GENEVA SCHOOL OF ECONOMICS AND MANAGEMENT

UNIVERSITÉ DE GENÈVE

Saddlepoint techniques for the statistical analysis of time series

Davide La Vecchia

University of Geneva

28-March-2024, ESSEC (Paris)

Three-fold aim of the talk

In the setting of time series with short- or long-memory, I:

Three-fold aim of the talk

In the setting of time series with short- or long-memory, I:

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

Three-fold aim of the talk

In the setting of time series with short- or long-memory, I:

- Illustrate that first-order asymptotic theory suffers from finite sample distortions.
- Develop saddlepoint techniques (for pdf/cdf approximation, p-values, and testing) which perform well in small samples and feature higher-order accuracy.

Three-fold aim of the talk

In the setting of time series with short- or long-memory, I:

- Illustrate that first-order asymptotic theory suffers from finite sample distortions.
- Develop saddlepoint techniques (for pdf/cdf approximation, *p*-values, and testing) which perform well in small samples and feature higher-order accuracy.
- Compare the saddlepoint density approximation to Edgeworth expansion and/or resampling methods, which represent the main competitors for finite sample analysis.

Three-fold aim of the talk

In the setting of time series with short- or long-memory, I:

- Illustrate that first-order asymptotic theory suffers from finite sample distortions.
- Develop saddlepoint techniques (for pdf/cdf approximation, *p*-values, and testing) which perform well in small samples and feature higher-order accuracy.
- Compare the saddlepoint density approximation to Edgeworth expansion and/or resampling methods, which represent the main competitors for finite sample analysis.

$\Rightarrow$ **The need for saddlepoint techniques is rooted in both the theory and practice of statistics and other disciplines.**

Theoretical statistics motivation

Theorem (Karlin-Rubin, as stated in Casella-Berger)

Consider testing

$$\mathcal{H}_0 : \theta \leq \theta^0 \quad \textit{versus} \quad \mathcal{H}_1 : \theta > \theta^0.$$

Suppose that T is a sufficient statistic for θ and the family of pdfs or pmfs $\{g(t \mid \theta) : \theta \in \Theta\}$ of T has a Monotone Likelihood Ratio. Then for any t_0, the test that rejects $\mathcal{H}_0$ if and only if $T > t_0$ is a UMP level α test, where

$$\alpha = P_{\theta^0}(T > t_0).$$

Financial motivation

Diffusions-type processes

$$dY(t) = \mu(Y_t)dt + \sigma(Y_t)dW_t + J_t dN_t$$

where N_t is a Poisson process, J_t is the jump size, W_t is a BM.

Financial motivation

Diffusions-type processes

$$dY(t) = \mu(Y_t)dt + \sigma(Y_t)dW_t + J_t dN_t$$

where N_t is a Poisson process, J_t is the jump size, W_t is a BM.

1. calculation of Value at Risk (VaR) or option prices: see e.g. Ait-Sahalia & Yu (2006, JoE), Glasserman & Kim (2009, JED&C), Rogers & Zane (1999, AoAP), Ait-Sahalia & Leaven (2023). For VaR we need the CDF:

$$P(Y_{t+\Delta} \leq a_0 | Y_t = x)$$

Financial motivation

Diffusions-type processes

$$dY(t) = \mu(Y_t)dt + \sigma(Y_t)dW_t + J_t dN_t$$

where N_t is a Poisson process, J_t is the jump size, W_t is a BM.

1. calculation of Value at Risk (VaR) or option prices: see e.g. Ait-Sahalia & Yu (2006, JoE), Glasserman & Kim (2009, JED&C), Rogers & Zane (1999, AoAP), Ait-Sahalia & Leaven (2023). For VaR we need the CDF:

$$P(Y_{t+\Delta} \leq a_0 | Y_t = x)$$

2. transition density for time interval $\Delta > 0$ and for $\tau \in \mathbb{R}$ (by Fourier inversion, $i^2 = -1$)

$$p(y|x,\Delta) = \frac{1}{2\pi i} \int_{\tau - i\infty}^{\tau + i\infty} \exp\{ \boxed{K_{y|x}}(\Delta, z; x) - zy\} dz$$

needed for inference on the model parameter; see e.g. Bibby et al. (Handbook of Fin. Econ., 2010), La Vecchia & Trojani, (JASA, 2012)

Financial motivation

Diffusions-type processes

$$dY(t) = \mu(Y_t)dt + \sigma(Y_t)dW_t + J_t dN_t$$

where N_t is a Poisson process, J_t is the jump size, W_t is a BM.

1. calculation of Value at Risk (VaR) or option prices: see e.g. Ait-Sahalia & Yu (2006, JoE), Glasserman & Kim (2009, JED&C), Rogers & Zane (1999, AoAP), Ait-Sahalia & Leaven (2023). For VaR we need the CDF:

$$P(Y_{t+\Delta} \leq a_0 | Y_t = x)$$

2. transition density for time interval $\Delta > 0$ and for $\tau \in \mathbb{R}$ (by Fourier inversion, $i^2 = -1$)

$$p(y|x,\Delta) = \frac{1}{2\pi i} \int_{\tau - i\infty}^{\tau + i\infty} \exp\{ \boxed{K_{y|x}}(\Delta, z; x) - zy\} dz$$

needed for inference on the model parameter; see e.g. Bibby et al. (Handbook of Fin. Econ., 2010), La Vecchia & Trojani, (JASA, 2012)

3. distribution of RV estimators, see e.g. Zhang et al. (2011, JoE), to improve on Gaussian approx

Financial motivation

Diffusions-type processes

$$dY(t) = \mu(Y_t)dt + \sigma(Y_t)dW_t + J_t dN_t$$

where N_t is a Poisson process, J_t is the jump size, W_t is a BM.

1. calculation of Value at Risk (VaR) or option prices: see e.g. Ait-Sahalia & Yu (2006, JoE), Glasserman & Kim (2009, JED&C), Rogers & Zane (1999, AoAP), Ait-Sahalia & Leaven (2023). For VaR we need the CDF:

$$P(Y_{t+\Delta} \leq a_0 | Y_t = x)$$

2. transition density for time interval $\Delta > 0$ and for $\tau \in \mathbb{R}$ (by Fourier inversion, $i^2 = -1$)

$$p(y|x, \Delta) = \frac{1}{2\pi i} \int_{\tau - i\infty}^{\tau + i\infty} \exp\{ \boxed{K_{y|x}}(\Delta, z; x) - zy\} dz$$

needed for inference on the model parameter; see e.g. Bibby et al. (Handbook of Fin. Econ., 2010), La Vecchia & Trojani, (JASA, 2012)

3. distribution of RV estimators, see e.g. Zhang et al. (2011, JoE), to improve on Gaussian approx
4. ...

Theoretical statistics motivation

Typical statistical problem: For a given statistic $T : \operatorname{dom} T \to \mathbb{R}$ or an estimator $\hat{\theta}_n$, tail probabilities or quantiles at different levels are needed to carry out statistical inference (essentially, tests and confidence intervals).

Theoretical statistics motivation

Typical statistical problem: For a given statistic $T : \operatorname{dom} T \to \mathbb{R}$ or an estimator $\hat{\theta}_n$, tail probabilities or quantiles at different levels are needed to carry out statistical inference (essentially, tests and confidence intervals).

Unless the (test) statistic T or the estimator have a simple form (e.g. linear in the observations) 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.

Theoretical statistics motivation

Typical statistical problem: For a given statistic $T : \operatorname{dom} T \to \mathbb{R}$ or an estimator $\hat{\theta}_n$, tail probabilities or quantiles at different levels are needed to carry out statistical inference (essentially, tests and confidence intervals).

Unless the (test) statistic T or the estimator have a simple form (e.g. linear in the observations) 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 approximations

Theoretical statistics motivation

We can approximate tail probabilities via

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

Theoretical statistics motivation

We can approximate tail probabilities via

- *Asymptotic theory*: use of Central Limit Theorem to get a Gaussian approximation in large samples
- *Analytical techniques*: use of expansions (Edgeworth, saddlepoint) to get an approximation in small samples

Theoretical statistics motivation

We can approximate tail probabilities via

- *Asymptotic theory*: use of Central Limit Theorem to get a Gaussian approximation in large samples
- *Analytical techniques*: use of expansions (Edgeworth, saddlepoint) to get an approximation in small samples
- *Resampling techniques*: use of resampling (bootstrap, subsampling) computer-aided methods to get an approximation in small samples

Theoretical statistics motivation

We can approximate tail probabilities via

- *Asymptotic theory*: use of Central Limit Theorem to get a Gaussian approximation in large samples
- *Analytical techniques*: use of expansions (Edgeworth, saddlepoint) to get an approximation in small samples
- *Resampling techniques*: use of resampling (bootstrap, subsampling) computer-aided methods to get an approximation in small samples

Analytical and resampling techniques can achieve higher order refinements over the first order asymptotic theory

Theoretical statistics motivation

The use of asymptotic techniques is twofold. First, they enable us to find approximate tests and confidence intervals [practical use]. Second, they can be applied to study the properties of statistical procedures [theoretical use].

[A.W. van der Vaart]

Theoretical statistics motivation

The use of asymptotic techniques is twofold. First, they enable us to find approximate tests and confidence intervals [practical use]. Second, they can be applied to study the properties of statistical procedures [theoretical use].

[A.W. van der Vaart]

The purpose of asymptotic theory in statistics is simple: to provide usable approximations *before passage to the limit.*

[J. Tukey]

Theoretical statistics motivation

Let $X \sim \mu$ with measure absolutely continuous w.r.t. the Lebesgue measure and having density f_X. We are given a random sample $\boldsymbol{X} = (X_1, ..., X_n)$ of i.i.d. copies of X, whose cumulant generating function (cgf):

$$\mathcal{K}(\upsilon) = \ln E_\mu[\exp(\upsilon X)], \;\; \upsilon \in \mathbb{R} \;\; \text{and} \;\; M(\upsilon) = E_\mu[\exp(\upsilon X)]$$

is the well-defined and $E_\mu[X] = 0$. The standardized mean (statistic, $T(\boldsymbol{X})$) has expression:

$$\sqrt{n}\bar{X}_n = \sum_{i=1}^{n} \frac{X_i}{\sqrt{n}}.$$

Theoretical statistics motivation

Let $X \sim \mu$ with measure absolutely continuous w.r.t. the Lebesgue measure and having density f_X. We are given a random sample $\boldsymbol{X} = (X_1, ..., X_n)$ of i.i.d. copies of X, whose cumulant generating function (cgf):

$$\mathcal{K}(\upsilon) = \ln E_\mu[\exp(\upsilon X)], \; \upsilon \in \mathbb{R} \;\text{ and }\; M(\upsilon) = E_\mu[\exp(\upsilon X)]$$

is the well-defined and $E_\mu[X] = 0$. The standardized mean (statistic, $T(\boldsymbol{X})$) has expression:

$$\sqrt{n}\bar{X}_n = \sum_{i=1}^{n} \frac{X_i}{\sqrt{n}}.$$

Edgeworth expansion to approx the density f_n of the standardized mean: 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.

