Dynamic Factor Models and Reduced Rank Regression in High-dimensional Time Series

Tze Leung Lai* Ka Wai Tsang† Hongsong Yuan‡

Abstract
We give a brief review of dynamic factor models and their financial and econometric applications. We then describe new methods to address some long-standing difficulties in choosing the factors or more targeted predictors and modeling their dynamics in high-dimensional time series.

Key Words: Canonical correlation analysis, dynamic factor models, macroeconomic time series, multivariate stochastic regression, group orthogonal matching pursuit, rank selection.

1. Introduction

Factor models are widely used for multivariate econometric and financial time series data. In finance, factor models are widely used for stock returns following the Capital Asset Pricing Model (CAPM) introduced by Sharpe (1964) and the Arbitrage Pricing Theory (APT) introduced by Ross (1976). Section 2 describes some important developments, particularly time series models of these factors and their applications to forecasting portfolio returns involving a large number of stocks. In econometrics, factor models have a long tradition in macroeconomics, dating back to Burns and Mitchell (1947) and subsequent developments of economic indicators at the National Bureau of Economic Research. Section 3 gives a review of dynamic factor models in macroeconomics, culminating in developments in the big data era of the past decade in which the increasing computational power has enabled one to evaluate a large number of variables that summarize macroeconomic activities and to extract from them a much smaller number of factors for time series modeling and prediction. Section 4 describes an alternative approach that uses more targeted predictors than dynamic factors. Further discussion and concluding remarks are given in Section 5.

2. Dynamic Factor Models of Asset Returns

Lai and Xing (2008, Section 3.3 and 3.4) give an overview of CAPM and APT and the underlying financial theory and statistical methods – linear regression and factor (or principal component) analysis. Chamberlain and Rothschild (1983) point out that Ross (1976) has only presented a heuristic justification of APT that claims the absence of arbitrage in a market with a very large number of assets to imply that “asset prices are approximately linear functions of factor loadings." They use a Hilbert space with the mean-square inner product to formulate (a) the prices of assets and their portfolios and (b) conditions for lack of arbitrage opportunities. In this framework they define a strict K-factor structure for the returns of the first N assets by $\mathbf{\Sigma}_N = \mathbf{B}_N\mathbf{B}_N^T + \mathbf{D}_N$, where $\mathbf{\Sigma}_N$ is the covariance matrix of these returns, $\mathbf{B}_N$ is $N \times K$ and $\mathbf{D}_N$ is a diagonal matrix whose elements are uniformly bounded by some constant ρ. They also define an approximate K-factor structure if $\mathbf{D}_N$ is replaced by a nonnegative definite matrix $\mathbf{R}_N$ such that $\sup_N \lambda_{\max}(\mathbf{R}_N) < \infty$. They prove that an approximate K-factor structure is sufficient for Ross’s theory. Moreover, they also ahow

*Stanford University, Stanford, CA, USA
†Chinese University of Hong Kong in Shenzhen, Shenzhen, China
‡Shanghai University of Finance & Economics, Shanghai, China

that if $\mathbf{\Sigma}_N$ has only K unbounded eigenvalues, then an approximate K-factor structure exists and is unique, and that the corresponding K unit-length eigenvectors converge and play the role of factor loadings.

Under the classical assumption of i.i.d. returns, factor models as elucidated by Chamberlain and Rothschild have played a very important role in dimension reduction and estimation of the covariance matrix $\mathbf{\Sigma}_N$. However, stylized facts about asset returns have shown clear deviations from i.i.d. model, especialy for the volatilities of the returns; see Chapter 6 of Lai and Xing (2008). Hence time series models of the factors and the idiosyncratic (asset-specific) residuals have been proposed for various financial applications. One such model is given in Lai et al. (2011). Section 5.3.2 and Section 9.4.2 of Lai and Xing (2008) have also given other examples of such models.

