# Optimal Linear Compression for Multivariate Time Series with Applications to Index Construction

Tucker McElroy[1]
U.S. Census Bureau

BRIN Mathematics Research Center
Recent Advances in Time Series Analysis
March 21-22, 2024

[1]This presentation is released to inform interested parties of research and to encourage discussion. The views expressed on statistical issues are those of the authors and not those of the U.S. Census Bureau (USCB). All time series analyzed in this presentation are from public or external data sources.

# Ben Kedem and Census

- Summer at Census 2011: "Integration of Information from Multiple Sources"
- Donald Martin [student]: Consultant, Summer at Census 2009
- Victor De Oliveira [student]: Summer at Census 2018
- Rich Gagnon [student]: time series staff (2003-2006)

# Outline

1. Background: the IDEA
2. Optimal Linear Compression and Canonical Analysis
3. Solution to Scalar Case
4. Examples and Applications
5. Open Problems

# Background: the IDEA

In 2023 the U.S. Census Bureau (USCB) began publishing a new data product: the IDEA (InDex of Economic Activity).

- Goal: to summarize 15 key monthly economic indicators (construction, retail, manufacturing, et al.) in a single time series, or index.
- Index would be updated daily according to the various release schedules.
- IDEA would provide a "snapshot of the combined movement of these" principal economic indicators.

# Visualization of the IDEA

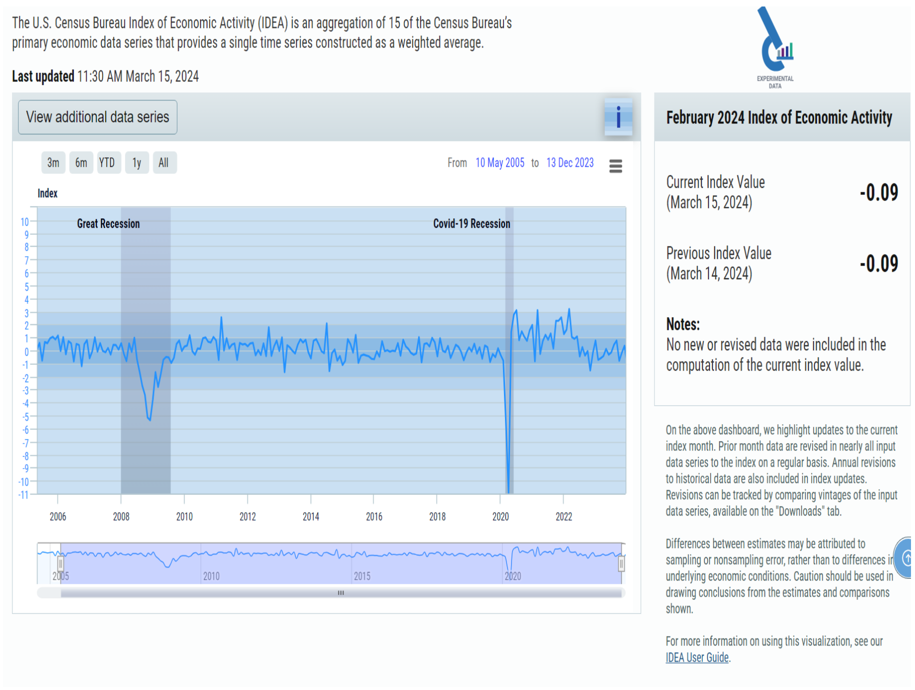


Figure 1: IDEA: `https://www.census.gov/economic-indicators/`

United States Census Bureau

# Background: Other Indexes

- Chicago Fed National Activity Index (Brave, 2008) of the Federal Reserve Bank of Chicago
- Weekly Economic Index (Lewis, Mertens, Stock, and Trivedi (2022)) of the Federal Reserve Bank of New York.

Both use Principal Components (PCA) methodology; so does IDEA.

# Background: Extending PCA

PCA seeks to compress the input time series (indicators) by finding a linear combination with maximal variance.

