Using High-Order Moments to Estimate Linear Independent Factor Models

Stéphane Bonhomme
CEMFI, Madrid

Jean-Marc Robin
Université de Paris I, Panthéon-Sorbonne,
University College London and
Institute for Fiscal Studies

Linear factor models

Examples:

- Measurement error model:

$$\begin{cases} y_i = a + bx_i^* + u_i, \\ x_i = x_i^* + v_i. \end{cases}$$

- Panel data models with one or more individual effect, with time-individual interactions:

$$y_{it} = z_{it}^T \beta + \lambda_{t1} x_{i1} + ... + \lambda_{tK} x_{iK} + u_{it}.$$

  Usually small T, large N.

- Structural VAR models:

$$y_t = By_{t-1} + Cu_t, \quad y_t, u_t \text{ vectors.}$$

- Finance (APT).

Questions

- Identification and estimation of parameters b, λ and C (factor loadings)?
- Identification and estimation of factor distributions? Prediction of factors and common components. (See companion paper.)

Orthogonal factor analysis

Model structure:

$$Y = \Lambda X + U, \quad \text{where} \begin{cases} Y : L\text{-vector} \\ X : K\text{-vector (zero mean)} \\ U : L\text{-vector (zero mean)} \\ \Lambda : L \times K \text{ matrix} \end{cases}$$

Assumptions: orthogonal factors and errors, and $\text{Var}(X) = I_K$ (normalisation).

Identifying restrictions: Matrix Λ identified from second-order restrictions:

$$\text{Var}(Y) = \Lambda\Lambda^T + \text{Var}(U).$$

Fundamental nonidentification result: Λ identified up to a multiplicative orthonormal matrix (as $\Lambda\Lambda^T = \Lambda Q Q^T \Lambda^T$ for all orthonormal Q).

Principal Component Analysis: ML + normal errors $\Rightarrow$ one particular normalisation.

Independent component analysis

Like OFA with "independence" instead of "orthogonality" + **no noise.**

Identification based on second and fourth-order moments.

Very commonly used in the literature on blind signal separation and image processing.

Many algorithms: e.g. Cardoso and Souloumiac's (1993) JADE algorithm (based on structural restrictions on matrices of fourth-order cumulants of data), Hyvärinen's FastICA algorithm (find w maximizing the non-gaussianity of the projection $w^T Y$).

Quasi-JADE

In this paper,

We develop a two-stage estimation algorithm for **noisy linear independent factor models**:

- First stage estimates error moments;
- Second stage applies JADE.

We show formal identification results,

We run Monte-Carlo simulations,

We provide an empirical application.

Empirical application

Y : log hourly wage (residual of a regression on background variables and age)

D : age at the end of school

D^* : median of D given certified highest diploma

OLS:

- Regress Y on $D = 4.37\%$;
- Regress Y on $D^* = 6.03\%$.

Model:

$$\text{One-factor model} \quad \begin{cases} Y = \lambda_{11} X_1 + U_1 \\ D = \lambda_{21} X_1 + U_2 \\ D^* = \lambda_{31} X_1 + U_3 \end{cases} \quad \text{or} \quad \text{Two-factor model} \quad \begin{cases} Y = \lambda_{11} X_1 + \lambda_{12} X_2 + U_1 \\ D = \lambda_{21} X_1 + \lambda_{22} X_2 + U_2 \\ D^* = \lambda_{31} X_1 + \lambda_{32} X_2 + U_3 \end{cases}$$

	$K=1$ PCA	$K=1$ quasi-JADE(4)	$K=1$ quasi-JADE(3,4)	$K=2$ quasi-JADE(4)	$K=2$ quasi-JADE(3,4)
$\widehat{\lambda}_{11}$	.141 (.138,.145)	.154 (.136,166)	.142 (.137,.148)	.172 (.146,.200)	.166 (.145,.182)
$\widehat{\lambda}_{21}$	2.15 (2.12,2.19)	2.09 (2.02,2.18)	2.13 (2.09,2.20)	2.05 (1.96,2.16)	2.09 (2.02,2.19)
$\widehat{\lambda}_{31}$	2.01 (1.98,2.03)	2.05 (1.95,2.14)	2.03 (1.96,2.11)	2.02 (1.93,2.12)	2.02 (1.93,2.10)
$\frac{\widehat{\lambda}_{11}}{\widehat{\lambda}_{21}}$	6.6%	7.4%	6.7%	8.5%	7.9%
$\widehat{\lambda}_{12}$	-	-	-	-.138 (-.212,-.067)	-.136 (-.209,-.040)
$\widehat{\lambda}_{22}$	-	-	-	.360 (.009,.561)	.316 (.091,.459)
$\widehat{\lambda}_{32}$	-	-	-	.475 (.310,.660)	.381 (.131,.484)
$\widehat{V}(U_1)$	.066 (.065,.067)	.052 (.041,.070)	.066 (.060,.069)	.038 (.000,.060)	.040 (.010,.063)
$\widehat{V}(U_2)$	2.31 (2.22,2.40)	2.56 (2.06,2.90)	2.43 (2.04,2.65)	2.61 (1.85,3.04)	2.50 (1.92,2.84)
$\widehat{V}(U_3)$	.672 (.604,.745)	.426 (.000,.850)	.586 (.177,.867)	.385 (.000,.766)	.500 (.089,.889)
$\kappa_3(X_1)$		-	1.34 (1.29,1.39)	-	1.17 (1.08,1.30)
$\kappa_3(X_2)$		-	-	-	.087 (-.709,6.10)
$\kappa_4(X_1)$		.612 (.391,.854)	.741 (.354,1.02)	.627 (.439,.768)	.665 (.445,.841)
$\kappa_4(X_2)$		-	-	13.6 (3.58,196)	15.5 (4.28,580)

- We thus obtain the following factor structure:

$$\begin{cases} Y = .17X_1 - .14X_2 + U_1 \\ D = 2X_1 + .4X_2 + U_2 \\ D^* = 2X_1 + .4X_2 + U_3 \end{cases}$$

- Interestingly, this model is consistent with a classical Mincer equation:

$$\begin{cases} Y = \alpha E + V, \\ E = 2X_1 + .4X_2, \\ V = (.17 - 2\alpha)X_1 - (.14 + .4\alpha)X_2 + U_1. \end{cases}$$

  where E can be interpreted as "true education", measured with error by D and D^* ($\text{Var}(E) = 5.6$, $\text{Var}(U_2) = 2.6$ and $\text{Var}(U_3) = .4$).