3. Dynamic Factor Models in Macroeconomics

Stock and Watson (2011) give a survey (up to 2010) of dynamic factor models (DFMs) in macroeconomics following the seminal works by Sargent and Sius (1977) and Geweke (1977), toward a dynamic generalization of classical factor analysis models by using frequency domain methods to identify a dynamic factor structure which, however, could not be used for forecasting. They describe the works of Engle and Watson (1981, 1983) and Sargent (1989) as the first generation of time-domain DFMs for forecasting, expressing the DFM as a linear Gaussian state-space model and using maximum likelihood (via the EM algorithm) to estimate the parameters of the state-space model and the Kalman filter to estimate the latent state vectors whose components are factors. Specially, the state-space model is

$$\mathbf{\Phi}(L)\mathbf{f}_t = \boldsymbol{\eta}_t, \quad \mathbf{X}_t = \mathbf{C}\mathbf{f}_t + \boldsymbol{\varepsilon}_t, \quad t = 1, \ldots, T, \tag{1}$$

where $\mathbf{\Phi}(L) = \mathbf{I} + \mathbf{A}_1 L + \ldots + \mathbf{A}_k L^k$ is a matrix polynomial in the lag operation L, $\mathbf{f}_t$ is the latest state vector consisting of the factors $f_{t1}, \ldots, f_{tq}$ at time t and $\boldsymbol{\eta}_t$ are i.i.d. random errors with mean $\mathbf{0}$ and diagonal covariance matrix so that the first equation of (1) is the state equation and the second equation relates the state to the observation vector $\mathbf{X}_t \in \mathbb{R}^p$ and the i.i.d. random error $\boldsymbol{\varepsilon}_t$ with mean $\mathbf{0}$ and covariance matrix $\mathbf{\Sigma}$. The basic idea is that the number q of factors is manageably small for time series modeling. However, because of the large number p of time series in $\mathbf{X}_t$, estimation of $\mathbf{C}$ by EM "was prohibitive in large system."

Stock and Watson (2011) describe factor estimation by cross-sectional weighted averaging of $\mathbf{X}_t$, using a $p \times r$ matrix $\mathbf{W}$ such that $\mathbf{W}^T\mathbf{W}/p = \mathbf{I}$ to weight $\mathbf{X}_t$ via $\mathbf{W}^T\mathbf{X}_t$, which has been proposed by Stock and Watson (2002a,b) and Bai and Ng (2002) and further developed by Forni et al. (2004,2005) and Bai and Ng (2006), as the second generation of DFMs. The key assumptions are $\lim_{p\to\infty} p^{-1}\mathbf{C}^T\mathbf{C} = \mathbf{D}$ for some positive definite $r \times r$ matrix and $\sup_{p\geq 1} \lambda_{\max}(\mathbf{\Sigma}) < \infty$. One such estimator of $\mathbf{f}_t$ is the principal components estimator that chooses $\mathbf{W}$ to be the matrix of unit-length eigenvectors associated with the r largest eigenvalues of the sample covariance matrix $T^{-1}\Sigma_{t=1}^{T}\mathbf{X}_t\mathbf{X}^T$. The third generation of DFMs "merges the statistical efficiency of the state-space approach with the robustness and convenience of the principal components approach" in the works of Giannone et al. (2008) and Reiss and Watson (2010); Section 2.3 of Stock and Watson (2011).

The preceding discussion has focused on estimation of the factors when the number r of factors is known. Moreover, we have assumed for simplicity that all factors are dynamic. Section 3 of Stock and Watson (2011) reviews developements in the problem of determining the number of factors and allows in addition the inclusion of static factors whose number s is also to be determined. It describes the frequency-domain procedure of

Hallin and Liska (2007) for estimating r and the information criterion used by Bai and Ng (2002) and Amengual and Watson (2007). It also describes the use of scree plots in principal component analysis and random matrix theory for the asymptotic distribution theory of scree plots as $T \to \infty$ and $p \to \infty$, developed by Onatski (2008,2009) for visual diagnostic and formal tests concerning s. Lai and Tsang (2016) have recently developed an alternative approach to determination of the number of factors, which will be discussed in Section 4.

