Универзитет у Београду
Математички факултет

Уопштени метод момената и примена у регресионим моделима

МАСТЕР РАД

Аутор:
Мирјана Вељовић 1012/2018

Ментор:
доц. др Бојана Милошевић

Београд, 2019.

Садржај

1 Увод **2**

1.1 УММ као уопштење МНК и ММВ 7

2 Оцењивање инструментима у линеарним регресионим моделима 10

2.1 Оцена параметра и декомпозиција популационог услова момената . 13

2.2 Асимптотска својства оцене параметра и оцене узорачког момента . 16

2.3 Оптимални избор тежинске матрице 19

2.4 Грешке у спецификацији модела 20

2.5 Закључак 22

3 УММ у коректно спецификованим моделима **23**

3.1 Популациони услов момената и идентификација параметра 23

3.2 Оцена параметра 25

3.3 $\mathcal{I}$ и $\mathcal{O}$ услови 25

3.4 Асимптотска својства оцене параметра и оцене узорачког момента . 27

3.4.1 Постојаност 27

3.4.2 Асимптотска нормалност 29

3.4.3 Оцена узорачког момента 31

3.5 Оцењивање S 32

3.6 Оптимални избор тежинске матрице 34

3.7 Закључак 36

4 УММ у погрешно спецификованим моделима **37**

5 Тестирање хипотеза **42**

5.1 Тест $\mathcal{O}$ услова (J-тест) 42

5.2 Тестирање хипотеза о вектору параметара 42

6 Понашање на коначним узорцима **45**

7 Примери у R-у **46**

8 Закључак **65**

Литература **66**

1 Увод

Уопштени метод момената (УММ) представљен је осамдесетих година XX века и од тада је изазвао велико интересовање, због своје широке примењивости у разним областима статистике и економетрије. УММ је у међувремену постао један од главних статистичких алата, а његов развој је имао највећи утицај у макроекономији и финансијама, где се највише и примењује. Његов велики значај је и у томе што обухвата многе друге статистичке методе, попут методе најмањих квадрата (МНК), методе максималне веродостојности (ММВ) и методе оцењивања инструментима.

У овом раду представићемо основне појмове уопштене методе момената, испитати њихове особине и приказати оцењивање и методе закључивања. Фокусираћемо се на примену методе на временским серијама[1]. Почећемо од методе оцењивања инструментима у линеарним регресионим моделима и показати главне идеје, а затим их проширити на примену у нелинеарним моделима. Приказаћемо асимптотске особине оцена добијених УММ, попут постојаности и асимптотске нормалности и представити неке од тестова. Додатно, позабавићемо се неким питањима својственим овој методи, попут оцењивања асимптотске коваријационе матрице узорачког момента и утицаја погрешне спецификације модела.

Природно је да се запитамо зашто је увођење уопштене методе момената изазвало толико интересовање, јер знамо да је у то време већ била доступна ММВ која даје најефикасније оцене. Оптималност ММВ потиче од тога што се метода базира на функцији веродостојности, односно на заједничкој расподели података, међутим, то некада може бити више мана него предност. Два проблема у веза са ММВ мотивишу коришћење УММ. Први проблем је осетљивост статистичких својстава на претпоставку о расподели - жељена статистичка својства оцене се постижу само ако је претпоставка о расподели тачна. Овде видимо посебан значај примене УММ у економији, јер економска теорија ретко обезбеђује потпуну спецификацију расподеле података. Други проблем је што је извођење закључака методом МВ за многе моделе који се срећу у пракси превише рачунски компликовано.

Насупрот томе, УММ обезбеђује рачунски погодну методу за извођење закључака без потребе да се одреди расподела података и сама функција веродостојности. УММ се заснива на скупу такозваних популационих услова момената, који се изводе из претпоставки модела. Природа ових услова варира од случаја до случаја, а њихова тачност је кључна за особине резултујуће оцене.

[1] Оне представљају податке који се скупљају за један ентитет током времена. Међутим, УММ се може примењивати и на податке који се прикупљају за више ентитиета у фиксираном временском тренутку, као и на такозване *панел* податке - податке који се прикупљају за више ентитета током времена.

Потенцијал коришћења услова момената у оцењивању је први пут препознат увођењем добро познате технике оцењивања - методе момената. Међутим, пре увођења УММ, статистичка теорија је била ограничена на услове момената одређене форме. Један од главних доприноса УММ је што уобличава заједничку структуру ових претходних анализа и развија статистичку теорију која се може користити за било који скуп услова момената. Према томе, пошто се УММ заснива на овим ранијим анализама, корисно је да се осврнемо на његове статистичке претходнике.

Прва важна метода која је мотивисала развој УММ је добро позната метода момената. Присетимо се, пошто су популациони моменти функције од непознатих параметара, параметри се у оквиру ММ оцењују вредностима које задовољавају одговарајуће узорачке аналогоне. Размотримо једноставан пример да се присетимо проблема који се могу јавити. Рецимо да желимо да оценимо параметре нормалне расподеле - m_0 и σ_0^2. Они задовољавају популационе услове

$$E(v_t - m_0) = E(v_t) - m_0 = 0$$

$$E(v_t - m_0)^2 = E(v_t^2) - m_0^2 = \sigma_0^2 \implies E(v_t^2) - (m_0^2 + \sigma_0^2) = 0.$$

Оцене су вредности $\hat{m}_n$ и $\hat{\sigma}_n$ које задовољавају одговарајуће узорачке услове (претпоставимо да је n обим узорка)

$$\frac{1}{n}\sum_{t=1}^{n} v_t - \hat{m}_n = 0, \text{ тј. } \hat{m}_n = \frac{1}{n}\sum_{t=1}^{n} v_t$$

$$\frac{1}{n}\sum_{t=1}^{n} v_t^2 - (\hat{m}_n^2 + \hat{\sigma}_n^2) = 0, \text{ тј. } \hat{\sigma}_n^2 = \frac{1}{n}\sum_{t=1}^{n}(v_t - \hat{m}_n)^2.$$

Међутим, ако бисмо, на пример, желели да базирамо оцењивање на прва три момента v_t, односно ако бисмо додали и услов

$$E(v_t - m_0)^3 = E(v_t^3) - 3E(v_t^2)m_0 + 3E(v_t)m_0^2 - m_0^3 = 0$$

на основу узорачких аналогона добили бисмо систем од 3 једначине са 2 непознате, који нема решење. Дакле, ММ је немоћна у овом случају. Према томе, јасно је да је била потребна нека модификација да би се оценило p параметара на основу више од p популационих услова момената.

Друга значајна метода која је имала утицај на развој УММ је минимум хи-квадрат метода. Метода је предложена како би се олакшало закључивање о томе да ли је узорак добијен из одређене расподеле, међутим, основна идеја која стоји иза ње се може применити на велики број проблема, укључујући и оцењивање (m_0, σ_0^2) на основу прва три услова момената.

Нејман и Пирсон су разматрали моделовање вероватноће да исход експеримента припада једном од s дисјунктних скупова (чија је унија простор исхода). Ако p_j означава вероватноћу да је исход у j-том скупу, онда је нулта хипотеза од интереса да је $p_j = h(j; \theta_0)$, где је $h(.)$ неки функционал индексиран параметром θ_0. Питање је како тестирати ову хипотезу. Пирсон је показао да се закључци могу базирати на статистици[2]

$$GF_n(\theta_0) = \sum_{j=1}^{s} \frac{(n_j - nh(j; \theta_0))^2}{n_j}$$

где је n_j учесталост исхода у j-том скупу из узорка обима n. Показано је да при нултој хипотези ова статистика има апроксимативно χ^2_{s-1} расподелу. Ако је θ_0 непознато, показано је да се ова статистика може користити за оцењивање θ_0, тако што се за оцену $\hat{\theta}_n$ узме вредност θ која минимизује $GF_n(\theta)$. Тада $GF_n(\hat{\theta}_n)$ има апроксимативно χ^2_{s-1-p} расподелу при нултој хипотези, где је p димензија вектора θ_0.

Иако на први поглед делује да ово нема никакве везе са оцењивањем параметара нормалне расподеле на основу прва три услова момената, испоставља се да се слична идеја може применити и у том случају. Да бисмо открили везу, уводимо скуп индикатора $\{I_t(j); j = \overline{1,s}, t = \overline{1,n}\}$ где $I_t(j)$ узима вредност 1 ако је t-ти исход у j-том скупу, а 0 иначе. Ако важи нулта хипотеза, онда је $P(I_t(j) = 1) = h(j; \theta_0)$ па је $E(I_t(j)) = h(j; \theta_0)$. Према томе, нулта хипотеза имплицира скуп од s популационих услова момената[3]

$$E \begin{pmatrix} I_t(1) - h(1; \theta_0) \\ I_t(2) - h(2; \theta_0) \\ . \\ . \\ . \\ I_t(s) - h(s; \theta_0) \end{pmatrix} = 0.$$

Ако је $s - 1 \geq p$, онда се ови услови могу користити за оцењивање θ_0. Узорачки аналогон је дат са

$$\begin{pmatrix} \hat{p}_1 - h(1; \theta) \\ \hat{p}_2 - h(2; \theta) \\ . \\ . \\ . \\ \hat{p}_s - h(s; \theta) \end{pmatrix} = 0$$

[2]Goodness of fit statistic.

[3]Пошто је $\sum_{j=1}^{s} (I_t(j) - h(j; \theta_0)) = 0$, потребно је само $s-1$ услова да нам обезбеди јединствене информације о θ_0, али ћемо задржати свих s да бисмо нагласили везу са $GF_n(\theta_0)$.

где је $\hat{p}_j = \frac{n_j}{n}$, релативна учесталост исхода у j-том скупу. Статистику $GF_n(\theta_0)$ можемо записати на следећи начин

$$GF_n(\theta_0) = \sum_{j=1}^{s} \frac{(n_j - nh(j;\theta_0))^2}{n_j} = n \sum_{j=1}^{s} \frac{(\hat{p}_j - h(j;\theta_0))^2}{\hat{p}_j}.$$

Ако је $s-1=p$ (исти број јединствених популационих услова момената и непознатих параметара), оцена ММ, у ознаци $\hat{\theta}_n$, задовољава да је $\hat{p}_j - h(j;\hat{\theta}_n) = 0$, за $j = \overline{1,s}$. Ово имплицира да је $GF_n(\hat{\theta}_n) = 0$, а пошто је $GF_n(\theta) \geq 0$, $\forall \theta$, важи да $\hat{\theta}_n$ минимизује $GF_n(\theta)$. Дакле, ако је $s-1=p$, оцена ММ и минимум хи-квадрат оцена се поклапају.

Ако је $s-1>p$, не можемо добити оцену ММ, али се минимум хи-квадрат метода и даље може применити.

