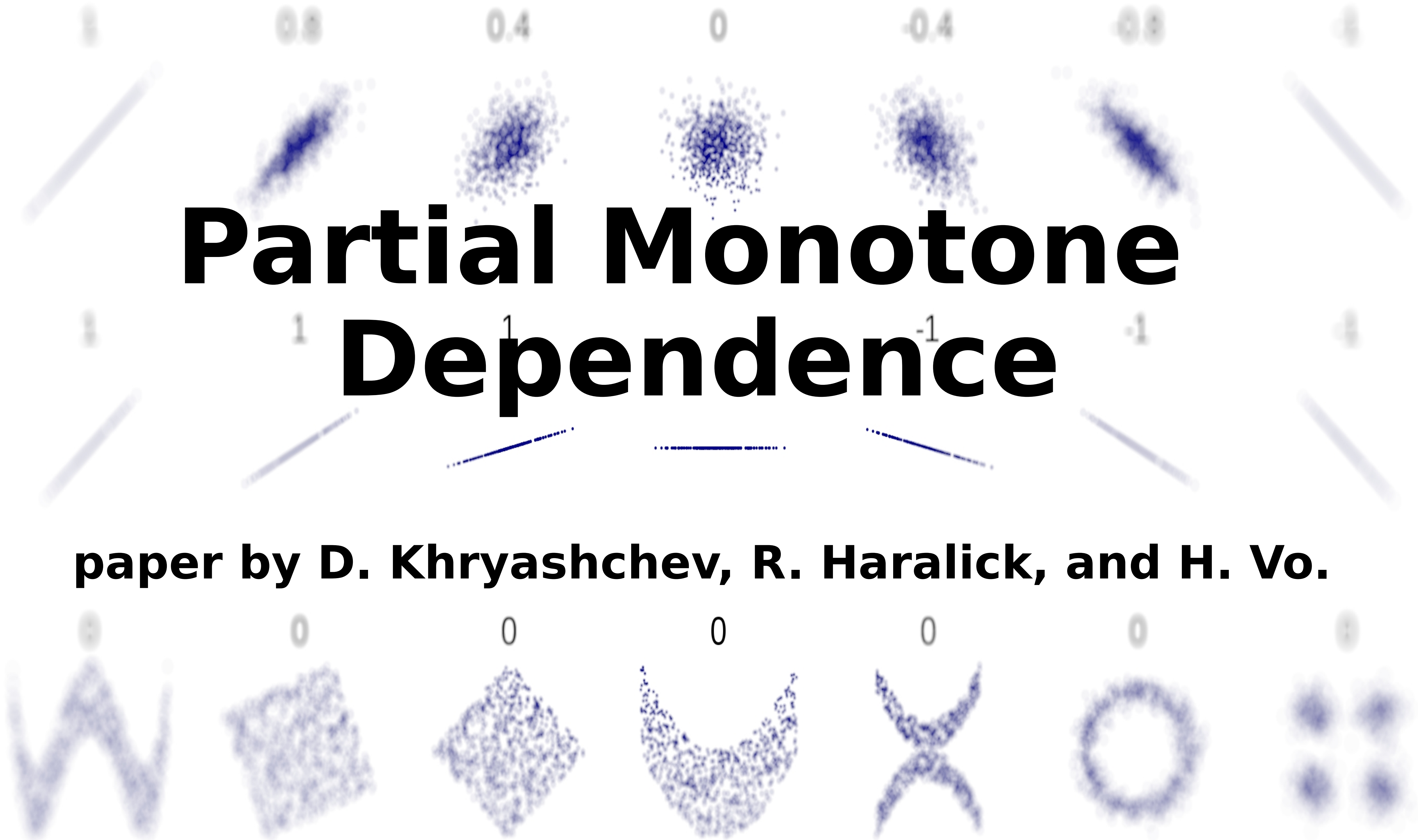
Partial Monotone Dependence

paper by D. Khryashchev, R. Haralick, and H. Vo.



