Efficient estimation of regression models with spillovers

Flexible parametric and semi-parametric approaches

Vincenzo Verardi
LouRIM-LSM
UCLouvain

Joint with
Nicolas Debarsy (CNRS, LEM UMR 9221)
and
Catherine Vermandele (Université libre de Bruxelles)

Stata Economic Symposium
November 7, 2024

Introduction

Structure of the **Presentation**:

- **Local Asymptotic Normality** Property (Le Cam 1960)
- Linear model with **spillovers**
- **Flexible Parametric** Estimator
- **Semiparametric** Estimator
- **Stata** Implementation
- Monte Carlo **Simulations**
- **Illustration**
- **Conclusion**

Remark: This presentation is based on two papers prepared while working as an Associate Researcher of the **FNRS** at **University of Namur**.

The **Stata** codes presented here will only be made available via SSC after the original papers have been validated through the peer-review process.

Introduction

Spillover regression models capture interactions between cross-sectional units, often used in **social networks**, corporate finance, and public policies, where outcomes are interdependent.

Technically, they are represented as:

$$y_i = \lambda \boldsymbol{W}_{i.}\boldsymbol{y} + \boldsymbol{x}_i^T \beta + \varepsilon_i$$

where

- y_i is the dependent variable for the $i-th$ unit and $\boldsymbol{y}$ the vector of the dependent variable for all the sampled units
- λ is the autoregressive parameter capturing the degree of interaction between units
- $\boldsymbol{W}_{i.}$ denotes the $i-th$ row of the weight matrix indicating the structure of the interactions
- $\boldsymbol{x}_i$ is a vector of explanatory variables for unit i
- β is the parameter vector measuring the effects of these variables
- ε_i is the error term capturing unexplained variability.

Introduction

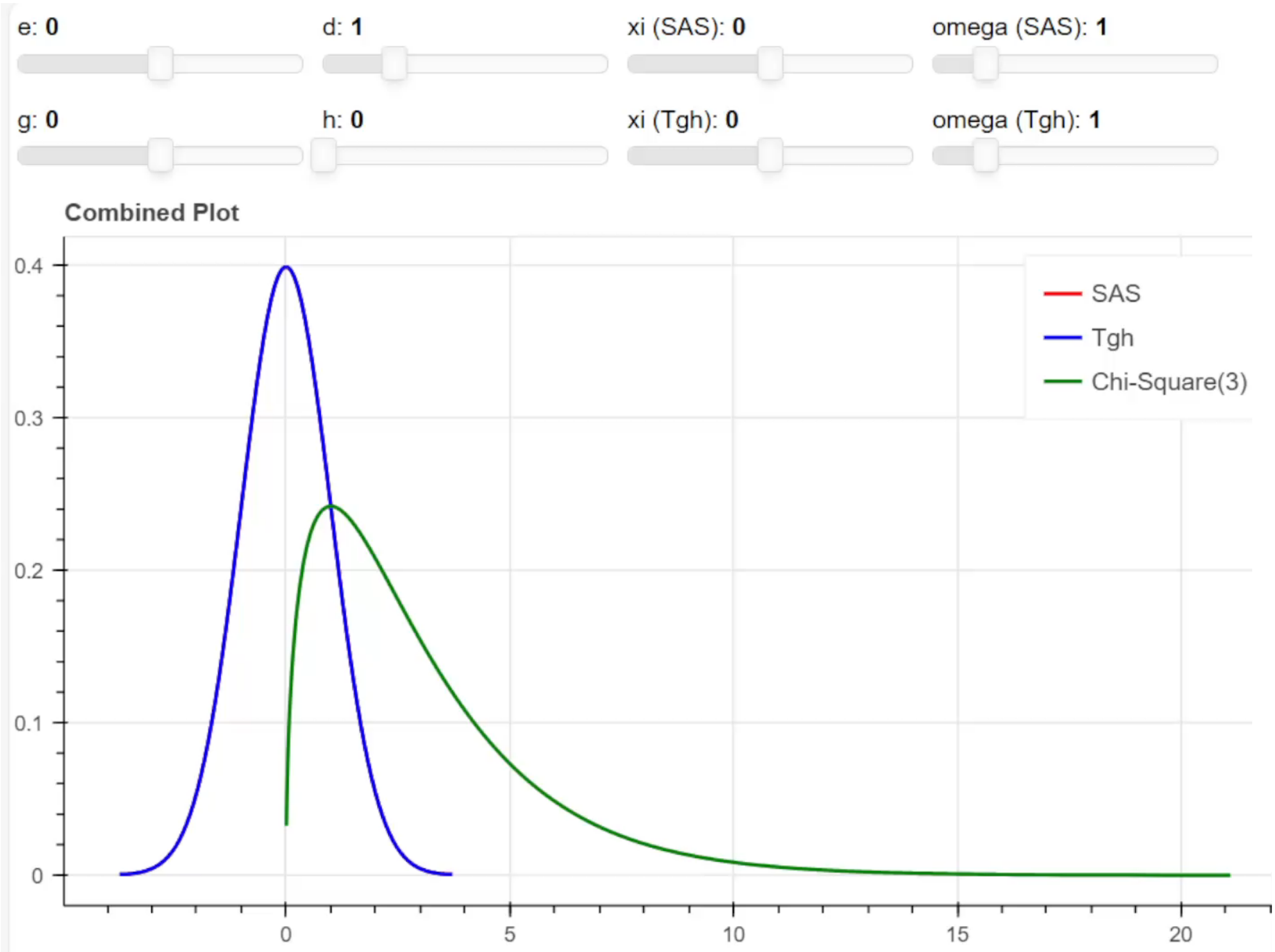
Ordinary Least Squares (OLS) is **unsuitable** for estimating models with **spillover effects** due to simultaneity among individual interactions. Many alternative estimators are available in the literature:

- **Two-Stage Least Squares** (see for example Kelejian and Prucha,1998 or Bramoullé et al., 2009)
- **Generalized Method of Moments** (see Liu at al. 2010)
- **(Quasi-)Maximum Likelihood** (under normality assumption) (see Lee, 2004)
- **Adaptive Estimation** (see Robinson, 2010)

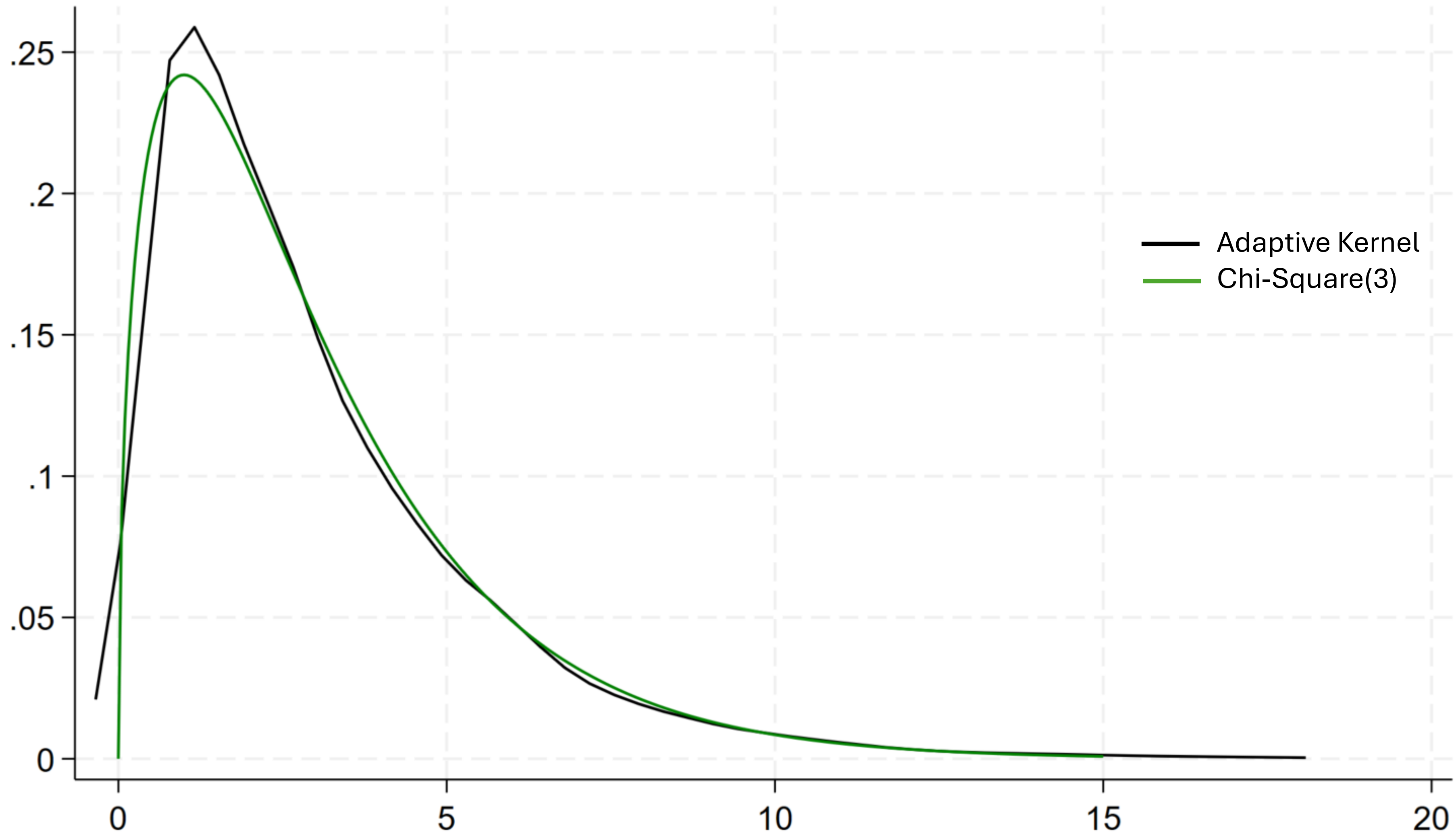
In this presentation, we will discuss **two** alternative, **efficient estimators** derived from the Local Asymptotic Normality (LAN) theory introduced by Le Cam (1960):

- A **flexible parametric** estimator
- A **semiparametric** estimator

Flexible Parametric



Non-parametric



LAN

Local Asymptotic Normality (LAN)

The **LAN property** suggests that, **under** certain **conditions**, **complex models** behave **similarly** to simpler **Gaussian models** in a local sense.

It examines **model's behavior** for **small perturbations** around a **true parameter** value.

As the **sample size increases**, the **log-likelihood ratio** between the true parameter and nearby values **approximates a normal distribution**.

This **helps** in **estimating parameters** and testing hypotheses by using **well-known properties** of the **normal distribution**.

Importantly, **LAN** only requires **Quadratic Mean Differentiability** (**QMD**), which is a less stringent assumption than what is necessary for maximum likelihood estimation. For instance, models with parameter constraints or irregularities at certain points may still satisfy QMD, while MLE assumptions may fail.

