

MANFRED DEISTLER
AND THE GENERAL DYNAMIC FACTOR MODEL APPROACH
TO THE ANALYSIS OF HIGH-DIMENSIONAL TIME SERIES

Marc Hallin
ECARES and Department of Mathematics
Université libre de Bruxelles

Abstract

For more than half a century, Manfred Deistler has been contributing to the construction of the rigorous theoretical foundations of the statistical analysis of time series and more general stochastic processes. Half a century of unremitting activity is not easily summarized in a few pages. In this short note, we chose to concentrate on a relatively little-known aspect of Manfred's contribution which nevertheless had quite an impact on the development of one of the most powerful tools of contemporary time series and econometrics: dynamic factor models.

Keywords: High-dimensional time series; General Dynamic Factor Models; spiked covariance model; reduced-rank process.

JEL subject classification : C0, C01, E0.

1 Introduction

Manfred Deistler is justly famous for his landmark contribution to the theoretical foundations of time series analysis. In this short note, however, we deliberately chose to focus on a lesser known aspect of his activity, which nevertheless had quite an impact on the theory and practice of one of the most powerful tools of contemporary time series and econometrics: dynamic factor models.

Dynamic factor models¹ were developed, in the econometric literature mostly, as a response the need to analyze and forecast time series in high dimension. Increasingly often, indeed, datasets of econometric interest take the form

$$\mathbf{X}_{N,T} := \{X_{it} | i = 1, \dots, N \ t = 1, \dots, T\}$$

of a large number N of time series observed over a period of time T —the finite $N \times T$ realization of double-indexed process

$$\mathcal{X} := \{X_{it} | i \in \mathbb{N} \ t \in \mathbb{Z}\}$$

¹As we shall see in Section 3 below, this traditional terminology is somewhat of a misnomer, as the *general dynamic factor model* (6)–(7) (of which other factor models are particular cases) follows as a representation result rather than constituting a statistical *model*, and does not necessarily involve *factors*; see Hallin and Lippi (2014).

with arbitrarily intricate cross-sectional and serial dependence structures.

Even for moderate values of N , traditional (parametric) methods of multivariate time series analysis are running into the theoretical and numerical problems related with the curse of dimensionality. The need for an alternative approach became evident in the late 1970s, leading to the first factor model proposals by Sargent and Sims (1977), Geweke (1977), Chamberlain (1983), and Chamberlain and Rothschild (1983). These four papers can be considered as early forerunners of the modern literature on factor models—a literature that had a new start in the early 2000's, with four papers, essentially, that triggered most subsequent developments: Forni et al. (2000), Bai and Ng (2002), Stock and Watson (2002a), and Stock and Watson (2002b). Chamberlain (1983) and Chamberlain and Rothschild (1983) were particularly influential as an early example of high-dimensional time-series asymptotics where both the dimension N and the series length T tend to infinity.

Econometricians, of course, were not the only ones facing inference problems in high-dimensional spaces. Interestingly, a couple of years later, mathematical statisticians, in the more restricted context of Gaussian i.i.d. observations, (a very particular case of time series, thus) independently adopted a somewhat different approach leading to the so-called *spiked covariance model*. We are showing here how spiked covariance models and factor models, while sharing some common features, nevertheless differ on an essential point, and explain why factor models are both more general and statistically more successful.

Finally, we show how Manfred Deistler, by providing the missing final piece of the general dynamic factor model jigsaw, had a decisive impact in the area that deserves being better known.

Outline of the paper. Section 2 is dealing with the spiked covariance model developed in the probability and mathematical statistics literatures. This model, indeed, bears some relations to the factor model approach, with an essential difference which helps understand the benefits of the latter. Section 3 features a brief history of the factor model approach and introduces the General Dynamic Factor Model (GDFM). Section 4 highlights the importance of Manfred Deistler's contribution in the ultimate development of the GDFM methodology.

2 Spiked covariance models: a needle in a growing haystack

While econometricians were facing the time series version of high-dimensional observations and the curse of dimensionality, mathematical statisticians also were dealing with high-dimensional asymptotics in the more restricted framework of i.i.d. samples where only cross-sectional dependencies are present. Interestingly, while sharing some common features with factor models, the models they developed are leading to strikingly different conclusions.