- For stationary time series this variance does not depend on time.
- PCA is designed for i.i.d. multivariate data, and does not utilize temporal dynamics.
- Why not extend to frequency domain PCA (Chapter 9 of Brillinger 2001)? Or Dynamic Factor Model (DFM)?

# Background: Feature Extraction

Comments by R. Kulik on ISI 2023 presentation:

- Does the index extract long-term features of the economy [policy] or short-term features [sentiment]?
- How to interpret? What data should be included?
- What sampling frequency is appropriate for objectives?
- What models are used? Nonlinear models?
- How to handle extremes? Discard for policy, retain for sentiment?

# Background: Feature Extraction

**Approach**: define "features" of interest through a "proxy" process, which exhibits the desired phenomena and can be related to input indicators through a joint time series model.

**Example**: Let GDP growth and the Unemployment Rate (UR) be the bivariate proxy. Then we seek a composite of the indicators that has the dynamics of GDP growth and UR.

# Optimal Linear Compression

**Elements**:

- Index $\{y_t\}$
- Indicators $\{x_t\}$
- Proxy $\{w_t\}$
- Compression: $y_t$ has lower dimension than $x_t$.

# Optimal Linear Compression

**Real-time**: $y_t$ is a function of $x_{t:} = \{x_s\}_{s \leq t}$.

**Linear**: $y_t = \Psi(B) x_t$ where $B$ is the backshift operator and $\Psi(z) = \sum_{k \geq 0} \psi_k z^k$ (the transfer function of the linear filter) is a power series (or causal Laurent).

**Optimal**: smallest trace mean squared error (MSE)

# Canonical Analysis

The problem of linear compression is equivalent to the canonical analysis of time series (Chapter 10 of Brillinger 2001).

- Compress $\{x_t\}$ unto $\{y_t\}$ via a (non-causal) linear filter, so as to closely approximate a target $\{w_t\}$.

- Canonical analysis generalizes frequency domain PCA (the case that $w_t = x_t$); can be used to construct indexes (Bowley, 1920).

- Brillinger (2001) solves non-causal case; we solve the causal (real-time) case.

# Canonical Analysis

**Framework**:

- A "de-compression filter" $\Omega(B)$ partially undoes the compression of $\Psi(B)$

- Seek to minimize trace mean square of $w_t - \Omega(B) y_t$.

- Equivalent to minimizing trace $V[w_t | y_{t:}] - V[w_t | x_{t:}]$, the linear conditional variances for estimating $w_t$ on the basis of $y_{t:}$ and $x_{t:}$, respectively.

**Heuristic**: we seek the best compression $y_t$ such that the loss is minimal in using $y_{t:}$ rather than $x_{t:}$ to describe $w_t$.

# Solution to Scalar Case

**Dimensions**:

- $x_t$ has dimension $n$
- $w_t$ has dimension $r$
- $y_t$ is scalar

**Scalar Case**: assume $r = 1$ (a scalar proxy); case $r > 1$ is work in progress.

# Solution to Scalar Case

**Canonical Analysis**: for stationary (mean zero) time series, seek $r \times 1$ filter $\Omega(B)$ and $1 \times n$ filter $\Psi(B)$ such that trace $V(w_t - \Omega(B)\Psi(B)x_t)$ is minimized.

**Real-time Case**: in canonical analysis, $\Omega(z)$ and $\Psi(z)$ are Laurent. In real-time case, they are multivariate power series (i.e., causal in $z$).

## Solution to Scalar Case

**Notation**: for two jointly stationary processes $\{a_t\}$ and $\{b_t\}$,

- their lag $h$ cross-covariance is $\text{Cov}(a_{t+h}, b_t) = \Gamma_{ab}(h)$;
- their multivariate cross-covariance generating function is denoted $f_{ab}(z) = \sum_{h \in \mathbb{Z}} \Gamma_{ab}(h) z^h$;
- setting $z = e^{-i\lambda}$, we obtain the cross-spectral density $f_{ab}(e^{-i\lambda})$;
- so $f_{aa}(z)$ is the autocovariance generating function of $\{a_t\}$.

