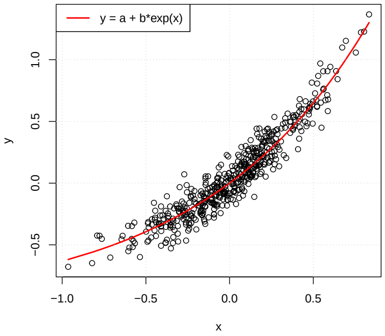
Метрични свойства на условни моменти с експоненциални тегла при проверка за адекватност на модел

Стоян В. Стоянов

Charles Schwab
stoyan.stoyanov@schwab.com

27 февруари 2023

Пример



- При дадени наблюдения (y_i, x_i) оценяваме модел $y_i = a + b\exp(x_i) + \epsilon_i$. Оценките на неизвестните параметри са $\hat{a} = -0.016$ и $\hat{b} = 1.017$.
- Описва ли функцията $f(x) = \hat{a} + \hat{b}\exp(x)$ адекватно условната средна $\mathbb{E}[y|x]$?

Основни резултати

1. Предлага се нова формулировка на тест за адекватност на модел с интегрирани условни моменти, който има асимптотични свойства сходни на сродни тестове (Bierens & Ploberger, 1997 и Bierens, 2016), но предлага по-лесна апроксимация на асимптотичното разпределение при нулевата хипотеза.
2. Функционалът зад статистиката има метрични свойства, които се използват за аналитична оценка на скоростта на сходимост към линеен модел, когато данните се агрегират във времето при хипотеза за независимост и еднакво разпределение. Числени примери за такова поведение са представени в Granger & Lee (1999) с помощта на Монте Карло симулации.
3. Текстът на статията може да се намери в SSRN: Stoyanov, S. (2023) "Metric Properties of Conditional Moments with Exponential Weights for Model Adequacy."

Съдържание

Основна хипотеза

- Наблюдаваме случаен вектор $(y_t, \mathbf{x}_t) \in \mathbb{R} \times \mathbb{R}^d$, $t \in \mathbb{N}$. За условната средна винаги можем да напишем $\mathbb{E}(y_t|\mathbf{x}_t) = h(\mathbf{x}_t)$, където $h(x): \mathbb{R}^d \to \mathbb{R}$ е неизвестна за нас (измерима) функция. Следователно е в сила тривиалното представяне,
$$y_t = h(\mathbf{x}_t) + \xi_t,$$
където $\mathbb{E}(\xi_t|\mathbf{x}_t) \stackrel{a.s.}{=} 0$.
- На практика работим с параметрична хипотеза,
$$y_t = m(\mathbf{x}_t, \theta) + \varepsilon_t,$$
където $\theta \in \Theta \subset \mathbb{R}^k$ е вектор от параметри и $m(x, z): \mathbb{R}^d \times \mathbb{R}^k \to \mathbb{R}$.
- При дадена извадка искаме да тестваме дали за някое θ_0, $m(x, \theta_0) = h(x)$ в някакъв вероятностен смисъл, или дали $\mathbb{E}(\varepsilon_t|\mathbf{x}_t) = 0$. Статистическата хипотеза е следната
$$H_0: \ \Pr[\mathbb{E}(y_t|\mathbf{x}_t) = m(\mathbf{x}_t, \theta)] = 1, \text{ за някое } \theta \in \Theta$$
$$H_1: \ \Pr[\mathbb{E}(y_t|\mathbf{x}_t) = m(\mathbf{x}_t, \theta)] < 1, \text{ за всички } \theta \in \Theta.$$

Главната идея зад голям клас от тестове

- Лесно се показва, че ако $\mathbb{E}(\varepsilon_t|\mathbf{x}_t) = 0$, то тогава $\mathbb{E}[\varepsilon_t f(\mathbf{x}_t)] = 0$ за всяка измерима f:

$$\mathbb{E}[\varepsilon_t f(\mathbf{x}_t)] = \mathbb{E}[\mathbb{E}[\varepsilon_t f(\mathbf{x}_t)|\mathbf{x}_t]] = \mathbb{E}[f(\mathbf{x}_t)\mathbb{E}[\varepsilon_t|\mathbf{x}_t]] = 0$$

- Статистическият тест се основава на твърдение в обратната посока. Може да се намери клас от "теглови" функции $\mathscr{F}$ такъв, че ако $\mathbb{E}[\varepsilon_t f(\mathbf{x}_t)] = 0$ за всяко $f \in \mathscr{F}$, то тогава $\mathbb{E}(\varepsilon_t|\mathbf{x}_t) = 0$.