Theoretical statistics motivation

Let $X \sim \mu$ with measure absolutely continuous w.r.t. the Lebesgue measure and having density f_X. We are given a random sample $\boldsymbol{X} = (X_1, ..., X_n)$ of i.i.d. copies of X, whose cumulant generating function (cgf):

$$\mathcal{K}(\upsilon) = \ln E_\mu[\exp(\upsilon X)], \ \upsilon \in \mathbb{R} \quad \text{and} \quad M(\upsilon) = E_\mu[\exp(\upsilon X)]$$

is the well-defined and $E_\mu[X] = 0$. The standardized mean (statistic, $T(\boldsymbol{X})$) has expression:

Ec
Ta
ar
in


PROJECT euclid
BROWSE RESOURCES ABOUT
Statist. Sci.
Vol.38 • No. 1 • February 2023
Home >
STATISTICAL SCIENCE
February 2023
On Some Connections Between Esscher's Tilting Saddlepoint Approximations, and Optimal Transportation: A Statistical Perspective
Davide La Vecchia, Elvezio Ronchetti, Andrej Ilievski
Author Affiliations +

Theoretical statistics motivation

This yields an expansion of the density of $\sqrt{n}\bar{X}_n$ in powers of $n^{-1/2}$, where the leading term is the normal density and higher order terms correct for skewness, kurtosis:

$$g_{\text{Edg}}(s) = \phi(s)$$

Theoretical statistics motivation

This yields an expansion of the density of $\sqrt{n}\bar{X}_n$ in powers of $n^{-1/2}$, where the leading term is the normal density and higher order terms correct for skewness, kurtosis:

$$g_{\mathrm{Edg}}(s) = \phi(s) + n^{-1/2}\frac{\lambda_3}{6}(s^3 - 3s)\phi(s)$$

Theoretical statistics motivation

This yields an expansion of the density of $\sqrt{n}\bar{X}_n$ in powers of $n^{-1/2}$, where the leading term is the normal density and higher order terms correct for skewness, kurtosis:

$$\begin{aligned} g_{\mathrm{Edg}}(s) &= \phi(s) + n^{-1/2}\frac{\lambda_3}{6}(s^3 - 3s)\phi(s) \\ &+ n^{-1}\left[\frac{\lambda_4}{24}(s^4 - 6s^2 + 3) + \frac{\lambda_3^2}{72}(s^6 - 15s^4 + 45s^2 + 15)\right]\phi(s), \end{aligned}$$

with λ_3 and λ_4 being the standardized cumulants of X of order three and four, while ϕ is the pdf of a standard normal.

Theoretical statistics motivation

This yields an expansion of the density of $\sqrt{n}\bar{X}_n$ in powers of $n^{-1/2}$, where the leading term is the normal density and higher order terms correct for skewness, kurtosis:

$$\begin{aligned} g_{\text{Edg}}(s) &= \phi(s) + n^{-1/2}\frac{\lambda_3}{6}(s^3 - 3s)\phi(s) \\ &+ n^{-1}\left[\frac{\lambda_4}{24}(s^4 - 6s^2 + 3) + \frac{\lambda_3^2}{72}(s^6 - 15s^4 + 45s^2 + 15)\right]\phi(s), \end{aligned}$$

with λ_3 and λ_4 being the standardized cumulants of X of order three and four, while ϕ is the pdf of a standard normal.

Remark

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

Theoretical statistics motivation

This yields an expansion of the density of $\sqrt{n}\bar{X}_n$ in powers of $n^{-1/2}$, where the leading term is the normal density and higher order terms correct for skewness, kurtosis:

$$
\begin{aligned}
g_{\text{Edg}}(s) &= \phi(s) + n^{-1/2}\frac{\lambda_3}{6}(s^3 - 3s)\phi(s) \\
&+ n^{-1}\left[\frac{\lambda_4}{24}(s^4 - 6s^2 + 3) + \frac{\lambda_3^2}{72}(s^6 - 15s^4 + 45s^2 + 15)\right]\phi(s),
\end{aligned}
$$

with λ_3 and λ_4 being the standardized cumulants of X of order three and four, while ϕ is the pdf of a standard normal.

Remark

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

- *they can be inaccurate in the tails*

Theoretical statistics motivation

This yields an expansion of the density of $\sqrt{n}\bar{X}_n$ in powers of $n^{-1/2}$, where the leading term is the normal density and higher order terms correct for skewness, kurtosis:

$$
\begin{aligned}
g_{\mathrm{Edg}}(s) &= \phi(s) + n^{-1/2}\frac{\lambda_3}{6}(s^3 - 3s)\phi(s) \\
&+ n^{-1}\left[\frac{\lambda_4}{24}(s^4 - 6s^2 + 3) + \frac{\lambda_3^2}{72}(s^6 - 15s^4 + 45s^2 + 15)\right]\phi(s),
\end{aligned}
$$

with λ_3 and λ_4 being the standardized cumulants of X of order three and four, while ϕ is the pdf of a standard normal.

Remark

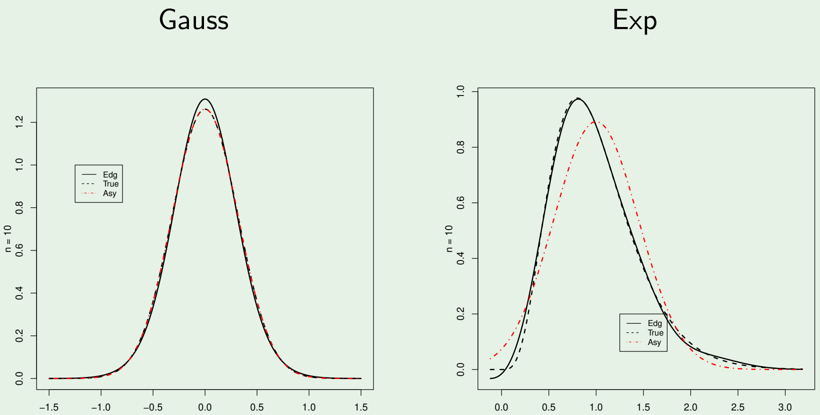
By construction, Edgeworth expansions provide in general a good approximation in the center of the density, **BUT**

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

Theoretical statistics motivation

Example (Sample mean)

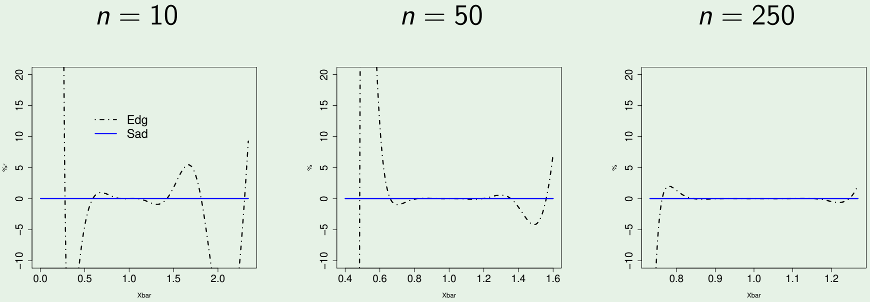
For Asy and Edg, consider $\bar{X}_n$ for $n = 10, 50, 250$, for $X_i \sim \mathcal{N}(0,1)$ and $X_i \sim \exp(1)$



Theoretical statistics motivation

Example (cont'd)

for the exponential case, rel. err. $= 100 \cdot$ (true -approx)/true



Any other higher oder technique to cope with these issues? saddlepoint approx...

Theoretical statistics motivation

Example (cont'd)

In this example about $\bar{X}_n$, we know the c.g.f. and the saddlepoint density approx $g_n(s)$ is (Daniels (1954)):

$$g_n(s) = \left[\frac{n}{2\pi \mathcal{K}'' \{v(s)\}}\right]^{1/2} \exp\left(n\left[\mathcal{K}\{v(s)\} - v(s)s\right]\right) \quad (1)$$

and $v(s)$ (saddlepoint) is the solution to

$$\mathcal{K}'(v) - s = 0,$$

Theoretical statistics motivation

Example (cont'd)

In this example about $\bar{X}_n$, we know the c.g.f. and the saddlepoint density approx $g_n(s)$ is (Daniels (1954)):

$$g_n(s) = \left[\frac{n}{2\pi \mathcal{K}'' \{v(s)\}} \right]^{1/2} \exp\left(n \left[\mathcal{K} \{v(s)\} - v(s)s \right] \right) \tag{1}$$

and $v(s)$ (saddlepoint) is the solution to

$$\mathcal{K}'(v) - s = 0,$$

namely, we look for $v(s)$ such that X has expected value equal to s.

Theoretical statistics motivation

Example (cont'd)

- To find the saddlepoint we need to solve

$$\mathcal{K}'(\upsilon) - s = 0 \iff \sup_{\upsilon \in \mathbb{R}} [\mathcal{K}\{\upsilon(s)\} - \upsilon(s)s] = -\mathcal{K}^{\dagger}(s),$$

with $\mathcal{K}^{\dagger}(s)$ being the Legendre transform of $\mathcal{K}\{\upsilon(s)\}$

Theoretical statistics motivation

Example (cont'd)

- To find the saddlepoint we need to solve

$$\boxed{\mathcal{K}'}(v) - s = 0 \iff \sup_{v \in \mathbb{R}}[\boxed{\mathcal{K}}\{v(s)\} - v(s)s] = -\boxed{\mathcal{K}^{\dagger}}(s),$$

  with $\boxed{\mathcal{K}^{\dagger}}(s)$ being the Legendre transform of $\boxed{\mathcal{K}}\{v(s)\}$

- $g_n(s)$ is a "Gaussian-type" integral with both mean and variance that depends on s: it is a density-like object that cannot take on negative values ($\neq$ Edg).

