Maintaining prior distributions across evolving eigenspaces
An application to portfolio construction

Kevin R. Keane and Jason J. Corso
Department of Computer Science and Engineering
University at Buffalo, The State University of New York
338 Davis Hall, Buffalo, NY 14260-2500 USA
Email: {krkeane, jcorso}@buffalo.edu

Abstract—Temporal evolution in the generative distribution of nonstationary sequential data is challenging to model. This paper presents a method for retaining the information in prior distributions of matrix variate dynamic linear models (MVDLMs) as the eigenspace of sequential data evolves. The method starts by constructing sliding windows — matrices composed of a fixed number of columns containing the most recent point-in-time multivariate observation vectors. Characteristic time series, the right singular vectors, are extracted from a window using singular value decomposition (SVD). Then, a sequence of matrices capturing the rotation and scaling of the eigenspace is specified as a function of adjacent windows' characteristic time series. The method is tested on observations derived from daily US stock prices spanning 25 years. The results indicate that models constructed using sliding window SVD and MVDLMs, as extended in this paper, are resistant to over-fitting and perform well when used in portfolio construction applications.

Keywords-Online regression methods; unsupervised methods; applications of dynamic, online, incremental learning.

I. INTRODUCTION

We present a comprehensive, self-contained, unsupervised approach to modeling nonstationary sequential distributions using state space models with multivariate Gaussian latent variables. We require greater flexibility than the linear dynamical systems (LDS) discussed by Bishop [1, Ch. 13.3], where the models are intended for stationary sequential distributions. In particular, we specify time-varying transition and emission matrices. A static transition matrix is typically used to deterministically update components of a state vector, for example $(position, velocity, acceleration)^\mathsf{T}$, as the sequence progresses. In a novel extension, we specify *time-varying transition matrices* to accommodate data sequences with an evolving eigenspace. In essence, the definition of the state vector evolves as the sequence advances. We use transition matrices to transform a posterior distribution from the coordinate system of a previous window, into a prior distribution in the coordinate system of a current window. We use *time-varying emission matrices*, specified by characteristic time series extracted using singular value decomposition in the current window, to model common variation across the multivariate sequence. Figure 1 illustrates the
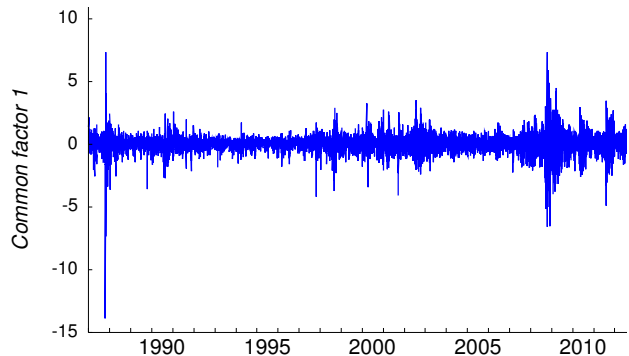

Figure 1. The dominant common factor in a high dimensional nonstationary sequential data set, the first difference of daily log prices ("returns") for approximately 1500 large US stocks, January 1987 to July 2012. The 25+ year daily time series is constructed by splicing characteristic time series — right singular vectors extracted using SVD — from sliding windows 252 trading days (approximately 1 year) in length. The construction maintains the relative magnitude of adjacent values throughout the sequence. The crash of '87, dot-com era, and recent financial crisis are clearly visible. Scaling shows the most recent 252 trading days with unit variance.

extraction of the dominant common source of variation from a data sequence we use to demonstrate our approach, a real nonstationary multivariate data set obtained from the MarketMap Analytic Platform *North American Pricing* database [2]. The sequential observations are the first difference of log prices, $\log(p_{i,t}/p_{i,t-1})$, for approximately 1500 large US stocks over a 25 year period (6300 trading days). We show evidence that our approach successfully resists over-fitting, yielding parameter estimates with a significantly greater density of zero entries. This approximately sparse estimated parameter structure arises naturally, without resorting to the forced imposition of a sparse graphical structure. We use the parameter estimates to construct evolving covariance matrices for the data stream; and, ultimately we use the sequence of estimated covariance matrices to construct two strategy portfolios each day. We report attractive back test performance of the portfolios built using our daily estimated parameters.

II. Background

A. Investment management

Successful portfolio construction requires robust multivariate statistical models to quantify the individual and joint distribution of asset price movements. Investment management seeks to maximize return and to minimize risk. *Return* is the change in an investment's value over a period of time, calculated as the first difference of log prices, $\log(p_t/p_{t-1})$, a transformation that yields a distribution approximately normal [3, Ch. 3]. *Factor returns* are common sources of returns. The return of a stock is modeled as a linear function of factor returns and residual stock specific return. *Risk* is the standard deviation of return. *Risk models* are estimated return covariance matrices. The trade-off between maximizing return and minimizing risk is initially explored by Markowitz [4], and combined in a statistic known as the *Sharpe ratio* [5] that measures return per unit of risk.

