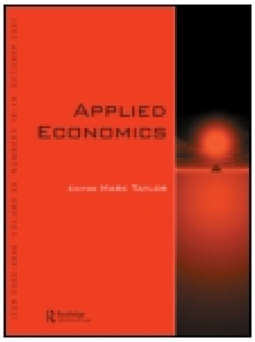

Applied Economics

ISSN: 0003-6846 (Print) 1466-4283 (Online) Journal homepage: https://www.tandfonline.com/loi/raec20

The nonlinear multidimensional relationship between stock returns and the macroeconomy

Pian Chen & Aaron Smith

To cite this article: Pian Chen & Aaron Smith (2013) The nonlinear multidimensional relationship between stock returns and the macroeconomy, Applied Economics, 45:35, 4985-4999, DOI: 10.1080/00036846.2013.806785

To link to this article: https://doi.org/10.1080/00036846.2013.806785

Published online: 23 Jul 2013.

Applied Economics, 2013
Vol. 45, No. 35, 4985–4999, http://dx.doi.org/10.1080/00036846.2013.806785

The nonlinear multidimensional relationship between stock returns and the macroeconomy

Pian Chen[a] and Aaron Smith[b,*]

[a]*Bates White, LLC, Washington, DC 20005, United States*
[b]*Department of Agricultural and Resource Economics, University of California, Davis CA 95616, United States*

We use nonparametric dimension-reduction methods to extract from a set of 15 macroeconomic variables the risk factors that are priced in the stock market. The dominant factor moves with the business cycle but, because it is a nonlinear function of observed macroeconomic variables, it captures a rich set of interactions. Low-credit risk and low-inflationary expectations have a greater positive effect on stock returns when leading macroeconomic indicators are high relative to current economic activity, i.e. early in the business cycle as the economy emerges from recession. High-stock returns also arise in periods when the economy is booming relative to its leading indicators, but such periods tend to portend crashes.

Keywords: asset pricing; business cycle; curse of dimensionality; data visualization

JEL Classification: C39; G12

I. Introduction

Asset pricing theory purports that risk premia depend on the sensitivity of each asset to aggregate risk. The finance literature has proposed and tested numerous measures of aggregate risk. In a seminal paper, Chen *et al*. (1986) showed that inflation shocks, industrial production, the spread between high- and low-grade bonds, and the spread between long- and short-term interest rates were priced in expected returns. More recent additions to the list of priced-risk factors include growth in various components of consumption (e.g. Parker and Julliard, 2005; Hansen *et al*., 2006; Yogo, 2006), investment growth (e.g. Cochrane, 1996), GDP growth (e.g. Vassalou, 2003), liquidity risk (e.g. Pastor and Stambaugh, 2003; Acharya and Pedersen, 2005) and the consumption-to-wealth ratio (Lettau and Ludvigson, 2004). For a more comprehensive list of priced macroeconomic risk factors, see Lewellen *et al*. (2010).

These papers typically start by proposing a theory that a particular variable may be a priced-risk factor and then testing empirically whether it is priced or not. The proposed factors are often tested in isolation or against a small subset of alternatives. In this article, we take the opposite approach by allowing the data to choose among a large set which macroeconomic variables are most strongly related to the cross section of stock returns. Moreover, we impose no functional form on the relationships.

Our approach has two steps. First, we apply inverse regression (Li, 1991) to reduce the macroeconomic variables to a few linear combinations known as inverse regression variates. To our knowledge, inverse regression has not been applied in finance, although it has been applied successfully in fields such as marketing

*Corresponding author. E-mail: adsmith@ucdavis.edu

(e.g. Naik *et al.*, 2010), bioinformatics (e.g. Li and Li, 2004) and chemometrics (e.g. Amato *et al.*, 2006). After obtaining the inverse regression variates, we apply the method of Donald (1997) to estimate a set of nonparametric common factors. These factors constitute the priced aggregate risk factors and are functions of the macroeconomic variables.

Our empirical approach also enables visualization of the high-dimensional relationships among our 25 stock portfolios and 15 macroeconomic variables. Data visualization can often identify interesting relationships among variables, but a high-dimensional system with an unknown joint distribution cannot be fully represented by arbitrary bi-variate plots (Carroll and Li, 1992; Cook, 1994). However, by reducing the system to a set of orthogonal nonparametric factors and a set of uncorrelated inverse regression variates, we can fully characterize a high-dimensional system using a minimum number of bi-variate plots. These plots reveal underlying structure previously hidden in high-dimensional data and guide parsimonious model specification.

The rest of the article is organized as follows. In Section II, we introduce inverse regression for variable space reduction and compare it with principal components analysis and canonical correlations, which are common dimension reduction methods. In Section III, we discuss function space reduction and high-dimensional data visualization. In Section IV, we extract the aggregate risk factors that explain the returns on 25 stock portfolios, where each risk factor is a nonparametric function of 15 macroeconomic variables. We reduce the 15 explanatory variables to three inverse regression variates, and we reduce the 25 regression functions to three nonparametric factors. Section V concludes the article.

II. Model Specification

We study the nonparametric model

$$Y_i = F_0(X_i) + U_i \quad (i = 1, 2, \cdots, n) \tag{1}$$

where Y_i represents a G-vector of excess stock returns, X_i denotes J macroeconomic variables, $F_0(X_i)$ is a G-vector of unknown functions of X_i and $E(U_i|X_i) = \underline{0}$. We reduce the dimension of the function space by decomposing $F_0(X_i)$ as

$$F_0(X_i) = AH(X_i) \tag{2}$$

where A is a $G \times L$ constant matrix and $H(X_i)$ is an L-vector of unknown functions with $L \leq G$. This decomposition implies that $F_0(X_i)$ falls into a function space spanned by L basis functions $(H_1, \ldots, H_L)$. Following Donald (1997), we call these unknown basis functions nonparametric factors. Equation 2 generalizes the linear parametric models of authors such as Chen *et al.* (1986) and Fama and French (1992, 1993) by allowing the aggregate risk factors to be nonparametric functions of a potentially large set of macroeconomic variables. Outside finance, other examples of the model in (2) include demand systems and nonparametric instrumental variables (Donald, 1997).

To estimate $(H_1, \ldots, H_L)$, we need to estimate $F_0(X_i)$ nonparametrically. In applications with many X variables, such estimation is hindered by the curse of dimensionality. To overcome this problem, we use inverse regression, which is based on the premise that a small number of linear combinations of X can capture the information in X that is relevant for Y (Li, 1991). Therefore, we can replace J-dimensional X with d linear combinations of X, i.e. $F_0(X_i) \equiv F_1(X_iB)$ and redefine the nonparametric factors as unknown functions of X_iB, i.e.

$$F_1(X_iB) = AH(X_iB) \tag{3}$$

where $B = (\beta_1, \ldots, \beta_d)$ is a $J \times d$ constant matrix with $d \leq J$. [1] Replacement of X_i with X_iB entails an orthogonal projection that transforms a large set of correlated X variables into a small set of uncorrelated inverse regression variates XB.

The matrix B cannot be uniquely identified from the data, but the space spanned by B is identifiable. Li (1991) proposes the inverse regression method, which generates root-n consistent estimates of this dimension reduction space without knowledge of the link function. Because the inverse regression estimates converge at a faster rate than standard nonparametric statistics, substituting the estimated inverse regression variates for the observed X variables does not affect the asymptotic distribution of nonparametric statistics in general. Consequently, in addition to the function space reduction problem, we treat in this article, inverse regression can be applied to many other testing and estimation problems in nonparametric econometrics.

Variable space reduction using inverse regression

Inverse regression is a dimension reduction method first proposed by Li (1991). To our knowledge, this method has not been applied in finance, so in this section we review the fundamental theory, outline the estimation procedure

[1] For notational convenience, we define X_i as the $1 \times J$ row vector that inhabits the rows of the $n \times J$ matrix X.

and compare the inverse regression to principal component analysis, linear regression and canonical correlations.

Inverse regression: theory and estimation

Write Y_i as a function d linear combinations of X_i, i.e.

$$Y_i = F(X_i\beta_1, X_i\beta_2, \cdots, X_i\beta_d, U_i) \tag{4}$$

where $d \leq J$, $B = (\beta_1, \beta_2, \cdots, \beta_d)$ are J-dimensional column vectors known as 'effective dimension reduction' (*e.d.r.*) directions, the link function F is unknown, X_iB are called inverse regression variates and U_i is independent of X_i. Without further structure on F, the vectors $\beta_1, \beta_2, \cdots, \beta_d$ are only identified up to an orthogonal rotation, i.e. the model would be identical if B were replaced by BC, for some $d \times d$ full rank matrix C. Inverse regression methods therefore focus on estimating the space spanned by B.

If $d = J$, then the representation in (4) places no restrictions on the functional form of the relationship between Y and X. However, in a typical case, d is much less than J and the model achieves parsimony by reducing F from a possibly nonlinear function of a large number of X variables to a possibly nonlinear function of a few inverse regression variates. If $d = 1$ and F is linear, then (4) is simply a linear regression model, in which one linear combination of X is sufficient to explain Y. On the other hand, if $d = 1$ and F is nonlinear, (4) reduces to the single-index model, which has been studied extensively in econometrics (Powell *et al.*, 1989; Ichimura, 1993; Klein and Spady, 1993). The model in (4) generalizes these special cases by allowing multiple indexes in the explanatory variables.