Theoretical statistics motivation

Example (cont'd)

- To find the saddlepoint we need to solve

$$\mathcal{K}'(v) - s = 0 \iff \sup_{v \in \mathbb{R}} [\mathcal{K}\{v(s)\} - v(s)s] = -\mathcal{K}^{\dagger}(s),$$

  with $\mathcal{K}^{\dagger}(s)$ being the Legendre transform of $\mathcal{K}\{v(s)\}$

- $g_n(s)$ is a "Gaussian-type" integral with both mean and variance that depends on s: it is a density-like object that cannot take on negative values ($\neq$ Edg).
  E.g. for i.i.d. standard Gaussian rvs: $\mathcal{K}(v) = \frac{\nu^2}{2}$, $\mathcal{K}'(v) = v$ and $\mathcal{K}''(v) = 1$, the saddlepoint is defined by $\mathcal{K}'(v) = s$, thus $v(s) = s$ and

Theoretical statistics motivation

Example (cont'd)

- To find the saddlepoint we need to solve

$$\boxed{\mathcal{K}'}(v) - s = 0 \iff \sup_{v\in\mathbb{R}}[\boxed{\mathcal{K}}\{v(s)\} - v(s)s] = -\boxed{\mathcal{K}^\dagger}(s),$$

  with $\boxed{\mathcal{K}^\dagger}(s)$ being the Legendre transform of $\boxed{\mathcal{K}}\{v(s)\}$

- $g_n(s)$ is a "Gaussian-type" integral with both mean and variance that depends on s: it is a density-like object that cannot take on negative values ($\neq$ Edg).
  E.g. for i.i.d. standard Gaussian rvs: $\boxed{\mathcal{K}}(v) = \frac{\nu^2}{2}$, $\boxed{\mathcal{K}'}(v) = v$ and $\boxed{\mathcal{K}''}(v) = 1$, the saddlepoint is defined by $\boxed{\mathcal{K}'}(v) = s$, thus $v(s) = s$ and

- for $\bar{X}_n$, $g_n(s) = \left(\frac{n}{2\pi}\right)^{1/2} e^{-\frac{ns^2}{2}}$ pdf of $\mathcal{N}\left(0, \frac{1}{n}\right)$

Theoretical statistics motivation

Example (cont'd)

- To find the saddlepoint we need to solve

$$\mathcal{K}'(v) - s = 0 \iff \sup_{v \in \mathbb{R}}[\mathcal{K}\{v(s)\} - v(s)s] = -\mathcal{K}^{\dagger}(s),$$

  with $\mathcal{K}^{\dagger}(s)$ being the Legendre transform of $\mathcal{K}\{v(s)\}$

- $g_n(s)$ is a "Gaussian-type" integral with both mean and variance that depends on s: it is a density-like object that cannot take on negative values ($\neq$ Edg).
  E.g. for i.i.d. standard Gaussian rvs: $\mathcal{K}(v) = \frac{v^2}{2}$, $\mathcal{K}'(v) = v$ and $\mathcal{K}''(v) = 1$, the saddlepoint is defined by $\mathcal{K}'(v) = s$, thus $v(s) = s$ and

- for $\bar{X}_n$, $g_n(s) = \left(\frac{n}{2\pi}\right)^{1/2} e^{-\frac{ns^2}{2}}$ pdf of $\mathcal{N}\left(0, \frac{1}{n}\right)$
- for $\sqrt{n}\bar{X}_n$, (by Jacobian formula) $g_n(s) = \left(\frac{1}{2\pi}\right)^{1/2} e^{-\frac{s^2}{2}}$ pdf of $\mathcal{N}(0, 1)$

Theoretical statistics motivation

Example (cont'd)

- To find the saddlepoint we need to solve

$$\mathcal{K}'(v) - s = 0 \iff \sup_{v \in \mathbb{R}}[\mathcal{K}\{v(s)\} - v(s)s] = -\mathcal{K}^{\dagger}(s),$$

with $\mathcal{K}^{\dagger}(s)$ being the Legendre transform of $\mathcal{K}\{v(s)\}$

- $g_n(s)$ is a "Gaussian-type" integral with both mean and variance that depends on s: it is a density-like object that cannot take on negative values ($\neq$ Edg).
E.g. for i.i.d. standard Gaussian rvs: $\mathcal{K}(v) = \frac{\nu^2}{2}$, $\mathcal{K}'(v) = v$ and $\mathcal{K}''(v) = 1$, the saddlepoint is defined by $\mathcal{K}'(v) = s$, thus $v(s) = s$ and

- for $\bar{X}_n$, $g_n(s) = \left(\frac{n}{2\pi}\right)^{1/2} e^{-\frac{ns^2}{2}}$ pdf of $\mathcal{N}\left(0, \frac{1}{n}\right)$
- for $\sqrt{n}\bar{X}_n$, (by Jacobian formula) $g_n(s) = \left(\frac{1}{2\pi}\right)^{1/2} e^{-\frac{s^2}{2}}$ pdf of $\mathcal{N}(0, 1)$

Theoretical statistics motivation

Example (cont'd)

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

$$f_n(s) = g_n(s)\left\{1 + O\left(n^{-1}\right)\right\} \tag{2}$$

Theoretical statistics motivation

Example (cont'd)

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

$$f_n(s) = g_n(s)\left\{1 + O\left(n^{-1}\right)\right\} \tag{2}$$

- The density g_n is obtained by approximating the Fourier inversion of M^n, which yields f_n:

$$\begin{aligned} f_n(s) &= \frac{n}{2\pi}\int_{-\infty}^{\infty} e^{-i\upsilon n s} M^n(i\upsilon)d\upsilon \overset{(z=i\upsilon)}{=} \frac{n}{2\pi i}\int_{\mathcal{I}} e^{-nzs} M^n(z)dz \\ &= \frac{n}{2\pi i}\int_{\tau-i\infty}^{\tau+i\infty} \exp\{n(\mathcal{K}(z) - zs)\}dz, \tau \in \mathbb{R} \end{aligned}$$

Theoretical statistics motivation

Example (cont'd)

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

$$f_n(s) = g_n(s)\left\{1 + O\left(n^{-1}\right)\right\} \tag{2}$$

- The density g_n is obtained by approximating the Fourier inversion of M^n, which yields f_n:

$$\begin{aligned} f_n(s) &= \frac{n}{2\pi}\int_{-\infty}^{\infty} e^{-i\upsilon n s} M^n(i\upsilon)d\upsilon \overset{(z=i\upsilon)}{=} \frac{n}{2\pi i}\int_{\mathcal{I}} e^{-nzs} M^n(z)dz \\ &= \frac{n}{2\pi i}\int_{\tau-i\infty}^{\tau+i\infty} \exp\{n(\mathcal{K}(z) - zs)\}dz, \tau \in \mathbb{R} \end{aligned}$$

Theoretical statistics motivation

The sadd approx is obtained via the method of the steepest descent: this is a general technique to compute asymptotic expansions of integrals

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

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

Theoretical statistics motivation

The sadd approx is obtained via the method of the steepest descent: this is a general technique to compute asymptotic expansions of integrals

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

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). See Daniels (AoMS, 1954).

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

Theoretical statistics motivation

Alternative: derive f_n via convex analysis.

Theoretical statistics motivation

Alternative: derive f_n via convex analysis.

Idea (Sadd from Edg)

We rely on the method of the conjugate density or tilted Edgeworth:

Theoretical statistics motivation

Alternative: derive f_n via convex analysis.

Idea (Sadd from Edg)

We rely on the method of the conjugate density or tilted Edgeworth:

- **by means of $\upsilon(s)$, recenter/Esscher tilt** *the density of X: we embed the original density f_X into an exponential family, and then define the (conjugate) density h_s such that it centers at s the density of the rv ($f_X \mapsto h_s$ via $\upsilon(s)$)*

Theoretical statistics motivation

Alternative: derive f_n via convex analysis.

Idea (Sadd from Edg)

We rely on the method of the conjugate density or tilted Edgeworth:

- **by means of $\upsilon(s)$, recenter/Esscher tilt** *the density of X: we embed the original density f_X into an exponential family, and then define the (conjugate) density h_s such that it centers at s the density of the rv ($f_X \mapsto h_s$ via $\upsilon(s)$)*
- *compute a low-order* **Edgeworth expansion** *for the tilted density (centered at s, so it works well!) to obtain $g_n(s)$*

Theoretical statistics motivation

Alternative: derive f_n via convex analysis.

Idea (Sadd from Edg)

We rely on the method of the conjugate density or tilted Edgeworth:

- **by means of $\upsilon(s)$, recenter/Esscher tilt** *the density of X: we embed the original density f_X into an exponential family, and then define the (conjugate) density h_s such that it centers at s the density of the rv ($f_X \mapsto h_s$ via $\upsilon(s)$)*
- *compute a low-order* **Edgeworth expansion** *for the tilted density (centered at s, so it works well!) to obtain $g_n(s)$*
- **repeat** *this procedure* **for every** $s \in \mathbb{R}$

Theoretical statistics motivation

Alternative: derive f_n via convex analysis.

Idea (Sadd from Edg)

We rely on the method of the conjugate density or tilted Edgeworth:

- **by means of $\upsilon(s)$, recenter/Esscher tilt** *the density of X: we embed the original density f_X into an exponential family, and then define the (conjugate) density h_s such that it centers at s the density of the rv ($f_X \mapsto h_s$ via $\upsilon(s)$)*
- *compute a low-order* **Edgeworth expansion** *for the tilted density (centered at s, so it works well!) to obtain $g_n(s)$*
- **repeat** *this procedure* **for every** $s \in \mathbb{R}$

$\Rightarrow$ *saddlepoint density approximation is a sequence of low-order local approximations; see Easton & Ronchetti (1986), JASA and Wang (1992).*

Practical motivations for dependent data

Many macroeconomic time series display a persistent time trend and contain only a few observations recorded at annual frequency. Much controversy in macroeconometrics has revolved around the suitability of ARIMA models; see the seminal paper of Nelson and Plosser (1982) and Gil-Alana and Robinson (1997) for a review of the literature.

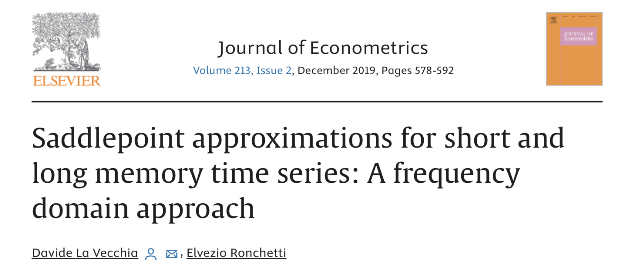
Practical motivations for dependent data

Many macroeconomic time series display a persistent time trend and contain only a few observations recorded at annual frequency. Much controversy in macroeconometrics has revolved around the suitability of ARIMA models; see the seminal paper of Nelson and Plosser (1982) and Gil-Alana and Robinson (1997) for a review of the literature.

Within this setting, to model the slow decay of the autocorrelation function displayed by many macroeconomic time series, the use of (Gaussian) FARIMA models and first order Gaussian asymptotic theory (Wald-type test statistics) is routinely applied for confidence intervals and testing statistical hypotheses.

Practical motivations for dependent data

Many macroeconomic time series display a persistent time trend and contain only a few observations recorded at annual frequency. Much controversy in macroeconometrics has revolved around the suitability of ARIMA models; see the seminal paper of Nelson and Plosser (1982) and Gil-Alana and Robinson (1997) for a review of the literature.