Према томе, да бисмо оценили (m_0, σ_0^2) на основу прва три момента нормалне расподеле, можемо да формулишемо оцењивање у терминима минимизације. Да бисмо имплементирали такву стратегију, морамо одабрати одговарајућу статистику коју треба минимизовати. Вратимо се на минимум хи-квадрат статистику да стекнемо идеју о облику одговарајуће статистике. Статистику $GF_n(\theta)$ можемо записати на следећи начин[4]

$$GF_n(\theta) = n \begin{pmatrix} \hat{p}_1 - h(1;\theta) \\ \hat{p}_2 - h(2;\theta) \\ . \\ . \\ . \\ \hat{p}_s - h(s;\theta) \end{pmatrix}^T \begin{pmatrix} \hat{p}_1^{-1} & 0 & . & . & . & 0 \\ 0 & \hat{p}_2^{-1} & . & . & . & 0 \\ . & & & & & . \\ . & & & & & . \\ . & & & & & . \\ 0 & 0 & . & . & . & \hat{p}_s^{-1} \end{pmatrix} \begin{pmatrix} \hat{p}_1 - h(1;\theta) \\ \hat{p}_2 - h(2;\theta) \\ . \\ . \\ . \\ \hat{p}_s - h(s,\theta) \end{pmatrix}.$$

Овај запис нам даје идеју како да сличан приступ применимо на било који проблем када имамо више услова момената него параметара које треба оценити. За оцењивање (m_0, σ_0^2) на основу прва три услова момената, дефинишемо статистику коју треба минимизовати на следећи начин

$$\begin{pmatrix} M_v(1) - m \\ M_v(2) - (m^2 + \sigma^2) \\ M_v(3) - 3M_v(2)m + 3M_v(1)m^2 - m^3 \end{pmatrix}^T M_n \begin{pmatrix} M_v(1) - m \\ M_v(2) - (m^2 + \sigma^2) \\ M_v(3) - 3M_v(2)m + 3M_v(1)m^2 - m^3 \end{pmatrix}$$

где је M_n позитивно дефинитна матрица која може да зависи од n, а $M_v(j) = \frac{1}{n}\sum_{t=1}^{n} v_t^j$. Направили смо две модификације у односу на $GF_n(\theta)$ - фактор n је изостављен, јер нема утицаја на минимизацију, а разлика је и у томе што сада нисмо навели тачну форму матрице M_n - то може бити било

[4] Матрица у средини је позитивно дефинитна ($GF_n(\theta)$ није дефинисана ако није $\hat{p}_j > 0$, за $\forall j$), па осигурава да је $GF_n(\theta) \geq 0$.

која позитивно дефинитна матрица. Оцене од (m_0, σ_0^2) су вредности (m, σ^2) које минимизују наведену статистику.

Наведени проблем у вези са методом момената дао је први наговештај да је потребно формулисати општију методу, на основу које ћемо моћи да изводимо закључке када имамо више услова момената него непознатих параметара. У економетрији је ова ситуација честа - обично су нам задати неки услови момената који важе на популацији, али не знамо читаву расподелу података, па не можемо спровести ММВ, те нам је потребан неки метод који ће спровести оцењивање тако да искористи све доступне информације и доведе до најбољих могућих оцена. Када имамо више услова момената него непознатих параметара (а обично желимо да укључимо што више њих у закључивање, јер пружају више информација, па воде до бољих закључака), не можемо спровести ММ, али нам минимум хи-квадрат метода даје идеју како да конструишемо оцену - као вредност параметра која је најближа задовољавању узорачких услова момената. Дакле, проблеми у вези ММ нам дају мотивацију за увођење УММ, док нам минимум хи-квадрат метода даје идеју како да конструишемо оцену и спроводимо закључивање.

Уводимо следеће дефиниције:

Дефиниција 1.1: Популациони услов момената
Нека је θ_0 вектор непознатих параметара које треба оценити, v_t вектор случајних величина и $g(.)$ вектор функција. Популациони услов момената је облика

$$E(g(v_t, \theta_0)) = 0, \text{ за } \forall t.$$

Дефиниција 1.2: Оцена УММ
Оцена УММ базирана на наведеном услову момената је вредност θ која минимизује

$$Q_n(\theta) = \Big(\frac{1}{n}\sum_{t=1}^{n} g(v_t, \theta)\Big)^T W_n \Big(\frac{1}{n}\sum_{t=1}^{n} g(v_t, \theta)\Big)$$

где је W_n позитивно семи-дефинитна матрица која може да зависи од података, али конвергира у вероватноћи ка позитивно дефинитној матрици константи.

Позитивна семи-дефинитност имплицира да је $Q_n(\theta) \geq 0$, за свако θ и да је $Q_n(\hat{\theta}_n) = 0$, ако је $\frac{1}{n}\sum_{t=1}^{n} g(v_t, \hat{\theta}_n) = 0$.

Пример:

Претпоставимо да имамо модел

$$y_i = X_i^T \theta_0 + e_i, \; E(Z_i e_i) = 0, \; i = \overline{1, n}.$$

Популациони услов момената на коме се заснива оцењивање УММ је облика

$$E(g(\theta_0)) = E(Z_i(y_i - X_i^T \theta_0)) = 0.$$

Узорачки моменти су, према томе, облика

$$g_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} g_i(\theta) = \frac{1}{n}\sum_{i=1}^{n} Z_i(y_i - X_i^T\theta) = \frac{1}{n}Z^T(y - X\theta)$$

где је $Z = (Z_1, ..., Z_n)^T$, $X = (X_1, ..., X_n)^T$ и $y = (y_1, ..., y_n)^T$. Претпоставимо да се за оцењивање УММ користи тежинска матрица $W_n = (\frac{1}{n}Z^T Z)^{-1}$. Тада је оцена параметра θ_0 уопштеном методом момената

$$\hat{\theta}_n = \underset{\theta}{argmin}\Big(\frac{1}{n}(y - X\theta)^T Z\Big)\big(\frac{1}{n}Z^T Z\big)^{-1}\Big(\frac{1}{n}Z^T(y - X\theta)\Big)$$

$$= \underset{\theta}{argmin}(y - X\theta)^T Z(Z^T Z)^{-1} Z^T(y - X\theta).$$[5]

УММ настаје као потреба за уопштењем ММ, међутим, овај метод као своје специјалне случајеве обухвата и многе друге познате методе закључивања, што нам омогућава да уочимо заједничке особине и испитујемо својства наизглед веома различитих оцена. Међу најзначајнијим примерима су ММВ и МНК, али и оцењивање инструментима, које се најчешће користи код проблема ендогености у линеарним регресионим моделима. За почетак, осврнимо се на интерпретацију ММВ и МНК у овом контексту, док ћемо се на инструменте, као важан специјални случај, вратити касније.

1.1 УММ као уопштење МНК и ММВ

Велики број оцена се добија оптимизацијом скалара облика

$$\sum_{t=1}^{n} T_t(\theta).$$

Ако је $T_t(\theta)$ диференцијабилна, оцена је вредност $\hat{\theta}_n$ која задовољава једначину

$$\sum_{t=1}^{n} \frac{dT_t(\hat{\theta}_n)}{d\theta} = 0.$$

Ова једначина нам говори да је $\hat{\theta}_n$ заправо оцена методом момената на основу услова

$$E\Big(\frac{dT_t(\theta_0)}{d\theta}\Big) = 0$$

па је истовремено и оцена уопштеном методом момената на основу овог популационог услова момената (у овом случају је $g(v_t, \theta) = \frac{dT_t(\theta)}{d\theta}$, док тежинска матрица не игра улогу у анализи). Ово можемо искористити да добијемо интерпретацију

[5] Ова оцена позната је у литератури као $2SLS$ оцена (Two-stage least squares estimator).

МНК и ММВ.

Пример 1: Оцењивање МНК у линеарном регресионом моделу

Размотримо линеарни регресиони модел

$$y_t = x_t^T \theta_0 + e_t,\ t = \overline{1, n}.$$

Оцена МНК је вредност θ која минимизује $\displaystyle\sum_{t=1}^{n} e_t(\theta)^2 = \sum_{t=1}^{n} (y_t - x_t^T \theta)^2$. Дакле, у контексту претходне дискусије

$$T_t(\theta) = (y_t - x_t^T \theta)^2.$$

Према томе, оцена МНК се може интерпретирати као оцена УММ на основу популационог услова момената

$$E(x_t(y_t - x_t^T \theta_0)) = 0^6.$$

Пример 2: Оцењивање ММВ

Нека је $p(v_t; \theta_0)$ густина расподеле непрекидног стационарног случајног вектора v_t. Оцена ММВ од θ_0 је вредност θ која максимизује функцију веродостојности, односно њен логаритам, тј.

$$\log L_n(\theta) = \log \prod_{t=1}^{n} p(v_t; \theta) = \sum_{t=1}^{n} \log(p(v_t; \theta))$$

што се уклапа у претходну дискусију за

$$T_t(\theta) = \log(p(v_t; \theta))$$

па се оцена ММВ може интерпретирати као оцена УММ на основу услова

$$E\left(\frac{d \log(p(v_t; \theta_0))}{d\theta}\right) = 0.$$

Већ смо споменули да је главна слабост ММВ њена зависност од избора расподеле вероватноће. Тај проблем се лако уочава ако користимо ову интерпретацију ММВ. Оцена ће бити постојана[7], ако је $E\big(\frac{d \log(p(v_t;\theta_0))}{d\theta}\big) = 0$. Лако се показује да је овај услов аутоматски задовољен ако је расподела тачна.
Густина расподеле задовољава да је

$$\int_V p(v_t; \theta_0) dv_t^T = 1.$$

[6] Примећујемо да нам овај услов говори да су регресор и грешка некорелисани.

[7] Ово ћемо показати касније у општем случају оцена УММ.

Ако диференцирамо по θ, добијамо да је

$$\frac{d}{d\theta}\Big(\int_V p(v_t;\theta_0)dv_t^T\Big) = 0.$$

Ако $p(.)$ задовољава релативно благе услове за замену места диференцирању и интеграцији, онда имамо да је

$$\int_V \frac{dp(v_t;\theta_0)}{d\theta}dv_t^T = 0$$

односно

$$\int_V \frac{1}{p(v_t;\theta_0)}\frac{dp(v_t;\theta_0)}{d\theta}p(v_t;\theta_0)dv_t^T = 0.$$

Ако је претпостављена расподела заиста расподела од v_t, онда се ово може написати као $E\big(\frac{d\log(p(v_t;\theta_0))}{d\theta}\big) = 0$[8]. Међутим, ако претпоставка о расподели није тачна, онда се ово не може интерпретирати као очекивање и не имплицира да је $E\big(\frac{d\log(p(v_t;\theta_0))}{d\theta}\big) = 0$. То не значи да овај услов не важи никада ако је расподела погрешна, али ако постоји могућност грешке, теоријско оправдање више не важи.

[8] $\frac{d\log(p(\theta))}{d\theta} = \frac{1}{p(\theta)}\frac{dp(\theta)}{d\theta}$.

