Statistical analysis of network data: learning from small samples

Olivier Scaillet

[1]University of Geneva,
[2]Swiss Finance Institute.

September 18th, 2020

Three-fold aim of the talk

For small samples network (spatial panels) data we:

Three-fold aim of the talk

For small samples network (spatial panels) data we:

- illustrate that first-order asymptotic theory suffers from finite sample distortions.

Three-fold aim of the talk

For small samples network (spatial panels) data we:

- illustrate that first-order asymptotic theory suffers from finite sample distortions.

- develop saddlepoint techniques which perform well in small samples and feature higher-order accuracy.

Three-fold aim of the talk

For small samples network (spatial panels) data we:

- illustrate that first-order asymptotic theory suffers from finite sample distortions.
- develop saddlepoint techniques which perform well in small samples and feature higher-order accuracy.
- compare the saddlepoint density approximation to the Edgeworth expansion and asymptotic approximation.

Three-fold aim of the talk

For small samples network (spatial panels) data we:

- illustrate that first-order asymptotic theory suffers from finite sample distortions.

- develop saddlepoint techniques which perform well in small samples and feature higher-order accuracy.

- compare the saddlepoint density approximation to the Edgeworth expansion and asymptotic approximation.

This talk is based on the working paper *Saddlepoint approximations for spatial panel data models*, joint work with

Three-fold aim of the talk

For small samples network (spatial panels) data we:

- illustrate that first-order asymptotic theory suffers from finite sample distortions.

- develop saddlepoint techniques which perform well in small samples and feature higher-order accuracy.

Jiang (RCS)

La Vecchia (RCS)

Ronchetti (RCS)

The problem in spatial econometrics terms

Feldstein and Horioka (1980) found a large, positive impact of domestic saving rates on domestic investment rates for Organisation for Economic Co-operation and Development (OECD) countries.

The problem in spatial econometrics terms

Feldstein and Horioka (1980) found a large, positive impact of domestic saving rates on domestic investment rates for Organisation for Economic Co-operation and Development (OECD) countries.
Investment and Saving rates for 24 OECD countries between 1960 and 2000 (41 yrs) are considered.

The problem in spatial econometrics terms

Feldstein and Horioka (1980) found a large, positive impact of domestic saving rates on domestic investment rates for Organisation for Economic Co-operation and Development (OECD) countries.
Investment and Saving rates for 24 OECD countries between 1960 and 2000 (41 yrs) are considered. Thus,

$$\{\mathrm{Inv}_{it}\} \qquad \text{and} \qquad \{\mathrm{Sav}_{it}\}$$

for $i = 1, .., n$ (cross-sectional dimension) and $t = 1, ... T$ (time series dimension).

The problem in spatial econometrics terms

Feldstein and Horioka (1980) found a large, positive impact of domestic saving rates on domestic investment rates for Organisation for Economic Co-operation and Development (OECD) countries.
Investment and Saving rates for 24 OECD countries between 1960 and 2000 (41 yrs) are considered. Thus,

$$\{\mathrm{Inv}_{it}\} \qquad \text{and} \qquad \{\mathrm{Sav}_{it}\}$$

for $i = 1, .., n$ (cross-sectional dimension) and $t = 1, \dots T$ (time series dimension).

Aim

Test for the presence of spatial autocorrelation (between country i and country j, $i \neq j$) in the investment-saving relationship, e.g. using p-value, as in the finance/economics literature.

The problem in spatial econometrics terms

Feldstein and Horioka (1980) found a large, positive impact of domestic saving rates on domestic investment rates for Organisation for Economic Co-operation and Development (OECD) countries.
Investment and Saving rates for 24 OECD countries between 1960 and 2000 (41 yrs) are considered. Thus,

$$\boxed{\{\text{Inv}_{it}\} \qquad \text{and} \qquad \{\text{Sav}_{it}\}}$$

for $i = 1, .., n$ (cross-sectional dimension) and $t = 1, ... T$ (time series dimension).

Aim

Test for the presence of spatial autocorrelation (between country i and country j, $i \neq j$) in the investment-saving relationship, e.g. using p-value, as in the finance/economics literature.

The problem in spatial econometrics terms

Feldstein and Horioka (1980) found a large, positive impact of domestic saving rates on domestic investment rates for Organisation for Economic Co-operation and Development (OECD) countries.
Investment and Saving rates for 24 OECD countries between 1960 and 2000 (41 yrs) are considered. Thus,

$$\boxed{\{\mathsf{Inv}_{it}\} \qquad \stackrel{\forall i}{\Longleftrightarrow} \qquad \{\mathsf{Sav}_{it}\}}$$

for $i = 1, .., n$ (cross-sectional dimension) and $t = 1, ... T$ (time series dimension).

Aim

Test for the presence of spatial autocorrelation (between country i and country j, $i \neq j$) in the investment-saving relationship, e.g. using p-value, as in the finance/economics literature.

The problem in spatial econometrics terms

Feldstein and Horioka (1980) found a large, positive impact of domestic saving rates on domestic investment rates for Organisation for Economic Co-operation and Development (OECD) countries.
Investment and Saving rates for 24 OECD countries between 1960 and 2000 (41 yrs) are considered. Thus,

$$\boxed{\{\text{Inv}_{it}\} \overset{?}{\Longleftrightarrow} \{\text{Inv}_{kt}\} \overset{?}{\Longleftrightarrow} \{\text{Inv}_{jt}\} \overset{?}{\Longleftrightarrow} \{\text{Inv}_{it}\} \Longleftrightarrow \{\text{Sav}_{it}\}}$$

for $i, j, k \in \mathbb{N}$ and $t = 1, \dots T$.

Aim (rephrased)

Test if a change in the saving rate in one country, i say, affects the investment rate of that country, which in turn affects the investment rates of other countries (j, k say), which then feed back to the investment rate of country i.

The problem in spatial econometrics terms

In the econometric literature, the problem is tackled using spatial panel data models; see e.g. Lee (2004), Lee and Yu (2010).

The problem in spatial econometrics terms

In the econometric literature, the problem is tackled using spatial panel data models; see e.g. Lee (2004), Lee and Yu (2010).

We consider a Gaussian random field described by the SARAR(1,1) model

$$\begin{aligned} \mathrm{Inv}_{nt} &= \lambda_0 W_n \,\mathrm{Inv}_{nt} + \beta_0 \mathrm{Sav}_{nt} + c_{n0} + E_{nt}, \\ E_{nt} &= \rho_0 M_n E_{nt} + V_{nt}, \qquad\qquad t = 1, 2, \cdots, T \end{aligned}$$

The problem in spatial econometrics terms

In the econometric literature, the problem is tackled using spatial panel data models; see e.g. Lee (2004), Lee and Yu (2010).