ELSEVIER

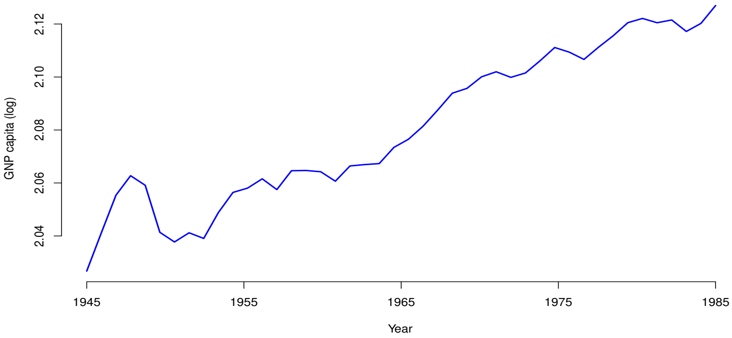
Journal of Econometrics
Volume 213, Issue 2, December 2019, Pages 578-592

Saddlepoint approximations for short and long memory time series: A frequency domain approach

Davide La Vecchia, Elvezio Ronchetti

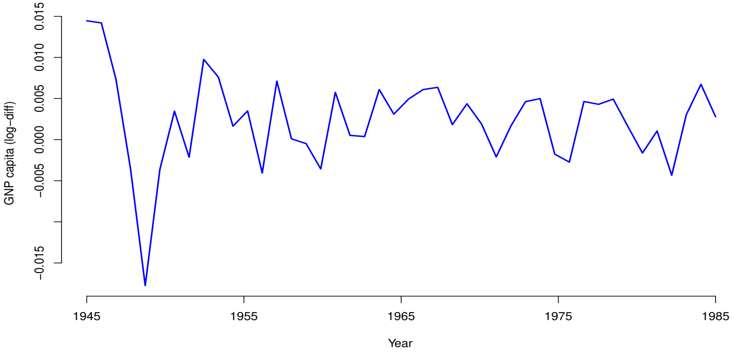
Practical motivations for dependent data

Focus on the extended Nelson and Plosser data set: plot log-GNP per capita (other time series available in the JoE paper)



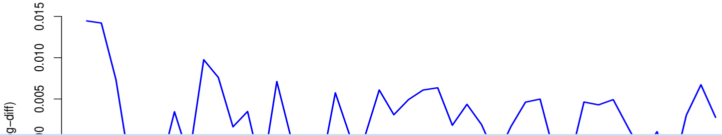
Practical motivations for dependent data

Focus on the extended Nelson and Plosser data set: plot log-diff GNP per capita (other time series available in the JoE paper)



Practical motivations for dependent data

Focus on the extended Nelson and Plosser data set: plot log-diff GNP per capita (other time series available in the JoE paper)



Remark

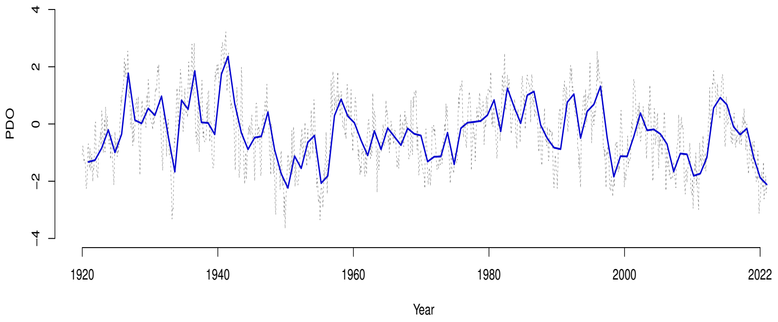
In the literature one is typically testing for the presence of long memory: ARFIMA models and

$$\mathcal{H}_0: \; d = 0 \quad vs \quad \mathcal{H}_1: \; d > 0$$

we resort on an M-estimator (Whittle), which is asymptotically χ^2 *Wald-type test statistics are applied when* $n = 44$*. Is this a sensible procedure? Is the asymptotics suffering from size distortion due to the small sample size?*

Practical motivations for dependent data

The Pacific Decadal Oscillation (PDO) index measures the climatological situation of the Southern hemisphere: its extremes correspond to episodes of abnormal weather conditions.



Practical motivations for dependent data

The Pacific Decadal Oscillation (PDO) index measures the climatological situation of the Southern hemisphere: its extremes correspond to episodes of abnormal weather conditions.

Remark

Whiting et al. (2003) model the time series by an ARFIMA$(0, d, 0)$*. Data analysis and inference is conducted using* **annual data**, *from 1920 to 2022, so* $n = 122$, *relying on M-estimator (Whittle), which yields Wald-type statistic from first order asymptotic theory to test*

$$\mathcal{H}_0 : \ d = 0 \quad vs \quad \mathcal{H}_1 : \ d > 0.$$

Practical motivations for dependent data

Example (ARFIMA synthetic data)

Let $\{Y_t,\ t \in \mathbb{Z}\}$ be an ARFIMA(p, d, q), having dynamics

$$\theta(L)(1 - L)^d Y_t = \phi(L)\epsilon_t, \tag{3}$$

where $\forall t$, the $\{\epsilon_t\}$ are i.i.d. with zero mean and known $\sigma_\epsilon^2 = 1$.

Practical motivations for dependent data

Example (ARFIMA synthetic data)

Let $\{Y_t, \ t \in \mathbb{Z}\}$ be an ARFIMA(p, d, q), having dynamics

$$\theta(L)(1 - L)^d Y_t = \phi(L)\epsilon_t, \tag{3}$$

where $\forall t$, the $\{\epsilon_t\}$ are i.i.d. with zero mean and known $\sigma_\epsilon^2 = 1$.

- We focus on an ARFIMA of order $p = 2$, $d = 0$ and $q = 0$ process, with AR coefficients $\theta_1 = 0.1$, $\theta_2 = 0.2$ and Gaussian errors.

Practical motivations for dependent data

Example (ARFIMA synthetic data)

Let $\{Y_t,\ t \in \mathbb{Z}\}$ be an ARFIMA(p, d, q), having dynamics

$$\theta(L)(1 - L)^d Y_t = \phi(L)\epsilon_t, \tag{3}$$

where $\forall t$, the $\{\epsilon_t\}$ are i.i.d. with zero mean and known $\sigma_\epsilon^2 = 1$.

- We focus on an ARFIMA of order $p = 2$, $d = 0$ and $q = 0$ process, with AR coefficients $\theta_1 = 0.1$, $\theta_2 = 0.2$ and Gaussian errors.
- We consider different increasing values of the sample size $n = 250, 2500, 5000$.

Practical motivations for dependent data

Example (ARFIMA synthetic data)

Let $\{Y_t,\ t \in \mathbb{Z}\}$ be an ARFIMA(p, d, q), having dynamics

$$\theta(L)(1-L)^d Y_t = \phi(L)\epsilon_t, \tag{3}$$

where $\forall t$, the $\{\epsilon_t\}$ are i.i.d. with zero mean and known $\sigma^2_\epsilon = 1$.

- We focus on an ARFIMA of order $p = 2$, $d = 0$ and $q = 0$ process, with AR coefficients $\theta_1 = 0.1$, $\theta_2 = 0.2$ and Gaussian errors.
- We consider different increasing values of the sample size $n = 250, 2500, 5000$.
- We estimate θ via the routinely applied Whittle's M-estimator, as implemented in the routine `WhittleEst` available in the R package `longmemo`.

Practical motivations for dependent data

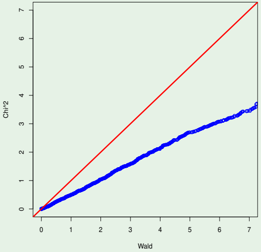
Example (cont'd)

The goal of our inference is to test

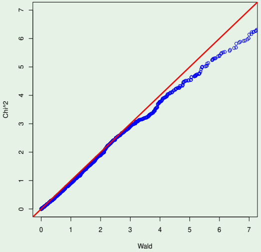
$$\mathcal{H}_0 : d = 0 \text{ vs. } \mathcal{H}_1 : d > 0,$$

and we resort on the Wald test statistic for Whittle's estimator, as available in the statistical software, comparing χ^2 quantiles to the true (as obtained by MC simulation).

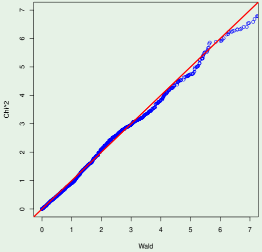
$n = 250$



$n = 2500$



$n = 5000$



Practical motivations for dependent data

Example (cont'd)

The goal of our inference is to test

$$\mathcal{H}_0 : d = 0 \text{ vs. } \mathcal{H}_1 : d > 0,$$

and we resort on the Wald test statistic for Whittle's estimator, as available in the statistical software, comparing χ^2 quantiles to the true (as obtained by MC simulation).

$n = 250$

$n = 2500$

$n = 5000$

Remark

As conjectured, the first order asymptotic theory suffers from size distortion. Any **saddlepoint techniques?**

Practical motivations for dependent data

Feldstein and Horioka (1980) investigated the following points:

Practical motivations for dependent data

Feldstein and Horioka (1980) investigated the following points:

- Does capital flow among industrial countries to equalise the yield to investors?

Practical motivations for dependent data

Feldstein and Horioka (1980) investigated the following points:

- Does capital flow among industrial countries to equalise the yield to investors?
- Alternatively, does the saving that originates in a country remain to be invested there?

Practical motivations for dependent data

Feldstein and Horioka (1980) investigated the following points:

- Does capital flow among industrial countries to equalise the yield to investors?
- Alternatively, does the saving that originates in a country remain to be invested there?
- Or does the truth lie somewhere between these two extremes?

Practical motivations for dependent data

Feldstein and Horioka (1980) investigated the following points:

- Does capital flow among industrial countries to equalise the yield to investors?
- Alternatively, does the saving that originates in a country remain to be invested there?
- Or does the truth lie somewhere between these two extremes?

Eventually, they found a large, positive (cor)relation of domestic saving rates on domestic investment rates for Organisation for Economic Co-operation and Development (OECD) countries. They interpret this finding as evidence of a high degree of frictions reducing capital flows between countries.

Practical motivations for dependent data

Feldstein and Horioka (1980) investigated the following points:

- Does capital flow among industrial countries to equalise the yield to investors?
- Alternatively, does the saving that originates in a country remain to be invested there?
- Or does the truth lie somewhere between these two extremes?

Eventually, they found a large, positive (cor)relation of domestic saving rates on domestic investment rates for Organisation for Economic Co-operation and Development (OECD) countries. They interpret this finding as evidence of a high degree of frictions reducing capital flows between countries. Re-considering FH study, Debarsy & Ertur (2010) analyzed Investment and Saving rates for 24 OECD countries between 1960 and 2000 (41 yrs):

$$\{\text{Inv}_{i,t}\} \qquad \text{and} \qquad \{\text{Sav}_{i,t}\}$$

for $i = 1, .., n$ (cross-sectional dimension, $n = 24$) and $t = 1, ..., T$ (time series dimension, $T = 41$).

