Sparse Recovery under Matrix Uncertainty

Alexandre Tsybakov,
joint work with Mathieu Rosenbaum

Laboratoire de Statistique, CREST
and
Laboratoire de Probabilités et Modèles Aléatoires,
Université Paris 6

Berlin, December 6, 2008

Matrix uncertainty (MU) model

Consider the model

$$\begin{aligned} y &= X\theta^* + \xi, \\ Z &= X + \Xi. \end{aligned}$$

- The random vector $y \in \mathbb{R}^n$ and the random $n \times p$ matrix Z are observed
- The $n \times p$ matrix X is unknown
- Ξ is an $n \times p$ random noise matrix, $\xi \in \mathbb{R}^n$ is a noise independent of Ξ
- $\theta^* = (\theta_1^*, \ldots, \theta_p^*)$ is an unknown vector of parameters.
- Possibly $p \gg n$ and θ^* is s-sparse.

Assumptions on the model

We assume that ξ and Ξ are deterministic and satisfy the assumptions:

$$\left|\frac{1}{n} Z^T \xi\right|_\infty \leq \varepsilon, \quad (1)$$

$$|\Xi|_\infty \leq \delta \quad (2)$$

for some $\varepsilon \geq 0, \delta \geq 0$. Here $|\cdot|_\infty$ stands for the maximum of components norm.

Assumptions on the model

We assume that ξ and Ξ are deterministic and satisfy the assumptions:

$$\left|\frac{1}{n}Z^T\xi\right|_\infty \leq \varepsilon, \tag{1}$$

$$|\Xi|_\infty \leq \delta \tag{2}$$

for some $\varepsilon \geq 0, \delta \geq 0$. Here $|\cdot|_\infty$ stands for the maximum of components norm.

If ξ and Ξ are random, these assumptions are satisfied with a probability close to 1 in many interesting cases.

Noise levels ε and δ

- $\xi \sim \mathcal{N}(0, \sigma^2 I) \Longrightarrow$ take $\varepsilon = A\sigma\sqrt{\frac{\log p}{n}}$ for some $A > \sqrt{2}$.
  Then condition (1) holds with probability at least $1 - p^{1-A^2/2}$.
  Similar choice of ε for subgaussian ξ.
- The components ξ_i of ξ are with $E(\xi_i) = 0$, $E(\xi_i^2) \le \sigma^2 < \infty$;

$$\frac{1}{n}\sum_{i=1}^{n} \max_{j=1,\dots,p} |X_{ij}|^2 \le c < \infty$$

where X_{ij} are entries of X. Then condition (1) holds with

$$\varepsilon = A\sqrt{\frac{(\log p)^{1+\gamma}}{n}},$$

with probability at least $1 - O\left((\log p)^{-\gamma}\right)$ (Lounici, 2008).

Noise levels ε and δ

- Noise level δ. Models with repeated measurements: Z is either an average of several observed matrices with mean X, or Z an empirical covariance matrix, with X as a population covariance matrix (in the latter case $p = n$). Then the threshold δ is defined in similar terms as ε.
- Noise level δ. Models with missing data.

Example 1. Models with missing data

Assume that the elements Z_{ij} of matrix Z satisfy

$$Z_{ij} = X_{ij}\eta_{ij} \tag{3}$$

where X_{ij} are the elements of X and η_{ij} are i.i.d. Bernoulli random variables taking value 1 with probability $1 - \pi$ and 0 with probability π, $0 < \pi < 1$.

- The data X_{ij} is missing if $\eta_{ij} = 0$, which happens with probability π. We are mainly interested in the case of small π.
- In practice it is easy to estimate π by the empirical probability of occurrences of zeros in the sample of Z_{ij}, so it is realistic to assume that π known.

Example 1. Models with missing data

We can rewrite (3) in the form

$$Z'_{ij} = X_{ij} + \xi'_{ij}$$

where $Z'_{ij} = Z_{ij}/(1-\pi)$, $\xi'_{ij} = X_{ij}(\eta_{ij} - E(\eta_{ij}))/(1-\pi)$. Thus, we can reduce the model with missing data (3) to the form

$$Z' = X + \Xi'$$

where the entries ξ'_{ij} of Ξ' are bounded random variables with zero means and variance $O(\pi|X|_\infty)$ for small π. Thus, with a probability close to 1, we have that $|\Xi|_\infty = O(\pi\sqrt{\log(pn)}|X|_\infty)$ for small π.