4. Targeted Predictors and Reduced Rank Regression

Section 4.1 of Stock and Watson (2011) reviews the application of dynamic factors to “second-stage regression" for multi-step ahead forecasts of key macroeconomic variables for the development and evaluation of monetary and other economic policies. Because multi-step ahead forecasts are involved in the analysis and prediction of the effect of a policy innovation on the economy, vector autoregression (VAR) has been widely used to provide such forecasting models since the seminal works of Bernanke and Blinder (1992) and Sims (1992). However, Bernanke, Boivin and Eliasz (2005) have pointed out that “to converse degrees of freedom, standard VARs rarely employ more than siz to eight variables," leading to at least three problems with the forecasts thus constructed. The first is the “price puzzle" that predicts an increase (instead of the anticipated decrease) in price level following a contradictionary monetary policy shock. The second is that “it requires taking a stand on specific observable measures" in the sparse information sets in VAR analysis (e.g. industrial production or real GPD as representative measure of economic activity). The third is that “impulse responses can only be observed for the included variables, which generally constitute only a small subset of variables that the researcher and policy-makers care about" but does not consider how the effect of policy shocks on other variables may in turn impact on the included variables. They therefore introduce a factor-augmented vector autoregression (FAVAR) approach that “combines the standard VAR analysis with factor analysis," for which the methods of Stock and Watson (2002a) can be used. The factor or principal component analysis, which basically relates to the covariance matrix of the vector $\mathbf{x}_t$ of macroeconomic indicators, in the FAVAR approach is used to determine predictors that augment the vector $\mathbf{y}_t$ of basic economic variables in the VAR analysis. Since the goal is to forecast future values of $\mathbf{y}_s$, we have recently developed a more efficient alternative to factor analysis by using reduced rank regression and pathwise variable selection methods to find the vector $\mathbf{f}_t$ of additional predictors. Pathwise variable selection methods such as LASSO and Elastic Net in statistical machine learning have been used by Bai and Ng (2008) to find “targeted predictors" before applying PCA to them to find more efficient factors for prediction. Instead of carrying out separately these machine learning methods that have been developed for i.i.d. samples and PCA for the selected variables, our alternative approach uses the high-dimensional multivariate stochastic regression modeling approach, developed by Lai and Tsang (2016), which consists of (a) orthogonal matching pursuit (which is a simplified version of forward stepwise inclusion of regressors) to exploit the “coefficient sparsity" (referring to mostly very small regression coefficients) and (b) reduced rank regression to exploit the inherent low-rank approximations (similar to the relatively small number of factors if PCA is used). To use (b) effectively in conjunction with (a), Lai and Tsang (2016) in fact apply a group orthogonal greedy algorithm (GOGA) similar to group LASSO to obtain the estimator $\widehat{\mathbf{B}}^{GOGA}$.

Specifically let $\mathbf{x}_t = (x_{t1}, \ldots, x_{tp})^T$ be a vector of observed variables (regressors) and $\mathbf{y}_t = (y_{t1}, \ldots, y_{tq})^T$ vector of basic variables whose multi-step ahead forecast are to be constructed. Consider the regression model

$$\mathbf{y}_t = \mathbf{B}^T \mathbf{x}_t + \varepsilon_t, \tag{2}$$

in which $\mathbf{x}^T$ includes 1 (for the intercept term) and $\mathbf{B} = (B_{ji})_{1\le j\le p, 1\le i\le q}$ is the coefficient matrix. It is assumed that $\mathbf{x}_t$ is $\mathcal{F}_{t-1}$-measurable and that the unobservable random errors ε_t form a martingale difference sequence with respect to $\{\mathcal{F}_t\}$ (hence "stochastic regression"), where $\mathcal{F}_s$ is the information set (σ-algebra) consisting of the observations up to time s. The regression model (2) can be used to carry out one-step ahead forecasts of $\mathbf{y}_t$. Macroeconomic forecasting typically involves not only much longer forecasting horizons but also a range of future times t for which the effect of a policy innovation on $\mathbf{y}_t$ has to be considered. This is the motivation underlying the widespread use of VAR and FAVAR models for macroeconomic forecasting. In particular, the FAVAR model of Bernanke et al. (2005) is a VAR model for the factor-augmented time series