# Solution to Scalar Case

Such a generating function $g(z)$ is a Laurent series in $z$.

- Let $\langle g \rangle_h$ denote its coefficient of index $h$ (corresponding to $z^h$);
- let $[g]_p^q = \sum_{h=p}^q \langle g \rangle_h z^h$ for $-\infty \leq p \leq q \leq \infty$ denote a portion of the full Laurent series;
- by the Fourier inversion theorem,

$$\langle g \rangle_h = (2\pi)^{-1} \int_{-\pi}^{\pi} e^{i\lambda h} g(e^{-i\lambda}) \, d\lambda.$$

# Solution to Scalar Case

**Other Notation**:

- $g^{-1}$ is a Laurent series, the matrix inverse of $g(z)$;
- $g'$ is transpose and $g^*$ is the conjugate transpose;
- tr denotes the trace.

# Solution to Scalar Case

**Terminology**:

- A Laurent series is *causal* when it is a power series;
- it is *strongly causal* if there are no constant terms;
- it is *anti-causal* if it is a power series in $z^{-1}$.

**Example**: if $g$ is Laurent, then $[g]_0^\infty$ is causal, $[g]_1^\infty$ is strongly causal, and $[g]_{-\infty}^0$ is anti-causal.

# Solution to Scalar Case

**Spectral Factorization**: an autocovariance generating function $g$ is Hermitian (i.e., $g^* = g$); if it is (strictly) positive definite, then $g$ can expressed as the product of a causal and an anti-causal factor:

$$g = \Theta \, \Sigma \, \Theta^*,$$

where $\Theta$ is a power series with initial coefficient given by the identity $I$ (i.e., $\Theta(0) = I$) and $\Sigma$ is an invertible constant matrix of the same dimension. We say that $\Theta$ is the spectral factor of $g$, with prediction error variance $\Sigma$.

## Solution to Scalar Case

**Strategy**: set $y_t = \Psi(B) x_t$, where $\Psi(z)$ is to be determined.

1. Find the causal filter $\widehat{\Omega}(z)$ which minimizes tr$V(w_t - \Omega(B) y_t)$. (This solution clearly depends on the dynamics of $\{y_t\}$.)

2. Find the causal filter $\widehat{\Psi}(z)$ which minimizes tr $V(w_t - \widehat{\Omega}(B)\Psi(B) x_t)$.

The following (more general) result addresses step 1.

# Solution to Scalar Case

**Proposition 1.** *Suppose that* $\{y_t\}$ *is a multivariate stationary time series with positive definite autocovariance function* $f_{yy}$*, with spectral factor* $\Theta_y$ *and prediction error variance* $\Sigma_y$*. The minimization of* $\mathrm{tr}\, V(w_t - \Omega(B) y_t)$ *over the class of power series* $\Omega(z)$ *is solved via*

$$\widehat{\Omega} = [f_{wy} \Theta_y^{-1*}]_0^\infty \Sigma_y^{-1} \Theta_y^{-1}. \tag{1}$$

**Remark**: observe (1) is causal. The same minimization over the broader Laurent class is solved by $\widehat{\Omega}$ given by removing the brackets, i.e., $f_{wy} f_{yy}^{-1}$. Cf. Bell and Martin (2004, JTSA).

# Solution to Scalar Case

**Towards Step 2**: with this choice of $\widehat{\Omega}$,

$$\mathrm{tr}\, V(w_t - \widehat{\Omega}(B)\Psi(B)x_t) = \mathrm{tr}\, \langle f_{ww} \rangle_0 - \mathrm{tr}\, \langle [f_{wy} \Theta_y^{-1*}]_0^\infty \Sigma_y^{-1} [\Theta_y^{-1} f_{yw}]_{-\infty}^0 \rangle_0 .$$

