Realization on Linear Regression Model

Petr Lachout

Charles University in Prague

ROBUST - September 10-14, 2012

Linear equations

We observe couples $(y_1, x_1), (y_2, x_2), (y_3, x_3), \ldots\ldots,$
where $y_t \in \mathbb{R}$, $x_t \in \mathbb{R}^d \quad \forall\ t \in \mathbb{N}$.
We suppose that members of couples are connected by linear equations

$$y_t = x_t^\top \beta_0 + \mathrm{e}_t \quad \forall\ t \in \mathbb{N},$$

where disturbances $\mathrm{e}_t \in \mathbb{R}$ are unknown and $\beta_0 \in \Theta \subset \mathbb{R}^d$ is unknown parameter common for all equations.
Our aim is to reasonably approximate β_0 from couples (y_t, x_t), $t \in \{1, 2, \ldots, T\}$, where $T \in \mathbb{N}$.

Linear equations

Inaccuracy in any equation will be measured using a nonnegative Borel measurable function

$$\rho : \mathbb{R} \to \mathbb{R}, \ \ \rho \geq 0.$$

As approximation of β_0, we take a parameter $\beta \in \Theta$ for which the corresponding disturbances are “small”.

Linear equations

More precisely, we possess given positive numbers $\varepsilon_T > 0$, $T \in \mathbb{N}$ with

$$\limsup_{T \to +\infty} \varepsilon_T \quad = \quad \bar{\varepsilon}, \; 0 \leq \bar{\varepsilon} < +\infty$$

and we find $\hat{\beta}_T \in \Theta$ such that

$$\frac{1}{T} \sum_{t=1}^{T} \rho(y_t - x_t^{\top} \hat{\beta}_T) \quad < \quad \Delta_T + \varepsilon_T,$$

where

$$\Delta_T \quad = \quad \inf \left\{ \frac{1}{T} \sum_{t=1}^{T} \rho(y_t - x_t^{\top} \beta) \; : \; \beta \in \Theta \right\}.$$

Functional setting

The problem can be generalized using a function defined for all $\beta \in \Theta$ and for all μ probability Borel measures on $\mathbb{R}^{d+1}$ by the formula

$$\mathrm{f}\left(\beta\,;\mu\right) \;=\; \int \rho\left(x^\top\left(\beta_0-\beta\right)+e\right)\mu(\,\mathrm{d}x,\,\mathrm{d}e).$$

The integral is always well-defined since ρ is a nonnegative Borel measurable function and μ is a probability Borel measures.

Generalized problem

Let us consider a sequence $\mu_T,\ T \in \mathbb{N}$ of probability Borel measures on $\mathbb{R}^{d+1}$.
We find a parameter $\hat{\beta}_T \in \Theta$ such that

$$\int \rho\left(e + x^\top\left(\beta_0 - \hat{\beta}_T\right)\right)\mu_T(\,\mathrm{d}x,\,\mathrm{d}e). \quad < \quad \Delta_T + \varepsilon_T,$$

where

$$\Delta_T \quad = \quad \inf\left\{\int \rho\left(x^\top\left(\beta_0 - \beta\right) + e\right)\mu_T(\,\mathrm{d}x,\,\mathrm{d}e)\ :\ \beta \in \Theta\right\}.$$

Generalized problem

Setting

$$\mu_T \quad = \quad \sum_{t=1}^{T} \frac{1}{T}\, \delta_{(y_t, \mathrm{e}_t)}$$

we receive precisely the original problem

$$\mathrm{f}\left(\beta\,;\mu_T\right) \quad = \quad \frac{1}{T}\sum_{t=1}^{T} \rho\left(\mathrm{e}_t + x_t^\top\left(\beta_0 - \beta\right)\right) = \frac{1}{T}\sum_{t=1}^{T} \rho(y_t - x_t^\top \beta).$$

LEMMA:
For any ν probability Borel measures on $\mathbb{R}^{d+1}$ and any $\Delta > 0$ we have

$$\lim_{\kappa \to +\infty} \inf_{\|\gamma\|=1} \nu\left(\left\{(x,e) \,:\, \kappa \left|x^\top \gamma\right| \geq \Delta + |e|\right\}\right) =$$

$$= \inf\left\{\nu\left(\left\{(x,e) \in \mathbb{R}^{d+1} \,:\, x^\top \gamma \neq 0\right\}\right) \,:\, \|\gamma\| = 1, \gamma \in \mathbb{R}^d\right\}.$$