$$\mathbf{A}(L)^T \mathbf{y}_t = \mathbf{B}(L)^T \mathbf{f}_{t-1} + \boldsymbol{\eta}_t, \tag{3}$$

$$\boldsymbol{\Phi}(L)^T \mathbf{f}_t = \boldsymbol{\Psi}(L)^T \mathbf{y}_{t-1} + \mathbf{w}_t, \tag{4}$$

where $\mathbf{A}(L) = \mathbf{I} - \sum_{s=1}^{d_1} \mathbf{A}_s L^s$, $\mathbf{B}(L) = \sum_{s=0}^{d_2} \mathbf{B}_s L^s$, $\boldsymbol{\Phi}(L) = \mathbf{I} - \sum_{s=1}^{d_3} \boldsymbol{\Phi}_s L^s$, and $\boldsymbol{\Psi}(L) = \sum_{s=0}^{d_4} \boldsymbol{\Psi}_s L^s$. The major difference of our approach from that of Bernanke et al. (2005) is that we use targeted predictors instead of factors to define $\mathbf{f}_t$. Specifically, define the $\hat{r} \times 1$ vector $\mathbf{f}_t$ by $\mathbf{f}_t^T = \mathbf{x}_t^T \widehat{\mathbf{B}}^{GOGA}(\mathbf{v}_1, \ldots, \mathbf{v}_{\hat{r}})$ for $t = 1, \ldots, T$, where $\mathbf{v}_1, \ldots, \mathbf{v}_{\hat{r}}$ are the $\hat{r}$ right singular vectors of $\mathbf{X}\widehat{\mathbf{B}}^{GOGA}$, and use it as the "targeted predictor" in lieu of the original high-dimensional vector $\mathbf{x}_t$ that covaries with the basic variables $\mathbf{y}_s$ whose future joint distribution is to be predicted.

To fit the multivariate regression model (2) under coefficient and rank sparsity, Lai and Tsang (2016) use the following variable and rank selection procedure along the GOGA path, which can be described as follows, letting $\mathbf{X} = (\mathbf{x}_1, \cdots, \mathbf{x}_T)^T$, $\mathbf{Y} = (\mathbf{y}_1, \cdots, \mathbf{y}_T)^T$, $\mathbf{X}_j$ be the jth column of $\mathbf{X}$, $\mathbf{X}_J$ be submatrix of $\mathbf{X}$ consisting of columns $\{\mathbf{X}_j : j \in J \subset \{1, \cdots, p\}\}$, and $\|\cdot\|_F$ be the squared Frobenius norm of a matrix.

1. Initialize with $\mathbf{U}^0 = \mathbf{Y}$, $\widehat{I}_0 = \emptyset$ and empty matrices $\mathbf{Q}_0$ and $\mathbf{R}_0$.
   For $k = 1$ to m do:
2. Choose $\hat{i}_k = \arg\min_{1\le i\le p} \left(\min_{\beta\in\mathbb{R}^q} ||\mathbf{U}^{k-1} - \mathbf{X}_i \boldsymbol{\beta}^T||_F^2\right)$
3. Update $\widehat{I}_k = \widehat{I}_{k-1} \cup \{\hat{i}_k\}$ and compute the QR decomposition

$$\mathbf{X}_{\widehat{I}_k} = [\mathbf{X}_{\widehat{I}_{k-1}} \mathbf{X}_{\hat{i}_k}] = \left[\begin{array}{cc} \mathbf{Q}_{k-1} & \mathbf{q}_k \end{array}\right] \left[\begin{array}{cc} \mathbf{R}_{k-1} & \vdots \\ 0\cdots 0 & r_k \end{array}\right] = \mathbf{Q}_k \mathbf{R}_k$$