On the other hand, optimality ensures that for $\widehat{\Upsilon} = [f_{wx} \Theta_x^{-1*}]_0^\infty \Sigma_x^{-1} \Theta_x^{-1}$,

$$V[w_t - \widehat{\Omega}(B)\Psi(B)x_t] = V[w_t - \widehat{\Upsilon}(B)x_t] + V[(\widehat{\Upsilon}(B) - \widehat{\Omega}(B)\Psi(B))x_t].$$

# Solution to Scalar Case

The first term on the right hand side is $\langle f_{ww} - G \rangle_0$, where

$$G = [f_{wx} \Theta_x^{-1*}]_0^{\infty} \Sigma_x^{-1} [\Theta_x^{-1} f_{xw}]_{-\infty}^{0}. \qquad (2)$$

So our task is to find a power series $\Psi$ to minimize

$$\text{tr}\, V[(\widehat{\Upsilon}(B) - \widehat{\Omega}(B)\Psi(B))] = \text{tr}\, \langle G - [f_{wy} \Theta_y^{-1*}]_0^{\infty} \Sigma_y^{-1} [\Theta_y^{-1} f_{yw}]_{-\infty}^{0} \rangle_0, \qquad (3)$$

which is trace $V[w_t | y_{t:}] - V[w_t | x_{t:}]$.

# Solution to Scalar Case

**Proposition 2.** *For $r = 1$, (3) is minimized over the class of power series by*

$$\widehat{\Psi} = [f_{wx} \Theta_x^{-1*}]_0^\infty \Sigma_x^{-1} \Theta_x^{-1},$$

*and the corresponding real-time optimal compression variance is $\langle f_{ww} - G \rangle_0$.*

**Remark**: the scalar case of canonical analysis (Theorem 10.3.1 in Brillinger 2001) indicates that the minimal variance is $\langle f_{ww} - f_{wx} f_{xx}^{-1} f_{xw} \rangle_0$; this corresponds to the result of Proposition 2 if we remove the brackets from $G$ in (2).

# Examples and Applications

**Assessment and Interpretation**: without defining features, how do we interpret an index, or assess how well it is working?

**Construct an Index**: defining the proxy gives us a well-defined notion of "feature", and we can assess how well the index is working.

**Proxies**: if proxy is published, why not use it instead? Why bother creating an index?

# Examples and Applications

Some scenarios where an index may be preferable to the proxy:

- The proxy is not publicly available.
- The proxy is not observable; for instance, the proxy represents some signal or other portions of the dynamics in the indicators.
- The proxy is not timely; perhaps the proxy represents future values of the indicators, so that the index corresponds to a forecasting or nowcasting application.

# Examples and Applications

- The proxy is multivariate; then the single index in a sense distills both the proxy and the indicators.

- The proxy is of a lower sampling frequency; then the index represents a higher frequency approximation to the proxy.

- The proxy is a binary process corresponding to expansion and contraction (e.g., by recession-dating), and the index is a continuous approximation to the proxy. Equivalently, the binary proxy can be encoded as a zero-crossing process, where zero denotes the transition between expansion and contraction.

# Examples and Applications

**Example 1.** In the non-scalar case ($r > 1$), suppose that proxy and indicators are identical: $w_t = x_t$. Then the compression amounts to finding a best scalar summary of the indicators, and the problem is identical to frequency domain PCA (Chapter 9 of Brillinger 2001); writing $x_t = \widehat{\Omega}(B) y_t + \epsilon_t$ (where $\{\epsilon_t\}$ is simply the error between the proxy-indicators and the application of the de-compression filter to the index), we have a DFM interpretation.

# Examples and Applications

**Example 2.** The proxy is a common signal of the indicators. In particular, let $x_t = \beta s_t + n_t$, where $\{s_t\}$ is scalar and $\beta$ is a vector of common effects. For instance, if $\{s_t\}$ is a scalar cycle (stationary) process, and $\{n_t\}$ is stationary, then the indicators follow a common cycle model. Setting $w_t = s_t$, the proxy corresponds to the common driving cycle process, and the indicators will be compressed so as to extract this feature. Clearly $f_{wx} = f_{ss}\beta'$, and

$$\widehat{\Psi} = [f_{ss}\beta'\Theta_x^{-1*}]_0^{\infty}\Sigma_x^{-1}\Theta_x^{-1},$$

which is the concurrent cycle extraction filter.