2 Оцењивање инструментима у линеарним регресионим моделима

Регресиони модели често пате од проблема *ендогености*, односно корелисаности предиктора и грешке. Таква ситуација настаје када промене зависне променљиве мењају вредност бар једног од предиктора, када постоје изостављене променљиве које утичу и на зависну променљиву и на предикторе, или када постоје грешке мерења. Предиктори који пате од бар једног од ових проблема називају се *ендогеним* предикторима. У овом случају, оцене МНК су пристрасне и непостојане, па оцењивање одговарајућих регресионих коефицијената методом најмањих квадрата води до лоших закључака. Стандардна процедура која се спроводи у таквим ситуацијама је метода оцењивања инструментима, која омогућава добијање постојаних оцена и у оваквим ситуацијама. Ова метода такође представља специјални случај уопштене методе момената.

Инструмент је „помоћна" променљива, променљива која сама по себи најчешће не припада једначини модела, али је корелисана са ендогеним предикторима под условом других предиктора. У линеарним моделима, постоје два главна захтева која променљива треба да испуњава да би се могла користити као инструмент:

1) Инструмент мора бити корелисан са ендогеним предикторима, под условом осталих предиктора. Пожељно је да ова корелација буде што јача, јер слаба корелација може водити до лоших закључака о оценама параметара и стандардних грешака.

2) Инструмент не сме бити корелисан са грешком у једначини модела, под условом осталих предиктора. Другим речима, инструмент не сме патити од истог проблема којег је имао почетни предиктор.

Ради илустрације, размотримо следећи систем једначина

$$q_t^{pot} = \alpha_0 p_t + e_t^{pot}$$
$$q_t^{pon} = \beta_{01}^T n_t + \beta_{02} p_t + e_t^{pon}$$

где q_t^{pot} и q_t^{pon} представљају потражњу и понуду у години t, p_t је цена робе у тој години, а n_t је вектор који садржи факторе који утичу на понуду. Претпоставља се да је $q_t^{pot} = q_t^{pon}$ и да је укупна произведена количина означена са q_t ($q_t^{pot} = q_t^{pon} = q_t$). Размотримо проблем оцењивања α_0 на основу узорка од n опсервација q_t и p_t. Метода најмањих квадрата води до пристрасних оцена (због истовремене узрочности q_t и p_t), па је предложено следеће решење - претпоставимо да постоји променљива z_t (за коју имамо податке) која је повезана са ценом, али таква да је $cov(z_t, e_t^{pot}) = 0$. Пример такве променљиве је неки од фактора који утиче на понуду, попут цене материјала или процента шкартова. Пошто је $q_t = \alpha_0 p_t + e_t^{pot}$, следи да је

$$cov(z_t, q_t) = \alpha_0 cov(z_t, p_t) \implies cov(z_t, q_t) - \alpha_0 cov(z_t, p_t) = 0.$$

Обично се претпоставља да је $E(e_t^{pot}) = 0$, па је $E(q_t) = \alpha_0 E(p_t)$, одакле добијамо да је

$$E(z_t q_t) - \alpha_0 E(z_t p_t) = 0.$$

Ово нам обезбеђује услов момената који се може користити као база за оцењивање, јер укључује променљиве за које имамо податке и непознати параметар α_0. Метод момената нас води оцењивању параметра вредностима које задовољавају аналогни узорачки моменат, тј.

$$\hat{\alpha}_n = \frac{\sum_{t=1}^{n} z_t q_t}{\sum_{t=1}^{n} z_t p_t}.$$

Рекли смо да метода оцењивања инструментима представља специјални случај уопштене методе момената, па сада желимо да је опишемо у контексту УММ, како бисмо интуитивно разумели кључне елементе УММ и створили базу за уопштење на нелинеарне моделе. Посматрајмо линеарни регресиони модел

$$y_t = x_t^T \theta_0 + e_t,\ t = \overline{1,n}$$

где је y_t зависна променљива, x_t је $p \times 1$ вектор предиктора, а e_t је грешка. Желимо да изведемо закључке о непознатом параметру θ_0 (односно $p \times 1$ вектору непознатих параметара θ_0). Нека је $e_t(\theta) = y_t - x_t^T \theta$. Дакле, $e_t(\theta_0) = e_t$.

Метода оцењивања инструментима заснива се на постојању вектора инструмената z_t, који су некорелисани са грешком e_t, тј. на постојању инструмената z_t који задовољавају $E(z_t e_t) = 0$. Дакле, то је специјални случај УММ, када се оцењивање базира на популационом услову момената $E(z_t e_t(\theta_0)) = 0$. Анализу ћемо почети овим специјалним случајем, па уводимо основне дефиниције и особине, које ћемо, уз одређене услове, касније уопштити.

Претпоставимо да је вектор инструмената $q \times 1$ вектор z_t. Да бисмо извели закључке и спровели анализу биће нам потребно неколико ограничења на променљиве. За почетак, уводимо следећу дефиницију:

Дефиниција 2.1: Строго стационаран процес
Нека је $\mathcal{N}_n = \{1, 2, ..., n\}$. Кажемо да је случајни процес $\{v_t, t \in \mathcal{N}_n\}$ строго стационаран, ако заједничка функција расподеле, $F(.)$, задовољава да је

$$F(v_{t_1}, ..., v_{t_k}) = F(v_{t_1+h}, ..., v_{t_k+h})$$

за сваки избор $t_1, ..., t_k \in \mathcal{N}_n$, свако $k \in \mathbb{N}$, такво да је $t_1, ..., t_k \in \mathcal{N}_n$ и свако $h \in \mathbb{Z}$, такво да је $t_1 + h, ..., t_k + h \in \mathcal{N}_n$.

Захтевамо да важи следеће:

Претпоставка 2.1: Строга стационарност
Случајни вектори $v_t = (x_t^T, z_t^T, e_t)^T$ чине строго стационаран процес.

Оцењивање θ_0 заснива се на следећем услову:

Претпоставка 2.2: Популациони услов момената
Вектор инструмената z_t задовољава: $E(z_t e_t(\theta_0)) = 0$.

У примеру смо навели да се инструменти могу користити за оцењивање параметара у једначинама понуде и потражње. Пошто се оцењивање заснива на методи момената, неопходно је само да пронађемо одговарајуће инструменте. Предложили смо два кандидата: цену материјала (z_{1t}) и проценат шкартова (z_{2t}). Интуиција налаже да је уместо произвољног избора између ових инструмената можда боље укључити оба кандидата у оцењивање. Ово води до 2×1 популационог услова момената

$$E(z_t(q_t - \alpha_0 p_t)) = 0$$

где је $z_t = (z_{1t}, z_{2t})^T$. Ово се уклапа у оквир претходне претпоставке, што видимо када поставимо $y_t = q_t$, $x_t = p_t$ и $\theta_0 = \alpha_0$.

Претпоставка о важењу популационог услова момената нам даје информације на којима се заснива оцењивање, међутим, оцењивање ће бити успешно само ако нам популациони услов момената обезбеђује довољно информација да јединствено одредимо θ_0, што не мора увек бити случај. Вектор параметара θ_0 је јединствено одређен[9] овим условом, ако важи да је $E(z_t e_t(\theta)) \neq 0$, $\forall \theta \in \Theta$, $\theta \neq \theta_0$. Пошто је

$$E(z_t e_t(\theta)) = E(z_t(y_t - x_t^T\theta)) = E(z_t(x_t^T\theta_0 + e_t - x_t^T\theta)) = E(z_t e_t(\theta_0)) + E(z_t x_t^T)(\theta_0 - \theta)$$

и претпоставили смо да је $E(z_t e_t(\theta_0)) = 0$, ово се своди на

$$E(z_t e_t(\theta)) = E(z_t x_t^T)(\theta_0 - \theta).$$

Према томе, θ_0 је јединствено одређено ако је $E(z_t x_t^T)(\theta_0 - \theta) \neq 0$, $\forall \theta \neq \theta_0$.
Ова неједнакост ће важити ако је матрица $E(z_t x_t^T)$ ранга p. Према томе, добијамо следећи услов идентификације параметра:

Претпоставка 2.3: Услов идентификације
$rank(E(z_t x_t^T)) = p$.

Популациони услов момената и услов идентификације обезбеђују кључне информације на којима се заснива оцењивање θ_0 - они имплицирају да постоји

[9]Кажемо да је параметар θ_0 *идентификован* овим популационим условом момената.

јединствена вредност θ у параметарском простору за коју је $E(z_t e_t(\theta)) = 0$[10].
Неуспех у идентификацији параметра се може десити из два разлога. Прво, ако постоји мање услова момената него параметара, тј. ако је $q < p$, јер је немогуће извући p информација потребних да јединствено одреде θ_0 из мање од p популационих услова момената[11].
Ипак, може се десити и да је $q \geq p$, а да параметар не може да се идентификује, јер популациони услови момената и даље не дају довољно информација да би се из њих могло јединствено одредити θ_0.

Јасно је да је веза између p и q веома важна. Кажемо да вектор параметара θ_0 *није идентификован* популационим условом момената, ако идентификација није успела. Ако је параметар идентификован и $p = q$, онда кажемо да је *тачно идентификован*, јер у овом случају популациони услов момената даје p информација потребних за идентификацију p чланова који чине θ_0. Ако је параметар идентификован и $q > p$, онда кажемо да је θ_0 *преидентификован* популационим условом момената (имамо више информација него што је потребно за идентификацију параметра). Надаље, претпостављамо да је параметар или тачно идентификован или преидентификован.

2.1 Оцена параметра и декомпозиција популационог услова момената

У уводу смо навели општу дефиницију оцене УММ, коју сада желимо да пренесемо на наш конкретан случај. Имали смо да је

$$y_t = x_t^T \theta_0 + e_t, t = \overline{1,n}$$

односно, матрично записано, да је

$$y = X\theta_0 + e$$

где је y $n \times 1$ вектор са t-тим чланом y_t, e $n \times 1$ вектор са t-тим чланом e_t, X $n \times p$ матрица са t-тим редом x_t^T, а θ_0 $p \times 1$ вектор непознатих параметара. Нека је Z $n \times q$ матрица са t-тим редом z_t^T. Дефинишемо, аналогно као раније, $e(\theta) = y - X\theta$. Ако извршимо одговарајуће измене у општој дефиницији[12], добијемо да у овом случају треба минимизовати

$$Q_n(\theta) = \big(\frac{1}{n} e(\theta)^T Z\big) W_n \big(\frac{1}{n} Z^T e(\theta)\big).$$

[10] Назвали смо је θ_0, међутим, ове претпоставке нам не говоре ништа о овој вредности осим да је јединствена.

[11] Тада је $rank(E(z_t x_t^T)) \leq q < p$.

[12] $g(v_t, \theta) = z_t e_t(\theta)$.