We consider a Gaussian random field described by the SARAR(1,1) model

$$\begin{aligned} \mathrm{Inv}_{nt} &= \lambda_0 W_n \, \mathrm{Inv}_{nt} + \beta_0 \mathrm{Sav}_{nt} + c_{n0} + E_{nt}, \\ E_{nt} &= \rho_0 M_n E_{nt} + V_{nt}, \qquad\qquad t = 1, 2, \cdots, T \end{aligned} \tag{1}$$

- Inv_{nt} is the $n \times 1$ vector of investment rates for all countries
- Sav_{nt} is the $n \times 1$ vector of saving rates
- V_{nt} is an $n \times 1$ vector and each element v_{it} in it is i.i.d across i and t, having Gaussian distribution with zero mean and variance σ_0^2.
- c_{n0} is an $n \times 1$ vector of fixed effects
- we assume $W_n = M_n$
- while the scalars λ_0 and ρ_0 control the spatial dependency.

The problem in spatial econometrics terms

Remark

The matrix W_n (M_n) is an $n \times n$ nonstochastic spatial weight matrix (also called contiguity/connectivity matrix) that generates the spatial dependence among cross sectional units.

The problem in spatial econometrics terms

Remark

The matrix W_n (M_n) is an $n \times n$ nonstochastic spatial weight matrix (also called contiguity/connectivity matrix) that generates the spatial dependence among cross sectional units. For instance, setting $n = 4$ we may define

$$W_n = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

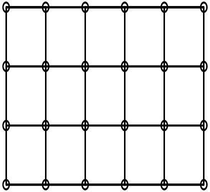
or row normalized (sum by row is equal to one)

$$W_n = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

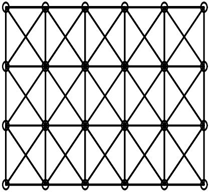
The problem in spatial econometrics terms

Example [Lattice geometry]: W_n, for $n = 24$:

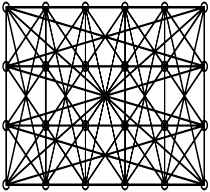
Rook



Queen



Queen torus



The problem in spatial econometrics terms

The goal is to test for spatial dependency in the investments, thus:

$$\underbrace{\mathcal{H}_0 : \ \lambda_0 = 0}_{\text{no spacial dep.}} \quad \text{vs} \quad \underbrace{\mathcal{H}_1 : \ \lambda_0 > 0}_{\text{pos. spacial dep.}} .$$

In the econometric literature, standard asymptotic theory is applied and a (Gaussian) p-value is computed.

The problem in spatial econometrics terms

The goal is to test for spatial dependency in the investments, thus:

$$\underbrace{\mathcal{H}_0 : \ \lambda_0 = 0}_{\text{no spacial dep.}} \quad \text{vs} \quad \underbrace{\mathcal{H}_1 : \ \lambda_0 > 0}_{\text{pos. spacial dep.}} .$$

In the econometric literature, standard asymptotic theory is applied and a (Gaussian) p-value is computed.

Issue

Because of macroeconomic reasons (deregulating financial markets), the period can be divided into shorter sub-periods; see Debarsy & Ertur (2010).

The problem in spatial econometrics terms

The goal is to test for spatial dependency in the investments, thus:

$$\underbrace{\mathcal{H}_0 : \ \lambda_0 = 0}_{\text{no spacial dep.}} \quad \text{vs} \quad \underbrace{\mathcal{H}_1 : \ \lambda_0 > 0}_{\text{pos. spacial dep.}}.$$

In the econometric literature, standard asymptotic theory is applied and a (Gaussian) p-value is computed.

Issue

Because of macroeconomic reasons (deregulating financial markets), the period can be divided into shorter sub-periods; see Debarsy & Ertur (2010). Is the use of asymptotics legitimate? Is the asymptotics suffering from size distortion due to the small samples, in both the cross section and in time?

The problem in spatial econometrics terms

The goal is to test for spatial dependency in the investments, thus:

$$\underbrace{\mathcal{H}_0 : \ \lambda_0 = 0}_{\text{no spacial dep.}} \quad \text{vs} \quad \underbrace{\mathcal{H}_1 : \ \lambda_0 > 0}_{\text{pos. spacial dep.}}.$$

In the econometric literature, standard asymptotic theory is applied and a (Gaussian) p-value is computed.

Issue

Because of macroeconomic reasons (deregulating financial markets), the period can be divided into shorter sub-periods; see Debarsy & Ertur (2010). Is the use of asymptotics legitimate? Is the asymptotics suffering from size distortion due to the small samples, in both the cross section and in time? If this is the case, can we rely on small sample techniques, like the saddlepoint approximations?

The problem in spatial econometrics terms

The goal is to test for spatial dependency in the investments, thus:

$$\underbrace{\mathcal{H}_0 : \ \lambda_0 = 0}_{\text{no spacial dep.}} \quad \text{vs} \quad \underbrace{\mathcal{H}_1 : \ \lambda_0 > 0}_{\text{pos. spacial dep.}}.$$

In the econometric literature, standard asymptotic theory is applied and a (Gaussian) p-value is computed.

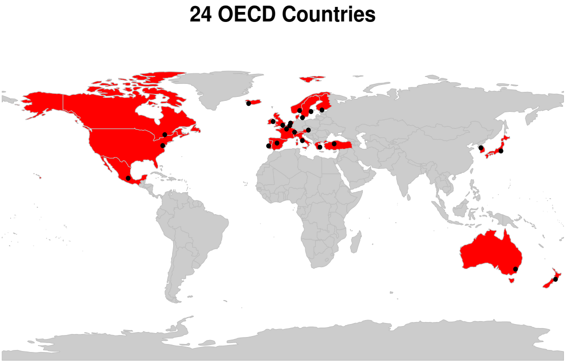
Issue

Because of macroeconomic reasons (deregulating financial markets), the period can be divided into shorter sub-periods; see Debarsy & Ertur (2010). Is the use of asymptotics legitimate? Is the asymptotics suffering from size distortion due to the small samples, in both the cross section and in time? If this is the case, can we rely on small sample techniques, like the saddlepoint approximations?

$\Rightarrow$ The issue is that saddlepoint approximations are available only for independent observations...We need to develop a new theory to tackle our problem for spatial panel data...

The problem in network analysis terms

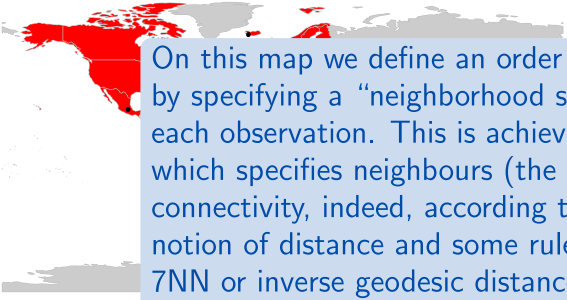
The problem can be reformulated in terms of network analysis:



The problem in network analysis terms

The problem can be reformulated in terms of network analysis:

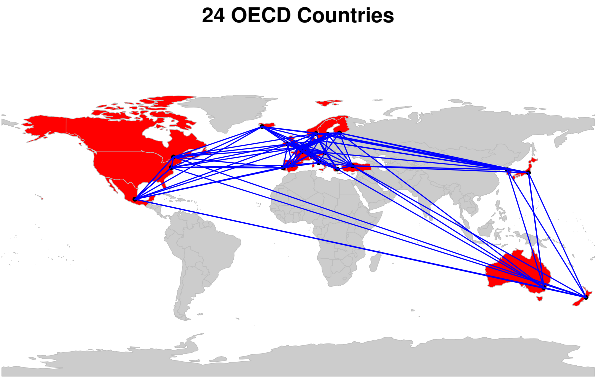
24 OECD Countries



On this map we define an order structure by specifying a "neighborhood set" for each observation. This is achieved by W_n, which specifies neighbours (the connectivity, indeed, according to some notion of distance and some rule, e.g. 7NN or inverse geodesic distance) for each capital city.