For instance, the **Laplace** distribution is **QMD** and thus enjoys the **LAN property** while parameters **cannot** be estimated **by ML**

Local Asymptotic Normality (LAN)

The concept is very well explained by **Canay (2021)**. Suppose you have probability distribution $P_\theta = N\left(\theta, \sigma^2\right)$, with known σ. Under P_θ,

$$\log\left[\frac{dP^{(n)}_{\theta+\tau^{(n)}/\sqrt{n}}}{dP^{(n)}_\theta}\right] = -\frac{1}{2\sigma^2}\sum_{i=1}^n\left(X_i^{(n)} - \theta - \frac{\tau^{(n)}}{\sqrt{n}}\right)^2 + \frac{1}{2\sigma^2}\sum_{i=1}^n\left(X_i^{(n)} - \theta\right)^2$$

$$= \frac{1}{\sigma^2}\sum_{i=1}^n\left(X_i^{(n)} - \theta\right)\frac{\tau^{(n)}}{\sqrt{n}} - \frac{\tau^{(n)2}}{2\sigma^2}$$

$$= \underbrace{\tau^{(n)}\sqrt{n}\left(\frac{\bar{X}^{(n)} - \theta}{\sigma}\right)\frac{1}{\sigma}}_{\Delta^{(n)}(\theta)\sim N(0, I(\theta))} - \frac{\tau^{(n)2}}{2}\underbrace{1/\sigma^2}_{I(\theta)}$$

$$= \tau^{(n)}\Delta^{(n)}(\theta) - \frac{1}{2}\tau^{(n)2} I(\theta) \sim N\left(-\frac{1}{2}\frac{\tau^2}{\sigma^2}, \frac{\tau^2}{\sigma^2}\right) \text{ under } P_\theta$$

where $I(\theta) = \frac{1}{\sigma^2}$ is the Fisher information matrix. Generalizing to the vector case we have

$$\log\left[\frac{dP^{(n)}_{\theta+\tau^{(n)}/\sqrt{n}}}{dP^{(n)}_\theta}\right] = (\tau^{(n)})^{\mathrm{T}}\boldsymbol{\Delta}^{(n)}(\theta) - \tfrac{1}{2}(\tau^{(n)})^{\mathrm{T}}\mathbf{I}(\theta)\tau^{(n)} \sim N\left(-\tfrac{1}{2}\tau^\top\mathbf{I}(\theta)\tau, \tau^\top\mathbf{I}(\theta)\tau\right)$$

Local Asymptotic Normality

Definition

Consider a sequence of statistical models $\left\{P_\theta^{(n)}\right\}$ indexed by a parameter $\theta \in \Theta$ and the sample size n. The model sequence exhibits LAN at a true parameter value θ if there exist sequences of random vectors $\mathbf{\Delta}^{(n)}(\theta)$ and symmetric positive semi-definite matrices $\mathbf{I}(\theta)$ such that the log-likelihood ratio satisfies the following approximation:

$$\log \frac{dP^{(n)}_{\theta+\tau/\sqrt{n}}}{dP^{(n)}_{\theta}} = (\tau^{(n)})^{\mathrm{T}} \mathbf{\Delta}^{(n)}(\theta) - \frac{1}{2}(\tau^{(n)})^{\mathrm{T}} \mathbf{I}(\theta) \tau^{(n)} + o_{\mathrm{P}}(1)$$

where (i) τ is a fixed vector representing the local perturbation of the parameter θ, (ii)$\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta})$ converges in distribution to a normal random vector $\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}) \sim N(0, \mathbf{I}(\theta))$ and (iii) $\mathbf{I}(\theta)$ converges to the Fisher information matrix $\mathbf{I}(\theta)$.

Important Implication of LAN

An **implication of the LAN** property is that if one has $\widetilde{\theta}^{(n)}$, a $\sqrt{n}$-**consistent estimator** of θ, then

$$\widehat{\theta}^{(n)} = \widetilde{\theta}^{(n)} + \frac{1}{\sqrt{n}} \left[\mathbf{I}\left(\widetilde{\theta}^{(n)} \right) \right]^{-1} \boldsymbol{\Delta}^{(n)} \left(\widetilde{\theta}^{(n)} \right)$$

is an **asymptotically efficient** estimator of θ where $\boldsymbol{\Delta}^{(n)}(\theta)$ is called the central sequence for θ (based on the score function):

$$\boldsymbol{\Delta}^{(n)}(\theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial \ln f_{\theta}(y_i)}{\partial \theta} \xrightarrow{\mathscr{L}} N(\mathbf{0}, \mathbf{I}(\theta))$$

and **information matrix** $\mathbf{I}(\theta)$ is

$$\mathbf{I}(\theta) = \mathrm{E}\left[\left(\frac{\partial \ln f_{\theta}(y_i)}{\partial \theta} \right) \left(\frac{\partial \ln f_{\theta}(y_i)}{\partial \theta} \right)^{\mathrm{T}} \right]$$

Spillovers

Spillovers-effect regression model

Consider the following **linear model** with **endogenous effects**.
For $i = 1, \cdots, n$

$$y_i^{(n)} = \left(x_i^{(n)}\right)^{\mathrm{T}} \beta + \lambda \sum_{j \neq i}^{n} w_{ij}^{(n)} y_j^{(n)} + \varepsilon_i^{(n)}$$

$\varepsilon_1^{(n)}, \ldots, \varepsilon_n^{(n)}$ are i.i.d. error terms with **unknown distribution** function F and density f assumed, without any loss of generality, to have a median of zero (to identify β_0).

Assumptions similar to those of **Lee (2004)**.
Excepted for the assumptions on distribution of $\varepsilon_i^{(n)}$:

- $\mu_f = \int_{-\infty}^{\infty} ef(e)de < \infty$ and $0 < \nu_f = \int_{-\infty}^{\infty} e^2 f(e)de < \infty$

- f is absolutely continuous with (almost everywhere) derivative f' and finite Fisher information for location $\mathscr{I}_f = \int_{\infty}^{\infty} \phi_f^2(e) f(e)de$, where $\phi_f(\cdot) = -\frac{f'(\cdot)}{f(\cdot)}$
- f is **strongly unimodal**

Efficient estimation

Assume first f is **known and parametric** (and belongs to the family of zero median unimodal distributions, denoted $\mathscr{F}_0$)

$$\ln L\left(\theta \mid \boldsymbol{y}^{(n)}, \boldsymbol{W}^{(n)}, \boldsymbol{X}^{(n)}\right) = \ln\left|\det\left(\boldsymbol{I}_n - \lambda \boldsymbol{W}^{(n)}\right)\right| + \sum_{i=1}^{n} \ln f\left(e_i^{(n)}(\theta)\right)$$

As it enjoys **LAN property**, MLE can be based on the **central sequence**:

$$\boldsymbol{\Delta}_f^{(n)}(\theta) = \begin{pmatrix} \boldsymbol{\Delta}_{f;\beta}^{(n)}(\theta) \\ \Delta_{f;\lambda}^{(n)}(\theta) \end{pmatrix} = \frac{1}{\sqrt{n}} \frac{\partial}{\partial \theta} \left\{ \ln L\left(\theta \mid \mathrm{y}^{(n)}, \boldsymbol{W}^{(n)}, \boldsymbol{X}^{(n)}\right) \right\}$$

$$= \begin{pmatrix} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \phi_f\left(e_i^{(n)}(\theta)\right) \boldsymbol{x}_i^{(n)} \\ -\frac{1}{\sqrt{n}} \operatorname{tr} \underbrace{\left(\boldsymbol{G}^{(n)}(\lambda)\right)}_{\boldsymbol{W}^{(n)}\left(\boldsymbol{I}_n - \lambda \boldsymbol{W}^{(n)}\right)^{-1}} + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \phi_f\left(e_i^{(n)}(\theta)\right) \boldsymbol{W}_{i.}^{(n)} \boldsymbol{y}^{(n)} \end{pmatrix}$$

Efficient estimation

However, **if** f is **misspecified** ($\neq g$, the true distribution), $\hat{\theta}$ is generally **not** $\sqrt{n}$ **- consistent**

- **Flexible parametric solution**: Consider a very flexible distribution to approximate well f
  - We consider **Tukey g-and-h** and **Jones and Pewsey SAS** distributions (many others could be considered)
- **Semiparametric solution**: Don't make any assumption on the distribution (apart those minimal above) and use a function **of error terms** in the construction of the central sequence.
  - This function **preserves** the i**nformation** contained in the **error term constant across distributions**.
  - In our framework, the **best function** (maximum invariant) that can be used is **based on ranks and signs** (**Hallin and Werker, 2003**)

Flexible Parametric

Flexible Parametric Solution

If we incorporate the parameters of the flexible distribution (γ), the central sequence is, up to $o_{\mathrm{P}}(1)$ given by

$$\boldsymbol{\Delta}^{(n)}(\theta) = \begin{pmatrix} \boldsymbol{\Delta}_{\beta}^{(n)}(\theta) \\ \Delta_{\lambda}^{(n)}(\theta) \\ \boldsymbol{\Delta}_{\gamma}^{(n)}(\theta) \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{n}} \frac{\partial}{\partial \beta} \left\{ \ln L\left(\theta \mid \mathbf{y}^{(n)}, \mathbf{W}^{(n)}, \mathbf{X}^{(n)}\right) \right\} \\ \frac{1}{\sqrt{n}} \frac{\partial}{\partial \lambda} \left\{ \ln L\left(\theta \mid \mathbf{y}^{(n)}, \mathbf{W}^{(n)}, \mathbf{X}^{(n)}\right) \right\} \\ \frac{1}{\sqrt{n}} \frac{\partial}{\partial \gamma} \left\{ \ln L\left(\theta \mid \mathbf{y}^{(n)}, \mathbf{W}^{(n)}, \mathbf{X}^{(n)}\right) \right\} \end{pmatrix}$$

Naturally, the dimension of $\boldsymbol{\Delta}_{\gamma}^{(n)}(\theta)$ depends on the number of parameters characterizing the flexible density function.

Flexible Parametric Solution

The **distribution of the error** term can be **characterized by** its distribution function F_γ or, equivalently, by its **quantile function**

$$Q_\gamma : (0,1) \to \mathbb{R} : u \mapsto Q_\gamma(u) = F_\gamma^{-1}(u)$$

Consequently, since

$$f_\gamma(e) = \frac{dF_\gamma(e)}{de} = \frac{d}{de}\left\{Q_\gamma^{-1}(e)\right\} = \frac{1}{Q'_\gamma\left(Q_\gamma^{-1}(e)\right)}$$

with $Q'_\gamma(u) = \frac{dQ_\gamma(u)}{du}$, the log-likelihood function may be written as :

$$\ln L\left(\theta \mid \mathbf{y}^{(n)}, \mathbf{W}^{(n)}, \mathbf{X}^{(n)}\right) = \ln\left|\det\left(\mathbf{I}_n - \lambda \mathbf{W}^{(n)}\right)\right| - \sum_{i=1}^{n} \ln Q'_\gamma\left(Q_\gamma^{-1}\left(e_i^{(n)}(\beta, \lambda)\right)\right)$$

This is **particularly useful** when the probability **density functions** are only **implicitly defined**, as is the case with Tukey's g-and-h distribution.

Flexible Parametric Solution

We therefore have the equivalent formulas:

$$\mathbf{\Delta}_{\beta}^{(n)}(\theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \widetilde{Q}_{\gamma} \left(\underbrace{Q_{\gamma}^{-1}\left(e_i^{(n)}(\beta,\lambda)\right)}_{u_i^{(n)}(\theta)} \right) \mathbf{x}_i^{(n)}$$

$$\Delta_{\lambda}^{(n)}(\theta) = -\frac{1}{\sqrt{n}} \operatorname{tr}\left(G^{(n)}(\lambda)\right) + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \widetilde{Q}_{\gamma}\left(u_i^{(n)}(\theta)\right) \mathbf{W}_{i.}^{(n)} \mathbf{y}^{(n)}$$

and

$$\mathbf{\Delta}_{\gamma_r}^{(n)}(\theta) = -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial}{\partial \gamma_r} \left\{ \ln Q'_{\gamma}\left(u_i^{(n)}(\theta)\right) \right\}$$

where

$$\widetilde{Q}_{\gamma}(u) \equiv \frac{Q''_{\gamma}(u)}{\left[Q'_{\gamma}(u)\right]^2}$$

Preliminary Consistent Estimator

As explained earlier, the LAN property allows us to achieve a **highly efficient estimator** by **refining** a **consistent preliminary estimator**.

For the **regression coefficients**, we consider a **SAR model** estimated by **2SLS** that gives a preliminary residuals $e_i^{(n)}$.

For the **parameters** of the **density functions**, we use **quantile least squares**, as proposed by **Xu et al. (2014)**.

The error terms $\varepsilon_i^{(n)} (i = 1, \ldots, n)$ are assumed to be *i.i.d.* with a distribution characterized by the quantile function $Q_\gamma(\cdot)$. Hence, considering, for $i = 1, \ldots, n$,

$$
\begin{aligned}
e_i^{(n)}\left(\widetilde{\beta}^{(n)}, \tilde{\lambda}^{(n)}\right) &= y_i^{(n)} - \left(\mathbf{x}_i^{(n)}\right)^{\mathrm{T}} \widetilde{\beta}^{(n)} - \widetilde{\lambda}^{(n)} \mathbf{W}_{i.}^{(n)} \mathbf{y}^{(n)} \\
&= y_i^{(n)} - \sum_{k=1}^{K} \widetilde{\beta}_k^{(n)} x_{ik}^{(n)} - \widetilde{\lambda}^{(n)} \mathbf{W}_{i.}^{(n)} \mathbf{y}^{(n)}
\end{aligned}
$$

Preliminary Consistent Estimator

We may **search** for a **preliminary estimate** $\widetilde{\gamma}^{(n)}$ of γ by **minimizing** the (squared) **distance** between some **residual quantiles** and the **corresponding theoretical quantiles** of the **distribution characterized** by the parameter γ :

$$\widetilde{\gamma}^{(n)} = \arg\min_{\gamma \in \Gamma} \sum_{p_i=1}^{m} \left[e_{p_i}^{(n)} \left(\widetilde{\beta}^{(n)}, \widetilde{\lambda}^{(n)} \right) - \zeta_{p_i} \right]^2$$

where $e_{p_i}^{(n)} \left(\widetilde{\beta}^{(n)}, \tilde{\lambda}^{(n)} \right)$ is the $p_i{}^{th}$ chosen sample **quantile** of the **residuals** fitted using the **preliminary consistent estimator**, ζ_{p_i} is the **corresponding theoretical quantile** and m is the number of chosen quantiles.

