGOVT	Log growth	All Employees: Government
CES0600000007	Log growth	Average Weekly Hours: Goods Producing
AWOTMAN	Log growth	Average Weekly Overtime Hours: Manufacturing
AWHMAN	Log growth	Average Weekly Hours: Manufacturing
HOUST	Log	Housing Starts: Total New Privately Owned
HOUSTNE	Log	Housing Starts: Total New Privately Owned, Northeast
HOUSTMW	Log	Housing Starts: Total New Privately Owned, Midwest
HOUSTS	Log	Housing Starts: Total New Privately Owned, South
HOUSTW	Log	Housing Starts: Total New Privately Owned, West
PERMIT	Log	New Private Housing Permits
PERMITNE	Log	New Private Housing Permits, Northeast
PERMITMW	Log	New Private Housing Permits, Midwest
PERMITS	Log	New Private Housing Permits, South
PERMITW	Log	New Private Housing Permits, West
AMDMNOx	Log growth	New Orders for Durable Goods
AMDMUOx	Log growth	Unfilled Orders for Durable Goods
BUSINVx	Log growth	Total Business Inventories

Table 1 (continued)

(1) FRED ticker	(2) Transformation	(3) Description
ISRATIOx	Difference	Total Business: Inventories to Sales Ratio
M1SL	Log growth	M1 Money Stock
M2SL	Log growth	M2 Money Stock
M2REAL	Log growth	Real M2 Money Stock
AMBSL	Log growth	St. Louis Adjusted Monetary Base
TOTRESNS	Log growth	Total Reserves of Depository Institutions
NONBORRES	Log growth	Reserves Of Depository Institutions
BUSLOANS	Log growth	Commercial and Industrial Loans
REALLN	Log growth	Real Estate Loans at All Commercial Banks
NONREVSL	Log growth	Total Nonrevolving Credit
CONSPI	Difference	Ratio of Nonrevolving Consumer Credit to Personal Income
DIVYLD	-	S&P 500 Dividend Yield
FEDFUNDS	-	Effective Federal Funds Rate
CP3Mx	-	3-Month AA Financial Commercial Paper Rate
TB3MS	-	3-Month Treasury Bill Rate
TB6MS	-	6-Month Treasury Bill Rate
GS1	-	1-Year Treasury Rate
GS5	-	5-Year Treasury Rate
GS10	-	10-Year Treasury Rate
AAA	-	Moody's Seasoned Aaa Corporate Bond Yield
BAA	-	Moody's Seasoned Baa Corporate Bond Yield
COMPAPFFx	-	CP3Mx Minus FEDFUNDS
TB3SMFFM	-	TB3MS Minus FEDFUNDS
TB6SMFFM	-	TB6MS Minus FEDFUNDS
T1YFFM	-	GS1 Minus FEDFUNDS
T5YFFM	-	GS5 Minus FEDFUNDS
T10YFFM	-	GS10 Minus FEDFUNDS
AAAFFM	-	AAA Minus FEDFUNDS
BAAFFM	-	BAA Minus FEDFUNDS
EXSZUSx	Log growth	Switzerland / U.S. Foreign Exchange Rate

Table 1 (continued)

(1) FRED ticker	(2) Transformation	(3) Description
EXJPUSx	Log growth	Japan / U.S. Foreign Exchange Rate
EXUSUKx	Log growth	U.S. / U.K. Foreign Exchange Rate
EXCAUSx	Log growth	Canada / U.S. Foreign Exchange Rate
WPSFD49207	Log growth	Producer Price Index: Finished Goods
WPSFD49502	Log growth	Producer Price Index: Finished Consumer Goods
WPSID61	Log growth	Producer Price Index: Intermediate Materials
WPSID62	Log growth	Producer Price Index: Crude Materials
OILPRICEx	Log growth	Crude Oil, Spliced WTI and Cushing
PPICMM	Log growth	Producer Price Index: Metals and metal products
CPIAUCSL	Log growth	Consumer Price Index: All Items
CPIAPPSL	Log growth	Consumer Price Index: Apparel
CPITRNSL	Log growth	Consumer Price Index: Transportation
CPIMEDSL	Log growth	Consumer Price Index: Medical Care
CUSR0000SAC	Log growth	Consumer Price Index: Commodities
CUSR0000SAD	Log growth	Consumer Price Index: Durables
CUSR0000SAS	Log growth	Consumer Price Index: Services
CPIULFSL	Log growth	Consumer Price Index: All Items Less Food
CUSR0000SA0L2	Log growth	Consumer Price Index: All items Less Shelter
CUSR0000SA0L5	Log growth	Consumer Price Index: All Items Less Medical Care
PCEPI	Log growth	Personal Consumption Expenditures Deflator
DDURRG3M086SBEA	Log growth	Personal Consumption Expenditures Deflator: Durable Goods
DNDGRG3M086SBEA	Log growth	Personal Consumption Expenditures Deflator: Nondurable Goods
DSERRG3M086SBEA	Log growth	Personal Consumption Expenditures Deflator: Services
CES0600000008	Log growth	Average Hourly Earnings: Goods Producing
CES2000000008	Log growth	Average Hourly Earnings: Construction
CES3000000008	Log growth	Average Hourly Earnings: Manufacturing
MZMSL	Log growth	MZM Money Stock
DTCOLNVHFNM	Log growth	Consumer Motor Vehicle Loans Outstanding
DTCTHFNM	Log growth	Total Consumer Loans and Leases Outstanding
INVEST	Log growth	Securities in Bank Credit at All Commercial Banks