Формулировка на тест

- Ако разгледаме

$$\mathbf{a}(\theta) = \int_{\mathbb{R}^d} \Big|\mathbb{E}\epsilon_t \underbrace{e^{i\langle u,\mathbf{x}_t\rangle}}_{f_u(x) = e^{i\langle u,x\rangle}} \Big|^2 \underbrace{g(u)}_{g(u)>0,\ g(u)=g(-u)} \, du,$$

къдетo $\langle x, y\rangle = \sum_i x_i y_i$, тогава може да се покаже, че $\mathbf{a}(\theta) = 0 \Leftrightarrow \mathbb{E}[\epsilon_t|\mathbf{x}_t] = 0$.

- Този функционал не е разглеждан в литературата.
- При дадена извадка от n наблюдения, статистиката приема вида

$$\hat{\mathbf{a}}_n(\hat{\theta}) = \int_{\mathbb{R}^d} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} [y_t - m(\mathbf{x}_t, \underbrace{\hat{\theta}}_{\text{напр. по метода на най-малките квадрати}})] e^{i\langle u,\mathbf{x}_t\rangle} \right|^2 g(u)du$$

$$= \frac{1}{n} \sum_{i,j} \hat{\epsilon}_i \hat{\epsilon}_j \underbrace{\varphi}_{\varphi(u) = \int_{\mathbb{R}^d} e^{i\langle u,x\rangle} g(x)dx} (\mathbf{x}_i - \mathbf{x}_j),$$

Асимптотично разпределение

Теорема 1

Асимптотичното разпределение на $\hat{\mathbf{a}}_n$ *при* H_0 *е*

$$\hat{\mathbf{a}}_n \xrightarrow{d} Q = \sum_{i=1}^{\infty} \kappa_i U_i^2,$$

където $U_i \in \mathcal{N}(0,1)$ *са независими,* $\kappa_1, \kappa_2, \ldots$ *са ненулевите собствени стойности на матрицата* $\{c_{i,j}\}_{i,j=1}^{\infty}$ *подредени в намаляващ ред, където* $c_{i,j} = \sqrt{\lambda_i \lambda_j}\sigma_{i,j}$ *и*

$$\sigma_{i,j} = \sigma_\epsilon^2 \operatorname{cov}\left(\psi_i(\mathbf{x}_1) - \boldsymbol{\mu}_i^{\mathsf{T}} \mathbf{A}^{-1} \dot{\mathbf{m}}(\mathbf{x}_1, \theta_0), \psi_j(\mathbf{x}_1) - \boldsymbol{\mu}_j^{\mathsf{T}} \mathbf{A}^{-1} \dot{\mathbf{m}}(\mathbf{x}_1, \theta_0)\right),$$

където $\boldsymbol{\mu}_i = \mathbb{E}[\dot{\mathbf{m}}(\mathbf{x}_1, \theta_0)\psi_i(\mathbf{x}_1)]$, $\lambda_1, \lambda_2, \ldots$ *и* $\{\psi_i\}_{i=0}^{\infty}$ *са собствените стойности и собствените функции в представянето* $\varphi(x-y) = \sum_{\ell=0}^{\infty} \lambda_\ell \psi_\ell(x)\psi_\ell(y)$ *и* $\varphi(u) = \int_{\mathbb{R}^d} e^{iux} g(x) dx$.

- Предимството е ясната връзката между κ_i и g. При подходящ избор на g, κ_i може да се пресметнат.

Апроксимация

Предложение 1

Означаваме частичната сума с $Q_s = \sum_{i=1}^{s} \kappa_i U_i^2$. *Хар. функция на* Q_s *се задава с*

$$\phi_{Q_s}(v) = \prod_{j=1}^{s} (1 - 2i\kappa_j v)^{-1/2}$$

и Q_s *клони по разпределение към* Q *когато* $s \to \infty$.