"Blind" Lasso/Dantzig selector

Lasso estimator:

$$\widehat{\theta}^L = \arg\min_{\theta\in\mathbb{R}^p} \left\{ |y - \mathbf{Z}\theta|_2^2 + r|\theta|_1 \right\},$$

where $|\theta|_q^q = \sum_{j=1}^q |\theta_j|$, $r > 0$ a tuning parameter, typically $r \sim 2\varepsilon$.

Dantzig selector (Candes and Tao, 2007):

$$\widehat{\theta}_D \triangleq \arg\min \left\{ |\theta|_1 : \left| \frac{1}{n}\mathbf{Z}^T(y - \mathbf{Z}\theta) \right|_\infty \leq 2\varepsilon \right\}.$$

Missing data

	$\pi = 0$	$\pi = 0.01$	$\pi = 0.02$	$\pi = 0.03$
$s = 1$	1.00 (0.00)	18.00 (25.36)	43.75 (27.04)	51.56 (25.49)
$s = 2$	2.00 (0.00)	41.29 (26.70)	61.40 (16.74)	65.62 (16.78)
$s = 3$	3.00 (0.00)	50.21 (24.31)	65.96 (15.66)	75.03 (10.12)

Empirical mean and standard deviation of the number of coefficients bigger than 10^{-2} for the Lasso

Missing data

	$\pi = 0$	$\pi = 0.01$	$\pi = 0.02$	$\pi = 0.03$
$s = 1$	1.00	0.58	0.15	0.12
$s = 2$	1.00	0.22	0.02	0.01
$s = 3$	1.00	0.08	0.00	0.00

Proportion of simulations where the sparsity pattern is exactly recovered, Lasso estimator.

	$\pi = 0$	$\pi = 0.02$	$\pi = 0.04$	$\pi = 0.06$
$s = 1$	1.00	0.21	0.02	0.01
$s = 3$	1.00	0.01	0.00	0.00
$s = 5$	1.00	0.00	0.00	0.00

Proportion of simulations where the sparsity pattern is exactly recovered, Dantzig selector.

Example 2. Portfolio replication

Replicating a hedge fund portfolio means obtaining a profit and loss profile similar to those of the hedge fund without investing in it.

We observe the daily absolute returns y_i, $i = 1, \dots, T$, of a portfolio (difference between the close price and the open price on day i). Theoretically:

$$y_i = \sum_{j=1}^{p} \theta_j X_{ij},$$

where p is the total number of assets in the portfolio, X_{ij} is the return of the j-th asset belonging to the portfolio on day i and θ_j its quantity. In practice: a measurement error between y_i and $\sum_{j=1}^{p} \theta_j X_{ij}$, which leads us to linear regression + measurement error in the matrix of returns $X = (X_{ij})_{i,j}$.

Example 3. Inverse problems with unknown operator

Recover an unknown function f that belongs to a Hilbert space H based on

$$Y = Af + \zeta$$

where $A : H \to V$ is a linear operator, V is a Hilbert space, ζ is a random variable with values in V.

Expansion with two bases (ϕ_j), (ψ_i) + truncation $\Longrightarrow$

$$Y = X\theta^* + \xi$$

$X = ((A\phi_j, \psi_i)_{i=1,\dots,n, j=1,\dots,p})$, and the vectors $y = (Y_1, \dots, Y_n)$, $\xi = (\xi_1, \dots, \xi_n)$.

Example 3. Inverse problems with unknown operator

In applications operator A is often not known but its action on any given function in H can be observed with a relatively small noise. Thus, we have noisy observations of the matrix $X \Longrightarrow$ Matrix Uncertainty model.

Efromovich/Koltchinskii (2001), Cavalier/Hengartner (2005), Cavalier/Raimondo (2007), Hoffmann/Reiss (2008), Marteau (2007) consider the case $n = p$ and non-degenerate X. Not always satisfying to assume, especially if n and p are very large. Our approach covers $n = p$ with degenerate matrices X that satisfy some regularity assumptions. It also covers the case $p \gg n$, which is a useful extension because by taking a large p we can assure that the residual r is indeed negligible.