Table 2
Principal component weights for 120 macro variables

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
RPI	0.01	−0.08	−0.07	0.05	−0.05	−0.01	0.01	0.02	0.24	−0.09
W875RX1	0.02	−0.10	−0.07	0.05	−0.06	−0.03	−0.01	0.01	0.23	−0.09
DPCERA3M086SBEA	0.01	−0.07	−0.04	0.04	0.03	0.01	0.04	0.15	0.05	−0.14
CMRMTSPLx	0.02	−0.07	0.01	−0.02	0.04	−0.06	−0.04	−0.16	−0.01	0.27
RETAILx	0.05	−0.05	0.05	0.02	0.06	−0.02	0.01	0.17	0.02	−0.22
INDPRO	0.06	−0.19	0.06	0.17	0.01	0.12	0.001	−0.01	−0.01	−0.01
IPFPNSS	0.07	−0.18	0.05	0.18	0.02	0.14	0.02	0.02	0.01	−0.05
IPFINAL	0.06	−0.16	0.05	0.19	0.01	0.16	0.03	0.004	0.02	−0.05
IPCONGD	0.04	−0.14	0.04	0.21	0.05	0.17	0.07	0.01	−0.003	−0.06
IPDCONGD	0.03	−0.13	0.05	0.19	0.05	0.12	0.005	−0.07	0.04	−0.06
IPNCONGD	0.03	−0.08	0.01	0.13	0.03	0.15	0.11	0.09	−0.04	−0.02
IPBUSEQ	0.07	−0.15	0.05	0.10	−0.05	0.10	−0.03	0.005	0.05	−0.01
IPMAT	0.04	−0.17	0.06	0.14	0.003	0.09	−0.02	−0.04	−0.02	0.03
IPDMAT	0.05	−0.17	0.08	0.12	0.02	0.08	−0.07	−0.08	0.03	0.03
IPNMAT	0.04	−0.12	0.04	0.09	0.03	0.05	0.02	0.04	−0.08	0.04
IPMANSICS	0.07	−0.19	0.07	0.16	0.03	0.10	−0.02	−0.01	0.001	−0.02
IPB51222S	−0.005	0.004	−0.01	0.05	−0.01	0.13	0.10	0.01	−0.03	0.11
IPFUELS	0.002	−0.02	−0.003	0.04	0.01	0.04	0.05	0.04	−0.06	−0.04
CUMFNS	0.05	−0.18	0.08	0.18	0.05	0.09	−0.06	−0.03	−0.01	−0.03
HWI	0.01	−0.07	0.02	−0.01	−0.01	−0.06	0.01	−0.03	−0.13	0.24
HWIURATIO	0.01	−0.10	0.01	−0.02	−0.02	−0.07	0.001	−0.03	−0.12	0.25
CLF16OV	0.05	−0.02	−0.02	−0.03	0.005	−0.10	0.05	0.17	−0.02	−0.24
CE16OV	0.06	−0.10	−0.002	−0.03	−0.02	−0.16	−0.03	0.08	−0.08	−0.18
UNRATE	−0.03	0.13	−0.03	−0.003	0.03	0.12	0.12	0.11	0.09	−0.06
UEMPMEAN	−0.02	0.03	0.04	0.10	0.08	0.06	0.03	−0.03	0.16	−0.05
UEMPLT5	0.003	0.03	−0.02	−0.04	−0.03	−0.004	0.04	0.10	0.07	−0.03
UEMP5TO14	−0.01	0.07	−0.03	−0.01	−0.004	0.06	0.06	0.05	−0.04	−0.05
UEMP15OV	−0.03	0.11	−0.004	0.07	0.10	0.14	0.13	0.05	0.16	−0.07
UEMP15T26	−0.02	0.08	−0.02	0.01	0.04	0.09	0.10	0.03	0.13	−0.03
UEMP27OV	−0.02	0.08	0.01	0.09	0.10	0.12	0.09	0.03	0.10	−0.06

The table reports weights for the first eight principal components extracted from 120 macro variables (listed in Table 1) from the FRED-MD database. The sample period is 1960:02 to 2018:06. The variable name in the first column corresponds to its FRED ticker.

Table 2 (continued)