- Плътността на Q_s може да се пресметне с бързото преобразуване на Фурие.

Експоненциално ядро

Ако изберем многомерно нормално разпределение $g(u) = \frac{1}{(2\pi)^{d/2}} \exp(-\frac{1}{2}\langle u, u\rangle)$, тогава хар. функция е $\varphi(x) = \exp(-\frac{1}{2}\|x\|_2^2)$.

Теорема 2

В сила е следното представяне

$$\exp\left(-\frac{\gamma\|x-y\|_2^2}{2(1-\gamma^2)}\right) = \sum_{|\nu|=0}^{\infty} \gamma^{|\nu|}\psi_\nu(x)\psi_\nu(y),$$

където $\gamma \in (0,1)$, $\psi_\nu(x) = (1-\gamma^2)^{\frac{d}{4}} \frac{H_\nu(x)}{\sqrt{\nu!}} e^{-\frac{1}{2}\frac{\gamma}{1+\gamma}\|x\|_2^2}$ *и*

$$H_\nu(x) = e^{\frac{1}{2}\|x\|_2^2}(\partial^{|\nu|} e^{-\frac{1}{2}\|x\|_2^2}) = \prod_{k=1}^{d} \mathrm{He}_{\nu_k}(x_k),$$

където $\mathrm{He}_m(x_i)$ *означава полином на Ермит от ред* m*. Функциите* $\psi_\nu(x)$ *са ортонормални по отношение на мярката* $d\zeta(x) = \frac{1}{(2\pi)^{d/2}} \exp\left(-\frac{1-\gamma}{2(1+\gamma)}\langle x, x\rangle\right) dx$*.*

Съдържание

Проверка за конкретни модели със симулации

Дефиниции на модели	
Autoregressive (AR)	$Y_t = 0.6Y_{t-1} + \epsilon_t$
Bilinear (BL)	$Y_t = 0.7Y_{t-1}\epsilon_{t-2} + \epsilon_t$
Threshold Autoregressive (TAR)	$Y_t = 0.9Y_{t-1}I_{\lvert Y_{t-1}\rvert \leq 1} + 0.3Y_{t-1}I_{\lvert Y_{t-1}\rvert > 1} + \epsilon_t$
Sign Autoregressive (SGN)	$Y_t = \mathrm{sgn}(Y_{t-1}) + \epsilon_t$
Non-Linear Autoregressive (NAR)	$Y_t = 0.7\lvert Y_{t-1}\rvert/(\lvert Y_{t-1}\rvert + 2) + \epsilon_t$
Square bivariate (SQ)	$Y_t = X_t^2 + \epsilon_t,\ X_t = 0.6X_{t-1} + e_t$
Exponential bivariate (EXP)	$Y_t = \exp(X_t) + \epsilon_t,\ X_t = 0.6X_{t-1} + e_t$

Таблица: Грешките са независими и $\epsilon_t, e_t \in N(0,1)$. AR има линейно условно очакване (в червено). H_0 е линейна. Независимата променлива се избира $X_t = Y_{t-1}$.

Сравнение с други тестове

Модели от Блок 1, честота на отхвърляне (в %)							
	AR	BL	TAR	SGN	NAR	SQ	EXP
Neural	5.2	58	79.8	98.9	20.4	100	100
White	5.5	99.5	4.2	12.1	5.5	79.5	60.0
Reset	5.2	40.2	7.0	32.7	26.2	95.4	59.5
Tsay	4.7	39.3	4.2	15.0	23.8	100	100
$\hat{\mathbf{a}}_n$	**5**	**42.7**	**98**	**99.5**	**22.5**	**100**	**100**

Таблица: Процент на отхърляне на линейната хипотеза за моделите от Блок 1 със $n = 200$ и 1000 повторения при ниво на доверие от 95% (Lee et al, 1993). Моделите с линейна условна средна са в червено.

- Честотата на отхвърляне при AR е близо до очкваната стойност от 5%.
- Новият тест се представя добре съвместно при TAR, SGN, SQ, и SQP.
- Представянето при BL е по-слабо.