Према томе, оцена параметра θ_0 је дефинисана са

$$\hat{\theta}_n = \underset{\theta\in\Theta}{argmin} Q_n(\theta).$$

Када заменимо $e(\theta) = y - X\theta$ у $Q_n(\theta)$, добијамо

$$Q_n(\theta) = \frac{1}{n^2}\big(y^T Z W_n Z^T y + \theta^T X^T Z W_n Z^T X\theta - y^T Z W_n Z^T X\theta - \theta^T X^T Z W_n Z^T y\big).$$

Дакле, $\hat{\theta}_n$, као минимум, мора да задовољава да је[13]

$$\big(\frac{1}{n}X^T Z\big)W_n\big(\frac{1}{n}Z^T y\big) = \big(\frac{1}{n}X^T Z\big)W_n\big(\frac{1}{n}Z^T X\big)\hat{\theta}_n$$

па, уз услов да је $\big(\frac{1}{n}X^T Z\big)W_n\big(\frac{1}{n}Z^T X\big)$ несингуларна, имамо да је

$$\hat{\theta}_n = \Big(\big(\frac{1}{n}X^T Z\big)W_n\big(\frac{1}{n}Z^T X\big)\Big)^{-1}\big(\frac{1}{n}X^T Z\big)W_n\big(\frac{1}{n}Z^T y\big).$$

Ако претходну једначину запишемо на други начин

$$\big(\frac{1}{n}X^T Z\big)W_n\frac{1}{n}Z^T e(\hat{\theta}_n) = 0$$

добијамо да је $\hat{\theta}_n$ заправо оцена методом момената заснована на услову

$$E(x_t z_t^T)WE(z_t e_t(\theta_0)) = 0.$$

Ова интерпретација је корисна јер нам омогућава да увидимо везу између информација које се користе за добијање оцене параметра и информација које носи популациони услов момената. Минимизација $Q_n(\theta)$ по θ представља оцењивање базирано на информацијама да је p линеарних комбинација од $E(z_t e_t(\theta_0))$ у претходном изразу једнако 0. Овај израз имплицира да, ако је $p = q$, оцена УММ и оцена методом момената су еквивалентне, јер је у овом случају $E(x_t z_t^T)W$ несингуларна, па мора бити $E(z_t e_t(\theta_0)) = 0$. Тада тежинска матрица више не игра никакву улогу[14] и оцена УММ је дата са

$$\hat{\theta}_n = \big(\frac{1}{n}Z^T X\big)^{-1}\big(\frac{1}{n}Z^T y\big).$$

Ако је $p < q$, избор тежинске матрице је важан, јер одређује природу линеарних комбинација од $E(z_t e_t(\theta_0))$ за које је $E(x_t z_t^T)WE(z_t e_t(\theta_0)) = 0$. Такође, у овом случају се информације које носи популациони услов момената

[13] $\frac{dQ_n(\hat{\theta}_n)}{d\theta} = 0$.

[14] Ако је $p = q$, онда је $\Big(\big(\frac{1}{n}X^T Z\big)W_n\big(\frac{1}{n}Z^T X\big)\Big)^{-1} = \big(\frac{1}{n}Z^T X\big)^{-1}W_n^{-1}\big(\frac{1}{n}X^T Z\big)^{-1}$, због постојања инверза на десној страни.

$E(z_t e_t(\theta_0)) = 0$ разликују од информација које се заиста користе у оцењивању - $E(x_t z_t^T) W E(z_t e_t(\theta_0)) = 0$. Ово мотивише поделу популационог услова момената на део који се користи у оцењивању - услови које ћемо називати $\mathcal{I}$ *условима (идентификационим условима)* и остатак - услови које ћемо називати $\mathcal{O}$[15] *условима (условима преидентификације)*.

У уводу смо рекли да је матрица W позитивно дефинитна, па мора постојати $q \times q$ несингуларна матрица $W^{\frac{1}{2}}$, таква да је $(W^{\frac{1}{2}})^T W^{\frac{1}{2}} = W$. Дакле, можемо писати

$$E(x_t z_t^T) W E(z_t e_t(\theta_0)) = F^T W^{\frac{1}{2}} E(z_t e_t(\theta_0)) = 0$$

где је $F = W^{\frac{1}{2}} E(z_t x_t^T)$. Кажемо да су задовољени $\mathcal{I}$ услови ако важи

$$F(F^T F)^{-1} F^T W^{\frac{1}{2}} E(z_t e_t(\theta_0)) = 0.$$

Овај израз се састоји од q једначина по $W^{\frac{1}{2}} E(z_t e_t(\theta_0))$, али је $rank(F(F^T F)^{-1} F^T) = rank(F) \leq p$, па нису све линеарно независне. Да бисмо осигурали идентификацију, претпоставили смо да је ранг баш једнак p, што нам омогућава да уочимо важну везу између оцењивања и идентификације: p параметара модела је идентификовано само ако се оцењивање базира на p линеарно независних једначина. Из овога директно следи да је део $W^{\frac{1}{2}} E(z_t e_t(\theta_0))$ који се не користи у оцењивању

$$(I_q - F(F^T F)^{-1} F^T) W^{\frac{1}{2}} E(z_t e_t(\theta_0)) = 0$$

при чему је $rank((I_q - F(F^T F)^{-1} F^T)) = q - p$, па ово даје скуп од $q - p$ линеарно независних једначина по $W^{\frac{1}{2}} E(z_t e_t(\theta_0))$. Ако овај израз важи, кажемо да су задовољени $\mathcal{O}$ услови. Дакле, $q \times 1$ вектор популационих услова момената се може разложити на p $\mathcal{I}$ услова, који представљају део који се користи у оцењивању и $q - p$ $\mathcal{O}$ услова, који представљају остатак.

Размотримо сада понашање узорачких аналогона ових популационих вредности. Пошто $\mathcal{I}$ услови представљају информације на којима се базира оцењивање, очекујемо да су њихови узорачки аналогони задовољени у $\hat{\theta}_n$. Лако показујемо да ово важи, јер је $\left(\frac{1}{n} X^T Z\right) W_n \frac{1}{n} Z^T e(\hat{\theta}_n) = 0$, па важи да је

$$F_n (F_n^T F_n)^{-1} F_n^T W_n^{\frac{1}{2}} \frac{1}{n} Z^T e(\hat{\theta}_n) = 0$$

где је $F_n = W_n^{\frac{1}{2}} \left(\frac{1}{n} Z^T X\right)$ и $W_n = (W_n^{\frac{1}{2}})^T W_n^{\frac{1}{2}}$.

Насупрот томе, $\mathcal{O}$ услови се занемарују у оцењивању, па очекујемо да они не важе на нивоу узорка у општем случају. Они играју сличну улогу остатка на

[15] Identifying and overidentifying restrictions.

узорку, а из претходног следи да је

$$(I_q - F_n(F_n^T F_n)^{-1} F_n^T) W_n^{\frac{1}{2}} \frac{1}{n} Z^T e(\hat{\theta}_n) = W_n^{\frac{1}{2}} \frac{1}{n} Z^T e(\hat{\theta}_n)$$

па је (трансформисана) оцена узорачког момента заправо узорачки аналогон функцији података у $\mathcal{O}$ условима, што нас доводи до корисне интерпретације $Q_n(\theta)$. Пошто је $Q_n(\theta) = \big(\frac{1}{n} e(\theta)^T Z\big) W_n \big(\frac{1}{n} Z^T e(\theta)\big)$, ако погледамо последњи добијени израз видимо да минимална вредност, $Q_n(\hat{\theta}_n)$, мери колико је узорак далеко од тога да задовољи $\mathcal{O}$ услове. Ова интерпретација ће нам касније бити од помоћи у формирању тестова исправне спецификације модела.

2.2 Асимптотска својства оцене параметра и оцене узорачког момента

У овом поглављу показаћемо да је оцена параметра добијена овом методом постојана и да има асимптотски нормалну расподелу, што омогућава креирање асимптотских интервала поверења за елементе вектора θ_0. Такође, као што смо поменули, оцена узорачког момента игра важну улогу у формирању тестова, па ћемо истражити и њено асимптотско понашање. Закључке ћемо изводити уз помоћ Централне граничне теореме и Слабог закона великих бројева, па ћемо их за почетак формулисати (у најопштијем могућем облику, а услове ћемо наводити како нам буду били потребни):

Лема 2.1: Слаби закон великих бројева (СЗВБ)
Нека је $\{v_t, t = \overline{1,n}\}$ низ строго стационарних $s \times 1$ случајних вектора таквих да је $E(v_t) = \mu$. Тада, под одређеним условима регуларности

$$\frac{1}{n} \sum_{t=1}^{n} v_t \xrightarrow{P} \mu.$$

Лема 2.2: Централна гранична теорема (ЦГТ):
Нека је $\{v_t, t = \overline{1,n}\}$ низ строго стационарних $s \times 1$ случајних вектора таквих да је $E(v_t) = \mu$. Тада, под одређеним условима регуларности

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} (v_t - \mu) \xrightarrow{D} \mathcal{N}(0, S)$$

где $\mathcal{N}(0, S)$ означава s-димензиону нормалну расподелу са очекивањем 0 и позитивно дефинитном коваријационом матрицом

$$S = \lim_{n \to \infty} D\Big(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} (v_t - \mu)\Big).$$

Пошто претпоставка о строгој стационарности није довољна да обезбеди важење ових лема, а циљ нам је да илуструјемо основне идеје, погодно је да уведемо следећу претпоставку:

Претпоставка 2.4: Независност
Вектор $v_t = (x_t^T, z_t^T, e_t)^T$ *је независан од* v_{t+s}*, за свако* $s \neq 0$.

Ова претпоставка нам, уз претпоставку о строгој стационарности, говори да v_t чини iid (независан и једнако расподељен) процес.

Покажимо прво да је оцена параметра θ_0 постојана. Имамо да је

$$\hat{\theta}_n = \Big(\big(\frac{1}{n} X^T Z\big) W_n \big(\frac{1}{n} Z^T X\big) \Big)^{-1} \big(\frac{1}{n} X^T Z\big) W_n \big(\frac{1}{n} Z^T y\big)$$

и $y = X\theta_0 + e$, па је

$$\hat{\theta}_n = \theta_0 + \Big(\big(\frac{1}{n} X^T Z\big) W_n \big(\frac{1}{n} Z^T X\big) \Big)^{-1} \big(\frac{1}{n} X^T Z\big) W_n \big(\frac{1}{n} Z^T e\big).$$

Из теореме Слуцког, добијамо да је[16]

$$\mathrm{lim_p} \hat{\theta}_n = \theta_0 + \mathrm{lim_p} \Big(\big(\frac{1}{n} X^T Z\big) W_n \big(\frac{1}{n} Z^T X\big) \Big)^{-1} \big(\frac{1}{n} X^T Z\big) W_n \big(\frac{1}{n} Z^T e\big)$$

$$= \theta_0 + \Big(\mathrm{lim_p} \big(\frac{1}{n} X^T Z\big) \mathrm{lim_p}(W_n) \mathrm{lim_p} \big(\frac{1}{n} Z^T X\big) \Big)^{-1} \mathrm{lim_p} \big(\frac{1}{n} X^T Z\big) \mathrm{lim_p}(W_n) \mathrm{lim_p} \big(\frac{1}{n} Z^T e\big).$$

Знамо да је $\mathrm{lim_p}(W_n) = W$, позитивно дефинитна матрица. Гранично понашање осталих матрица добијамо коришћењем СЗВБ[17]

$$\frac{1}{n} Z^T X = \frac{1}{n} \sum_{t=1}^{n} z_t x_t^T \xrightarrow{P} E(z_t x_t^T)$$

$$\frac{1}{n} Z^T e = \frac{1}{n} \sum_{t=1}^{n} z_t e_t \xrightarrow{P} E(z_t e_t).$$

Претпоставили смо да је $rank(E(z_t x_t^T)) = p$, па постоји инверз од $E(x_t z_t^T) W E(z_t x_t^T)$ и важи да је

$$\mathrm{lim_p} \hat{\theta}_n = \theta_0 + M E(z_t e_t)$$

где је $M = (F^T F)^{-1} F^T W^{\frac{1}{2}}$ и $F = W^{\frac{1}{2}} E(z_t x_t^T)$. Популациони услов момената нам говори да је $E(z_t e_t) = 0$, па закључујемо да је $\hat{\theta}_n$ постојана оцена за θ_0.