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
CLAIMSx	−0.01	0.10	−0.02	−0.09	−0.05	0.03	0.04	0.04	−0.01	0.005
PAYEMS	0.10	−0.18	0.01	−0.02	−0.08	−0.15	0.04	0.07	0.03	0.02
USGOOD	0.08	−0.19	0.04	−0.0005	−0.09	−0.11	−0.03	−0.01	−0.01	−0.03
CES1021000001	0.03	−0.01	0.05	−0.01	−0.06	−0.01	0.01	0.003	0.03	−0.07
USCONS	0.05	−0.13	−0.01	−0.05	−0.01	−0.10	−0.10	0.10	−0.03	−0.16
MANEMP	0.07	−0.18	0.06	0.04	−0.10	−0.10	0.02	−0.06	0.003	0.05
DMANEMP	0.07	−0.18	0.05	0.04	−0.10	−0.07	0.02	−0.08	0.02	0.04
NDMANEMP	0.06	−0.14	0.06	0.02	−0.08	−0.15	0.04	0.05	−0.05	0.07
SRVPRD	0.10	−0.12	−0.02	−0.05	−0.06	−0.17	0.12	0.13	0.07	0.06
USTPU	0.10	−0.13	0.01	−0.05	−0.06	−0.21	0.02	0.09	0.03	0.03
USWTRADE	0.11	−0.12	0.01	−0.07	−0.09	−0.14	0.04	0.04	0.02	0.04
USTRADE	0.09	−0.11	−0.005	−0.03	−0.02	−0.21	0.04	0.09	0.06	0.01
USFIRE	0.11	−0.08	−0.08	−0.05	−0.003	−0.09	0.10	0.02	0.02	0.02
USGOVT	0.03	−0.02	−0.05	−0.06	−0.02	0.004	0.27	0.11	0.10	0.09
CES0600000007	−0.05	−0.12	0.08	−0.10	−0.21	−0.01	−0.14	0.04	−0.06	−0.04
AWOTMAN	0.02	−0.07	0.05	0.09	0.05	0.06	−0.03	−0.07	0.01	0.03
AWHMAN	−0.05	−0.12	0.09	−0.12	−0.20	0.01	−0.14	0.02	−0.06	−0.01
HOUST	0.13	−0.09	−0.17	−0.15	0.13	0.09	0.01	−0.01	−0.04	−0.02
HOUSTNE	0.11	−0.08	−0.14	−0.12	0.08	0.04	0.19	0.03	0.05	0.02
HOUSTMW	0.11	−0.08	−0.13	−0.13	0.12	0.09	0.14	0.02	0.02	0.01
HOUSTS	0.11	−0.09	−0.17	−0.13	0.11	0.09	−0.12	−0.05	−0.08	−0.05
HOUSTW	0.11	−0.08	−0.16	−0.15	0.15	0.09	−0.04	−0.01	−0.05	−0.01
PERMIT	0.10	−0.10	−0.16	−0.17	0.14	0.11	−0.09	−0.03	−0.08	−0.03
PERMITNE	0.10	−0.09	−0.14	−0.14	0.09	0.07	0.17	0.01	0.02	0.01
PERMITMW	0.10	−0.10	−0.13	−0.17	0.13	0.11	0.10	0.02	−0.02	−0.01
PERMITS	0.07	−0.07	−0.13	−0.13	0.10	0.10	−0.28	−0.06	−0.12	−0.06
PERMITW	0.11	−0.08	−0.16	−0.15	0.15	0.10	−0.06	−0.01	−0.07	−0.02
AMDMNOx	0.03	−0.07	0.04	0.07	0.02	0.04	−0.002	−0.02	0.01	−0.10
AMDMUOx	0.09	−0.07	0.02	−0.07	−0.12	0.01	0.03	−0.10	−0.03	−0.04
BUSINVx	0.11	−0.004	0.04	−0.05	−0.14	−0.06	0.07	0.10	0.08	−0.10

Table 2 (continued)

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
ISRATIOx	−0.05	0.05	−0.04	0.01	−0.11	0.02	0.05	0.23	0.04	−0.31
M1SL	−0.02	0.02	−0.02	0.07	0.05	−0.19	0.02	−0.23	−0.07	−0.07
M2SL	0.00	0.003	−0.11	0.06	0.11	−0.15	0.12	−0.30	−0.08	−0.15
M2REAL	−0.06	−0.06	−0.19	0.07	0.03	−0.10	0.08	−0.21	−0.05	−0.09
AMBSL	−0.03	0.03	−0.08	0.08	−0.04	−0.09	0.05	−0.11	−0.01	0.00
TOTRESNS	−0.03	0.03	−0.08	0.05	−0.04	−0.09	0.08	−0.15	−0.03	−0.09
NONBORRES	0.001	−0.0002	0.002	−0.005	0.005	0.01	−0.02	−0.02	−0.02	0.09
BUSLOANS	0.03	−0.03	−0.04	−0.08	−0.17	−0.05	0.06	−0.08	−0.05	−0.06
REALLN	0.05	−0.05	−0.15	−0.08	0.04	0.03	0.06	−0.05	−0.06	−0.03
NONREVSL	0.06	−0.06	−0.05	−0.10	−0.05	−0.05	−0.11	−0.18	0.38	0.02
CONSPI	−0.03	−0.03	−0.03	−0.09	−0.04	−0.004	−0.13	−0.15	0.40	−0.02
DIVYLD	0.00	−0.004	0.02	−0.02	−0.07	−0.05	0.06	−0.09	−0.06	−0.05
FEDFUNDS	0.09	0.09	−0.10	0.09	−0.05	−0.01	−0.07	0.04	−0.01	0.03
CP3Mx	0.09	0.09	−0.10	0.10	−0.04	−0.02	−0.04	0.04	−0.01	0.03
TB3MS	0.08	0.08	−0.11	0.10	−0.03	−0.03	−0.07	0.05	−0.01	0.03
TB6MS	0.08	0.08	−0.11	0.10	−0.02	−0.04	−0.06	0.06	−0.005	0.03
GS1	0.08	0.08	−0.11	0.11	−0.01	−0.05	−0.07	0.06	−0.0001	0.04
GS5	0.08	0.08	−0.11	0.13	0.05	−0.08	−0.11	0.09	0.01	0.05
GS10	0.08	0.08	−0.11	0.15	0.07	−0.10	−0.14	0.09	0.02	0.06
AAA	0.09	0.09	−0.11	0.16	0.08	−0.10	−0.19	0.08	0.002	0.04
BAA	0.11	0.11	−0.11	0.17	0.09	−0.10	−0.19	0.07	−0.002	0.03
COMPAPFFx	−0.07	−0.07	0.02	−0.04	0.13	−0.11	0.26	0.05	0.05	−0.02
TB3SMFFM	−0.12	−0.12	0.04	−0.04	0.17	−0.10	0.08	0.04	0.03	0.01
TB6SMFFM	−0.12	−0.12	0.03	−0.04	0.19	−0.13	0.10	0.06	0.04	0.01
T1YFFM	−0.10	−0.10	−0.01	0.02	0.24	−0.19	0.05	0.11	0.06	0.04
T5YFFM	−0.08	−0.08	0.01	0.05	0.26	−0.16	−0.05	0.11	0.06	0.05
T10YFFM	−0.07	−0.07	0.04	0.04	0.24	−0.15	−0.08	0.08	0.06	0.04
AAAFFM	−0.05	−0.05	0.05	0.03	0.21	−0.11	−0.12	0.04	0.03	0.01
BAAFFM	−0.02	−0.02	0.03	0.07	0.22	−0.12	−0.14	0.03	0.02	−0.001
EXSZUSx	0.002	−0.003	0.003	0.05	−0.02	−0.05	−0.02	−0.13	−0.09	−0.27