Съдържание

Нови означения

- Предполагаме, че $(\Delta y_k, \Delta \mathbf{x}_k)$ са независими и еднакво разпределени като сл. в. $\mathbf{v}_{\Delta t} = (y_{\Delta t}, \mathbf{x}_{\Delta t})$, който е нарастване на процес на Леви върху интервал с дължина Δt и $\mathbb{E} v_{\Delta t} = 0$, $\mathbb{E}\, \|v_{\Delta t}\|_2^2 < \infty$.
- Разглеждаме

нормираме, за да може да се търси граница по n

агрегиране на сл. в.

$$\widetilde{\mathbf{v}}_{n\Delta t} = \frac{1}{\sqrt{n}} \left(\sum_{k=1}^{n} \Delta y_k \, , \, \sum_{k=1}^{n} \Delta \mathbf{x}_k \right) = \frac{1}{\sqrt{n}} (y_{n\Delta t}, \mathbf{x}_{n\Delta t}).$$

- Използваме следния функционал

$$\mathbf{m}(\mathbf{v}_{\Delta t}) = \inf_{\theta} \sup_{u \in \mathbb{R}^d} \left| \mathbb{E}(y_{\Delta t} - \mathbf{x}_{\Delta t}^{\mathsf{T}} \theta) e^{i\langle u, \mathbf{x}_{\Delta t} \rangle} \right| = \inf_{\theta} \mathbf{m}^{\theta}(\mathbf{v}_{\Delta t}). \tag{1}$$

- В сила е: $\mathbf{m}(\mathbf{v}_{\Delta t}) = 0 \Leftrightarrow \mathbb{E}[y_{\Delta t} | \mathbf{x}_{\Delta t}] = \mathbf{x}_{\Delta t}^{\mathsf{T}} \theta$ за някое θ.

Факторизация на средната стойност

- Да означим хар. функция на $\mathbf{x}_{\Delta t}$ с $\phi_{\mathbf{x}_{\Delta t}}(u) = \mathbb{E}e^{i\langle u, \mathbf{x}_{\Delta t}\rangle}$, оценката по метода на най-малките квадрати с $\theta^* = (\mathbb{E}[\mathbf{x}_{\Delta t}\mathbf{x}_{\Delta t}^{\mathsf{T}}])^{-1}\mathbb{E}[y_{\Delta t}\mathbf{x}_{\Delta t}]$ и очакването в (1) с $h_n(u) = n^{-1/2}\mathbb{E}[\epsilon_{n\Delta t}e^{i\langle u, n^{-1/2}\mathbf{x}_{n\Delta t}\rangle}]$. Функцията $h_n(u)$ може да се представи:

$$h_n(u) = \overbrace{\sqrt{n}h_1\left(\frac{u}{\sqrt{n}}\right)}^{\text{определя асимптотиката на } h_n(u)} \overbrace{\left[\phi_{\mathbf{x}_{\Delta t}}\left(\frac{u}{\sqrt{n}}\right)\right]^{n-1}}^{\text{границата е } e^{-\frac{1}{2}u'\mathbf{C}u}},$$

  Ако развием първата функция в ред, получаваме:

$$\sqrt{n}h_1\left(\frac{u}{\sqrt{n}}\right) = \underbrace{\mathbb{E}[\epsilon_{\Delta t}]}_{=0} + i\langle \underbrace{\mathbb{E}[\epsilon_{\Delta t}\mathbf{x}_{\Delta t}]}_{=0,\text{ заради избора на } \theta^*}, u\rangle + o\left(n^{-1/2}\|u\|_2\right) = o\left(n^{-1/2}\|u\|_2\right)$$

- За произволна теглова функция $f_u(x)$ горната факторизация е възможна единствено и само ако $f_u(x) = \exp(i\langle u, x\rangle)$.