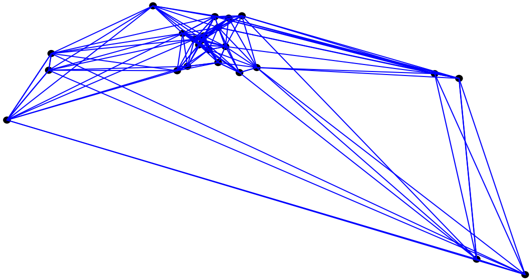
The problem in network analysis terms

The problem can be reformulated in terms of network analysis:



The problem in network analysis terms

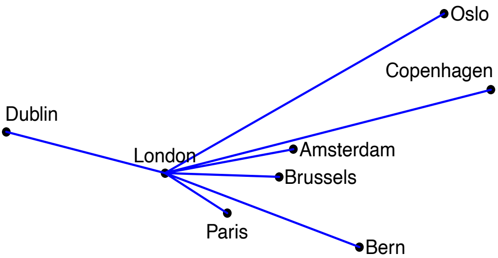
For each t, an undirected graph $\mathcal{G}$ describes the network:



and $\{\mathrm{Inv}_{it}\}$ is recorded at each location (arbitrarily numbered)

The problem in network analysis terms

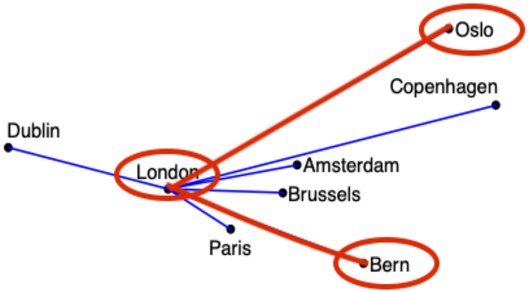
For each t, we may focus on one vertex (London, say) with the 7NN criterion



so...

The problem in network analysis terms

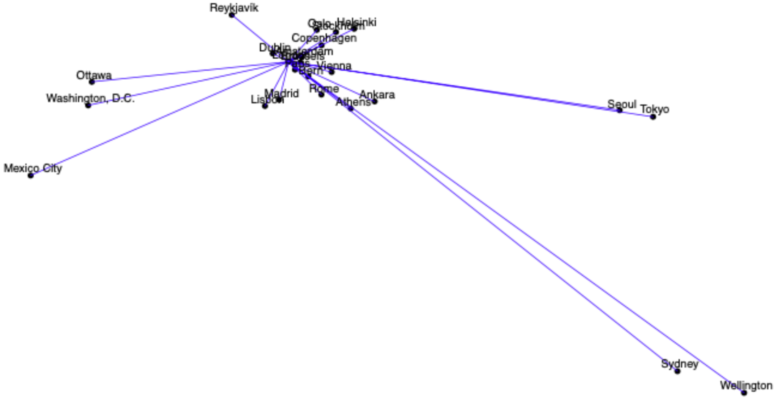
For each t, we may focus on one vertex (London, say) with the 7NN criterion



so…there is a *walk (or a path)* from Oslo to Bern via London

The problem in network analysis terms

For each t, we may focus on one vertex (London) with the Inverse Geodesic Distance criterion to specify W_n



The problem in network analysis terms

Some comments are in order:

- W_n is a weighted adjacency matrix, as in Graph Signal Processes literature; see e.g. Shuman, Vandergheynst et al. (2013) for an introduction with application to learning problems

The problem in network analysis terms

Some comments are in order:

- W_n is a weighted adjacency matrix, as in Graph Signal Processes literature; see e.g. Shuman, Vandergheynst et al. (2013) for an introduction with application to learning problems
- $\mathcal{G} = (V, E)$, where V is the set of vertices (capital cities) and E is the set of edges (connections between capital cities). We have $|V| = n$ and W_n is such that if there is an edge $e = (i, j)$ connecting vertices i and j, the entry w_{ij} represents the weight of the edge; otherwise, $w_{ij} = 0$. We assume no self-edges, thus $w_{ii} = 0$.

The problem in network analysis terms

Some comments are in order:

- W_n is a weighted adjacency matrix, as in Graph Signal Processes literature; see e.g. Shuman, Vandergheynst et al. (2013) for an introduction with application to learning problems
- $\mathcal{G} = (V, E)$, where V is the set of vertices (capital cities) and E is the set of edges (connections between capital cities). We have $|V| = n$ and W_n is such that if there is an edge $e = (i, j)$ connecting vertices i and j, the entry w_{ij} represents the weight of the edge; otherwise, $w_{ij} = 0$. We assume no self-edges, thus $w_{ii} = 0$.
- the graph topology can be related to a lattice in an Euclidean space (e.g. $\mathbb{Z}^2$, with latitude and longitude) or to a general manifold (as in the case of the earth, which is a "sphere")

The problem in network analysis terms

Some comments are in order:

- W_n is a weighted adjacency matrix, as in Graph Signal Processes literature; see e.g. Shuman, Vandergheynst et al. (2013) for an introduction with application to learning problems
- $\mathcal{G} = (V, E)$, where V is the set of vertices (capital cities) and E is the set of edges (connections between capital cities). We have $|V| = n$ and W_n is such that if there is an edge $e = (i, j)$ connecting vertices i and j, the entry w_{ij} represents the weight of the edge; otherwise, $w_{ij} = 0$. We assume no self-edges, thus $w_{ii} = 0$.
- the graph topology can be related to a lattice in an Euclidean space (e.g. $\mathbb{Z}^2$, with latitude and longitude) or to a general manifold (as in the case of the earth, which is a "sphere")
- we model a Markovian (in the space) Gaussian random field $\{\text{Inv}_{it}\}$ taking values on a graph, randomness is induced by a Gaussian random field for the disturbances

The problem in network analysis terms

Some comments are in order:

- W_n is a weighted adjacency matrix, as in Graph Signal Processes literature; see e.g. Shuman, Vandergheynst et al. (2013) for an introduction with application to learning problems
- $\mathcal{G} = (V, E)$, where V is the set of vertices (capital cities) and E is the set of edges (connections between capital cities). We have $|V| = n$ and W_n is such that if there is an edge $e = (i, j)$ connecting vertices i and j, the entry w_{ij} represents the weight of the edge; otherwise, $w_{ij} = 0$. We assume no self-edges, thus $w_{ii} = 0$.
- the graph topology can be related to a lattice in an Euclidean space (e.g. $\mathbb{Z}^2$, with latitude and longitude) or to a general manifold (as in the case of the earth, which is a "sphere")
- we model a Markovian (in the space) Gaussian random field $\{\text{Inv}_{it}\}$ taking values on a graph, randomness is induced by a Gaussian random field for the disturbances
- we add some vertex specific covariates (saving rate of i-th county)