Most of the probability and statistics literature in the area revolves around the so-called *spiked covariance models*—a terminology that was coined, apparently, by Johnstone (2001). In that model, which has attracted much interest in the recent years, the observation is of the form

$$\mathbf{X}_{N,T} := \{X_{it} | i = 1, \dots, N \ t = 1, \dots, T\} \quad (1)$$

where

$$\mathbf{X}_t := (X_{1t}, \dots, X_{Nt})', \ t = 1, \dots, T \text{ are i.i.d. } \mathcal{N}(\mathbf{0}, \mathbf{C}), \quad (2)$$

with, denoting by \mathcal{S}_{N-1} the unit sphere in \mathbb{R}^N ,

$$\mathbf{C}_N = \mathbf{I}_N + \sum_{k=1}^q \lambda_k \mathbf{v}_k^{(n)} \mathbf{v}_k^{(n)'} , \quad q \geq 1, \quad \lambda_1 > \lambda_2 > \dots > \lambda_q > 0, \quad \mathbf{v}_k^{(n)} \in \mathcal{S}_{N-1}, k = 1, \dots, q \text{ unspecified} \quad (3)$$

and N and T tending to infinity with $N/T \rightarrow \kappa$, the so-called *phase transition threshold*. To simplify the discussion, let $q = 1$, that is, $\mathbf{C}_N = \mathbf{I}_N + \lambda \mathbf{v}_N \mathbf{v}_N'$ for some $\mathbf{v}_N \in \mathcal{S}_N$.

That model leads to a number of mathematically beautiful but statistically puzzling asymptotic results: the sample covariance eigenvalues pack together, filling the support of the *Marchenko-Pastur density*; the distribution of any finite number of centered and normalized largest sample covariance eigenvalues converges to the multivariate *Tracy-Widom law* irrespective of the values of $(1 + \lambda)$ in $[1, \kappa)$; the sequence of distributions of \mathbf{X}_t under $\lambda = 0$ (no spike) is contiguous to the corresponding sequence with $0 < \lambda < \kappa$, albeit with contiguity rate $n^0 = 1$ —which, in particular, precludes consistent estimation of λ (see Onatski et al. (2013, 2014) for details). The statistical value of such results is, to say the least, somewhat limited—all the more so that in practice $N = N_0$ and $T = T_0$ are not tending to infinity, so that the value of κ , the role of which is crucial, is completely arbitrary and bears no relation to the observed sample.² That spiked covariance literature, thus, has little to offer to econometricians who have to produce forecasts and, moreover, are facing serially dependent and mostly non-Gaussian observations.

These (intriguing but weird) results all are to be blamed on the choice of the asymptotic scheme itself. Recall that asymptotics are a mathematical fiction by which limiting results (as N and T tend to infinity) are expected to provide an approximation to the actual fixed- $(N = N_0)$, fixed- $(T = T_0)$ problem. The scenario of that mathematical fiction is not in the data: it entirely depends on the statistician's choice. That choice, thus, should aim at optimizing the quality of that approximation and is not meant to describe any actual real-world situation: a scenario under which the “cross-sectional future” resembles the actual observation is likely to achieve that objective much better than a “worst-case one”. In traditional time series asymptotics with fixed-dimension $(N = N_0)$, where only $T \rightarrow \infty$, stationarity with respect to t is the reasonable usual choice. While the specification, for $N > N_0$ of a fictitious yet sensible “cross-sectional future” is more delicate, the choice leading to the spiked covariance model (2)-(3) definitely has the flavor of a “catastrophe scenario” which is unlikely to provide a good approximation to the finite-dimensional, finite-sample situation.

Below are two types of data-generating processes (two sequences of unit vectors \mathbf{v}_N) leading to the single-spiked-covariance model—the $N \times N$ covariance matrix (3):

$$(a) \quad \begin{cases} X_{1t} &= \chi_{1t} + \xi_{1t} \\ X_{it} &= \xi_{it}, \quad i = 2, \dots, N \end{cases} \quad \mathbf{v}_N = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad \mathbf{C}_N = \mathbf{I} + \lambda \begin{pmatrix} 1 & 0 \dots 0 \\ 0 & 0 \dots 0 \\ \vdots & \ddots \\ 0 & 0 \dots 0 \end{pmatrix},$$

where χ_{1t} i.i.d. $\mathcal{N}(0, \lambda)$, and ξ_{it} i.i.d. $\mathcal{N}(0, 1)$, $i = 1 \dots, N$, $t = 1, \dots, T$;

²The value of the actual ratio N_0/T_0 is usually chosen for want of anything better.

$$(b) \begin{cases} X_{1t} &= \chi_{1t} + \xi_{1t} \\ X_{it} &= \chi_{it} + \xi_{it}, \quad i = 2, \dots, N \end{cases} \quad \mathbf{v}_N = \begin{pmatrix} \frac{1}{\sqrt{N}} \\ \frac{1}{\sqrt{N}} \\ \vdots \\ \frac{1}{\sqrt{N}} \end{pmatrix} \quad \mathbf{C}_N = \mathbf{I} + \lambda \begin{pmatrix} \frac{1}{N} & \cdots & \frac{1}{N} \\ \vdots & \ddots & \vdots \\ \frac{1}{N} & \cdots & \frac{1}{N} \end{pmatrix}$$

where $\chi_{it} = \sqrt{\frac{\lambda}{N}} u_t$, u_t and ξ_{it} i.i.d. $\mathcal{N}(0, 1)$, $i = 1, \dots, N$, $t = 1, \dots, T$ (for $\mathbf{v}_N = (\frac{1}{\sqrt{N}} \dots \frac{1}{\sqrt{N}})'$).