B. Risk models

The construction of risk models is challenging due to the non-stationary, high dimensional nature of financial time series [6]. Furthermore, the number of stocks n_t is not constant, evolving as companies arise through IPOs or spin-offs, and disappear due to events such as mergers, takeovers, buyouts and bankruptcy. We subscript n_t to emphasize the fact that the number of companies varies over time. The characteristics of individual companies also change due to innovation and obsolesce of products, acquisitions, and divestitures. At the portfolio level, factors broadly impacting prices vary as well. Pervasive evolution forces a local focus in the return series, reducing the effective number of observations τ. The combined effects of non-stationarity, high stock dimension n_t, and low number of observations τ, $\tau \ll n_t$, makes the direct estimation of the $n_t \times n_t$ return covariance matrix impractical. Practitioners resort to factor models and structured covariance matrices [7, Ch. 9] to overcome these challenges.

C. Factor sources

A variety of sources for factor returns are considered in the literature. A comparison of macroeconomic, fundamental, and statistical factors appears in [8]. Macroeconomic factor models use observable economic series for the factor returns, and estimate factor loadings. Fundamental factor models use observable attributes such as industry, earnings and dividends for the factor loadings, and estimate factor returns cross sectionally. Statistical factor models extract latent times series such as those identified by principal components analysis (PCA) or singular value decomposition (SVD) for use as factor returns, and estimate factor loadings by some form of regression. Financial index models [9] are somewhat a hybrid of fundamental factor and statistical factor models, in that the explanatory time series arise from portfolios constructed based on observable stock attributes;

but, the factor loadings are determined by regression. Using the Connor [8] taxonomy, our approach is statistical. We extract explanatory series from a data stream, and model the dependence of individual assets on these evolving characteristic series using a matrix variate dynamic linear model.

D. Constructing portfolios

An analytic methodology for maximizing the Sharpe ratio is presented by Chen et al. [10]. Given an estimated return vector $\hat{\boldsymbol{\alpha}}$ and an estimated covariance matrix $\hat{\boldsymbol{\Sigma}}$, the asset weight vector $\boldsymbol{w}$ defining the portfolio with the maximum Sharpe ratio is:

$$\boldsymbol{w} \propto \hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\alpha}} \quad . \tag{1}$$

Constraining the sum of the weights, $\mathbf{1}^\mathsf{T}\boldsymbol{w} = 1$, we identify a unique vector $\boldsymbol{w}$ given $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\Sigma}}$.

E. Matrix variate dynamic linear models

We will describe our application primarily in the framework of matrix variate dynamic linear models (MVDLMs) [9, 11], due to the established notation and extensive literature for modeling nonstationary sequential data, including the routine treatment of time varying transition and emission matrices. Table I provides a quick summary of notational differences between the two equivalent frameworks, LDS and DLMs. The matrix variate DLM form allows consideration of multiple observation series simultaneously. The observed response $\boldsymbol{Y}_t$ is an n_t-vector of asset returns $Y_{i,t}$. The time t *observation equation* is:

$$\boldsymbol{Y}_t^\mathsf{T} = \boldsymbol{F}_t^\mathsf{T}\boldsymbol{\Theta}_t + \boldsymbol{\epsilon}_t^\mathsf{T}, \quad \boldsymbol{\epsilon}_t \sim \mathrm{N}\left[0, \boldsymbol{\Psi}_t\right] \tag{2}$$

Table I
TRANSLATING THE NOTATION OF LDS TO MVDLMs

	Linear dynamical systems		Matrix variate dynamic linear models
$\boldsymbol{z}_t$	latent variables	$\boldsymbol{\Theta}_t$	$p \times n_t$ system matrix
$\boldsymbol{A}$	transition matrix	$\boldsymbol{G}_t$	$p \times p$ evolution matrix
$\boldsymbol{\Gamma}$	transition variance matrix	$\boldsymbol{W}_t$	$p \times p$ left variance matrix
		$\boldsymbol{\Psi}_t$	$n_t \times n_t$ right variance matrix
	transition distribution		system distribution
	$(\boldsymbol{z}_t\|\boldsymbol{z}_{t-1}) \sim$		$(\boldsymbol{\Theta}_t\|\boldsymbol{\Theta}_{t-1}) \sim$
	$\mathrm{N}\left[\boldsymbol{A}\boldsymbol{z}_{t-1}, \boldsymbol{\Gamma}\right]$		$\mathrm{N}\left[\boldsymbol{G}_t\boldsymbol{\Theta}_{t-1}, \boldsymbol{W}_t, \boldsymbol{\Psi}_t\right]$
			matrix normal distribution
			[11, Ch. 16.4.2]
$\boldsymbol{x}_t$	observed variables	$\boldsymbol{Y}_t$	$n_t \times 1$ observation matrix
$\boldsymbol{C}$	emission matrix	$\boldsymbol{F}_t$	$p \times 1$ design matrix
$\boldsymbol{\Sigma}$	emission variance matrix	$\boldsymbol{\Psi}_t$	$n_t \times n_t$ observation variance
			same as right variance ↑
	emission distribution		observation distribution
	$(\boldsymbol{x}_t\|\boldsymbol{z}_t) \sim$		$(\boldsymbol{Y}_t^\mathsf{T}\|\boldsymbol{\Theta}_t) \sim$
	$\mathrm{N}\left[\boldsymbol{C}\boldsymbol{z}_t, \boldsymbol{\Sigma}\right]$		$\mathrm{N}\left[\boldsymbol{F}_t^\mathsf{T}\boldsymbol{\Theta}_t, \boldsymbol{\Psi}_t\right]$
$\boldsymbol{\mu}_0$	initial latent variable mean	$\boldsymbol{m}_0$	initial information mean
$\boldsymbol{V}_0$	initial latent variable variance	$\boldsymbol{C}_0$	initial information variance