---

[16] $\mathrm{lim_p}$ означава лимес у вероватноћи.
[17] Уз претпоставку да очекивања на десној страни постоје.

Сада желимо да покажемо да оцена има асимптотски нормалну расподелу. Из претходног, имамо да је

$$\sqrt{n}(\hat{\theta}_n-\theta_0)=\Big(\big(\frac{1}{n}X^TZ\big)W_n\big(\frac{1}{n}Z^TX\big)\Big)^{-1}\big(\frac{1}{n}X^TZ\big)W_n\big(\frac{1}{\sqrt{n}}Z^Te\big)=M_nh_n$$

где је $M_n=\Big(\big(\frac{1}{n}X^TZ\big)W_n\big(\frac{1}{n}Z^TX\big)\Big)^{-1}\big(\frac{1}{n}X^TZ\big)W_n$. Показали смо да M_n конвергира у вероватноћи матрици константи $M=\Big(E(x_tz_t^T)WE(z_tx_t^T)\Big)^{-1}E(x_tz_t^T)W$, а применом ЦГТ добијамо да

$$h_n=\frac{1}{\sqrt{n}}Z^Te=\frac{1}{\sqrt{n}}\sum_{t=1}^{n}z_te_t\xrightarrow{D}\mathcal{N}(0,S)$$

где је $S=\lim\limits_{n\to\infty}D\Big(\frac{1}{\sqrt{n}}\sum\limits_{t=1}^{n}z_te_t\Big)$ и очекивање је 0 због претпоставке о важењу популационог услова момената. Према томе, добијамо да

$$\sqrt{n}(\hat{\theta}_n-\theta_0)\xrightarrow{D}\mathcal{N}(0,MSM^T).$$

Ако је $p=q$, M се своди на $(E(z_tx_t^T))^{-1}$, па је $MSM^T=(E(z_tx_t^T))^{-1}S(E(x_tz_t^T))^{-1}$, те не зависи од тежинске матрице.

Пошто смо оправдали асимптотску нормалност оцене, можемо изводити апроксимативне интервале поверења за компоненте вектора $\theta_0=(\theta_{01},...,\theta_{0p})^T$ - $100(1-\alpha)\%$-ни апроксимативни интервал поверења за θ_{0i}, дат је са

$$\hat{\theta}_{ni}\pm z_{\frac{\alpha}{2}}\sqrt{\frac{[\hat{D}_{a,n}]_{ii}}{n}}$$

где је $\hat{\theta}_{ni}$ i-ти елемент вектора $\hat{\theta}_n$, $[\hat{D}_{a,n}]_{ii}$ (i,i)-ти елемент постојане оцене матрице MSM^T и $z_{\frac{\alpha}{2}}=\Phi^{-1}(1-\frac{\alpha}{2})$. Постојану оцену за MSM^T можемо добити одређивањем постојаних оцена појединачних компонената, јер из теореме Слуцког, ако $\hat{M}_n\xrightarrow{P}M$ и $\hat{S}_n\xrightarrow{P}S$, онда и $\hat{M}_n\hat{S}_n\hat{M}_n^T\xrightarrow{P}MSM^T$.
Раније смо показали да $M_n=\Big(\big(\frac{1}{n}X^TZ\big)W_n\big(\frac{1}{n}Z^TX\big)\Big)^{-1}\big(\frac{1}{n}X^TZ\big)W_n\xrightarrow{P}M$, па постављамо $\hat{M}_n=M_n$. Остаје још да нађемо постојану оцену за S.

Претпоставили смо да је z_te_t *iid* процес са средњом вредношћу нула, па је $E(e_te_sz_tz_s^T)=E(e^2zz^T)$, ако је $t=s$, а 0 ако је $t\neq s$.
Према томе,

$$S=\lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\sum_{s=1}^{n}E(e_te_sz_tz_s^T)=E(e^2zz^T).$$

Постојана оцена за S је дата са

$$\hat{S}_n = \frac{1}{n}\sum_{t=1}^{n} \hat{e}_t^2 z_t z_t^T, \text{ где је } \hat{e}_t = y_t - x_t^T\hat{\theta}_n.$$

На крају, желимо да израчунамо асимптотску расподелу оцене узорачког момента. Уобичајено је да се посматра трансформисана статистика која се добија множењем слева са $W_n^{\frac{1}{2}}$

$$W_n^{\frac{1}{2}}\frac{1}{\sqrt{n}}Z^T e(\hat{\theta}_n) = W_n^{\frac{1}{2}}\frac{1}{\sqrt{n}}Z^T e - W_n^{\frac{1}{2}}\frac{1}{n}Z^T X\sqrt{n}(\hat{\theta}_n - \theta_0).$$

Пошто је $\sqrt{n}(\hat{\theta}_n - \theta_0) = M_n h_n = \Big(\big(\frac{1}{n}X^T Z\big)W_n\big(\frac{1}{n}Z^T X\big)\Big)^{-1}\big(\frac{1}{n}X^T Z\big)W_n h_n$, следи да је

$$W_n^{\frac{1}{2}}\frac{1}{\sqrt{n}}Z^T e(\hat{\theta}_n) = (I_q - P_n)W_n^{\frac{1}{2}}\Big(\frac{1}{\sqrt{n}}Z^T e\Big) = T_n h_n$$

где је $T_n = (I_q - P_n)W_n^{\frac{1}{2}}$, $P_n = F_n(F_n^T F_n)^{-1}F_n^T$, а подсетимо се, $F_n = W_n^{\frac{1}{2}}\big(\frac{1}{n}Z^T X\big)$. T_n конвергира у вероватноћи ка матрици константи $T = (I_q - P)W^{\frac{1}{2}}$, $P = F(F^T F)^{-1}F^T$, а h_n конвергира вектору са $\mathcal{N}(0, S)$ расподелом. Дакле, имамо сличну ситуацију као код $\sqrt{n}(\hat{\theta}_n - \theta_0)$, па закључујемо да

$$W_n^{\frac{1}{2}}\frac{1}{\sqrt{n}}Z^T e(\hat{\theta}_n) \xrightarrow{D} \mathcal{N}(0, TST^T).$$

Поменули смо раније везу између оцене узорачког момента и $\mathcal{O}$ услова. Она се манифестује и у асимптотској расподели, јер имамо да је

$$W_n^{\frac{1}{2}}\frac{1}{\sqrt{n}}Z^T e(\hat{\theta}_n) = (I_q - P)W^{\frac{1}{2}}\Big(\frac{1}{\sqrt{n}}Z^T e\Big) + o_p(1)$$

па је асимптотско понашање $W_n^{\frac{1}{2}}\frac{1}{\sqrt{n}}Z^T e(\hat{\theta}_n)$ одређено функцијом података у $\mathcal{O}$ условима - јасно је да асимптотска расподела оцене узорачког момента има очекивање нула само ако су $\mathcal{O}$ услови задовољени у θ_0.

2.3 Оптимални избор тежинске матрице

Тежинска матрица игра важну улогу у анализи, јер одређује тачну природу функције коју треба минимизовати, $Q_n(\theta)$. Као оптимална тежинска матрица дефинише се она која минимизује асимптотску дисперзију оцене параметра (у матричном смислу).

Знамо да $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{D} \mathcal{N}(0, MSM^T)$ и $M = \big(E(x_t z_t^T)WE(z_t x_t^T)\big)^{-1}E(x_t z_t^T)W$. Дакле, прво треба да одредимо оптимални избор W, а онда да видимо како да конструишемо матрицу која конвергира у вероватноћи ка њој. Асимптотска дисперзија је облика

$$D_a(W) = \big(E(x_t z_t^T) W E(z_t x_t^T)\big)^{-1} E(x_t z_t^T) W S W E(z_t x_t^T) \big(E(x_t z_t^T) W E(z_t x_t^T)\big)^{-1}.$$

Оптималан избор за W је матрица W_{min} таква да минимизује $D_a(W)$ у матричном смислу, тј. таква да је $D_a(W) - D_a(W_{min})$ позитивно семи-дефинитна матрица за било који валидан избор матрице W. Показано је да је такво $W_{min} = S^{-1}$.[18] Ако заменимо ову вредност у израз за асимптотску дисперзију, добијамо ефикасну границу дисперзије за оцењивање θ_0 на основу услова $E(z_t e_t(\theta_0)) = 0$ (јер је за било који други избор W $D_a(W)$ бар толико велика)

$$D_a(S^{-1}) = \big(E(x_t z_t^T) S^{-1} E(z_t x_t^T)\big)^{-1}.$$

Да бисмо конструисали оцену УММ која достиже ефикасну границу, довољно је да поставимо $W_n = \hat{S}_n^{-1}$, где је $\hat{S}_n$ постојана оцена за S. Али како да одредимо $\hat{S}_n$? Знамо да $\hat{S}_n$ зависи од $\hat{\hat{\theta}}_n$, па је немогуће одредити је, а да прво не одредимо оцену параметра. Иако делује да упадамо у проблем, ово се срећом лако решава, јер је за постојаност $\hat{S}_n$ потребно само да буде конструисана коришћењем постојане оцене од θ_0 (не и оптималне). Према томе, користи се следећа процедура за конструисање оптималне оцене параметра: У првом кораку, оценимо θ_0 (постојаном) оценом на основу било ког валидног избора W_n[19] и ову оцену искористимо да конструишемо $\hat{S}_n$. У следећем кораку, постављамо $W_n = \hat{S}_n^{-1}$ и добијамо нову оцену за θ_0. Ова два корака су довољна да добијемо оптималну оцену (називамо је *двостепена (оптимална) оцена*), у смислу да је њена асимптотска дисперзија $D_a(S^{-1})$. Међутим, често је корисно наставити ову процедуру, јер делује да би могло бити бољитака у понашању на коначним узорцима, док ће асимптотско понашање оцене остати исто у сваком кораку - овако добијена оцена назива се *итерирана (оптимална) оцена*.

