Efficient estimation of regression models with spillovers

Flexible parametric and semi-parametric approaches

Nicolas Debarsy [1] Vincenzo Verardi [2] Catherine Vermandele [3]

[1]CNRS, LEM UMR 9221

[2]Université catholique de Louvain

[3]Université Libre de Bruxelles

Stata Conference, KU Leuven
September 9 2025

Introduction

- SAR models aimed at explicitly accounts for interactions/spillovers between cross-sectional units
- Ex: Social networks, corporate finance, local public policies

$$y_i = \lambda \boldsymbol{W}_{.\boldsymbol{i}}\boldsymbol{y} + \boldsymbol{x}_{\boldsymbol{i}}'\beta + \boldsymbol{W}_{.\boldsymbol{i}}\boldsymbol{X}\kappa + \varepsilon_i$$

- W models interactions between observations (geography, peers, socio-economic indicators, etc.)
- In the social network literature:
  - Endogenous effects (i's outcome depend on j's outcomes) $\boldsymbol{W}_{\boldsymbol{i}.}\boldsymbol{y} = \sum_j w_{ij} y_j$

  $\Rightarrow$ λ measures the intensity of interactions between units
  - Contextual effects (i's outcome depend on j's characteristics) $\boldsymbol{W}_{\boldsymbol{i}.}\boldsymbol{X}$

  $\Rightarrow$ κ quantify local spillovers

Introduction

Due to the endogenous effect $\boldsymbol{W_{i.} y}$, OLS cannot generally be used.

- If the innovation distribution is known, MLE is the best estimator
- If the distribution is unknown, MLE not feasible, while if the distribution is not correct, MLE is generally inconsistent

$\Rightarrow$ Lee (Ecta, 2004) derives a quasi-ML estimator, based on Normality, which allows for deviations from the normal distribution.

- Liu et al (2010) developed a GMM estimator which may bring a **gain in efficiency** compared to QML.
- Robinson (2010) develops an adaptive estimator, more efficient than QML.

We propose **two** alternative, **efficient estimators**, derived from the Local Asymptotic Normality (LAN) theory (Le Cam, 1960):

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

Local Asymptotic Normality

- 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 **is normally distributed**.
- This helps in estimating parameters and testing hypotheses by using the properties of the normal distribution.

Important implication of LAN property

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

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

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

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

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

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

Efficient SAR model without the normal distribution

- Consider now the following linear model with endogenous effects (contextual effects may be included). For $i = 1, \cdots, n$

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

- Assumptions wrt. the interaction matrix and regressors are identical to Lee (Ecta, 2004).
- However, we **relax the normality assumption** of the distribution of $\varepsilon_1^{(n)}, \ldots, \varepsilon_n^{(n)}$, but still **maintain iid**.
- We first propose an efficient estimator for any parametric f (satisfying some regularity conditions) and then relax the parametric assumption on f.

Efficient estimation

Assume first f is **known and parametric**

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

The **distribution of the error** term can be characterized by its **quantile function** (useful if density is not explicit)

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

Consequently,

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

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

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

Efficient parametric estimation

- Recall that we **don't** maximize the log-likelihood function!
- Instead, we will compute the one-step efficient estimator:

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

with $\boldsymbol{\Delta}^{(n)}(\theta)$ is **central sequence** (normalized score function) and $\mathbf{I}(\theta)$ is the information matrix computed as:

$$\lim_{n\to\infty} \mathbb{E}\left[\boldsymbol{\Delta}^{(n)}(\theta)\left(\boldsymbol{\Delta}^{(n)}(\theta)\right)'\right],$$

⇒ We only need a preliminary $\sqrt{n}-$ consistent estimator ($\widetilde{\theta}^{(n)}$), and the FOC and we are mostly done

(only requires additional integrals to compute the Information matrix)

Efficient parametric estimation

- If unsure about the correct parametric distribution, use **flexible** distributions

  We develop the formulas and code for

  1. The **Tukey g-and-h** distribution (non explicit density)
  2. The Jones and Pewsey SAS distribution (numerical convergence problems by classical ML)

  But many others could be considered

Tukey g-and-h error term distribution

- Transformation of the normal distribution, introduced by Tukey (1977), to accommodate **skewness** and **heavy tails**.
- Let Z be a random variable with standard normal distribution $N(0,1)$.

