Probabilistic Targeted Factor Analysis

Miguel C. Herculano Santiago Montoya-Blandon
University of Glasgow

CRETE ’26

Outline

- **Introduction**
- **Probabilistic Targeted Factor Analysis (PTFA)**
  - Model
  - EM Algorithm
  - Extensions: Missing Data, Stochastic Volatility, Dynamic Factor Models
- **Simulations and Empirical Results**
  - Simulation Study
  - Forecasting Applications
- **Conclusion**

Introduction

- High-dimensional data common in Economics and Finance.
- Popular dimensionality reduction methods: PCA, PLS.
- Partial Least Squares (PLS): Targets specific variables but lacks probabilistic structure.
- **PTFA:** Introduces probabilistic foundation to PLS.

Principal Component Analysis (PCA)

- **Unsupervised** dimensionality reduction method.
- **Goal**: Transform a set of correlated variables X into uncorrelated variables f called principal components (i.e. Extracts latent factors f maximizing variance in predictors X).
- Formally, the PCA finds the eigenvectors of the covariance matrix S that correspond to the largest eigenvalues as follows. the kth principal component direction v_k solves:

$$\max_{\alpha} \ \mathrm{Var}(\mathbf{X}\alpha) \quad \text{subject to} \quad \|\alpha\| = 1, \quad \alpha^T S v_l = 0, \quad l = 1, \ldots, k-1.$$

- Does **not** consider target variables explicitly.
- Deterministic (SVD) and sensitive to noise and outliers.

Partial Least Squares (PLS)

- **Supervised** dimensionality reduction method.
- **Goal**: Unlike PCA, PLS explicitly uses target variables Y in constructing components.
- Extracts latent factors from X maximizing covariance with Y, rather than maximizing variance. Formally, the kth PLS direction ϕ_k solves:

$$\max_{\alpha} \operatorname{Corr}^2(\mathbf{Y}, \mathbf{X}\alpha)\operatorname{Var}(\mathbf{X}\alpha) \quad \text{subject to} \quad \|\alpha\| = 1 \quad \alpha^T S \phi_l = 0, \quad l = 1, \ldots, k-1.$$

- Iterative component selection via NIPALS algorithm.
- Still deterministic and does not explicitly handle uncertainty or noise.

What is PTFA?

- Finds latent factors f maximizing predictability of specific targets Y.
- **Probabilistic interpretation**: handles uncertainty, noise, missing data, stochastic volatility.
- PTFA admits a dynamic form, giving rise to a Targeted Dynamic Factor Model specification.
- Estimation via Expectation-Maximization (EM) algorithm.

Model Specification

A simple statistical formulation for performing factor extraction that embodies the objective of PLS.

$$x = Pf + e_x, \quad e_x \sim N_p(0, \sigma_x^2 I_p)$$
$$y = Qf + e_y, \quad e_y \sim N_q(0, \sigma_y^2 I_q)$$

- Common latent factors $f \sim N_k(0, V_F)$,
- Factor loadings P, Q and e_x, e_y are isotropic Gaussian noise terms.
- Conditional likelihood: Gaussian with structured covariance.

$$p(x, y|f) = p(x|f)p(y|f) = N(\mu, \Sigma)$$

EM Algorithm for PTFA

1. **E-step**: derive the complete data log-likelihood $\log p(\mathbf{X}, \mathbf{Y}, \mathbf{F} \mid \theta)$ and compute

$$
\begin{aligned}
Q(\theta) &\coloneqq E_{\mathbf{F}|\mathbf{X},\mathbf{Y};\theta}[\log p(\mathbf{X}, \mathbf{Y}, \mathbf{F} \mid \theta)], \\
&= E_{\mathbf{F}|\mathbf{X},\mathbf{Y};\theta}\Big[\log p(\mathbf{X} \mid \mathbf{F}; \mathbf{P}, \sigma_x^2) + \log p(\mathbf{Y} \mid \mathbf{F}; \mathbf{Q}, \sigma_y^2) + \log p(\mathbf{F})\Big]. \qquad (1)
\end{aligned}
$$

   Compute posterior $f|x, y \sim N_k(m, \Omega)$.
2. **M-step**: The maximization (M) step consists in optimizing $Q(\theta)$ (given an initial value) with a view of deriving updating rules for θ. Update parameters $(P, Q, \sigma_x^2, \sigma_y^2)$ via closed-form updates.