Practical motivations for dependent data

Feldstein and Horioka (1980) investigated the following points:

- Does capital flow among industrial countries to equalise the yield to investors?
- Alternatively, does the saving that originates in a country remain to be invested there?
- Or does the truth lie somewhere between these two extremes?

Eventually, they found a large, positive (cor)relation of domestic saving rates on domestic investment rates for Organisation for Economic Co-operation and Development (OECD) countries. They interpret this finding as evidence of a high degree of frictions reducing capital flows between countries. Re-considering FH study, Debarsy & Ertur (2010) analyzed Investment and Saving rates for 24 OECD countries between 1960 and 2000 (41 yrs):

$$\boxed{\{\text{Inv}_{i,t}\} \qquad \text{and} \qquad \{\text{Sav}_{i,t}\}}$$

for $i = 1, .., n$ (cross-sectional dimension, $n = 24$) and $t = 1, ..., T$ (time series dimension, $T = 41$).

Practical motivations for dependent data

Feldstein and Horioka (1980) investigated the following points:

- Does capital flow among industrial countries to equalise the yield to investors?
- Alternatively, does the saving that originates in a country remain to be invested there?
- Or does the truth lie somewhere between these two extremes?

Eventually, they found a large, positive (cor)relation of domestic saving rates on domestic investment rates for Organisation for Economic Co-operation and Development (OECD) countries. They interpret this finding as evidence of a high degree of frictions reducing capital flows between countries. Re-considering FH study, Debarsy & Ertur (2010) analyzed Investment and Saving rates for 24 OECD countries between 1960 and 2000 (41 yrs):

$$\boxed{\underbrace{\{\text{Inv}_{i,t}\} \quad \overset{\forall i}{\Longleftrightarrow} \quad \{\text{Sav}_{i,t}\}}_{\text{as in FH}}}$$

Practical motivations for dependent data

Feldstein and Horioka (1980) investigated the following points:

- Does capital flow among industrial countries to equalise the yield to investors?
- Alternatively, does the saving that originates in a country remain to be invested there?
- Or does the truth lie somewhere between these two extremes?

Eventually, they found a large, positive (cor)relation of domestic saving rates on domestic investment rates for Organisation for Economic Co-operation and Development (OECD) countries. They interpret this finding as evidence of a high degree of frictions reducing capital flows between countries. Re-considering FH study, Debarsy & Ertur (2010) analyzed Investment and Saving rates for 24 OECD countries between 1960 and 2000 (41 yrs):

$$\boxed{\{\mathsf{Inv}_{i,t}\} \overset{?}{\Longleftrightarrow} \{\mathsf{Inv}_{k,t}\} \overset{?}{\Longleftrightarrow} \{\mathsf{Inv}_{j,t}\} \overset{?}{\Longleftrightarrow} \underbrace{\{\mathsf{Inv}_{i,t}\} \Longleftrightarrow \{\mathsf{Sav}_{i,t}\}}_{\text{as in FH}}}$$

Practical motivations for dependent data

Feldstein and Horioka (1980) investigated the following points:

- Does capital flow among industrial countries to equalise the yield to investors?
- Alternatively, does the saving that originates in a country remain to be invested there?
- Or does the truth lie somewhere between these two extremes?

Eventually, they found a large, positive (cor)relation of domestic saving rates on domestic investment rates for Organisation for Economic Co-operation and Development (OECD) countries. They interpret this finding as evidence of a high degree of frictions reducing capital flows between countries. Re-considering FH study, Debarsy & Ertur (2010) analyzed Investment and Saving rates for 24 OECD countries between 1960 and 2000 (41 yrs):

$$\boxed{\underbrace{\{\mathrm{Inv}_{i,t}\} \overset{?}{\Longleftrightarrow} \{\mathrm{Inv}_{k,t}\} \overset{?}{\Longleftrightarrow} \{\mathrm{Inv}_{j,t}\} \overset{?}{\Longleftrightarrow} \{\mathrm{Inv}_{i,t}\}}_{\text{spillover}} \Longleftrightarrow \{\mathrm{Sav}_{i,t}\}}$$

Practical motivations for dependent data

Aim

Test for the presence of spillover (spatial autocorrelation) between country i and country j, $i \neq j$, in the investment-saving relationship, e.g. using p-value and the quantiles of Wald-type statistics for SARMA, where the parameter λ controls the spatial dependence (spillover effect), thus:

$$\mathcal{H}_0 : \lambda = 0 \quad vs \quad \mathcal{H}_1 : \lambda > 0,$$

as in the common in the spatial econometrics literature.

Practical motivations for dependent data

Aim

Test for the presence of spillover (spatial autocorrelation) between country i and country j, $i \neq j$, in the investment-saving relationship, e.g. using p-value and the quantiles of Wald-type statistics for SARMA, where the parameter λ controls the spatial dependence (spillover effect), thus:

$$\mathcal{H}_0 : \ \lambda = 0 \quad vs \quad \mathcal{H}_1 : \ \lambda > 0,$$

as in the common in the spatial econometrics literature.

To achieve this aim, the extant approach resorts on **first order Gaussian asymptotic theory**; see Debarsy & Ertur (2010).

Practical motivations for dependent data

Aim

Test for the presence of spillover (spatial autocorrelation) between country i and country j, $i \neq j$, in the investment-saving relationship, e.g. using p-value and the quantiles of Wald-type statistics for SARMA, where the parameter λ controls the spatial dependence (spillover effect), thus:

$$\mathcal{H}_0 : \ \lambda = 0 \quad vs \quad \mathcal{H}_1 : \ \lambda > 0,$$

as in the common in the spatial econometrics literature.

To achieve this aim, the extant approach resorts on **first order Gaussian asymptotic theory**; see Debarsy & Ertur (2010).

Is the use of **first order asymptotics sensible** (small cross-sectional n and time T dimension)? Can we rely on analytical techniques, like the saddlepoint approximations?

Practical motivations for dependent data

Aim

Test for the presence of spillover (spatial autocorrelation) between country i and country j, $i \neq j$, in the investment-saving relationship, e.g. using p-value and the quantiles of Wald-type statistics for SARMA, where the parameter λ controls the spatial dependence (spillover effect), thus:

saddlepoint approximations.

Menu

- Literature: a bird's-eye view

Menu

- Literature: a bird's-eye view
- Time series
  - Some elements of spectral analysis
  - Setting: SRD & LRD

Menu

- Literature: a bird's-eye view
- Time series
  - Some elements of spectral analysis
  - Setting: SRD & LRD
  - Saddlepoint techniques

Menu

- Literature: a bird's-eye view
- Time series
    - Some elements of spectral analysis
    - Setting: SRD & LRD
    - Saddlepoint techniques
        - exponential-based (density approx and test in the presence of nuisance parameter)

Menu

- Literature: a bird's-eye view
- Time series
  - Some elements of spectral analysis
  - Setting: SRD & LRD
  - Saddlepoint techniques
    - exponential-based (density approx and test in the presence of nuisance parameter)
    - empirical version (density approx and test in the presence of nuisance parameter, connection to EL & ET)

Menu

- Literature: a bird's-eye view
- Time series
  - Some elements of spectral analysis
  - Setting: SRD & LRD
  - Saddlepoint techniques
    - exponential-based (density approx and test in the presence of nuisance parameter)
    - empirical version (density approx and test in the presence of nuisance parameter, connection to EL & ET)
  - Monte Carlo results

Menu

- Literature: a bird's-eye view
- Time series
  - Some elements of spectral analysis
  - Setting: SRD & LRD
  - Saddlepoint techniques
    - exponential-based (density approx and test in the presence of nuisance parameter)
    - empirical version (density approx and test in the presence of nuisance parameter, connection to EL & ET)
  - Monte Carlo results
- Conclusion: take home message

Literature: a bird's-eye view

(i) Most of the results on **saddlepoint techniques** are available for the **iid setting:** see Field & Ronchetti (1990), Jensen (1995), Kolassa (2006), Butler (2007), or Brazzale et al. (2007) for book-length presentation.

Literature: a bird's-eye view

(i) Most of the results on **saddlepoint techniques** are available for the **iid setting:** see Field & Ronchetti (1990), Jensen (1995), Kolassa (2006), Butler (2007), or Brazzale et al. (2007) for book-length presentation.

(ii) In time series, some results for **saddlepoint density and tail area approximations for $M-$estimators** in the noncircular **Gaussian AR(1) process**: see Daniels (Annals, 1956), Phillips (Biometrika, 1978), Cox & Solomon (Biometrika, 1988), Wang (Biometrika, 1992), Butler & Paolella (Bernoulli, 1998), Pereira et al. (Stats & Prob Letters, 2008), Lozada-Can & Davison (American Statistician, 2010),and Field & Robinson (Annals, 2013); for a book-length discussion see Taniguchi & Kakitzawa (2001).

Literature: a bird's-eye view

(i) Most of the results on **saddlepoint techniques** are available for the **iid setting:** see Field & Ronchetti (1990), Jensen (1995), Kolassa (2006), Butler (2007), or Brazzale et al. (2007) for book-length presentation.

(ii) In time series, some results for **saddlepoint density and tail area approximations for $M-$estimators** in the noncircular **Gaussian AR(1) process**: see Daniels (Annals, 1956), Phillips (Biometrika, 1978), Cox & Solomon (Biometrika, 1988), Wang (Biometrika, 1992), Butler & Paolella (Bernoulli, 1998), Pereira et al. (Stats & Prob Letters, 2008), Lozada-Can & Davison (American Statistician, 2010),and Field & Robinson (Annals, 2013); for a book-length discussion see Taniguchi & Kakitzawa (2001).

(iii) **Higher order techniques in frequency domain (spectral analysis) for time series** are available: see Taniguchi (JMA, 1987, Edgeworth for Whittle under SRD), Franke & Härdle (Annals, 1992, FDB), Dahlhaus & Janas (Annals, 1996, FDB), Andrews & Lieberman (Econometrica, 2005, Edgeworth for Whittle under LRD).

Some elements of spectral analysis

Let us start from a peculiar function of time series data: the autocovariance function

$$\gamma_Y(h) = cov(Y_{t+h}, Y_t) = E[(Y_{t+h} - \mu)(Y_t - \mu)]$$

for all h and with $E(Y_t) = \mu$, $\forall t$.