Table 2 (continued)

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
EXJPUSx	0.01	0.01	0.03	0.01	−0.04	−0.01	−0.02	−0.11	−0.05	−0.22
EXUSUKx	0.01	−0.01	0.03	−0.07	0.04	0.10	−0.02	0.13	0.12	0.26
EXCAUSx	−0.001	0.003	−0.02	0.002	−0.03	−0.15	0.06	−0.09	−0.02	−0.17
WPSFD49207	0.11	0.05	0.20	−0.11	0.08	−0.004	−0.005	−0.07	0.002	−0.05
WPSFD49502	0.10	0.04	0.21	−0.12	0.09	0.004	−0.03	−0.06	−0.001	−0.05
WPSID61	0.12	0.03	0.20	−0.11	0.06	0.01	−0.03	−0.06	0.01	−0.04
WPSID62	0.06	−0.01	0.17	−0.09	0.09	0.03	−0.06	−0.04	0.005	−0.02
OILPRICEx	0.02	−0.001	0.04	−0.03	0.001	0.08	−0.04	0.02	−0.002	0.14
PPICMM	0.05	−0.04	0.09	−0.07	0.05	0.03	−0.07	−0.01	0.03	0.06
CPIAUCSL	0.16	0.09	0.16	−0.04	0.07	−0.02	0.01	−0.02	−0.01	−0.03
CPIAPPSL	0.08	0.03	0.06	0.05	−0.01	−0.06	0.11	−0.03	−0.02	0.09
CPITRNSL	0.10	0.04	0.23	−0.08	0.13	−0.01	−0.03	−0.01	0.02	−0.04
CPIMEDSL	0.10	0.10	−0.04	0.13	0.03	−0.05	0.03	0.05	−0.0003	0.02
CUSR0000SAC	0.13	0.06	0.22	−0.08	0.12	−0.01	0.01	−0.03	0.005	−0.04
CUSR0000SAD	0.12	0.07	0.04	0.11	0.04	−0.11	0.11	0.03	0.05	0.03
CUSR0000SAS	0.14	0.10	−0.003	0.05	−0.04	−0.02	0.04	0.001	0.005	0.01
CPIULFSL	0.15	0.09	0.16	−0.02	0.07	−0.02	0.01	−0.01	0.002	−0.02
CUSR0000SA0L2	0.15	0.07	0.19	−0.06	0.10	−0.02	0.01	−0.03	−0.004	−0.04
CUSR0000SA0L5	0.16	0.08	0.16	−0.05	0.07	−0.02	0.02	−0.03	−0.01	−0.04
PCEPI	0.17	0.09	0.13	−0.02	0.08	−0.04	0.04	−0.02	−0.02	−0.02
DDURRG3M086SBEA	0.12	0.07	0.01	0.08	0.03	−0.11	0.15	0.01	0.04	0.06
DNDGRG3M086SBEA	0.12	0.05	0.22	−0.11	0.11	0.01	0.0001	−0.04	−0.01	−0.06
DSERRG3M086SBEA	0.15	0.09	−0.01	0.07	0.04	−0.06	0.02	−0.003	−0.04	0.01
CES0600000008	0.10	0.03	−0.02	0.07	0.02	−0.02	0.19	−0.23	0.07	0.04
CES2000000008	0.03	0.05	−0.02	−0.01	−0.005	0.01	0.18	−0.19	0.07	0.13
CES3000000008	0.10	0.01	−0.004	0.11	0.03	0.01	0.15	−0.22	0.05	0.05
MZMSL	−0.03	0.01	−0.09	0.07	0.14	−0.09	−0.04	−0.22	−0.11	−0.11
DTCOLNVHFNM	0.04	−0.0003	−0.04	−0.02	0.02	−0.02	−0.16	−0.13	0.29	−0.06
DTCTHFNM	0.05	−0.01	−0.06	−0.06	−0.01	−0.004	−0.16	−0.11	0.41	−0.01
INVEST	−0.01	0.01	−0.02	0.07	0.15	0.002	0.02	−0.04	−0.06	−0.09

Table 3
Sparse principal component weights for 120 macro variables

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	SPC1	SPC2	SPC3	SPC4	SPC5	SPC6	SPC7	SPC8	SPC9	SPC10
Variable	Yields	Inflation	Production	Housing	Employment	Yield spreads	Wages	Optimism	Money	Credit
RPI	-	-	-	-	-	-	-	0.33	-	0.11
W875RX1	-	-	-	-	-	-	-	0.35	-	0.13
DPCERA3M086SBEA	-	-	-	-	-	-	-	0.34	-	-
CMRMTSPLx	-	-	-	-	-	-	-	-	-	0.10
RETAILx	-	-	-	-	-	-	-	0.31	-	-
INDPRO	-	-	0.32	-	-	-	-	-	-	-
IPFPNSS	-	-	0.32	-	-	-	-	-	-	-
IPFINAL	-	-	0.31	-	-	-	-	-	-	-
IPCONGD	-	-	0.27	-	-	-	-	-	-	-
IPDCONGD	-	-	0.27	-	-	-	-	-	-	-
IPNCONGD	-	-	-	-	-	-	-	0.21	-	-
IPBUSEQ	-	-	0.26	-	-	-	-	-	-	-
IPMAT	-	-	0.27	-	-	-	-	-	-	-
IPDMAT	-	-	0.29	-	-	-	-	-	-	-
IPNMAT	-	-	-	-	-	-	-	0.32	-	-
IPMANSICS	-	-	0.32	-	-	-	-	-	-	-
IPB51222S	-	-	-	-	-	-	-	-	-	-
IPFUELS	-	-	-	-	-	-	-	-	-	-
CUMFNS	-	-	0.32	-	-	-	-	-	-	-
HWI	-	-	-	-	-	-	-	0.24	-	-
HWIURATIO	-	-	-	-	-	-	-	0.29	-	-
CLF16OV	-	-	-	-	-	-	-	-	-	-
CE16OV	-	-	-	-	0.24	-	-	-	-	-
UNRATE	-	-	-	-	−0.24	-	-	-	-	-
UEMPMEAN	-	-	-	-	-	-	-	-	-	-
UEMPLT5	-	-	-	-	-	-	-	-	-	-
UEMP5TO14	-	-	-	-	-	-	-	−0.22	-	-
UEMP15OV	-	-	-	-	−0.21	-	-	-	-	-
UEMP15T26	-	-	-	-	-	-	-	-	-	-
UEMP27OV	-	-	-	-	-	-	-	-	-	−0.17