PROOF:
Let us denote

$$\mathrm{M} = \inf\left\{\nu\left(\left\{(x,e)\in\mathbb{R}^{d+1} : x^\top\gamma\neq 0\right\}\right) : \|\gamma\|=1, \gamma\in\mathbb{R}^d\right\}.$$

1)
Let $\gamma\in\mathbb{R}^d$ and $k\in\mathbb{N}$, then

$$\left\{(x,e) : k\left|x^\top\gamma\right|\geq\Delta+|e|\right\}\subset\left\{(x,e) : x^\top\gamma\neq 0\right\}.$$

Therefore,

$$\lim_{\kappa\to+\infty}\inf_{\|\gamma\|=1}\nu\left(\left\{(x,e) : \kappa\left|x^\top\gamma\right|\geq\Delta+|e|\right\}\right)\leq\mathrm{M}.$$

Proof 1

2)
Consider a sequence $\gamma_k \in \mathbb{R}^d$, $\|\gamma_k\| = 1$ for all $k \in \mathbb{N}$ such that

$$\nu\left(\left\{(x,e)\,:\,k\left|x^\top\gamma_k\right| \geq \Delta + |e|\right\}\right) <$$
$$< \inf_{\|\gamma\|=1} \nu\left(\left\{(x,e)\,:\,k\left|x^\top\gamma\right| \geq \Delta + |e|\right\}\right) + \frac{1}{k}.$$

The sequence belongs to a compact, therefore, it contains a convergent subsequence

$$\lim_{j\to+\infty} \gamma_{k_j} = \hat{\gamma},\ \|\hat{\gamma}\| = 1.$$

Proof 2

Hence,

$$\left\{(x,e)\,:\,x^\top\hat{\gamma}\neq 0\right\}$$

$$= \bigcup_{J=1}^{+\infty}\bigcap_{j=J}^{+\infty}\left\{(x,e)\,:\,\left|x^\top\hat{\gamma}\right|\geq\left|x^\top(\hat{\gamma}-\gamma_{k_j})\right|+\frac{1}{k_j}(\Delta+|e|)\right\}$$

$$\subset \bigcup_{J=1}^{+\infty}\bigcap_{j=J}^{+\infty}\left\{(x,e)\,:\,k_j\left|x^\top\gamma_{k_j}\right|\geq\Delta+|e|\right\}.$$

Proof 3

Because of σ-additivity of the measure ν, we have

$$
\begin{aligned}
\mathsf{M} &\leq \nu\left(\left\{(x,e)\ :\ x^\top\hat{\gamma}\neq 0\right\}\right) \\
&\leq \lim_{J\to+\infty} \nu\left(\bigcap_{j=J}^{+\infty}\left\{(x,e)\ :\ k_j\left|x^\top\gamma_{k_j}\right|\geq \Delta+|e|\right\}\right) \\
&\leq \liminf_{J\to+\infty} \nu\left(\left\{(x,e)\ :\ k_J\left|x^\top\gamma_{k_J}\right|\geq \Delta+|e|\right\}\right) \\
&\leq \liminf_{J\to+\infty}\left[\inf_{\|\gamma\|=1}\nu\left(\left\{(x,e)\ :\ \kappa\left|x^\top\gamma\right|\geq \Delta+|e|\right\}\right)+\frac{1}{k_J}\right] \\
&= \inf_{\|\gamma\|=1}\nu\left(\left\{(x,e)\ :\ \kappa\left|x^\top\gamma\right|\geq \Delta+|e|\right\}\right).
\end{aligned}
$$

Q.E.D.