Скорост на сходимост

Теорема 3

Да предположим, че $\mathbf{v}_{\Delta t} = (y_{\Delta t}, \mathbf{x}_{\Delta t})$ *е нарастване на процес на Леви с* $\mathbb{E}|y_{\Delta t} - \mathbf{x}_{\Delta t}^{\mathsf{T}}\theta^*| \, \|\mathbf{x}_{\Delta t}\|_1^{k+\delta} < \infty$, *където* $k \geq 1$, $0 < \delta \leq 1$ *и* $\theta^* = (\mathbb{E}[\mathbf{x}_{\Delta t}\mathbf{x}_{\Delta t}^{\mathsf{T}}])^{-1}\mathbb{E}[y_{\Delta t}\mathbf{x}_{\Delta t}]$ *съществува. Допускаме, че* $\mathbf{x}_{\Delta t}$ *има достатъчно гладка плътност така, че* $\sup_{u \in \mathbb{R}^d} \|u\|_1^{k+\delta} |\phi_{\mathbf{x}_{\Delta t}}(u)| = B_{k+\delta} < \infty$ *и следните моментни условия са в сила*

$$\mathbb{E}[(y_{\Delta t} - \mathbf{x}_{\Delta t}^{\mathsf{T}}\theta^*)\mathbf{x}_{\Delta t}^{\alpha}] = 0, \qquad \forall \alpha : \, |\alpha| \leq k,$$

където α *е мултииндекс,* $|\alpha| = \sum_{i=1}^{d} \alpha_i$. *При тези предположения,*

$$\mathbf{m}(\widetilde{\mathbf{v}}_{n\Delta t}) \leq \mathbf{m}^{\theta^*}(\widetilde{\mathbf{v}}_{n\Delta t}) \leq C_{k,\delta} n^{-\frac{k+\delta-1}{2}},$$

където за константата $C_{k,\delta}$ *е в сила*

$$C_{k,\delta} = A_{\delta} B_{k+\delta} \frac{\Gamma(\delta+1)}{\Gamma(k+\delta+1)} \mathbb{E}\left(|y_{\Delta t} - \mathbf{x}_{\Delta t}^{\mathsf{T}}\theta^*| \, \|\mathbf{x}_{\Delta t}\|_1^{k+\delta}\right),$$

където $A_{\delta} = \sup_{y \in \mathbb{R}} y^{-\delta}|e^{iy} - 1|$ *и* $\Gamma(x) = \int_0^{\infty} u^{x-1}e^{-u}du$.

Каква е най-ниската скорост на сходимост?

- Да разгледаме случая $k = 1$. Условието $\mathbb{E}[(y_{\Delta t} - \mathbf{x}_{\Delta t}^{\intercal}\theta^{*})\mathbf{x}_{\Delta t}] = 0$ е гарантирано от избора на θ^{*} (това са условията от първи ред в задачата за най-малките квадрати). Получаваме оценката

$$\mathbf{m}(\widetilde{\mathbf{v}}_{n\Delta t}) \leq C_{1,\delta} n^{-\frac{\delta}{2}}, \qquad \delta \in (0,1]$$

- От неравенството

$$\mathbb{E}\left(|y_{\Delta t} - \mathbf{x}_{\Delta t}^{\intercal}\theta^{*}| \, \|\mathbf{x}_{\Delta t}\|_{1}^{k+\delta}\right) \leq \max(1, \|\theta^{*}\|_{\infty})\mathbb{E}\, \|\mathbf{v}_{\Delta t}\|_{1}^{k+1+\delta}$$

се вижда, че $\mathbb{E}\, \|\mathbf{v}_{\Delta t}\|_{1}^{2+\delta} < \infty$ е достатъчно условие за да съществува лявата страна.

Кога имаме по-висока скорост на сходимост?

- Да разгледаме случая $k = 2$. Критичните условия са $\mathbb{E}[(y_{\Delta t} - \mathbf{x}_{\Delta t}^{\intercal}\theta^*)\mathbf{x}_{\Delta t}^{\alpha}] = 0$, $|\alpha| = 2$. Записани по-просто, новите условия са $\mathbb{E}[(y_{\Delta t} - \mathbf{x}_{\Delta t}^{\intercal}\theta^*)\mathbf{x}_{i,\Delta t}\mathbf{x}_{j,\Delta t}] = 0$, $\forall i, j$.
- Можем да намерим иконометрична интерпретация. Това са условията от първи ред в следната задача

$$(\theta_0, \beta_0) = \arg\min_{(\theta,\beta)} \mathbb{E}\left(y_{\Delta t} - \sum_{k=1}^{d} \theta_k \mathbf{x}_{k,\Delta t} - \sum_{l,m} \beta_{l,m}\mathbf{x}_{l,\Delta t}\mathbf{x}_{m,\Delta t}\right)^2$$