Extensions

- **Missing Data**: Handles Missing-at-Random and Mixed-Frequency scenarios.
- **Stochastic Volatility**: EWMA method for updating volatility estimates.
- **Factor Dynamics**: Introduces persistence via Dynamic Factor Models (DFM).

Missing Data (MaR)

Data could be missing at random (MaR) or, there may be an underlying pattern for missingness (eg. mixed-frequency).

- **Missing at Random**: For any row $t \in \{1, \ldots, T\}$ and feature $j \in \{1, \ldots, p\}$, we let $\tau_{tj}^{(X)}$ be an indicator for whether that particular observation is missing in the feature matrix $\mathbf{X}$.

$$\tilde{\mathbf{X}}_{ij} = \sum_{c=1}^{k} \tilde{\mathbf{M}}_{ic} \tilde{P}_{jc} \quad \text{if observation is missing in } X,$$

where $\tilde{\mathbf{M}}$ is the matrix of estimated posterior means of the latent factors for each observation, computed in the E-step of the EM algorithm.

Missing Data (Mixed-Frequency)

There could also be 'patterns' of missingness, namely mixed-frequency data (e.g., monthly predictors with quarterly targets).

- **Mixed-Frequency**: we introduce latent targets at higher frequencies, y^*, which are aggregated to match lower-frequency observations:

$$y = \frac{1}{L}\sum_{l=1}^{L} y^{*(l)}, \quad \text{with} \quad y^{*(l)} = Qf^{(l)} + e_y^{*(l)},$$

  where L is the ratio of high-to-low frequency observations.
- This aggregation generates a conditional likelihood adjusted directly into the EM steps, allowing us to jointly estimate parameters at mixed frequencies.
- The posterior distributions are computed for these latent higher-frequency factors, considering correlations within lower-frequency periods

Stochastic Volatility via EWMA

Because the innovations are isotropic, we can simplify volatility modelling by applying simulation-free, recursive variance matrix discounting.

We estimate $\sigma_x^2(t)$ and $\sigma_y^2(t)$ using Exponentially Weighted Moving Averages (EWMA):

$$\sigma_x^2(t) = (1-\lambda_x) \cdot \hat{\sigma}_x^2(t) + \lambda_x \cdot \sigma_x^2(t-1),$$

$$\sigma_y^2(t) = (1-\lambda_y) \cdot \hat{\sigma}_y^2(t) + \lambda_y \cdot \sigma_y^2(t-1),$$

where $\hat{\sigma}_x^2(t)$ and $\hat{\sigma}_y^2(t)$ are the per-period variance estimates from the model.

In practice:

- λ_x, λ_y are set close to 1 to smooth volatility and enable adaptive learning.
- Setting $\lambda_x = \lambda_y = 0$ removes stochastic volatility, reducing the model to the static case with variances computed as time averages.

Targeted Dynamic Factor Models

Factor dynamics are conceptually straightforward to introduce and are attractive for the applied econometrician.

- We extend PTFA to a Dynamic Factor Model (DFM) by assuming:

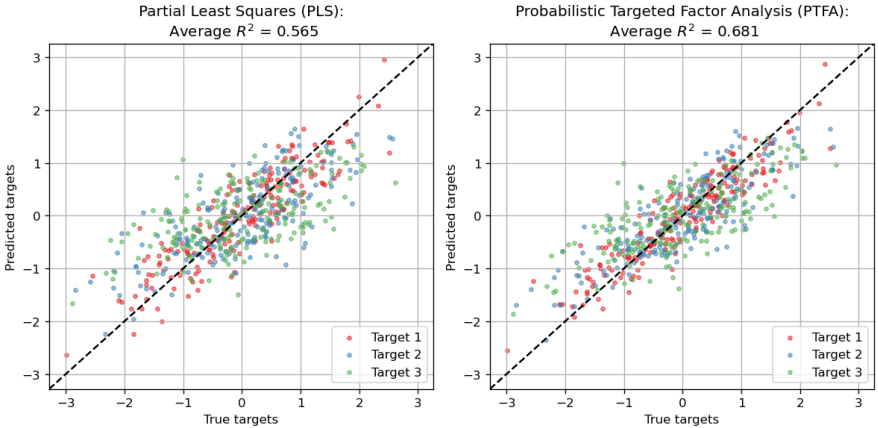
$$
\begin{aligned}
x_t &= Pf_t + e_{x,t}, \\
y_t &= Qf_t + e_{y,t}, \\
f_t &= Af_{t-1} + v_t,
\end{aligned}
$$

- where A is a $k \times k$ transition matrix and $v_t \sim \mathcal{N}(0, V_F)$.
- This VAR(1) structure introduces serial dependence in the factors.
- Parameters A and f_0 are estimated via EM using posterior moments from a banded precision approach (Durrande et al., 2019).
- Posterior moments $\mathbb{E}[F]$, $\mathbb{E}[FF']$ replace Kalman filtering for efficiency.