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

where $\tau_{g,h}(z) = \frac{1}{g}(\exp(gz) - 1) \exp\left(hz^2/2\right)$

- $\xi \in \mathbb{R}, \sigma \in \mathbb{R}_0^+$,
- $g \in \mathbb{R}$ **controls** the **skewness**

For $g = 0$, $\tau_{0,h}(z) = \lim_{g \to 0} \tau_{g,h}(z) = z \exp\left(hz^2/2\right))$
- $h \in [0, 0.25)$ **controls** the **tail heaviness**.
- $h < 0.25$ to guarantee 4^{th} moment exists (Martinez & Iglewicz, 1984)
- Implicitly defined density functions. $\Rightarrow$ Classical ML demanding

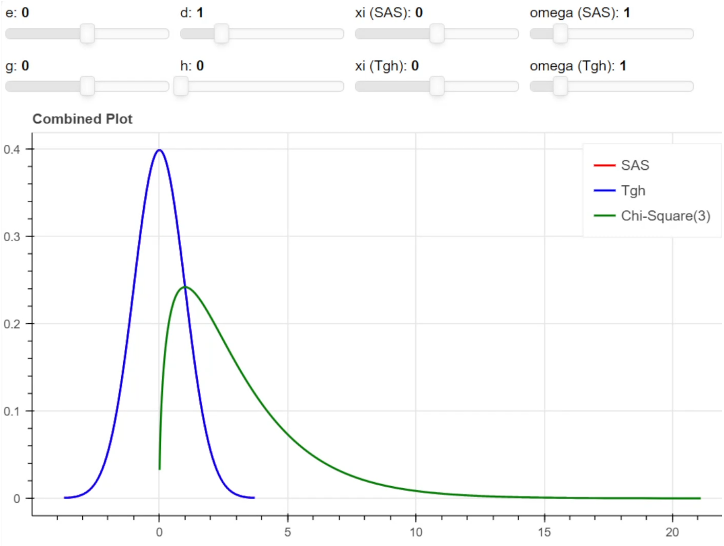
sinh-arcsinh (SAS) distribution

- Alternative transformation of a **standard normal** developed by Jones and Pewsey (2005)
- Let Z be a random variable with a standard normal distribution $N(0,1)$.

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

  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 $\sigma \in \mathbb{R}^+_0$.
- $X \sim \text{SAS}_{\varepsilon,\delta}(\xi,\sigma)$.
- $\varepsilon \in \mathbb{R}$ **controls** the **skewness**
- $\delta \in \mathbb{R}^+_0$ **controls** the **heaviness of the tails** (tail weight decreases when δ increases).

Approximating a distribution



Semiparametric Solution

- Stay agnostic on the distribution of the error term and rely on function of the residuals, which only uses the information invariant wrt. the underlying distribution.
- In our framework, the **best function** (maximum invariant) that can be used is **based on ranks and signs** (**Hallin and Werker, 2003**)
- The idea is to condition the parametric central sequence $\boldsymbol{\Delta}_f^{(n)}(\theta)$ on the maximum invariant 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 $\mathscr{F}_0$.
- We finally use a consistent estimator of f to obtain fully semiparametric efficient estimator (i.e. efficient not only under f)

Codes in Stata

- These estimators are coded in **Stata** (as well as in Matlab and R). They will be made available (on SSC) only after the underlying theoretical papers have been 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 **weighting matrix** that defines a **connectivity lag** of the dependent variable (like in spregress).
  Data have to be spset.

Stata commands

Syntax

sarnp *depvar* *indepvars* [*if*] [*in*] **,** dvarlag(*spatname*) [*options*]

dvarlag(*spatname*): the user declares a **sarnp** weighting matrix that defines a connectivity lag of the dependent variable

Dependencies

The package **moremata** by Ben Jann is used by sarnp. Please type (or follow the link) **ssc install moremata** to install it.