2.4 Грешке у спецификацији модела

До сада смо сматрали да је претпостављени статистички модел тачан, али у пракси, нажалост, то често није случај, па је важно испитати како грешке у спецификацији утичу на асимптотска својства оцене параметра и оцене узорачког момента. Интутитивно, делује да ће овакве грешке имати штетан утицај на закључке, што је мотивисало развијање тестова за испитивање коректности спецификације модела.

Нека $\mathcal{M}$ означава претпостављени статистички модел. Према томе,

$$\mathcal{M} \implies E(z_t e_t(\theta_0)) = 0, \text{ за } \forall t \text{ и неко јединствено } \theta_0 \in \Theta.$$

[18] За сада само наводимо резултат, док ћемо доказ приказати у општијем случају у наредном поглављу.

[19] Најчешће се користи $W_n = I_q$ или $W_n = (\frac{1}{n} Z^T Z)^{-1}$.

Ово имплицира да су и $\mathcal{I}$ и $\mathcal{O}$ услови задовољени у θ_0, па је $\hat{\theta}_n$ постојана оцена за θ_0, $\sqrt{n}(\hat{\theta}_n - \theta_0)$ конвергира нормалној расподели са очекивањем 0 и $\frac{1}{\sqrt{n}} Z^T e(\hat{\theta}_n)$ конвергира нормалној расподели са очекивањем 0.

Ако посумњамо да $\mathcal{M}$ не важи, природно се намећу две могућности. Прво, прави модел је $\mathcal{M}_1$, који, иако различит од $\mathcal{M}$, задовољава

$$\mathcal{M}_1 \implies E(z_t e_t(\theta_*)) = 0, \text{ за } \forall t \text{ и неко јединствено } \theta_* \in \Theta$$

или је прави модел $\mathcal{M}_2$, такав да

$$\mathcal{M}_2 \implies \nexists \theta \in \Theta \text{ такво да је } E(z_t e_t(\theta)) = 0, \text{ за } \forall t.$$

Ако је $p = q$, онда $E(z_t e_t(\theta)) = 0$ представља скуп од p једначина са p непознатих, па мора имати решење. Према томе, други случај може бити истинит само ако је $p < q$.

Претпоставимо сада да је тачан модел $\mathcal{M}_1$. $\mathcal{M}$ и $\mathcal{M}_1$ имају исте претпоставке о $E(z_t e_t(\theta))$ и једина разлика је у вредности параметра за коју је популациони услов момената задовољен. Ако је тачан модел $\mathcal{M}_1$, $\mathcal{I}$ и $\mathcal{O}$ услови су задовољени у θ_*, па $\hat{\theta}_n \xrightarrow{P} \theta_*$, $\sqrt{n}(\hat{\theta}_n - \theta_*)$ конвергира нормалној расподели са очекивањем 0 и $\frac{1}{\sqrt{n}} Z^T e(\hat{\theta}_n)$ конвергира нормалној расподели са очекивањем 0. Дакле, једина разлика између ова два модела је у вредности параметра којој оцена конвергира. Међутим, како ниједан модел не говори ништа о вредности θ која задовољава популациони услов момената осим да је јединствена, не могу се разликовати $\mathcal{M}$ и $\mathcal{M}_1$ на основу само $E(z_t e_t(\theta))$.

Међутим, ако је тачан модел $\mathcal{M}_2$, онда нема вредности θ која задовољава популациони услов момената, али мора постојати решење $\mathcal{I}$ услова, јер формирају скуп од p једначина са p непознатих. Ако је то решење θ_*, онда $\hat{\theta}_n \xrightarrow{P} \theta_*$. Могуће је описати и асимптотско понашање оцене, али је анализа доста компликованија него при моделу $\mathcal{M}$. Главна разлика се огледа у понашању оцене узорачког момента - испоставља се да $W_n^{\frac{1}{2}} \frac{1}{\sqrt{n}} Z^T e(\hat{\theta}_n)$ не конвергира у расподели.

Јасно је да су нам неопходни тестови који ће испитивати да ли је претпостављени модел тачан. Из дискусије видимо да је немогуће разликовати $\mathcal{M}$ и $\mathcal{M}_1$ само на основу $\frac{1}{\sqrt{n}} Z^T e(\hat{\theta}_n)$, међутим, очекујемо да ћемо моћи да разликујемо $\mathcal{M}$ и $\mathcal{M}_2$ на основу оцене узорачког момента, па се одатле и родила идеја за креирање тестова који се заснивају на $\mathcal{O}$ условима. Тестови се базирају на $Q_n(\hat{\theta}_n)$, јер смо рекли да ова статистика мери колико је далеко узорак од задовољавања $\mathcal{O}$ услова. Дефинишемо тест заснован на $\mathcal{O}$ условима (назива се и Ј-тест), који открива када је прави модел $\mathcal{M}_2$, са

$$J_n = nQ_n(\hat{\theta}_n) = \Big(\frac{1}{\sqrt{n}} e(\hat{\theta}_n)^T Z\Big) \hat{S}_n^{-1} \Big(\frac{1}{\sqrt{n}} Z^T e(\hat{\theta}_n)\Big).$$

При нултој хипотези

$$H_0 : E(z_t e_t(\theta_0)) = 0$$

$J_n \xrightarrow{D} \chi^2_{q-p}$. Испоставиће се да се исти тест користи и да исти резултат важи и у општијем, нелинеарном случају.

2.5 Закључак

У овом поглављу смо увели главне елементе уопштене методе момената користећи пример оцењивања инструментима у линеарним регресионим моделима. Овај специјални случај нам омогућава да истакнемо кључне аспекте методе, које ћемо испитивати у општијем случају у наредним поглављима:

1) **Идентификација:** Да би оцењивање било успешно, популациони услов момената мора бити тачан и мора обезбедити довољно информација да се јединствено одреди вектор параметара.

2) $\mathcal{I}$ **и** $\mathcal{O}$ **услови:** Оцењивање код преидентификованих модела укључује декомпозицију услова момената на $\mathcal{I}$ и $\mathcal{O}$ услове. $\mathcal{I}$ услови садрже информације које се користе у оцењивању, а $\mathcal{O}$ услови представљају остатак.

3) **Оцена параметра:** Оцена параметра је постојана и има граничну нормалну расподелу[20].

4) **Оцена узорачког момента:** Оцена узорачког момента има граничну нормалну расподелу, чије особине зависе од функције података у $\mathcal{O}$ условима.

5) **Оцењивање асимптотске коваријационе матрице:** Да бисмо могли да изводимо практичне закључке на основу асимптотске нормалности, неопходно је да постојано оценимо асимптотску дисперзију узорачког момента.

6) **Оптимални избор тежинске матрице:** Оптимални избор тежинске матрице зависи од асимптотске дисперзије узорачког момента, па је потребно спроводити двостепено или итерирано оцењивање.

7) **Дијагностика модела:** J-тест обезбеђује базу за тестирање коректности спецификације модела коришћењем оцене узорачког момента.

У наредном поглављу показаћемо како се УММ примењује у нелинеарним моделима, ослањајући се на темеље постављене у овом поглављу и испитати како се ових седам кључних тачака мењају у том случају.

[20]Када је скалирана на одговарајући начин.

3 УММ у коректно спецификованим моделима

У претходном поглављу смо увели основне појмове УММ, извели основне закључке и навели неке од проблема који могу настати. Иако се претходно поглавље заснивало на линеарним моделима, интуиција се преноси и на компликованије случајеве, односно на нелинеарне моделе. Нажалост, у пракси се најчешће срећемо са оваквим моделима, који воде до популационих услова момената који укључују нелинеарне функције података и параметара. У овом поглављу извешћемо оцене, испитати њихово асимптотско понашање и видети који проблеми могу настати (током овог поглавља претпостављамо да је модел коректно спецификован).

3.1 Популациони услов момената и идентификација параметра

Анализу ћемо започети формулисањем популационог услова момената, без покушаја да карактеризујемо процес генерисања података који стоји иза њега. Овај услов укључује функцију $g(.,.)$, чији су аргументи $r \times 1$ случајни вектор v_t и $p \times 1$ вектор непознатих параметара θ_0. Пре него што формулишемо популациони услов момената, неопходно је да поставимо нека ограничења на податке и функцију g.

Претпоставка 3.1: Строга стационарност
Случајни вектори $\{v_t \in V \subset \mathbb{R}^r, t \in \mathbb{R}\}$ формирају строго стационаран случајни процес.

Претпоставка 3.2: Услови регуларности за g
Функција $g : V \times \Theta \to \mathbb{R}^q$, $q < \infty$, задовољава:
1) *Непрекидна је на Θ за свако $v_t \in V$;*
2) *$E(g(v_t, \theta))$ постоји и коначно је за свако $\theta \in \Theta$;*
3) *$E(g(v_t, \theta))$ је непрекидно на Θ.*

Претпоставка 3.3: Популациони услов момената
Случајни вектор v_t и вектор параметара θ_0 задовољавају $q \times 1$ популациони услов момената $E(g(v_t, \theta_0)) = 0$.

Као и код линеарних модела, овај услов се може користити за оцењивање само ако обезбеђује довољно информација да се јединствено одреди θ_0. Међутим, овде је ситуација доста компликованија него код линеарних модела, јер смо тамо идентификацију параметра могли испитати једноставним условом који зависи само од података. У нелинеарним моделима, идентификација може бити неуспешна због особина података, v_t, али и због особина g као функције по θ, или комбинације ова два разлога. Према томе, овде морамо проширити анализу, па уводимо појмове *глобалне* и *локалне* идентификације.

Претпоставка 3.4: Глобална идентификација
$E(g(v_t, \theta)) \neq 0, \forall\theta \in \Theta$, такво да је $\theta \neq \theta_0$.