The problem in network analysis terms

the model for $\{\mathrm{Inv}_{it}\}$ is a supervised learning model, where the LHS variable is expressed as a (spatial) weighted sum of its values in nearby vertices (similar to Graph Neural Network), which are the RHS variables

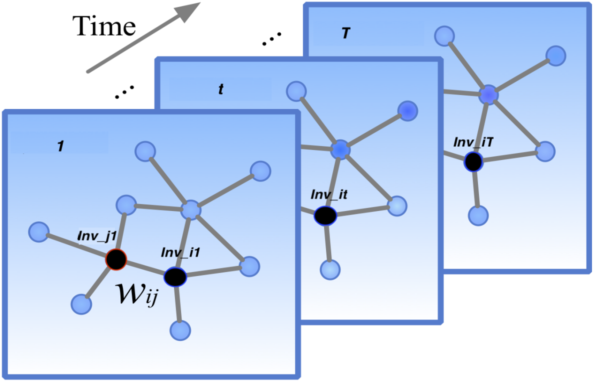
The problem in network analysis terms

- the model for $\{\mathrm{Inv}_{it}\}$ is a supervised learning model, where the LHS variable is expressed as a (spatial) weighted sum of its values in nearby vertices (similar to Graph Neural Network), which are the RHS variables
- we have repeated observations (one for each time point t) of the random field on the graph.

The problem in network analysis terms

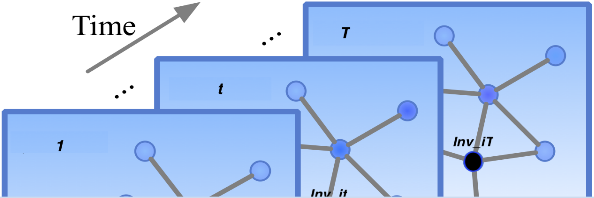
the model for $\{\mathrm{Inv}_{it}\}$ is a supervised learning model, where the LHS variable is expressed as a (spatial) weighted sum of its values in nearby vertices (similar to Graph Neural Network), which are the RHS variables

we have repeated observations (one for each time point t) of the random field on the graph.



The problem in network analysis terms

- the model for $\{\text{Inv}_{it}\}$ is a supervised learning model, where the LHS variable is expressed as a (spatial) weighted sum of its values in nearby vertices (similar to Graph Neural Network), which are the RHS variables
- we have repeated observations (one for each time point t) of the random field on the graph.



⇒ **Although the whole paper is written in terms of spatial econometrics, we may re-interpret the results in terms of statistical learning for network data, with focus on small sample problems.**

Outline of the talk

Inference problem formulation

Outline of the talk

Inference problem formulation

Spatial econometrics

Outline of the talk

Inference problem formulation

Spatial econometrics

Conclusion: take home message

Testing in small samples

General problem: For a given statistics $\hat{\theta}_{n,T}$, tail probabilities

$$P[\hat{\theta}_{n,T} > x]$$

are needed to carry out statistical inference (essentially, tests and confidence intervals).

Unless the (test) statistic $\hat{\theta}_{n,T}$ has a simple form (e.g. linear) and/or the underlying distribution of data has a particular form (e.g. normal), tail probabilities (more generally the whole distribution) cannot be computed exactly.

Testing in small samples

General problem: For a given statistics $\hat{\theta}_{n,T}$, tail probabilities

$$P[\hat{\theta}_{n,T} > x]$$

are needed to carry out statistical inference (essentially, tests and confidence intervals).

Unless the (test) statistic $\hat{\theta}_{n,T}$ has a simple form (e.g. linear) and/or the underlying distribution of data has a particular form (e.g. normal), tail probabilities (more generally the whole distribution) cannot be computed exactly.

$\Rightarrow$ we have to rely on asymptotic approximations

Asymptotic theory versus finite sample techniques

We can approximate tail probabilities via

> *Asymptotic theory*: use of Central Limit Theorem to get a Gaussian approximation in large samples,

Asymptotic theory versus finite sample techniques

We can approximate tail probabilities via

Asymptotic theory: use of Central Limit Theorem to get a Gaussian approximation in large samples,

Resampling techniques: use of resampling (bootstrap, subsampling) to get an approximation in small samples,

Analytical techniques: use of expansions (Edgeworth, saddlepoint) to get an approximation in small samples.

Edgeworth expansion

Construction: Taylor expansion of the characteristic function of the statistic of interest around 0, i.e., at the center of the distribution, followed by a Fourier inversion.

Edgeworth expansion

Construction: Taylor expansion of the characteristic function of the statistic of interest around 0, i.e., at the center of the distribution, followed by a Fourier inversion. This leads to expansions of the density in powers of $n^{-1/2}$, where the leading term is the normal density with absolute error of order $O(n^{-1/2})$. Typically, the expansion is computed up to order $O(n^{-1})$.

Edgeworth expansion

Construction: Taylor expansion of the characteristic function of the statistic of interest around 0, i.e., at the center of the distribution, followed by a Fourier inversion. This leads to expansions of the density in powers of $n^{-1/2}$, where the leading term is the normal density with absolute error of order $O(n^{-1/2})$. Typically, the expansion is computed up to order $O(n^{-1})$.

By construction, Edgeworth expansions provide in general a good approximation in the center of the density, **BUT**

Edgeworth expansion

Construction: Taylor expansion of the characteristic function of the statistic of interest around 0, i.e., at the center of the distribution, followed by a Fourier inversion. This leads to expansions of the density in powers of $n^{-1/2}$, where the leading term is the normal density with absolute error of order $O(n^{-1/2})$. Typically, the expansion is computed up to order $O(n^{-1})$.

By construction, Edgeworth expansions provide in general a good approximation in the center of the density, **BUT**

they can be inaccurate in the tails

Edgeworth expansion

Construction: Taylor expansion of the characteristic function of the statistic of interest around 0, i.e., at the center of the distribution, followed by a Fourier inversion. This leads to expansions of the density in powers of $n^{-1/2}$, where the leading term is the normal density with absolute error of order $O(n^{-1/2})$. Typically, the expansion is computed up to order $O(n^{-1})$.

By construction, Edgeworth expansions provide in general a good approximation in the center of the density, **BUT**

- they can be inaccurate in the tails
- they can even become negative in the tails.

Edgeworth expansion

Construction: Taylor expansion of the characteristic function of the statistic of interest around 0, i.e., at the center of the distribution, followed by a Fourier inversion. This leads to expansions of the density in powers of $n^{-1/2}$, where the leading term is the normal density with absolute error of order $O(n^{-1/2})$. Typically, the expansion is computed up to order $O(n^{-1})$.

By construction, Edgeworth expansions provide in general a good approximation in the center of the density, **BUT**

- they can be inaccurate in the tails
- they can even become negative in the tails.