Let's consider **the Tukey** g**-and-**h distribution as an example

Tukey g-and-h error term distribution

Tukey (1977) introduced the g-and-h distributions, a **flexible** family of (**implicitly defined**) **density functions** derived from **transformations** of the **normal** distribution to accommodate **skewness** and **heavy tails**.

Definition

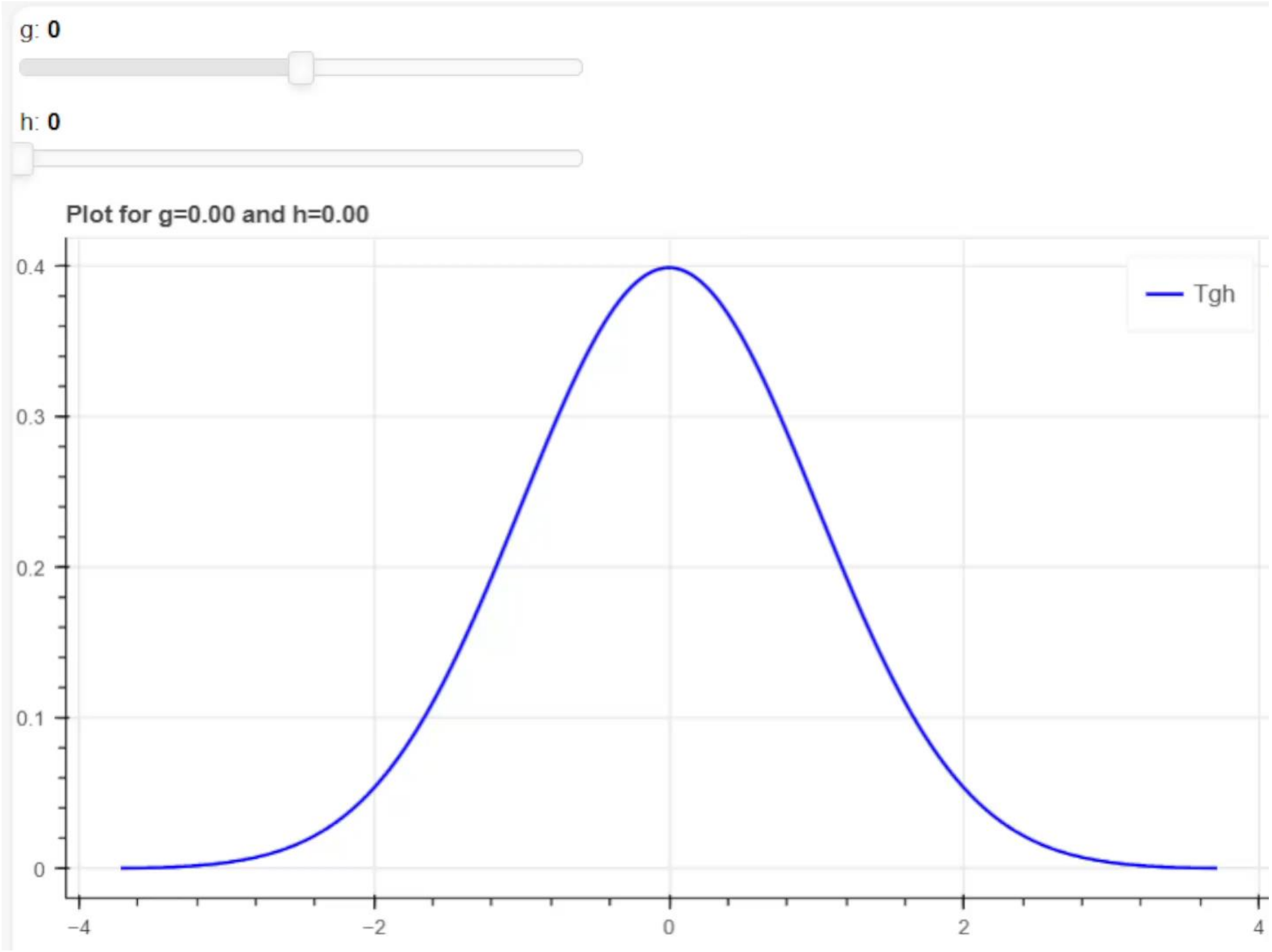
Let Z be a random variable with standard normal distribution $N(0,1)$. Define random variable X through the transformation

$$X = \xi + \sigma \tau_{g,h}(Z)$$

where $\xi \in \mathbb{R}, \sigma \in \mathbb{R}_0^+$, and $\tau_{g,h}(z) = \frac{1}{g}(\exp(gz) - 1)\exp\left(hz^2/2\right)$ with $g \in \mathbb{R}$ (for $g = 0$, $\tau_{0,h}(z) = \lim_{g \to 0} \tau_{g,h}(z) = z\exp\left(hz^2/2\right)$) and $h \in [0, 0.25)$.

Parameter ***g*** **controls** the **skewness** , while ***h*** **controls** the **tail heaviness**. To guarantee the existence of 4th order moment, **Martinez** & **Iglewicz (1984)** have shown that h should be smaller than 0.25 .

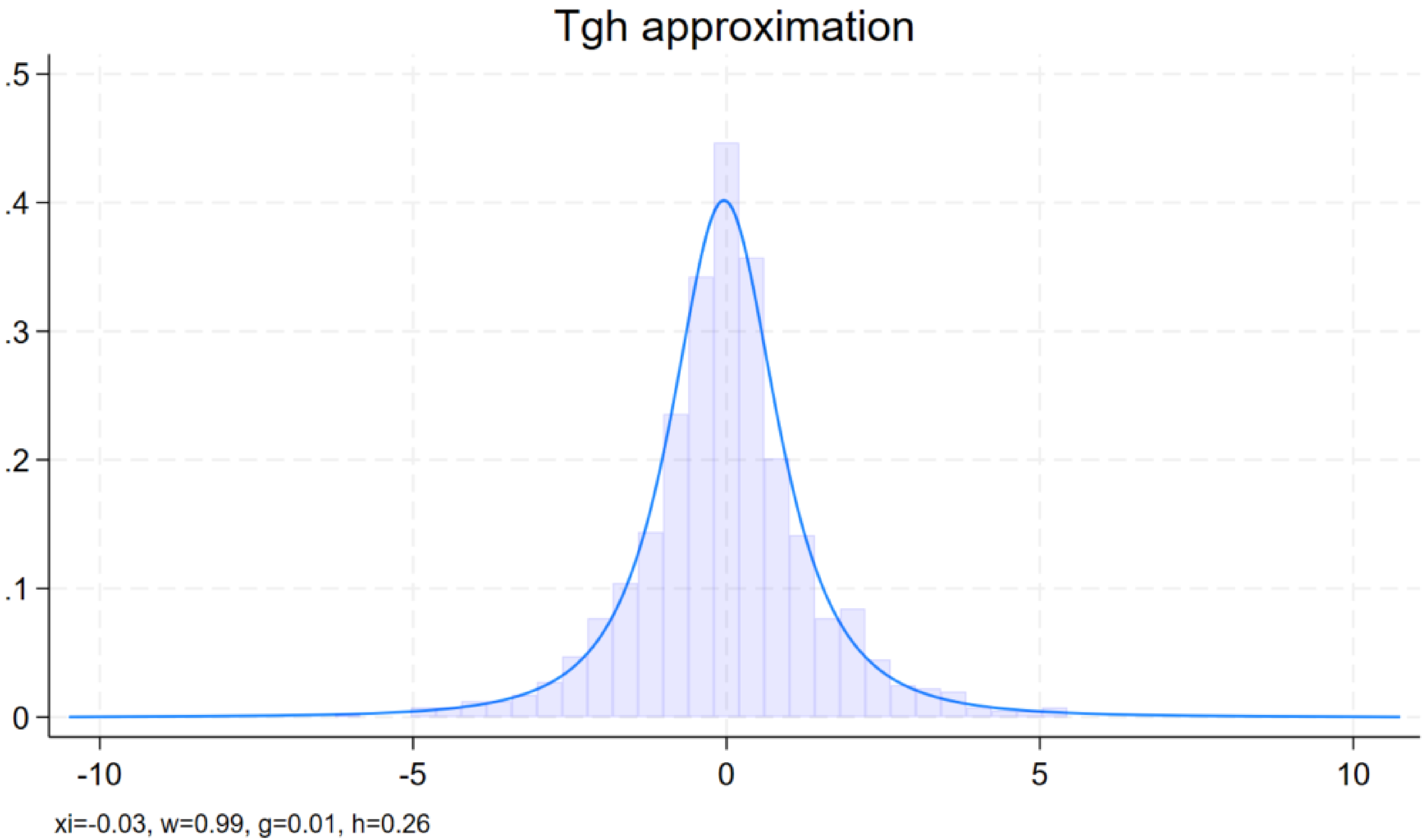
Tukey g-and-h error term distribution



Tukey g-and-h error term distribution

```
clear
set obs 1000
set seed 12345
gen x=rlaplace(0,1)

cumul(x), gen(F)
gen Z=invnorm(F)
pctile x2=x, nquantile(20) gen(Z2)
replace Z2=invnorm(Z2/100)

nl (x2={xi}+({w}/{g})*((exp({g}*Z2)-1)*exp(0.5*{h}*Z2^2))), init(w 1 g  0.1 h 0.1)
```

Tukey g-and-h error term distribution



Tukey g-and-h error term distribution

Expressions for $\Delta^{(n)}(\theta)$ and $\mathbf{I}(\theta)$ are **explicit**

$$Q_\gamma(u) = \sigma\tau_{g,h}(z_u) = \sigma\tau_{g,h}\left(\Phi^{-1}(u)\right)$$

$$Q'_\gamma(u) = \frac{\sigma}{\phi(z_u)}\left[\exp\left(\frac{hz_u^2}{2}+gz_u\right)+hz_u\tau_{g,h}(z_u)\right],$$

$$Q''_\gamma(u) = \frac{\sigma}{\phi^2(z_u)}\left[(2hz_u+z_u+g)\exp\left(\frac{hz_u^2}{2}+gz_u\right)+h\left(1+z_u^2+hz_u^2\right)\tau_{g,h}(z_u)\right],$$

$$\frac{\partial Q_\gamma(u)}{\partial\gamma_1} = \frac{\partial Q_\gamma(u)}{\partial\sigma} = \tau_{g,h}(z_u), \frac{\partial Q'_\gamma(u)}{\partial\gamma_1} = \frac{\partial Q'_\gamma(u)}{\partial\sigma} = \frac{1}{\sigma}Q'_\gamma(u)$$

$$\frac{\partial Q_\gamma(u)}{\partial\gamma_2} = \frac{\partial Q_\gamma(u)}{\partial g} = \frac{\sigma}{g}\left[z_u\exp\left(\frac{hz_u^2}{2}+gz_u\right)-\tau_{g,h}(z_u)\right],$$

$$\frac{\partial Q'_\gamma(u)}{\partial\gamma_2} = \frac{\partial Q'_\gamma(u)}{\partial g} = \frac{\sigma z_u}{\phi(z_u)}\left[\left(1+\frac{hz_u}{g}\right)\exp\left(\frac{hz_u^2}{2}+gz_u\right)-\frac{h}{g}\tau_{g,h}(z_u)\right],$$

$$\frac{\partial Q_\gamma(u)}{\partial\gamma_3} = \frac{\partial Q_\gamma(u)}{\partial h} = \frac{\sigma z_u^2}{2}\tau_{g,h}(z_u),$$

$$\frac{\partial Q'_\gamma(u)}{\partial\gamma_3} = \frac{\partial Q'_\gamma(u)}{\partial h} = \frac{\sigma z_u}{\phi(z_u)}\left[\frac{z_u}{2}\exp\left(\frac{hz_u^2}{2}+gz_u\right)+\left(1+\frac{hz_u^2}{2}\right)\tau_{g,h}(z_u)\right].$$

Jones and Pewsey SAS-distribution

Jones and Pewsey(2005) introduced a flexible family of density functions (called sinh-arcsinh or **SAS**) constructed by **transforming** a **standard normal**, to model **skewness** and **tails heaviness**.