Summary

- A non-asymptotic approach to errors-in-variables models, free of classical indentifiability constraints, $p \gg n$.
- Extension of ℓ_1-based sparse recovery beyond the often prohibitive restricted isometry/restricted eigenvalue conditions.
- Simple and efficient way of sparse recovery in several specific problems, such as models with missing data, inverse problems with unknown operator or some financial models (portfolio selection, portfolio replication).

Linear equation with noisy matrix

No noise in observations, $\xi = 0$. Thus, we solve

$$y = X\theta,$$

where X is an unknown matrix such that we can observe its noisy values

$$Z = X + \Xi,$$

where Ξ satisfies $|\Xi|_\infty \leq \delta$.

- Let Θ be a given convex subset of $\mathbb{R}^p$.
- We will assume that there exists an s-sparse solution θ_s of $y = X\theta$ such that $\theta_s \in \Theta$.

Linear equation with noisy matrix, *MU*-selector

Define the estimator $\hat{\theta}$ of θ_s by:

$$\hat{\theta} = \mathrm{argmin}\{|\theta|_1 : \ \theta \in \Theta, \ |y - Z\theta|_\infty \le \delta|\theta|_1\}. \tag{4}$$

(We denote by $|x|_r$, $r \ge 1$, the ℓ_r-norm of $x \in \mathbb{R}^d$ whatever is $d \ge 1$.)

This is a convex minimization problem. If $\Theta = \mathbb{R}^p$ or if Θ is a linear subspace of $\mathbb{R}^p$, a simplex, a cone, we have a linear programming problem.

We will call solutions of (4) the non-noisy (or pure) matrix uncertainty selectors (shortly **non-noisy *MU*-selectors**).

MU-selector: Existence

The feasible set of problem (4):

$$\Theta_1 = \{\theta \in \Theta : |y - Z\theta|_\infty \le \delta|\theta|_1\}$$

is non-empty. In fact, Θ_1 contains at least θ_s, since

$$|y - Z\theta_s|_\infty = |\Xi\theta_s|_\infty \le |\Xi|_\infty |\theta_s|_1 \le \delta|\theta_s|_1.$$

Thus, there always exists a solution $\hat{\theta}$ of (4). But it is not necessarily unique.

Restricted eigenvalue assumption

For a vector $\mathbf{\Delta} = (a_j)_{j=1,\dots,M}$ and a subset of indices $J \subseteq \{1, \dots, M\}$ write

$$\mathbf{\Delta}_J = (a_j \mathbf{1}\{j \in J\})_{j=1,\dots,M}.$$

The Gram matrix: $\Psi = X^T X / n$.

Assumption RE(s). (Bickel, Ritov and T., 2007)

There exists $\kappa > 0$:

$$\mathbf{\Delta}^T \Psi \mathbf{\Delta} \geq \kappa |\mathbf{\Delta}_J|_2^2$$

for all $J \subseteq \{1, \dots, p\}$ such that $|J| \leq s$ and $|\mathbf{\Delta}_{J^c}|_1 \leq |\mathbf{\Delta}_J|_1$.

More specific assumptions

Assumption RE is more general than several other assumptions on the Gram matrix:

- Coherence assumption (Donoho/Elad/Temlyakov),
- Restricted Isometry, "Uniform uncertainty principle" (Candes/Tao),
- Incoherent design assumption (Meinshausen/Yu, Zhang/Huang).

Coherence assumption

Assumption C. All the diagonal elements of the matrix $\Psi = X^T X/n$ are qual to 1 and all its off-diagonal elements $\Psi_{ij}, i \neq j$, satisfy the coherence condition:

$$\max_{i \neq j} |\Psi_{ij}| \leq \rho$$

with some $\rho < 1$.