$\boldsymbol{\Theta}_t$ is the $p \times n_t$ latent parameter matrix. $\boldsymbol{F}_t$ is a $p \times 1$ design matrix. $\boldsymbol{\epsilon}_t$ is an n_t-vector of observation errors. $\boldsymbol{\Psi}_t$ is a $n_t \times n_t$ diagonal observation variance matrix. (2) expresses the view that we observe a linear function of the system matrix $\boldsymbol{\Theta}_t$, a design matrix $\boldsymbol{F}_t$, and noise vector $\boldsymbol{\epsilon}_t$.

In our application described below, $\boldsymbol{\Lambda}_t$, the covariance of the design matrix $\boldsymbol{F_t}$ is the identity matrix by construction. The time t asset covariance matrix is:

$$\boldsymbol{\Sigma}_t = \boldsymbol{\Theta}_t^{\mathsf{T}} \boldsymbol{\Lambda}_t \boldsymbol{\Theta}_t + \boldsymbol{\Psi}_t = \boldsymbol{\Theta}_t^{\mathsf{T}} \boldsymbol{\Theta}_t + \boldsymbol{\Psi}_t \qquad (3)$$

The time t *evolution equation* for the system matrix is:

$$\boldsymbol{\Theta}_t = \boldsymbol{G}_t \boldsymbol{\Theta}_{t-1} + \boldsymbol{\Omega}_t, \quad \boldsymbol{\Omega}_t \sim \mathrm{N}\left[0, \boldsymbol{W}_t, \boldsymbol{\Psi}_t\right], \qquad (4)$$

where $\boldsymbol{\Omega}_t$ is *matrix normal* [12] with row covariance $\boldsymbol{W}_t$ and column covariance $\boldsymbol{\Psi}_t$. (4) describes the evolution of the system matrix $\boldsymbol{\Theta}_t$. $\boldsymbol{G}_t$ is a deterministic component, the specification of which we will propose and discuss in detail in §III-E. $\boldsymbol{\Omega}_t$ captures stochastic evolution of the system matrix.

III. METHODOLOGY

A. Constructing sliding windows

Subsequent to each trading day t, we construct a sliding window, the $n_t \times \tau$ matrix of asset returns $\boldsymbol{X}_t$. Columns 1 to τ correspond to time periods $t - \tau + 1$ to t. Rows 1 to n_t contain transposed individual asset return vectors, scaled to unit length for the window. We use $\tau = 252$ periods (approximately 1 year), $n_t \approx 1500$ assets, and $p = 10$ common factors. The value of n_t is determined by including the top 95% of the market capitalization of US stocks from the historical pricing data base [2] on day t.

B. Singular value decomposition (SVD)

The font in (5) *is to emphasize that the SVD scale matrix $\boldsymbol{\mathcal{S}}_t$ is unrelated to the DLM estimated observation variance matrix $\boldsymbol{S}_t$ used elsewhere in this paper.* We use SVD [13] to extract characteristic time series from the sliding window $\boldsymbol{X}_t$. Our approach to identifying characteristic time series is similar to [14], although Connor and Korajczyk use asymptotic principal components. Connor and Korajczyk propose a two pass approach, down-weighting outlier asset return series in the second pass. They note the results change little with the second pass when the characteristic series are extracted from a large matrix of asset returns. Informally, this implies that a few noisy time series do not materially degrade the identification of *pervasive* time series; and, that the inherent noise reduction of a PCA or SVD approach is effective. A discussion of SVD and its relation to PCA appears in [13]. Using SVD, we factor $\boldsymbol{X}_t$:

$$\boldsymbol{X}_t = \boldsymbol{\mathcal{U}}_t \boldsymbol{\mathcal{S}}_t \boldsymbol{\mathcal{V}}_t^{\mathsf{T}} \qquad (5)$$

The first p columns of right singular vectors $\boldsymbol{\mathcal{V}}_t$ correspond to the p largest singular values in diagonal matrix $\boldsymbol{\mathcal{S}}_t$.

These p vectors are orthonormal, and explain the maximum variation in the rows of $\boldsymbol{X}_t$ possible when limiting an approximation to rank p.