Remark

To solve these problems, saddlepoint approximations (based on the method of the steepest descent) have been introduced.

Method of Steepest descent

General technique to compute asymptotic expansions of integrals

$$\int_{\mathcal{P}} e^{v \cdot w(z)} \xi(z) dz,$$

with $v \in \mathbb{R}^+$ is large, ξ and w being analytic functions of $z \in \mathbb{C}$.

Method of Steepest descent

General technique to compute asymptotic expansions of integrals

$$\int_{\mathcal{P}} e^{v \cdot w(z)} \xi(z) dz,$$

with $v \in \mathbb{R}^+$ is large, ξ and w being analytic functions of $z \in \mathbb{C}$.

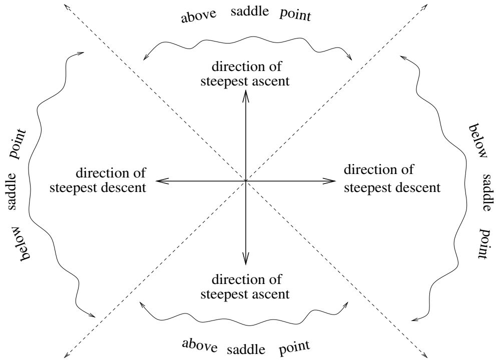
Idea

Deform the path of integration (Cauchy's theorem) so that the new path of integration passes through the so-called saddlepoint, namely the zero of the derivative $w'(z)$. Then, we approximate the resulting integral using a series expansion (Watson's lemma).

Loosely speaking, we do a "Laplace-type approx" on $\mathbb{C}$. Jump to Laplace

Method of Steepest descent (Geometry)

The function w is constant in the imaginary direction and has an extremum in the real direction: it decreases rapidly from the extreme point.



Steepest ascent and descent are plotted as solid line vectors, while directions of constant elevation are dashed line vectors; from Small (2010).

Saddlepoint density approx for the mean

As in Daniels (1954), let $X_1, \ldots, X_n$ be **iid** random variables from a distribution F with $M(\nu) = E[e^{\nu} X]$ being the moment generating function such that

$$\mathcal{K}(\nu) = \log M(\nu)$$

is the cumulant generating function.

Inferential problem

Approximate, for every n, the density f_n (say) of the

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i,$$

which is an M-estimator of location.

Saddlepoint techniques for the mean (cont.)

The method of the steepest descent yields the saddlepoint approx $p_n(\alpha)$:

$$f_n(\alpha) = p_n(\alpha)\left\{1 + O\left(n^{-1}\right)\right\},$$

where

$$p_n(\alpha) = \left[\frac{n}{2\pi \mathcal{K}^{('')}\{\nu(\alpha)\}}\right]^{1/2} \exp\left(n\left[\mathcal{K}\{\nu(\alpha)\} - \nu(\alpha)\alpha\right]\right),$$

and $\nu(\alpha)$ (saddlepoint) is the solution to $\boxed{\mathcal{K}^{(')}(\nu) - \alpha = 0}$.

Remark

The approximation p_n features relative error of order $O(n^{-1})$ over the whole $\mathbb{R}$:

$$rel.err. = \frac{p_n - f_n}{f_n} = O(n^{-1}).$$

General References (books)

Barndorff-Nielsen, O., Cox D.R. (1989), *Asymptotic Techniques for Use in Statistics*, Chapman and Hall.

Field, C., Ronchetti, E.(1990) *Small Sample Asymptotics*, Institute of Mathematical Statistics Monograph Series, Hayward (CA).

Jensen, J.L.(1995) *Saddlepoint Approximations*, Oxford University Press.

Kolassa, J.E. (1997) *Series Approximation Methods in Statistics*, Springer, New York.

Ullah, A. (2004) *Finite Sample Econometrics*, Oxford University Press.

Butler, R.W. (2007) *Saddlepoint Approximations with Applications*, Cambridge University Press.

Small, C.(2010), *Expansions and Asymptotics for Statistics*, CRC press.

Spatial Autoregressive models with covariates (SARAR)

For user-specified spatial matrices W_n and M_n, we consider

$$\begin{aligned} Y_{nt} &= \lambda_0 W_n Y_{nt} + X_{nt}\beta_0 + c_{n0} + E_{nt}, \\ E_{nt} &= \rho_0 M_n E_{nt} + V_{nt}, \quad t = 1, ..., T. \end{aligned}$$

where $Y_{nt} = (y_{1t}, y_{2t}, ..., y_{nt})$, X_{nt} is an matrix of non stochastic time-varying regressors, c_{n0} is an $n \times 1$ vector of fixed effects and $V_{nt} = (v_{1t}, v_{2t}, .., v_{nt})'$ are $n \times 1$ vectors and $v_{it} \sim \mathcal{N}(0, \sigma_0^2)$, i.i.d. across i and t (Lee and Yu (2010)).

We label this model SARAR(1,1) to emphasize the spatial dependence in both the response variable (Y_{nt}) and in the error term (E_{nt}). The model parameter is $\theta := (\lambda_0, \beta_0, \rho_0, \sigma_0^2)$ and we estimate it using the Gaussian MLE ($\hat{\theta}_{n,T}$).

Main theoretical result

We derive the saddlepoint density approximation for $q(\hat{\theta}_{n,T})$ at point α, where $q : \Theta \to \mathbb{R}$ and $\theta \in \Theta \subset \mathbb{R}^p, p \geq 1$:

$$p_{n,T}(\alpha) = \left[\frac{n}{2\pi \tilde{\mathcal{K}}^{(\prime\prime)}_{n,T}(\nu)} \right]^{1/2} \exp\left\{ n \left[\tilde{\mathcal{K}}_{n,T}(\nu) - \nu\alpha \right] \right\},$$

with relative error of order $O(m^{-1})$, for $m = n(T-1)$, and $\nu := \nu(\alpha)$ is the saddlepoint defined by

$$\boxed{\tilde{\mathcal{K}}^{(\prime)}_{n,T}(\nu) - \alpha = 0.}$$

See paper for the explicit expression of $\tilde{\mathcal{K}}_{n,T}$ (approximate c.g.f. of $q(\hat{\theta}_{n,T})$).