Some elements of spectral analysis

Let us start from a peculiar function of time series data: the autocovariance function

$$\gamma_Y(h) = cov(Y_{t+h}, Y_t) = E[(Y_{t+h} - \mu)(Y_t - \mu)]$$

for all h and with $E(Y_t) = \mu$, $\forall t$.
Under suitable assumptions, we have (for $i \in \mathbb{C}$)

$$\gamma_Y(h) = \int_{-1/2}^{1/2} \exp\{2\pi i \lambda h\} f(\lambda) d\lambda, \quad h = 0, \pm 1, \pm 2 \ldots$$

as the inverse Fourier transform of the spectral density $f(\cdot)$:

$$f(\lambda) = \sum_{h=-\infty}^{\infty} \gamma_Y(h) \exp\{-i 2\pi \lambda h\}, \quad -1/2 \leq \lambda \leq 1/2.$$

Some elements of spectral analysis

Definition

Given time series data $Y_1, \ldots, Y_n$, the discrete Fourier transform (DFT) is

$$d(\lambda_j) = n^{-1/2} \sum_{t=1}^{n} Y_t \exp\{-2\pi i \lambda_j t\},$$

for $j = 0, 1, \ldots, n-1$, where the frequencies $\lambda_j = j/n$ are called Fourier or fundamental frequencies. The periodogram at λ_j is $I(\lambda_j) = |d(\lambda_j)|^2$.

We have that

$$I(\lambda_j) = \sum_{h=-(n-1)}^{n-1} \hat{\gamma}_Y(h) \exp\{-2i\pi\lambda_j h\},$$

where $\hat{\gamma}_Y(h)$ is the empirical covariance and $\bar{Y}$ is the sample average.

Some elements of spectral analysis

Property 1. The periodogram is an asymptotically unbiased (nonparametric) estimator of the spectral density $f(\lambda)$. To reduce the finite sample bias, tapering and smoothing (essentially, averaging) are routinely applied.

Some elements of spectral analysis

Property 1. The periodogram is an asymptotically unbiased (nonparametric) estimator of the spectral density $f(\lambda)$. To reduce the finite sample bias, tapering and smoothing (essentially, averaging) are routinely applied.

Property 2. The periodogram ordinates are such that

$$I(\lambda) \xrightarrow{d} i.d.\ \xi f(\lambda), \quad \xi \sim \exp(1) \tag{4}$$

Remark

The asymptotic iid-ness of the standardized periodogram ordinates allows to transform problems for dependent data into problems for iid data.

Some elements of spectral analysis

Property 2 allows to derive a frequency domain likelihood and parameter estimation is obtained maximazing this likelihood.

Some elements of spectral analysis

Property 2 allows to derive a frequency domain likelihood and parameter estimation is obtained maximazing this likelihood.

This idea goes back to Whittle (1951): if there is a parametric model for $f(\lambda, \theta)$, then we may work on:

$$\mathcal{L}_W(\theta) = \frac{1}{2\pi}\left[\int_{-\pi}^{\pi} \ln f(\lambda,\theta) d\lambda + \int_{-\pi}^{\pi} \frac{I(\lambda)}{f(\lambda,\theta)} d\lambda\right], \quad (5)$$

which is obtained making use of Property 2 (λ is in radians, from now on).

Some elements of spectral analysis

Property 2 allows to derive a frequency domain likelihood and parameter estimation is obtained maximazing this likelihood.

This idea goes back to Whittle (1951): if there is a parametric model for $f(\lambda, \theta)$, then we may work on:

$$\mathcal{L}_W(\theta) = \frac{1}{2\pi}\left[\int_{-\pi}^{\pi} \ln f(\lambda, \theta) d\lambda + \int_{-\pi}^{\pi} \frac{I(\lambda)}{f(\lambda, \theta)} d\lambda\right], \tag{5}$$

which is obtained making use of Property 2 (λ is in radians, from now on).

The optimization of $L_W(\theta)$ (the Riemann-discretized version of $\mathcal{L}_W$):

$$\hat{\theta}_n = \arg\max_{\theta} L_W(\theta)$$

(or $\nabla_\theta L_W(\hat{\theta}_n) = 0$) defines an **M-estimator in the frequency domain**. Then,

$$\mathcal{V}_n = \sqrt{n}(\hat{\theta}_n - \theta^0)$$

and we want an approximation to its density $f_{\hat{\theta}_n}$.

Some elements of spectral analysis

Property 2 allows to derive a frequency domain likelihood and parameter estimation is
ob

Th
we

wh

Th

(or

$$\mathcal{V}_n = \sqrt{n}(\theta_n - \theta^0)$$

and we want an approximation to its density $f_{\hat{\theta}_n}$.

Indeed, for each $\lambda \in (-\pi, \pi]$, treating the periodogram ordinates as independent rvs, we have $I(\lambda) \sim \xi f(\lambda, \theta)$ and it has pdf

$$p(z, \theta) = \frac{1}{f(\lambda, \theta)} e^{-\frac{z}{f(\lambda,\theta)}}.$$

Thus, taking the log on both sides, we have

$$\ln p(z, \theta) = -\ln f(\lambda, \theta) - \frac{z}{f(\lambda, \theta)}.$$

The sum/integral of these quantities defines the (negative) log-likelihood.

Setting: SRD and LRD

Suppose that $\{Y_t\}$ is a linear and second order stationary process

$$Y_t = \sum_{r=0}^{\infty} a_r \varepsilon_{t-r},$$

for $t \in \mathbb{Z}$,

Setting: SRD and LRD

Suppose that $\{Y_t\}$ is a linear and second order stationary process

$$Y_t = \sum_{r=0}^{\infty} a_r \varepsilon_{t-r},$$

for $t \in \mathbb{Z}$, with spectral density function

$$f(\lambda, \theta) = |\lambda|^{-2d} L(\lambda, \vartheta), \quad \lambda \in \Pi = (-\pi, \pi] \tag{6}$$

where $d \in [0, 0.5)$, $\vartheta \in \mathbb{R}^p$ with $p \geq 1$ and $\theta = (d, \vartheta)$.

Setting: SRD and LRD

Suppose that $\{Y_t\}$ is a linear and second order stationary process

$$Y_t = \sum_{r=0}^{\infty} a_r \varepsilon_{t-r},$$

for $t \in \mathbb{Z}$, with spectral density function

$$f(\lambda, \theta) = |\lambda|^{-2d} L(\lambda, \vartheta), \quad \lambda \in \Pi = (-\pi, \pi] \tag{6}$$

where $d \in [0, 0.5)$, $\vartheta \in \mathbb{R}^p$ with $p \geq 1$ and $\theta = (d, \vartheta)$.

Definition

We classify the process $\{Y_t\}$ as short-range dependent (SRD) or long-range dependent (LRD)

- when $d = 0$ and the function $L(\cdot, \vartheta)$ is bounded with $L(0, \vartheta) \neq 0$, then the process $\{Y_t\}$ features SRD
- Otherwise, the process $\{Y_t\}$ features LRD—f has a pole at $\lambda = 0$.

Saddlepoint approximation (exponential-based)

First order asymptotic theory implies

$$\mathcal{V}_n \xrightarrow{\mathcal{D}} \mathcal{N}(0, V).$$

To have a better density approximation, we may derive the saddlepoint density approximation $g_{\hat{\theta}_n}$ treating the periodogram ordinates as independently and exponentially distributed r.v.'s: we use it to approximate the c.g.f. and its general Legendre transform.

Saddlepoint approximation (exponential-based)

First order asymptotic theory implies

$$\mathcal{V}_n \xrightarrow{\mathcal{D}} \mathcal{N}(0, V).$$

To have a better density approximation, we may derive the saddlepoint density approximation $g_{\hat{\theta}_n}$ treating the periodogram ordinates as independently and exponentially distributed r.v.'s: we use it to approximate the c.g.f. and its general Legendre transform.

Remark

The saddlepoint approximation can be **easily** *derived treating the periodogram ordinates $\{I(\lambda)\}$ as independent rvs, exponentially distributed. It features:*

- **SRD**: *relative of order $o(n^{-1/2})$*

Saddlepoint approximation (exponential-based)

First order asymptotic theory implies

$$\mathcal{V}_n \xrightarrow{\mathcal{D}} \mathcal{N}(0, V).$$

To have a better density approximation, we may derive the saddlepoint density approximation $g_{\hat{\theta}_n}$ treating the periodogram ordinates as independently and exponentially distributed r.v.'s: we use it to approximate the c.g.f. and its general Legendre transform.

Remark

The saddlepoint approximation can be **easily** *derived treating the periodogram ordinates $\{I(\lambda)\}$ as independent rvs, exponentially distributed. It features:*

- **SRD**: *relative of order $o(n^{-1/2})$*
- **LRD**: *relative error of order $O(n^{-1/2})$.*

Saddlepoint approximation (exponential-based)

Specifically:

- Whittle's estimating function is

$$\psi_j\left(I(\lambda_j), \theta\right) = \left(\frac{I(\lambda_j)}{f(\lambda_j, \theta)} - 1\right) \nabla_\theta \ln f(\lambda_j, \theta),$$

Saddlepoint approximation (exponential-based)

Specifically:

- Whittle's estimating function is

$$\psi_j\left(I(\lambda_j),\theta\right)=\left(\frac{I(\lambda_j)}{f(\lambda_j,\theta)}-1\right)\nabla_\theta \ln f(\lambda_j,\theta),$$

- for $m=\lfloor (n-1)/2\rfloor$, Whittle's M-estimator $\hat{\theta}_n$ is the solution to

$$\sum_{j=1}^{m}\psi_j(I(\lambda_j),\hat{\theta}_n)=0.$$

Saddlepoint approximation (exponential-based)

Specifically:

- Whittle's estimating function is

$$\psi_j\left(I(\lambda_j),\theta\right)=\left(\frac{I(\lambda_j)}{f(\lambda_j,\theta)}-1\right)\nabla_\theta \ln f(\lambda_j,\theta),$$

- for $m=\lfloor (n-1)/2\rfloor$, Whittle's M-estimator $\hat{\theta}_n$ is the solution to

$$\sum_{j=1}^{m}\psi_j(I(\lambda_j),\hat{\theta}_n)=0.$$

- define $\mathcal{K}^*_{\mathcal{V}_n}(\upsilon,s)=\sum_j K^*_{\psi_j}(\upsilon,s)$, where

$$K^*_{\psi_j}(\upsilon,s)=\ln\left(E^*\left[\exp\{\upsilon\psi_j(I(\lambda_j),s)\}\right]\right),$$

with E^* computed treating $I(\lambda_j)/f(\lambda_j,\theta^0)\sim\exp(1)$.