- $\text{Cov}(E, V) = 0$ if and only if $\alpha = 6.8\%$. Same as 2SLS (controls for measurement error).

- Next, suppose that $\text{Cov}(E, V) \neq 0$. Identification of α requires instruments Z.

- **Except** if one assumes that Z is independent of V. Then, $Z = X_1$ or X_2 and $\alpha = 8.5\%$ or $\alpha = -35\%$. Only $\alpha = 8.5\%$ is reasonable.

- Yields following decomposition: ICA - OLS = 4.1%, 2.4% is due to measurement error and 1.7% reflects unobserved heterogeneity.

Cumulants

Univariate cumulants of centred random variables:

$$\begin{aligned}\kappa_2(Z) &= \mathrm{Cum}(Z,Z) = \mathrm{Var}(Z) = \mathbb{E}Z^2,\\ \kappa_3(Z) &= \mathrm{Cum}(Z,Z,Z) = \mathbb{E}Z^3,\\ \kappa_4(Z) &= \mathrm{Cum}(Z,Z,Z,Z) = \mathbb{E}\left(Z^4\right) - 3\mathbb{E}\left(Z^2\right)^2.\end{aligned}$$

Multivariate cumulants of centred random vectors:

$$\begin{aligned}\mathrm{Cum}(Y_i,Y_j) &= \mathbb{E}(Y_iY_j),\\ \mathrm{Cum}(Y_i,Y_j,Y_\ell) &= \mathbb{E}(Y_iY_jY_\ell),\\ \mathrm{Cum}(Y_i,Y_j,Y_\ell,Y_m) &= \mathbb{E}(Y_iY_jY_\ell Y_m) - \mathbb{E}(Y_iY_j)\mathbb{E}(Y_\ell Y_m)\\ &\quad -\mathbb{E}(Y_iY_\ell)\mathbb{E}(Y_jY_m) - \mathbb{E}(Y_iY_m)\mathbb{E}(Y_jY_\ell).\end{aligned}$$

Tensor or multi-linear structure.

Moment restrictions

Second order:

$$\begin{aligned}\mathrm{Cum}\left(Y_i, Y_j\right) &= \sum_{k=1}^{K} \lambda_{ik}\lambda_{jk} + \mathrm{Cov}\left(U_i, U_j\right) \\ &\Leftrightarrow \Sigma_Y = \Lambda\Lambda^T + \Sigma_U.\end{aligned}$$

Third order:

$$\mathrm{Cum}\left(Y_i, Y_j, Y_\ell\right) = \sum_{k=1}^{K} \lambda_{ik}\lambda_{jk}\lambda_{\ell k}\kappa_3\left(X_k\right) + \mathbf{1}\left\{i = j = \ell\right\}\kappa_3\left(U_i\right)$$

Fourth-order:

$$\mathrm{Cum}\left(Y_i, Y_j, Y_\ell, Y_m\right) = \sum_{k=1}^{K} \lambda_{ik}\lambda_{jk}\lambda_{\ell k}\lambda_{mk}\kappa_4\left(X_k\right) + \mathbf{1}\left\{i = j = \ell = m\right\}\kappa_4\left(U_i\right).$$

Tensor or multi-linear structure.

Matrix restrictions

Multilinear restrictions of order 2, 3, 4 in matrix form.

Let

$$\Gamma_Y(\ell) = \left[\mathrm{Cum}\left(Y_i, Y_\ell, Y_j\right); (i,j) \in \{1,...,L\}^2\right] \in \mathbb{R}^{L\times L}, \quad \ell \in \{1...L\}.$$

Then

$$\Gamma_Y(\ell) = \Lambda D_3 \,\mathrm{diag}\left(\Lambda_\ell\right) \Lambda^T + \kappa_3\left(U_\ell\right) \mathrm{Sp}_{L,\ell},$$

where

- $\Lambda_\ell^T \in \mathbb{R}^{K\times 1}$ is the ℓth row of Λ,
- $D_3 = \mathrm{diag}\left(\kappa_3\left(X_1\right), ..., \kappa_3\left(X_K\right)\right)$,
- and $\mathrm{Sp}_{L,\ell}$ is the $L \times L$ sparse matrix with only one 1 in position (ℓ,ℓ).

Let

$$\Omega_Y(\ell,m) = \left[\mathrm{Cum}\left(Y_i, Y_\ell, Y_m, Y_j\right)\right]_{i\times j} \in \mathbb{R}^{L\times L}, \quad (\ell \leq m).$$

Then,

$$\Omega_Y(\ell,m) = \Lambda D_4 \,\mathrm{diag}\left(\Lambda_\ell \odot \Lambda_m\right) \Lambda^T + \delta_{\ell m}\kappa_4\left(U_\ell\right) \mathrm{Sp}_{L,\ell},$$

where

- $D_4 = \mathrm{diag}\left(\kappa_4\left(X_1\right), ..., \kappa_4\left(X_K\right)\right)$,
- and $\odot$ is the Hadamard (element by element) matrix product.