Li (1991) shows how to obtain root-n consistent estimates of the space spanned by B without the knowledge of the link function F. His key insight comes from the fact that Y is related only to d linear combinations of the X variables. It follows directly that estimating B requires the finding of the linear combinations of X that relate to Y. Li begins by first estimating nonlinear regressions of each X variable on Y. The resulting regression functions must be linearly dependent because Y is only related to a few linear combinations of X. Thus, rather than regressing Y on X (forward regression), Li estimates B by finding common factors in regressions of each X variable on Y (inverse regression).

To obtain this estimator, Li assumes that the joint distribution of X_i is such that $E(X_i\, b | X_i\beta_1, X_i\beta_2, \cdots, X_i\beta_d)$ is linear in $X_i\beta_1, X_i\beta_2, \cdots, X_i\beta_d$ for any vector $b \in R^J$. This assumption is known as the linear design condition, and it holds if the distribution of X_i is elliptically symmetric (e.g. the normal distribution). However, elliptical symmetry is not necessary for the condition to hold. Hall and Li (1993) prove under general conditions that deviations from the linear design condition converge to zero as the dimension of X_i increases. In other words, the linear design condition holds asymptotically for high-dimensional data. Under the linear design condition and the model defined in (4), Li (1991) shows that the centred inverse regression curve $E(X_i|Y_i) - E(X_i)$ is contained in the linear subspace spanned by $\Sigma_{XX}B$.

To understand this result, consider a vector $b \in R^J$ defined such that $b'\Sigma_{XX}\beta_k = 0$ for all $k = 1, 2, \ldots, d$, where Σ_{XX} denotes the covariance matrix of X_i. This vector b is orthogonal to the linear subspace spanned by $\Sigma_{XX}B$. Li shows that $E(b'X_i|Y_i) - E(b'X_i) = 0$ with probability one for all such vectors b, which implies that $b'Cov[E(X_i|Y_i)]b = 0$. It follows that the covariance matrix $Cov[E(X_i|Y_i)]$ is degenerate in any direction orthogonal to $(\Sigma_{XX}\beta_1, \Sigma_{XX}\beta_2, \cdots, \Sigma_{XX}\beta_d)$. Therefore, $(\Sigma_{XX}\beta_1, \Sigma_{XX}\beta_2, \cdots, \Sigma_{XX}\beta_d)$ are the eigenvectors associated with the d nonzero eigenvalues of $Cov[E(X_i|Y_i)]$. These expressions simplify slightly if the explanatory variables are standardized to $Z_i = \Sigma_{XX}^{-1/2}[X_i - E(X_i)]$. In that case, if $(\eta_1, \eta_2, \cdots, \eta_d)$ are the eigenvectors associated with the d nonzero eigenvalues of $Cov[E(Z_i|Y_i)]$, then the *e.d.r.* directions are $\beta_k = \Sigma_{XX}^{-1/2}\,\eta_k (k = 1, 2, \cdots, d)$.

The eigenvalue decomposition of $Cov[E(Z_i|Y_i)]$ requires an estimate of $E(Z_i|Y_i)$. For univariate Y_i, Li (1991) suggests dividing the range of Y_i into multiple 'slices'. Within each slice, the mean value of Z_i provides a crude estimate of the expected value of Z given a value of Y in that slice. In other words, $E(Z_i|Y_i)$ is estimated from Nadaraya–Watson kernel regression using the naïve kernel (see Pagan and Ullah, 1991, p. 8). The sliced inverse regression (SIR) estimates of B are thus obtained from an eigenvalue decomposition of

$$\hat{M}_{SIR} = \sum_{h=1}^{H} \hat{p}_h \hat{m}_h \hat{m}_h' \tag{5}$$

where $\hat{p}_h = n^{-1}\sum_{i=1}^{n} 1(S_{h-1} \leq Y_i \leq S_h)$, $\hat{m}_h = (\hat{p}_h n)^{-1} \sum_{i=1}^{n} 1(S_{h-1} \leq Y_i \leq S_h)Z_i$, $1(S_{h-1} \leq Y_i \leq S_h)$ is an indicator function and $S_0, S_1, \ldots, S_H$ denotes the cut-off points for the slices. The precision of SIR is determined by the precision of the estimate of $Cov[E(Z_i|Y_i)]$, and it is quite insensitive to the number of slices. Changing the number of slices makes one component of the covariance more precise while making another component less precise. For example, using a small number of slices allows for a large number of observations per slice and therefore generates relatively precise estimates of $E(Z_i|Y_i)$, but leaves few slices over which to estimate the covariance. At the other extreme, a large number of slices produces imprecise $E(Z_i|Y_i)$ estimates, but leaves many terms with which to precisely estimate the covariance.

Slicing provides one method for estimating $E(Z_i|Y_i)$. Other proposed inverse regression estimators estimate $E(Z_i|Y_i)$ by kernel regression (Fang and Zhu, 1996), nearest neighbours (Hsing, 1999) and parametric regression

(Bura and Cook, 2001). Methods to estimate B using the second moment of $X|Y$ have also been proposed, including the Sliced Average Variance Estimate (SAVE, Cook and Weisberg, 1991) and SIR-II (Li, 1991). In addition, SIR-α (Li, 1991) is an attempt to combine the information from both the first and second moments. These second-moment based methods may capture additional information that is orthogonal to the first moment of $X|Y$ (Cook and Weisberg, 1991).

SIR is easy to implement if the dependent variable Y_i is univariate. When Y_i is of high dimension, the slicing step of SIR becomes impractical because too many cells are empty. However, Hsing's (1999) nearest neighbour inverse regression (NNIR) proves quite robust to multi-dimensional Y_i.

To expound NNIR, let $i \in \{1, 2, \cdots, n\}$ and $i^* \in \{1, 2, \cdots, n\} - \{i\}$ be the indices for which

$$\gamma(Y_i, Y_{i*}) = \min_{j \neq i,\ 1 \leq j \leq n} \gamma(Y_i, Y_j) \quad (6)$$

where $\gamma(\cdot, \cdot)$ is a metric such as Euclidean distance, i.e. Y_{i*} is the nearest neighbour of Y_i. Let X_{i*} be the concomitant of Y_{i*}. Hsing (1999) estimates

$$Cov[E(Z_i|Y_i)] = 0.5E\left[Z_i' E(Z_i|Y_i)\right] + 0.5E\left[E(Z_i'|Y_i)Z_i\right] \quad (7)$$

using

$$\hat{M}_{NNIR} = (2n)^{-1} \sum_{i=1}^{n} (\hat{Z}_i' \hat{Z}_{i*} + \hat{Z}_{i*}' \hat{Z}_i) \quad (8)$$

where $\hat{Z}_i$ and $\hat{Z}_{i*}$ are standardized X_i and X_{i*}, respectively. Although $\hat{Z}_{i*}$ provides a noisy estimate of $E(Z_i|Y_i)$, the space spanned by the first d eigenvectors of $\hat{M}_{NNIR}$ is root-n consistent for the *e.d.r.* space under general conditions. We use NNIR in our application (see Section IV).

To our knowledge, the asymptotic distribution of the d smallest eigenvalues of $\hat{M}_{NNIR}$ has not been derived. Thus, hypothesis tests based on a statistic like (5) cannot be used to estimate the number of significant *e.d.r.* directions. However, cross-validation (CV) and permutation tests (Cook and Weisberg, 1991; Cook and Yin, 2001) can be useful for determining d in practice. In this article, we use CV to determine d (see Section Comparison to canonical correlations for details).

Comparison to principal components analysis

Rather than using inverse regression, it is common to apply principal component analysis (PCA) on X and keep the first few principal components for modelling the relationship between Y and X (e.g. Stock and Watson, 2002). The main drawback of this approach is that PCA does not take into account the relationship between Y and X when estimating the principal components. Thus, for any two different Y variables, PCA reduces the data to the same linear combinations of X, even if the relationships between Y and X are different across the two Y variables. Consequently, PCA can misspecify the fundamental relationship between Y and X by excluding relevant components and introducing irrelevant components for Y.

A simple example illustrates this point. Consider the linear regression model $Y_i = X_i\beta + U_i$. The first principal component of X is the linear combination that explains as much as possible of the variance of X. In general, there is no correspondence between this principal component and β. For example, suppose X contains several highly positively correlated variables measuring activity in the real economy and a single monetary variable that is weakly correlated with the real variables. PCA will be successful if the real variables have large effects on Y, with coefficients of the same sign. If a single monetary variable matters most for Y, then PCA will miss this relationship because the first principal component loads heavily on the block of real variables. Similarly, if the real variables matter for Y but enter the regression with opposing signs, then the first principal component will perform poorly.