Definition

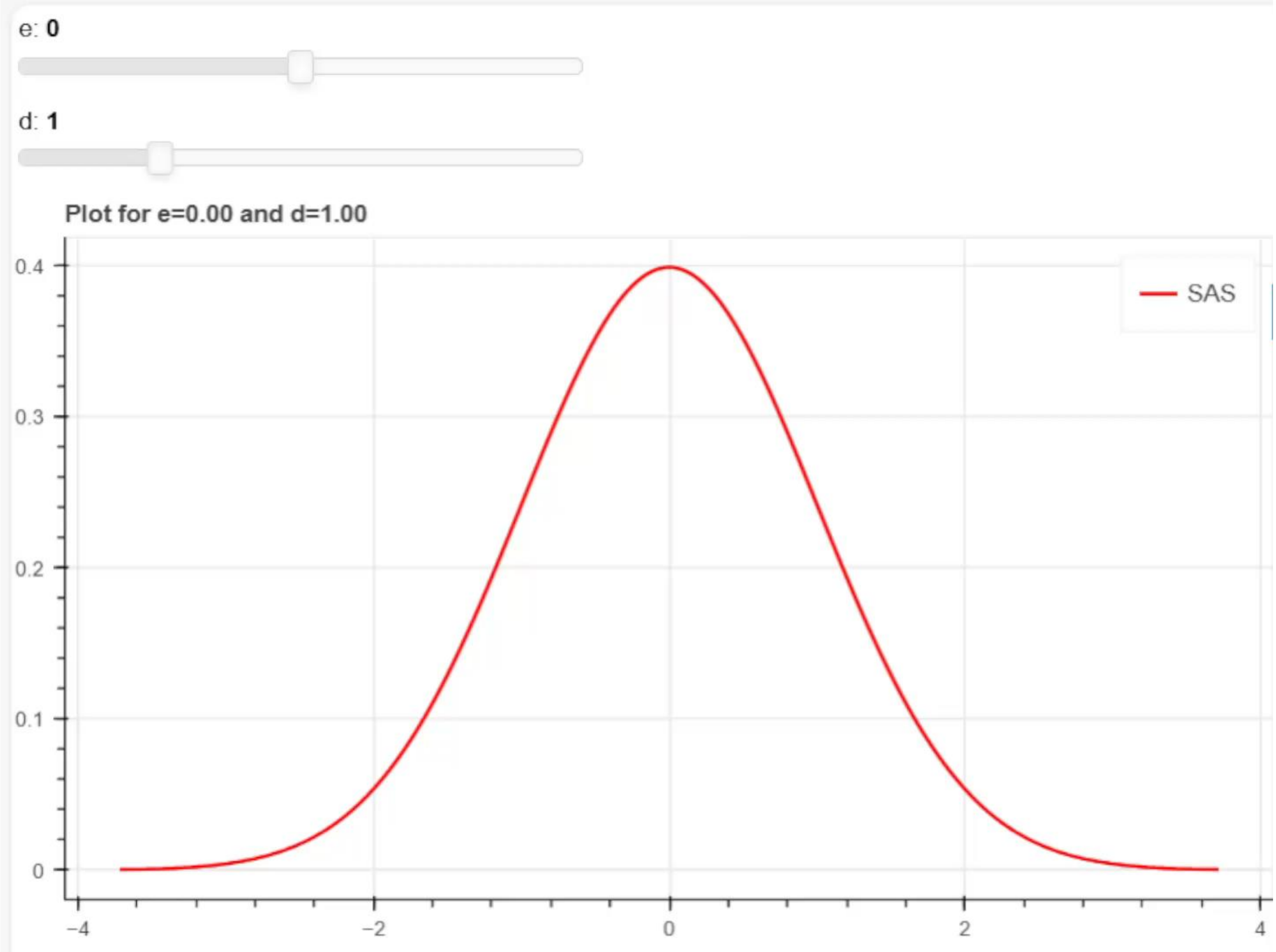
Let Z be a random variable with a standard normal distribution $N(0,1)$. Define the random variable X through the transformation

$$X = \xi + \sigma S^{-1}_{\varepsilon,\delta}(Z)$$

where $S^{-1}_{\varepsilon,\delta}(v) = S_{-\varepsilon/\delta,1/\delta}(v) = \sinh\left(\frac{1}{\delta}\sinh^{-1}(v) + \frac{\varepsilon}{\delta}\right)$, $\xi \in \mathbb{R}$ and $\sigma \in \mathbb{R}^+_0$.Variable X is said to follow a sinh−arcsinh normal distribution with location parameter ξ, scale parameter σ, and shape parameters ε and δ : $X \sim \mathrm{SAS}_{\varepsilon,\delta}(\xi,\sigma)$.

Parameter $\varepsilon \in \mathbb{R}$ **controls** the **skewness** and parameter $\delta \in \mathbb{R}^+_0$ **controls** the **heaviness of the tails** (tail weight decreases when δ increases).

SAS error term distribution



SAS error term distribution

Expressions for $\Delta^{(n)}(\theta)$ and $\mathbf{I}(\theta)$ are **explicit**

$$Q_\gamma(u) = \sigma S_{-\frac{\varepsilon}{\delta},\frac{1}{\delta}}(z_u) = \sigma S_{-\frac{\varepsilon}{\delta},\frac{1}{\delta}}\left(\Phi^{-1}(u)\right),$$

$$Q'_\gamma(u) = \frac{\sigma}{\delta\phi(z_u)\sqrt{1+z_u^2}} C_{-\frac{\varepsilon}{\delta},\frac{1}{\delta}}(z_u), C_{\varepsilon,\delta}(t) = \cosh\left(\delta\sinh^{-1}(t)-\varepsilon\right)$$

$$Q''_\gamma(u) = \frac{\sigma}{\delta^2\phi^2(z_u)(1+z_u^2)}\left[S_{-\frac{\varepsilon}{\delta},\frac{1}{\delta}}(z_u) + \delta C_{-\frac{\varepsilon}{\delta},\frac{1}{\delta}}(z_u)\frac{z_u^3}{\sqrt{1+z_u^2}}\right],$$

$$\frac{\partial Q_\gamma(u)}{\partial\gamma_1} = \frac{\partial Q_\gamma(u)}{\partial\sigma} = S_{-\frac{\varepsilon}{\delta},\frac{1}{\delta}}(z_u), \frac{\partial Q'_\gamma(u)}{\partial\gamma_1} = \frac{\partial Q'_\gamma(u)}{\partial\sigma} = \frac{1}{\sigma}Q'_\gamma(u)$$

$$\frac{\partial Q_\gamma(u)}{\partial\gamma_2} = \frac{\partial Q_\gamma(u)}{\partial\varepsilon} = \frac{\sigma}{\delta}C_{-\frac{\varepsilon}{\delta},\frac{1}{\delta}}(z_u),$$

$$\frac{\partial Q'_\gamma(u)}{\partial\gamma_2} = \frac{\partial Q'_\gamma(u)}{\partial\varepsilon} = \frac{\sigma}{\delta^2\phi(z_u)\sqrt{1+z_u^2}} S^{\partial}_{-\varepsilon/\delta,1/\delta(z_u)},$$

$$\frac{\partial Q_\gamma(u)}{\partial\gamma_3} = \frac{\partial Q_\gamma(u)}{\partial\delta} = \frac{(-\sigma)}{\delta}C_{-\frac{\varepsilon}{\delta},\frac{1}{\delta}}(z_u)\left(\frac{1}{\delta}\sinh^{-1}(z_u)+\frac{\varepsilon}{\delta}\right),$$

$$\frac{\partial Q'_\gamma(u)}{\partial\gamma_3} = \frac{\partial Q'_\gamma(u)}{\partial\delta} = \frac{(-\sigma)}{\delta^2\phi(z_u)\sqrt{1+z_u^2}}\left[C_{-\frac{\varepsilon}{\delta},\frac{1}{\delta}}(z_u) + S_{-\frac{\varepsilon}{\delta},\frac{1}{\delta}}(z_u)\left(\frac{1}{\delta}\sinh^{-1}(z_u)+\frac{\varepsilon}{\delta}\right)\right]$$

Semiparametric

Semiparametric Solution

Transform the **error terms** in the **central sequence** construction using a **distribution-independent** function.

For $f \in \mathscr{F}_0$, **ranks and signs** are the **best transformation**, i.e. yielding a **semiparametrically efficient** estimator (under f) (**Hallin and Werker, 2003**).

The idea is to condition $\boldsymbol{\Delta}_f^{(n)}(\theta)$ on ranks and signs to obtain a semiparametric central sequence, which leads to a semi-parametric efficient estimator of θ, under f : $E\left[\boldsymbol{\Delta}_f^{(n)}(\theta) \mid \mathrm{N}^{(n)}(\theta), \mathrm{R}^{(n)}(\theta)\right] = \tilde{\boldsymbol{\Delta}}_f^{(n)*}(\theta) + o_{\mathrm{P}}(1)$ with $\widetilde{\boldsymbol{\Delta}}_f^{(n)*}(\theta) \sim N(\mathbf{0}, I_f^*(\theta))$, for any distribution in $\mathscr{F}_0$.

To obtain fully semiparametric efficient estimator (i.e. efficient not only under f) we need a **semiparametric central sequence** $\widetilde{\boldsymbol{\Delta}}^{(n)*}(\theta)$ **asymptotically equivalent** to $\widetilde{\boldsymbol{\Delta}}_f^{(n)*}(\theta)$ (under f), but **which does not depend on** f anymore. Density f is replaced by a kernel estimate.

Semiparametric Solution

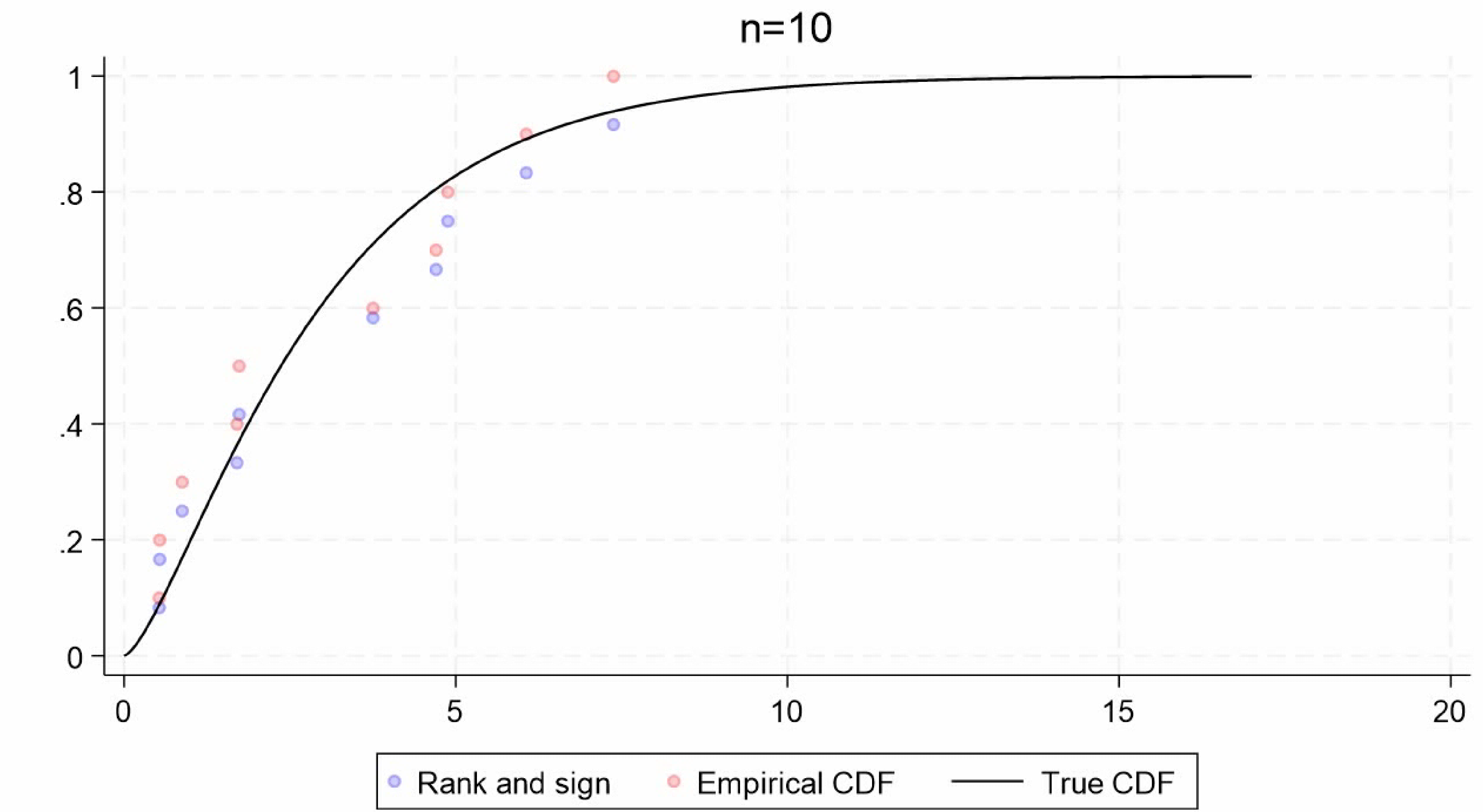
Relying on a **preliminary consistent** (and potentially inefficient) **estimator, sign-and-rank of residuals** can be calculated as :