C. Specifying the design matrix

The first p entries from the last row of $\boldsymbol{\mathcal{V}}_t$, transposed and scaled by $\sqrt{\tau}$ to attain unit variance, define our $p \times 1$ design matrix $\boldsymbol{F_t}$. In §III-E, we will need to access several rows from $\boldsymbol{\mathcal{V}}_t$ similarly scaled and transposed. We will refer to the larger $p \times \tau$ design matrix as $\boldsymbol{\mathcal{F}}_t$.

D. Achieving continuity across sliding windows

The SVD's right singular vectors, interpreted as characteristic time series (factors) in this application, rotate, scale, permute order, and rise or fall from consideration as the data window slides forward in time. In a novel contribution of this paper, we accommodate these transformations of parameter space using the evolution matrices $\boldsymbol{G}_t$. This permits the latent parameter posterior distributions from one period to be carried forward as prior distributions in the next period. PCA or SVD typically identify a distinct eigen basis in each window. Discussions in the literature focus on eigenspace tracking [15]. We want to track the current eigenspace *and* retain accumulated knowledge in the form of prior distributions as the eigenspace evolves. We address the second requirement by specifying a series of evolution matrices $\boldsymbol{G}_t$.

E. Specifying the evolution matrices

It is desirable to require approximately equal predictions in adjacent windows, when using each window's distinct basis, parameter estimates, and design matrix:

$$\boldsymbol{\mathcal{F}}_t^{\mathsf{T}} \boldsymbol{a}_t \approx \boldsymbol{\mathcal{F}}_{t-1}^{\mathsf{T}} \boldsymbol{m}_{t-1} \quad . \qquad (6)$$

As discussed in §III-F below, matrix $\boldsymbol{a}_t$ is the mean of the prior distribution at time t; and $\boldsymbol{m}_t$ is the mean of the posterior distribution at time $t - 1$. Matrix $\boldsymbol{\mathcal{F}}_t$ provides characteristic returns for the period $t - \tau + 1$ to t. Matrix $\boldsymbol{\mathcal{F}}_{t-1}$ provides characteristic returns for the period $t - \tau$ to $t - 1$. To avoid excessively intricate notation in (6), we refer implicitly to the *overlapping* $(\tau - 1) \times p$ matrices of scaled right singular vectors from adjacent data windows, for the common period $t - \tau + 1$ to $t - 1$.

Pre-multiplying (6) by the left inverse of $\boldsymbol{\mathcal{F}}_t^{\mathsf{T}}$, we express the time t prior distribution mean $\boldsymbol{a}_t$ as a function of the matrices $\boldsymbol{\mathcal{F}}_{t-1}$, $\boldsymbol{\mathcal{F}}_t$, and $\boldsymbol{m}_{t-1}$:

$$\boldsymbol{a}_t \approx \left(\boldsymbol{\mathcal{F}}_t \boldsymbol{\mathcal{F}}_t^{\mathsf{T}} \right)^{-1} \boldsymbol{\mathcal{F}}_t \boldsymbol{\mathcal{F}}_{t-1}^{\mathsf{T}} \boldsymbol{m}_{t-1} \quad , \qquad (7)$$

and identify the evolution matrix $\boldsymbol{G}_t$ required by (4) and (9) as:

$$\boldsymbol{G}_t = \left(\boldsymbol{\mathcal{F}}_t \boldsymbol{\mathcal{F}}_t^{\mathsf{T}} \right)^{-1} \boldsymbol{\mathcal{F}}_t \boldsymbol{\mathcal{F}}_{t-1}^{\mathsf{T}} \quad . \qquad (8)$$

F. Estimating latent parameters

1) Prior distribution: Following [11, Ch. 16.4] and [9, §7], the joint matrix normal / inverse Wishart *prior* with $\delta\nu_{t-1}$ degrees of freedom for $\boldsymbol{\Theta}_t$ and $\boldsymbol{\Psi}_t$ at time t is:

$$(\boldsymbol{\Theta}_t, \boldsymbol{\Psi}_t | D_{t-1}) \sim \text{NW}_{\delta\nu_{t-1}}^{-1} [\boldsymbol{a}_t, \boldsymbol{R}_t, \delta\boldsymbol{S}_{t-1}] \quad \text{where}$$
$$(9)$$
$$\boldsymbol{a}_t = \boldsymbol{G}_t \boldsymbol{m}_{t-1} \text{ and } \boldsymbol{R}_t = \boldsymbol{G}_t \boldsymbol{C}_{t-1} \boldsymbol{G}_t^\mathsf{T} + \boldsymbol{W}_t.$$

Particular to our approach, $\boldsymbol{G}_t$ performs a change of basis operation on the posterior distribution from the previous period. Then, as is standard, the prior variance is inflated with the addition of the $\boldsymbol{W}_t$ term to capture evolution variance entering the system. In this application, we specify $\boldsymbol{W}_t \approx 4 \times 10^{-6} \boldsymbol{I}$.