To illustrate this last scenario in a nonlinear context, we generate 200 observations using $Y_i = \sin\ (X_i\beta_1) + U_i$, where $U_i \sim iid\ N(0,\ 0.01^2)$, $\beta_1 = [0.25\ -0.35\ -0.57\ 0.70]'$ and $X_i \sim iid\ N(\mu,\ \Sigma_{XX})$, with mean vector $\mu = [5\quad 5\quad 5\quad 5]$ and covariance matrix

$$\Sigma_{XX} = \begin{bmatrix} 1 & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho \\ \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & 1 \end{bmatrix}$$

In this example, the sine link function is nonlinear in the four X variables, and the four X variables can be reduced to one linear component denoted by $X_i\beta_1$. We experiment with three correlation coefficients $\rho = 0$, 0.5 and 0.95, and apply both SIR and PCA to estimate β_1. We use ten slices to obtain the SIR estimates. In Fig. 1, we plot Y against the first SIR variate in the left panel and plot Y against the first principal component in the right panel. It is evident that SIR correctly identifies the nonlinear sine pattern while PCA finds no pattern for all three correlation coefficients. When ρ increases, the X variables become highly collinear and SIR's ability to identify the *e.d.r.* direction decreases.

Comparison to canonical correlations

Under the assumption of a linear link function, inverse regression is identical to canonical correlation. Consider the multivariate linear regression model $Y_i = Z_i\eta\xi' + u_i$, where the coefficient matrix $\beta = \eta\xi'$ may be of reduced rank and $Z_i = \Sigma_{XX}^{-1/2}[X_i - E(X_i)]$ denotes the standardized

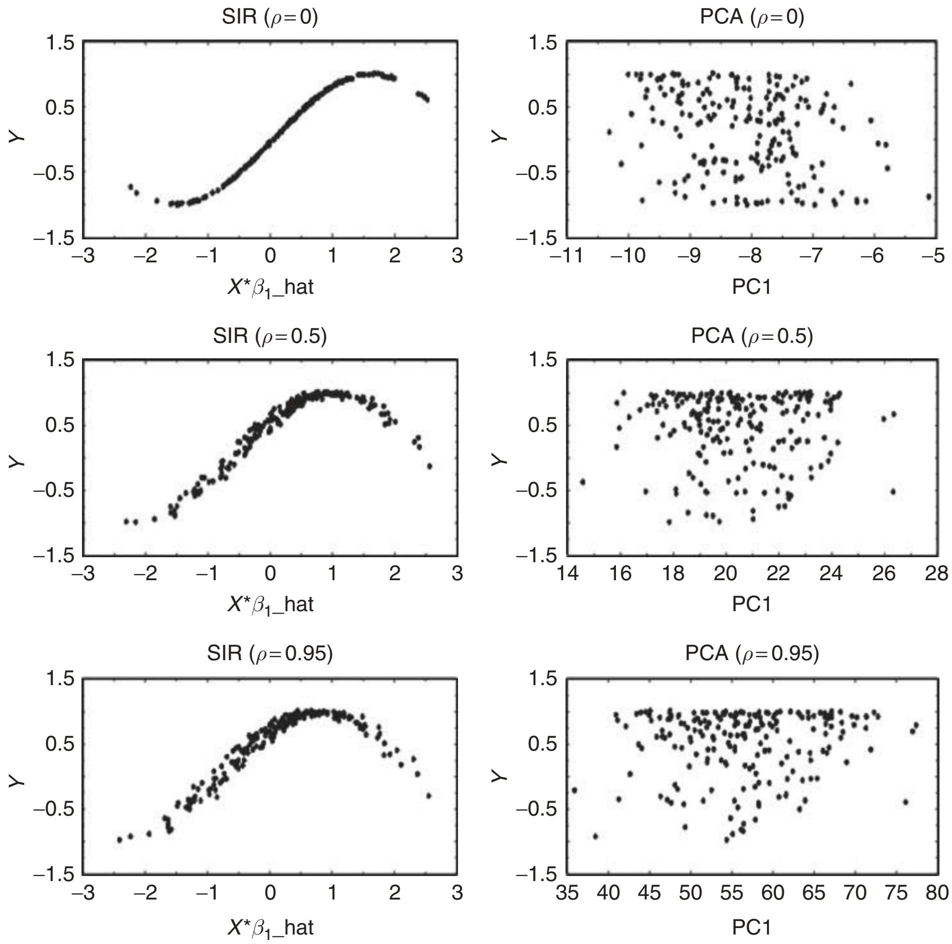

Fig. 1. Sliced inverse regression versus principal component analysis
Note: We generate 200 observations using $Y_i = \sin(X_i\beta_1) + U_i$, where $U_i \sim iid\ N(0,\ 0.01^2)$, $\beta_1 = [0.25 \quad -0.35 \quad -0.57 \quad 0.70]'$, $X_i \sim iid\ N(\mu,\ \Sigma_{XX})$ and $\rho \equiv corr(X_{ij}, X_{ik})$ for $j \neq k$. The graphs show Y_i plotted against $X\hat{\beta}$ for $\hat{\beta}$ estimated as the first SIR *e.d.r.* direction and the first principal component.

regressors. Canonical correlation analysis finds the most highly correlated linear combinations of Y_i and X_i.

Consider the canonical correlation between $Z_i\eta$, which is a linear combination of Z_i, and $Y_i\alpha$, which is a linear combination of Y_i. For identification, we use the normalizations $\eta'\Sigma_{ZZ}\eta = \eta'\eta = 1$ and $\alpha'\Sigma_{YY}\alpha = 1$, which imply that the canonical correlation coefficients η equal the eigenvectors of $\Sigma_{ZY}\Sigma_{YY}^{-1}\Sigma_{YZ}$ (Hamilton, 1994, pp. 630–5). Inverse regression calculates the eigenvectors of $Cov[E(Z_i|Y_i)]$. Linearity implies that $E(Z_i|Y_i) = \Sigma_{ZY}\Sigma_{YY}^{-1}Y_i$, so it follows that $Cov[E(Z_i|Y_i)] = \Sigma_{ZY}\Sigma_{YY}^{-1}\Sigma_{YZ}$. Because both methods perform an eigenvalue decomposition of $\Sigma_{ZY}\Sigma_{YY}^{-1}\Sigma_{YZ}$, canonical correlation and inverse regression are identical given a linear relationship between Y and Z.

Moreover, when estimated by maximum likelihood, a reduced rank regression model $Y_i = Z_i\eta\xi' + \varepsilon_i$ with a normally distributed ε_i has the first-order condition $(\Sigma_{ZY}\Sigma_{YY}^{-1}\Sigma_{YZ} - \lambda^2)\eta = 0$ (see Camba-Mendez *et al.*, 2003). This condition is the same as the one for canonical correlation, and therefore makes inverse regression identical to maximum likelihood for this case. Although all four methods use information from Y, inverse regression generalizes linear regression, canonical correlation and reduced rank regression because it imposes no structure on link function and error term. The inverse regression curve is estimated nonparametrically and the error term need not be additive.

III. Function Space Reduction Using Nonparametric Factors

In this section, we shift our attention from the variable space to the function space. First, we discuss how to estimate the regression function after reducing the dimension of the variable space. Second, we illustrate function space reduction. The idea of function space reduction is similar to that of variable space reduction. That is, the information in a large number of unknown regression functions may be well captured by a small number of orthogonal basis functions, known as nonparametric factors. Identifying nonparametric factors enables function space visualization, which in turn can reveal structure previously hidden in the data. It could also guide parsimonious model specification.

Nonparametric regression after inverse regression

Without reducing the dimension of the variable space, nonparametric estimation of $F_0(X_i)$ is subject to the *curse of dimensionality*. For example, consider the Nadaraya–Watson kernel estimator

$$\hat{F}_0(X_i) = \sum_{j=1}^{n} K\left(h^{-1}(X_j - X_i)\right) Y_j \Big/ \sum_{j=1}^{n} K\left(h^{-1}(X_j - X_i)\right) \tag{9}$$

where K represents a kernel weighting function and X_i is of dimension J. The optimal bandwidth $h_{opt} = O(n^{-1/(J+4)})$ results in the smallest mean squared error (MSE) of $O(n^{-4/(J+4)})$ under the regularity conditions (A2)-(A7) in Pagan and Ullah (1999, pp. 96–104). It follows that the MSE converges to zero slowly when J is large, creating the *curse of dimensionality*.

To mitigate the *curse of dimensionality*, we regress Y_i on the d inverse regression variates X_iB, rather than Y_i on the J-dimensional X_i. The reduced variates X_iB capture all the information in X_i that is relevant for Y_i, i.e. $F_0(X_i) \equiv F_1(X_iB)$, which implies that we can estimate $F_0(X_i)$ using

$$\hat{F}_1(X_iB) = \sum_{j=1}^{n} K_{ij}\, Y_j \Big/ \sum_{j=1}^{n} K_{ij} \tag{10}$$

where

$$K_{ij} = \prod_{k=1}^{d} K\left((X_i - X_j)\beta_k / h_k\right) \tag{11}$$

is a product kernel (Hall *et al.,* 2007). The MSE of this kernel regression is $O(n^{-4/(d+4)})$, which is smaller than $O(n^{-4/(J+4)})$ if $d < J$.