The command **fvstrip** by Mark E. Schaffer is used by sarnp. Please type (or follow the link) **ssc install fvstrip** to install it.

Options

options	Description
Model	
init(*string*)	choose between "2sls" and "ml" as preliminary estimator.
impower(#)	order of instrumental-variable approximation. Set to 2 by default. If ml is selected, this option has no effect.
nolog	do not show iteration logs
maxit(#)	set the maximum number of iterations. 100 by default.
tol	set the tolerance for the logL convergence. By default it is set to 1e-4
preliminary	show the preliminary estimator
noomit	keeps b/n/o factor variable operators
Reporting	
level(*cilevel*)	set the confidence level

Stata commands

Title

sarflex Flexible parametric SAR regression estimator

Syntax

sarflex *depvar* *indepvars* [*if*] [*in*] **,** dvarlag(*spatname*) [*options*]

dvarlag(*spatname*): the user declares a weighting matrix that defines a connectivity lag of the dependent variable

Dependencies

The package **moremata** by Ben Jann is used by sarflex. Please type (or follow the link) **ssc install moremata** to install it.

The command **fvstrip** by Mark E. Schaffer is used by sarflex. Please type (or follow the link) **ssc install fvstrip** to install it.

Options

options	Description
Model	
est(*string*)	choose between "Tgh" (default), "SAS" and "L1" estimators. "L1" is only meaningful if the error term is Laplace distributed.
init(*string*)	choose between "2sls" and "ml" as initial estimator.
nquantiles(#)	defines the number of quantiles needed for the preliminary estimation of distribution parameters
impower(#)	order of instrumental-variable approximation. Set to 2 by default. If ml is selected, this option has no effect.
nolog	do not show iteration logs
maxit(#)	set the maximum number of iterations. 100 by default.
tol	set the tolerance for the logL convergence. By default it is set to 1e-4
preliminary	show the preliminary estimator

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) (default).

`direct` generates the **direct effects**.

`indirect` generates the **indirect effects**.

`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 obs 500
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

spregress y x*, gs2sls dvarlag(W) impower(2)
est store ML

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

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

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

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

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)

