

Results on the PASCAL PROMO challenge

Ivan Markovsky

University of Southampton

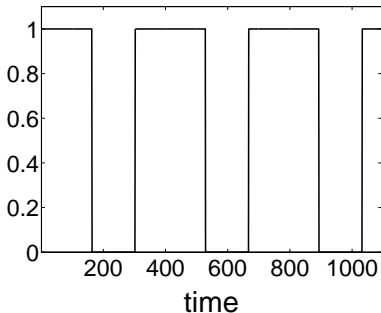
The challenge

Data: consists of two (simulated) time series

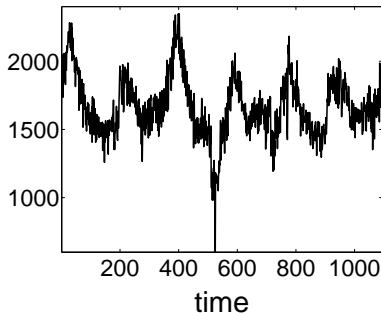
$$\begin{aligned} u_d(1), \dots, u_d(1095) &\in \{0, 1\}^{1000} && \text{promotions} \\ y_d(1), \dots, y_d(1095) &\in \mathbb{R}^{100} && \text{product sales} \end{aligned}$$

Challenge: find ≤ 50 promotions that affect each product sales

7th promotion



3rd product sales



Comments

- time series nature of the data \implies **dynamic phenomenon** (the current output may depend on past inputs and outputs)
- it is natural to think of the promotions as inputs (causes) and the sales as outputs (effects)
- **multivariable data**: $m = 1000$ inputs, $p = 100$ outputs
- $T = 1095$ data points—very few, relative to m and p
- even static linear model $y = Au$ is **unidentifiable** (A can not be recovered uniquely from (u_d, y_d)) for $T < T_{\min} := 10^5$
- prior knowledge that a few (≤ 50) inputs affect each output helps ($T_{\min} = 5000$) but doesn't recover identifiability
- this prior knowledge makes the problem **combinatorial**

Proposed model

Main assumptions:

1. **static input-output relation** $y_j(t) = a_j u(t)$
(this implies that one output can not affect other outputs)
2. there is offset and seasonal component, which is sine, *i.e.*,

$$\text{Base line: } y_{\text{bl},j}(t) := b_j + c_j \sin(\omega_j t + \phi_j)$$

The model is

$$y_j(t) = y_{\text{bl},j}(t) + Au(t)$$

or, with $Y := [y(1) \ \cdots \ y(T)]$, $U := [u(1) \ \cdots \ u(T)]$, *etc.*,

$$Y = Y_{\text{bl}}(b, c, \omega, \phi) + AU$$

Identification problem

Parameters:

- $A \in \mathbb{R}^{p \times m}$ — input/output (feedthrough) matrix
- $\mathbf{b} := (b_1, \dots, b_p) \in \mathbb{R}^p$ — vector of offsets
- $\mathbf{c} := (c_1, \dots, c_p) \in \mathbb{R}^p$ — vector of amplitudes
- $\boldsymbol{\omega} := (\omega_1, \dots, \omega_p) \in \mathbb{R}^p$ — vector of frequencies
- $\boldsymbol{\phi} := (\phi_1, \dots, \phi_p) \in [-\pi, \pi]^p$ — vector of phases

Identification problem:

minimize over the parameters $\|Y_d - Y_{bl}(\mathbf{b}, \mathbf{c}, \boldsymbol{\omega}, \boldsymbol{\phi}) - A\mathbf{U}_d\|$
subject to each row of A has at most 50 nonzero elements.

combinatorial, constrained, nonlinear, least squares problem

Solution approach

Model: $\hat{y}_j(t) = b_j + c_j \sin(\omega_j t + \phi_j) + Au(t)$

Linear in A, b, c . Nonlinear in ω, ϕ . Combinatorial in A .

Our approach: Split the problem into two stages:

1. **Baseline estim.:** minimize over b, c, ω, ϕ , assuming $A = 0$.
Nonlinear LS problem. We use local optimization.
2. **I/O function etim.:** minimize over A, b, c , with ω, ϕ fixed.
This is a combinatorial problem. We use the ℓ_1 heuristic.