$$\tilde{R}_i(\widetilde{\theta}) = \mathbb{I}\left[s_i(\widetilde{\theta}) = -1\right]\left\{\frac{1}{2}\frac{R_i(\widetilde{\theta})}{N_-(\widetilde{\theta})+1}\right\} + \mathbb{I}\left[s_i(\widetilde{\theta}) = +1\right]\left\{\frac{1}{2}+\frac{1}{2}\frac{R_i(\widetilde{\theta}) - \left(n - N_+(\widetilde{\theta})\right)}{N_+(\widetilde{\theta})+1}\right\}$$

where $R_i(\widetilde{\theta})$ is the rank of residual $e_i(\widetilde{\theta})$ and $N_-(\widetilde{\theta})$ and $N_+(\widetilde{\theta})$ are respectively the **number of negative and positive** residuals

The **sign-and-rank** and the standard **empirical CDF converge almost surely** (for $n \to \infty$) to the **exact CDF**.

Semiparametric Solution



Thanks to **Robert Grant** to show how easy it is to generate animated graphs in Stata:

Grant, R. (2012). “Producing animated graphs from Stata without having to learn any specialized software”, *United Kingdom Stata Users' Group Meetings*. https://ideas.repec.org/p/boc/usug12/02.html

Stata

Codes in Stata

We have **programmed** these estimators **in Stata** (as well as in Matlab).

They will be **made available** (on SSC) **only after** the underlying **theoretical papers** have been **peer-reviewed and published**.

The general **syntax** is the following:

cmd depvar indepvars , dvarlag(*spatname*) [options]

where *cmd* is the command name (***sarnp*** for the **semiparametric** estimator and ***sarflex*** for the **flexible parametric**)

In the dvarlag(*spatname*) option, the user declares a npspat **weighting matrix** that defines a **connectivity lag** of the dependent variable (like in spregress).

Data have to be spset.

Codes in Stata

Several **command specific options** are available. In the **flexible parametric**, `est()` option allows to declare TGH (the default), SAS or L1.

For **post estimation**, several predictions are possible:

`rform` generates the **reduced form** (or total effect) mean (default).

`direct` generates the **direct mean**.

`indirect` generates the **indirect mean**.

`xb` generates the **linear prediction**.

`naive` generates the **naive prediction**.

`residuals` generates **residuals**.

`wy` generates the **spillover** effect of the dependent variable.

The `margins` commands can thus be applied where applicable.

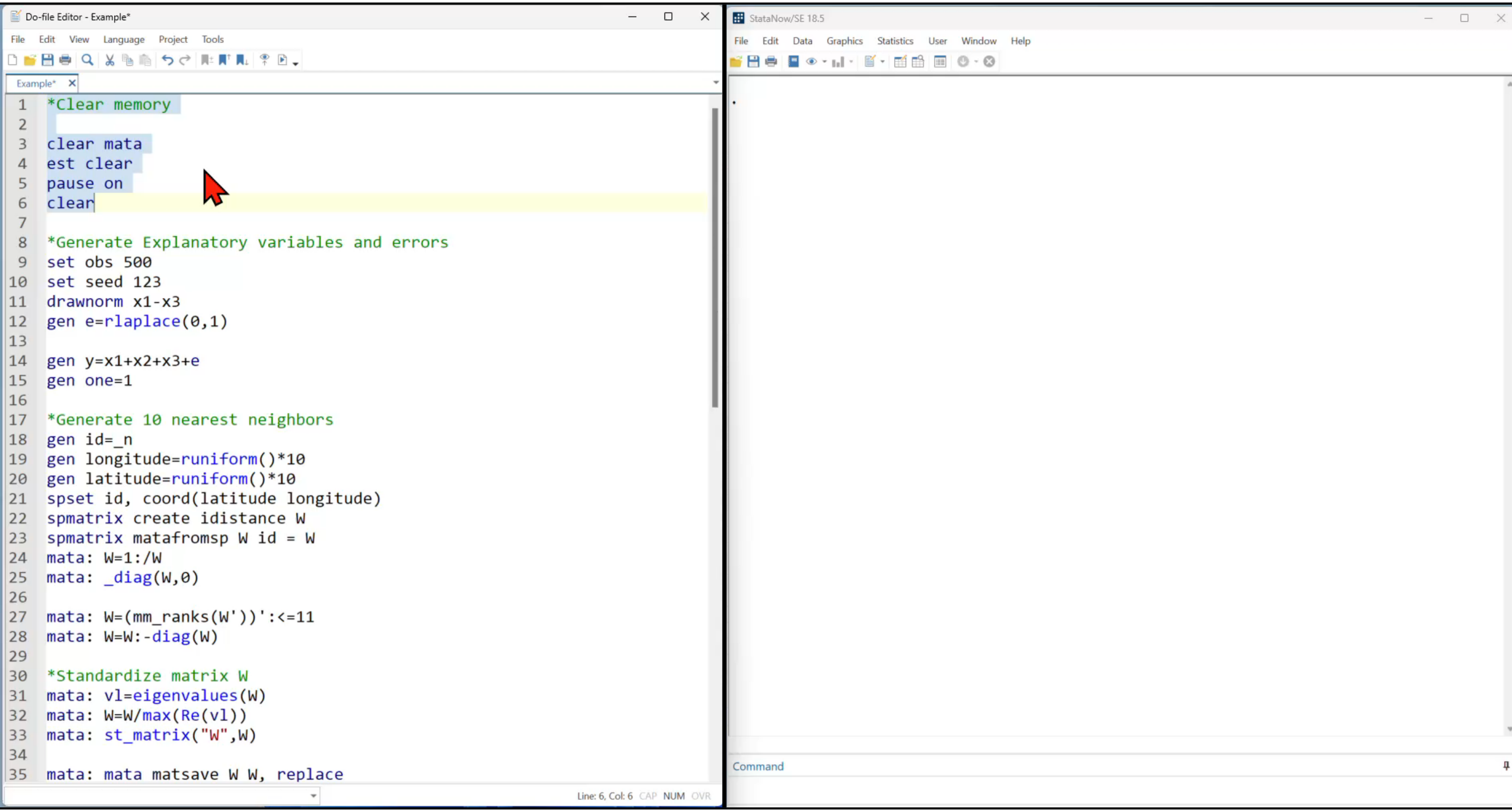
Codes in Stata

```
set seed 123
drawnorm x1-x3
gen e=rlaplace(0,1)

gen y=x1+x2+x3+e
gen one=1

gen id=_n
gen longitude=runiform()*10
gen latitude=runiform()*10
spset id, coord(latitude longitude)
spmatrix create idistance W
spmatrix matafromsp W id = W
mata: W=1:/W
mata: _diag(W,0)

mata: W=(mm_ranks(W'))':<=11
mata: W=W:-diag(W)

mata: vl=eigenvalues(W)
mata: W=W/max(Re(vl))
mata: st_matrix("W",W)

mata: mata matsave W W, replace
mata: mata matsave id id, replace

putmata x* one e, replace

mata {
rho=0.6
y=luinv(I(rows(W))-rho*W)*(1:+x1:+x2:+x3:+e)
}

getmata y, replace
```

Codes in Stata

```
spset

spregress y x*, ml dvarlag(W)
est store ML
pause

sarnp y x*, dvarlag(W)
est store NP
pause

sarflex y x*, dvarlag(W)
est store TGH
pause

sarflex y x*, dvarlag(W) est(SAS)
est store SAS
pause


estout *, cells(b(star fmt(%9.3f)) se(par)) legend label collabels(none) varlabels(_cons Constant) replace
```

Codes in Stata



Codes in Stata

	ML	NP	TGH	SAS
x1	1.018***	1.025***	1.033***	1.032***
	(0.068)	(0.056)	(0.063)	(0.062)
x2	1.116***	1.093***	1.094***	1.102***
	(0.068)	(0.057)	(0.063)	(0.063)
x3	0.989***	1.012***	1.002***	1.002***
	(0.063)	(0.052)	(0.059)	(0.058)
Constant	1.341***	0.917**	1.194***	1.145***
	(0.363)	(0.294)	(0.315)	(0.321)
Wy	0.487***	0.658***	0.543***	0.548***
	(0.140)	(0.106)	(0.116)	(0.115)

* $p<0.05$, ** $p<0.01$, *** $p<0.001$

Simulations

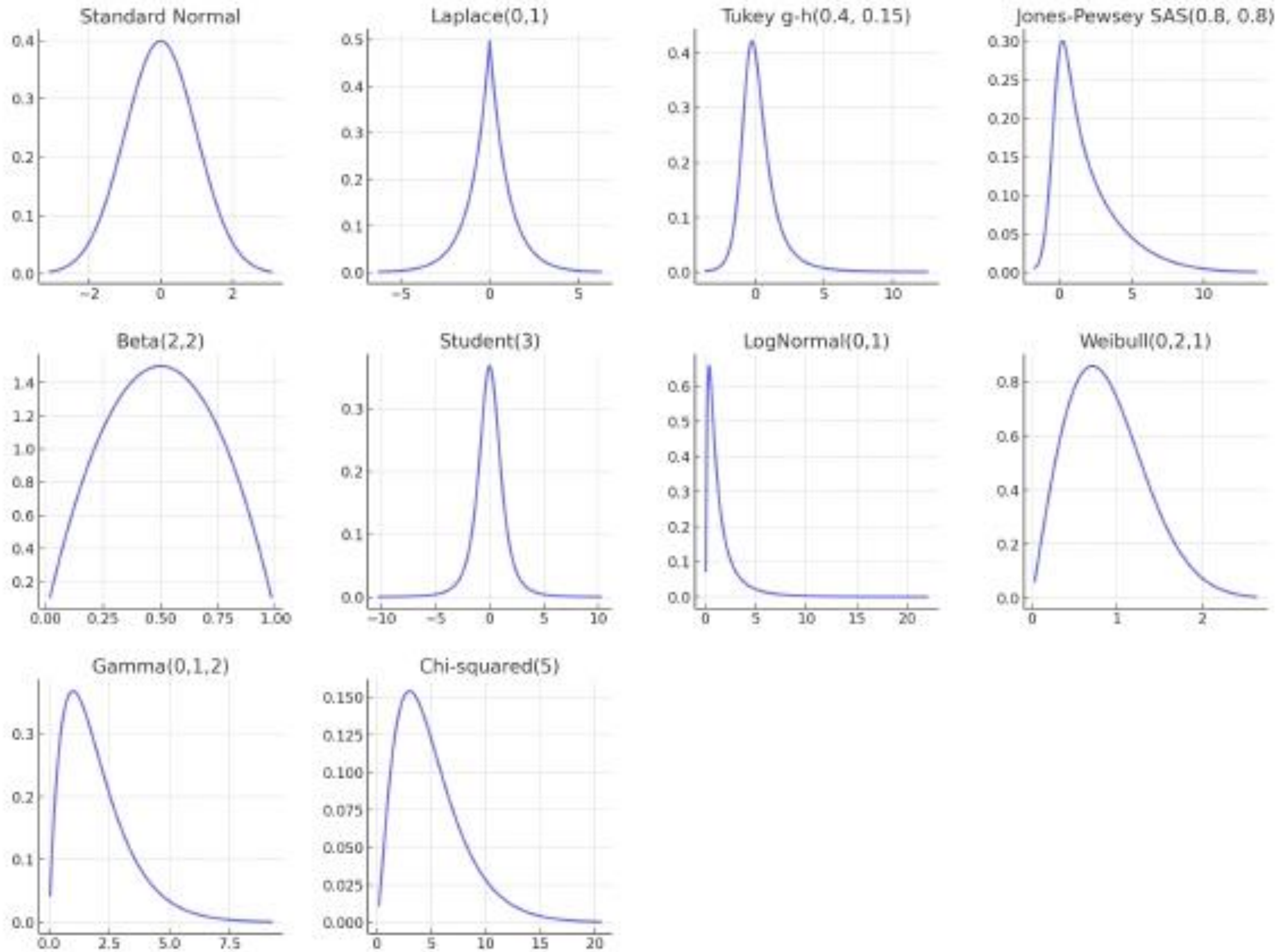
Simulations

Simulations setup

$$y_i = \lambda W_{i.} y + 1 + x_i + \varepsilon_i$$
$$X \sim N(0,1), n = 300; 500; 900$$