Notation and assumptions

Without the loss of generality we assume that all numerically valued random variables X and Y are standardized

$$E_X[X] = E_Y[Y] = 0 \text{ and } E_X\left[X^2\right] = E_Y\left[Y^2\right] = 1.$$

All of the transformations f, g are Borel-measurable functions, such that

$$f, g\colon \mathbb{R} \to \mathbb{R},\ E[f(\cdot)] = E[g(\cdot)] = 0 \text{ and } E\left[f(\cdot)^2\right] = E\left[g(\cdot)^2\right] = 1.$$

Notation and assumptions

Without the loss of generality we assume that all numerically valued random variables X and Y are standardized

$$E_X[X] = E_Y[Y] = 0 \text{ and } E_X\left[X^2\right] = E_Y\left[Y^2\right] = 1.$$

All of the transformations f, g are Borel-measurable functions, such that

$$f, g: \mathbb{R} \to \mathbb{R},\ E[f(\cdot)] = E[g(\cdot)] = 0 \text{ and } E\left[f(\cdot)^2\right] = E\left[g(\cdot)^2\right] = 1.$$

We will denote Pearson product-moment (linear) correlation as

$$\rho(X, Y) = E_{XY}[XY]$$

Maximal Correlation

The first measure of dependence that fits almost all of the requirements of Renyi postulates was proposed by Gebelein in 1941 [2]

$$\rho_{max}(X,Y) = \max_{f,g} \rho\big(f(X)g(Y)\big)$$

Maximal Correlation

The first measure of dependence that fits almost all of the requirements of Renyi postulates was proposed by Gebelein in 1941 [2]

$$\rho_{max}(X,Y) = \max_{f,g} \rho\big(f(X)g(Y)\big)$$

Which following our assumptions simplifies to

$$\rho_{max}(X,Y) = \max_{f,g} E_{XY}[f(X)g(Y)]$$

Monotone Correlation

Kimeldorf and Sampson [3] demonstrated that maximization over all Borel-measurable functions is too broad and, in some cases, leads to $\rho_{max}(X, Y) > 0$ for independent random variables.

Monotone Correlation

Kimeldorf and Sampson [3] demonstrated that maximization over all Borel-measurable functions is too broad and, in some cases, leads to $\rho_{max}(X,Y) > 0$ for independent random variables.

They introduced the monotone correlation coefficient satisfying all the postulates

$$\rho_{mono}(X,Y) = \max_{f,g:monotone} \rho\big(f(X)g(Y)\big).$$

Monotone Correlation

Kimeldorf and Sampson [3] demonstrated that maximization over all Borel-measurable functions is too broad and, in some cases, leads to $\rho_{max}(X,Y) > 0$ for independent random variables.