This approach simplifies the solution but leads to **suboptimality**.

Identification of the autonomous term

The problem **decouples into p independent problems:**

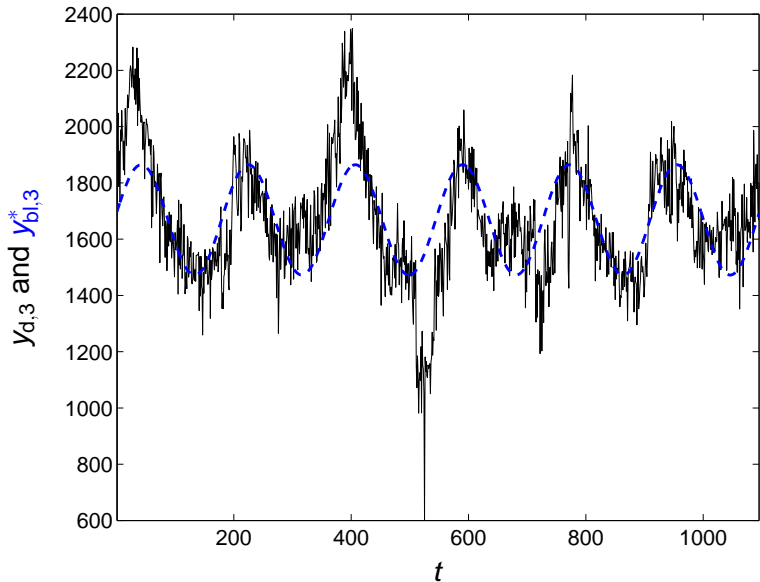
$$\begin{aligned} \text{minimize} \quad & \text{over } \mathbf{b}_j, \mathbf{c}_j, \omega_j \in \mathbb{R}, \phi_j \in [-\pi, \pi] \quad \|\mathbf{y}_{d,j} - \mathbf{y}_{bl,j}(\mathbf{b}_j, \mathbf{c}_j, \omega_j, \phi_j)\|_2 \\ & (1) \\ & (\mathbf{y}_{d,j} \text{ — } j\text{th row of } \mathbf{Y}_d, \quad \mathbf{y}_{bl,j} \text{ — } j\text{th row of } \mathbf{Y}_{bl}) \end{aligned}$$

A special case of the line spectral estimation problem, for which solution subspace and maximum likelihood (ML) methods exist.

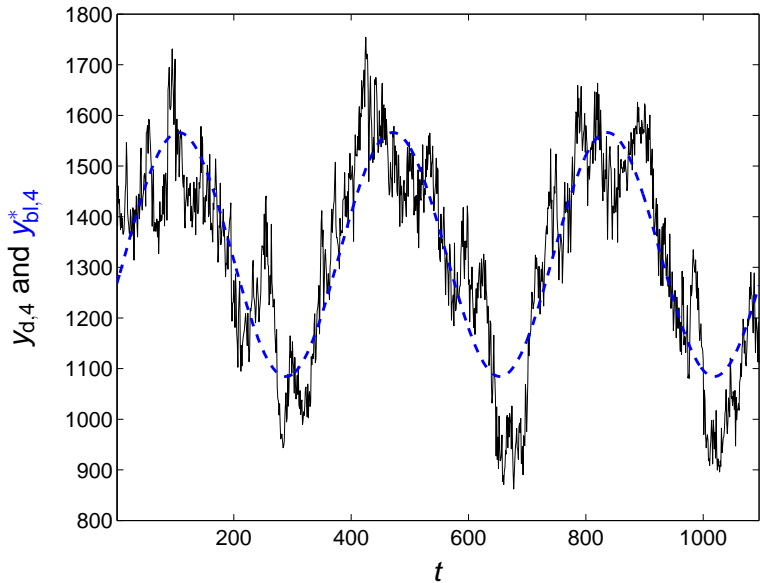
We use the ML approach, *i.e.*, **local optimization**, assuming $\omega_j = 12\pi/T$ (one year period) or $6\pi/T$ (half year period).