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

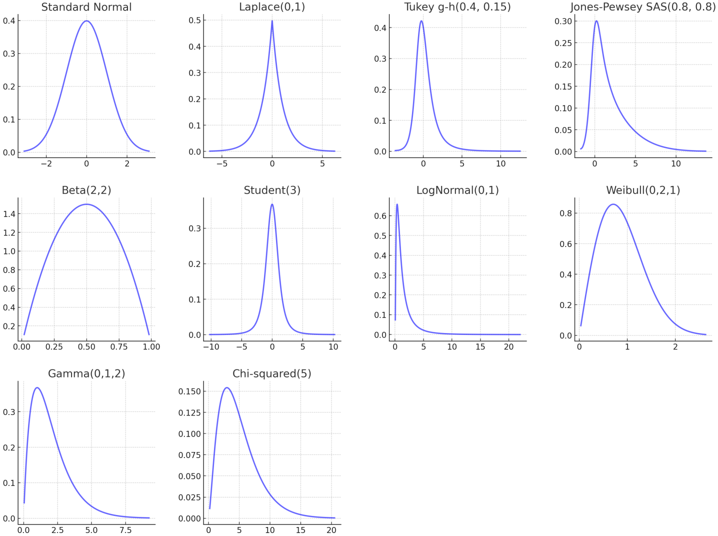
Simulations

Simulations setup

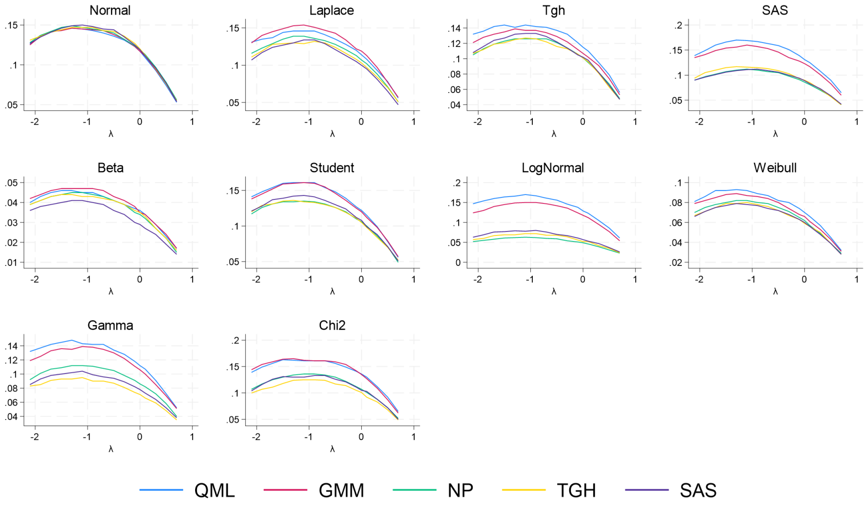
$$y_i = \lambda W_{i.} y + 1 + x_i + \varepsilon_i$$
$$X \sim N(0,1), n = 300; 500; 900$$

- 1000 replications
- 10 Nearest Neighbors interaction matrix
- **Preliminary estimator**: 2SLS $\left(\widetilde{\theta}^{(n)}\right)$
- Measured Bias: $med(\theta_r - \theta_{true}),\ r = 1, \cdots, 1000$
- Measured Variability: Interquartile range divided by 1.349