Actually, (a) and (b) coincide up to a rotation in \mathbb{R}^N (an orthogonal matrix \mathbf{P} with first column of the form $(\frac{1}{\sqrt{N}} \dots \frac{1}{\sqrt{N}})'$ is turning (a) into (b)): they only differ by the choice of a coordinate system and all their eigenvalues (theoretical and empirical) coincide.

Under (a), a bounded *spike* λ (justifying the *spiked* terminology) is “hidden” under a growing number $N-1$ of uninformative white noises—a finite needle buried in an ever-growing haystack. Growing N clearly does not bring any information (only growing T does). The fact that the needle gets undetected when its size λ is small relative to the asymptotic value κ of the ratio N/T (the larger that ratio, the faster the haystack growth) thus is hardly surprising.

Model (b) takes the form of a factor model decomposition, with a cross-sectionally pervasive “common shock” u_t loaded (with loadings $\sqrt{\lambda/N}$ tending to zero as $N \rightarrow \infty$) by all components X_{it} and an idiosyncratic ξ_{it} which is Gaussian white noise. While cross-sectionally pervasive, however, u_t is not loaded strongly enough for the largest eigenvalue of \mathbf{C}_N , which is $1 + \lambda$, to diverge as $N \rightarrow \infty$.

The situation, however, improves dramatically if the size of the needle grows with the dimension N . Letting χ_{it} of the form $\chi_{it} = \sqrt{\lambda/N^{1-\delta}} u_t$ in (b), with $\delta \in (0, 1)$ arbitrarily small (loadings $\sqrt{\lambda/N^{1-\delta}}$ still tending to zero as $N \rightarrow \infty$, at slightly slower rate, though), \mathbf{C}_N ’s largest eigenvalue is $1 + N^\delta \lambda$, which tends to infinity as $N \rightarrow \infty$. With χ_{it} of the form $\chi_{it} = \sqrt{c\lambda} u_t$ ($c > 0$ arbitrarily small), the loadings are $\sqrt{c\lambda}$ and no longer tend to zero as $N \rightarrow \infty$; \mathbf{C}_N ’s largest eigenvalue is $1 + c\lambda N$, which linearly tends to infinity as $N \rightarrow \infty$.

All problems then disappear: with such loadings, (b), indeed, yields a (very) special case of the dynamic factor models developed by the econometricians and described in Section 3, where consistent estimation of the loadings—hence of λ —is possible.

Now, the same conclusions hold about (b) if we let

$$\chi_{it} = \sqrt{\frac{\lambda}{N}} u_t \text{ for } i = 1, \dots, N \text{ and } \chi_{it} = \sqrt{\lambda/N^{1-\delta}} u_t \text{ for } i \geq N+1 \quad (4)$$

or

$$\chi_{it} = \sqrt{c\lambda} u_t \text{ for } i = 1, \dots, N \text{ and } \chi_{it} = \sqrt{\lambda/N^{1-\delta}} u_t \text{ for } i \geq N+1. \quad (5)$$

Assuming (b) with χ_{it} of the form (4) or (5) instead of the original formulation (2)-(3), clearly, is tantamount to adopting alternative asymptotic scenarios, as it only modifies the postulated but not-to-be-observed form of the “cross-sectional future” (namely, the form of the cross-sectional components with index $i \geq N+1$), which is an arbitrary choice of the statistician. Under these alternative scenarios, the spike $\lambda = \lambda_N$, say, keeps growing with N (viz., $\lambda_N = \lambda N^\delta$ under (4), $\lambda_N = c\lambda N$ under (5)), thus balancing the impact of a growing dimension N : the needle now is growing—be it arbitrarily slowly—with the haystack.

The asymptotic scenario (5) and, actually, any scenario of loadings leading to a linearly exploding

largest eigenvalue for \mathbf{C}_N (all other eigenvalues remaining bounded) is particularly appealing, as it consists in assuming that the fictitious “cross-sectional future” resembles the observed one—a form of cross-sectional stability which is much more likely to provide a good approximation to the finite- (N, T) problem under study than the classical spiked-covariance-model asymptotics underlying (2)-(3).