4. Update $\mathbf{U}^k = \mathbf{U}^{k-1} - \mathbf{q}_k \boldsymbol{\beta}_k^T$, where $\boldsymbol{\beta}_k^T = \mathbf{q}_k^T \mathbf{U}^{k-1}$.
5. End for, with $\hat{i}_k$th row of $\hat{\mathbf{B}} \in \mathbb{R}^{p\times q}$ equal to the kth row of $\mathbf{R}_m^{-1}[\boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_m]^T$ and the other rows equal to $\mathbf{0}^T$.

Let $K_T = O(\{T/\log(p_T q_T)\}^{1/2})$ be a prescribed upper bound on the number of GOGA iterations. Here subscripts of p and q are introduced to denote the dependence on series length T. A "high-dimensional information criterion" is used to choose the number of input variables along the GOGA path:

$$\text{HDIC}(k) = T \log((T q_T)^{-1} \|\mathbf{U}^k\|_F^2) + k w_T \log(p_T q_T).$$

In particular, for $w_T = \log T$, $q_T \mathrm{HDIC}(k)$ corresponds to HDBIC. For $w_T = c$, $q_T \mathrm{HDIC}(k)$ corresponds to HDAIC. After obtaining the selected subset of $\hat{k}_T$ predictors, a BIC-type criterion

$$\mathrm{IC}(r) = Tq \log \hat{\sigma}^2(r) + rc(T+q) \log(\frac{Tq}{T+q})$$

is used to choose the rank of $\mathbf{B}$, where $\hat{\sigma}^2(r) = \|\mathbf{Y} - \mathbf{X}_S \mathbf{B}(r)\|_F^2/(Tq)$. Lai and Tsang (2016) have shown that $\mathrm{IC}(r)$ gives consistent rank estimation for any choice of c under rank sparsity.

Besides choice of $\hat{r}$ for reduced rank regression, we also have to choose d_1, d_2, d_3, d_4 for the lag polynomials in (3) and (4). Instead of an information (or entropy-based) criterion, we use a prediction-based criterion, which aims at getting optimal multi-step ahead forecasting performance. Since we are making forecasts of future $\mathbf{y}_t$, performance involves functionals of the joint distribution of $\mathbf{y}_t$. Choosing the functional appropriately can be circumvent the price puzzle mentioned earlier in this section. Details are presented elsewhere.

5. Conclusion

Factor models have played a major role in econometric and financial modeling. The factors provide major dimension-reduction building blocks for time series modeling and forecasting, particularly in the big data era that features a very large number of economic and financial time series. It is still an active area of research and we have described some of our ongoing work in this area.

REFERENCES

Akaike, H. (1969), “Fitting autoregressive models for prediction," *Ann. Inst. Statist.*, 21, 243–247.

Akaike, H. (1977), “On entropy maximisation principle," *Applications of Statistics, Ed. P. R. Krishntiah*, 27–41, Amsterdam: North Holland.

Amengual, D., and Watson, M. W. (2007), “Consistent Estimation of the Number of Dynamic Factors in a Large N and T Panel," *Journal of Business and Economic Statistics*, 25, 91–96.

Bai, J. (2003), “Inferential theory for factor models of large dimensions," *Econometrica*, 71, 135–172.

Bai, J., and Ng, S. (2002), “Determining the number of factors in approximate factor models," *Econometrica*, 70, 191–221.

Bai, J., and Ng, S. (2008), “Large dimensional factor analysis," *Foundations and Trends in Econometrics*,3(2), 89–163.

Bernanke, B., and Blinder, A. (1992), “The federal funds rate and the channels of monetary transmission," *Amer. Econ. Rev.*, 82, 901–921.