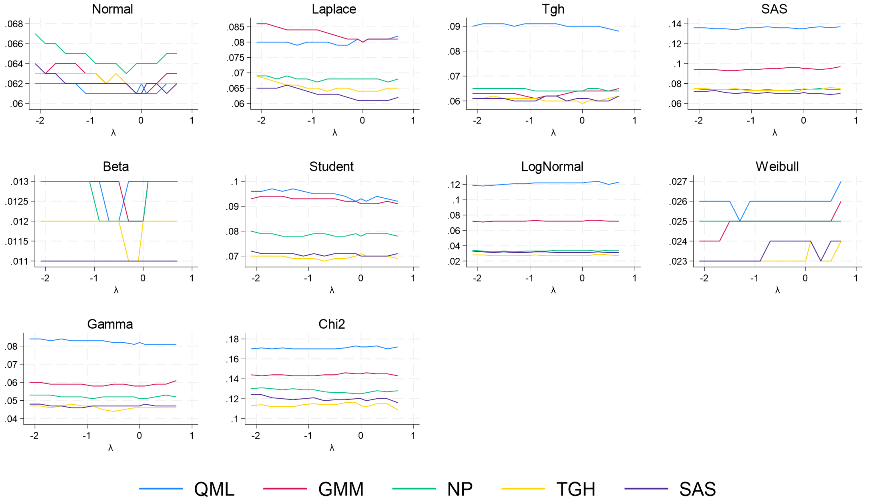
Distributions



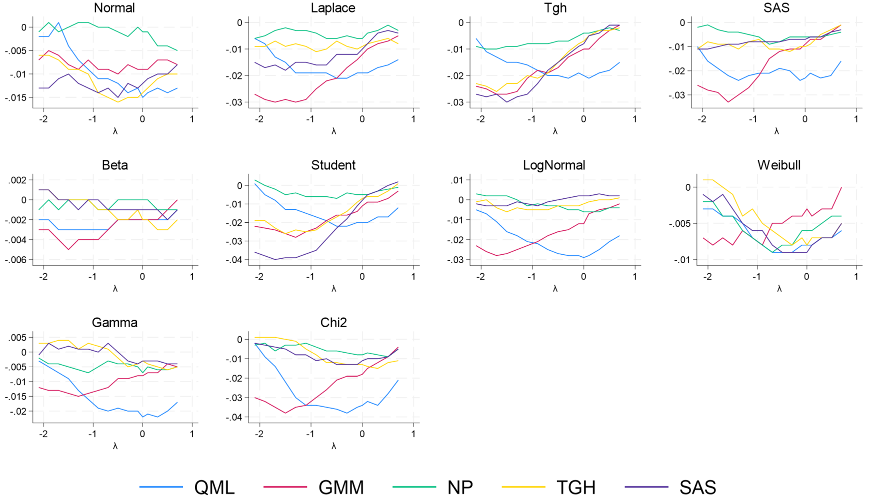
Efficiency of λ, $n = 300$



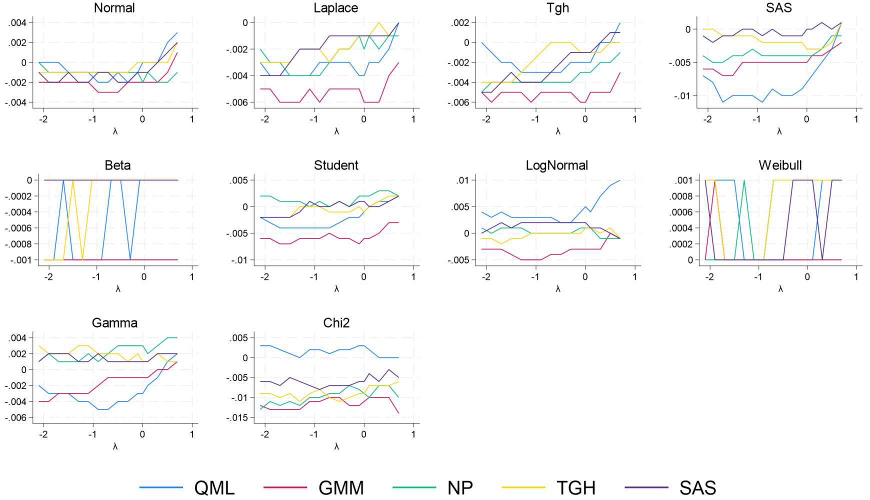
Efficiency of β_1, $n = 300$



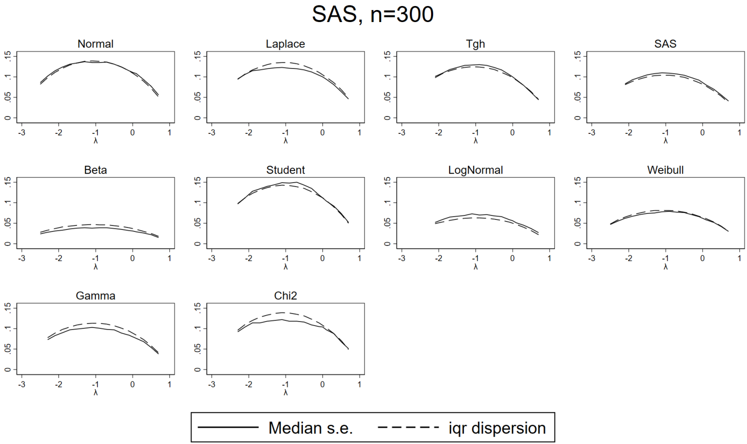
Bias of λ, $n = 300$



Bias of β_1, $n = 300$



Comparison between IQR and median S.E. for λ, n = 300, SAS distribution



Comparison between IQR and median S.E. for λ, n = 300, SAS distribution

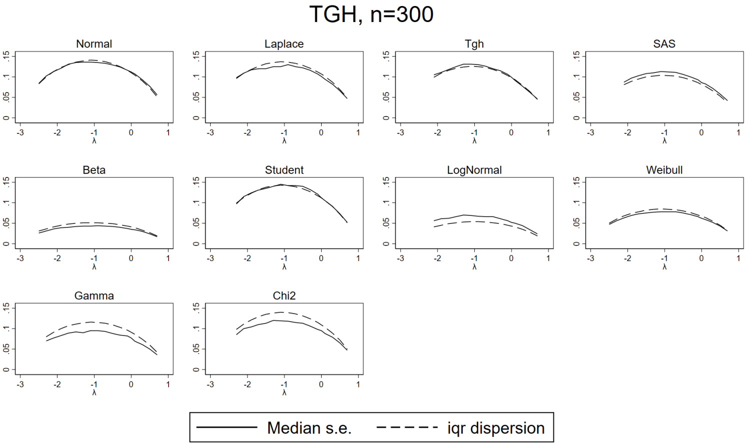


Illustration on a trade equation

- Behrens, Ertur and Koch (2012) (BEK) 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
- Estimation of the following model:

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

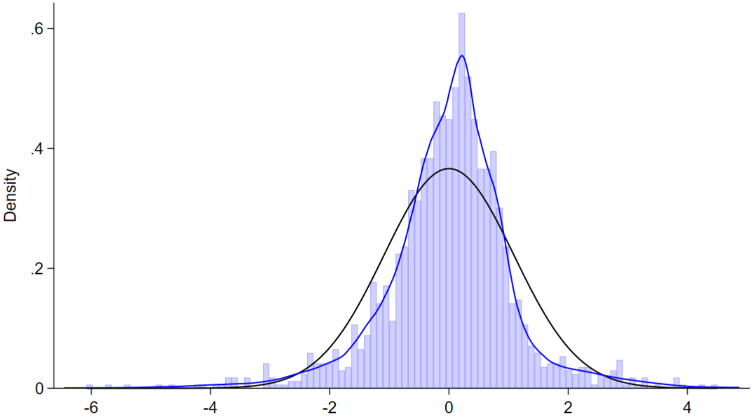
  - Z_{ij}: Manufacturing exports from i to j, standardized by GDP
  - d_{ij}: Distance between i and j (includes a measure of internal distance)
  - w_i: Average hourly manufacturing wage in region i
  - b_{ij}: 1 if i is CAN and j is US and vice-versa
  - L_i: Population in region i.
- Objective: Compare the results obtained under normality with those relying on more flexible distributions

Results comparison

	QML	TSLS	GMM	TGH	SAS	R&S
Constant	-13.869	-14.378	-12.274	-12.752	-12.700	-12.230
	(0.709)	(0.683)	(0.691)	(0.523)	(0.510)	(0.472)