# Examples and Applications

**Example 3.** The proxy is some linear combination $A$ of a future value of the indicators, say $h$ steps ahead. Then $w_t = A x_{t+h}$, and

$$\widehat{\Psi} = [\cdot^{-h} A \Theta_x \Sigma_x]_0^{\infty} \Sigma_x^{-1} \Theta_x^{-1},$$

where $\cdot^{-h}$ is the Laurent series $z^{-h}$.

## Examples and Applications

**Example 4.** The proxy is observed at a lower sampling frequency than the indicators, so that $\Delta$ time units of the indicator correspond to each time unit of the proxy. Let the index $t$ correspond to the lower sampling frequency (of the proxy), and redefine the indicators by stacking $\Delta$ of them that occur in each time $t$, writing the whole as $x_t$. Jointly model the proxy and stacked indicators at the low sampling frequency, obtain $\widehat{\Psi}$, and apply to $\{x_t\}$; by un-stacking, we obtain a filter of the original high frequency indicators.

# Examples and Applications

The resulting compression will be a high frequency synthesis of the indicators that targets the low frequency proxy, and hence can be used as an "advance" or "more timely" version of the proxy.

For example, the proxy could be quarterly and the indicators monthly, so that $\Delta = 3$. Then $x_t$ would be a 3-vector consisting of all the three monthly values of the various indicators that fall in quarter $t$.

# Examples and Applications

**A Basic Case**: suppose both proxy and indicators are available to the curator. To compute the optimal real-time compression filter, fit a joint VAR($p$) model to $\{w_t, x_t\}$, from which $f_{wx}$ and $f_{xx}$ are easily obtained, and the latter quantity can be spectrally factorized.

# Open Problems

1. Extend Proposition 2 to $r > 1$.

2. Extend framework to multiple indexes (so $\Psi(z)$ is $m \times n$ for $m > 1$).

3. What is the source of a particular observed movement in the index? [Client]

4. Extend to non-linear compression; use higher order polyspectra? [Kulik]

5. How to handle extremes? [Kulik]

# Open Problems

**Problem 1**: In analogy with Proposition 2 and Brillinger's canonical analysis, for $r > 1$ we might conjecture

$$\widehat{\Psi} = e_1' U^* [f_{wx} \Theta_x^{-1*}]_0^\infty \Sigma_x^{-1} \Theta_x^{-1},$$

where $G = U^* \Lambda U$ is the eigen-decomposition and $e_1$ is the first unit vector. However, $e_1' U^*$ need not be causal, so $\widehat{\Psi}(z)$ above may not be a power series.

**One Idea**: there exists a unitary $W$ such that $e_1' U^* W$ is causal (proof of spectral factorization in Janashia et al. (2011)).

# Open Problems

**Problem 3**: we propose to compute $n$ all-but-one compressions, where we compute an optimal real-time compression based on $n$ sets of indicators corresponding to leaving one out.

- Each sub-index uses less information; if an individual indicator is somewhat redundant with the others – for the purpose of predicting the proxy – then the sub-index where that indicator has been omitted will tend to track the main index.

- Conversely, a large discrepancy between sub-index and main index means that the particular indicator was important at that time $t$.

- Gauge size of importance through variance of difference of the sub-index and the main index.

# Open Problems

**Problem 5**: for IDEA we eliminate any time points corresponding to extremes – such as Covid epoch – for the purpose of computing PCA weights. (IDEA only depends on sample mean and sample covariance calculations, so these can be adapted for NAs.)

Could extreme values be part of the desired features? What proxy process would capture this?

# Conclusion

Can we put index creation on a rigorous foundation of optimal compression?

Thank you!

Contact: tucker.s.mcelroy@census.gov