Bernanke, B. S., Boivin J., and Eliasz, P. (2005), “Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach," *Quart. J. Econ.*, 120, 387–422.

Boivin, J., and Ng, S. (2006), “Are more data always better for factor analysis," *Journal of Econometrics*, 132, 169–194.

Box, G. E. P. (1976), “Science and statistics," *J. Amer. Statist. Assoc.*, 71, 791–799.

Burns, A. F., and Mitchell, W. C. (1946), *Measuring Business Cycles*, New York, New York: National Bureau of Economic Research.

Chamberlain, G., and Rothschild, M. (1983), “Arbitrage Factor Structure, and MeanVariance Analysis of Large Asset Markets," *Econometrica*, 51, 1281–1304.

Christiano L., Eichenbaum M., and Evans, C. (2000), “Monetary policy shocks: what have we learned and to what end?," *J. Taylor and M.Woodford, eds., Handbook of Macroeconomics*, Amsterdam: North-Holland 2000.

Connor, G., and Korajzcyk, R. (1986), “Performance measurement with the arbitrage pricing theory: a new framework for analysis," *J. Finan. Econ.*, 15, 373–394.

Connor, G., and Korajzcyk, R. (1998). Risk and return in an equilibrium apt application of a new test methodology. *J. Finan. Econ.* **21**, 225-289.

Engle, R. F., and Watson, M. W. (1981), “A one-factor multivariate time series model of metropolitan wage rates," *J. Amer. Statist. Assoc.*, 76, 774–781.

Engle, R. F., and Watson, M. W. (1983), “Alternative algorithms for estimation of dynamic mimic, factor, and time varying coefficient regression models," *J. Econometrics*, 23, 385–400.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2004), “The Generalized Factor Model: Consistency and Rates," *Journal of Econometrics*, 119, 231–255.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2005), “The generalized dynamic factor model: one-sided estimation and forecasting," *J. Amer. Statist. Assoc.*, 100, 830–839.

Geweke, J. (1977), The Dynamic Factor Analysis of Economic Time-Series Models, *Latent Variables in Socio-Economic Models*, North-Holland, New York.

Giannone, D., Reichlin, L., and Small, D. (2008), “Nowcasting: The Real-Time Informational Content of Macroeconomic Data," *Journal of Monetary Economics*, 55, 665–676.

Hallin, M., and Liska, R. (2007), “The Generalized Dynamic Factor Model: Determining the Number of Factors," *Journal of the American Statistical Association*, 102, 603–617.

Hannan, E. J. (1987), “Rational transfer function approximation," *Stat. Sc.*, 2, 135–161.

Hannan, E. J., and Quinn, B. G. (1979), “The determination of the order of an autoregression," *J. Roy. Statist. Soc. B*, 41, 190–195.

Hannan, E. J., and Rissanen, J. (1982), “Recursive estimation of ARMA order," *Biometrika*, 69, 81–94.

Huang, D., and Guo, L. (1990), “Estimation of nonstationary ARMAX models based on the Hannan-Rissanen methods," *Ann. Statist.*, 18, 1729–1756.

Ing, C. K., and Lai, T. L. (2011), “A stepwise regression method and consistent model selection for high-dimensional sparse linear models," *Statist. Sinica*, 21, 1473–1513.

Ing, C. K., and Lai, T. L. (2014), “An effcient pathwise variable selection criterion in weakly sparse ARX models," *Tech. Report*, Dept. Statistics, Stanford Univ.

Klein, L. R. (1950), “Economic fluctuations in the United States, 1921-1941," *Cowless Commission for Research in Economics*, New Haven.

Lai, T. L., and Tsang, K. W. (2015), “Multivariate stochastic regression in time series modeling," *Statist. Sinica*, DOI: 10.5705/ss.2014.211t.

Lai, T. L., and Xing, H. (2008), *Statistical Models and Methods for Financial Markets*, Springer Texts in Statistics. Springer, New York.