	[-19.496]	[-21.049]	[-17.767]	[-24.376]	[-24.915]	[-25.898]
d_{ij}	-1.255	-1.219	-1.280	-1.198	-1.206	-1.208
	(0.035)	(0.034)	(0.033)	(0.026)	(0.025)	(0.024)
	[-35.984]	[-35.708]	[-39.278]	[-46.164]	[-48.073]	[-50.753]
$\ln(w_i)$	-1.163	-1.149	-1.759	-1.187	-1.196	-1.264
	(0.176)	(0.177)	(0.170)	(0.135)	(0.130)	(0.124)
	[-6.631]	[-6.482]	[-10.370]	[-8.795]	[-9.174]	[-10.220]
b_{ij}	-1.046	-1.056	-0.804	-1.191	-1.188	-1.199
	(0.066)	(0.066)	(0.063)	(0.050)	(0.049)	(0.046)
	[-15.961]	[-16.044]	[-12.726]	[-23.641]	[-24.433]	[-25.972]
λ	0.033	0.005	0.045	0.091	0.090	0.109
	(0.029)	(0.027)	(0.029)	(0.021)	(0.021)	(0.019)
	[1.012]	[0.186]	[1.577]	[4.241]	[4.285]	[5.682]

Notes: standard errors between parentheses and t-stats between square brackets.

Distribution of the residuals



Conclusion

- Maximum likelihood estimation is efficient when the error distribution is known, but becomes infeasible with unknown distributions; quasi-maximum likelihood remains consistent (under normality) but not efficient.
- We propose two estimators based on Local Asymptotic Normality, achieving efficiency.
- These estimators can also be used for classical linear models (simplification required)
- 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, 773–794.