2) Forecast distribution: Using the most recently observed $p \times 1$ design matrix $\boldsymbol{F}_t$ discussed in §III-C, the multivariate normal conditional / multivariate T marginal one-step *forecast* distributions for Y_t are:

$$(\boldsymbol{Y}_t | \boldsymbol{\Psi}_t, D_{t-1}) \sim \text{N}[\boldsymbol{f}_t, Q_t \boldsymbol{\Psi}_t] \text{ and}$$

$$(\boldsymbol{Y}_t | D_{t-1}) \sim \text{T}_{\delta\nu_{t-1}}[\boldsymbol{f}_t, \delta Q_t \boldsymbol{S}_{t-1}] \quad \text{where} \quad (10)$$

$$\boldsymbol{f}_t = \boldsymbol{a}_t^\mathsf{T} \boldsymbol{F}_t \text{ and } Q_t = \boldsymbol{F}_t^\mathsf{T} \boldsymbol{R}_t \boldsymbol{F}_t + 1.$$

3) Posterior distribution: The joint matrix normal / inverse Wishart *posterior* with ν_t degrees of freedom for $\boldsymbol{\Theta}_t$ and $\boldsymbol{\Psi}_t$ at time t is:

$$(\boldsymbol{\Theta}_t, \boldsymbol{\Psi}_t | D_t) \sim \text{NW}_{\nu_t}^{-1}[\boldsymbol{m}_t, \boldsymbol{C}_t, \boldsymbol{S}_t] \quad \text{where}$$

$$\boldsymbol{m}_t = \boldsymbol{a}_t + \boldsymbol{A}_t \boldsymbol{e}_t^\mathsf{T}, \quad \boldsymbol{C}_t = \boldsymbol{R}_t - \boldsymbol{A}_t \boldsymbol{A}_t^\mathsf{T} Q_t,$$

$$\nu_t = \delta\nu_{t-1} + 1, \quad \boldsymbol{S}_t = \delta\boldsymbol{S}_{t-1} + \nu_t^{-1} \boldsymbol{e}_t \boldsymbol{e}_t^\mathsf{T} / Q_t,$$

$$\boldsymbol{A}_t = \boldsymbol{R}_t \boldsymbol{F}_t / Q_t, \quad \text{and } \boldsymbol{e}_t = \boldsymbol{Y}_t - \boldsymbol{f}_t.$$
$$(11)$$

$$(\boldsymbol{\Theta}_{t-1}, \boldsymbol{\Psi}_{t-1} | D_{t-1}) \sim \text{NW}_{\nu_{t-1}}^{-1}[\boldsymbol{m}_{t-1}, \boldsymbol{C}_{t-1}, \boldsymbol{S}_{t-1}]$$
$$(12)$$

Given our assumption that $\boldsymbol{\Psi}_t$ is diagonal, we only maintain the diagonal elements of $\boldsymbol{S}_t$.

G. Constructing risk models

We construct an $n_t \times n_t$ structured covariance matrix after each trading day t using estimates from (9); and, to obtain the next period's estimated observation variance matrix $\hat{\boldsymbol{\Sigma}}_{t+1}$ *before* observing the $p \times 1$ vector $\boldsymbol{F}_{t+1}$, we use the random regression vector DLM framework of Wang and Carvalho [9, §7]. We assume $\boldsymbol{G}_{t+1} = \boldsymbol{I}$ (no change to the eigenspace), and following [9, Thrm. 2.(iii)], obtain:

$$\hat{\boldsymbol{\Sigma}}_{t+1} = \boldsymbol{a}_{t+1}^\mathsf{T} \boldsymbol{a}_{t+1} + (1 + \text{tr}(\boldsymbol{R}_{t+1})) \boldsymbol{S}_t \quad . \quad (13)$$

Comparing (13) with (3), it is apparent we increase the scale of the estimated observation variance matrix $\boldsymbol{S}_t$ to

account for parameter uncertainty expressed in the prior variance $\boldsymbol{R}_{t+1}$. Note that we really need $\hat{\boldsymbol{\Sigma}}_{t+1}^{-1}$ rather than $\hat{\boldsymbol{\Sigma}}_{t+1}$ for portfolio construction as noted in (1) and (14). To invert the structured covariance matrix (13), we exploit the ease of inverting diagonal matrix $\boldsymbol{S}_t$, and use the Sherman-Morrison-Woodbury formula [16].

IV. DISCUSSION OF MVDLM FACTOR LOADINGS

The matrix variate DLM appears to resist over-fitting. When reviewing the distribution of elements in the estimated latent parameter matrices, the increase in the density of zero elements in Figure 2 relative to the estimates obtained from PCA is striking. While PCA seeks to concisely duplicate the data presented, the matrix variate DLM appears to be well regularized by its prior distribution. Note the distribution of loadings on the first characteristic time series are similar across methods. For the remaining factor loadings, the matrix variate DLM output is significantly more concentrated around zero.