Identification of factor loadings – noisy model ($U \neq 0$)
First case: kurtic factor distributions

Let

$$\Omega_Y = [\mathrm{Cum}\,(Y_i, Y_j, Y_\ell, Y_m)]_{(i \leq j) \times (\ell < m)} \in \mathbb{R}^{\frac{L(L+1)}{2} \times \frac{L(L-1)}{2}}.$$

Then

$$\mathrm{Cum}\,(Y_i, Y_j, Y_\ell, Y_m) = \sum_{k=1}^{K} \lambda_{ik}\lambda_{jk}\lambda_{\ell k}\lambda_{mk}\kappa_4\,(X_k)$$

implies that

$$\Omega_Y = \overline{Q} D_4 Q^T,$$

where

$$\overline{Q} \equiv \overline{Q}\,(\Lambda) = [\lambda_{i1}\lambda_{j1}, ..., \lambda_{iK}\lambda_{jK}]_{(i \leq j) \times k} \in \mathbb{R}^{\frac{L(L+1)}{2} \times K}$$
$$Q \equiv Q\,(\Lambda) = [\lambda_{\ell 1}\lambda_{m1}, ..., \lambda_{\ell K}\lambda_{mK}]_{(\ell < m) \times k} \in \mathbb{R}^{\frac{L(L-1)}{2} \times K}.$$

Lemma 1 *Assume that (i) $K \leq \frac{L(L-1)}{2}$, (ii) Q has rank K and (iv) factor variables have non zero kurtosis excess. Then matrix Ω_Y has rank K.*

First case: kurtic factor distributions (cont'ed)

Remark that, for any diagonal matrix $D = \mathrm{diag}\,(d)$, $\mathrm{vech}\left(\Lambda D \Lambda^T\right) = \overline{Q} d$, where vech stacks the non redundant elements of a symmetric matrix.

For example,

$$\begin{aligned} \mathrm{vech}\left(\Gamma_Y\left(\ell\right)\right) &= \mathrm{vech}\left(\Lambda D_3\, \mathrm{diag}\left(\Lambda_\ell\right) \Lambda^T + \kappa_3\left(U_\ell\right) \mathrm{Sp}_{L,\ell}\right) \\ &= \overline{Q} D_3 \Lambda_\ell + \kappa_3\left(U_\ell\right) \mathrm{vech}\left(\mathrm{Sp}_{L,\ell}\right). \end{aligned}$$

Let $\overline{C} \in \mathbb{R}^{\frac{L(L+1)}{2} \times \left(\frac{L(L+1)}{2} - K\right)}$ be a basis of the null space of Ω_Y. Then,

$$\overline{C}^T\, \mathrm{vech}\left(\Gamma_Y\left(\ell\right)\right) = \kappa_3\left(U_\ell\right) \overline{C}_{(\ell,\ell)},$$

where $\overline{C}_{(\ell,\ell)}$ is the (ℓ, ℓ)th column of $\overline{C}^T$.

More generally, we have the following lemma:

Lemma 2 *Assume that (i) $K \leq \frac{L(L-1)}{2}$, (ii) Q has rank K and (iv) factor variables have non zero kurtosis excess. Let $\overline{C} \in \mathbb{R}^{\frac{L(L+1)}{2} \times \left(\frac{L(L+1)}{2} - K\right)}$ be a basis of the null space of Ω_Y. Then the following propositions hold true.*

1. $\mathrm{Var}\,(U_\ell)$, $\kappa_3\,(U_\ell)$ *and* $\kappa_4\,(U_\ell)$ *solve the system:*

$$\begin{aligned}
\overline{C}^T \operatorname{vech}\left(\Sigma_Y\right) &= \sum_{\ell=1}^{L} \mathrm{Var}\left(U_\ell\right) \overline{C}_{(\ell,\ell)}, \\
\overline{C}^T \operatorname{vech}\left(\Gamma_Y\left(\ell\right)\right) &= \kappa_3\left(U_\ell\right) \overline{C}_{(\ell,\ell)}, \\
\overline{C}^T \operatorname{vech}\left(\Omega_Y\left(\ell,\ell\right)\right) &= \kappa_4\left(U_\ell\right) \overline{C}_{(\ell,\ell)}.
\end{aligned}$$

   where $\overline{C}^T_{(\ell,\ell)}$ denotes the (ℓ,ℓ)th row of $\overline{C}$.

2. *Matrix $\left[\overline{C}_{(1,1)}, ..., \overline{C}_{(L,L)}\right]$ is full rank and* $\mathrm{Var}\,(U_\ell)$, $\kappa_3\,(U_\ell)$ *and* $\kappa_4\,(U_\ell)$ *are uniquely defined.*

First case: kurtic factor distributions (cont'ed)

The following theorem then follows straightforwardly.

Theorem 3 ***(Sufficient conditions for parametric identification when $K \leq L$)*** *Assume that (i) $K \leq \min\left\{L, \frac{L(L-1)}{2}\right\}$, (ii) Λ is full column rank, (iii) $Q(\Lambda)$ has rank K, and (iv) factor variables have non zero kurtosis excess. Then, factor loadings are identified from second and fourth-order moments.*

Identification of factor loadings – noisy model ($U \neq 0$)
Second case: nonkurtic factor distributions

If some or all factor distributions may have zero kurtosis excess, search identification in third-order moments. Let

$$\begin{aligned}\Omega_Y(j) &= \left[\mathrm{Cum}\left(Y_i, Y_j, Y_\ell, Y_m\right)\right]_{i\times(\ell<m)} \in \mathbb{R}^{L\times\frac{L(L-1)}{2}} \\ &= \Lambda\,\mathrm{diag}\left(\Lambda_j\right) D_4 Q^T,\end{aligned}$$

$$\begin{aligned}\Gamma_Y &= \left[\mathrm{Cum}\left(Y_i, Y_\ell, Y_m\right)\right]_{i\times(\ell<m)} \in \mathbb{R}^{L\times\frac{L(L-1)}{2}} \\ &= \Lambda D_3 Q^T\end{aligned}$$

$$\Xi_Y = \left[\Gamma_Y, \Omega_Y(1), \ldots, \Omega_Y(L)\right]$$

Lemma 4 *Assume that (i)* $K \leq \min\left\{L, \frac{L(L-1)}{2}\right\}$*, (ii)* Λ *and* $Q(\Lambda)$ *have full column rank* K *and (iii) each factor distribution is either skewed or kurtic. Then, matrix* Ξ_Y *has rank* K*.*

Second case: nonkurtic factor distributions and... $K \leq L-1$

The identifiability of error moments then comes at the price of some additional assumptions on the matrix of factor loadings.