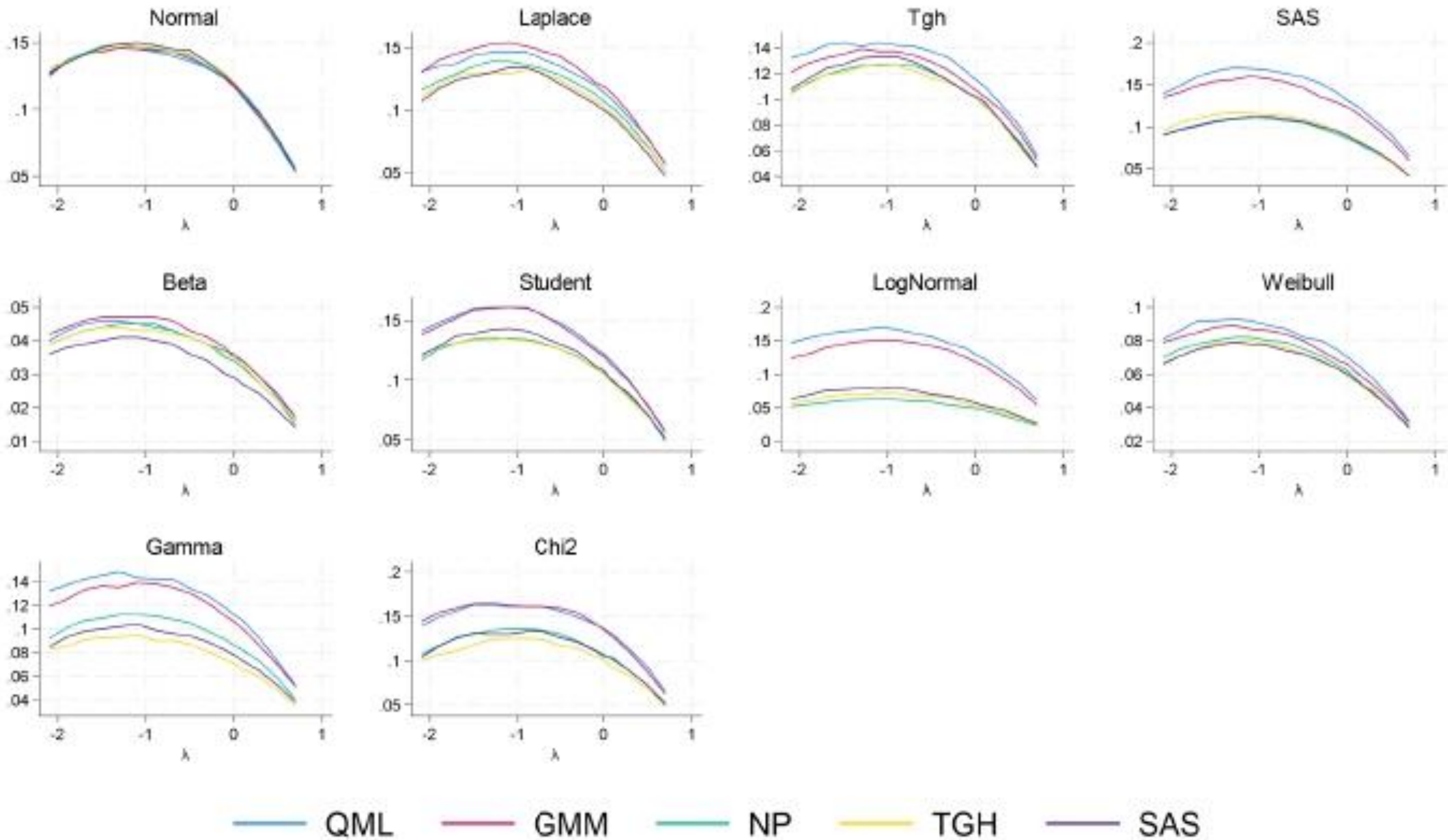
- **1000 replications**
- **10 Nearest neighbors** interaction matrix
- **20 quantiles** used to compute the **initial value** of the **distribution parameters**
- All proposed estimators, start with **2SLS**
- **Measured Bias**: **difference** between the **true value** and the **median value of estimators** over the replications
- **Measured Variability**: **Interquartile range** of the estimators divided by 1.349