In practice, we do not observe the true *e.d.r.* directions, so we estimate B using inverse regression. Inverse regression methods produce root-n consistent estimates, whose convergence rate is faster than those of nonparametric methods. Therefore, replacing B with $\hat{B}$ does not affect the asymptotic properties of nonparametric estimates of $F_0(X_i)$ in general.

Identifying nonparametric factors

The matrix A and the nonparametric factors in (3) are not separately identifiable because $F_1(X_iB) = AH(X_iB) = A\,\Pi^{-1} \cdot \Pi H(X_iB) = A^*H^*(X_iB)$ holds for any $L \times L$ full rank matrix Π. Although $H(X_iB)$ and $H^*(X_iB) = \Pi H(X_iB)$ are different functions, they span the same function space because $H^*(X_iB)$ is a linear transformation of $H(X_iB)$. Consequently, we could span the function space using either $H(X_iB)$ or $H^*(X_iB)$. This fact implies that we can consistently estimate the function space spanned by the nonparametric factors using

$$\hat{H}(X_i\hat{B}) = \Xi\, \hat{F}_1(X_i\hat{B}) \tag{12}$$

where Ξ denotes an $L \times G$ matrix with full rank L.

To identify the nonparametric factors, we impose the orthonormal restriction

$$\Xi \left(n^{-1} \sum_{i=1}^{n} \hat{F}_1(X_i\hat{B})\hat{F}_1(X_i\hat{B})' \right) \Xi' = I_L \tag{13}$$

This restriction generates orthogonal nonparametric factors, and it implies that we can estimate Ξ using an eigenvalue decomposition. Specifically, we set $\xi = \Lambda^{-1/2}C'$, where Λ is a diagonal matrix containing the L largest eigenvalues of $n^{-1}\sum_{i=1}^{n} \hat{F}_1(X_i\hat{B})\hat{F}_1(X_i\hat{B})'$ and C is a $G \times L$ matrix containing the corresponding eigenvectors.

The factor loading matrix, $\hat{A} = \Xi'(\Xi\Xi')^{-1}$, measures the sensitivity of the dependent variables to the nonparametric factors. Because (13) implies $n^{-1}\sum_{i=1}^{n} \hat{H}(X_i\hat{B})\hat{H}(X_i\hat{B})' = I_L$, the resulting L nonparametric factors are orthogonal and have a unit norm. Consequently, the magnitude of $\hat{A}$ indicates the relative contribution of each nonparametric factor to the variation of $F_1(X_iB)$.

To visualize the function space, we use the $L \times d$ scatter plots of the estimated nonparametric factors, $[\hat{H}_1(X_i\hat{B}), \cdots, \hat{H}_L(X_i\hat{B})]$, against the estimated inverse regression variates, $(X_i\hat{\beta}_1, \cdots, X_i\hat{\beta}_d)$. Such bi-variate plots can effectively reveal features of the data that would be imperceptible from the bi-variate plots of each Y against each X or from the plots of estimated nonparametric regression functions $\hat{F}_1(X_i\hat{B})$ against $X_i\hat{B}$. We illustrate function space visualization in Section IV.

Order selection through cross-validation (CV)

Function space visualization requires the user to identify the number of nonparametric factors as well as the number of inverse regression variates. We use CV to jointly determine the number of significant nonparametric factors L, the number of significant NNIR variates d and the optimal bandwidths $h = (h_1, \cdots, h_d)$ for kernel regression.

Specifically, we select (d, h, L) by minimizing the estimated prediction error (EPE). We calculate the EPE using

$$EPE(d,h,L) = \frac{\sum_{i=1}^{n} \omega(X_i\hat{B})\left(Y_i - \tilde{A}\tilde{H}_{-i}(X_i\hat{B})\right)'\left(Y_i - \tilde{A}\tilde{H}_{-i}(X_i\hat{B})\right)}{\sum_{i=1}^{n} \omega(X_i\hat{B})} \tag{14}$$

where $\omega(X_i\hat{B}) \in \{0,1\}$ denotes a trimming function that removes outliers in $X_i\hat{B}$, thereby reducing sensitivity of EPE to outliers. We specify this function to equal zero for all observations for which the maximum value of $X_i\hat{B}$ exceeds its mean by more than a pre-set threshold. This approach enables us to trim the same observations for all candidate values of d, h and L. In our application, we set the threshold to remove 1% of the observations in the first four inverse regression variates.

The terms $\tilde{A}$ and $\tilde{H}_{-i}$ in (14) are leave-one-out estimates of factor loadings and nonparametric factors with the i^{th} observation being deleted while performing the estimation. Specifically, we first estimate $F_1(X_iB)$ using the leave-one-out estimator

$$\tilde{F}_{-i}(X_iB) = \sum_{j\neq i}^{n} K_{ij}\, Y_j \Big/ \sum_{j\neq i}^{n} K_{ij} \tag{15}$$

Next, we use an eigenvalue decomposition to estimate $\tilde{\Xi}$ from

$$\tilde{\Xi}\left(n^{-1}\sum_{i=1}^{n}\tilde{F}_{-i}(X_i\hat{B})\tilde{F}_{-i}(X_i\hat{B})'\right)\tilde{\Xi}' = I_L$$

which is the orthonormal restriction in (13). Finally, we estimate the factor loadings matrix as $\tilde{A} = \tilde{\Xi}'(\tilde{\Xi}\tilde{\Xi}')^{-1}$ and the nonparametric factors as $\tilde{H}_{-i}(X_i\hat{B}) = \tilde{\Xi}\,\tilde{F}_{-i}(X_i\hat{B})$.

As advocated by Hall *et al.* (2007), we use a product kernel in the nonparametric regression. This kernel allows a different bandwidth for each inverse regression variate, and therefore can remove an insignificant variate from the model by selecting a very large bandwidth for it. When $(X_i - X_j)\beta_k$ is divided by a large bandwidth, it is close to zero, and consequently the individual kernel $K\left((X_i - X_j)\beta_k / h_k\right)$ is close to a constant that cancels out the corresponding term in the denominator.

The EPE statistic in Equation 14 may overestimate d because it does not account explicitly for parameter estimation error in estimating B. The leave-one-out nonparametric regression in (15) uses the same $\hat{B}$ for each observation i and that estimate is obtained using all the data. An alternative would be to leave observation i out of the inverse regression estimate that is used to predict observation i. This alternative is unlikely to make a difference because the order of magnitude of the error in the nonparametric fit exceeds the order of magnitude of the error in estimating B. In our application, we find no material difference in the EPE or the selected model from using this more computationally expensive alternative.

IV. Asset Pricing Using Inverse Regression and Nonparametric Factors

Data

For asset returns, we use 25 portfolios of monthly stock returns constructed by Fama and French.[2] These portfolios comprise the intersections of five portfolios formed on firm size and five portfolios formed on the ratio of book equity to market equity. We model excess returns, which we calculate by subtracting the monthly average of the 1-month Treasury bill rate from each portfolio return. For the explanatory variables, we use a selection of 15 monthly macroeconomic variables measured over the period 1960:2–2008:12. Table 1 lists these variables and the data source for each one.

We use eight real and seven nominal macroeconomic variables. The first seven variables measure the monthly flow of real economic activity, and the eighth variable, unemployment, provides a snapshot of the employment situation during the month. We also include an inflation variable and several interest rate variables. We create the term spread and Treasury bond yield series as the difference between the yield on the last day of the month and the last day of the previous month. Our credit spread variable equals the difference between Moody's Seasoned Baa Corporate bond yield and the 10-year Treasury bond yield. Because daily corporate bond data go back only to 1983, we use monthly data to calculate credit spreads. The reported monthly data measure average daily yields during the month, whereas we would like to measure the yield at the end of the month, which is the same time we measure stock returns. To account for the fact that this variable measures interest rates with a lag, we include both the contemporaneous and first lead of credit spreads. The final variable is a survey of consumer expectations, for which we also include a 1-month lead to account for the potential staleness of each monthly observation by the end of the month.

[2] http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

Table 1. Macroeconomic variables

	Variable description	Source	Transform	Seas Adj
1	Consumption-wealth ratio (cay)	Lettau	2	yes
2	Real personal cons exp per capita: durables	FRED(pcedc96/pop)	1	yes
3	Real personal cons exp per capita: nondurables	FRED(pcendc96/pop)	1	yes
4	Real personal cons exp per capita: services	FRED(pcesc96/pop)	1	yes
5	Real personal income per capita	FRED(dspic96/pop)	1	yes
6	Industrial production	FRED(indpro)	1	yes
7	Housing starts	FRED(houst)	1	yes
8	Unemployment rate	FRED(unrate)	2	yes
9	Inflation shock (deviation from 12-month ave.)	FRED(cpiaucsl)	3	yes
10	Term spread	FRED(dgs10,dtb3m)	2	no
11	10-year T-bond yield	FRED(dgs10)	2	no
12	Credit spread	FRED(baa,gs10)	2	no
13	Credit spread (1 month lead)	FRED(baa,gs10)	2	no
14	UM index of consumer expectation	BCD-83/UM	2	no
15	UM index of cons expectation (1 month lead)	BCD-83/UM	2	no

Notes: (1) We obtained the cay variable from http://faculty.haas.berkeley.edu/lettau/data_cay.html (Lettau's website) and converted it from quarterly to monthly frequency using linear interpolation.
(2) The consumer expectations variable is as used in Stock and Watson (2002) under the moniker 'hhsntn'. We updated the consumer expectations series using the final column of Table 3 published at http://www.sca.isr.umich.edu (University of Michigan Survey Research Center website).
(3) The variables are transformed according to the following code: 1 = log difference (growth), 2 = first difference, and $3 = \pi_t - \sum_{i=1}^{12} \pi_{t-i}/12$.