Lemma 5 *Assume that (i) $K \leq L-1$, (ii) every submatrix of Λ made of a selection of $L-1$ rows has rank K, (iii) $Q(\Lambda)$ have full column rank K and (iv) each factor distribution is either skewed or kurtic. Let $C \in \mathbb{R}^{L \times (L-K)}$ be a basis of the null space of Ξ_Y. Let C_ℓ^T denote the ℓth row of C. The following propositions hold true.*

1. $\operatorname{Var}(U_\ell)$, $\kappa_3(U_\ell)$ *and* $\kappa_4(U_\ell)$ *solve the system:*

$$C^T \begin{pmatrix} \operatorname{Cum}(Y_1, Y_\ell) & \operatorname{Cum}(Y_1, Y_\ell, Y_\ell) & \operatorname{Cum}(Y_1, Y_\ell, Y_\ell, Y_\ell) \\ \vdots & \vdots & \vdots \\ \operatorname{Cum}(Y_L, Y_\ell) & \operatorname{Cum}(Y_L, Y_\ell, Y_\ell) & \operatorname{Cum}(Y_L, Y_\ell, Y_\ell, Y_\ell) \end{pmatrix} = [\operatorname{Var}(U_\ell), \kappa_3(U_\ell), \kappa_4(U_\ell)]\, C_\ell.$$

2. *No column of C is nil ($C_\ell \neq 0$, $\forall \ell$) and* $\operatorname{Var}(U_\ell)$, $\kappa_3(U_\ell)$ *and* $\kappa_4(U_\ell)$ *are identified.*

Note that if, in particular, all factor distributions are skewed then one can define C as the null space of Γ_Y.

Second case: nonkurtic factor distributions $K \leq L - 1$

The following theorem then follows immediately.

Theorem 6 ***(Sufficient conditions for parametric identification when*** $K \leq L - 1$***)*** *Assume that (i)* $K \leq L - 1$*, (ii) every submatrix of* Λ *made of a selection of* $L - 1$ *rows has rank* K*, (iii) matrix* $Q(\Lambda)$ *has rank* K*, (iv) each factor is either skewed or kurtic. Then, factor loadings are parametrically identified from second, third and fourth-order moments.*

Again, if all factors are skewed then factor loadings are parametrically identified from second and third-order moments.

Corollary 7 ***(Sufficient conditions for parametric identification from second and third-order moments when*** $K \leq L - 1$***)*** *Assume that (i)* $K \leq L - 1$*, (ii) every submatrix of* Λ *made of a selection of* $L - 1$ *rows has rank* K*, (iii) matrix* $Q(\Lambda)$ *has rank* K*, (iv) all factor distributions are skewed. Then, factor loadings are parametrically identified from second and third-order moments.*

Example: the measurement error model

Model:

$$\begin{cases} Y_1 = \lambda_{11} X_1 + U_1, \\ Y_2 = \lambda_{21} X_1 + U_2, \end{cases}$$

where factor X_1 has a non symmetric distribution: $\mathbb{E}\left(X_1^3\right) \neq 0$.

Using second and third-order restrictions yields:

$$\begin{aligned} \lambda_{11} &= \sqrt{\mathbb{E}(Y_1 Y_2) \frac{\mathbb{E}(Y_1 Y_1 Y_2)}{\mathbb{E}(Y_1 Y_2 Y_2)}}, \\ \lambda_{21} &= \sqrt{\mathbb{E}(Y_1 Y_2) \frac{\mathbb{E}(Y_1 Y_2 Y_2)}{\mathbb{E}(Y_1 Y_1 Y_2)}}. \end{aligned}$$

Interestingly,

$$\frac{\lambda_{21}}{\lambda_{11}} = \frac{\mathbb{E}(Y_1 Y_2 Y_2)}{\mathbb{E}(Y_1 Y_1 Y_2)}.$$

Replacing expectations by sample means, we obtain Geary's (1942) estimator for the measurement error model:

Regress Y_2 on Y_1, with no intercept, by 2SLS, using $Y_1 Y_2$ as an instrument for Y_1.

Estimation
Number of factors

We apply Robin and Smith's (2000) rank test to various matrices.

1. **Estimating K when $K \leq \frac{L(L-1)}{2}$ and all factors are kurtic.** Assuming that Q is full column rank and that factor variables have non zero kurtosis, then

$$\text{rank}\left(\Omega_Y\right) = K, \text{ for all } K \leq \frac{L\left(L-1\right)}{2}.$$

This allows to test whether $K \leq L$ or not.

- **Refinement.** Based on matrices

$$\Omega_Y\left(\ell, m\right) = \Lambda D_4 \,\text{diag}\left(\Lambda_\ell \odot \Lambda_m\right) \Lambda^T, \quad \ell < m.$$

Let $w = (w_{1,2}, ..., w_{L-1,L})$ be a vector of $\frac{L(L-1)}{2}$ positive weights. As no column of Q is identically zero, then

$$\Omega_{Y,w} \equiv \sum_{\ell<m} w_{\ell,m} \Omega_Y\left(\ell, m\right) = \Lambda D_4 \,\text{diag}\left(Q^T w\right) \Lambda^T \quad \text{has rank } \min\left\{K, L\right\} \text{ for almost all } w.$$

2. **Estimating K when $K \leq \min\left\{L, \frac{L(L-1)}{2}\right\}$ and all factors are skewed.** Assuming that Λ and Q are full column rank and that factor variables have non zero skewness, then

$$\operatorname{rank}\left(\Gamma_Y\right) = K, \text{ for all } K \leq \min\left\{L, \frac{L\left(L-1\right)}{2}\right\}.$$

3. **Estimating K when $K \leq \min\left\{L, \frac{L(L-1)}{2}\right\}$ and all factors are either skewed or kurtic.** Assuming Λ and Q full column rank (so $K \leq \min\left\{L, \frac{L(L-1)}{2}\right\}$) and that each factor is either skewed or kurtic, then apply rank test to matrix Ξ_Y.