The construction at a glance

Our construction relies on four steps:

The construction at a glance

Our construction relies on four steps:

(i) eliminating the fixed effect without creating dependence—to this end, we adopt the transformation approach of Lee & Yu (2010) and deriving the Gaussian MLE

The construction at a glance

Our construction relies on four steps:

(i) eliminating the fixed effect without creating dependence—to this end, we adopt the transformation approach of Lee & Yu (2010) and deriving the Gaussian MLE

(ii) defining a second-order asymptotic (von Mises) expansion for the MLE of θ and using it to define an asymptotic expansion for $q(\hat{\theta}_{n,T})$

The construction at a glance

Our construct

(i) eliminati
we adopt
deriving

(ii) defining
MLE of

Gaussian MLE

The log-likelihood reads as:

$$\begin{aligned}\ell_{n,T}(\theta) &= -\frac{n(T-1)}{2}\ln(2\pi\sigma^2) + (T-1)[\ln|S_n(\lambda)| \\ &+ \ln|R_n(\rho)|] - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\tilde{V}'_{nt}(\zeta)\tilde{V}_{nt}(\zeta),\end{aligned}$$

where, for $S_n(\lambda) = I_n - \lambda W_n$ and $R_n(\rho) = I_n - \rho M_n$, the **transformed errors, which are iid**

$$\tilde{V}_{nt}(\zeta) = R_n(\rho)[S_n(\lambda)\tilde{Y}_{nt} - \tilde{X}_{nt}\beta],$$

The con

Our construct

(i) eliminati
we adopt
deriving

(ii) defining
MLE of

Gaussian MLE

The log-likelihood reads as:

$$\begin{aligned}\ell_{n,T}(\theta) &= -\frac{n(T-1)}{2}\ln(2\pi\sigma^2) + (T-1)[\ln|S_n(\lambda)| \\ &+ \ln|R_n(\rho)|] - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\tilde{V}'_{nt}(\zeta)\tilde{V}_{nt}(\zeta),\end{aligned}$$

where, for $S_n(\lambda) = I_n - \lambda W_n$ and $R_n(\rho) = I_n - \rho M_n$, the **transformed errors, which are iid**

$$\tilde{V}_{nt}(\zeta) = R_n(\rho)[S_n(\lambda)\tilde{Y}_{nt} - \tilde{X}_{nt}\beta],$$

are obtained using the transformed variables and transformed covariates $\tilde{Y}_{nt} = Y_{nt} - \sum_{t=1}^{T} Y_{nt}/T$ and $\tilde{X}_{nt} = X_{nt} - \sum_{t=1}^{T} X_{nt}/T$.

The con... at a glance

Our construct...

(i) eliminati... we adopt... deriving...

(ii) defining... MLE of ...

Gaussian MLE

The log-likelihood reads as:

$$\begin{aligned}\ell_{n,T}(\theta) &= -\frac{n(T-1)}{2}\ln(2\pi\sigma^2) + (T-1)[\ln|S_n(\lambda)| \\ &+ \ln|R_n(\rho)|] - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\tilde{V}'_{nt}(\zeta)\tilde{V}_{nt}(\zeta),\end{aligned}$$

where, for $S_n(\lambda) = I_n - \lambda W_n$ and $R_n(\rho) = I_n - \rho M_n$, the **transformed errors, which are iid**

$$\tilde{V}_{nt}(\zeta) = R_n(\rho)[S_n(\lambda)\tilde{Y}_{nt} - \tilde{X}_{nt}\beta],$$

are obtained using the transformed variables and transformed covariates $\tilde{Y}_{nt} = Y_{nt} - \sum_{t=1}^{T} Y_{nt}/T$ and $\tilde{X}_{nt} = X_{nt} - \sum_{t=1}^{T} X_{nt}/T$.

The MLE $\hat{\theta}_{n,T}$ for θ is obtained solving $\hat{\theta}_{n,T} = \arg\max_{\theta\in\Theta}\ell_{n,T}(\theta)$. This is an **M-estimator** related to the log-likelihood score function, which can be written as a quadratic form in $\tilde{V}_{nt}(\zeta)$. Under suitable assumptions (see next slides), we have, for $\theta_0 \in \text{int}\Theta$

$$\sqrt{n(T-1)}(\hat{\theta}_{n,T} - \theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0,\Sigma).$$

The construction at a glance

Our construction relies on four steps:

(i) eliminating the fixed effect without creating dependence—to this end, we adopt the transformation approach of Lee & Yu (2010) and deriving the Gaussian MLE

(ii) defining a second-order asymptotic (von Mises) expansion for the MLE of θ and using it to define an asymptotic expansion for $q(\hat{\theta}_n)$

The construction at a glance

Our construction relies on four steps:

(i) eliminating the fixed effect without creating dependence—to this end, we adopt the transformation approach of Lee & Yu (2010) and deriving the Gaussian MLE

(ii) defining a second-order asymptotic (von Mises) expansion for the MLE of θ and using it to define an asymptotic expansion for $q(\hat{\theta}_n)$

(iii) identifying the corresponding U-statistic, deriving its $\boxed{\tilde{\mathcal{K}}_{n,T}}$

The construction at a glance

Our construction relies on four steps:

(i) eliminating the fixed effect without creating dependence—to this end, we adopt the transformation approach of Lee & Yu (2010) and deriving the Gaussian MLE

(ii) defining a second-order asymptotic (von Mises) expansion for the MLE of θ and using it to define an asymptotic expansion for $q(\hat{\theta}_n)$

(iii) identifying the corresponding U-statistic, deriving its $\boxed{\tilde{\mathcal{K}}_{n,T}}$

(iv) using (iii) to derive its Edgeworth expansion (Bickel et al, 1986), whose tilted version is the saddlepoint density (Gatto & Ronchetti, 1996).

The construction at a glance

Our construct

(i) eliminati we adopt deriving

(ii) defining MLE of

(iii) identifyir

(iv) using (iii whose til 1996).

Intuition for tilted-Edgeworth (BNS & Cox, 89, Ch 4)

Let us consider again the sample average $\bar{X}$ in the i.i.d. setting. Its Edgeworth expansion reads as

$$p^{\text{Edg}}(\alpha) = \phi(\alpha)\left\{1 + \frac{1}{\sqrt{n}}\kappa_3(\alpha^3 - 3\alpha) + O(n^{-1})\right\},$$

where κ_3 is the standardized cumulant of order 3.

The construction at a glance

Our construct

(i) eliminati
we adopt
deriving

(ii) defining
MLE of

(iii) identifyi

(iv) using (iii
whose til
1996).

Intuition for tilted-Edgeworth (BNS & Cox, 89, Ch 4)

Let us consider again the sample average $\bar{X}$ in the i.i.d. setting. Its Edgeworth expansion reads as

$$p^{\mathrm{Edg}}(\alpha) = \phi(\alpha)\left\{1 + \frac{1}{\sqrt{n}}\kappa_3(\alpha^3 - 3\alpha) + O(n^{-1})\right\},$$

where κ_3 is the standardized cumulant of order 3. Computing this approx at the center ($\alpha = 0$), we kill the term of order $n^{-1/2}$ and we are left with an approx with relative error $O(n^{-1})$.

The construction at a glance

Our construct

(i) eliminati
we adopt
deriving

(ii) defining
MLE of

(iii) identifyi

(iv) using (iii
whose ti
1996).

Intuition for tilted-Edgeworth (BNS & Cox, 89, Ch 4)

Let us consider again the sample average $\bar{X}$ in the i.i.d. setting. Its Edgeworth expansion reads as

$$p^{\text{Edg}}(\alpha) = \phi(\alpha)\left\{1 + \frac{1}{\sqrt{n}}\kappa_3(\alpha^3 - 3\alpha) + O(n^{-1})\right\},$$

where κ_3 is the standardized cumulant of order 3. Computing this approx at the center ($\alpha = 0$), we kill the term of order $n^{-1/2}$ and we are left with an approx with relative error $O(n^{-1})$. To exploit that, we tilt the underling distribution in such a way that X_i is centered at α, then we compute an Edgeworth which is always at the zero!