Variable and function space reduction

The econometric model is $Y_i = F_1(X_iB) + U_i$, where Y_i denotes the vector of excess stock returns, $F_1(X_iB) = AH(X_iB)$ and $E(U_i|X_i) = \underline{0}$. In this model, excess stock returns are linear in the nonparametric factors H, which in turn, are unknown functions of the observed X variables. This model differs from existing asset pricing models because it includes a large number of X variables, and it allows for the factors to be nonlinear functions of X. For convenience, we divide each X variable by its SD before we reduce dimensions for both variable space and function space.

To reduce the variable space, we estimate the *e.d.r.* directions, B, using the NNIR method introduced in Section II. To indicate the relative importance of NNIR variates, we report in Table 2 the eigenvalues of $\hat{M}_{NNIR}$ from the inverse regression algorithm. The first two eigenvalues appear notably greater than the next three, which appear notably greater than the remaining eigenvalues. These results suggest that between two and five NNIR variates may be needed to explain the variation in the 25 stock portfolios. We use the CV procedure introduced in the Section 'Comparison to canonical correlations' to select jointly the number of significant NNIR variates, d, along with the optimal kernel regression bandwidths, h and the number of significant nonparametric factors, L.

The CV procedure suggests that the smallest EPE, 581.3, is achieved when L = 3, d = 3 and the optimal

Table 2. Eigenvalues from variable and function space reduction

Variable space (NNIR)		Function space (F′F)	
1	0.34	1	208.20
2	0.27	2	4.91
3	0.19	3	2.95
4	0.16	4	0.72
5	0.13	5	0.51
6	0.07	6	0.31
7	0.04	7	0.27
8	0.02	8	0.20
9	−0.01	9	0.18
10	−0.03	10	0.16
11	−0.06	11	0.14
12	−0.08	12	0.13
13	−0.11	13	0.10
14	−0.15	14	0.10
15	−0.16	15	0.09
		16	0.07
		17	0.06
		18	0.05
		19	0.04
		20	0.03
		21	0.03
		22	0.03
		23	0.02
		24	0.02
		25	0.02

Note: We computed the eigenvalues for function space using the first three NNIR variates, which were selected by cross-validation (see Table 3).

Table 3. Cross validation for optimal (*d*, *h*, *L*)

Models	L	EPE	Fit (%)	h_1	h_2	h_3	h_4
Mean-only	0	775.6					
Single-index model	1	663.5	14.4	0.88	–	–	–
	2	662.2	14.6	0.88	–	–	–
	3	662.5	14.6	0.89	–	–	–
	4	662.5	14.6	0.89	–	–	–
	25	663.1	14.5	0.89	–	–	–
Double-index model	1	617.8	20.4	1.04	0.56	–	–
	2	615.6	20.6	1.04	0.56	–	–
	3	615.8	20.6	1.05	0.56	–	–
	4	616.0	20.6	1.05	0.56	–	–
	25	618.2	20.3	1.04	0.62	–	–
Multi-index model	1	583.8	24.7	1.03	0.62	2.31	1000
	2	581.3	25.1	1.03	0.62	2.31	1000
	3	581.3	25.1	1.03	0.62	2.31	1000
	4	581.6	25.0	1.03	0.62	2.31	1000
	25	583.6	24.8	1.04	0.62	2.31	1000

Notes: 1% trimming, Epanechnikov kernel, Fit is per cent EPE reduction over mean-only model. Because EPE is calculated using leave-one-out kernel regression, the fit is a pseudo out-of-sample statistic.

bandwidths for the first three NNIR variates are ($h_1 = 1.03$, $h_2 = 0.62$, $h_3 = 2.31$). In Table 3, we report detailed CV results when trimming 1% of the data. We obtain similar results when trimming up to 5% of the data. The table reports the fit obtained by each specification, which is a pseudo out-of-sample statistic because EPE is calculated using leave-one-out kernel regression. The model explains 25.1% of the variation in returns when $d = 3$, compared to 14.6% for a single-index model ($d = 1$). Setting $d = 2$ achieves a fit of 20.6%. The relatively large increment in fit when increasing d from one to two, along with the relatively large bandwidth on the third NNIR variate, shows that the first two NNIR variates contribute more to the model than the third. These results show a substantial improvement in fit from moving beyond a single-index model.

The CV procedure minimizes EPE at three nonparametric factors. The eigenvalues in Table 2 show that the first of these nonparametric factors dominates. A single nonparametric factor explains vastly more of the variation in returns than the remaining 24 factors; the largest eigenvalue equals 208.2 and the next largest is 4.91. The explanatory power of the first factor can also be seen in the EPE values in Table 3. In the multi-index model, a single factor causes the EPE to drop from 775.6 to 583.8, but adding a second factor reduces EPE only by a further 2.5 points to 581.3. Nonetheless, EPE does decline as we increase the number of nonparametric factors to three, which suggests that the second and third factors are statistically significant.

Table 4. Cross validation for canonical correlations

Number of CC	EPE	Fit (%)
1	625.4	19.4
2	625.0	19.4
3	625.2	19.4
4	626.8	19.2
15	647.4	16.5

We compare the fit of our dimension reduction model with that of canonical correlations, which allows dimension reduction but imposes linearity (Camba-Mendez *et al.*, 2003). To do so, we apply CV to determine the number of significant canonical correlations between Y and X. To make the results comparable to those in Table 3, we trim the same observations when calculating EPE. The EPE values in Table 4 indicate that two canonical correlations are statistically significant. However, even the smallest EPE in Table 4 (625.0) is larger than the EPE of any multi-index model in Table 3, although it is smaller than the EPE for the single-index models.

To further assess whether the nonlinearity is statistically significant, we apply Anderson and Vahid's (1998) test for common nonlinear components in multiple-time series. The test begins with a candidate nonlinear function of X, for which we use $W_i = \left[(X_i\hat{\beta}_1)^2, (X_i\hat{\beta}_2)^2, (X_i\hat{\beta}_3)^2\right]$, where $X_i\hat{\beta}_1$, $X_i\hat{\beta}_2$ and $X_i\hat{\beta}_3$ denote the first three NNIR variates. The next step is to compute the residuals from regressions of Y on X and W on X to remove the linear components of the relationship between Y and X. In the final step, we calculate the canonical correlations between these two sets of residuals. We obtain p-values of 0.08, 0.01 and 0.00 for tests of the null hypotheses that there are at most 2, 1 and 0 common nonlinear components. Using a 5% significance level, these statistics suggest the presence of two common nonlinear components and reinforce the CV result that the link function is nonlinear.

The fact that the canonical correlations model fits significantly worse than the multi-index model implies that significant nonlinearity exists in the relationship between stock returns and the macroeconomic variables. In the next sub-section, we compare the loadings of NNIR variates and canonical correlations to show how they differ.

Interpretation of NNIR variates and comparison with canonical correlations

In Fig. 2, we plot the estimated betas of the first three NNIR variates as well as corresponding canonical correlations. Because the standardized X variables have SD equal to 1 and each beta vector is constrained to have unit length, the betas are scale independent and therefore indicate the relative importance of each variable. Moreover, because the NNIR variates, XB, are arguments in a nonparametric