   - **Refinement.** Based on matrices Γ_Y and $\Omega_Y\left(j\right)$:

$$\Xi_{Y,w} \equiv \Gamma_Y + \sum_{j=1}^{L} b_j \Omega_Y\left(j\right) = \Lambda\left[D_3 + D_4 \operatorname{diag}\left(\Lambda^T \beta\right)\right] Q^T \quad \text{has rank } K \text{ for almost all } w.$$

Cardoso and Souloumiac's JADE algorithm ($U = 0,\ K = L$)

Assuming no noise, theory implies that there exist diagonal matrices $D_{4,\ell,m}$ (unspecified) such that

$$\Omega_Y(\ell, m) = \Lambda D_{4,\ell,m}\Lambda^T, \quad (\ell \leq m), \quad \text{and} \quad \Sigma_Y = \Lambda\Lambda^T.$$

Joint Diagonalisation algorithm:

1. "Whiten" the data, i.e. compute $\widetilde{Y} = P^{-1}Y$, where P is a $L \times L$ such that $PP^T = \Sigma_Y$.

2. Compute $\Omega_{\widetilde{Y}}(\ell, m)$, for all $\ell \leq m$. These matrices satisfy the restrictions:

$$V^T \Omega_{\widetilde{Y}}(\ell, m) V = D_{4,\ell,m},$$

   where $V = P^{-1}\Lambda$ is an orthonormal matrix of dimensions L.

3. Compute V as an orthonomal matrix minimising the sum of squares of the off-diagonal elements of matrices $V^T \Omega_{\widetilde{Y}}(\ell, m) V$. Then, $\Lambda = PV$.
   Cardoso and Souloumiac (1993) develop a simple and efficient algorithm to do this optimisation (inspired from standard algorithms for PCA).

Given an i.i.d. sample, the **JADE algorithm** (Joint Approximate Diagonalisation of Eigenmatrices) applies the JD algorithm to matrices of empirical moments.

Asymptotic theory for JADE

- Let $\widehat{A}_1, ..., \widehat{A}_J$ be root-N consistent and asymptotically normal estimators of J symmetric $K \times K$ matrices $A_1, ..., A_J$.
  Let $\widehat{A} = \left[\widehat{A}_1, ..., \widehat{A}_J\right]$ and $A = [A_1, ..., A_J]$.
- The JADE estimator is
$$\widehat{V} = \arg\min_{V \in \mathcal{O}_K} \sum_{j=1}^{J} \text{off}(V^T \widehat{A}_j V),$$
  where $\text{off}(M) = \sum_{i \neq j} m_{ij}^2$ and $\mathcal{O}_K$ is the set of orthonormal $K \times K$ matrices.
- Assume that $\exists! V \in \mathcal{O}_K$, $\forall j$, $V^T A_j V = D_j$, where $D_j = \text{diag}\,(d_{j1}, ..., d_{jK})$.
- Define the $K^2 \times K^2$ matrices:
$$R(D_j) = \left[\frac{(d_{jk} - d_{jm})}{\sum_{j'=1}^{J}(d_{j'k} - d_{j'm})^2}; k, m = 1, ..., J\right],$$
  and let W be the following $K^2 \times JK^2$ matrix:
$$W = [\text{diag}(\text{vec}(R(D_1))), ..., \text{diag}(\text{vec}(R(D_J)))]\,.$$

Theorem 8 *Assume that* $\sum_{j=1}^{J}(d_{jk} - d_{jm})^2 \neq 0$ *for all* $k \neq m$. *Then*

$$N^{1/2}\left(\text{vec}(\widehat{V}) - \text{vec}(V)\right) \underset{N\to\infty}{\overset{L}{\to}} \mathcal{N}\left(0, \text{Var}\left(\text{vec}(\widehat{V})\right)\right),$$

where:

$$\text{Var}\left(\text{vec}(\widehat{V})\right) = (I_K \otimes V)W(I_J \otimes V^T \otimes V^T)\,\text{Var}\left(\text{vec}(\widehat{A})\right)(I_J \otimes V \otimes V)W^T(I_K \otimes V^T).$$

- When $J = 1$, yields variance-covariance matrix of the eigenvectors of a symmetric matrix (*e.g.* Anderson, 1963).
  The diagonal coefficients of matrix W are equal to $1/(d_{1k} - d_{1m})$, for $k \neq m$.
  The variance of eigenvectors thus blows up (model not identified) when two eigenvalues of A_1 get close to each other.

- If $J > 1$, variance blows up if $\sum_j (d_{jk} - d_{jm})^2 \to 0$.
  For example, if $A_j \equiv \Omega_Y(\ell, m)$, then $D_j = D_4\,\text{diag}(\Lambda_\ell \odot \Lambda_m)$.
  Variance blows up if $\exists k, k'$ such that $d_{jk} = d_{jk'}$ for all j, or

$$\lambda_{\ell k}\lambda_{mk}\kappa_4(X_k) = \lambda_{\ell k'}\lambda_{mk'}\kappa_4(X_{k'}), \forall \ell, m.$$

  This cannot happen if model is identified, i.e. **at most one factor has zero kurtosis excess** and if **no couple of columns of** Λ **are proportional to each other**.