3 Dynamic Factor Models: the blessing of dimensionality

As mentioned in the Introduction, econometricians faced the needle-in-the haystack problem long time before probabilists and mathematical statisticians did—long time before the expression “big data” was coined. Econometrics require operational solutions: the asymptotic scenario of a needle that, at the end of the day, cannot be found is somewhat inappropriate in the econometric context. Moreover, economic data seldom are i.i.d. Gaussian; they generally are serially auto- and cross-correlated, and often heavy-tailed (typically yielding infinite fourth-order moments). An econometric theory of high-dimensional time series therefore needs to address much more general situations than covered by Gaussian spiked covariance models.

The econometric theory of factor models for high-dimensional observations arose from that need and takes various forms; the spirit of the approach, as initiated by Chamberlain (1983) and Chamberlain and Rothschild (1983), is easily explained, however, from a multispiked (q spikes) extension of (b) yielding q exploding eigenvalues $\lambda_{1;N}, \dots, \lambda_{q;N}$ —namely, q needles which, in a sense, are growing with the haystack.

Under their most general form (the General Dynamic Factor Model³ or GDFM introduced in Forni et al. (2000)), factor models proceed as follows. The observation is still of the form $\mathbf{X}_{N,T}$ described in (1), but the assumptions are much more general, allowing for serial auto- and cross-correlations and non-Gaussian densities. Assuming that they exist, denote by $\Sigma_N(\theta)$, $\theta \in [0, \pi]$ the $N \times N$ spectral density matrices of the N -dimensional processes $\{\mathbf{X}_{N,t} := (X_{1t}, \dots, X_{Nt})' | t \in \mathbb{Z}\}$. These density matrices are nested as $N \rightarrow \infty$. Some eigenvalues— q of them, say—are exploding ($\theta - a.e.$) and the other ones remain bounded ($\theta - a.e.$) as $N \rightarrow \infty$. This yields a decomposition of $\Sigma_N(\theta)$ into

$$\Sigma_N(\theta) = \Sigma_N^X(\theta) + \Sigma_N^\xi(\theta), \quad \theta \in [0, \pi]$$

where $\Sigma_N^X(\theta)$, called the *common spectral density* has reduced rank q and q diverging dynamic eigenvalues,⁴ while $\Sigma_N^\xi(\theta)$, called the *idiosyncratic spectral density* only has bounded dynamic eigenvalues.

That decomposition in turn induces a decomposition of the observations X_{it} into

$$X_{it} = \chi_{it} + \xi_{it} \quad i = 1, \dots, N, \quad t = 1, \dots, T, \quad N \in \mathbb{N}, \quad T \in \mathbb{Z} \quad (6)$$

where χ_{it} , with spectral density $\Sigma_N^X(\theta)$, and ξ_{it} , with spectral density $\Sigma_N^\xi(\theta)$, are mutually orthogonal at all leads and lags; χ_{it} is called the *common component*, ξ_{it} the *idiosyncratic component*. Since the spectral density $\Sigma_N^X(\theta)$ has reduced rank q , the common component χ_{it} is driven by $q \ll n$ mutually orthogonal white noises:

$$\chi_{it} = B_i(L)\mathbf{u}_t, \quad \mathbf{u}_t = (u_{1t}, \dots, u_{qt})' \quad (7)$$

³The terminology Generalized Dynamic Factor Model is used equivalently.

⁴The terminology *dynamic eigenvalues* used for the eigenvalues of a spectral density matrix was coined by Brillinger (1964, 1981), who introduced the concept.

(L , as usual, stands for the lag operator), while the idiosyncratic component ξ_{it} , having bounded dynamic eigenvalues, is only mildly cross-correlated (it can be strongly auto-correlated, though); see Hallin and Lippi (2019) or Lippi et al. (2022) for recent surveys.

Further constraints can be imposed on the decomposition. Among them, the static loading assumption (Stock and Watson (2002a, b), Bai and Ng (2002), Bai (2003), and many others)

$$\chi_{it} = \mathbf{B}_i' \mathbf{u}_t, \quad \mathbf{u}_t = (u_{1t}, \dots, u_{qt})' \quad (8)$$

(\mathbf{B}_i a $q \times 1$ real vector) under which the shocks \mathbf{u}_t are loaded in a *static*, contemporaneous way while in (7) the loadings are filters $B_i(L)$.