Furthermore, we eliminate the “linear” parameters $\mathbf{b}_j, \mathbf{c}_j$ by projection \rightsquigarrow **VARPRO method**

Baseline



Baseline



Identification of the term involving the inputs

Problem:

$$\begin{aligned} & \text{minimize} \quad \text{over } b_j, c_j, a_j \quad \|y_{d,j} - y_{bl,j}(b_j, c_j, \phi_j^*, \omega_j^*) - a_j^\top U_d\|_2 \\ & \text{subject to} \quad a_j \text{ has at most 50 nonzero elements} \end{aligned} \tag{2}$$

Proposed heuristic:

$$\begin{aligned} & \text{minimize} \quad \text{over } b_j, c_j, a_j \quad \|y_{d,j} - y_{bl,j}(b_j, c_j, \phi_j^*, \omega_j^*) - a_j^\top U_d\|_2 \\ & \text{subject to} \quad \|a_j\|_1 \leq \gamma_j \end{aligned} \tag{3}$$

$\gamma_j > 0$ is parameter controlling the **sparsity vs accuracy trade-off**

Choice of the regularization parameter γ_j

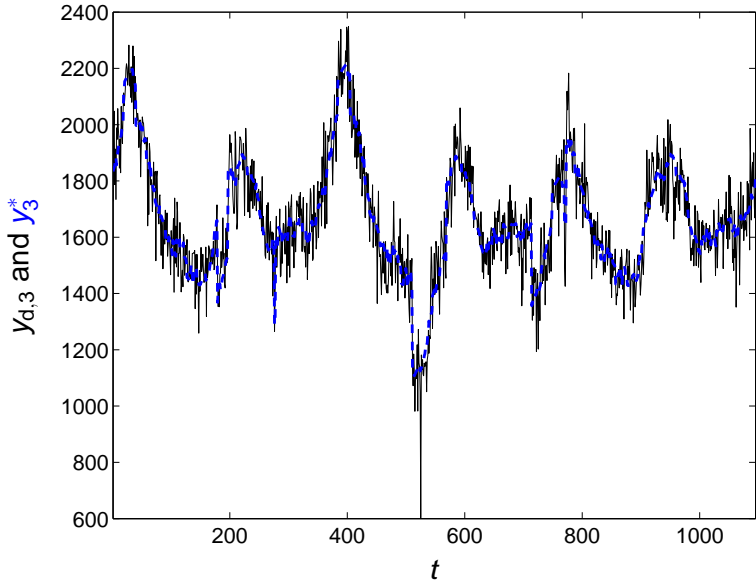
If we fix the nonzero elements to be the first 10 elements, the optimal solution (with this choice of the nonzero elements) is

$$\mathbf{a}_j := [(\mathbf{y}_{d_j} - \mathbf{y}_{bl,j}) \mathbf{U}_d(1:10, :)^+ \quad \mathbf{0}_{1 \times (m-10)}]$$

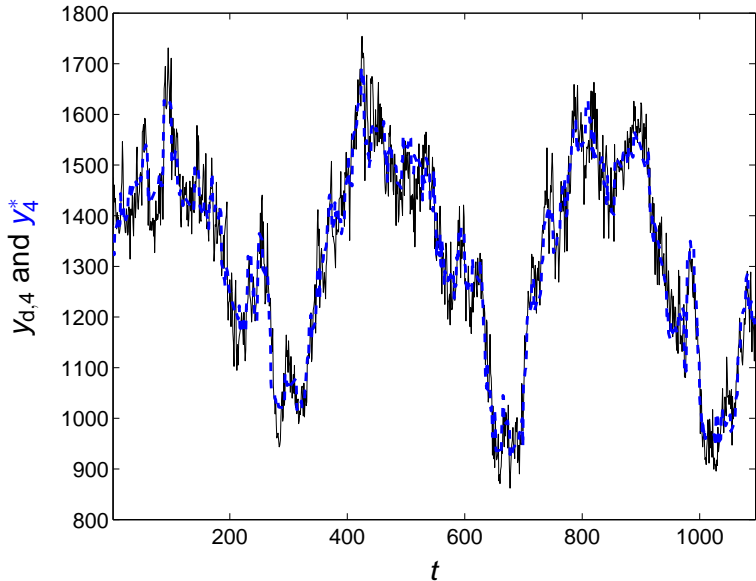
Let \mathbf{a}^* be the optimal solution over all choices of the nonzero elements.