Practical recommendation:

- Use bootstrap to compute stds of estimates. Matrix Var $\left(\mathrm{vec}(\widehat{A})\right)$ difficult to compute or imprecisely estimated by resampling:

N	500	1000	5000	10000	∞
κ_3	4.51 (1.98)	5.01 (2.36)	5.73 (2.65)	5.89 (2.02)	6.18
κ_4	36.1 (38.4)	48.6 (62.4)	77.0 (132.3)	83.3 (104.7)	110.9

Log-normal distribution

- Weight each matrix $\widehat{A}_j$ by average precision (one over the sqrt of the sum of the variances of the elements of $\widehat{A}_j$).

Quasi-JADE algorithm

1. Estimate matrices $C \in \mathbb{R}^{L\times(L-K)}$ and/or $\overline{C} \in \mathbb{R}^{\#\overline{\Delta}_{L,2}\times\left(\#\overline{\Delta}_{L,2}-K\right)}$ of Lemmas 2 and 5 by Singular Value Decomposition.

2. Estimate $\mathrm{Var}\,(U_\ell)$, $\kappa_3\,(U_\ell)$ and/or $\kappa_4\,(U_\ell)$ using the restrictions in Lemmas 2 and 5.

3. Proceed to the JD of matrices

$$P^{-}\left[\Gamma_Y\,(\ell) - \kappa_3\,(U_\ell)\,\mathrm{Sp}_{L,\ell}\right]P^{-T} \quad \text{and/or} \quad P^{-}\left[\Omega_Y\,(\ell,m) - \delta_{\ell m}\kappa_4\,(U_\ell)\,\mathrm{Sp}_{L,\ell}\right]P^{-T},$$

   where P is a full column rank $L\times K$ matrix such that

$$\Sigma_Y - \Sigma_U = PP^T.$$

   Let V be the orthonormal matrix of joint eigenvectors. Then $\Lambda = PV$.

4. Estimate factor cumulants $\kappa_3\,(X_k)$ and $\kappa_4\,(X)$ by OLS from restrictions:

$$\begin{aligned}\left[V^TP^{-}\left[\Gamma_Y\,(\ell) - \kappa_3\,(U_\ell)\,\mathrm{Sp}_{L,\ell}\right]P^{-T}V\right]_{k,k} &= \lambda_{\ell k}\kappa_3\,(X_k)\,,\\ \left[V^TP^{-}\left[\Omega_Y\,(\ell,m) - \delta_{\ell m}\kappa_4\,(U_\ell)\,\mathrm{Sp}_{L,\ell}\right]P^{-T}V\right]_{k,k} &= \lambda_{\ell k}\lambda_{mk}\kappa_4\,(X)\,,\end{aligned}$$

   where we denote as $[A]_{i,j}$ the (i,j) entry of matrix A.

Simulations: Convergence of the quasi-JADE estimator

N	500	1000	5000	10000
λ_{11}	2.03 (.28)	2.03 (.17)	2.01 (.09)	2.01 (.06)
λ_{21}	.95 (.23)	.99 (.14)	1.00 (.07)	1.00 (.05)
λ_{31}	.95 (.23)	.99 (.15)	.99 (.07)	1.00 (.05)
λ_{12}	.98 (.23)	.98 (.15)	1.00 (.06)	1.00 (.05)
λ_{22}	2.05 (.27)	2.03 (.19)	2.01 (.08)	2.01 (.07)
λ_{32}	.97 (.23)	.98 (.17)	1.00 (.06)	1.00 (.05)
λ_{13}	.97 (.23)	.98 (.15)	.99 (.06)	1.00 (.05)
λ_{23}	.97 (.23)	.98 (.16)	1.00 (.06)	1.00 (.05)
λ_{33}	2.06 (.27)	2.02 (.19)	2.01 (.09)	2.00 (.05)
$\text{Var}(U_1)$	.77 (.59)	.87 (.43)	.96 (.20)	.98 (.16)
$\text{Var}(U_2)$	.76 (.57)	.87 (.43)	.98 (.20)	.98 (.17)
$\text{Var}(U_3)$	.74 (.56)	.86 (.42)	.96 (.20)	.98 (.16)

Quasi-JADE algorithm based on 2nd, 3rd and 4th moments, assuming all factors kurtic. Log-normal factors, standard normal errors.

Simulations: Robustness of Quasi-JADE to noise

$\text{Var}(U_\ell)$	.01	.25	1	4	.01	.25	1	4
λ_{11}	2.00 (.07)	2.11 (.08)	2.36 (.12)	2.81 (.46)	1.98 (.12)	2.01(.13)	2.03 (.17)	2.02 (.44)
λ_{21}	1.00 (.11)	1.00 (.12)	.95 (.24)	.72 (.86)	1.00 (.15)	.99 (.12)	.99 (.14)	.95 (.31)
λ_{31}	1.00 (.11)	1.03 (.14)	1.08 (.22)	1.05 (.77)	1.00 (.16)	.99 (.13)	.99 (.15)	.95 (.32)
λ_{12}	1.00 (.11)	1.00 (.12)	.97 (.24)	.78 (.86)	1.00 (.16)	.99 (.13)	.98 (.15)	.97 (.33)
λ_{22}	2.00 (.07)	2.11 (.07)	2.37 (.12)	2.86 (.32)	1.97 (.11)	2.02 (.11)	2.03 (.19)	2.02 (.41)
λ_{32}	1.00 (.12)	1.03 (.13)	1.08 (.22)	1.08 (.76)	.99 (.16)	.99 (.13)	.98 (.17)	.97 (.32)
λ_{13}	1.00 (.11)	.87 (.13)	.61 (.20)	.16 (.69)	1.00 (.16)	1.00 (.14)	.98 (.15)	.96 (.32)
λ_{23}	1.00 (.11)	.87 (.12)	.62 (.20)	.15 (.67)	1.00 (.16)	1.00 (.13)	.98 (.16)	.96 (.32)
λ_{33}	2.00 (.08)	2.02 (.09)	2.13 (.16)	2.52 (.43)	1.98 (.11)	2.02 (.11)	2.02 (.19)	2.01 (.42)
$\text{Var}(U_1)$					.04 (.11)	.18 (.22)	.87 (.43)	3.77 (.98)
$\text{Var}(U_2)$					.04 (.11)	.17 (.23)	.87 (.43)	3.77 (.94)
$\text{Var}(U_3)$					.04 (.11)	.17 (.22)	.86 (.42)	3.77 (.97)
	JADE				Quasi-JADE			

Log-normal factors (variance = 4.67), standard normal errors, $N = 1000$.