Remark: Assumption C with

$$\rho < \frac{1}{3\alpha s}$$

implies Assumption RE(s) with

$$\kappa = \sqrt{1 - 1/\alpha}.$$

Theorem 1

Assume that there exists an s-sparse solution $\theta_s \in \Theta$ of the equation $y = X\theta$. Then for any non-noisy MU-selector $\hat{\theta}$:

$$\frac{1}{n}|X(\hat{\theta}-\theta_s)|_2^2 \quad \le \quad 4\delta^2|\hat{\theta}|_1^2.$$

If Assumption RE(s) holds, then

$$|\hat{\theta}-\theta_s|_1 \quad \le \quad \frac{4\sqrt{s}\delta}{\kappa}|\hat{\theta}|_1.$$

If Assumption RE(2s) holds, then

$$|\hat{\theta}-\theta_s|_2 \quad \le \quad \frac{4\delta}{\kappa}|\hat{\theta}|_1.$$

Theorem 1 (cont'd)

If Assumption C holds with $\rho < \frac{1}{3\alpha s}$, $\alpha > 1$, then

$$|\hat{\theta} - \theta_s|_\infty \; < \; 2\left(1 + \frac{2}{3\sqrt{s\alpha(\alpha-1)}}\right)\delta|\hat{\theta}|_1.$$

Remarks

1. We can replace $|\hat{\theta}|_1$ by $|\theta_s|_1$ in all the inequalities of Theorem 1.

2. It is straightforward to deduce a bound for $|\hat{\theta} - \theta_s|_r$, $\forall\ r \geq 1$ from the bounds of Theorem 1.

Selection of sparsity pattern

Define the thresholded estimator $\tilde{\theta} = (\tilde{\theta}_1, \ldots, \tilde{\theta}_p)$ where

$$\tilde{\theta}_j = \hat{\theta}_j I\{|\hat{\theta}_j| > \tau\}, \quad j = 1, \ldots, p, \tag{5}$$

with the data-dependent threshold

$$\tau = C_*(\alpha)\delta|\hat{\theta}|_1$$

for $C_*(\alpha) = 2\left(1 + \frac{2}{3\sqrt{\alpha(\alpha-1)}}\right)$ and some $\alpha > 1$.

Since the *MU*-selector $\hat{\theta}$ is, in general, not unique, the thresholded estimator $\tilde{\theta}$ is neither necessarily unique.

Selection of sparsity pattern

Denote by $J(\theta)$ the set of non-zero coordinates of θ.

Theorem 2

Assume that $\theta_s \in \Theta$ is an s-sparse solution of $y = X\theta$, and that $\Theta \subseteq \{\theta \in \mathbb{R}^p : |\theta|_1 \le a\}$ for some $a > 0$. Let Assumption C hold with $\rho < (3\alpha s)^{-1}$ for some $\alpha > 1$. If

$$\min_{j \in J(\theta_s)} |\theta_{sj}| > C_*(\alpha)\delta a,$$

then

$$\operatorname{sign} \tilde{\theta}_j = \operatorname{sign} \theta_{sj}, \quad j = 1, \dots, p.$$

for all $\tilde{\theta}$.

Selection of sparsity pattern

Remark. Under Assumption C with $\rho < (3\alpha s)^{-1}$ as required in Theorem 2, the s-sparse solution is unique, cf. Lounici (2008), so that the right hand side of

$$\operatorname{sign} \tilde{\theta}_j = \operatorname{sign} \theta_{sj}, \quad j = 1, \ldots, p. \tag{6}$$

is uniquely defined. The estimator $\tilde{\theta}$ is not necessarily unique, nevertheless Theorem 2 assures that the sign recovery property (6) holds for all versions of $\tilde{\theta}$.

Noisy case: Definition of *MU*-selector

Let $\xi \neq 0$. Then define the *MU*-selector as:

$$\hat{\theta} = \mathrm{argmin}\{|\theta|_1 : \ \theta \in \Theta, \left|\frac{1}{n} Z^T (y - Z\theta)\right|_\infty \leq (1+\delta)\delta|\theta|_1 + \varepsilon\}. \quad (7)$$

- For $\delta = 0$ and $\Theta = \mathbb{R}^p$ we get the Dantzig selector.
- (7) is a convex minimization problem and it reduces to linear programming if $\Theta = \mathbb{R}^p$ or if Θ is a linear subspace of $\mathbb{R}^p$ or a simplex.
- The feasible set of (7) is non-empty since it contains the true vector θ^*.
- The solution of (7) is not necessarily unique.