The table reports nonzero weights for ten sparse principal components extracted from 120 macro variables (listed in Table 1) from the FRED-MD database. The sample period is 1960:02 to 2018:06. The variable name in the first column corresponds to its FRED ticker. The column headings provide descriptions of the sparse principal components based on the active elements of their weight vectors.

Table 3 (continued)

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	SPC1	SPC2	SPC3	SPC4	SPC5	SPC6	SPC7	SPC8	SPC9	SPC10
Variable	Yields	Inflation	Production	Housing	Employment	Yield Spreads	Wages	Optimism	Money	Credit
CLAIMSx	-	-	-	-	-	-	-	−0.32	-	-
PAYEMS	-	-	-	-	0.37	-	-	-	-	-
USGOOD	-	-	-	-	0.34	-	-	-	-	-
CES1021000001	-	-	-	-	-	-	-	-	-	-
USCONS	-	-	-	-	0.25	-	-	-	-	-
MANEMP	-	-	0.25	-	-	-	-	-	-	-
DMANEMP	-	-	0.25	-	-	-	-	-	-	0.09
NDMANEMP	-	-	0.08	-	0.29	-	-	-	-	−0.07
SRVPRD	-	-	-	-	0.32	-	-	-	-	-
USTPU	-	-	-	-	0.35	-	-	-	-	-
USWTRADE	-	-	-	-	0.31	-	-	-	-	-
USTRADE	-	-	-	-	0.30	-	-	-	-	-
USFIRE	-	-	-	0.19	-	-	-	-	-	-
USGOVT	-	-	-	-	-	-	-	-	-	-
CES0600000007	-	-	-	-	-	-	−0.37	-	-	-
AWOTMAN	-	-	-	-	-	-	-	0.25	-	-
AWHMAN	-	-	-	-	-	-	−0.39	-	-	-
HOUST	-	-	-	0.34	-	-	-	-	-	-
HOUSTNE	-	-	-	0.27	-	-	-	-	-	-
HOUSTMW	-	-	-	0.29	-	-	-	-	-	-
HOUSTS	-	-	-	0.31	-	-	-	-	-	-
HOUSTW	-	-	-	0.32	-	-	-	-	-	-
PERMIT	-	-	-	0.33	-	-	-	-	-	-
PERMITNE	-	-	-	0.28	-	-	-	-	-	-
PERMITMW	-	-	-	0.30	-	-	-	-	-	-
PERMITS	-	-	-	0.24	-	-	-	-	-	-
PERMITW	-	-	-	0.32	-	-	-	-	-	-
AMDMNOx	-	-	-	-	-	-	-	0.22	-	-
AMDMUOx	-	-	-	-	0.18	-	-	-	-	-
BUSINVx	-	-	-	-	-	−0.20	-	-	−0.12	-

Table 3 (continued)

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	SPC1	SPC2	SPC3	SPC4	SPC5	SPC6	SPC7	SPC8	SPC9	SPC10
Variable	Yields	Inflation	Production	Housing	Employment	Yield Spreads	Wages	Optimism	Money	Credit
ISRATIOx	-	-	-	-	-	-	-	-	-	-
M1SL	-	-	-	-	-	-	-	-	0.40	-
M2SL	-	-	-	-	-	-	0.17	-	0.44	-
M2REAL	-	−0.23	-	-	-	-	-	-	0.30	-
AMBSL	-	-	-	-	-	-	-	-	0.33	-
TOTRESNS	-	-	-	-	-	-	-	-	0.36	-
NONBORRES	-	-	-	-	-	-	-	-	-	-
BUSLOANS	-	-	-	-	-	−0.20	-	-	-	0.12
REALLN	-	-	-	0.22	-	-	-	-	-	-
NONREVSL	-	-	-	-	-	-	-	-	-	0.54
CONSPI	-	-	-	-	-	-	-	-	-	0.48
DIVYLD	-	-	-	-	-	-	-	-	-	-
FEDFUNDS	0.31	-	-	-	-	-	-	-	-	-
CP3Mx	0.31	-	-	-	-	-	-	-	-	-
TB3MS	0.31	-	-	-	-	-	-	-	-	-
TB6MS	0.31	-	-	-	-	-	-	-	-	-
GS1	0.31	-	-	-	-	-	-	-	-	-
GS5	0.31	-	-	-	-	-	-	-	-	-
GS10	0.30	-	-	-	-	-	-	-	-	-
AAA	0.30	-	-	-	-	-	-	-	-	-
BAA	0.29	-	-	-	-	-	-	-	-	-
COMPAPFFx	-	-	-	-	-	0.25	-	-	-	-
TB3SMFFM	−0.23	-	-	-	-	0.16	-	-	-	-
TB6SMFFM	−0.22	-	-	-	-	0.18	-	-	-	-
T1YFFM	-	-	-	-	-	0.35	-	-	-	-
T5YFFM	-	-	-	-	-	0.39	-	-	-	-
T10YFFM	-	-	-	-	-	0.39	-	-	-	-
AAAFFM	-	-	-	-	-	0.38	-	-	-	-
BAAFFM	-	-	-	-	-	0.36	-	-	-	-
EXSZUSx	-	-	-	-	-	-	-	-	-	-