Дакле, глобална идентификација нам говори да популациони услов момената важи за само једну вредност из целог параметарског простора. У претходном поглављу смо извели релативно једноставан услов за глобалну идентификацију, па бисмо свакако волели да развијемо нешто слично и за нелинеарне моделе. Нажалост, ово се обично не може урадити, јер је немогуће пронаћи корисну алтернативну репрезентацију $g(v_t, \theta)$ која важи за $\forall\theta \in \Theta$, као што смо то имали код линеарних модела, одакле смо извели услов који треба да буде задовољен да би параметар био идентификован[21]. Међутим, то је могуће урадити на некој погодно одабраној околини од θ_0. Цена овог приступа је што сада изводимо услове идентификације само на овој околини (ово се назива *локалном идентификацијом*). Наравно, локална идентификација не гарантује глобалну, али глобална не може важити без локалне. Да бисмо извели услов локалне идентификације, потребно је да дефинишемо следећи појам - скуп $D_\epsilon = \{\theta : ||\theta - \theta_0|| < \epsilon\} \subset \Theta$ називамо ϵ-околином од θ_0. Такође, уводимо следећу претпоставку:

Претпоставка 3.5: Услови регуларности за $\frac{dg(v_t,\theta)}{d\theta^T}$
1) *Матрица извода $\frac{dg(v_t,\theta)}{d\theta^T}$ постоји и непрекидна је на Θ за свако $v_t \in V$;*
2) *θ_0 је унутрашња тачка Θ;*
3) *$E(\frac{dg(v_t,\theta_0)}{d\theta^T})$ постоји и коначно је.*

Услов локалне идентификације изводимо из Тејлоровог развоја функције g на D_ϵ за довољно мало ϵ

$$g(v_t, \theta) = g(v_t, \theta_0) + \frac{dg(v_t, \theta_0)}{d\theta^T}(\theta - \theta_0).$$

Ако прођемо очекивањем, пошто нам важење популационог услова момената говори да је $E(g(v_t, \theta_0)) = 0$, уз претходну претпоставку добијамо да је

$$E(g(v_t, \theta)) = E\left(\frac{dg(v_t, \theta_0)}{d\theta^T}\right)(\theta - \theta_0).$$

Добијамо форму сличну оној коју смо имали код линеарних модела[22]. Према томе, користећи сличну логику, уводимо следећи услов локалне идентификације (као уопштење оног којег смо имали код линеарних модела):

[21]Они су специјалан случај за $g(v_t, \theta) = z_t e_t(\theta)$, а $E(z_t e_t(\theta)) = E(z_t x_t^T)(\theta_0 - \theta)$, одакле смо извели да за идентификацију мора бити $rank(E(z_t x_t^T)) = p$.

[22]Тамо је $g(v_t, \theta) = z_t e_t(\theta)$, па је $\frac{dg(v_t,\theta_0)}{d\theta^T} = -z_t x_t^T$ (и имамо глобалну идентификацију, јер развој важи за свако θ, а не само на околини θ_0).

Претпоставка 3.6: Локална идентификација
$rank\big(E\big(\frac{dg(v_t,\theta_0)}{d\theta^T}\big)\big) = p.$

Овај услов имплицира да се параметар не може идентификовати ако је $q < p$, што је у складу са оним што смо имали код линеарних модела, али није било одмах очигледно из услова глобалне идентификације. Присетимо се, идентификација може бити неуспешна и ако је $q \geq p$, ако нам популациони услов момената не обезбеђује довољно информација да јединствено одредимо θ_0.

3.2 Оцена параметра

У уводу смо навели да се оцена параметра уопштеном методом момената по дефиницији добија као вредност θ која минимизује

$$Q_n(\theta) = \big(\frac{1}{n}\sum_{t=1}^{n} g(v_t,\theta)\big)^T W_n\big(\frac{1}{n}\sum_{t=1}^{n} g(v_t,\theta)\big).$$

Према томе, оцена је дефинисана са

$$\hat{\theta}_n = \underset{\theta\in\Theta}{argmin} Q_n(\theta).$$

Претпоставка 3.7: Особине тежинске матрице
Матрица W_n је позитивно семи-дефинитна матрица која конвергира у вероватноћи позитивно дефинитној матрици константи W.

Ако важи претпоставка о регуларности $\frac{dg(v_t,\theta)}{d\theta^T}$, онда минимизација имплицира да мора бити $\frac{dQ_n(\hat{\theta}_n)}{d\theta} = 0$. Одавде добијамо да $\hat{\theta}_n$ задовољава

$$\big(\frac{1}{n}\sum_{t=1}^{n}\frac{dg(v_t,\hat{\theta}_n)}{d\theta^T}\big)^T W_n\big(\frac{1}{n}\sum_{t=1}^{n} g(v_t,\hat{\theta}_n)\big) = 0.$$

Код линеарних модела, могли смо ово да решимо и добијемо експлицитни израз за $\hat{\theta}_n$, међутим, у нелинеарним моделима, то најчешће није могуће. Срећом, напредовање компјутерске технологије је омогућило развој техника нумеричке оптимизације које се могу користити за израчунавање $\hat{\theta}_n$.

3.3 $\mathcal{I}$ и $\mathcal{O}$ услови

Сама дефиниција оцене УММ не захтева да g буде диференцијабилна по θ, међутим, ако јесте, оцену параметра можемо дефинисати као решење претходне једначине. Ово нам је важно због интерпретације УММ методом момената, јер она, као и код линеарних модела, води до декомпозиције популационог услова момената на $\mathcal{I}$ и $\mathcal{O}$ услове. Ово разлагање је корисно за разумевање особина методе, а касније игра кључну улогу у формирању тестова спецификације модела, па проширујемо ранију причу о декомпозицији на нелинеарне моделе (задржавамо претпоставку о регуларности извода).

Ако погледамо претходну једначину, видимо да се оцена УММ базирана на популационом услову момената $E(g(v_t, \theta_0)) = 0$ може интерпретирати као оцена методом момената базирана на

$$F(\theta_0)^T W^{\frac{1}{2}} E(g(v_t, \theta_0)) = 0$$

где је $F(\theta_0) = W^{\frac{1}{2}} E\big(\frac{dg(v_t,\theta_0)}{d\theta^T}\big)$. Ова једначина имплицира да је $rank(F(\theta_0))$ линеарних комбинација од $W^{\frac{1}{2}} E(g(v_t, \theta_0))$ једнако 0. Претпоставка о локалној идентификацији нам говори да је овај ранг баш p, па нам као и код линеарних модела, ова интерпретација даје везу између оцењивања и идентификације - p параметара модела је *локално* идентификовано, ако се оцењивање базира на p линеарно независних једначина. У овом случају, претходна једначина има добро дефинисано решење у θ_0. Међутим, могу постојати и друге тачке параметарског простора у којима ова једначина има добро дефинисано решење - ова могућност се искључује претпоставком о глобалној идентификацији.

Ако је $p = q$, тежинска матрица не игра улогу у анализи, јер је претходна једначина еквивалентна са $E(g(v_t, \theta_0)) = 0$. Међутим, ако је $p < q$, онда постоји разлика између информација које се користе у оцењивању и информација које носи популациони услов момената. Слично као код линеарних модела, можемо разложити популациони услов момената на $\mathcal{I}$ и $\mathcal{O}$ услове. $\mathcal{I}$ услови (део који се користи у оцењивању) су дати са

$$F(\theta_0)(F(\theta_0)^T F(\theta_0))^{-1} F(\theta_0)^T W^{\frac{1}{2}} E(g(v_t, \theta_0)) = 0$$

а $\mathcal{O}$ услови са

$$(I_q - F(\theta_0)(F(\theta_0)^T F(\theta_0))^{-1} F(\theta_0)^T) W^{\frac{1}{2}} E(g(v_t, \theta_0)) = 0.$$

Поново, особине ових услова се одражавају на њиховим узорачким аналогонима - пошто се $\mathcal{I}$ услови користе у оцењивању, њихови узорачки аналогони су задовољени у $\hat{\theta}_n$, док се $\mathcal{O}$ услови занемарују при оцењивању, па њихови узорачки аналогони нису задовољени у општем случају.

Као и код линеарних модела, имамо интересантну везу између $\mathcal{O}$ услова и оцене узорачког момента. Важи да је

$$W_n^{\frac{1}{2}}\big(\frac{1}{n}\sum_{t=1}^{n} g(v_t, \hat{\theta}_n)\big) = (I_q - F_n(\hat{\theta}_n)(F_n(\hat{\theta}_n)^T F_n(\hat{\theta}_n))^{-1} F(\hat{\theta}_n)^T) W_n^{\frac{1}{2}}\big(\frac{1}{n}\sum_{t=1}^{n} g(v_t, \hat{\theta}_n)\big)$$

где је $F_n(\theta) = W_n^{\frac{1}{2}}\left(\frac{1}{n}\sum_{t=1}^{n}\frac{dg(v_t,\theta)}{d\theta^T}\right)$, па видимо да је трансформисана оцена узорачког момента узорачки аналогон функцији података која се појављује у $\mathcal{O}$ условима. Према томе, $Q_n(\hat{\theta}_n)$ се може интерпретирати као мера колико је узорак далеко од тога да задовољи $\mathcal{O}$ услове.

3.4 Асимптотска својства оцене параметра и оцене узорачког момента

Да бисмо развили асимптотску теорију, неопходно је да поставимо додатне услове на v_t[23].

Претпоставка 3.8 Ергодичност
Случајни процес $\{v_t, t \in \mathbb{R}\}$ је ергодичан.

3.4.1 Постојаност

Да бисмо испитали постојаност, размотрићемо шта се дешава ако применимо сличан принцип минимизације на популациони аналогон од $Q_n(\theta)$

$$Q_0(\theta) = (E(g(v_t,\theta))^T W E(g(v_t,\theta)).$$

Популациони услов момената имплицира да је $Q_0(\theta_0) = 0$, а ако претпоставимо да важи услов глобалне идентификације, уз позитивну дефинитност W, важи да је $Q_0(\theta) > 0, \forall\theta \neq \theta_0$, па $Q_0(\theta)$ достиже јединствени минимум у $\theta = \theta_0$. Интуиција налаже да ако $\hat{\theta}_n$ минимизује $Q_n(\theta)$ и $Q_n(\theta)$ конвергира у вероватноћи ка функцији $Q_0(\theta)$, чији је јединствени минимум θ_0 - онда ће и $\hat{\theta}_n$ конвергирати у вероватноћи ка θ_0. Међутим, не мора нужно минимум низа функција конвергирати минимуму граничне функције, али је довољно да $Q_n(\theta)$ конвергира униформно ка $Q_0(\theta)$ да би то важило. Морамо претпоставити следеће:

Претпоставка 3.9: Компактност Θ
Скуп Θ је компактан.

Претпоставка 3.10: Ограниченост
$E(\sup_{\theta\in\Theta}||g(v_t,\theta)||) < \infty.$

Сада можемо добити униформну конвергенцију:

[23]Биће нам потребни СЗВБ и ЦГТ, а стационарност, сама по себи, није довољна за њихову примену.

Лема 3.1: Униформна конвергенција у вероватноћи $Q_n(\theta)$
Ако важе претпоставке 3.1, 3.2, 3.7-3.10, тада $\sup\limits_{\theta\in\Theta}|Q_n(\theta) - Q_0(\theta)| \xrightarrow{P} 0$.