Theorem 3

Let the true parameter $\theta^* = \theta_s$ be s-sparse and let $\theta^* \in \Theta$. Let all the diagonal elements of $X^T X/n$ be equal to 1. Set

$$\nu = 2(2+\delta)\delta|\theta_s|_1 + 2\varepsilon.$$

Then, under Assumption RE(s) for MU-selector $\hat{\theta}$:

$$\begin{aligned} |\hat{\theta} - \theta_s|_1 &\leq \frac{4\nu s}{\kappa^2}, \\ \frac{1}{n}|X(\hat{\theta} - \theta_s)|_2^2 &\leq \frac{4\nu^2 s}{\kappa^2}. \end{aligned}$$

Under Assumption RE(2s):

$$|\hat{\theta} - \theta_s|_r^r \leq \left(\frac{4\nu}{\kappa^2}\right)^r s, \quad \forall 1 \leq r \leq 2,$$

Theorem 3 (cont'd)

Under Assumption C with $\rho < \frac{1}{3\alpha s}$, $\alpha > 1$*:*

$$|\hat{\theta} - \theta_s|_\infty \quad < \quad \frac{3\alpha + 1}{3(\alpha - 1)}\,\nu. \tag{8}$$

Remarks

1. It is straightforward to get a bound for $|\hat{\theta} - \theta_s|_r$, $\forall\ r > 2$ from the bounds of Theorem 3.
2. If $\delta = 0$ and $\Theta = \mathbb{R}^p$ we retrieve the corresponding results in Bickel, Ritov and T. (2007) for Dantzig selector.

3. If $\Theta \subseteq \{\theta \in \mathbb{R}^p : |\theta|_1 \leq a\}$ for some $a > 0$, then (8) is less than

$$\tau = \frac{3\alpha + 1}{3(\alpha - 1)}\Big(2\varepsilon + 2(2 + \delta)\delta a\Big).$$

Selection of sparsity pattern: noisy case

Theorem 4

Let the true parameter $\theta^* = \theta_s$ be s-sparse and let $\theta^* \in \Theta$. Let $\Theta \subseteq \{\theta \in \mathbb{R}^p : |\theta|_1 \leq a\}$ for some $a > 0$ and all the diagonal elements of $X^T X/n$ be equal to 1. Let Assumption C hold with $\rho < (3\alpha s)^{-1}$ for some $\alpha > 1$. If

$$\min_{j \in J(\theta_s)} |\theta_{sj}| > \tau,$$

then

$$\operatorname{sign} \tilde{\theta}_j = \operatorname{sign} \theta_{sj}, \quad j = 1, \dots, p.$$

$$\tau = \tfrac{3\alpha+1}{3(\alpha-1)}\Big(2\varepsilon + 2(2+\delta)\delta a\Big).$$

Approximately s-sparse solutions

Let θ^* be arbitrary, not necessarily s-sparse. Then we can get bounds involving a residual term, the difference between θ^* and its s-sparse approximation θ_s. In particular, we can take θ_s as the best s-sparse approximation of θ^*, i.e., the vector that coincides with θ^* in its s largest in absolute value coordinates and has other coordinates that vanish.

Approximately s-sparse solutions

Let θ^* be arbitrary, not necessarily s-sparse. Then we can get bounds involving a residual term, the difference between θ^* and its s-sparse approximation θ_s. In particular, we can take θ_s as the best s-sparse approximation of θ^*, i.e., the vector that coincides with θ^* in its s largest in absolute value coordinates and has other coordinates that vanish.

Assumption RE($s, 2$)

There exists $\kappa > 0$ such that

$$\min_{\Delta \neq 0:\, |\Delta_{J^c}|_1 \leq 2|\Delta_J|_1} \frac{|X\Delta|_2}{\sqrt{n}|\Delta_J|_2} \geq \kappa$$

for all subsets J of $\{1, \ldots, p\}$ of cardinality $|J| \leq s$.