Considered error distributions

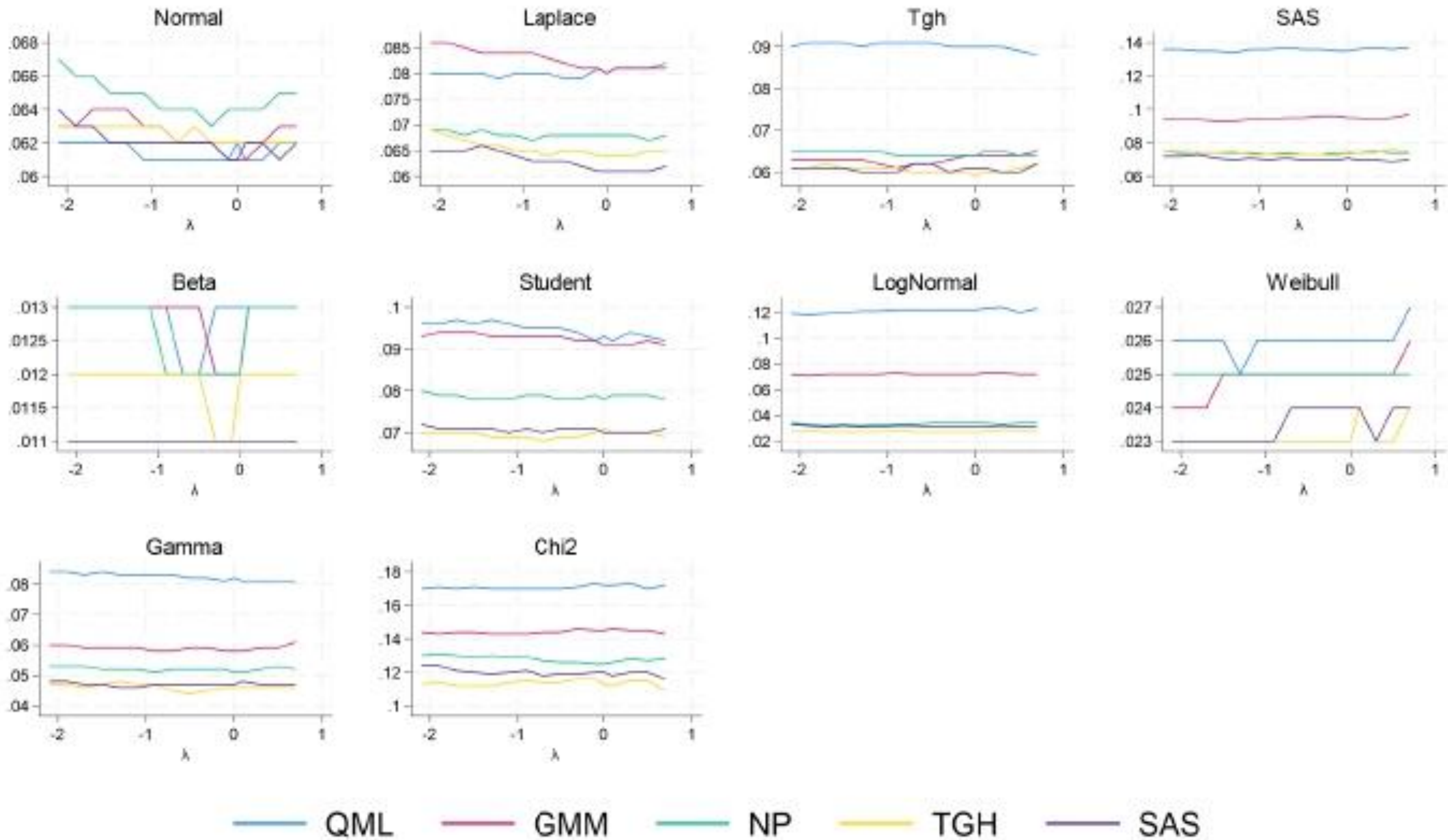


Simulations - Efficiency

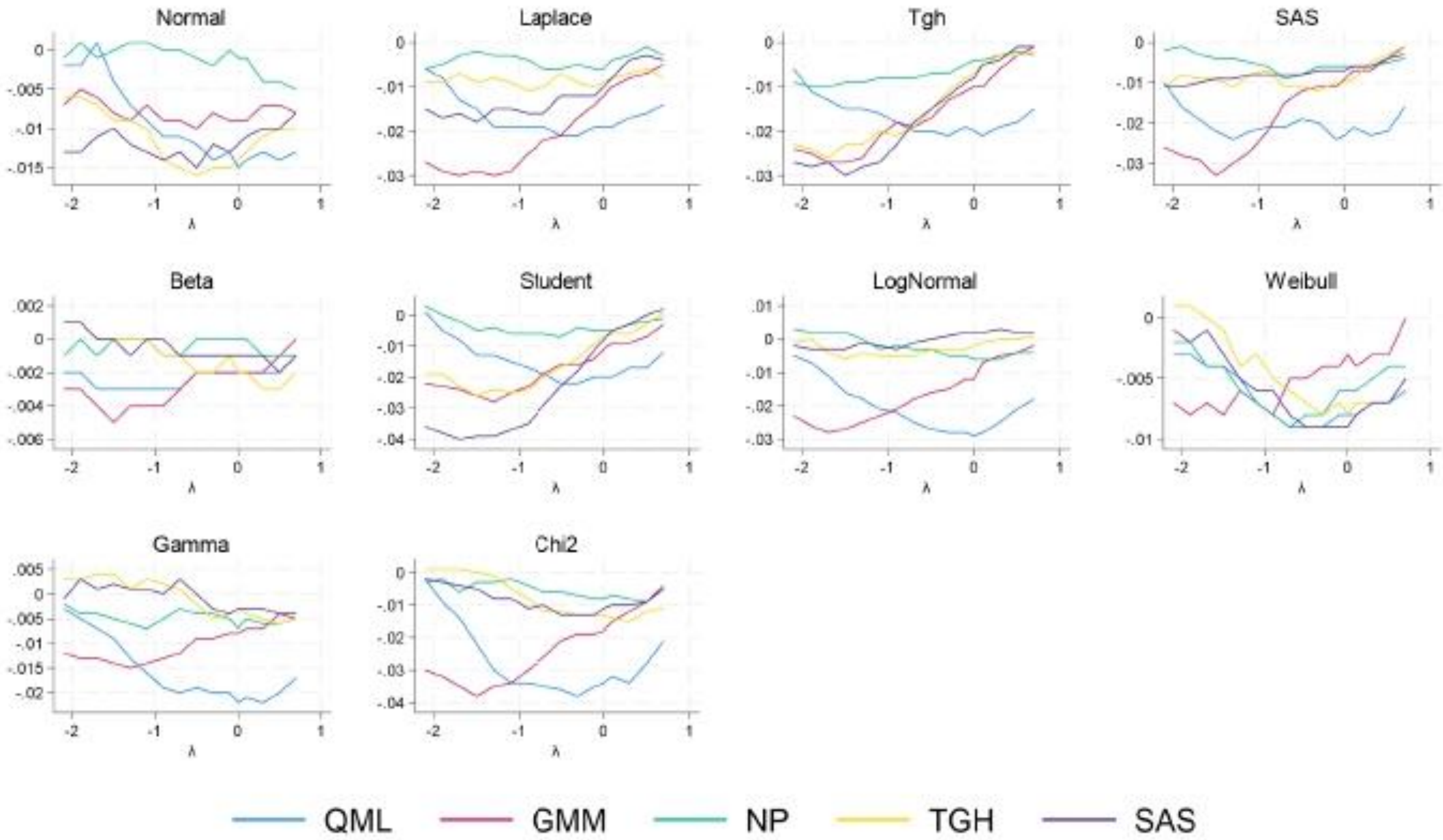
Autoregressive parameter, n=300



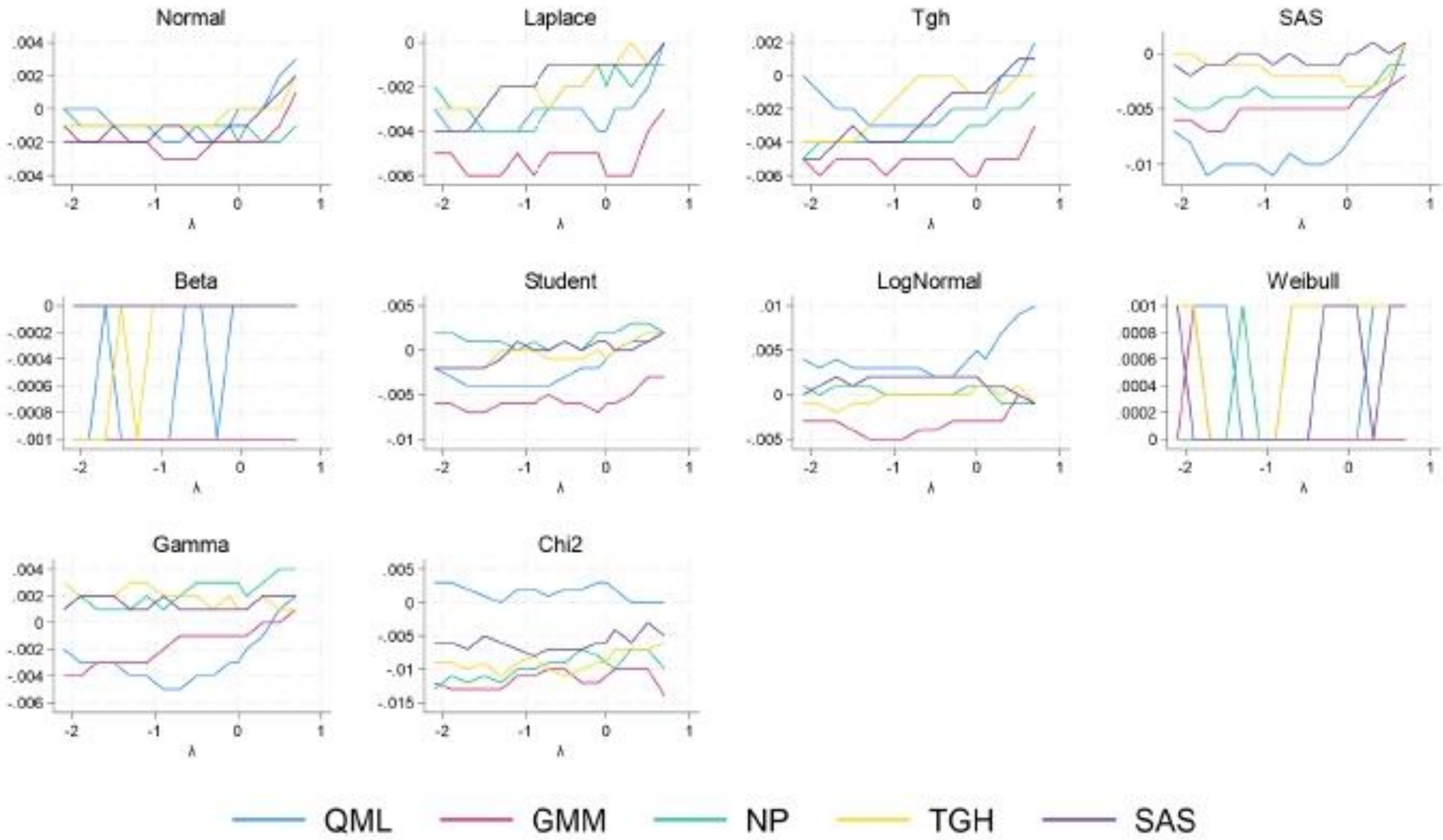
Regression parameters, n=300



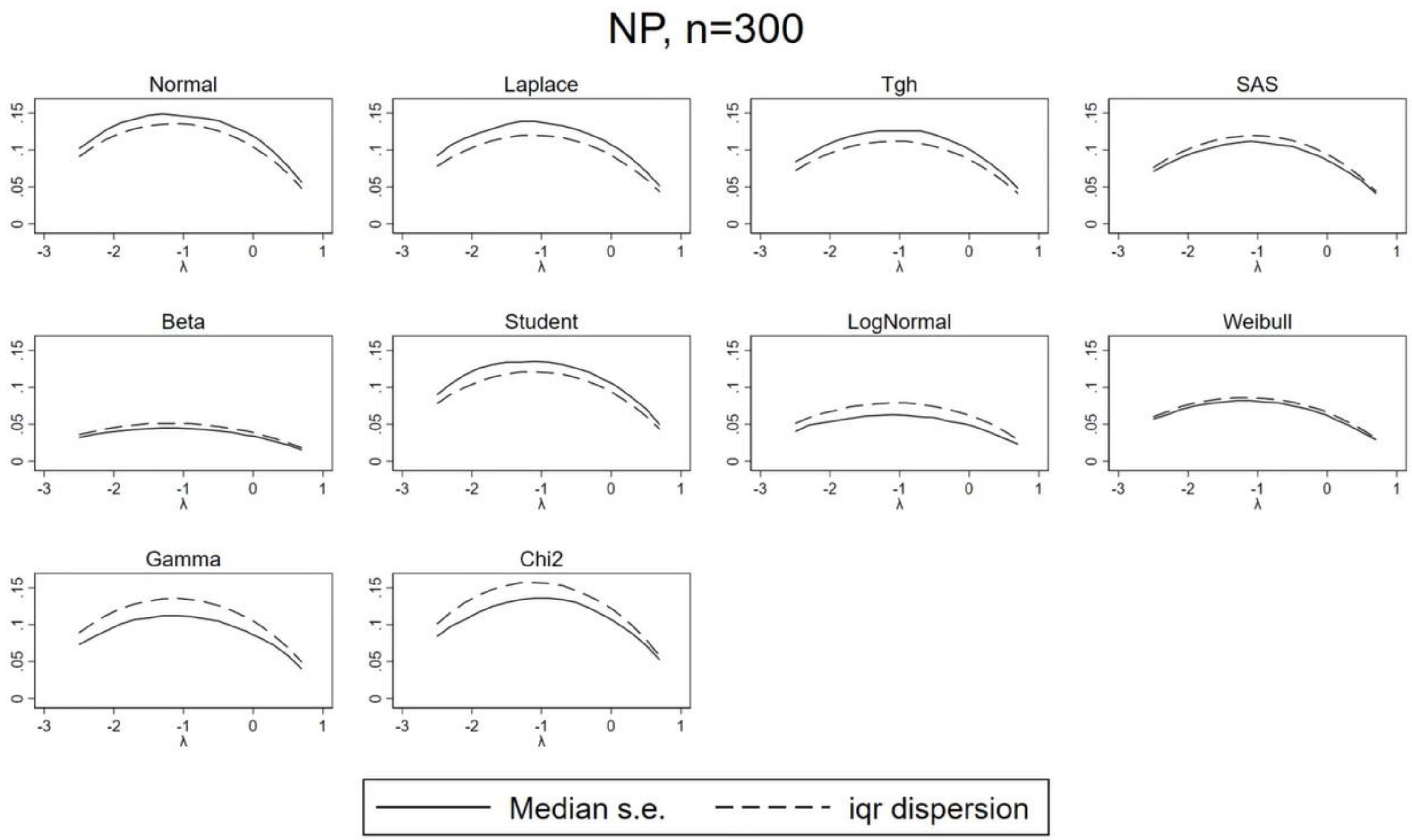
Autoregressive parameter, n=300



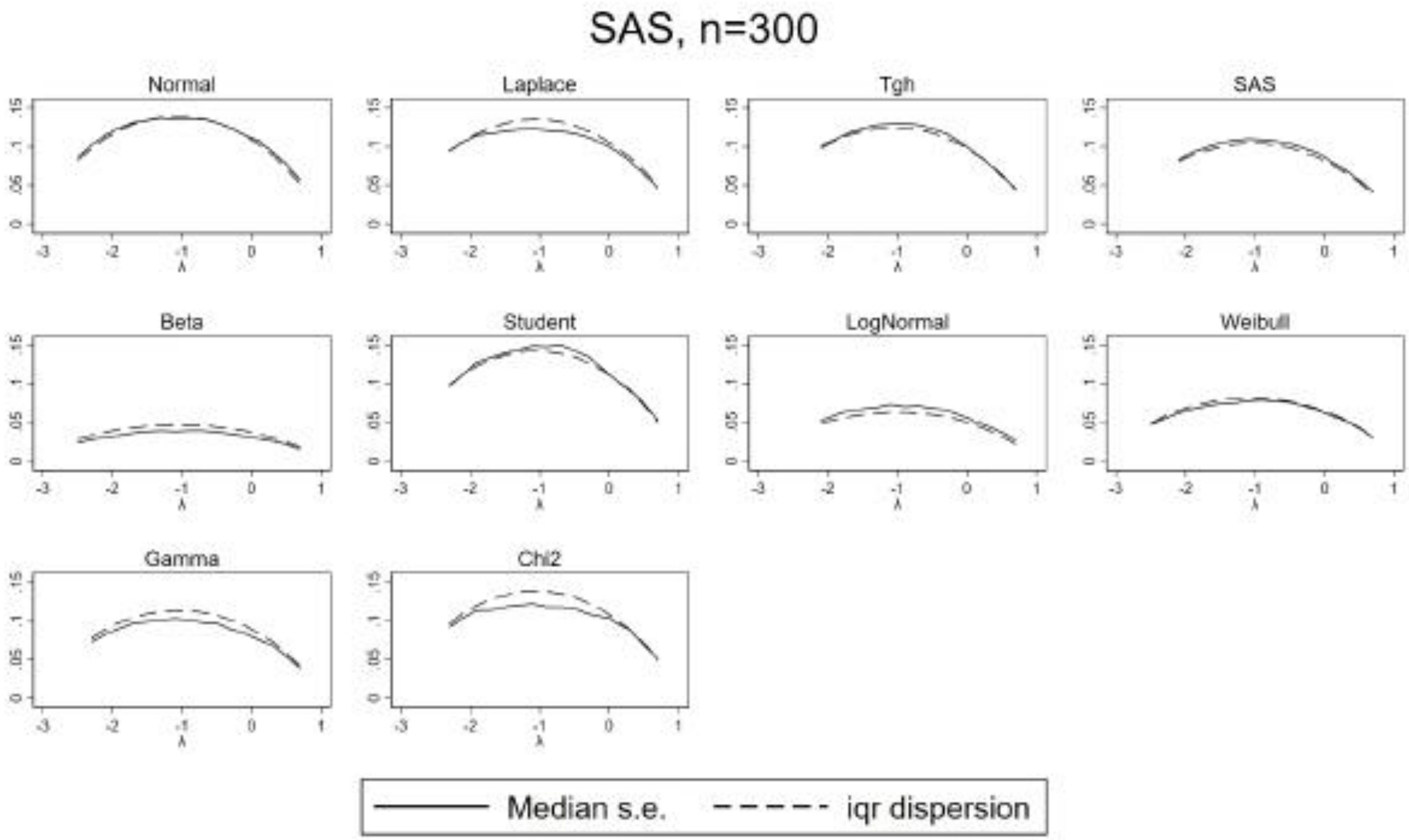
Regression parameters, n=300



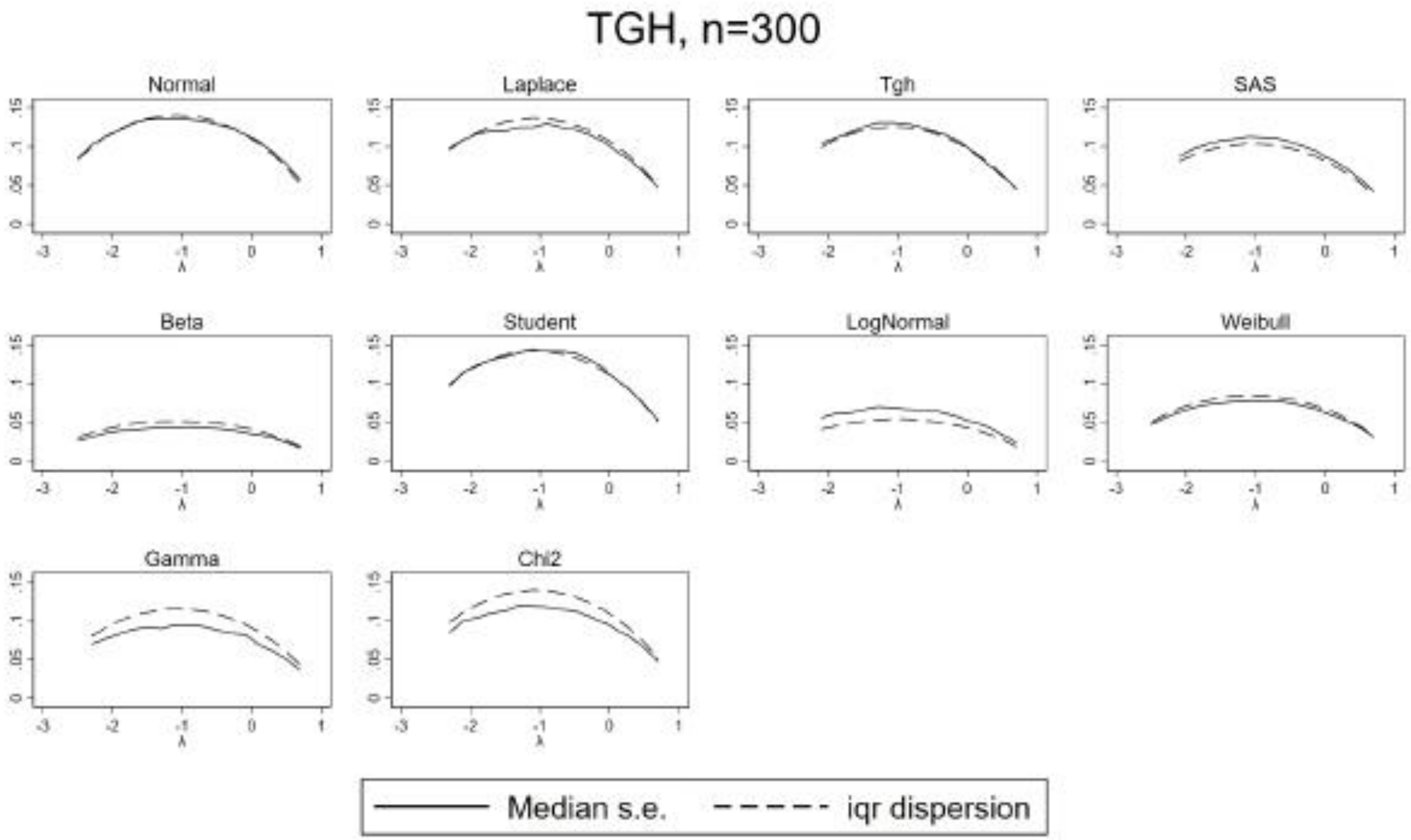
S.E. of the Autoregressive parameter, n=300, NP



S.E. of the Autoregressive parameter, n=300, SAS



S.E. of the Autoregressive parameter, n=300, Tgh



Illustration

Illustration

Behrens, Ertur and Koch (2012) estimate a trade model using **spatial econometrics** methods to assess the effect of the **Canada-US border** on **trade flows**.

Dataset on **40 regions** (30 US States and 10 CAN Provinces), **n=1600**.

Fitted **model**:

$$\ln(Z_{ij}) = \beta_0 + \beta_1 d_{ij} + \beta_2 \ln(w_i) + \beta_3 b_{ij} + \lambda \overbrace{\textstyle\sum_{k\neq i}^{n} \frac{L_k}{L} \ln\left(Z_{kj}\right)}^{\mathbf{Wy}} + \varepsilon_{ij}$$

- Z_{ij}: **Manufacturing exports** from i to j, standardized by GDP
- d_{ij}: **Distance** between i and j
- w_i: Average **hourly manufacturing wage** in region i
- b_{ij} : 1 if i is **CAN** and j is **US** and vice-versa
- L_i : **Population** in region i.

There is then an additional *internal distance* control variable.

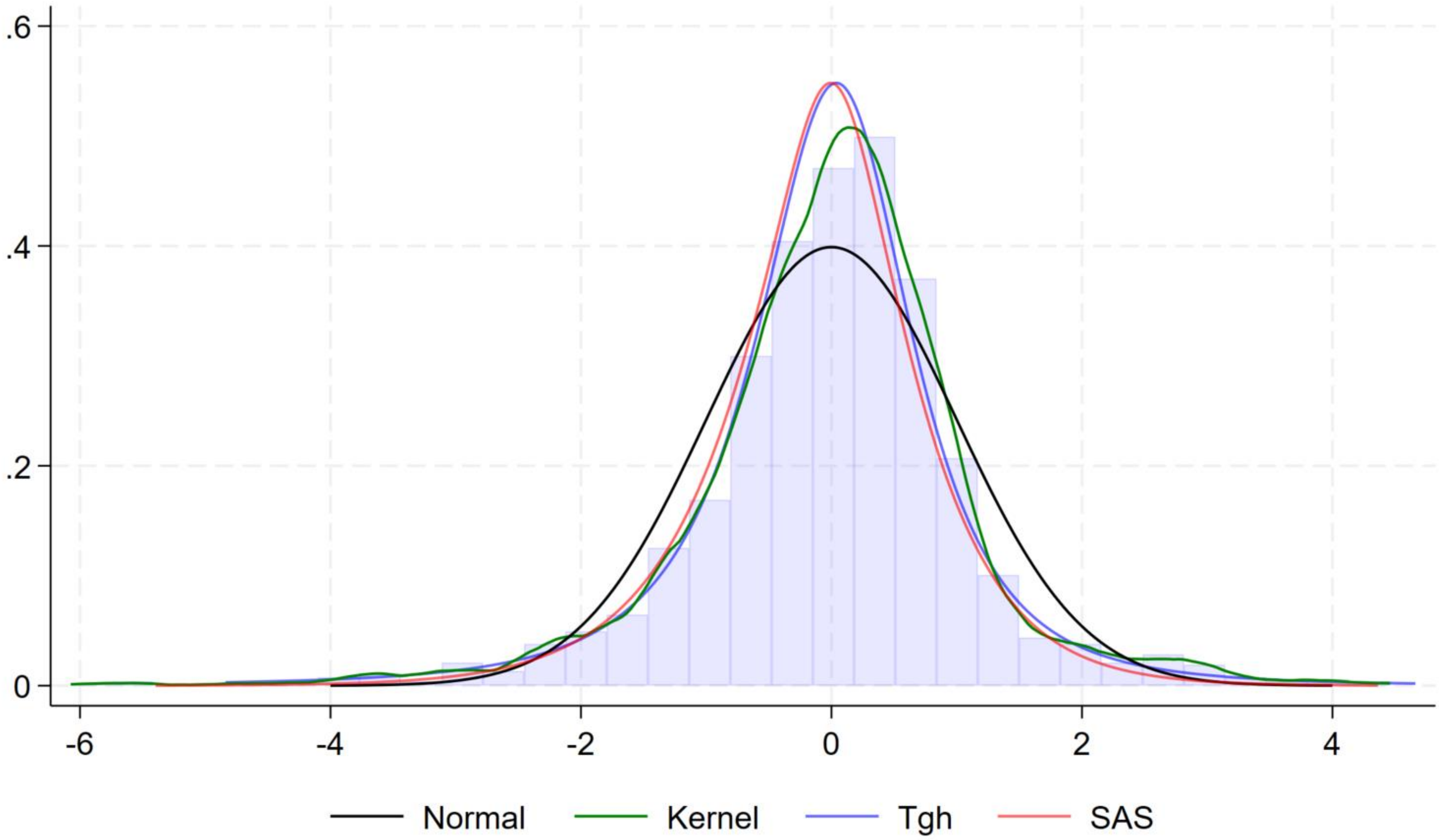
Illustration

The results are:

	TSLS	ML	NP	Tgh	SAS
d	-1.219***	-1.223***	-1.208***	-1.199***	-1.206***
	(0.034)	(0.034)	(0.024)	(0.026)	(0.025)
lnw	-1.150***	-1.173***	-1.264***	-1.188***	-1.196***
	(0.177)	(0.177)	(0.124)	(0.135)	(0.130)
b	-1.056***	-1.052***	-1.199***	-1.191***	-1.189***
	(0.066)	(0.066)	(0.046)	(0.050)	(0.049)
intdist	-16.979***	-16.950***	-17.689***	-17.144***	-17.060***
	(0.163)	(0.162)	(0.114)	(0.124)	(0.120)
Constant	-14.358***	-13.902***	-12.230***	-12.753***	-12.699***
	(0.682)	(0.667)	(0.472)	(0.523)	(0.510)
Wy (λ)	0.006	0.029	0.108***	0.090***	0.090***
	(0.027)	(0.026)	(0.019)	(0.021)	(0.021)

* $p<0.05$, ** $p<0.01$, *** $p<0.001$

Illustration



Conclusion

Estimating **linear models with spillovers** requires **specific techniques** like **two-stage least squares**, **generalized method of moments** or **quasi-maximum likelihood**, as ordinary least squares is generally not applicable.