Simulations: Role of factor kurtosis

Factors are normal mixtures ($N(0, 1/2)$ w.prob. ρ and $N(0, (2-\rho)/(2-2\rho))$ w.p. $1-\rho$). Normal errors. $N = 1000$.

$\kappa_4(\rho)$ (ρ)	$-6/5$ (-1)	$1/2$ $\left(\frac{2}{5}\right)$	1 $\left(\frac{4}{7}\right)$	5	10 $\left(\frac{40}{43}\right)$	100 $\left(\frac{400}{403}\right)$	≈ 110 (1)
λ_{11}	1.94 (.48)	1.66 (.78)	1.76 (.74)	2.03 (.33)	2.01 (.26)	2.01 (.19)	2.03 (20)
λ_{21}	.91 (.48)	.97 (.71)	.94 (.63)	.97 (.30)	.98 (.21)	.99 (.16)	.98 (.15)
λ_{31}	.92 (.48)	1.00 (.69)	.96 (.65)	.97 (.29)	.97 (.21)	.98 (.17)	.98 (.16)
λ_{12}	.97 (.49)	1.00 (.71)	.98 (.65)	.96 (.30)	.98 (.21)	.99 (.19)	.98 (.16)
λ_{22}	1.98 (.44)	1.71 (.69)	1.83 (.64)	2.02 (.35)	2.02 (.26)	2.01 (.18)	2.03 (.18)
λ_{32}	.98 (.49)	1.00 (.72)	.95 (.66)	.97 (.30)	.98 (.20)	.99 (.18)	.98 (.16)
λ_{13}	.96 (.49)	1.12 (.74)	1.05 (.70)	.97 (.29)	.99 (.20)	.99 (.17)	.98 (.15)
λ_{23}	.94 (.49)	1.12 (.75)	1.05 (.69)	.97 (.29)	.98 (.19)	.99 (.18)	.98 (.15)
λ_{33}	1.97 (.43)	1.83 (.57)	1.89 (.56)	2.03 (.32)	2.03 (.25)	2.02 (.18)	2.03 (.20)
$\text{Var}(U_1)$	.71 (.65)	.92 (.84)	.76 (.79)	.77 (.63)	.88 (.53)	.92 (.40)	.86 (.44)
$\text{Var}(U_2)$	.75 (.65)	.89 (.83)	.69 (.78)	.75 (.64)	.83 (.55)	.93 (.40)	.87 (.43)
$\text{Var}(U_3)$	.74 (.66)	.93 (.82)	.76 (.80)	.77 (.64)	.84 (.53)	.91 (.40)	.86 (.44)

Simulations: $K = 2,\ L = 3$

N	500	500	1000	1000	5000	5000
Cumulants	2,3,4	2,3	2,3,4	2,3	2,3,4	2,3
λ_{11}	1.95 (.28)	1.93 (.32)	1.98 (.19)	1.97 (.24)	2.00 (.08)	2.00 (.08)
λ_{21}	1.96 (.30)	1.91 (.37)	1.99 (.16)	1.96 (.23)	1.00 (.09)	2.00 (.05)
λ_{31}	.97 (.23)	.98 (.25)	.98 (.17)	.98 (.20)	1.00 (.08)	1.00 (.08)
λ_{12}	2.02 (.24)	2.03 (.27)	2.01 (.17)	2.01 (.20)	1.00 (.08)	2.00 (.08)
λ_{22}	1.02 (.28)	1.05 (.32)	1.00 (.18)	1.02 (.22)	2.00 (.09)	1.00 (.08)
λ_{32}	2.01 (.12)	1.99 (.14)	2.01 (.10)	2.00 (.11)	1.00 (.05)	2.00 (.05)
Var(U_1)	.98 (.21)	1.01 (.16)	.98 (.15)	1.00 (.13)	.97 (.09)	1.00 (.06)
Var(U_2)	.94 (.21)	.99 (.20)	.96 (.15)	1.00 (.15)	.97 (.08)	1.00 (.07)
Var(U_3)	.94 (.22)	1.00 (.20)	.96 (.15)	1.00 (.15)	.98 (.09)	1.00 (.07)

Simulations: Lots of factors

	$L=K=5$			$L=K=10$		
N	500	1000	5000	500	1000	5000
λ_{11}	2.06 (.41)	2.03 (.28)	2.01 (.13)	1.85 (.72)	1.97 (.56)	2.00 (.27)
λ_{21}	.95 (.35)	.98 (.25)	.99 (.12)	.89 (.52)	.90 (.43)	.98 (.22)
λ_{31}	.95 (.34)	.98 (.24)	1.00 (.12)	.88 (.53)	.90 (.45)	.98 (.23)
λ_{41}	.95 (.35)	.98 (.24)	.99 (.11)	.88 (.53)	.92 (.43)	.98 (.22)
λ_{51}	.95 (.34)	.98 (.24)	.99 (.12)	.88 (.53)	.90 (.43)	.98 (.22)
λ_{61}				.88 (.54)	.91 (.43)	.98 (.22)
λ_{71}				.89 (.53)	.90 (.44)	.98 (.22)
λ_{81}				.88 (.52)	.90 (.44)	.98 (.23)
λ_{91}				.87 (.53)	.91 (.44)	.98 (.23)
$\lambda_{10,1}$				.88 (.52)	.89 (.44)	.98 (.22)
$\mathrm{Var}(U_1)$	.58 (.56)	.81 (.44)	.95 (.20)	.40 (.55)	.49 (.53)	.88 (.28)

Simulations: Rank test based on Ω_Y, Size

The true value of Λ is $\begin{pmatrix} 2 & 2 \\ 2 & 1 \\ 1 & 2 \end{pmatrix}$. Factors are normal mixtures. Errors are normal. $N = 1000$.