Upon reflection, this property is perhaps expected. The explanatory time series are independent by construction. The *superposition principle* [11] permits the composition of independent univariate models. To the extent that the superposition principle is valid, we would expect a DLM to produce zero loadings on irrelevant series in univariate components; and, similarly to produce zero loadings in an equivalent multivariate composite. This approximately sparse output and its explanation deserve further study.

V. PORTFOLIO STRATEGIES

A. Obtaining the asset weight vectors

For each trading day during the 25 year period January 1987 to December 2011, we construct four portfolios: maximum Sharpe ratio (MSR), global minimum variance (GMV), equal weight (EW), and capitalization weight (CW). The first two portfolios require a risk model, the second two portfolios do not. Using our unsupervised approach, composed of sliding window SVD and matrix variate DLMs, we construct risk models each day for approximately 1500 assets for 25 years (approximately 6300 trading days) using a historical data set [2]. The risk models are then used to construct the GMV and MSR portfolios; and, we construct the CW and EW portfolios without a risk model. The CW weights are proportional to each asset's capitalization, i.e. shares outstanding $\times$ price per share. The EW weights are simply n_t^{-1} for each asset.

For GMV and MSR, we must specify an estimated return vector $\hat{\boldsymbol{\alpha}}_t$. The GMV strategy [17] assumes that all assets have equivalent expected returns, so we specify a constant positive value for all entries of $\hat{\boldsymbol{\alpha}}_{\text{GMV}}$. The MSR strategy [18] assumes that all assets have equivalent risk adjusted returns, $\frac{\alpha_i}{\sigma_i} = \frac{\alpha_j}{\sigma_j} \; \forall \; i, j$, so we specify the estimated individual standard deviations $\hat{\sigma_i}$ for the entries of $\hat{\boldsymbol{\alpha}}_{\text{MSR}}$.
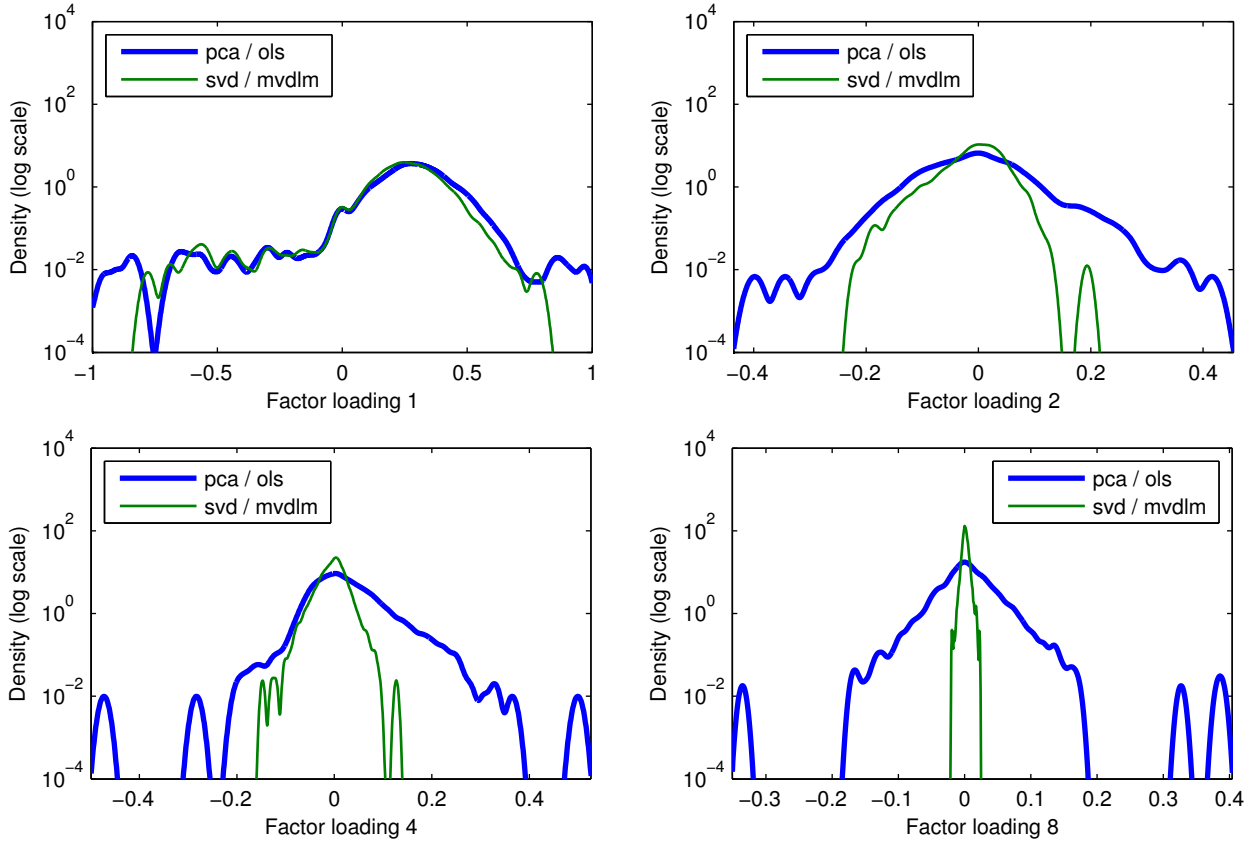
Figure 2. Comparison of the distribution of factor loading (latent parameter) estimates for December 30, 2011. Estimates obtained using the method of Connor and Korajczyk [14] are labeled *pca / ols*. Estimates obtained using the method proposed in this paper are labeled *svd / mvdlm*. Our method results in a greater density of zero factor loading estimates. Four of ten factor loading spectra shown. Omitted factor loading spectra show a similar progression towards sparsity.