As soon as the spectral density $\Sigma_N(\theta)$ exists and admits a finite number q of exploding eigenvalues,⁵ a GDFM representation (6) with dynamic loadings (7) exists. The existence of a factor model decomposition (6) with static loadings (8), however, is a strong assumption one should like to avoid. Multivariate economic time series, let alone the infinite-dimensional ones, indeed, typically involve leading and lagging series, loading the common shocks with various leads and lags. The GDFM, which allows for this and basically does not place any restrictions on the data-generating process, is much preferable in that respect.

The GDFM was introduced by Forni et al. (2000), who establish its asymptotic identifiability—the “blessing of dimensionality”—and propose a consistent⁶ estimation strategy based on Brillinger’s concept of *dynamic principal components*. The moot point with dynamic principal components is that their computation involves two-sided filters, i.e., implicates both the past and the future values of the observed X_{it} ’s. This is fine in the “center” of the observation period but not at the edges. In particular, the estimation, for forecasting purposes, of \mathbf{u}_T , is likely to be poor irrespective of N and T , as the future observations $\mathbf{X}_{N,T+1}, \mathbf{X}_{N,T+2}, \dots$ are not available.

The advantage of static loadings is that, under (8), q coincides with the number of exploding “static” eigenvalues—the eigenvalues of the $N \times N$ covariance matrix \mathbf{C}_N of $\mathbf{X}_{N,T}$. And Bai and Ng (2002) and Stock and Watson (2002a, b) propose an estimation method relying on a traditional principal component analysis of $\mathbf{X}_{N,T}$. These principal components, at time t , only require the contemporaneous observation $\mathbf{X}_{N,t} := (X_{1t}, \dots, X_{Nt})'$: no problems, thus, for $t = T$. This, and the popularity of traditional principal component methods, explains why practitioners regrettably prefer the static contemporaneous loading approach despite its lesser generality, its lack of parsimony, and the fact that the crucial and quite restrictive underlying assumption (8) may not hold.⁷

Up to this point, thus, the General Dynamic Factor Model, in view of its generality, is a brilliant idea the practical implementation of which is blocked, apparently, by the two-sidedness nature of Brillinger’s dynamic principal components. Here is, however, where Manfred Deistler enters into the picture.

⁵Although an infinite number of exploding eigenvalues in theory is not impossible, such cases are extremely artificial and contrived: see Hallin and Lippi (2014) for an example.

⁶Consistency here is not uniform in t , though: see Forni et al. (2000) for details.

⁷This latter fact is often dispelled by arguing that lagged values of the factors can be incorporated into a static loading scheme via stacking. This, which may very severely inflate the number of factors, is a flawed argument. Indeed, there is no guarantee that these lagged values enjoy the pervasiveness properties required from static factors, with the consequence that, in the traditional principal component estimation method, they get lost to the idiosyncratic.

4 Manfred Deistler and the General Dynamic Factor Model

Manfred Deistler's seminal contribution to the analysis of General Dynamic Factor Models originates in his interest in the properties of reduced rank processes. The following result—unrelated, at first sight, to factor models and high-dimensional time series—follows from Anderson and Deistler (2008a).

Anderson and Deistler call *tall* the transfer function $\mathbf{D}(L)$ of a Q -dimensional process $\{\mathbf{Y}_t | t \in \mathbb{Z}\}$ driven by a q -dimensional white noise $\{\mathbf{w}_t | t \in \mathbb{Z}\}$ where $q < Q$ —namely, a process of the form

$$\mathbf{Y}_t = \mathbf{D}(L)\mathbf{w}_t \quad t \in \mathbb{Z} \quad (9)$$

where $\mathbf{D}(L)$ is some $Q \times q$ filter. By abuse of language, we say that the process $\{\mathbf{Y}_t\}$ itself is tall.

Recall that a process satisfying (9) with $\mathbf{D}(L) = (D_{ij}(L))$ is called *rational* if the filter $\mathbf{D}(L)$ itself is rational, that is, if there exist $Q \times q$ matrix filters $(E_{ij}(L))$ and $(F_{ij}(L))$ and integers m and p such that, for all $i = 1, \dots, Q$ and $j = 1, \dots, q$, $F_{ij}(0) = 1$, the degree of $E_{ij}(L)$ is m , the degree of $F_{ij}(L)$ is p , and $D_{ij}(L) = E_{ij}(L)/F_{ij}(L)$; note that $\mathbf{D}(L)$ then involves a finite (although unspecified) number $P = Qq(m + p + 1)$ of real parameters: call it *rational of order P* . Also recall that a subset of a topological space is *generic* if it contains an open and dense subset. Denoting by Π_P the parameter space⁸ indexing the family of rational filters of order P , genericity below is meant for Π_P as a subset of \mathbb{R}^P : call it Π_P -genericity.