Selection of sparsity pattern: noisy case

Theorem 5

Assume that there exists a solution $\theta^* \in \Theta$ of the equation $y = X\theta$. Then for any non-noisy MU-selector $\hat{\theta}$,

$$\frac{1}{n}|X(\hat{\theta} - \theta^*)|_2^2 \quad \leq \quad 4\delta^2|\hat{\theta}|_1^2.$$

If Assumption RE(s,2) holds, then

$$|\hat{\theta} - \theta^*|_1 \quad \leq \quad \frac{4\sqrt{s}\delta}{\kappa}|\hat{\theta}|_1 + 6 \min_{J:|J|\leq s} |\theta^*_{J^c}|_1.$$

NB Assumption C with $\rho < \frac{1}{5\alpha s}$ for some $\alpha > 1$ implies Assumption RE$(s, 2)$ with $\kappa^2 = 1 - 1/\alpha$ (Bickel, Ritov and T., 2007).

Theorem 5 (cont'd)

If Assumption C holds with $\rho < \frac{1}{5\alpha s}$, $\alpha > 1$, *then*

$$|\hat{\theta} - \theta^*|_\infty \;<\; 2\left(1 + \frac{2}{5\sqrt{s\alpha(\alpha-1)}}\right)\delta|\hat{\theta}|_1 + \frac{6}{5\alpha s}\min_{J:|J|\le s}|\theta^*_{J^c}|_1.$$

Censored matrix

- matrix X of size 100×1000 ($n = 100, p = 1000$) which is the normalized version of a 100×1000 matrix with iid standard Gaussian entries.
- Get Z_{ij} by censoring of X_{ij}:

$$Z_{ij} = X_{ij} I\{|X_{ij}| \le t\} + t(\text{sign} X_{ij}) I\{|X_{ij}| > t\}, \quad t = 0.9.$$

- Choose randomly (uniformly) s non-zero elements in a vector θ of size 1000. The associated coefficients are $1 + |N_i|$, $i = 1, \dots, s$, where the N_i are iid standard Gaussian variables.
- We set $y = X\theta + \xi$, where ξ a normal random vector with zero mean and covariance matrix $\sigma^2 I$ with $\sigma = 0.05/1.96$ (so that for an element of ξ, the probability of being between -0.05 and 0.05 is 95%).

Censored matrix

- We compute the solution of (7) where we optimize over $\Theta = \mathbb{R}_+^{1000}$ for $\varepsilon = |\frac{1}{n} Z^T \xi|_\infty$ and different values of the parameter δ. We also compute the "blind" (i.e., based on (y, Z)) Lasso and Dantzig selector.
- Practical choice of δ is crucial. Since the matrix is normalized and $t = 0.9$, it is reasonable to take a value of δ whose order of magnitude $\lesssim 0.1$. We take $\delta = 0$ (ignoring the noise), and $\delta = 0.05, 0.1$.
- We make 100 replications for each couple (s, π).

Censored matrix

	ℓ_2-err	PredErr	Nb_1	Nb_2	Exact
Lasso	0.0670 (0.0106)	11.97 (1.785)	95.46 (2.017)	1 (0)	0
Dantzig	0.0464 (0.0075)	4.673 (1.040)	72.23 (4.751)	1 (0)	0
$\delta = 0$	0.0627 (0.0112)	9.297 (1.685)	74.43 (5.142)	1 (0)	0
$\delta = 0.05$	0.0131 (0.0026)	1.328 (0.257)	1.440 (0.711)	1 (0)	66
$\delta = 0.1$	0.0027 (0.0008)	0.275 (0.085)	1 (0)	1 (0)	100

Censored matrix, $s = 1$.

Censored matrix

	ℓ_2-err	PredErr	Nb_1	Nb_2	Exact
Lasso	0.1825 (0.0317)	34.56 (6.161)	96.8 (1.509)	3 (0)	0
Dantzig	0.1411 (0.0267)	14.55 (3.687)	84.59 (4.547)	3 (0)	0
$\delta = 0$	0.2115 (0.0415)	30.95 (6.027)	85.91 (4.404)	3 (0)	0
$\delta = 0.05$	0.0053 (0.0059)	0.526 (0.517)	3.140 (0.374)	3 (0)	87
$\delta = 0.1$	0.0382 (0.0162)	3.512 (1.120)	3 (0)	3 (0)	100

Censored matrix, $s = 3$.