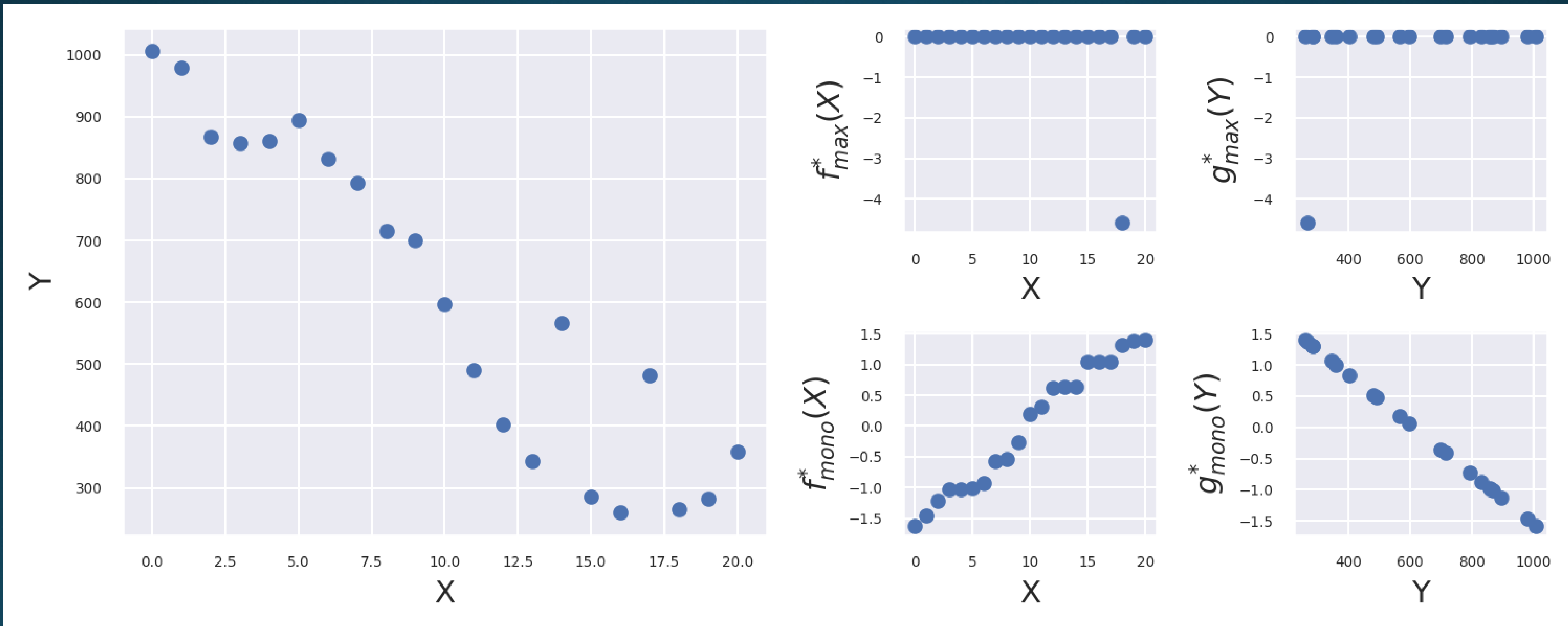
They introduced the monotone correlation coefficient satisfying all the postulates

$$\rho_{mono}(X,Y) = \max_{f,g:monotone} \rho\big(f(X)g(Y)\big).$$

Clearly,

$$|\rho(X,Y)| \leq \rho_{mono}(X,Y) \leq \rho_{max}(X,Y).$$

Limitations of ρ_{max} and ρ_{mono}



Partial Monotone Correlation

To mitigate the limitations of the Maximal and Monotone Correlation, we introduce Partial Monotone Correlation coefficient:

$$\rho_{p.mono}(X, Y, m, n) = \sup_{f_m, g_n} \rho\big(f_m(X) g_n(Y)\big),$$

Partial Monotone Correlation

To mitigate the limitations of the Maximal and Monotone Correlation, we introduce Partial Monotone Correlation coefficient:

$$\rho_{p.mono}(X, Y, m, n) = \sup_{f_m, g_n} \rho\big(f_m(X) g_n(Y)\big),$$

$$m = \left|\left\{i \middle| f_m\left(x_{(i)}\right) > f_m\left(x_{(i+1)}\right)\right\}\right|,$$
$$n = \left|\left\{j \middle| g_n\left(y_{(j)}\right) > g_n\left(y_{(j+1)}\right)\right\}\right|.$$

Approximation of $\rho_{p.mono}$

First, we compute $\rho_{p.mono}$ using Simultaneous Perturbation Stochastic Approximation algorithm with slight modifications [10].

Approximation of $\rho_{p.mono}$

First, we compute $\rho_{p.mono}$ using Simultaneous Perturbation Stochastic Approximation algorithm with slight modifications [10].

Our 1st problem is to find f_0 and g_0 that maximize $0 \leq E_{XY}[f_0(X)g_0(Y)] \leq 1$.

Approximation of $\rho_{p.mono}$

First, we compute $\rho_{p.mono}$ using Simultaneous Perturbation Stochastic Approximation algorithm with slight modifications [10].

Our 1st problem is to find f_0 and g_0 that maximize $0 \leq E_{XY}[f_0(X)g_0(Y)] \leq 1$.