As in Forni et al. (2015), the way the fundamental result of Anderson and Deistler (2008a) is presented here is geared towards its General Dynamic Factor Model application and slightly differs from the original formulation. Rather than a rational spectrum of order P (P unspecified but finite), the latter, actually, assumes a state space representation with finitely many parameters. For the sake of simplicity, the formulation below also slightly differs from the statements in Sections 3 and 4 of Forni et al. (2015) where more general and complete results can be found.

Proposition. *Let $\{\mathbf{Y}_t | t \in \mathbb{Z}\}$ be a tall process satisfying (9) for some rational $Q \times q$ filter $\mathbf{D}(L)$. Then, for Π_P -generic values of $\mathbf{D}(L)$ (P unspecified), \mathbf{Y}_t admits, for some $K < \infty$, a VAR(K) representation of the form*

$$\mathbf{A}(L)\mathbf{Y}_t = \sum_{k=1}^K \mathbf{A}_k \mathbf{Y}_{t-k} = \mathbf{R}\mathbf{w}_t \quad (10)$$

where $\mathbf{A}(L)$ is $Q \times Q$ and \mathbf{R} is $Q \times q$.

Refinements of this result can be found in Anderson and Deistler (2009), Chen et al. (2011), and Anderson et al. (2012), where some of its consequences also are discussed. The most elegant proof is found in Anderson et al. (2016). Anderson and Deistler (2008a) only briefly mention, without entering into details, the relevance of their result for the General Dynamic Factor Model (with dynamic loadings as in (7)). That relevance is further discussed in Anderson and Deistler (2008b, 2009), Deistler et al. (2010), and Forni and Lippi (2011); it is fully exploited in Forni et al. (2015, 2017), Forni et al. (2018), Barigozzi et al. (2021, 2022), and several subsequent papers.

That relevance stems from the fact that, by definition, for almost all values of i_1, \dots, i_{q+1} , the $(q+1)$ -dimensional vector of common components $\mathbf{Y}_t := (\chi_{i_1,t}, \dots, \chi_{i_{q+1},t})'$ is tall, with $Q = q + 1$ and

⁸A complete description of Π_P would require the filters $F_{ij}(L)$ and $E_{ij}(L)$ to be *stable* and having no common zeroes.

a q -dimensional white noise $\mathbf{w}_t = \mathbf{u}_t$. Under the very mild additional assumption that $\{\mathbf{Y}_t\}$ is rational (for some unspecified m and p), thus, the above proposition applies.

Suppose, for convenience (and without loss of generality), that $N = M(q + 1)$, where $M \in \mathbb{N}$, and write $(\chi_{1,t}, \dots, \chi_{N,t})$ as $(\boldsymbol{\chi}_t^1, \dots, \boldsymbol{\chi}_t^M)'$ where $\boldsymbol{\chi}_t^k \in \mathbb{R}^{q+1}$, $k = 1, \dots, M$. Assuming the rationality of all $\boldsymbol{\chi}_t^k$'s, the Anderson-Deistler result applies, so that

$$\begin{pmatrix} \mathbf{A}^1(L) & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{A}^2(L) & \cdots & \mathbf{0} \\ & & \ddots & \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{A}^M(L) \end{pmatrix} \begin{pmatrix} \boldsymbol{\chi}_t^1 \\ \boldsymbol{\chi}_t^2 \\ \vdots \\ \boldsymbol{\chi}_t^M \end{pmatrix} = \begin{pmatrix} \mathbf{R}^1 \\ \mathbf{R}^2 \\ \vdots \\ \mathbf{R}^M \end{pmatrix} \mathbf{v}_t,$$

that is, *generically*,⁹ for some $\mathbf{A}_n(L)$, \mathbf{R}_n and \mathbf{v}_t , we have the block-diagonal VAR representation

$$\mathbf{A}_n(L)\mathbf{X}_{nt} = \mathbf{R}_n\mathbf{v}_t$$

with $(q + 1) \times (q + 1)$ blocks $\mathbf{A}^k(L)$, $k = 1, \dots, M$, an $N \times q$ matrix \mathbf{R}_n , and a q -dimensional white noise \mathbf{v}_t .¹⁰

As a consequence,

$$\mathbf{A}_n(L)\mathbf{X}_{nt} = \mathbf{A}_n(L)(\mathbf{X}_{nt} + \boldsymbol{\xi}_{nt}) = \mathbf{R}_n\mathbf{v}_t + \mathbf{A}_n(L)\boldsymbol{\xi}_{nt} \quad (11)$$

where, being a linear transformation of idiosyncratic components, $\mathbf{A}_n(L)\boldsymbol{\xi}_{nt}$ itself is idiosyncratic. This is a static factor model (static contemporaneous loadings) for the filtered series $\mathbf{A}_n(L)\mathbf{X}_{nt}$: traditional principal components thus, which do not involve filters, can be used instead of dynamic ones in the estimation of (11) once consistent estimators¹¹ $\hat{\mathbf{A}}^k(L)$ are substituted for the unspecified filters $\mathbf{A}^k(L)$.