Table 3 (continued)

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	SPC1	SPC2	SPC3	SPC4	SPC5	SPC6	SPC7	SPC8	SPC9	SPC10
Variable	Yields	Inflation	Production	Housing	Employment	Yield spreads	Wages	Optimism	Money	Credit
EXJPUSx	-	-	-	-	-	-	-	-	-	-
EXUSUKx	-	-	-	-	-	-	-	-	−0.15	-
EXCAUSx	-	-	-	-	-	-	-	-	-	-
WPSFD49207	-	0.28	-	-	-	-	-	-	-	-
WPSFD49502	-	0.27	-	-	-	-	-	-	-	-
WPSID61	-	0.27	-	-	-	-	-	-	-	-
WPSID62	-	-	-	-	-	-	-	-	−0.18	-
OILPRICEx	-	-	-	-	-	-	-	-	-	-
PPICMM	-	-	-	-	-	-	-	-	−0.20	-
CPIAUCSL	-	0.31	-	-	-	-	-	-	-	-
CPIAPPSL	-	-	-	-	-	-	0.19	-	-	-
CPITRNSL	-	0.28	-	-	-	-	-	-	-	-
CPIMEDSL	-	-	-	-	-	-	0.33	-	-	-
CUSR0000SAC	-	0.31	-	-	-	-	-	-	-	-
CUSR0000SAD	-	-	-	-	-	-	0.35	-	−0.13	-
CUSR0000SAS	-	-	-	-	-	−0.25	0.21	-	-	-
CPIULFSL	-	0.29	-	-	-	-	-	-	-	-
CUSR0000SA0L2	-	0.31	-	-	-	-	-	-	-	-
CUSR0000SA0L5	-	0.31	-	-	-	-	-	-	-	-
PCEPI	-	0.29	-	-	-	-	0.13	-	-	-
DDURRG3M086SBEA	-	-	-	-	-	−0.16	0.30	-	-	-
DNDGRG3M086SBEA	-	0.31	-	-	-	-	-	-	-	-
DSERRG3M086SBEA	0.23	-	-	-	-	-	-	-	-	-
CES0600000008	-	-	-	-	-	-	0.35	-	-	-
CES2000000008	-	-	-	-	-	-	0.20	-	-	-
CES3000000008	-	-	-	-	-	-	0.32	-	-	-
MZMSL	-	-	-	-	-	-	-	-	0.41	-
DTCOLNVHFNM	-	-	-	-	-	-	-	-	-	0.36
DTCTHFNM	-	-	-	-	-	-	-	-	-	0.49
INVEST	-	-	-	-	-	-	-	-	0.16	−0.11

Table 4
Innovation correlations, 1960:03–2018:06

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Conventional PCA innovations									
PC	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9
PC2	−0.38								
PC3	0.78	−0.06							
PC4	0.13	−0.69	−0.08						
PC5	0.25	0.24	0.51	−0.21					
PC6	−0.02	−0.09	0.13	0.48	0.01				
PC7	0.04	0.10	−0.02	0.18	−0.19	0.06			
PC8	−0.09	−0.15	−0.05	−0.08	−0.12	0.04	−0.05		
PC9	−0.03	−0.03	−0.11	−0.04	−0.08	0.05	0.08	−0.08	
PC10	−0.19	0.20	−0.06	−0.11	−0.08	0.15	0.13	0.01	−0.02
Panel B: Sparse PCA innovations									
Description	Yields	Inflation	Production	Housing	Employment	Yield Spreads	Wages	Optimism	Money
Inflation	0.18								
Production	0.11	0.001							
Housing	0.01	0.04	0.19						
Employment	0.11	0.08	0.46	0.31					
Yield spreads	−0.58	−0.09	−0.14	0.13	−0.07				
Wages	0.19	0.19	−0.03	−0.05	−0.13	−0.07			
Optimism	0.10	−0.06	0.56	0.30	0.44	−0.05	−0.07		
Money	−0.11	−0.34	−0.09	0.05	−0.08	0.12	0.13	−0.08	
Credit	−0.004	−0.05	0.02	0.02	0.08	−0.05	−0.001	0.08	0.03

The table reports correlations for innovations to conventional and sparse principal components extracted from 120 macro variables (listed in Table 1) from the FRED-MD database. The innovations are computed by fitting a first-order vector autoregression to each set of principal components. The row and column headings in Panel B provide descriptions of the sparse principal components based on the active elements of their weight vectors.

Table 5
Three-pass regression results, 1963:07–2018:06

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Innovations to conventional PCA macro factors					Innovations to sparse PCA macro factors				
Factor	$\hat{\gamma}_g$	t-stat.	R_g^2	Sharpe	Factor	$\hat{\gamma}_g$	t-stat.	R_g^2	Sharpe
PC1	−0.001	−0.12	2.79%	−0.38	Yields	−0.008	−2.92**	6.95%	−0.85
PC2	−0.014	−1.55	2.72%	−0.51	Inflation	−0.002	−0.21	1.18%	0.08
PC3	−0.001	−0.09	1.55%	−0.03	Production	0.006	0.44	2.83%	0.14
PC4	−0.006	−0.61	2.48%	−0.19	Housing	0.011	3.68**	3.53%	1.02
PC5	0.008	1.01	0.83%	0.55	Employment	0.003	0.33	2.79%	0.79
PC6	0.009	0.77	4.30%	0.19	Yield spreads	0.009	1.80	1.28%	0.11
PC7	−0.009	−1.24	0.89%	−0.58	Wages	−0.010	−1.40	1.66%	−0.49
PC8	0.031	2.07*	3.59%	0.62	Optimism	0.032	2.29*	5.75%	0.53
PC9	0.014	1.13	1.26%	0.45	Money	−0.015	−1.00	3.79%	−0.33
PC10	−0.017	−1.08	5.04%	−0.29	Credit	0.008	0.71	1.26%	0.28
R_f^2	67.72%								