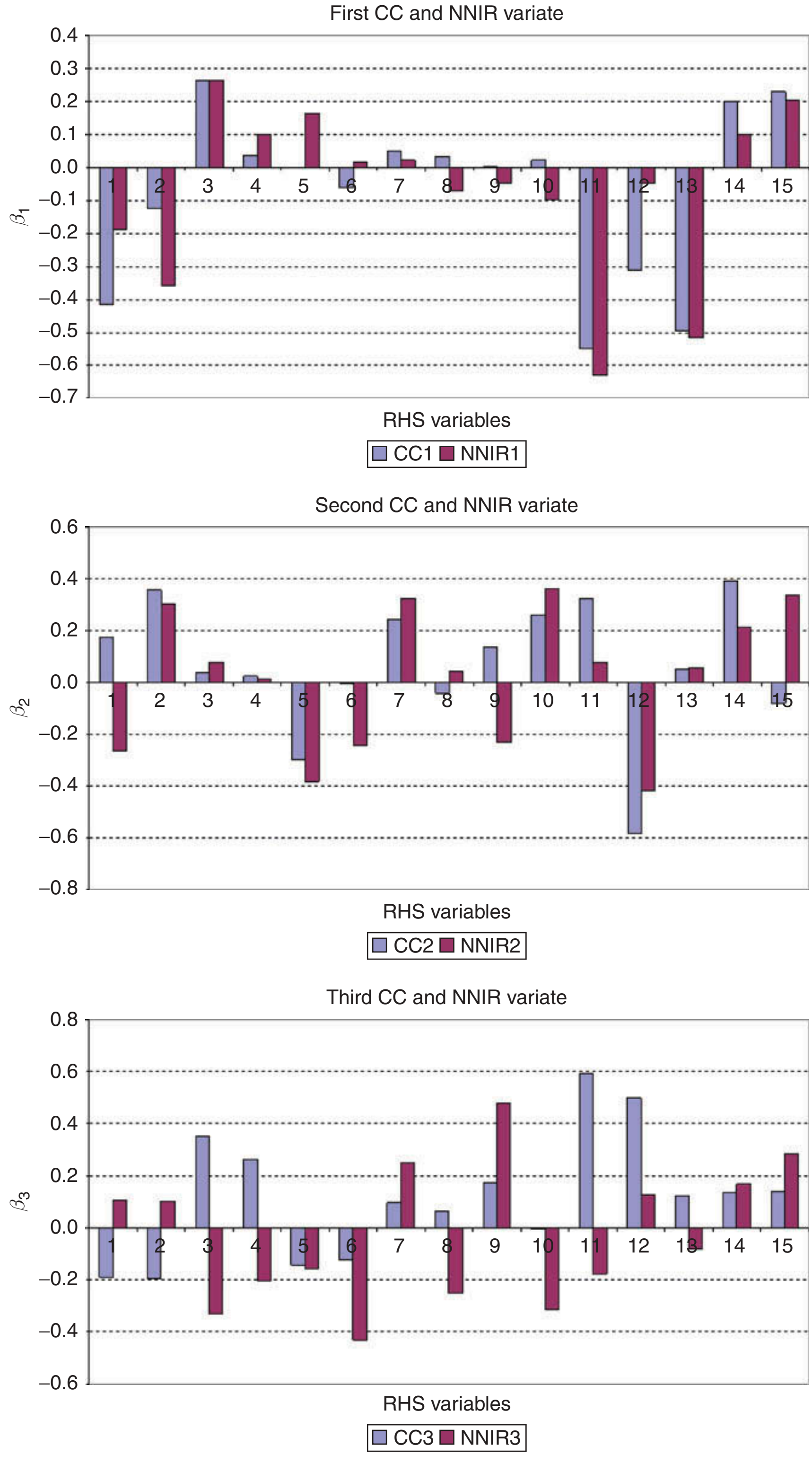

Fig. 2. Loadings on the first three NNIR variates and canonical correlations
Note: The variables are: 1 = cay; 2 = PCE-durables; 3 = PCE-nondurables; 4 = PCE services; 5 = personal income; 6 = industrial production; 7 = housing starts; 8 = unemployment rate; 9 = inflation shock; 10 = term spread; 11 = 10-yr T-bond yield; 12 = credit spread; 13 = lead of credit spread; 14 = UM index of cons. exp.; 15 = lead of UM index of cons. exp. See Table 1 for details.

function, the sign of XB is not identified. The model is the same whether we use XB or $-XB$. Thus, to aid interpretation, we impose a positive sign on the consumer expectation variable (#14), as we expect this variable to be positively correlated with returns.

The first NNIR variate is dominated by the 10-year Treasury bond yield (#11) and the credit spread (#13), which also featured prominently in the seminal macroeconomic asset pricing paper by Chen *et al*. (1986). The lead of the credit spread variable (#13) has a larger coefficient than the contemporaneous value (#12), suggesting that average credit spreads over a month contain much information that is stale for stock returns during the month. The Treasury bond yield and credit spread variables both measure long-run credit market conditions and reflect inflationary expectations and default risk, respectively.

The consumption variables have the next largest set of coefficients in the first NNIR variate. Consumption of nondurables (#3) has a positive sign and consumption of durables (#2) a negative sign, which implies that substitution from nondurable to durable consumption is negatively associated with this factor. One possible cause of such substitution is an increase in expected inflation. Barsky *et al*. (2007) use a New Keynesian model to show theoretically that a monetary expansion causes real consumption to increase in the sector with more flexible prices relative to the sticky price sector. If durable goods prices are stickier, then the expectation of high prices in the future causes consumers to make more nondurable goods purchases today before prices rise. It follows that a drop in durable goods consumption relative to nondurables may indicate a monetary expansion and higher expected future inflation. Overall, we interpret large values of the first NNIR variate as indicating low-credit risk and low-inflationary expectations.

The first canonical correlation also loads on the 10-year Treasury bond yield (#11) and credit spread (#13). For the consumption variables, it picks up nondurables (#3) but not durables (#2). It assigns high weights to both contemporaneous and lead of the credit spread variable (#12 and #13) as well as the consumption-wealth ratio (cay, #1). Overall, the first canonical correlation is similar to the first NNIR variate.

The second NNIR variate loads positively on four variables that foreshadow future increases in aggregate output. These variables are consumption of durables (#2), housing starts (#7), term spread (#10) and consumer expectations (#15). Moreover, it loads negatively on the credit spread (#12), which is negatively associated with future output increases. In contrast, this variate loads negatively on three measures of current economic activity: personal income (#5), industrial production (#6) and the inflation shock (#9). It also has a negative coefficient on the consumption wealth ratio (#1, *cay*, Lettau and Ludvigson, 2004), which measures current consumption of nondurables and services relative to income and wealth. Overall, the second variate loads positively on leading indicators of future macroeconomic growth and negatively on indicators of current macroeconomic activity.

The second canonical correlation picks up consumption of durables (#2), housing starts (#7), term spread (#10), credit spread (#12), personal income (#5) with the same sign and magnitude as those of the second NNIR variate. However, it differs from the second NNIR variate by being weakly related to consumer expectations (#15), industrial production (#6), the inflation shock (#9) and the consumption-wealth ratio (#1, *cay*).

The large CV bandwidth in Table 3 implies that the third NNIR variate relates weakly to returns. This variate is most strongly related to the inflation shock variable (#9) and industrial production (#6). The third canonical correlation is not statistically significant, and it favours the 10-year Treasury bond yield (#11) and the contemporaneous value of the credit spread variable (#12).

Function space visualization

Data visualization is often the first step of preliminary data analysis. If we had only a few variables, it would be convenient to use bivariate scatter plots to examine the relationships among them. However, plotting 25 returns against 15 macroeconomic variables would require 375 bivariate plots, which are too many to synthesise even if some plots share common features. Reducing the 15 macroeconomic variables to three NNIR variates reduces the number of bivariate plots to 75; reducing the 25 return functions to three nonparametric factors, further narrowing down the task of visualization to nine bivariate plots. We present these nine scatter plots in Fig. 3. Specifically, we plot each estimated nonparametric factor $\hat{H}_l(X_i\hat{\beta}_1, X_i\hat{\beta}_2, X_i\hat{\beta}_3)$ against each NNIR variate $X_i\hat{\beta}_k$ for $l = 1, 2, 3$ and $k = 1, 2, 3$.

The first factor is positively correlated with the first two NNIR variates ($X_i\hat{\beta}_1$ and $X_i\hat{\beta}_2$) and weakly correlated with the third variate $X_i\hat{\beta}_3$. The second factor is also positively correlated with $X_i\hat{\beta}_1$, but is negatively correlated with $X_i\hat{\beta}_2$. The third factor is negatively correlated with $X_i\hat{\beta}_1$, especially when $X_i\hat{\beta}_1$ is negative; it is weakly correlated with the other two variates. The weak relationship of all three factors to $X_i\hat{\beta}_3$ follows from the large bandwidth chosen for it by the CV procedure (see Table 3). As implied by the eigenvalues in Table 2, the first factor explains almost all the variation in equation-by-equation regression functions $\hat{F}_1(X_i\hat{B})$. Specifically, the correlation between the first nonparametric factor, $\hat{H}_1(X_i\hat{B})$, and the 25 columns of $\hat{F}_1(X_i\hat{B})$ ranges from 0.92 to 0.99, with a median of 0.98. Thus, because it explains almost all the variation