Based on this, Forni and Lippi (2011) and Forni et al. (2015, 2017) propose a winning strategy for a consistent one-sided reconstruction of the common components χ_{it} (hence, also the idiosyncratic components ξ_{it}), their impulse response functions, the common shocks \mathbf{u}_t , the loadings filters $B_i(L)$, etc. The corresponding asymptotic distributions are derived in Barigozzi et al. (2022). Surprisingly, the consistency rates are comparable to those obtained by Bai (2003) for the static method (without the preliminary filtering (11)—the validity of which, however, requires the much more stringent assumptions of the static model (8). The same results also apply in the identification and estimation of volatilities (Barigozzi and Hallin (2016, 2017, 2019)), of time-varying GDFMs (Barigozzi et al. (2021)), and in the prediction of conditional variances, values at risk, and expected shortfalls (Hallin and Trucíos (2022), Trucíos et al. (2022)).

The (asymptotic) validity of this strategy, of course, requires (11) to hold. This has to be assumed. Since, however, it holds *generically*, that assumption is extremely mild.

Numerical exercises (both Monte-Carlo and empirical) demonstrate the forecasting superiority of the resulting method, which seems to outperform all other methods proposed in the literature while remaining valid under much milder and more general assumptions on the data-generating process. Forni

⁹Genericity here is meant for any fixed N .

¹⁰It can be shown, moreover, that $\mathbf{v}_t = \mathbf{O}\mathbf{u}_t$ for some $q \times q$ orthogonal matrix \mathbf{O} .

¹¹Such estimators can be constructed from the decomposition of the estimated spectral density followed by inverse Fourier transforms and Yule-Walker VAR estimation in dimension $q + 1$: see Forni et al. (2015, 2017) for details.

et al. (2018) show that, even when the assumptions (8) of static loadings are satisfied, the GDFM method still performs better than the static one. Barigozzi et al. (2022) finalize the study of its asymptotic properties by establishing the corresponding asymptotic distributional results.

Manfred Deistler, thus, can be credited for unlocking the applicability of the General Dynamic Factor Model with dynamic loadings instead of the more restrictive, less performant, and less parsimonious model with static loadings.

References

- [1] Anderson, B.D.O. and Deistler, M. (2008a). Properties of zero-free transfer function matrices, *SICE Journal of Control, Measurement and System Integration* **1**, 284–292.
- [2] Anderson, B.D.O. and Deistler, M. (2008b). Generalized linear dynamic factor models - a structure theory, *47th IEEE Conference on Decision and Control*, 1980–1985.
- [3] Anderson, B.D.O. and Deistler, M. (2009). Properties of zero-free spectral matrices, *IEEE Transactions on Automatic Control* **54**, 2365–2375.
- [4] Anderson, B.D.O., Deistler, M., Chen, W. and Filler, A. (2012). Autoregressive models of singular spectral matrices, *Automatica* **48**, 2843–2849.
- [5] Anderson, B.D.O., Deistler, M., Felsenstein, E. and Koelbl, L. (2016). The structure of multivariate AR and ARMA systems: Regular and singular systems; the single and the mixed frequency case, *Journal of Econometrics* **192**, 366–373.
- [6] Bai, J.H. and Ng, S. (2002). Determining the number of factors in approximate factor models, *Econometrica* **70**, 191–221.
- [7] Barigozzi, M. and Hallin, M. (2016). Generalized dynamic factor models and volatilities: recovering the market volatility shocks, *The Econometrics Journal* **19**, 33–60.
- [8] Barigozzi, M. and Hallin, M. (2017). Generalized dynamic factor models and volatilities: estimation and forecasting, *Journal of Econometrics* **201**, 307–321.
- [9] Barigozzi, M., Hallin, M., von Sachs, R. and Soccorsi, S. (2021). Time-varying General Dynamic Factor Models and the measurement of financial connectedness, *Journal of Econometrics* **222**, 324–343.
- [10] Barigozzi, M., Hallin, M., Luciani, M. and Zaffaroni, P. (2021). Inferential theory for generalized dynamic factor models, Working Papers ECARES 2021-20, ULB – Université libre de Bruxelles.
- [11] Brillinger, D.R. (1964). A frequency approach to the techniques of principal components, factor analysis and canonical variates in the case of stationary time series, Royal Statistical Society Conference, Cardiff, available at <http://stat-www.berkeley.edu/users/brill/papers.html>.
- [12] Brillinger, D.R. (1981). *Time Series: Data Analysis and Theory*, San Francisco: Holden-Day.
- [13] Chamberlain, G. (1983). Funds, factors, and diversification in arbitrage pricing models. *Econometrica* **51**, 1281–1304.