Assumptions

Assumption 1: $\Theta \subset \mathbb{R}^d$ is a closed set and for each $\beta \in \Theta$

$$\int \rho(e)\,\nu\,(\,\mathrm{d}x,\,\mathrm{d}e) \leq \int \rho(e + x^\top(\beta_0 - \beta))\,\nu\,(\,\mathrm{d}x,\,\mathrm{d}e)\,.$$

Assumption 2: There is a function $\psi : \mathbb{R}_+ \to \mathbb{R}_+$ which is continuous, nondecreasing and fulfilling:

1. For all $t \in \mathbb{R}$ $\quad \rho(t) \leq \psi(|t|)$.
2. For all $t > 0$ $\quad \int \psi(|e| + t\|x\|)\nu\,(\,\mathrm{d}x,\,\mathrm{d}e) < +\infty$.
3. For all $t > 0$ $\quad \int \psi(|e| + t\|x\|)\mu_T(\,\mathrm{d}x,\,\mathrm{d}e) < +\infty$.

Assumptions

Assumption 3:

$$\mu_T \xrightarrow[T\to+\infty]{\psi-\mathcal{D}} \nu$$

which means

$$\forall f : \mathbb{R}^{d+1} \to \mathbb{R} \text{ bounded continuous}$$

$$\int f(x,e)\mu_T(\,\mathrm{d}x,\,\mathrm{d}e) \xrightarrow[T\to+\infty]{} \int f(x,e)\nu(\,\mathrm{d}x,\,\mathrm{d}e)\,,$$

$$\forall t>0 \int \psi(|e|+t\|x\|)\mu_T(\,\mathrm{d}x,\,\mathrm{d}e) \xrightarrow[T\to+\infty]{} \int \psi(|e|+t\|x\|)\nu(\,\mathrm{d}x,\,\mathrm{d}e)\,.$$

Assumptions

Assumption 4: Denoting

$$\mathsf{H}_\rho = \liminf_{\Delta \to +\infty} \inf \left\{ \rho(t) \,:\, |t| > \Delta, t \in \mathbb{R} \right\},$$

$$\mathsf{M} = \inf \left\{ \nu \left(\left\{ (x, e) \in \mathbb{R}^{d+1} \,:\, x^\top \gamma \neq 0 \right\} \right) \,:\, \|\gamma\| = 1, \gamma \in \mathbb{R}^d \right\},$$

we require $\mathsf{M} > 0$ and a balance

$$\mathsf{H}_\rho \mathsf{M} > \int \rho(e)\, \nu\left(\mathrm{d}x,\, \mathrm{d}e \right) + \limsup_{T \to +\infty} \varepsilon_T .$$

Notation

We denote

$$g(\beta) = \int \rho(e + x^\top(\beta_0 - \beta))\, \nu\,(\mathrm{d}x, \mathrm{d}e) \quad \forall \beta \in \Theta.$$

According to Assumption 1, we know that β_0 is a minimizer of g.
The set of all minimizers and the set of all ε-minimizers of g are denoted by

$$\begin{aligned} \Phi\,\langle g\rangle &= \{\beta \in \Theta \,:\, g(\beta) = g(\beta_0)\}\,, \\ \Psi\,\langle g; \varepsilon\rangle &= \{\beta \in \Theta \,:\, g(\beta) \le g(\beta_0) + \varepsilon\}\,. \end{aligned}$$

We recall that we possess given positive numbers $\varepsilon_T > 0$, $T \in \mathbb{N}$ with

$$\limsup_{T \to +\infty} \varepsilon_T = \bar{\varepsilon}\,,\ 0 \le \bar{\varepsilon} < +\infty.$$

THEOREM:

Let the previous **Assumptions 1-4** be fulfilled.
Then $\beta_0 \in \Phi \langle \mathrm{g} \rangle$ and $\hat{\beta}_T$ exists for every $T \in \mathbb{N}$.
The sequence $\hat{\beta}_T$, $T \in \mathbb{N}$ **is compact** and

$$\emptyset \neq \mathrm{Ls}\left\{ \hat{\beta}_T\,,\ T \in \mathbb{N} \right\} \subset \Psi \langle \mathrm{g}; \bar{\varepsilon} \rangle\,,$$

$$\lim_{T \to +\infty} \mathrm{d}\left(\hat{\beta}_T, \Psi \langle \mathrm{g}; \bar{\varepsilon} \rangle \right) = 0.$$

PROOF:

1)

We know that for each $\beta \in \Theta$, $T \in \mathbb{N}$

$$\mathrm{f}\left(\beta\,;\mu_T\right),\ \mathrm{f}\left(\beta\,;\nu\right) \in \mathbb{R}.$$