Since $\|\mathbf{a}_j^*\|_1 = \gamma_j$, a heuristic choice for γ_j is $\gamma_j := \|\mathbf{a}_j\|_1$.

Complete model (baseline and 24 inputs)



Complete model (baseline and 25 inputs)



Nonuniqueness of the solution

For uniqueness of A , we need U_d to be full row rank.

Special cases that lead to rank deficiency of U :

- **Zero inputs** can't affect the output. Removing them leads to an equivalent reduced model. For maximum sparsity, assign zero weights in A to those inputs.
- **Inputs that are multiples of other inputs** lead to essential nonuniqueness that can not be recovered by the sparsity.

Preprocessing step: remove redundant inputs.

Algorithm

1. **Input:** $U_d \in \mathbb{R}^{m \times T}$ and $Y_d \in \mathbb{R}^{p \times T}$.
2. **Preprocessing:** detect and remove redundant inputs.
3. For $j = 1$ to p
 - 3.1 **Identify the baseline** $\rightsquigarrow (\omega_j^*, \phi_j^*, \mathbf{c}_j^*, \mathbf{a}_j^*)$
 - 3.2 **Identify the I/O relation** $\rightsquigarrow (\mathbf{b}_j^*, \mathbf{c}_j^*, \mathbf{a}_j^*)$, sparsity pattern of \mathbf{a}_j^*
 - 3.3 **Solve (2) with fixed sparsity pattern**, $\phi_j = \phi_j^*$ and $\omega_j = \omega_j^*$
 $\rightsquigarrow (\mathbf{b}_j^*, \mathbf{c}_j^*, \mathbf{a}_j^*)$
4. **Postprocessing:** add zero rows in A^* corresponding to the removed inputs
5. **Output:** $Y_{bl}(\mathbf{b}^*, \mathbf{c}^*, \omega^*, \phi^*)$ and A^*

Identification of the baseline:

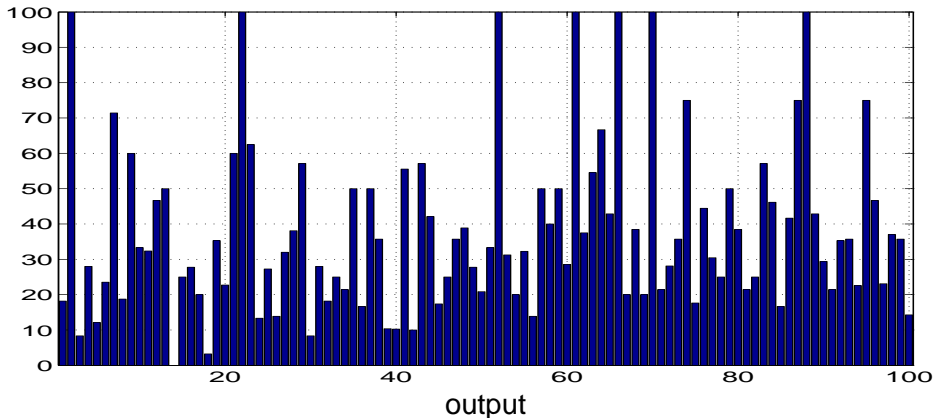
1. Let f', ϕ_j' be minimum value/point of (1) with $\omega_j = 6\pi/T$.
2. Let f'', ϕ_j'' be minimum value/point of (1) with $\omega_j = 12\pi/T$.
3. If $f' < f''$, $\omega_j^* := 6\pi/T$, $\phi_j^* := \phi_j'$, else $\omega_j^* := 12\pi/T$, $\phi_j^* := \phi_j''$.

Identification of the baseline:

1. Let $\gamma_j := \|(y_{d,j} - y_{bl,j})U_d(1:10, :)^+\|_1$.
2. Let a_j' be solution to (3) with $\phi_j = \phi_j^*$, $\omega_j = \omega_j^*$.
3. Determine the sparsity pattern of a_j' .

Results on the PROMO challenge

% correctly identified inputs



Total: 2321 true inputs, 1796 identified inputs, of which 507 correct.

Code: <http://www.ecs.soton.ac.uk/~im/challenge.tar>