The construction at a glance

Our construct

(i) eliminati
we adopt
deriving

(ii) defining
MLE of

(iii) identifyi

(iv) using (iii
whose ti
1996).

Intuition for tilted-Edgeworth (BNS & Cox, 89, Ch 4)

Let us consider again the sample average $\bar{X}$ in the i.i.d. setting. Its Edgeworth expansion reads as

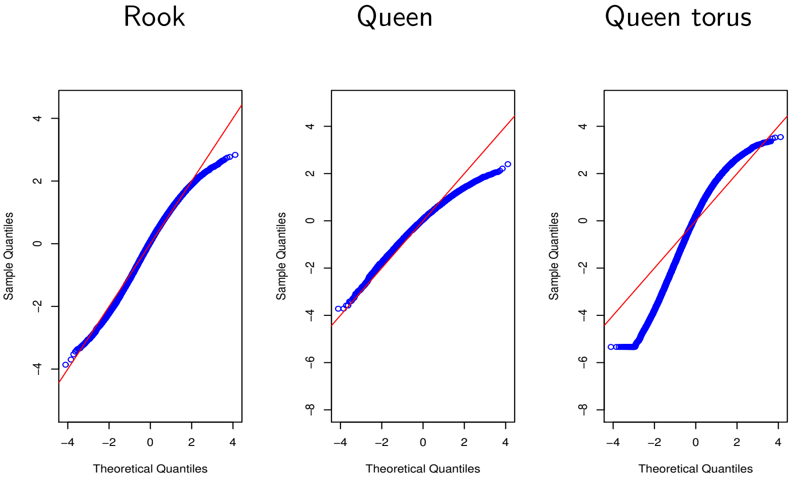
$$p^{\mathrm{Edg}}(\alpha) = \phi(\alpha)\left\{1 + \frac{1}{\sqrt{n}}\kappa_3(\alpha^3 - 3\alpha) + O(n^{-1})\right\},$$

where κ_3 is the standardized cumulant of order 3. Computing this approx at the center ($\alpha = 0$), we kill the term of order $n^{-1/2}$ and we are left with an approx with relative error $O(n^{-1})$. To exploit that, we tilt the underling distribution in such a way that X_i is centered at α, then we compute an Edgeworth which is always at the zero!

The resulting approx is the **saddlepoint density approx.**

Comparison with asymptotic theory

SAR(1): distribution of $\hat{\lambda}$ for $n = 24$ and $\lambda_0 = 0.2$



Comparison with asymptotic theory (cont'd)

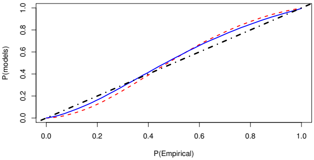
Is the saddlepoint density approximation doing any better?

Comparison with asymptotic theory (cont'd)

Is the saddlepoint density approximation doing any better?

Rook

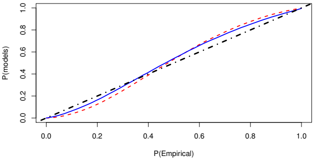
Queen



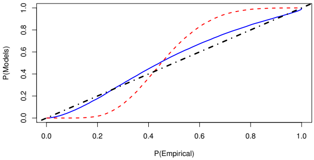
Comparison with asymptotic theory (cont'd)

Is the saddlepoint density approximation doing any better?

Rook



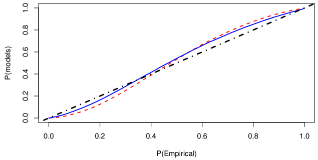
Queen



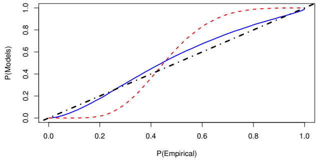
Comparison with asymptotic theory (cont'd)

Is the saddlepoint density approximation doing any better?

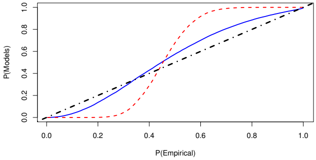
Rook



Queen

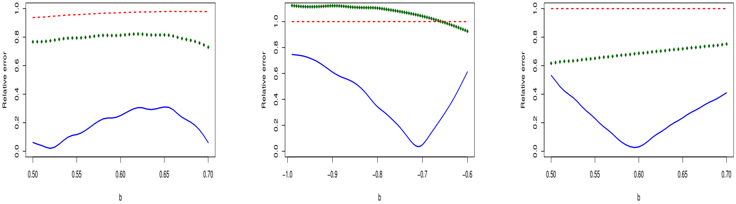


Queen torus



Comparison with Edgeworth expansion (relative error)

Rook Queen Queen torus



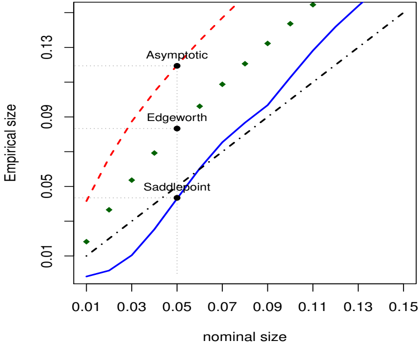
[Sad, Edgeworth, Asym, $n = 24$ and $\lambda_0 = 0.2$].

"All together now" ($n = 24$ and W_n is Rook)

We study the level of the test $\mathcal{H}_0 : \lambda_0 = 0$ vs $\mathcal{H}_1 : \lambda_0 > 0$.

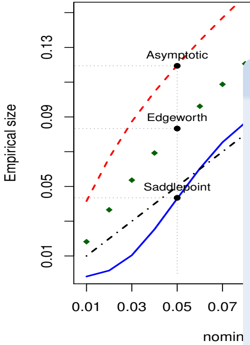
"All together now" ($n = 24$ and W_n is Rook)

We study the level of the test $\mathcal{H}_0 : \lambda_0 = 0$ vs $\mathcal{H}_1 : \lambda_0 > 0$.



“All together now” ($n = 24$ and W_n is Rook)

We study the level of the test $\mathcal{H}_0 : \lambda_0 = 0$ vs $\mathcal{H}_1 : \lambda_0 > 0$.

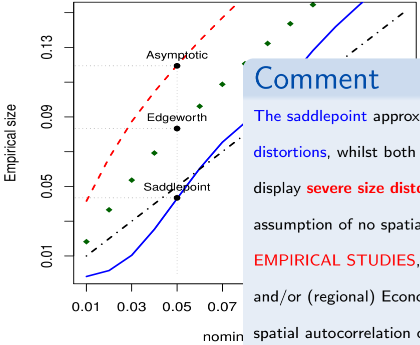


Comment

The saddlepoint approximation exhibits only mild size distortions, whilst both the asymptotic and the Edg approx display **severe size distortion**. Over-rejection of the assumption of no spatial correlation $\Rightarrow$ ALARM FOR EMPIRICAL STUDIES, e.g., in (applied) Econometrics and/or (regional) Economics.