Saddlepoint approximation (exponential-based)

The saddlepoint density approximation is:

$$g_{\hat{\theta}_n}(s) = \left(\frac{n}{2\pi \boxed{\mathcal{K}^{*''}_{\mathcal{V}_n}}(\upsilon_0, s)} \right)^{1/2} e^{\boxed{\mathcal{K}^{*}_{\mathcal{V}_n}}(\upsilon_0, s)}, \tag{7}$$

and the saddlepoint $\upsilon_0 = \upsilon_0(s)$ solves

$$\boxed{\mathcal{K}^{*'}_{\mathcal{V}_n}}(\upsilon, s) = 0.$$

Remark

The advantage of using $I(\lambda)/f(\lambda, \theta) \sim \exp(1)$ *is that* $\boxed{\mathcal{K}^{*}_{\mathcal{V}_n}}$ *is strictly convex, thus the saddlepoint equation admits a unique solution—which can be computed using standard methods, like the one based on the secant.*

Saddlepoint approximation (exponential-based)

Example (ASY and frequency domain bootstrap (FDB))

Let us consider the AR(1):

$$Y_t = \theta_0 Y_{t-1} + \epsilon_t, \quad \epsilon_t \sim t_6 \quad \theta^0 = 0.4.$$

and the Whittle's estimator $\hat{\theta}_n$. Goal: approximate $P_{\theta^0}(\hat{\theta}_n > t_0)$.

Saddlepoint approximation (exponential-based)

Example (ASY and frequency domain bootstrap (FDB))

Let us consider the AR(1):

$$Y_t = \theta_0 Y_{t-1} + \epsilon_t, \quad \epsilon_t \sim t_6 \quad \theta^0 = 0.4.$$

and the Whittle's estimator $\hat{\theta}_n$. Goal: approximate $P_{\theta^0}(\hat{\theta}_n > t_0)$.

	12.5%	10%	5%	2.5%

Saddlepoint approximation (exponential-based)

Example (ASY and frequency domain bootstrap (FDB))

Let us consider the AR(1):

$$Y_t = \theta_0 Y_{t-1} + \epsilon_t, \quad \epsilon_t \sim t_6 \quad \theta^0 = 0.4.$$

and the Whittle's estimator $\hat{\theta}_n$. Goal: approximate $P_{\theta^0}(\hat{\theta}_n > t_0)$.

	12.5%	10%	5%	2.5%
		$n = 36$		
SAD	12.2%	9.1%	4.4%	2.0%
ASY	15.0%	11.8%	6.4%	3.2%
FDB	—	—	—	—

Saddlepoint approximation (exponential-based)

Example (ASY and frequency domain bootstrap (FDB))

Let us consider the AR(1):

$$Y_t = \theta_0 Y_{t-1} + \epsilon_t, \quad \epsilon_t \sim t_6 \quad \theta^0 = 0.4.$$

and the Whittle's estimator $\hat{\theta}_n$. Goal: approximate $P_{\theta^0}(\hat{\theta}_n > t_0)$.

	12.5%	10%	5%	2.5%
		$n = 36$		
SAD	12.2%	9.1%	4.4%	2.0%
ASY	15.0%	11.8%	6.4%	3.2%
FDB	—	—	—	—
		$n = 150$		
SAD	12.7%	9.9%	4.9%	2.3%
ASY	12.1%	9.2%	4.4%	2.0%
FDB	13.5%	10.8%	5.6%	2.9%
$(q_1; q_3)$	(10.5%; 15.7%)	(8.0%; 12.7%)	(4.0%; 6.6%)	(2.0%; 3.5%)

Saddlepoint approximation (exponential-based)

More generally, let $\theta = (\theta^{(1)}, \theta^{(2)})$, where $\theta^{(2)} \in \mathbb{R}^{p_2}, 1 < p_2 < p$ and consider testing

$$\mathcal{H}_0 : \theta^{(2)} = 0 \ \text{ vs } \ \mathcal{H}_1 : \theta^{(2)} > 0$$

with $\theta^{(1)}$ being the nuisance parameter.

Saddlepoint approximation (exponential-based)

More generally, let $\theta = (\theta^{(1)}, \theta^{(2)})$, where $\theta^{(2)} \in \mathbb{R}^{p_2}, 1 < p_2 < p$ and consider testing

$$\mathcal{H}_0 : \theta^{(2)} = 0 \quad \text{vs} \quad \mathcal{H}_1 : \theta^{(2)} > 0$$

with $\theta^{(1)}$ being the nuisance parameter. Two options:

Saddlepoint approximation (exponential-based)

More generally, let $\theta = (\theta^{(1)}, \theta^{(2)})$, where $\theta^{(2)} \in \mathbb{R}^{p_2}, 1 < p_2 < p$ and consider testing

$$\mathcal{H}_0 : \theta^{(2)} = 0 \quad \text{vs} \quad \mathcal{H}_1 : \theta^{(2)} > 0$$

with $\theta^{(1)}$ being the nuisance parameter. Two options:

- $g_{\hat{\theta}_n}$ is available: construct the test using analytical marginalization techniques

Saddlepoint approximation (exponential-based)

More generally, let $\theta = (\theta^{(1)}, \theta^{(2)})$, where $\theta^{(2)} \in \mathbb{R}^{p_2}, 1 < p_2 < p$ and consider testing

$$\mathcal{H}_0 : \theta^{(2)} = 0 \quad \text{vs} \quad \mathcal{H}_1 : \theta^{(2)} > 0$$

with $\theta^{(1)}$ being the nuisance parameter. Two options:

- $g_{\hat{\theta}_n}$ is available: construct the test using analytical marginalization techniques
- adapt the univariate saddlepoint test statistic of Robinson et al (2003. AoS):

$$S(\hat{\theta}_n^{(2)}) = 2 \inf_{\theta^{(1)}} \left[\sup_{\upsilon} \{ - \sum_j K_{\psi_j}(\upsilon; (\theta^{(1)}, \hat{\theta}_n^{(2)})) \} \right],$$

where υ solves the saddlepoint equation. The distribution of $S(\hat{\theta}_n^{(2)})$ under the null, can be approximated by a $\chi^2_{p_2}$ and it is asymptotically first order equivalent to the Wald test.

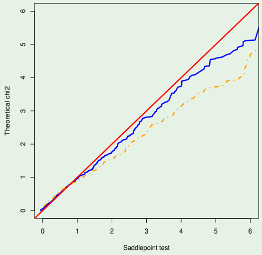
Saddlepoint approximation (exponential-based)

Example (Gaussian ARFIMA $(0, d, 0)$)

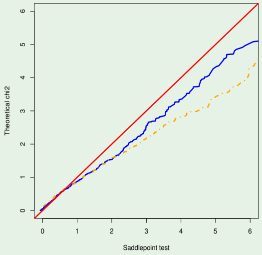
Testing about the long-memory (no nuisance, no need for the inf) for $n = 100, 250$:

$$\mathcal{H}_0 : d = d^0 \quad \text{vs} \quad \mathcal{H}_1 : d > d^0.$$

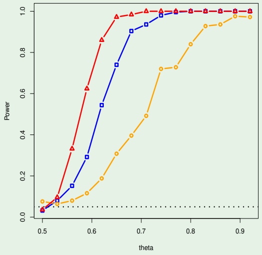
$d^0 = 0.1$



$d^0 = 0.35$



Power



Saddlepoint approximation (empirical version)

Remark

The c.g.f. may be approximated using the empirical distribution of the periodogram ordinates, keeping their independence but not relying on the exponential distribution.

Saddlepoint approximation (empirical version)

Remark

The c.g.f. may be approximated using the empirical distribution of the periodogram ordinates, keeping their independence but not relying on the exponential distribution.

- *Dahlhaus & Janas (1996. AoS) (FDB)*
- *Monti (1997, Biom.) (FDEL)*
- *Kakizawa (2013, JTSA) (FDGEL)*

Saddlepoint approximation (empirical version)

Cornell University

We gratefully acknowledge su[pport from the Simons Foundation and] member institu[tions]

arXiv > stat > arXiv:2403.12714

Search...
Help | Advanced Sea[rch]

Statistics > Methodology

[Submitted on 19 Mar 2024]

On the use of the cumulant generating function for inference on time series

Alban Moor, Davide La Vecchia, Elvezio Ronchetti

We introduce innovative inference procedures for analyzing time series data. Our methodology enables density approximation and composite hypothesis testing based on Whittle's estimator, a widely applied M–estimator in the frequency domain. Its core feature involves the (general Legendre transform of the) cumulant generating function of the Whittle likelihood score, as obtained using an approximated distribution of the periodogram ordinates. We present a testing algorithm that significantly expands the applicability of the state–of–the–art saddlepoint test, while maintaining the numerical accuracy of the saddlepoint approximation. Additionally, we demonstrate connections between our findings and three other prevalent frequency domain approaches: the bootstrap, empirical likelihood, and exponential tilting. Numerical examples using both simulated and real data illustrate the advantages and accuracy of our methodology.

Subjects: **Methodology (stat.ME)**; Computation (stat.CO)

Saddlepoint approximation (empirical version)

The empirical saddlepoint density approximation is

$$\hat{g}_{\hat{\theta}_n}(s) = \left(\frac{m}{2\pi}\right)^{p/2} \left|\det \hat{M}(s)\right| \left|\det \hat{\Sigma}(s)\right|^{-1/2} \exp\{m \boxed{\hat{K}(s)}\}, \tag{8}$$

where

$$\boxed{\hat{K}(s)} = \boxed{\hat{K}(\hat{\upsilon}, s)} = \ln\left[\frac{1}{m}\sum_{j=1}^{m} \exp\{\hat{\upsilon}^T \psi_j(I_j, s)\}\right], \tag{9}$$

$$\hat{M}(s) = \frac{1}{m}\exp\{-\hat{K}(s)\}\sum_{j=1}^{m} \nabla_w \psi_j(I_j, w)|_{w=s} \exp\{\hat{\upsilon}^T \psi_j(I_j, s)\},$$

$$\hat{\Sigma}(s) = \frac{1}{m}\exp\{-\hat{K}(s)\}\sum_{j=1}^{m} \psi_j(I_j, s)\psi_j(I_j, s)^T \exp\{\hat{\upsilon}^T \psi_j(I_j, s)\}$$

and the empirical saddlepoint $\hat{\upsilon}$ satisfies:

$$\sum_{j=1}^{m} \psi_j(I_j, s) \exp\{\hat{\upsilon}^T \psi_j(I_j, s)\} = 0. \tag{10}$$

Saddlepoint approximation (empirical version)

The empirical saddlepoint is based on the c.g.f. $\boxed{\hat{K}}$ as an approximation to the true c.g.f.: it is the key tool needed to compute $\hat{g}_{\hat{\theta}_n}$ and it unveils important connection with the FDEL.