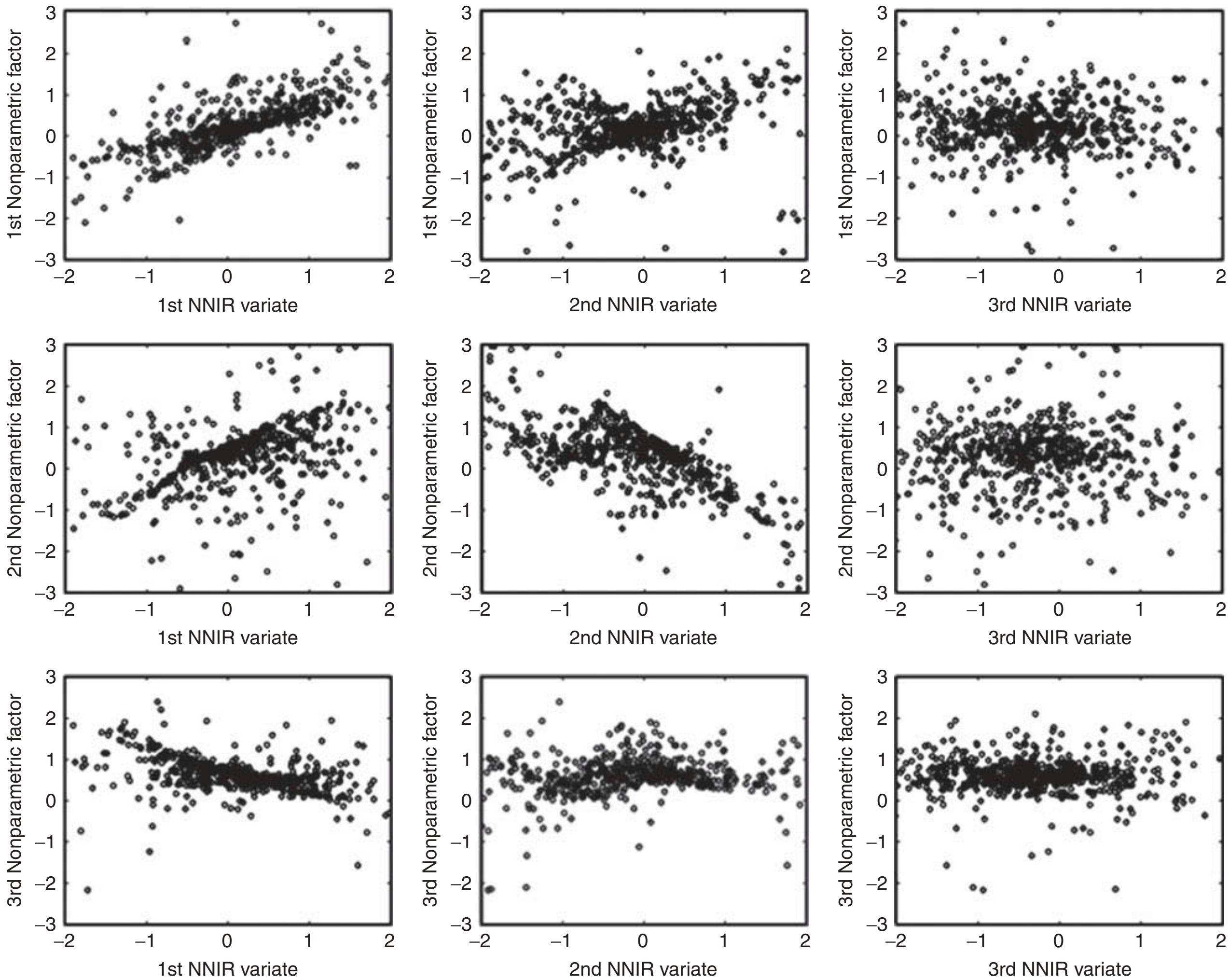

Fig. 3. Scatter plots of three nonparametric factors against three NNIR variates
Note: Nonparametric factors estimated as in (12) with bandwidth as selected by CV and shown in Table 3.

in the data, we focus our discussion on interpretation of the first nonparametric factor.

The first nonparametric factor is a nonlinear function of $X_i\hat{\beta}_1$ and $X_i\hat{\beta}_2$. In fact, the existence of more than one significant NNIR variate implies nonlinearity, because if the factor were linear in both $X_i\hat{\beta}_1$ and $X_i\hat{\beta}_2$, we would not be able to separately identify $\hat{\beta}_1$ and $\hat{\beta}_2$. To illustrate this nonlinearity and interpret this factor, we show in Fig. 4 two views of the relationship between this nonparametric factor and the first two NNIR variates. Panel A shows a surface plot of H_1 holding $X_i\hat{\beta}_3$ constant at its mean and varying $X_i\hat{\beta}_1$ and $X_i\hat{\beta}_2$ over a range that avoids the tails of the distribution. Panel B shows time series plots of H_{1i}, $X_i\hat{\beta}_1$ and $X_i\hat{\beta}_2$. For clarity, the figure shows calendar-year averages of these variables, rather than monthly observations. Panel B supports our leading indicator interpretation of $X_i\hat{\beta}_2$, as it tends to lead $X_i\hat{\beta}_1$.

The surface in Panel A of Fig. 4 is steepest when both variates are positive and flattest when both variates are close to zero. Thus, the two variates interact positively. Using our interpretation of the variates in the Section 'Interpretation of NNIR variates and comparison with canonical correlations', low-credit risk and low-inflationary expectations have a greater positive effect on stock returns when leading macroeconomic indicators are high relative to current economic activity. Panel B shows that such features arise early in the business cycle as the economy emerges from recession (e.g. 1976, 1982, 1991 and 2003).

As expansions proceed, we observe periods with positive $X_i\hat{\beta}_1$ and negative $X_i\hat{\beta}_2$ (e.g. mid 1960s, mid 1980s, mid-to-late 1990s and mid 2000s). In such periods, the economy is booming relative to its leading indicators, credit markets are loosening and

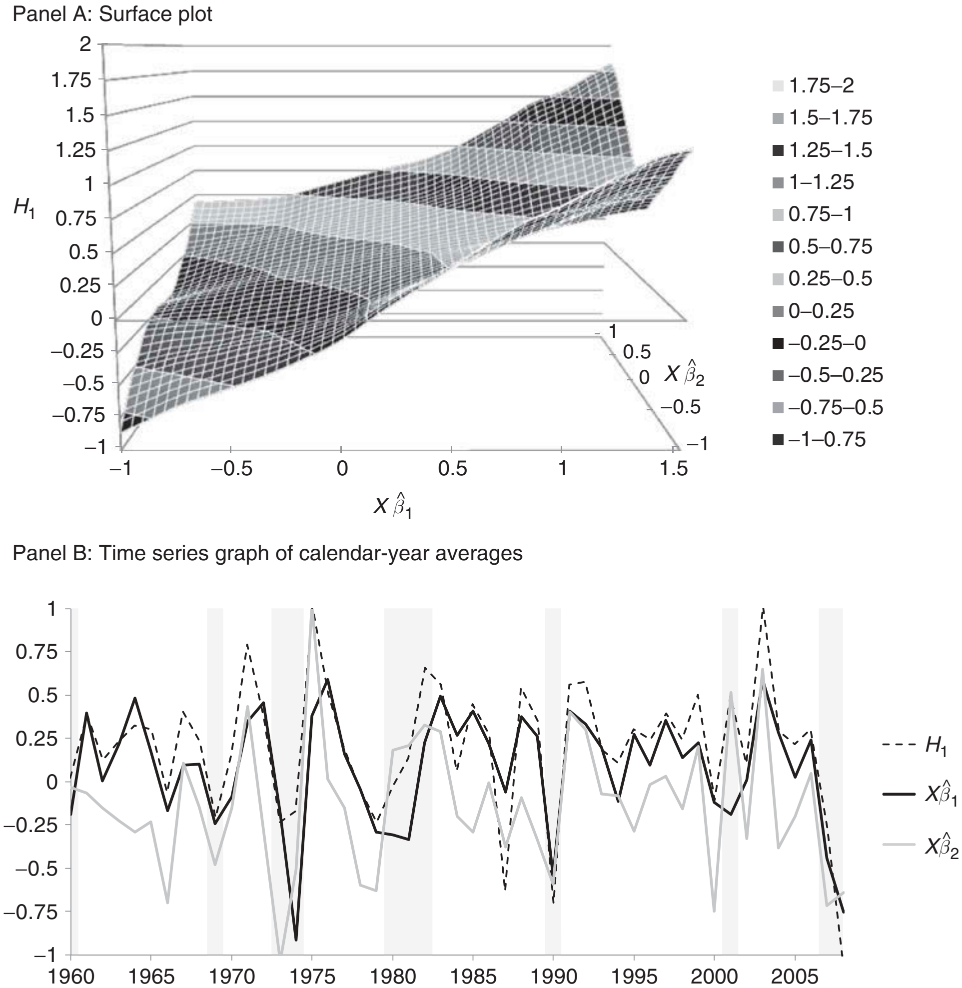

Fig. 4. Two views of the first nonparametric factor and the first two NNIR variates

inflationary expectations declining. The high values of the nonparametric factor in these periods reveals these are good times for stocks. Panel A shows that these periods coincide with larger values of the nonparametric factor than periods with $X_i\hat{\beta}_2$ close to zero. In other words, if $X_i\hat{\beta}_1$ is large and positive, then the nonparametric factor has a U-shaped relationship with $X_i\hat{\beta}_2$. Given that the region with positive $X_i\hat{\beta}_1$ and negative $X_i\hat{\beta}_2$ is often visited when the economy is booming, it may reflect the declining risk aversion. Reinforcing this conjecture is the fact that these episodes immediately preceded events such as the 1987 stock market crash, the 1996 'irrational exuberance' speech by Alan Greenspan, the 2000 dotcom crash and the recent subprime crisis.