We will look for f_0 and g_0 in the form $f_0(X) = X + \Delta^x$ and $g_0(Y) = Y + \Delta^y$.

Approximation of $\rho_{p.mono}$

Our strictly monotone constraint on f_0 and g_0 is

$$\forall i < M: x_{(i)} + \delta_i^x < x_{(i+1)} + \delta_{i+1}^x$$

$$\forall j < N: y_{(j)} + \delta_j^y < y_{(j+1)} + \delta_{j+1}^y.$$

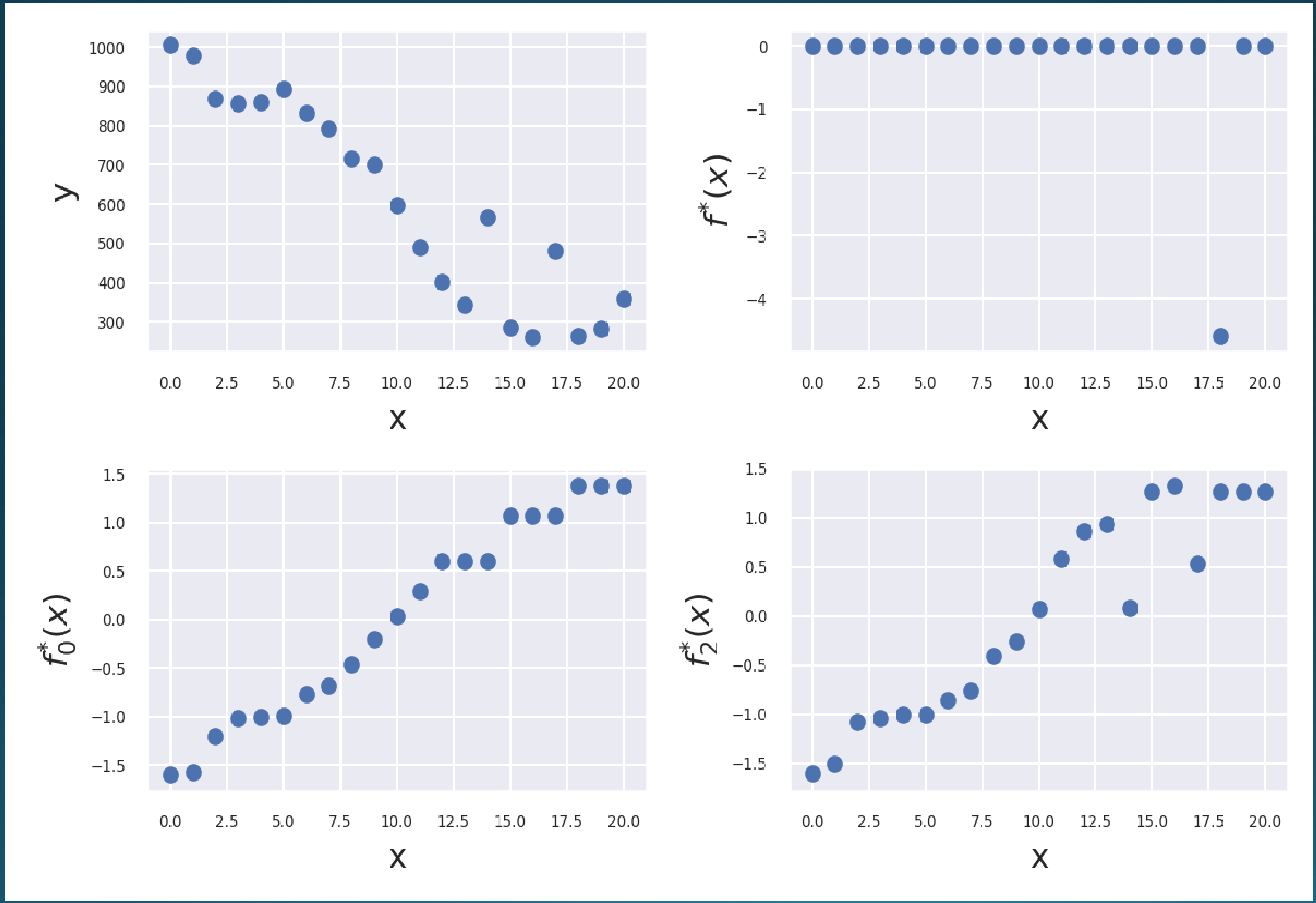
Approximation of $\rho_{p.mono}$

Our strictly monotone constraint on f_0 and g_0 is

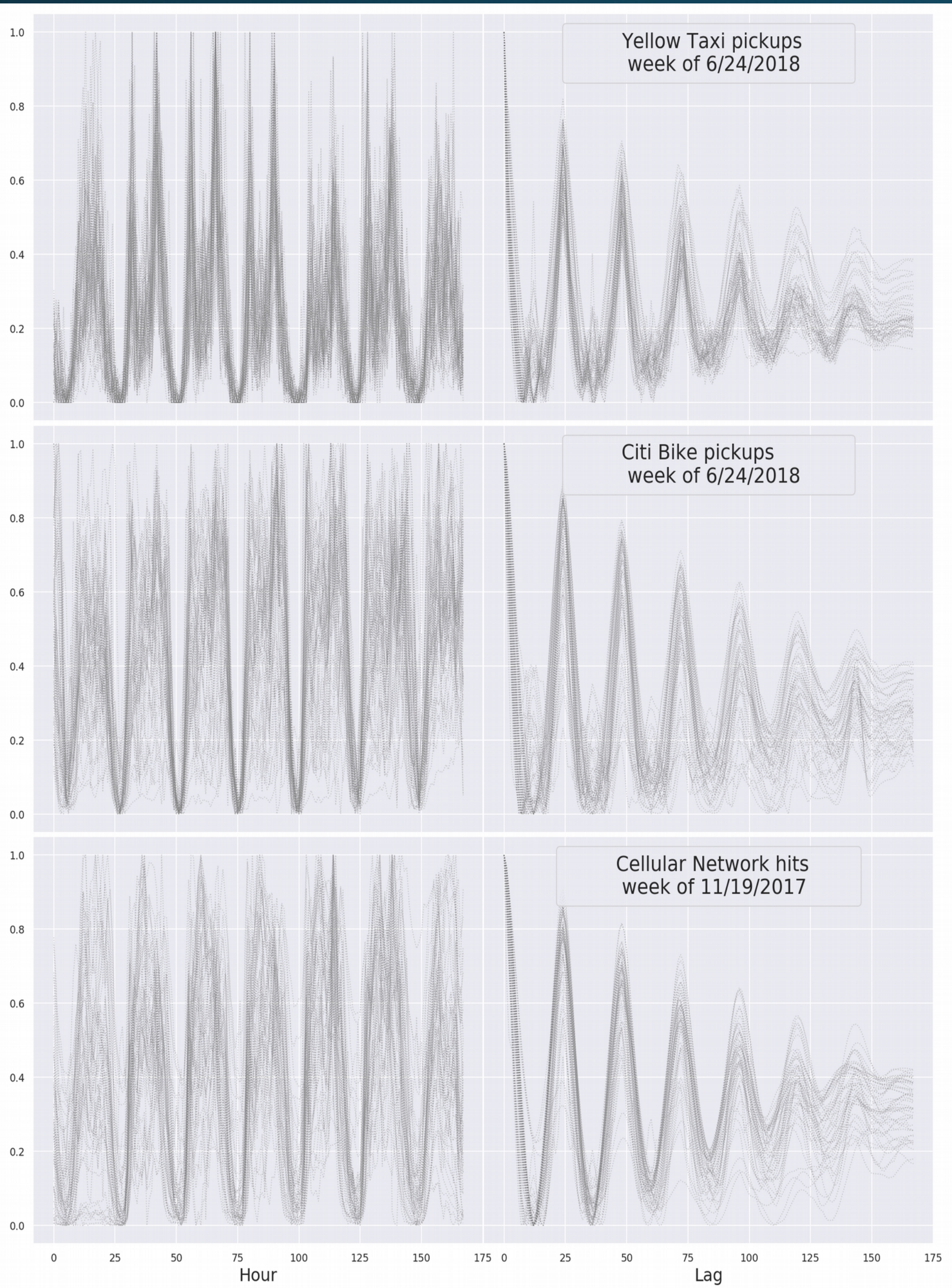
$$\forall i < M: x_{(i)} + \delta_i^x < x_{(i+1)} + \delta_{i+1}^x$$
$$\forall j < N: y_{(j)} + \delta_j^y < y_{(j+1)} + \delta_{j+1}^y.$$