- [14] Chamberlain, G. and M. Rothschild (1983). Arbitrage, factor structure and mean-variance analysis in large asset markets. *Econometrica* **51**, 1305–1324.
- [15] Chen, W., Anderson, B.D.O., Deistler, M. and Filler, A. (2011). Solutions of Yule-Walker equations for singular AR processes. *Journal of Time Series Analysis* **32**, 531–538.
- [16] Deistler, M., Anderson, B.D.O., Filler, A., Zinner, Ch. and Chen, W. (2010). Generalized linear dynamic factor models: an approach via singular autoregressions, *European Journal of Control* **16**, 211–224.
- [17] Deistler, M., Filler, A. and Funovits, B. (2011). AR systems and AR processes: The singular case. *Communications in Information and Systems* **11**, 225–236.
- [18] Forni, M., Giovannelli, A., Lippi, M., and Soccorsi, S. (2018). Dynamic factor models with infinite-dimensional factor space: forecasting, *Journal of Applied Econometrics* **33**, 625–642.
- [19] Forni, M., Hallin, M., Lippi, M., and Reichlin, R. (2000). The generalized dynamic factor model: identification and estimation, *Review of Economics and Statistics* **82**, 540–554.
- [20] Forni, M., Hallin, M., Lippi, M., and Zaffaroni, P. (2015). Dynamic factor models with infinite-dimensional factor space: one-sided representations. *Journal of Econometrics* **185**, 359–371.
- [21] Forni, M., Hallin, M., Lippi, M., and Zaffaroni, P. (2017). Dynamic factor models with infinite-dimensional factor space: asymptotic analysis, *Journal of Econometrics* **199**, 74–92.
- [22] Geweke, J. (1977). The dynamic factor analysis of economic time series, in Aigner, D.J. and Goldberger, A.S., Eds., *Latent Variables in Socio-Economic Models 1*, North-Holland, Amsterdam, pp. 365–383.
- [23] Hallin, M. and Lippi, M. (2014). Factor models in high-dimensional time series, a time-domain approach, *Stochastic Processes and their Applications* **123**, 2678–2695.
- [24] Hallin, M. and Lippi, M. (2019). High-dimensional time series and dynamic factor models, Preface to *The General Dynamic Factor Model: Time Series Analysis in High Dimension*, World Scientific, Singapore.
- [25] Hallin, M. and Trucíos, C. (2022). Forecasting Value-at-Risk and Expected Shortfall in large portfolios: a General Dynamic Factor Model approach, *Econometrics and Statistics*, to appear.
- [26] Johnstone, I.M. (2001). On the distribution of the largest eigenvalue in principal components analysis, *Annals of Statistics* **29**, 295–327.
- [27] Lippi, M., Deistler, M. and Anderson, B.D.O. (2022). High-dimensional dynamic factor models: a selective survey and lines of future research, *Econometrics and Statistics*, to appear.
- [28] Onatski, A., Moreira, M. J. and Hallin, M. (2013). Asymptotic power of sphericity tests for high-dimensional data, *Annals of Statistics* **41**, 1204–1231.
- [29] Onatski, A., Moreira, M. J. and Hallin, M. (2014). Signal detection in high dimension: the multi-spiked case, *Annals of Statistics* **42**, 225–254.

- [30] Sargent, T.J. and C.A. Sims (1977). Business cycle modelling without pretending to have too much a priori economic theory, in C.A. Sims, Ed., *New Methods in Business Cycle Research*, Federal Reserve Bank of Minneapolis, Minneapolis, pp. 45–109.
- [31] Stock, J.H. and Watson, M.W. (2002a). Macroeconomic forecasting using diffusion indexes, *Journal of Business and Economic Statistics* **20**, 147–162.
- [32] Stock, J.H. and Watson, M.W. (2002b). Forecasting using principal components from a large number of predictors, *Journal of the American Statistical Association* **97**, 1167–1179.
- [33] Trucíos, C., Mazzeu, J.H.G., Hallin, M., Hotta, L.K., Valls Pereira, P.L. and Zevallos, M. (2021). Forecasting conditional covariance matrices in high-dimensional time series with application to dynamic portfolio optimization: a General Dynamic Factor approach, *Journal of Business and Economic Statistics*, to appear.