The table reports three-pass regression results for innovations to conventional and sparse principal components extracted from 120 macro variables (listed in Table 1) from the FRED-MD database. The names in the sixth column provide descriptions of the sparse principal components based on the active elements of their weight vectors. The risk premium estimates are computed via the three-pass methodology using 202 equity portfolios as test assets. The first step estimates ten principal components for the excess return observations for the test assets. The second step estimates risk premia for the ten principal components from the first step via a cross-sectional regression that relates the betas for the ten principal components to the average excess returns of the test assets; R_f^2 is the R^2 for the cross-sectional regression. The third step estimates time-series regressions relating each innovation to the ten principal components from the first step; R_g^2 is the goodness-of-fit measure for the time-series regression. $\hat{\gamma}_g$ is the estimated risk premium for the innovation, which is based on the estimated cross-sectional and time-series slope coefficients in the second and third steps, respectively; t-statistics are computed using Newey and West (1987) standard errors; * and ** indicate significance at the 5% and 1% levels, respectively. "Sharpe" is the annualized Sharpe ratio for the Giglio and Xiu (2018) factor mimicking portfolio for the innovation.

Table 6
Pricing errors

(1) Multifactor model	(2) Average absolute alpha	(3) GRS statistic
Panel A: 63 anomaly portfolios, 1974:01–2016:12		
Carhart (1997) four-factor model	0.0037	2.85^{**}
Fama and French (2015) five-factor model	0.0028	2.61^{**}
Hou et al. (2015) q-factor model	0.0023	2.36^{**}
Sparse macro four-factor model	0.0024	2.53^{**}
Panel B: 30 industry portfolios, 1967:01–2016:12		
Carhart (1997) four-factor model	0.0018	2.36^{**}
Fama and French (2015) five-factor model	0.0024	2.56^{**}
Hou et al. (2015) q-factor model	0.0020	2.11^{**}
Sparse macro four-factor model	0.0017	2.35^{**}

The table reports pricing errors for two sets of test assets and four multifactor models based on the Gibbons et al. (1989) framework. The test assets in Panel A are 63 anomaly portfolios from Huang et al. (2018), while those in Panel B are 30 industry portfolios from Kenneth French's Data Library. The sparse macro four-factor model includes the market factor and mimicking portfolio returns for yields, housing, and optimism factors. The second column reports the average of the absolute values of the alphas. The third column reports the Gibbons et al. (1989) W_u statistic for testing the null hypothesis that the alphas are jointly zero; * and ** indicate significance at the 5% and 1% levels, respectively.