ρ	-	2/5	4/7	20/23	40/43	400/403
$\kappa_4(\rho)$	-6/5	1/2	1	5	10	100
α =.10	.90	.73	.82	.87	.85	.62
α =.20	.79	.57	.67	.74	.69	.43
α =.30	.67	.44	.54	.61	.57	.29
α =.40	.58	.33	.42	.50	.45	.19
α =.50	.47	.24	.32	.40	.35	.11
α =.60	.37	.16	.22	.32	.26	.05
α =.70	.27	.10	.13	.24	.19	.02
α =.80	.20	.05	.08	.15	.11	.01
α =.90	.10	.02	.04	.06	.04	.00

Simulations: Rank test for $K \leq L$, Size

The true value of Λ is $\begin{pmatrix} 2 & 2 \\ 2 & 1 \\ 1 & 2 \end{pmatrix}$. Log-normal factors. Standard normal errors. $N = 1000$.

Matrix	Ω_Y	$\sum_{\ell<m} w_{\ell m}\Omega_Y(\ell,m)$	Γ_Y
$\alpha = .10$	.56	.87	.90
$\alpha = .20$	.34	.71	.79
$\alpha = .30$	.20	.56	.69
$\alpha = .40$	.12	.44	.58
$\alpha = .50$	.08	.32	.48
$\alpha = .60$	.05	.21	.38
$\alpha = .70$	.02	.13	.29
$\alpha = .80$	.01	.06	.16
$\alpha = .90$	.00	.01	.07

Simulations: Rank test for $K \leq L$, Power

Tests $K = 2$ against $K = 3$. True value is $\Lambda = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}$. Factors are normal mixtures. Errors are normal. $N = 1000$.

$\kappa_4(\rho)$	$-6/5$	$1/2$	1	5	10	100
$\alpha = .10$	.99	.81	.81	1.00	1.00	.89
$\alpha = .20$	.99	.63	.66	1.00	1.00	.80
$\alpha = .30$	.98	.68	.51	.99	1.00	.72
$\alpha = .40$	.97	.36	.39	.99	1.00	.64
$\alpha = .50$	.96	.26	.29	.98	.99	.56
$\alpha = .60$	.94	.18	.22	.96	.98	.47
$\alpha = .70$	.93	.11	.16	.92	.96	.35
$\alpha = .80$	.89	.06	.10	.86	.90	.22
$\alpha = .90$	.83	.02	.04	.72	.77	.12

Empirical application: Stocks

Fama and French (1993) identify three factors explaining a large proportion of the variance of time-series of U.S. excess stock returns, $Y_\ell(t) = R_\ell(t) - R_F(t)$, $\ell = 1, ..., L$.

In addition to the market return ($R_M(t) - R_F(t)$, where $R_F(t)$ is the risk-free return), which is the unique factor of the CAPM model, they identify two additional factors:

- $SMB(t)$, or "small minus big", is the difference between the average of the returns on two stock portfolios: one containing firms with market value (price time number of shares) less than the median, and one containing firms with size above the median.
- $HML(t)$, or "high minus low", is the difference between the average of the returns on two stock portfolios: one gathering firms with book-to-market ratio (book value of capital divided by market value, denoted B/M) less that the 30th percentile and another one containing all firms with B/M ratio above the 70th percentile.

Fama and French show that these three factors explain monthly data on 25 portfolios formed by intersecting size and book-to-market quintiles remarkably well.

Size	B/M ratio	Factor 1	Factor 2	Factor 3	Error variance	R^2
Small	Low	.57	.79	-.37	.055	.95
	2	.50	.67	-.23	.054	.93
	3	.45	.56	-.14	.035	.93
	4	.41	.52	-.09	.029	.94
	High	.43	.52	-.05	.028	.94
2	Low	.70	.72	-.50	.102	.92
	2	.59	.58	-.26	.049	.94
	3	.54	.51	-.17	.066	.89
	4	.53	.47	-.10	.079	.84
	High	.61	.51	-.08	.118	.82
3	Low	.71	.61	-.58	.122	.90
	2	.62	.47	-.30	.050	.93
	3	.58	.40	-.15	.061	.88
	4	.60	.37	-.10	.080	.84
	High	.70	.40	-.06	.095	.85
4	Low	.74	.46	-.61	.094	.92
	2	.70	.33	-.28	.041	.94
	3	.69	.29	-.16	.038	.94
	4	.68	.28	-.08	.086	.84
	High	.77	.30	-.06	.146	.79

	X_1	X_2	X_3
$R_M - R_F$	.84	.24	-.41
SMB	-.49	.85	-.09
HML	-.11	.23	.90

a) Fama French

	X_1	X_2	X_3
$R_M - R_F$	.97	-.01	-.09
SMB	-.01	.98	-.06
HML	.09	.07	.93

b) Independent Fama French

Conclusion

Nonparametric estimation of factor densities. Done in a companion paper.

Allows to predict factors X_{kt}, $k = 1, ..., K, t = 1, ..., L$:

$$\mathbb{E}\left[g(X)|Y = y\right] = \frac{\int g(x) f_U(y - \Lambda x) f_X(x) dx}{\int f_U(y - \Lambda x) f_X(x) dx}.$$

In the future, we plan to pursue our research in the following directions.

- First, there is plenty room for efficiency improvements of quasi-JADE and of the test for the number of factors.
- A second direction of research concerns the extension of existing algorithms to deal with more factors than measurements ($K > L$). In the ICA literature, this case is refered to as *overcomplete* ICA.
- Address the issue of dynamics (MA factors and errors).
- Nonlinearities.