Simulation Setup

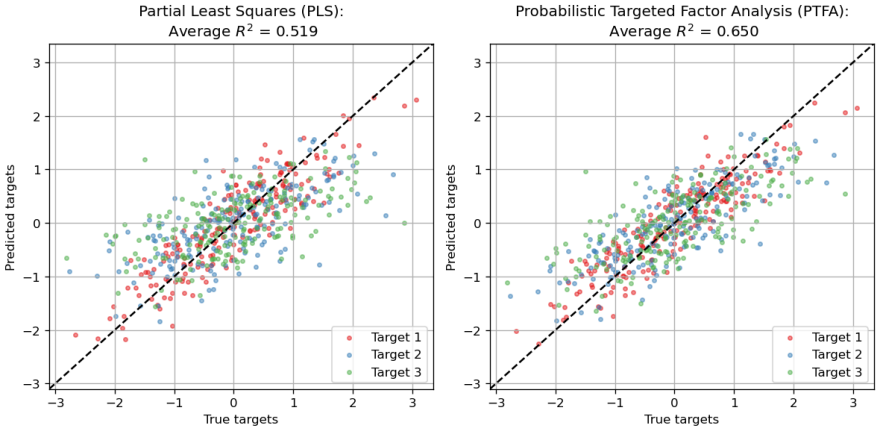
- Simulation Setup:
  - we set $T = 200$, $p = 10$, $q = 3$, and $k = 2$.
  - We first draw all entries in the loadings P and Q from a uniform distribution, between 0 and 1.
  - Then, we generate $f_t \sim N_k(0_k, I_k)$ independently for each time period $t \in \{1, \ldots, T\}$.
  - Given the factors and loadings, we finally generate features and targets according to model equations.
- We consider three DGPs which vary in the distribution assumed for e_x and e_y.
  - DGP.1: isotropic Gaussian error (fixing $\sigma_x = \sigma_y = 1$)
  - DGP.2: Gaussian errors with a correlation structure.
  - DGP.3: Non-Gaussian errors (t-student, qui-square).

Simulation Results



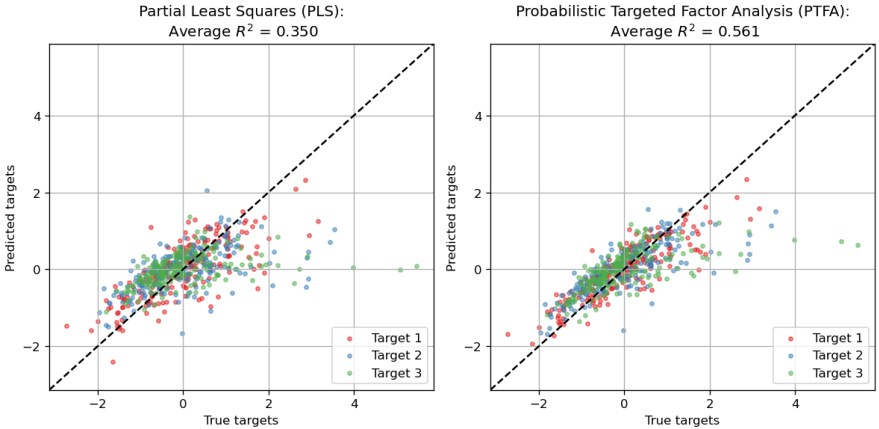
Comparison of PLS and PTFA on a single realization of simulated data with independent Gaussian errors.

Simulation Results



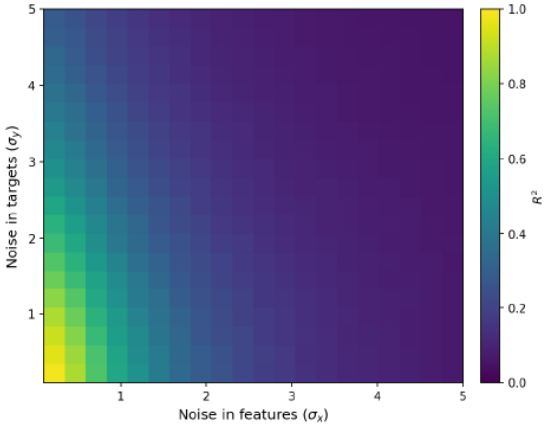
Comparison of PLS and PTFA on a single realization of simulated data with dependent Gaussian errors.