"All together now" ($n = 24$ and W_n is Rook)

We study the level of the test $\mathcal{H}_0 : \lambda_0 = 0$ vs $\mathcal{H}_1 : \lambda_0 > 0$.



Comment

The saddlepoint approximation exhibits only mild size distortions, whilst both the asymptotic and the Edg approx display **severe size distortion**. Over-rejection of the assumption of no spatial correlation $\Rightarrow$ ALARM FOR EMPIRICAL STUDIES, e.g., in (applied) Econometrics and/or (regional) Economics. In the OECD example, spatial autocorrelation can be spuriously detected/rejected by the asymptotic theory!

Test results for the OCDE countries

SARAR(1,1) model: *p*-values of Saddlepoint (SAD) and first-order asymptotic (ASY) approximation.

		1960-1970	1971-1985	1986-2000
Weight matrix: 7 nearest neighbours				
β	SAD	0.0000	0.0000	0.0000
	ASY	0.0000	0.0000	0.0000
λ	SAD	0.5768	0.0178	0.1674
	ASY	0.5681	0.0005	0.0072
ρ	SAD	1.0000	0.0000	0.3871
	ASY	0.8800	0.0413	0.5352

Test results for the OCDE countries

Remark

- *In the sub-period 86-00, there are large differences between p-values under the two approximations. We find that there is no spatial dependence of investing rates across countries for that period, and vice-versa for the asymptotic approximation. This is in line with the overrejection of the ASY that we find in the Monte Carlo experiments for* λ.

Test results for the OCDE countries

Remark

- *In the sub-period 86-00, there are large differences between p-values under the two approximations. We find that there is no spatial dependence of investing rates across countries for that period, and vice-versa for the asymptotic approximation. This is in line with the overrejection of the ASY that we find in the Monte Carlo experiments for λ.*
- *In the sub-period 71-85, both types of spatial dependence coexist, namely in Inv_{nt} and the disturbances E_{nt}, and thus we find an additional spillover through the contemporary shocks between countries.*

Test results for the OCDE countries

Remark

- *In the sub-period 86-00, there are large differences between p-values under the two approximations. We find that there is no spatial dependence of investing rates across countries for that period, and vice-versa for the asymptotic approximation. This is in line with the overrejection of the ASY that we find in the Monte Carlo experiments for λ.*
- *In the sub-period 71-85, both types of spatial dependence coexist, namely in Inv_{nt} and the disturbances E_{nt}, and thus we find an additional spillover through the contemporary shocks between countries.*
- *The spillover seems thus to go also through the innovations, i.e., through the unexpected part, a finding not documented in the literature.*

Take home message

First-order asymptotics and Edgeworth expansions may deliver poor inference in the setting of dependent processes (like e.g. spatial autoregressive models with and without covariates) in small samples since they exhibit severe absolute and relative distortions in the tail areas.

Take home message

First-order asymptotics and Edgeworth expansions may deliver poor inference in the setting of dependent processes (like e.g. spatial autoregressive models with and without covariates) in small samples since they exhibit severe absolute and relative distortions in the tail areas.

Saddlepoint techniques are fast (no resampling) and accurate, and provide a better alternative than first-order asymptotics, Edgeworth expansions and bootstrap methods (faster).

Thank you

For questions: olivier.scaillet@unige.ch

Laplace in brief

The Laplace method is typically applied to approximate integrals of type:

$$\int_a^b \exp^{v\, k(x)}\, dx,$$

where $k(\cdot)$ has unique maximum at $x_0 \in (a, b) \subset \mathbb{R}$ and $v \in \mathbb{R}^+$ is large.

Laplace in brief

The Laplace method is typically applied to approximate integrals of type:

$$\int_a^b \exp^{v \, k(x)} \, dx,$$

where $k(\cdot)$ has unique maximum at $x_0 \in (a, b) \subset \mathbb{R}$ and $v \in \mathbb{R}^+$ is large. A second-order Taylor expansion for $k(\cdot)$ yields

$$\int_a^b \exp^{v \, k(x)} \, dx \sim \exp^{v \, k(x_0)} \int_{x_0-\epsilon}^{x_0+\epsilon} \exp^{v \, k^{('')}(x_0) \frac{x^2}{2}} \, dx$$

Laplace in brief

The Laplace method is typically applied to approximate integrals of type:

$$\int_a^b \exp^{v\, k(x)}\, dx,$$

where $k(\cdot)$ has unique maximum at $x_0 \in (a, b) \subset \mathbb{R}$ and $v \in \mathbb{R}^+$ is large. A second-order Taylor expansion for $k(\cdot)$ yields

$$\int_a^b \exp^{v\, k(x)}\, dx \sim \exp^{v\, k(x_0)} \int_{x_0-\epsilon}^{x_0+\epsilon} \exp^{v\, k^{('')}(x_0)\frac{x^2}{2}}\, dx \sim \sqrt{\frac{2\pi}{-v\, k^{('')}(x_0)}} \exp^{v\, k(x_0)},$$

Laplace in brief

The Laplace method is typically applied to approximate integrals of type:

$$\int_a^b \exp^{v\, \boxed{k}(x)}\, dx,$$

where $k(\cdot)$ has unique maximum at $x_0 \in (a, b) \subset \mathbb{R}$ and $v \in \mathbb{R}^+$ is large. A second-order Taylor expansion for $k(\cdot)$ yields

$$\int_a^b \exp^{v\, \boxed{k}(x)}\, dx \sim \exp^{v\, \boxed{k}(x_0)} \int_{x_0-\epsilon}^{x_0+\epsilon} \exp^{v\, \boxed{k^{('')}}(x_0)\frac{x^2}{2}}\, dx \sim \sqrt{\frac{2\pi}{-v\, \boxed{k^{('')}}(x_0)}}\, \exp^{v\, \boxed{k}(x_0)},$$

where (i) for $\epsilon > 0$, we deform the path of integration $\int_a^b \mapsto \int_{x_0-\epsilon}^{x_0+\epsilon}$

Laplace in brief

The Laplace method is typically applied to approximate integrals of type:

$$\int_a^b \exp^{v\, k(x)}\, dx,$$

where $k(\cdot)$ has unique maximum at $x_0 \in (a, b) \subset \mathbb{R}$ and $v \in \mathbb{R}^+$ is large. A second-order Taylor expansion for $k(\cdot)$ yields

$$\int_a^b \exp^{v\, k(x)}\, dx \sim \exp^{v\, k(x_0)} \int_{x_0-\epsilon}^{x_0+\epsilon} \exp^{v\, k^{('')}(x_0)\frac{x^2}{2}}\, dx \sim \sqrt{\frac{2\pi}{-v\, k^{('')}(x_0)}} \exp^{v\, k(x_0)},$$

where (i) for $\epsilon > 0$, we deform the path of integration $\int_a^b \mapsto \int_{x_0-\epsilon}^{x_0+\epsilon}$ and (ii) we solve the Gaussian integral—getting an approx featuring relative error, under suitable assumptions.

Jump Back