Proof 1

2)
Let $\beta, \beta_T \in \Theta$ such that $\beta_T \xrightarrow[T \to +\infty]{} \beta$.
Let us fix $\varepsilon > 0$.
Then there exist Δ, Q and a compact $\bar{K} \subset \mathbb{R}^{d+1}$ such that

$$\|\beta_T - \beta_0\| \leq \Delta \quad \text{for all } T \in \mathbb{N},$$

$$\int_{\psi(|e|+\Delta\|x\|)>Q} (\psi(|e| + \Delta\|x\|) - Q)\, \nu(\mathrm{d}x, \mathrm{d}e) < \varepsilon,$$

$$\mu_T(\mathbb{R}^{d+1} \setminus \bar{K}) < \frac{\varepsilon}{Q} \quad \text{for all } T \in \mathbb{N}.$$

Proof 2

Now, we are able to derive a convergence.

$$\mathrm{f}\left(\beta_T\,;\mu_T\right) = \int \rho(e + x^\top(\beta_0 - \beta_T))\mu_T(\,\mathrm{d}x,\,\mathrm{d}e) =$$

$$= \int \min\{Q, \rho\left(x^\top(\beta_0 - \beta) + e\right)\}\mu_T(\,\mathrm{d}x,\,\mathrm{d}e) +$$

$$+ \int \min\{Q, \rho(e + x^\top(\beta_0 - \beta_T))\} - \min\{Q, \rho\left(x^\top(\beta_0 - \beta) + e\right)\}\mu_T(\,\mathrm{d}x,\,\mathrm{d}e$$

$$+ \int_{\rho(e + x^\top(\beta_0 - \beta_T)) > Q} \left(\rho(e + x^\top(\beta_0 - \beta_T)) - Q\right)\mu_T(\,\mathrm{d}x,\,\mathrm{d}e).$$

1. term

The function $\min\{Q,\rho\}$ is bounded and continuous. Therefore,

$$\int \min\{Q, \rho\left(x^\top(\beta_0-\beta)+e\right)\}\mu_T(\,\mathrm{d}x,\,\mathrm{d}e) \xrightarrow[T\to+\infty]{} \int \min\{Q, \rho\left(x^\top(\beta_0-\beta)+e\right)\}\nu\left(\,\mathrm{d}x,\,\mathrm{d}e\right).$$

2. term

The second term fulfills

$$
\begin{aligned}
&\left| \int \min\{Q, \rho(e + x^\top (\beta_0 - \beta_T))\} - \min\{Q, \rho\left(x^\top (\beta_0 - \beta) + e\right)\}\mu_T(\,\mathrm{d}x,\,\mathrm{d}e) \right| \\
&< 2\varepsilon + \left| \int_{\bar{K}} \min\{Q, \rho(e + x^\top (\beta_0 - \beta_T))\} - \min\{Q, \rho\left(x^\top (\beta_0 - \beta) + e\right)\}\mu_T(\,\mathrm{d} \right. \\
&\leq 2\varepsilon + \sup_{(x,e)\in\bar{K}} \left|\min\{Q, \rho(e + x^\top (\beta_0 - \beta_T))\} - \min\{Q, \rho\left(x^\top (\beta_0 - \beta) + e\right)\}\right| \\
&\xrightarrow[T\to+\infty]{} 2\varepsilon
\end{aligned}
$$

because ρ is continuous, and, hence, uniformly continuous on each compact set.

3. term

The third term is smaller than ε since

$$
\begin{aligned}
0 &\le \int_{\rho(e+x^\top(\beta_0-\beta_T))>Q} \left(\rho(e+x^\top(\beta_0-\beta_T)) - Q\right) \mu_T(\mathrm{d}x, \mathrm{d}e) \\
&\le \int_{\psi(|e|+T\|x\|)>Q} \left(\psi(|e| + T\|x\|) - Q\right) \mu_T(\mathrm{d}x, \mathrm{d}e) \\
&= \int \psi(|e| + T\|x\|)\mu_T(\mathrm{d}x, \mathrm{d}e) - \\
&- \int \min\{Q, \psi(|e| + T\|x\|)\}\mu_T(\mathrm{d}x, \mathrm{d}e) \\
&\xrightarrow[T\to+\infty]{} \\
&\int_{\psi(|e|+T\|x\|)>Q} \left(\psi(|e| + T\|x\|) - Q\right) \nu(\mathrm{d}x, \mathrm{d}e) < \varepsilon.
\end{aligned}
$$