Given $\hat{\boldsymbol{\Sigma}}_t$, $\hat{\boldsymbol{\alpha}}_{\text{GMV}}$, and $\hat{\boldsymbol{\alpha}}_{\text{MSR}}$, portfolio weights are computed using the appropriate choice for vector $\hat{\boldsymbol{\alpha}}_t$:

$$\boldsymbol{w}_t = \frac{\hat{\boldsymbol{\Sigma}}_t^{-1}\hat{\boldsymbol{\alpha}}_t}{\mathbf{1}^{\mathsf{T}}\hat{\boldsymbol{\Sigma}}_t^{-1}\hat{\boldsymbol{\alpha}}_t} \tag{14}$$

where $\mathbf{1}$ is a n_t-vector of ones, and the denominator scales the weights such that $\mathbf{1}^{\mathsf{T}}\boldsymbol{w}_t = 1$.
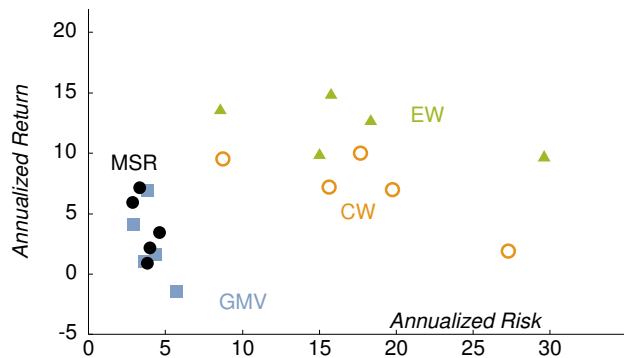
B. Results

We compare the Sharpe ratios of the four strategies during sub-periods of 5 years, and for the entire 25 year period. The results are for zero-investment strategies. We deduct funding expenses: the long positions are charged the risk-free rate plus 25 basis points (bp); cash from short positions receives the risk-free rate less 25 bp. We deduct transaction costs of 10 bp when rebalancing.

CW is the default diversified benchmark for the capital asset pricing model (CAPM) [19]. The current popularity of the other three strategies — EW, GMV, and MSR — is indicative of the fact that the efficient market hypothesis is questioned by practitioners and investors. In fact, all three alternative strategies beat the CW results on a risk adjusted

basis (Sharpe ratio) over the 25 year back test, as shown in Figure 3(b) and Table II. The GMV and MSR portfolios are broadly similar in risk, with MSR having generally higher returns than GMV. The CW and EW portfolios are similar in risk, with EW having generally higher return than CW. Relative to the CW and EW portfolios, the risk model exploited by the GMV and MSR portfolios resulted in dramatically lower risk in each of sub-period. Surprisingly, the ex post risk of the MSR portfolio was similar or lower than the GMV portfolio in several sub-periods and for the full period. The MSR portfolio lived up to its name, *maximum Sharpe ratio*, in 3 of the 5 periods and overall. When MSR did not dominate, it remained competitive with the dominant strategy for the period.

VI. Summary

We proposed a novel method to specify the evolution matrices required to convey prior information across an evolving eigenspace in matrix variate dynamic linear models. We demonstrated the effective implementation of this method in a portfolio construction application requiring the estimation of a high dimensional statistical model. We noted

(a) Return versus risk for 5-year sub-periods.
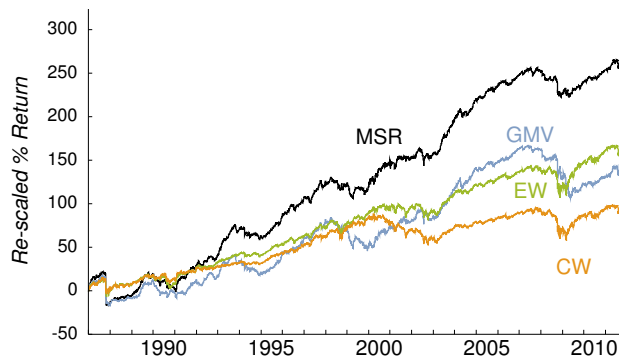
(b) Cumulative returns rescaled to 10% average volatility.