Simulation Results

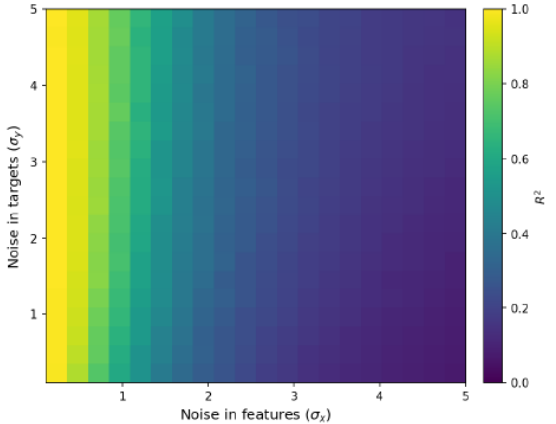


Comparison of PLS and PTFA on a single realization of simulated data with non-Gaussian errors.

Simulation Results



(a) Partial Least Squares (PLS)



(b) Probabilistic Targeted Factors (PTFA)

Comparison of the median R^2 statistics for PLS and PTFA across 1000 replications from DGP , varying noise in features (σ_x) and targets (σ_y).

Empirical Applications

- **Macroeconomic Forecasting**: McCracken and Ng (2016) FRED-MD dataset
- **Equity Premium Prediction**: Using Goyal et al. (2024) data
- **Financial Conditions Indices (FCI)**: NFCI dataset used for Targeted construction

Empirical Results: Macroeconomic Forecasting

Horizon	PCA	PLS	PTFA	PTFA-SV
Industrial Production				
1	0.7990	2.6746	0.6996	0.7003
4	0.8747	1.2164	0.8363	0.8367
6	0.9192	1.3295	0.8716	0.8739
12	0.9966	1.4466	0.9116	0.9138
CPI Inflation				
1	1.0155	1.2575	0.9903	0.9912
4	1.0460	1.6858	1.0033	1.0038
6	1.0277	1.8232	1.0018	1.0011
12	1.0368	1.7118	1.0119	1.0123
Unemployment Rate				
1	1.2254	1.0781	0.9048	0.9051
4	1.1147	1.3735	1.0069	1.0094
6	1.0569	2.2511	1.0117	1.0121
12	1.0507	1.5848	1.0089	1.0094

Notes: MSFE statistics calculated out-of-sample with a rolling window of 200 months. Forecast horizons: 1, 4, 6, and 12 months. 7 factors used consistently across models. Standardization applied to predictors and targets.

- PTFA outperforms all other methods, improving MSFE over prediction horizons
- PTFA-SV doesn't seem to improve upon the baseline homoskedastic model

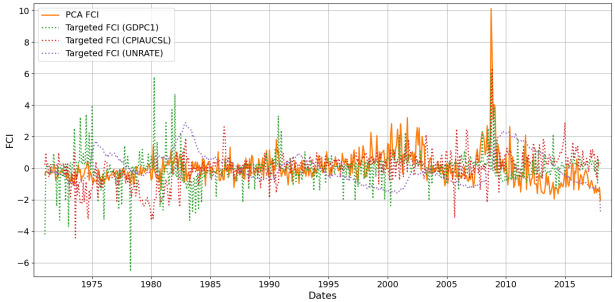
Empirical Results: Equity Premium Prediction

Model	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$
PCA	1.0678	1.1182	1.1261	1.1570	1.1780
PLS	1.1043	1.2110	1.3280	1.5425	1.7775
PTFA	1.0139	1.0363	1.0557	1.0680	1.1027
PTFA-SV	1.0136	1.0403	1.0648	1.0772	1.1335

Table: Out-of-Sample Performance of PTFA

- MSFE statistics calculated one-month-ahead out-of-sample on a rolling window of 60 monthly observations on a sample 1926M1-2023M12.
- PTFA outperforms all other methods, improving MSFE over prediction horizons, independent of the choice of k.

Empirical Results: Targeted Financial Conditions Indices



- Financial Condition Indices have been criticized for their lack of interpretability.
- Targeted FCIs offer fine-tuned indices that target specific economic variables of interest.

Conclusion

- Probabilistic approach significantly enhances interpretability and robustness.
- Empirical evidence demonstrates PTFA's practical value.
- Open-source implementation (Python package).
- Future directions: Full Bayesian treatment, sparsity priors, and further empirical applications.