Proof - convergence

Thus, we proved

$$\lim_{T\to+\infty} \mathrm{f}\left(\beta_T\,;\mu_T\right) \;=\; \int \rho\left(x^\top\left(\beta_0-\beta\right)+e\right)\nu\left(\,\mathrm{d}x,\,\mathrm{d}e\right)=\mathrm{f}\left(\beta\,;\nu\right).$$

It means that $\mathrm{f}\left(\beta\,;\mu_T\right)$ converge to $\mathrm{f}\left(\beta\,;\nu\right)$ uniformally on each compact.

Proof - tightness

According to Assumptions, there is a number $\Delta > 0$ such that

$$\inf_{|t|>\Delta} \rho(t) \cdot \mathrm{M} > \mathrm{f}\left(\beta_0\,;\nu\right) + \bar{\varepsilon}.$$

Hence according to Lemma, we are able to find Γ such that

$$\inf_{|t|>\Delta} \rho(t) \cdot \inf_{\|\gamma\|=1} \nu\left(\left\{(x,e)\,:\,\Gamma\left|x^\top\gamma\right| \geq \Delta + |e|\right\}\right) \quad > \quad \mathrm{f}\left(\beta_0\,;\nu\right) + \bar{\varepsilon}.$$

Then, we define a compact

$$K \quad = \quad \left\{\beta \in \Theta\,:\,\|\beta - \beta_0\| \leq \Gamma\right\}.$$

Proof - tightness

For $\beta \in \Theta$, $\|\beta - \beta_0\| > \Gamma$ we receive following chain of inequalities:

$$\begin{aligned}
\mathrm{f}\,(\beta\,;\mu_T) &= \int \rho\left(x^\top(\beta_0-\beta)+e\right)\mu_T(\mathrm{d}x,\,\mathrm{d}e) \\
&\geq \int_{|x^\top(\beta_0-\beta)+e|>\Delta} \rho\left(x^\top(\beta_0-\beta)+e\right)\mu_T(\mathrm{d}x,\,\mathrm{d}e) \\
&\geq \inf_{|t|>\Delta}\rho(t)\cdot\mu_T\left(\left\{(x,e)\,:\,|x^\top(\beta_0-\beta)+e|>\Delta\right\}\right) \\
&\geq \inf_{|t|>\Delta}\rho(t)\cdot\mu_T\left(\left\{(x,e)\,:\,|x^\top(\beta-\beta_0)|>\Delta+|e|\right\}\right) \\
&\geq \inf_{|t|>\Delta}\rho(t)\cdot\mu_T\left(\left\{(x,e)\,:\,\Gamma\left|x^\top\frac{\beta-\beta_0}{\|\beta-\beta_0\|}\right|>\Delta+|e|\right\}\right).
\end{aligned}$$

Proof - tightness

For any $\varepsilon > 0$, properly chosen sequence of γ_T, $\|\gamma_T\| = 1$ and its cluster point $\hat{\gamma}$, we have

$$\liminf_{T\to+\infty} \inf_{\beta\notin K} \mathrm{f}\left(\beta\,;\mu_T\right) \geq$$

$$\geq \inf_{|t|>\Delta} \rho(t) \cdot \liminf_{T\to+\infty} \inf \left\{ \mu_T\left(\left\{(x,e)\,:\,\Gamma\left|x^\top\gamma\right| > \Delta + |e|\right\}\right)\,:\,\|\gamma\| = 1 \right\}$$

$$\geq \inf_{|t|>\Delta} \rho(t) \cdot \liminf_{T\to+\infty} \mu_T\left(\left\{(x,e)\,:\,\Gamma\left|x^\top\gamma_T\right| > \Delta + |e|\right\}\right)$$