Lai, T. L., Xing, H., and Chen, Z. (2011), “Mean-Variance Portfolio Optimization when Means and Covariances are Unknown." *Ann. Appl. Statist.*, 5, 798–823.

Liu, X., and Chen, R. (2015), “Regime-switching factor models for high-dimensional time series," *Statist. Sinica*, DOI: 10.5705/ss.2014.265t.

Onatski, A. (2008), “The Tracy-Widom Limit for the Largest Eigenvalues of Singular Complex Wishart Matrices," *Annals of Applied Probability*, 18, 470–490.

Onatski, A. (2008), “Testing Hypotheses about the Number of Factors in Large Factor Models," *Econometrica*, 77, 1447–1479.

Pẽna, D., and Box, G. E. P. (1987), “Identifying a simplifying structure in time series," *J. Amer. Statist. Assoc.*, 82, 836–843.

Pẽna, D., and Poncela, P. (2006), “Nonstationary dynamic factor analysis," *J. Stat. Plan. Inference*, 136, 1237–1257.

Pẽna, D., and Yohai, V. J. (2016), “Generalized dynamic principal components," *J. Amer. Statist. Assoc.*, DOI: 10.1080/01621459.2015.1072542.

Quah, D., and Sargent, T. J. (1993), “A Dynamic Index Model for Large Cross Sections," *Business Cycles, Indicators, and Forecasting, ed. by J.H. Stock and M.W. Watson, Chicago: University of Chicago Press for the NBER*, 285–310.

Reiss, R., and Watson, M. W. (2010), “Relative Goods' Prices and Pure Inflation," *American Economic Journal: Macroeconomics*, 2, 128–157.

Rissanen, J. (1986), “Stochastic complexity and modeling," *Ann. Statist.*, 14, 1080–1100.

Ross, S. (1976), “The Arbitrage Theory of Capital Asset Pricing." *Journal of Economic Theory*, 13, 341–360.

Sargent, T. J. (1989), “Two models of measurements and the investment accelerator," *J. Polit. Economy*, 97, 251–287.

Sargent, T. J., and Sims, C. A. (1977), “Business cycle modeling without pretending to have too much a-priori economic theory," *New Methods in Business Cycle Research, ed. by C. Sims et al.*, Minneapolis: Federal Reserve Bank of Minneapolis.

Schwarz, G. (1978), “Estimating the dimension of a model," *Ann. Statist.*, 6, 461–464.

Sharpe, W. F. (1964), “Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk," *Journal of Finance*, 19, 425–442.

Sims, C. A. (1980), “Macroeconomics and Reality," *Econometrica*, 48(1), 1–48.

Sims, C. A. (1992), “Interpreting the macroeconomic time series facts: the effects of monetary policy," *Europ. Econ. Rev.*, 36, 975–1000.

Stock, J. H., and Watson, M. W. (2002), “Forecasting using principal components from a large number of predictors," *J. Amer. Statist. Assoc.*, 97, 1167–1179.

Stock, J. H., and Watson, M. W. (2002), “Macroeconomic Forecasting Using Diffusion Indexes," *Journal of Business and Economic Statistics*, 20, 147–162.

Stock, J. H. and Watson, M. W. (2004), “An empirical comparison of methods for forecasting using many predictors", Mimeo, Princeton University.

Stock, J. H., and Watson, M. W. (2011), “Dynamic Factor Models," *Oxford Handbook on Economic Forecasting*. Oxford: Oxford University Press.

Tsai, H., and Tsay, R. S. (2010), “Constrained factor models," *J. Amer. Statist. Assoc.*, 105, 1593–1605.

Wei, C. Z. (1992), “On predictive least squares principles," *Ann. Statist.*, 20, 1–42.

Xia, Y., and Tong, H. (2011), “Feature matching in time series modeling (with discussion)," *Stat. Sc.*, 26, 21–46.

Yuan, M., and Lin, Y. (2006), “Model selection and estimation in regression with grouped variables," *J. Roy. Statist. Soc. B*, 68, 49–67.