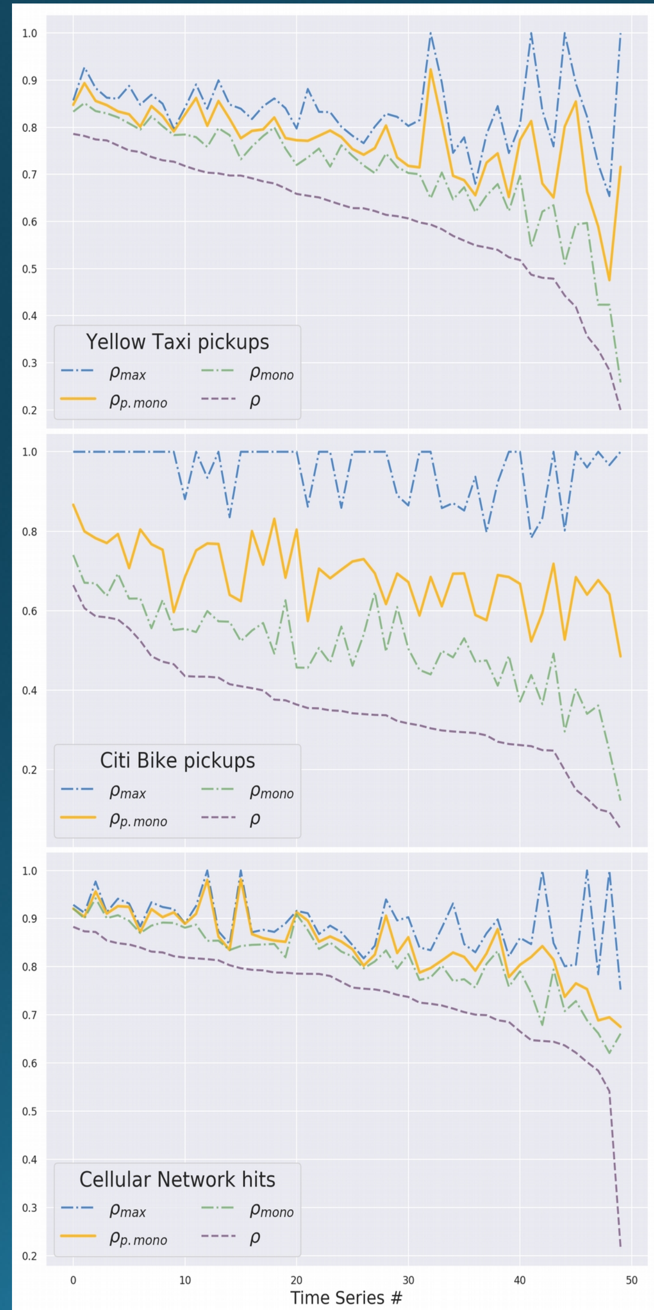
We pick a uniformly random direction through a point Z on $(M+N)$-dimensional sphere: $Z \sim N(0, I)$, $\|Z\| = 1$. The first M dimensions $Z_1^M = (z_1, \dots, z_M)^T$ correspond to the direction of change in Δ^x, the last N dimensions $Z_{M+1}^{M+N} = (z_{M+1}, \dots, z_{M+N})^T$ correspond to the direction of change in Δ^y.