$$\geq \inf_{|t|>\Delta} \rho(t) \cdot \liminf_{T\to+\infty} \mu_T\left(\left\{(x,e)\,:\,\Gamma\left|x^\top\hat{\gamma}\right| > \Delta + |e| + \Gamma\left|x^\top(\gamma_T - \hat{\gamma})\right|\right\}\right)$$

$$\geq \inf_{|t|>\Delta} \rho(t) \cdot \liminf_{T\to+\infty} \mu_T\left(\left\{(x,e)\,:\,\Gamma\left|x^\top\hat{\gamma}\right| > (1+\varepsilon)\Delta + |e|\right\}\right)$$

$$\geq \inf_{|t|>\Delta} \rho(t) \cdot \nu\left(\left\{(x,e)\,:\,\Gamma\left|x^\top\hat{\gamma}\right| > (1+\varepsilon)\Delta + |e|\right\}\right).$$

Proof - tightness

Letting ε vanish we have

$$\begin{aligned}\liminf_{T\to+\infty} \inf_{\beta\notin K} \mathrm{f}\left(\beta\,;\mu_T\right) &\geq \inf_{|t|>\Delta} \rho(t)\cdot\nu\left(\left\{(x,e)\;:\;\Gamma\left|x^\top\hat{\gamma}\right|\geq\Delta+|e|\right\}\right)\\ &> \mathrm{f}\left(\beta_0\,;\nu\right)+\bar{\varepsilon}.\end{aligned}$$

Proof - finish

We found that it is sufficient to consider $\beta \in K$, only.
We know that $\mathrm{f}(\beta\,;\mu_T)$ converge to $\mathrm{f}(\beta\,;\nu)$ uniformally on K.
That completes the proof.
Q.E.D.

Chen, X.R.; Wu, Y.H.: Strong consistency of *M*-estimates in linear model. J. Multivariate Analysis **27,1**(1988), 116-130.

Dodge, Y.; Jurečková, J.: *Adaptive Regression.* Springer-Verlag, New York, 2000.

Hoffmann-Jørgensen, J.: *Probability with a View Towards to Statistics I, II.* Chapman and Hall, New York, 1994.

Huber, P.J.: *Robust Statistics.* John Wiley & Sons, New York, 1981.

Jurečková, J.: Asymptotic representation of M-estimators of location. Math. Operat. Stat. Sec. Stat. **11,1**(1980), 61-73.

Jurečková, J.: Representation of M-estimators with the second-order asymptotic distribution. Statistics & Decision **3**(1985), 263-276.

Jurečková, J.; Sen, P.K.: *Robust Statistical Procedures.* John Wiley & Sons, Inc., New York, 1996.

Kelley, J.L.: *General Topology.* D. van Nostrand Comp., New York, 1955.

Knight, K.: Limiting distributions for L_1-regression estimators under general conditions. Ann. Statist. **26,2**(1998), 755-770.

Lachout, P.: Stability of stochastic optimization problem - nonmeasurable case. Kybernetika **44,2**(2008), 259-276.

Lachout, P.: Stochastic optimization sensitivity without measurability. In: Proceedings of the 15^{th} MMEI held in Herĺany, Slovakia (Eds.: K. Cechlárová, M. Halická, V. Borbeĺová, V. Lacko) (2007), 131-136.

Lachout, P.; Liebscher, E.; Vogel, S.: Strong convergence of estimators as ε_n-minimizers of optimization problems. Ann. Inst. Statist. Math. **57,2**(2005), 291-313.

Lachout, P.; Vogel, S.: On Continuous Convergence and Epi-convergence of Random Functions. Part I: Theory and Relations. Kybernetika **39,1**(2003), 75-98.

- Leroy, A.M.; Rousseeuw, P.J.: *Robust Regression and Outlier Detection.* John Wiley & Sons, New York, 1987.
- Robinson, S.M.: Analysis of sample-path optimization. Math. Oper. Res. **21,3**(1996), 513-528.
- Rockafellar, T.; Wets, R.J.-B.: *Variational Analysis.* Springer-Verlag, Berlin, 1998.
- Vajda, I.; Janžura, M.: On asymptotically optimal estimates for general observations. Stochastic Processes and their Applications **72,1**(1997), 27-45.
- van der Vaart, A.W.; Wellner, J.A.: *Weak Convergence and Empirical Processes* Springer, New York, 1996.