Maximum likelihood estimation is **efficient** when the error **distribution is known**, but becomes **infeasible** with **unknown distributions**; **quasi-maximum likelihood** remains **consistent** (unders normality) but **not efficient**.

Two **estimators** based on **Local Asymptotic Normality** are proposed, achieving high efficiency .

Monte Carlo experiments demonstrate that these **estimators outperforms existing methods** under **non-normal error** distributions.

Work in progress considers **extending** the model **to** incorporate **heteroskedasticity** and/or **non-independence** between **errors**.

References

Behrens, K., Ertur, C. and Koch, W. (2012), 'Dual gravity: using spatial econometrics to control for multilateral resistance', *Journal of Applied Econometrics*, 27, pp. 773–794.

Bramoullé, Y., Djebbari, H. & Fortin, B. (2009), 'Identification of peer effects through social networks', *Journal of Econometrics*, 150, pp. 41-55.

Canay, I. (2021), *Econ 481 - Lecture 8: Local Asymptotic Normality* [Video]. YouTube. https://www.youtube.com/watch?v=-S4s1JiMWnQ

Hallin and Werker, 2003, 'Semi-parametric efficiency, distribution-freeness and invariance', *Bernouilli*, 9(1), pp. 137–165.

Jones, C. & Pewsey, A. (2009), 'Sinh-arcsinh distributions', *Biometrika*, 96, pp. 761–780.

Kelejian, H. H. Prucha, I. R. (1998), 'A generalized spatial two stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances', *Journal of Real Estate Finance and Economics*, 17, pp. 99-121.

References

Le Cam, L. (1960), 'Locally asymptotically normal families of distributions', *University of California Publications in Statistics* 3, pp. 37–98.

Lee, L.-f. (2004), 'Asymptotic distributions of Quasi-Maximum Likelihood estimators for spatial autoregressive models', *Econometrica*, 72, pp. 1899–1925.

Liu, X., Lee, L.-f. & Bollinger, C. R. (2010), 'An efficient GMM estimator of spatial autoregressive models', *Journal of Econometrics* 159, pp. 302–319.

Martinez, J. & Iglewicz, B. (1984), 'Some properties of the tukey g-and-h family of distributions', *Communications in Statistics - Theory and Methods* 13, pp. 353–369.

Robinson, P. M. (2010), 'Efficient estimation of the semiparametric spatial autoregressive model', *Journal of Econometrics* 157, pp. 6–17.

Tukey, J. W. (1977), *Modern techniques in data analysis*, pp. 761–780.

Xu, Y., Iglewicz, B. & Chervoneva, I. (2014), 'Robust estimation of the parameters of g-and-h distributions, with applications to outlier detection', *Computational Statistics and Data Analysis* 75, pp. 66–80.