$\rho_{p.mono}$ vs ρ_{mono} vs ρ_{max}



Applications. Correlation





As expected, the values of the correlation coefficients are arranged as follows:

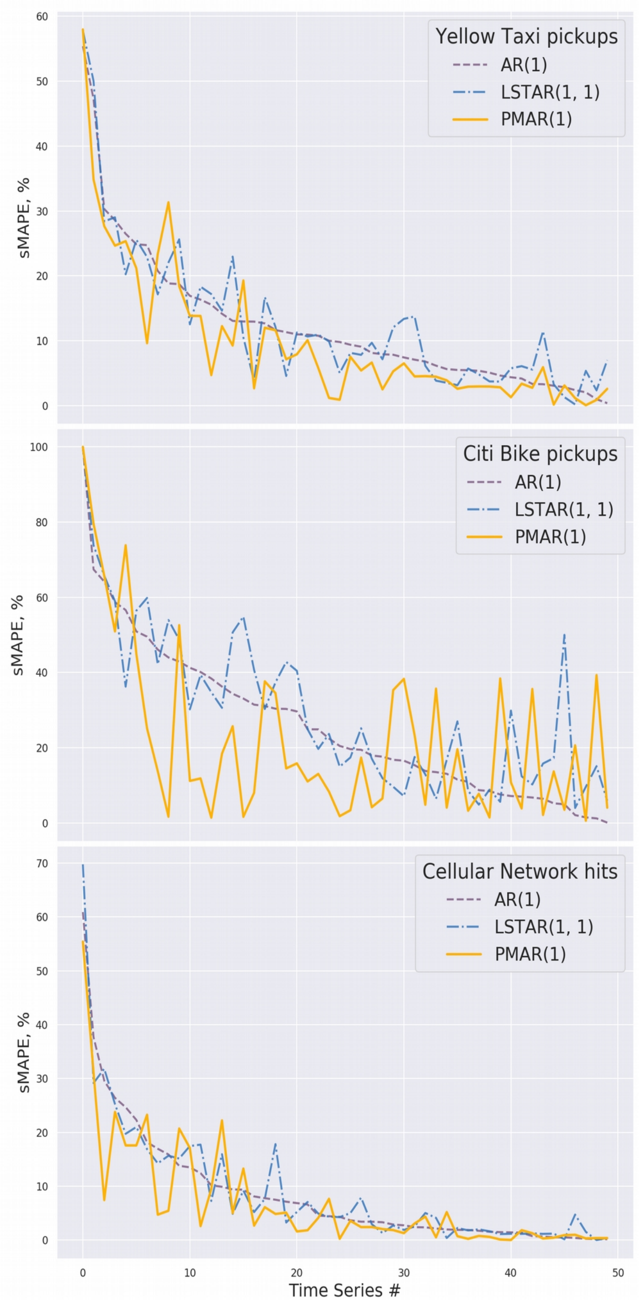
$$\rho \leq \rho_{mono} \leq \rho_{p.mono} \leq \rho_{max}$$

Applications. Forecasting

We apply $\rho_{p.mono}$ in a basic nonlinear autoregressive model, PMAR (Partial Monotone AutoRegression). Given time series $Z = \{z_t\}_1^N = \{z_1, \dots, z_N\}$ it is

$$g_0^*(z_t) = \alpha f_m^*(z_{t-1}) + \beta + \epsilon_t.$$

Results of forecasting



sMAPE			
	AR	**LSTAR**	**PMAR**
Taxi	12.27%	12.68%	9.86%
Bike	26.04%	29.55%	22.0%
Cellular	8.63%	8.93%	6.94%

bias			
	AR	**LSTAR**	**PMAR**
Taxi	-7.91	1.22	-16.17
Bike	1.30	1.39	0.54
Cellular	0.99	4.82	14.40