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 distribution, *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.

Appendix

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^2. Under P_θ,

$$
\begin{aligned}
\log\left[\frac{dP^{(n)}_{\theta_0+\tau^{(n)}/\sqrt{n}}}{dP^{(n)}_{\theta_0}}\right] &= -\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(X_i^{(n)}-\theta_0-\frac{\tau^{(n)}}{\sqrt{n}}\right)^2+\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(X_i^{(n)}-\theta_0\right)^2 \\
&= \frac{1}{\sigma^2}\sum_{i=1}^{n}\left(X_i^{(n)}-\theta_0\right)\frac{\tau^{(n)}}{\sqrt{n}}-\frac{{\tau^{(n)}}^2}{2\sigma^2} \\
&= \tau^{(n)}n^{1/2}\left(\bar{X}^{(n)}-\theta_0\right)/\sigma^2-\frac{{\tau^{(n)}}^2}{2}1/\sigma^2 \\
&= \tau^{(n)}\Delta^{(n)}-\frac{1}{2}{\tau^{(n)}}^2 I_{\theta_0}
\end{aligned}
$$

where $\Delta^{(n)} = n^{1/2}\left(\bar{X}^{(n)}-\theta_0\right)/\sigma^2 \sim N\left(0, I_{\theta_0}\right)$ and $I_{\theta_0} = 1/\sigma^2$

$$
\Rightarrow \log\left[\frac{dP^{(n)}_{\theta+\tau^{(n)}/\sqrt{n}}}{dP^{(n)}_{\theta}}\right] \sim N\left(-\frac{1}{2}\frac{\tau^2}{\sigma^2}, \frac{\tau^2}{\sigma^2}\right) \text{ under } P_\theta
$$

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_0 if there exist sequences of random vectors $\mathbf{\Delta}^{(n)}(\theta)$ and symmetric positive semi-definite matrices $\mathbf{I}(\theta_0)$ such that the log-likelihood ratio satisfies the following approximation:

$$\log \frac{dP^{(n)}_{\theta_0+\tau/\sqrt{n}}}{dP^{(n)}_{\theta_0}} = (\tau^{(n)})^{\mathrm{T}} \mathbf{\Delta}^{(n)}(\theta_{\mathbf{0}}) - \frac{1}{2}(\tau^{(n)})^{\mathrm{T}} \mathbf{I}(\theta_{\mathbf{0}}) \tau^{(n)} + o_{\mathrm{P}}(1)$$

where (i) τ is a fixed vector representing the local perturbation of the parameter $\theta_{\mathbf{0}}$, and (ii)$\mathbf{\Delta^{(n)}(\theta_0)}$ converges in distribution to a normal random vector $\mathbf{\Delta^{(n)}(\theta_0)} \sim N(0, \mathbf{I}(\theta_{\mathbf{0}}))$

Practical Implementation I

- The LAN property allows us to obtain an efficient estimator by **refining** a $\sqrt{n}-$**consistent** preliminary estimator.

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

1. For the regression coefficients, we consider a SAR model estimated by 2SLS that gives a preliminary residuals

$$e_i^{(n)} \left(\widetilde{\beta}^{(n)}, \tilde{\lambda}^{(n)} \right) = y_i^{(n)} - \left(\mathbf{x}_i^{(n)} \right)^{\mathrm{T}} \widetilde{\beta}^{(n)} - \widetilde{\lambda}^{(n)} \mathbf{W}_{i.}^{(n)} \mathbf{y}^{(n)}$$

Attention, the constant term is not directly estimated, but appear as the location parameter of the distribution.

Practical Implementation II

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

   Minimizing the (squared) distance between some **residual quantiles** and the **corresponding theoretical quantiles** of the considered distribution:

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

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

3. Evaluate the central sequence and the Information matrix at the preliminary estimates
4. Compute the one-step efficient estimator

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 :

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

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.

Tukey g-and-h error term distribution

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

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

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

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

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

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

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

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

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

SAS error term distribution

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

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

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

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

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

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

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

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

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