Saddlepoint approximation (empirical version)

The empirical saddlepoint is based on the c.g.f. $\hat{K}$ as an approximation to the true c.g.f.: it is the key tool needed to compute $\hat{g}_{\hat{\theta}_n}$ and it unveils important connection with the FDEL.
Indeed, FDEL solves the system of (tilted) estimating equations

$$\sum_{j=1}^{m} \psi_j(I_j, s)[1 + \hat{\xi}^T \psi_j(I_j; s)]^{-1} = 0, \tag{11}$$

where we use the shorthand notation $\hat{\xi} = \hat{\xi}(s)$. Then, Monti defines a FD version of Owen's statistics as

$$\hat{W}(s) = 2 \sum_{j=1}^{m} \ln\{1 + \hat{\xi}^T \psi_j(I_j; s)\}$$

Saddlepoint approximation (empirical version)

Now notice that

- the saddlepoint satisfies (Taylor expansion of the exp) the equation

$$\sum_{j=1}^{m} \psi_j(I_j; s)[1 + \hat{\upsilon}^T \psi_j(I_j; s)] = O_P(n^{-1}),$$

since $\hat{\upsilon} = O_P(n^{-1/2})$.

Saddlepoint approximation (empirical version)

Now notice that

- the saddlepoint satisfies (Taylor expansion of the exp) the equation

$$\sum_{j=1}^{m} \psi_j(I_j; s)[1 + \hat{\upsilon}^T \psi_j(I_j; s)] = O_P(n^{-1}),$$

  since $\hat{\upsilon} = O_P(n^{-1/2})$.
- a Taylor expansion of the equation defining the FDEL yields

$$\sum_{j=1}^{m} \psi_j(I_j; s)[1 - \hat{\xi}^T \psi_j(I_j; s)] = O_P(n^{-1}),$$

  since $\hat{\xi} = O_P(n^{-1/2})$.

Saddlepoint approximation (empirical version)

Now notice that

- the saddlepoint satisfies (Taylor expansion of the exp) the equation

$$\sum_{j=1}^{m} \psi_j(I_j; s)[1 + \hat{\upsilon}^T \psi_j(I_j; s)] = O_P(n^{-1}),$$

  since $\hat{\upsilon} = O_P(n^{-1/2})$.

- a Taylor expansion of the equation defining the FDEL yields

$$\sum_{j=1}^{m} \psi_j(I_j; s)[1 - \hat{\xi}^T \psi_j(I_j; s)] = O_P(n^{-1}),$$

  since $\hat{\xi} = O_P(n^{-1/2})$.

Remark

The empirical saddlepoint and the empirical likelihood solve at the order $O_P(n^{-1})$ the same equation.

Saddlepoint approximation (empirical version)

Building on this remark, we prove that:

$$-2n\hat{K}(s) = 2\hat{W}(s) - \frac{2m^{-1/2}}{3}\sum_{j=1}^{m}\left\{u^T\hat{M}^T\hat{\Sigma}^{-1}\psi_j(I_j;\hat{\theta}_n)\right\}^3 + R_n$$

where, under some conditions, $R_n = O_P(n^{-1})$, $\hat{\Sigma} = \hat{\Sigma}(\hat{\theta}_n)$ and $\hat{M} = \hat{M}(\hat{\theta}_n)$.

Saddlepoint approximation (empirical version)

Building on this remark, we prove that:

$$-2n\hat{K}(s) = 2\hat{W}(s) - \frac{2m^{-1/2}}{3}\sum_{j=1}^{m}\left\{u^T\hat{M}^T\hat{\Sigma}^{-1}\psi_j(I_j;\hat{\theta}_n)\right\}^3 + R_n$$

where, under some conditions, $R_n = O_P(n^{-1})$, $\hat{\Sigma} = \hat{\Sigma}(\hat{\theta}_n)$ and $\hat{M} = \hat{M}(\hat{\theta}_n)$.

Remark

The latter result has a threefold importance:

(i) *it connects our FDES to the FDEL*

Saddlepoint approximation (empirical version)

Building on this remark, we prove that:

$$-2n\hat{K}(s) = 2\hat{W}(s) - \frac{2m^{-1/2}}{3}\sum_{j=1}^{m}\left\{u^T\hat{M}^T\hat{\Sigma}^{-1}\psi_j(I_j;\hat{\theta}_n)\right\}^3 + R_n$$

where, under some conditions, $R_n = O_P(n^{-1})$, $\hat{\Sigma} = \hat{\Sigma}(\hat{\theta}_n)$ and $\hat{M} = \hat{M}(\hat{\theta}_n)$.

Remark

The latter result has a threefold importance:

(i) *it connects our FDES to the FDEL*

(ii) *it illustrates that the difference between* $\hat{K}$ *and* $\hat{W}$ *depends on the third moment of the Whittle's score: both correct the Wald statistic for the skewness but in a different way*

Saddlepoint approximation (empirical version)

Building on this remark, we prove that:

$$-2n\,\hat{K}(s) = 2\hat{W}(s) - \frac{2m^{-1/2}}{3}\sum_{j=1}^{m}\left\{u^T\hat{M}^T\hat{\Sigma}^{-1}\psi_j(I_j;\hat{\theta}_n)\right\}^3 + R_n$$

where, under some conditions, $R_n = O_P(n^{-1})$, $\hat{\Sigma} = \hat{\Sigma}(\hat{\theta}_n)$ and $\hat{M} = \hat{M}(\hat{\theta}_n)$.

Remark

The latter result has a threefold importance:

(i) *it connects our FDES to the FDEL*

(ii) *it illustrates that the difference between $\hat{K}$ and $\hat{W}$ depends on the third moment of the Whittle's score: both correct the Wald statistic for the skewness but in a different way*

(iii) *it yields a nonparametric approximation of the density of Whittle's estimator based on the FDEL*

Saddlepoint approximation (empirical version)

On the practical side: use the empirical saddlepoint under $\mathcal{H}_0$ to approximate the distribution of Wald-type (or EL, ET) test statistics, where

$$\mathcal{H}_0 : \theta = \theta^0 \text{ vs. } \mathcal{H}_1 : \theta \neq \theta^0.$$

Saddlepoint approximation (empirical version)

On the practical side: use the empirical saddlepoint under $\mathcal{H}_0$ to approximate the distribution of Wald-type (or EL, ET) test statistics, where

$$\mathcal{H}_0 : \theta = \theta^0 \text{ vs. } \mathcal{H}_1 : \theta \neq \theta^0.$$

To this end,

- We define the Wald-type statistic, with $\hat{V} = \hat{M}^{-1}\hat{\Sigma}\hat{M}^{-1}$ (estimate of asym var of Whittle estim.),
$$\tilde{W}_n(\theta) = n(\hat{\theta}_n - \theta)^T \hat{V}^{-1}(\hat{\theta}_n - \theta).$$
Typically, the distribution of $\tilde{W}_n$ is approximated by a χ^2.

Saddlepoint approximation (empirical version)

On the practical side: use the empirical saddlepoint under $\mathcal{H}_0$ to approximate the distribution of Wald-type (or EL, ET) test statistics, where

$$\mathcal{H}_0 : \theta = \theta^0 \text{ vs. } \mathcal{H}_1 : \theta \neq \theta^0.$$

To this end,

- We define the Wald-type statistic, with $\hat{V} = \hat{M}^{-1}\hat{\Sigma}\hat{M}^{-1}$ (estimate of asym var of Whittle estim.),
$$\tilde{W}_n(\theta) = n(\hat{\theta}_n - \theta)^T \hat{V}^{-1}(\hat{\theta}_n - \theta).$$
Typically, the distribution of $\tilde{W}_n$ is approximated by a χ^2.
- In contrast, we make use of $\hat{g}_{\hat{\theta}_n}$ to obtain
$$P[\tilde{W}_n(\theta^0) > \tilde{w}(\theta^0) \mid \mathcal{H}_0] \approx 1 - \int_{\mathcal{B}} \hat{g}_{\hat{\theta}_n}(\theta) d\theta, \tag{12}$$
where $\tilde{w}(\theta^0)$ is the observed value of the test statistic and
$$\mathcal{B} = \left\{ \theta \in \mathbb{R}^d \mid \tilde{W}_n(\theta) \geq \tilde{w}(\theta^0) \right\}.$$
- To compute the integral in (12), we suggest to use an importance sampling scheme based on an instrumental Gaussian distribution.

Saddlepoint approximation (empirical version)

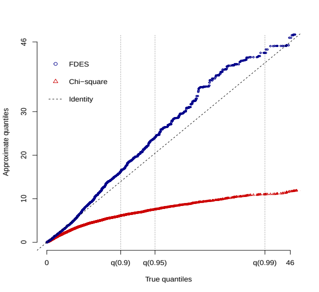
Example

We consider an ARFIMA(1,d,1) with $\theta^0 = (0.5, 0.25, 0.5)$ and test

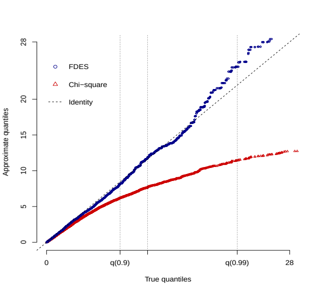
$$\mathcal{H}_0 : \theta = \theta^0 \text{ vs. } \mathcal{H}_1 : \theta \neq \theta^0$$

using the empirical saddlepoint. We compare the approx quantiles to true quantiles (as obtained by MC simulations), for the saddlepoint technique and first-order asymptotic theory (χ^2_3).

$n = 100$



$n = 500$



Saddlepoint approximation (empirical version)

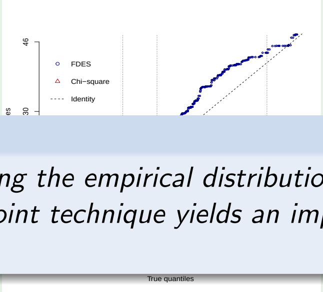
Example

We consider an ARFIMA(1,d,1) with $\theta^0 = (0.5, 0.25, 0.5)$ and test

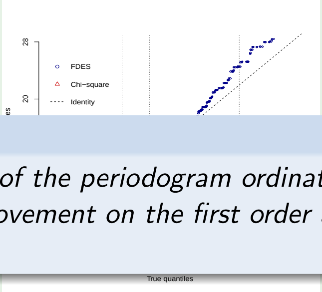
$$\mathcal{H}_0 : \theta = \theta^0 \text{ vs. } \mathcal{H}_1 : \theta \neq \theta^0$$

using the empirical saddlepoint. We compare the approx quantiles to true quantiles (as obtained by MC simulations), for the saddlepoint technique and first-order asymptotic theory (χ^2_3).

$n = 100$



$n = 500$



Remark

Also using the empirical distribution of the periodogram ordinates, the saddlepoint technique yields an improvement on the first order asymptotic theory.

Take home message

- First-order asymptotics and Edgeworth expansions may deliver poor inference in the setting of dependent data 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 data 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.

Thank you

For questions: `davide.lavecchia@unige.ch`

Laplace in brief

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

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

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:

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

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

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

Laplace in brief

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

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

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

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

Laplace in brief

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

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

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

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

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

Laplace in brief

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

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

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

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

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