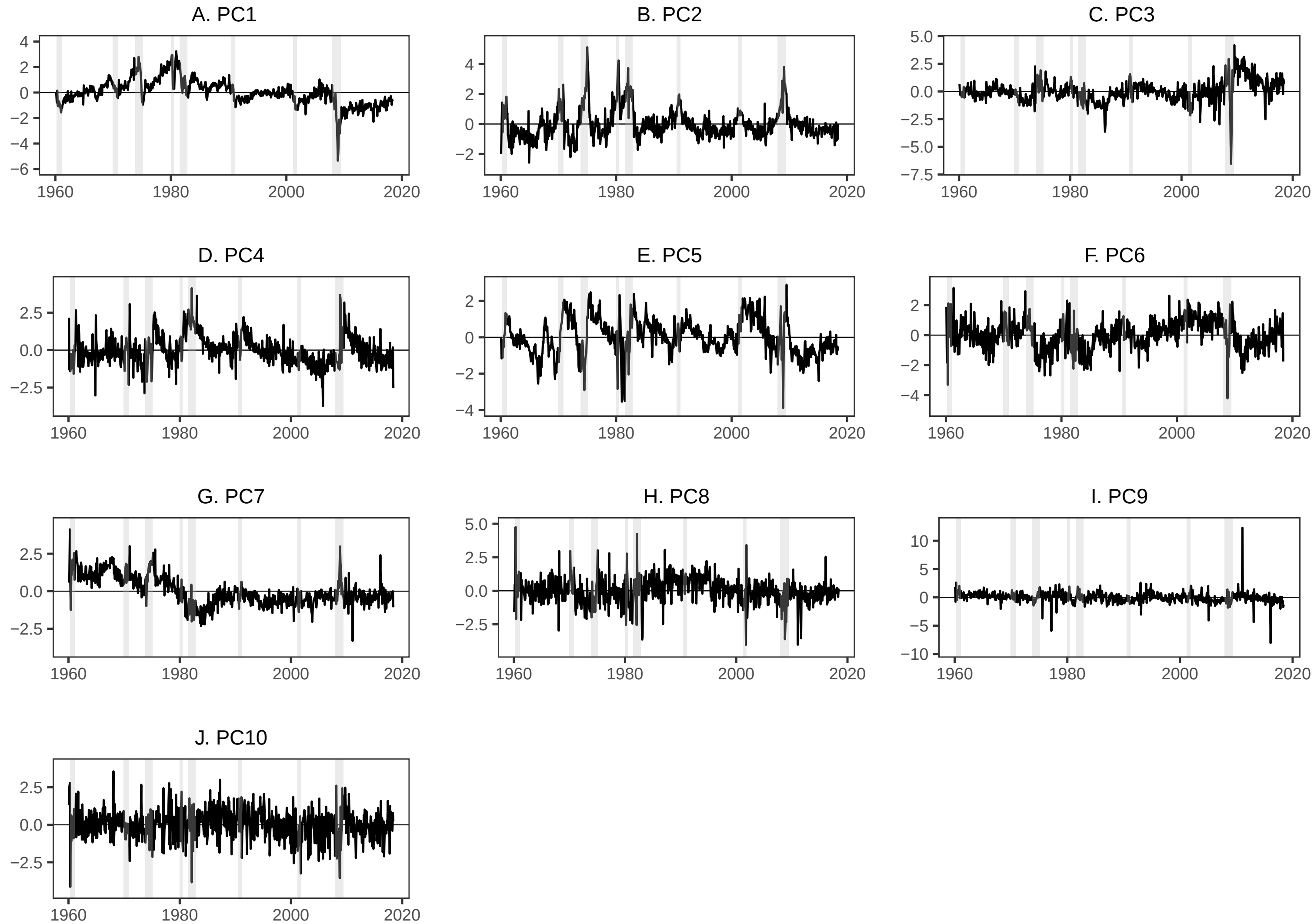

Figure 1. Conventional principal components, 1960:02–2018:06.

The figure depicts the first ten conventional principal components extracted from 120 macro variables (listed in Table 1) from the FRED-MD database. Each principal component is standardized to have zero mean and unit variance. Vertical bars delineate recessions as dated by the National Bureau of Economic Research.

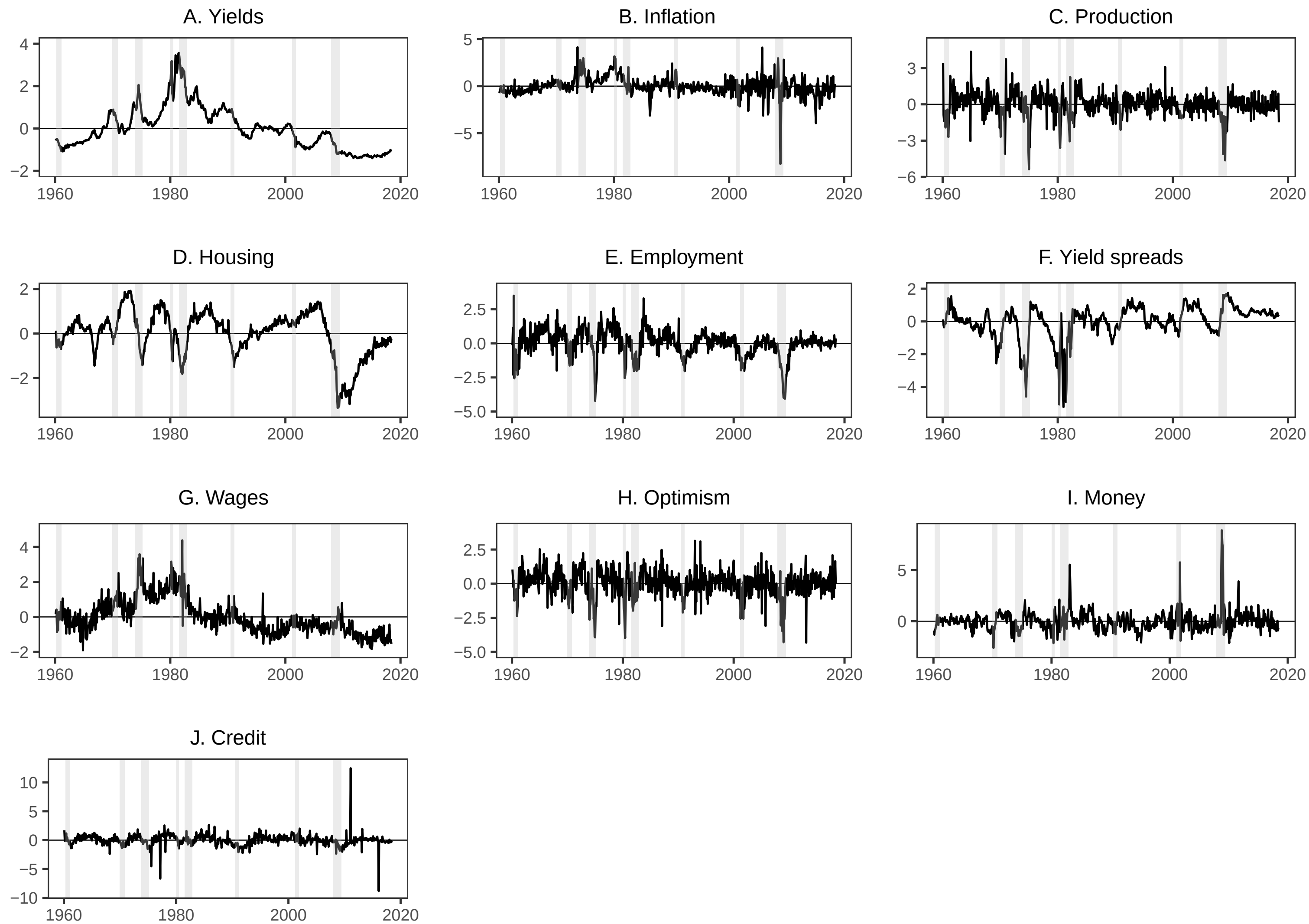

Figure 2. Sparse principal components, 1960:02–2018:06.

The figure depicts ten sparse principal components extracted from 120 macro variables (listed in Table 1) from the FRED-MD database. The panel headings provide descriptions of the sparse principal components based on the active elements of their weight vectors. Each sparse principal component is standardized to have zero mean and unit variance. Vertical bars delineate recessions as dated by the National Bureau of Economic Research.