Figure 3. Results of a 25 year back test for four portfolio strategies. CW – capitalization weighted portfolio; EW – equal weighted portfolio; GMV – global minimum variance portfolio; and, MSR – maximum Sharpe ratio portfolio. The CW and EW portfolios do not require a risk model. The GMV and MSR portfolios are constructed using the risk model specified in §III.

Table II
STRATEGY PERFORMANCE

	Maximum Sharpe Ratio Portfolio			Global Minimum Variance Portfolio			Equal Weight Portfolio			Capitalization Weight Portfolio		
	Annualized		Sharpe	Annualized		Sharpe	Annualized		Sharpe	Annualized		Sharpe
Dates	Return	Volatility	Ratio	Return	Volatility	Ratio	Return	Volatility	Ratio	Return	Volatility	Ratio
1987 - 1991	2.19	3.99	0.55	1.08	3.66	0.30	9.96	15.00	**0.66**	10.02	17.67	0.57
1992 - 1996	5.93	2.86	**2.07**	4.11	2.89	1.42	13.69	8.58	1.60	9.52	8.72	1.09
1997 - 2001	3.44	4.62	**0.74**	1.60	4.34	0.37	12.79	18.34	0.70	7.01	19.76	0.35
2002 - 2006	7.13	3.30	**2.16**	6.94	3.81	1.82	14.94	15.77	0.95	7.20	15.64	0.46
2007 - 2011	0.91	3.80	0.24	-1.41	5.72	-0.25	9.79	29.62	**0.33**	1.91	27.28	0.07
1987 - 2011	3.99	3.77	**1.06**	2.40	4.20	0.57	12.30	18.51	0.66	7.15	18.63	0.38

a high density of zeros in the estimated parameter matrices, particularly in contrast with existing methods. We hope our technique, in expanding the utility of standard tools such as SVD and matrix variate DLMs, will encourage their use in other challenging high dimensional, dynamic domains.

REFERENCES

[1] C. Bishop, *Pattern Recognition and Machine Learning.* Springer Science+Business Media, LLC., 2006.
[2] MarketMap Analytic Platform, *North American Pricing.* SunGard Data Systems, Inc., 2012.
[3] A. Meucci, *Risk and asset allocation.* Springer Verlag, 2009.
[4] H. Markowitz, "Portfolio selection," *The Journal of Finance*, vol. 7, no. 1, pp. 77–91, 1952.
[5] W. Sharpe, "The Sharpe ratio," *Journal of Portfolio Management*, vol. 21, no. 1, pp. 49–58, 1994.
[6] J. Bai and S. Shi, "Estimating high dimensional covariance matrices and its applications," *Annals of Economics and Finance*, vol. 12, no. 2, pp. 199–215, 2011.
[7] R. Johnson and D. Wichern, *Applied multivariate statistical analysis*, 4th ed. Prentice Hall, 1998.
[8] G. Connor, "The three types of factor models: A comparison of their explanatory power," *Financial Analysts Journal*, pp. 42–46, 1995.
[9] H. Wang and C. Carvalho, "Dynamic financial index models: Modeling conditional dependencies via graphs," *Bayesian Analysis*, vol. 6, no. 4, pp. 639–664, 2011.
[10] H. Chen, H. Tsai, and D. Lin, "Optimal mean-variance portfolio selection using Cauchy–Schwarz maximization," *Applied Economics*, vol. 43, no. 21, pp. 2795–2801, 2011.
[11] M. West and J. Harrison, *Bayesian forecasting and dynamic models.* Springer Verlag, 1997.
[12] A. Dawid, "Some matrix-variate distribution theory: notational considerations and a Bayesian application," *Biometrika*, vol. 68, no. 1, p. 265, 1981.
[13] M. Wall, A. Rechtsteiner, and L. Rocha, "Singular value decomposition and principal component analysis," *A Practical Approach to Microarray Data Analysis*, pp. 91–109, 2003.
[14] G. Connor and R. Korajczyk, "Risk and return in an equilibrium APT: Application of a new test methodology," *Journal of Financial Economics*, vol. 21, no. 2, pp. 255–289, 1988.
[15] S. Papadimitriou, J. Sun, and P. Yu, "Local correlation tracking in time series," in *Data Mining, 2006. ICDM'06. Sixth International Conference on.* IEEE, 2006, pp. 456–465.
[16] Bindel,D., "Lecture 12: Sherman-Morrison-Woodbury," http://www.cs.cornell.edu/~bindel/class/cs6210-f09/lec12.pdf, 2009.
[17] R. Clarke, H. de Silva, and S. Thorley, "Minimum-variance portfolios in the US equity market," *The Journal of Portfolio Management*, vol. 33, no. 1, pp. 10–24, 2006.
[18] L. Martellini, "Toward the design of better equity benchmarks," *The Journal of Portfolio Management*, vol. 34, no. 4, pp. 34–41, 2008.
[19] W. Sharpe, "Capital asset prices: A theory of market equilibrium under conditions of risk," *The Journal of Finance*, vol. 19, no. 3, pp. 425–442, 1964.