Factor loadings

We show in Table 5 the factor loading estimates ($\hat{A}$) that relate the nonparametric factors to the 25 stock returns. The magnitudes of the loadings on the first factor are much greater than those for the other factors, reflecting the fact that the first factor has most of the explanatory power. Small stocks are more sensitive to the first factor than large stocks; the average loading for the smallest quintile is 3.2 compared to 2.2 for the largest quintile. This result is consistent with the notion that small stocks are more risky, and therefore co-vary more strongly with the aggregate risk factor.

Growth stocks, which are represented in the leftmost column, have larger factor loadings than any column to the right. This result is consistent with the notion that growth stocks, which consist mostly of young companies with small-current earnings relative to expected-future earnings, are more susceptible to aggregate risk than other stocks. The loadings are nonmonotonic in value quintile. The loading for the highest quintile exceeds the loading for the second-highest quintile in all cases.

Table 5. Nonparametric-factor loadings by portfolio

	Value 1	Value 2	Value 3	Value 4	Value 5
Loadings on first nonparametric factor					
Size 1	3.65	3.33	3.07	2.92	3.21
Size 2	3.47	3.17	2.93	2.88	3.09
Size 3	3.32	3.02	2.67	2.68	2.94
Size 4	2.99	2.81	2.81	2.57	2.79
Size 5	2.41	2.28	2.07	2.17	2.25
Loadings on second nonparametric factor					
Size 1	−0.98	−0.73	−0.43	−0.30	−0.34
Size 2	−0.52	−0.29	−0.03	0.14	−0.01
Size 3	−0.34	0.05	0.21	0.26	0.20
Size 4	−0.02	0.35	0.48	0.48	0.40
Size 5	0.42	0.57	0.68	0.65	0.51
Loadings on third nonparametric factor					
Size 1	−0.43	0.19	0.24	0.39	0.51
Size 2	−0.48	−0.02	0.31	0.34	0.44
Size 3	−0.56	0.01	0.19	0.21	0.36
Size 4	−0.60	−0.25	−0.01	0.13	0.15
Size 5	−0.67	−0.33	−0.24	−0.06	0.17

Notes: Portfolios formed by intersections of five portfolios formed on firm size and five portfolios formed on the ratio of book equity to market equity. Sizes 1–5 represent small – large stocks; values 1–5 present growth – value stocks. We calculate excess returns by subtracting the monthly average of the 1-month Treasury bill rate from each portfolio return. Sample period: 1960:2–2008:12.

V. Conclusions

In this article, we apply a new and convenient framework to analyse the high-dimensional nonlinear relationship between stock returns and macroeconomic risk factors. Rather than imposing a parametric model *a priori*, we take a nonparametric approach and apply dimension-reduction techniques to evade the *curse of dimensionality*. We reduce dimension in both the variable space and the function space. For the variable space, we reduce a large number of macroeconomic explanatory variables to a few inverse regression variates. For the function space, we model a large number of regression functions using a small number of nonparametric factors.

We apply this framework to 25 stock return portfolios. We use inverse regression to extract from a set of 15 macroeconomic variables three variates that have substantial explanatory power for stock returns. There are substantial gains in fit from moving beyond a single-index model and from incorporating nonlinearity. The first two inverse regression variates can naturally be interpreted as representing low-credit risk and low-inflationary expectations (the first variate) and strong leading economic indicators relative to current-economic activity (the second variate). The third variate captures inflation shocks and makes only a small contribution to the model. We show that combining the three inverse regression variates into a single nonparametric factor explains 95% of the common variation in the stock returns. This nonparametric factor is nonlinear in the variates, with its most notable nonlinear feature revealing an increasing appetite for risk during the expansion stage of the business cycle.

References

Acharya, V. and Pedersen, L. (2005) Asset pricing with liquidity risk, *Journal of Financial Economics*, **77**, 375–410.

Amato, U., Antoniadis, A. and De Feis, I. (2006) Dimension reduction in functional regression with applications, *Computational Statistics and Data Analysis*, **50**, 2422–46.

Anderson, H. M. and Vahid, F. (1998) Testing multiple equation systems for common nonlinear components, *Journal of Econometrics*, **84**, 1–36.

Barsky, R. B., House, C. L. and Kimball, M. S. (2007) Sticky-price models and durable goods, *The American Economic Review*, **97**, 984–98.

Bura, E. and Cook, R. D. (2001) Extending SIR: the weighted Chi-Square test, *Journal of the American Statistical Association*, **96**, 996–1003.

Camba-Mendez, G., Kapetanios, G., Smith, R. J. *et al*. (2003) Tests of rank in reduced rank regression models, *Journal of Business and Economic Statistics*, **21**, 145–55.

Carroll, R. J. and Li, K. C. (1992) Measurement error regression with unknown link: dimension reduction and data visualization, *Journal of the American Statistical Association*, **87**, 1040–50.

Chen, C. H. and Li, K. C. (1998) Can SIR be as popular as multiple linear regression, *Statistica Sinica*, **8**, 289–316.

Chen, N. F., Roll, R. and Ross, S. (1986) Economic forces and the stock market, *Journal of Business*, **59**, 383–403.

Cochrane, J. (1996) A cross-sectional test of an investment-based asset pricing model, *Journal of Political Economy*, **104**, 572–621.

Cook, R. D. (1994) On the interpretation of regression plots, *Journal of the American Statistical Association*, **89**, 177–89.

Cook, R. D. and Weisberg, S. (1991) Discussion of 'Sliced inverse regression' by K. C. Li, *Journal of the American Statistical Association*, **86**, 328–32.

Cook, R. D. and Yin, X. (2001) Dimension reduction and visualization in discriminant analysis, *Australian & New Zealand Journal of Statistics*, **43**, 147–99.

Donald, S. (1997) Inference concerning the number of factors in a multivariate nonparametric relationship, *Econometrica*, **65**, 103–31.

Fama, E. F. and French, K. R. (1992) The cross-section of expected stock returns, *Journal of Finance*, **47**, 427–65.

Fama, E. F. and French, K. R. (1993) Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics*, **33**, 3–56.

Fang, K.-T. and Zhu, L.-X. (1996) Asymptotics for kernel estimate of sliced inverse regression, *Annals of Statistics*, **24**, 1053–68.

Hall, P. and Li, K. C. (1993) On almost linearity of low dimensional projections from high dimensional data, *Annals of Statistics*, **21**, 867–89.

Hall, P., Li, Q. and Racine, J. S. (2007) Nonparametric estimation of regression functions in the presence of irrelevant regressors, *Review of Economics and Statistics*, **89**, 784–89.

Hamilton, J. D. (1994) *Time Series Analysis*, Princeton University Press, Princeton, NJ.

Hansen, L., Heaton, J. and Li, N. (2006) Consumption strikes back?, measuring long-run risk, *Journal of Political Economy*, **116**, 260–302.

Hsing, T. (1999) Nearest neighbor inverse regression, *Annals of Statistics*, **27**, 697–731.

Ichimura, H. (1993) Semiparametric least squares (SLS) and weighted SLD estimation of single-index models, *Journal of Econometrics*, **58**, 71–120.

Klein, R. L. and Spady, R. H. (1993) An efficient semiparametric estimator for binary response models, *Econometrica*, **61**, 387–421.

Lettau, M. and Ludvigson, S. (2004) Understanding trend and cycle in Asset values: reevaluating the wealth effect on consumption, *American Economic Review*, **94**, 279–99.

Lewellen, J., Nagel, S. and Shanken, J. (2010) A skeptical appraisal of asset pricing tests, *Journal of Financial Economics*, **96**, 175–94.

Li, K. C. (1991) Sliced inverse regression for dimension reduction, *Journal of the American Statistical Association*, **86**, 316–33.

Li, L. and Li, H. (2004) Dimension reduction methods for microarrays with application to censored survival data, *Bioinformatics*, **20**, 3406–12.

Naik, P. A., Wedel, M. and Kamakura, W. (2010) Multi-index binary response analysis of large data sets, *Journal of Business and Economic Statistics*, **28**, 67–81.

Pagan, A. and Ullah, A. (1999) *Nonparametric Econometrics*, Cambridge University Press, Cambridge.

Parker, J. and Julliard, C. (2005) Consumption risk and the cross dection of expected returns, *Journal of Political Economy*, **113**, 185–222.

Pastor, L. and Stambaugh, R. (2003) Liquidity risk and expected tock returns, *Journal of Political Economy*, **111**, 642–85.

Powell, J., Stock, J. and Stoker, T. (1989) Semiparametric estimation of index coefficients, *Econometrica*, **57**, 1403–30.

Ross, S. A. (1976) The arbitrage theory of capital asset pricing, *Journal of Economic Theory*, **13**, 341–60.

Stock, J. H. and Watson, M. W. (2002) Macroeconomic forecasting using diffusion indexes, *Journal of Business and Economic Statistics*, **20**, 147–62.

Vassalou, M. (2003) News related to future GDP growth as a risk factor in equity returns, *Journal of Financial Economics*, **68**, 47–73.

Yogo, M. (2006) A consumption-based explanation of expected stock returns, *Journal of Finance*, **61**, 539–80.