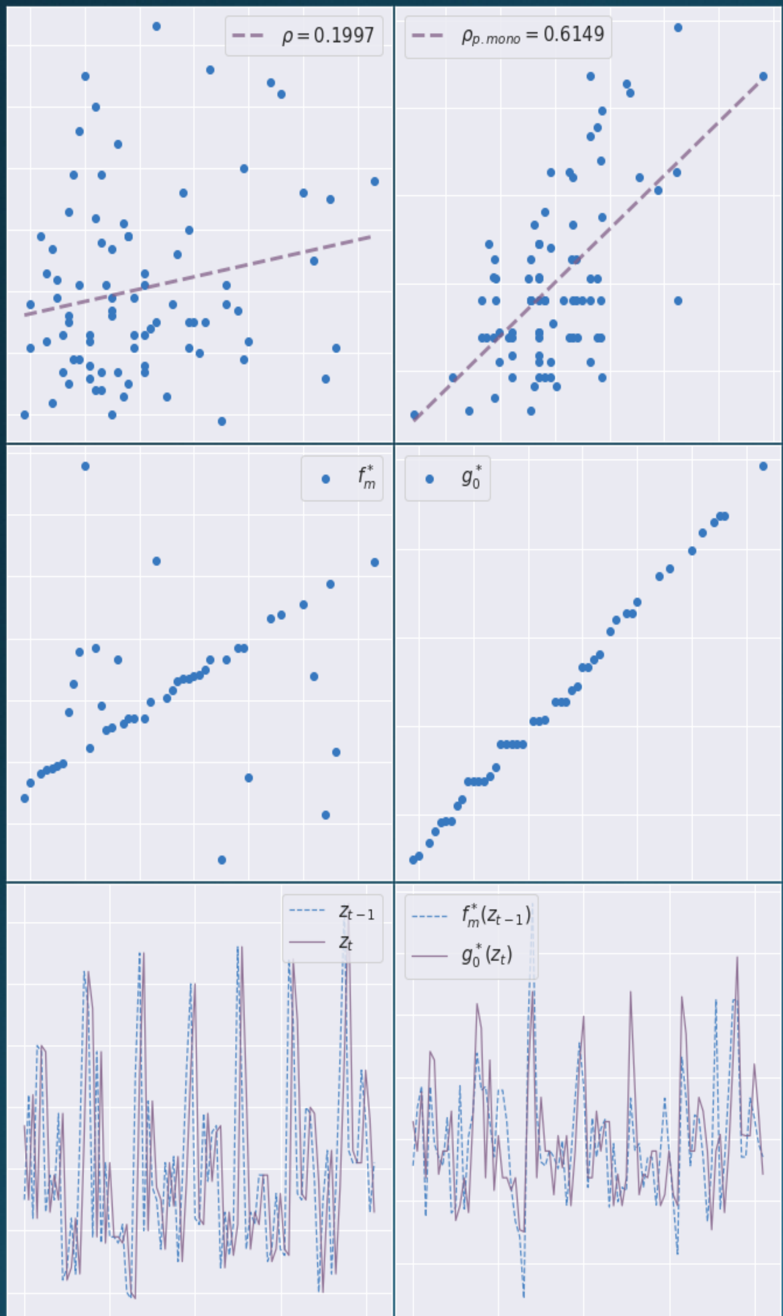
We compared the performances of the models in terms of

$$sMAPE = \frac{100\,\%}{N} \sum_{t=1}^{N} \frac{|\hat{z}_t - z_t|}{|\hat{z}_t| + |z_t|}$$

and

$$bias = \frac{1}{N} \sum_{t=1}^{N} \hat{z}_t - z_t$$

Yellow Taxi pickups transformed



We applied our $\rho_{p.mono}$ to a taxi pickup time series.

Top-left: scatter plot of original time series, z_t vs z_{t-1}.
Top-right: scatter plot of transformed time series, $g_0^*(z_t)$ vs $f_m^*(z_{t-1})$.

Middle-left: maximizing transformation f_m^*.
Middle-right: maximizing transformation g_0^*.

Bottom-left: original series z_t vs its lagged copy z_{t-1}.
Bottom-right: transformed time series $g_0^*(z_t)$ and its lagged copy $f_0^*(z_{t-1})$ aligned.