Missing data

The same parameters of experiment as above, except that the observed matrix Z is defined by $Z_{ij} = \eta_{ij} X_{ij}$, η_{ij} Bernoulli with $\pi = 0.1$.

	ℓ_2-err	PredErr	Nb_1	Nb_2	Exact
Lasso	0.0180 (0.0101)	2.204 (1.165)	94.22 (3.061)	1 (0)	0
Dantzig	0.0097 (0.0070)	0.963 (0.749)	66.18 (10.89)	1 (0)	0
$\delta = 0$	0.0151 (0.0105)	1.438 (0.953)	68.75 (10.92)	1 (0)	0
$\delta = 0.05$	0.0032 (0.0022)	0.272 (0.175)	7.460 (4.940)	1 (0)	12
$\delta = 0.1$	0.0043 (0.0017)	0.416 (0.128)	1.560 (1.194)	1 (0)	74

Missing data, $s = 1$.

Missing data

	ℓ_2-err	PredErr	Nb_1	Nb_2	Exact
Lasso	0.0719 (0.0275)	6.672 (2.108)	96.84 (1.270)	3 (0)	0
Dantzig	0.0529 (0.0233)	4.867 (2.183)	83.55 (5.038)	3 (0)	0
$\delta = 0$	0.0740 (0.0326)	5.536 (2.163)	84.76 (5.020)	3 (0)	0
$\delta = 0.05$	0.0314 (0.0177)	2.496 (0.848)	6.910 (3.108)	3 (0)	12
$\delta = 0.1$	0.0643 (0.0179)	6.099 (0.903)	3.290 (0.791)	3 (0)	84

Missing data, $s = 3$.

Portfolio replication

Data basis: Open and close prices of $p = 491$ assets of the Standard and Poors (S&P 500 index) for the $n = 251$ trading days of 2007.

- p_{ij}^o and p_{ij}^c are open and close prices of the j-th asset for the i-th day. The matrix $(\tilde{X})_{ij} = p_{ij}^c - p_{ij}^o$; X is a normalized matrix obtained from $\tilde{X}$.
- We pick s assets to build our portfolio. We compute the daily absolute return vector of our portfolio $X\theta$, with the coordinate in the vector $\theta \in \mathbb{R}^{491}$ of each chosen asset equal to $1/s$ and the other equal to 0 (note that in practice, if the j-th asset is in the portfolio, it means that its quantity is $1/(s\tilde{\sigma}_j)$, with $\tilde{\sigma}_j$ the empirical standard deviation of its absolute returns).

We consider the following six portfolios:

$s = 2$	$s = 3$
Boeing, Goldman Sachs	Boeing, Google, Goldman Sachs
Boeing, Coca Cola	Boeing, Google, Coca Cola
Boeing, Ford	Boeing, Google, Ford

- We compute $y = X\theta + \xi$, where ξ is the same noise as in the preceding application (ε will also be chosen in the same way as in the preceding application).
- We run the algorithm with $Z = X$ (no matrix uncertainty) and $\delta = 0.5$. We output the retrieved sparsity pattern.
- We run the algorithm with $\delta = 0.5$ and matrix uncertainty: Z is equal to X, up to one of its columns which is replaced by zeros. The column corresponds to one of the assets in the portfolio. We suppress a column associated to the asset different from Boeing and Google.

Initial Portfolio	Retrieved Portfolio, $\delta = 0.5$
B, Goldman Sachs B, Coca Cola B, Ford	B, Morgan Stanley, Merrill Lynch B, Pepsico B, General Motors
B, G, Goldman Sachs B, G, Coca Cola B, G, Ford	B, G, Morgan Stanley, Merrill Lynch B, G B, G, General Motors

Summary

- A non-asymptotic approach to errors-in-variables models, free of classical indentifiability constraints, $p \gg n$.
- Extension of ℓ_1-based sparse recovery beyond the often prohibitive restricted isometry/restricted eigenvalue conditions.
- Simple and efficient way of sparse recovery in several specific problems, such as models with missing data, inverse problems with unknown operator, some financial models.