  където оптималното решение е $(\theta_0, \beta_0) = (\theta^*, 0)$; т.е. промеливите $\mathbf{x}_{i,\Delta t}\mathbf{x}_{j,\Delta t}$ се оказват излишни, но това не означава, че условната средна е линейна!
- В този случай, ако $\mathbb{E}\|\mathbf{v}_t\|_1^{3+\delta} < \infty$ скоростта на сходимост е $\mathbf{m}(\widetilde{\mathbf{v}}_{n\Delta t}) \leq C_{2,\delta} n^{-\frac{1+\delta}{2}}$.

Какво се случва ако условията са изпълнени за всяко $k \geq 2$?

- Нека $\mathbb{E}[(y_{\Delta t} - \mathbf{x}_{\Delta t}^{\mathsf{T}}\theta^*)\mathbf{x}_{\Delta t}^{\alpha}] = 0$ за всеки избор на α. За всяка аналитична функция $h: \mathbb{R}^d \to \mathbb{C}$, за която съответните моменти съществуват е в сила

$$\begin{aligned}\mathbb{E}[(y_{\Delta t} - \mathbf{x}_{\Delta t}^{\mathsf{T}}\theta^*)h(X_{\Delta t})] &= \mathbb{E}\left[(y_{\Delta t} - \mathbf{x}_{\Delta t}^{\mathsf{T}}\theta^*)\sum_{|\alpha|=0}^{\infty}\frac{\partial^{\alpha}h(0)}{\alpha!}\mathbf{x}_{\Delta t}^{\alpha}\right] \\ &= \sum_{|\alpha|=0}^{\infty}\frac{\partial^{\alpha}h(0)}{\alpha!}\mathbb{E}\left[(y_{\Delta t} - \mathbf{x}_{\Delta t}^{\mathsf{T}}\theta^*)\mathbf{x}_{\Delta t}^{\alpha}\right] \\ &= 0.\end{aligned}$$

- Функциите $\{h_u(x) = \exp(i\langle u, x\rangle)\}$ са от този клас и следователно

$$\mathbb{E}[(y_{\Delta t} - \mathbf{x}_{\Delta t}^{\mathsf{T}}\theta^*)e^{i\langle u, \mathbf{x}_{\Delta t}\rangle}] = 0, \quad \forall u \in \mathbb{R}^d$$

от където следва, че $\mathbf{m}(\widetilde{\mathbf{v}}_{n\Delta t}) = 0$ за $\forall n$.

Съдържание

Обобщение

- Разглеждаме нов тест за адекватност на нелинеен модел, чието асимптотично разпределение е изведено при условие, че H_0 е вярна.
- С помощта на симулации за конкретни линейни и нелинейни модели показваме, че новият тест се представя добре в сравнение с други тестове в литературата. Предимството е в лесната числена работа.
- Намерени са условия които определят скоростта на сходимост към линеен модел на суми от сл. вектори при независими и еднакво разпределени събираеми.

Литература

[1] Bierens, H. J. (1982). *Consistent Model Specification Tests*, Journal of Econometrics 20: 105-134.

[2] Bierens, H. J. and Ploberger, W. (1997). *Asymptotic theory of integrated conditional moment tests*, Econometrica 65: 1129–1151.

[3] Bierens, H. J. (2016). *Econometric model specification*, World Scientific.

[4] Brock, W., D. Hsieh and B. LeBaron (1991). Nonlinear dynamics, chaos, and instability: statistical theory and economic evidence, MIT press

[5] Granger, C. and Lee, Tae-Hwy (1999). *The effect of aggregation on nonlinearity*, Econometric Reviews 18(3): 259–269

[6] Lee, T-H., H. White and C. Granger (1993). *Testing for neglected nonlinearity in time series models: A comparison of neural network methods and alternative tests*, Journal of Econometrics 56(3): 269–290

[7] Varberg, D. (1966). *Convergence of quadratic forms in independent random variables*, The Annals of Mathematical Statistics 37(3): 567–576