Теорема 3.1: Постојаност оцене параметра
Ако важе претпоставке 3.1-3.4, 3.7-3.10, тада $\hat{\theta}_n \xrightarrow{P} \theta_0$.
Доказ:
Доказ теореме ћемо извести из два корака.
<u>1.корак:</u> Показујемо да услови теореме имплицирају да $\lim\limits_{n\to\infty} P(0 \le Q_0(\hat{\theta}_n) < \epsilon) = 1$, *за* $\forall\epsilon > 0$, *што значи да* $\hat{\theta}_n$ *минимизује* $Q_0(\theta)$ *са вероватноћом 1, кад* $n \to \infty$.
Лема нам говори да разлика између $Q_n(\theta)$ *и* $Q_0(\theta)$ *нестаје са вероватноћом 1, кад* $n \to \infty$, *за* $\forall\theta \in \Theta$.
По дефиницији $\hat{\theta}_n \in \Theta$, *па је из леме* $\lim\limits_{n\to\infty} P(|Q_n(\hat{\theta}_n) - Q_0(\hat{\theta}_n)| < \frac{\epsilon}{3}) = 1$, *за* $\forall\epsilon > 0$.
Из овога следи да је

$$\lim_{n\to\infty} P(Q_0(\hat{\theta}_n) < Q_n(\hat{\theta}_n) + \frac{\epsilon}{3}) = 1.$$

Пошто $\hat{\theta}_n$ *минимизује* $Q_n(\theta)$, *следи да је*

$$\lim_{n\to\infty} P(Q_n(\hat{\theta}_n) < Q_n(\theta_0) + \frac{\epsilon}{3}) = 1.$$

Такође, $\theta_0 \in \Theta$, *па важи*

$$\lim_{n\to\infty} P(Q_n(\theta_0) < Q_0(\theta_0) + \frac{\epsilon}{3}) = 1.$$

Из комбинације прва два услова добијамо

$$\lim_{n\to\infty} P(Q_0(\hat{\theta}_n) < Q_n(\theta_0) + \frac{2\epsilon}{3}) = 1$$

а када ово комбинујемо са трећом тврдњом добијамо

$$\lim_{n\to\infty} P(Q_0(\hat{\theta}_n) < Q_0(\theta_0) + \epsilon) = 1.$$

Према томе, показали смо шта смо желели, јер претпоставка о важењу популационог услова момената имплицира да је $Q_0(\theta_0) = 0$, *а позитивна дефинитност* W *да је* $Q_0(\theta) \ge 0$, $\forall\theta$.

<u>2.корак:</u> Показујемо да део доказан у првом кораку имплицира постојаност оцене.
Нека је D *отворен подскуп* Θ, *такав да* $\theta_0 \in D$, *а* D^c *његов комплемент у односу на* Θ. D^c *је затворен подскуп компактног скупа, па је и сам компактан, а пошто је* $Q_0(\theta)$ *непрекидна функција, следи да она има инфимум на* D^c. *Пошто смо претпоставили да важи услов глобалне идентификације, следи да*

је овај инфимум строго позитиван (јер $\theta_0 \notin D^c$). Према томе, ако поставимо $\epsilon := \inf_{\theta \in D^c} Q_0(\theta)$, из првог корака имамо да

$$\lim_{n\to\infty} P(Q_0(\hat{\theta}_n) < \inf_{\theta \in D^c} Q_0(\theta)) = 1$$

што имплицира да је $\lim_{n\to\infty} P(\hat{\theta}_n \notin D^c) = 1$, па је $\lim_{n\to\infty} P(\hat{\theta}_n \in D) = 1$. Пошто ово важи за сваки избор отвореног скупа $D \subset \Theta$, без обзира колико мали он био, мора важити да је $\lim_{n\to\infty} P(\hat{\theta}_n = \theta_0) = 1$, што смо и желели да покажемо.

3.4.2 Асимптотска нормалност

Уведимо следеће ознаке: $g_n(\theta) = \dfrac{1}{n}\sum_{t=1}^{n} g(v_t,\theta)$ и $G_n(\theta) = \dfrac{1}{n}\sum_{t=1}^{n} \dfrac{dg(v_t,\theta)}{d\theta^T}$.

Желимо да покажемо да $\sqrt{n}(\hat{\theta}_n - \theta_0)$ тежи нормалној расподели, па нам је потребан одговарајући израз за $\sqrt{n}(\hat{\theta}_n - \theta_0)$ из кога ћемо моћи то да закључимо. Њега ћемо добити применом теореме о средњој вредности диференцијалног рачуна, па морамо претпоставити да важи претпоставка о регуларности извода. Из теореме добијамо да важи

$$g_n(\hat{\theta}_n) = g_n(\theta_0) + G_n(\hat{\theta}_n, \theta_0, \lambda_n)(\hat{\theta}_n - \theta_0)$$

где је $G_n(\hat{\theta}_n, \theta_0, \lambda_n)$ $q \times p$ матрица чији i-ти ред одговара i-том реду матрице $G_n(\overline{\theta}_n^{(i)})$, а $\overline{\theta}_n^{(i)} = \lambda_{ni}\theta_0 + (1-\lambda_{ni})\hat{\theta}_n$, за неко $\lambda_{ni} \in [0,1]$ и $\lambda_n = (\lambda_{n1}, ..., \lambda_{nq})^T$. Ако помножимо претходну једнакост са $G_n(\hat{\theta}_n)^T W_n$ слева, добијамо

$$0 = G_n(\hat{\theta}_n)^T W_n g_n(\hat{\theta}_n) = G_n(\hat{\theta}_n)^T W_n g_n(\theta_0) + G_n(\hat{\theta}_n)^T W_n G_n(\hat{\theta}_n, \theta_0, \lambda_n)(\hat{\theta}_n - \theta_0).$$

Према томе, добијамо да је

$$\hat{\theta}_n - \theta_0 = -\big(G_n(\hat{\theta}_n)^T W_n G_n(\hat{\theta}_n, \theta_0, \lambda_n)\big)^{-1} G_n(\hat{\theta}_n)^T W_n g_n(\theta_0)$$

па је

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = -\big(G_n(\hat{\theta}_n)^T W_n G_n(\hat{\theta}_n, \theta_0, \lambda_n)\big)^{-1} G_n(\hat{\theta}_n)^T W_n \sqrt{n} g_n(\theta_0) = -C_n \sqrt{n} g_n(\theta_0).$$

Добили смо исту структуру као код линеарних модела - случајна матрица C_n пута случајни вектор $\sqrt{n} g_n(\theta_0)$. Сада испитујемо сваку од ових компонената. Граничну расподелу случајног вектора $\sqrt{n} g_n(\theta_0)$ добијамо из верзије ЦГТ коју ћемо доле формулисати, а да би она могла да се примени, неопходно је да важи следећа претпоставка:

Претпоставка 3.11: Особине дисперзије узорачког момента
1) $E(g(v_t,\theta_0)g(v_t,\theta_0)^T)$ *постоји и коначно је;*
2) $\lim_{n\to\infty} D(\sqrt{n}g_n(\theta_0)) = S$ *постоји и* S *је позитивно дефинитна матрица са коначним вредностима.*

Лема 3.2: ЦГТ за $\sqrt{n}g_n(\theta_0)$
Ако важе претпоставке 3.1, 3.3, 3.8 и 3.11, онда $\sqrt{n}g_n(\theta_0) \xrightarrow{D} \mathcal{N}(0,S)$.

Следећи корак је да испитамо гранично понашање матрице $C_n = \big(G_n(\hat{\theta}_n)^T W_n G_n(\hat{\theta}_n,\theta_0,\lambda_n)\big)^{-1} G_n(\hat{\theta}_n)^T W_n$. Претпоставка је да $W_n \xrightarrow{P} W$. Погледајмо сада остале компоненте. Пошто $\hat{\theta}_n \xrightarrow{P} \theta_0$, а $\overline{\theta}_n^{(i)} = \lambda_{ni}\theta_0 + (1-\lambda_{ni})\hat{\theta}_n$, за неко $\lambda_{ni} \in [0,1]$, те стога лежи на дужи која спаја $\hat{\theta}_n$ и θ_0, мора важити да $\overline{\theta}_n^{(i)} \xrightarrow{P} \theta_0$, за $\forall i$. Да бисмо одредили чему теже $G_n(\hat{\theta}_n)$ и $G_n(\hat{\theta}_n,\theta_0,\lambda_n)$, потребне су нам следеће две претпоставке:

Претпоставка 3.12: Непрекидност $E\big(\frac{dg(v_t,\theta)}{d\theta^T}\big)$
$E\big(\frac{dg(v_t,\theta)}{d\theta^T}\big)$ *је непрекидна на некој* ϵ*-околини од* θ_0 D_ϵ.

Претпоставка 3.13: Униформна конвергенција $G_n(\theta)$
$\underset{\theta\in D_\epsilon}{sup}||G_n(\theta) - E\big(\frac{dg(v_t,\theta)}{d\theta^T}\big)|| \xrightarrow{P} 0.$[24]

Лема 3.3: Конвергенција $G_n(\hat{\theta}_n)$ ***и*** $G_n(\hat{\theta}_n,\theta_0,\lambda_n)$
Ако важе претпоставке 3.1-3.5, 3.7-3.10, 3.12 и 3.13 онда $G_n(\hat{\theta}_n) \xrightarrow{P} G_0$ *и* $G_n(\hat{\theta}_n,\theta_0,\lambda_n) \xrightarrow{P} G_0$*, где је* $G_0 = E\big(\frac{dg(v_t,\theta_0)}{d\theta^T}\big)$.

Коришћењем теореме Слуцког, закључујемо да $C_n \xrightarrow{P} (G_0^T W G_0)^{-1} G_0^T W = C$. Дакле, као и код линеарних модела, $\sqrt{n}(\hat{\theta}_n - \theta_0)$ је производ матрице која конвергира у вероватноћи матрици константи и случајног вектора који конвергира нормалној расподели.

Теорема 3.2: Асимптотска нормалност оцене параметра
Ако важе претпоставке 3.1-3.5 и 3.7-3.13 онда $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{D} \mathcal{N}(0, CSC^T)$.

Сада можемо извести апроксимативне интервале поверења за компоненте вектора θ_0. Теорема имплицира да је апроксимативни $100(1-\alpha)\%$-ни интервал поверења за θ_{0i}

$$\hat{\theta}_{ni} \pm z_{\frac{\alpha}{2}}\sqrt{\frac{[\hat{D}_{a,n}]_{ii}}{n}}$$

[24]За матрицу A, дефинишемо $||A|| = (tr(A^TA))^{\frac{1}{2}}$, а за вектор a, $||a|| = (a^Ta)^{\frac{1}{2}}$.

где је $\hat{\theta}_{ni}$ i-ти елемент вектора $\hat{\theta}_n$, $[\hat{D}_{a,n}]_{ii}$ (i,i)-ти елемент постојане оцене матрице CSC^T и $z_{\frac{\alpha}{2}} = \Phi^{-1}(1-\frac{\alpha}{2})$. Као и код линеарних модела, природно је да се оцена матрице CSC^T базира на постојаним оценама C и S. Међутим, сада се C_n не може искористити, јер, иако је постојана, вредности $\overline{\theta}_n^{(i)}$ су непознате. Овај проблем је срећом лако решив, коришћењем оцене $\hat{C}_n = \big(G_n(\hat{\theta}_n)^T W_n G_n(\hat{\theta}_n)\big)^{-1} G_n(\hat{\theta}_n)^T W_n$. Међутим, оцењивање S је доста компликованије и захтева посебну дискусију, на коју ћемо се вратити касније.

3.4.3 Оцена узорачког момента

Сада желимо да изведемо асимптотску расподелу за $W_n^{\frac{1}{2}}\sqrt{n}g_n(\hat{\theta}_n)$. Вратимо се поново на развој

$$g_n(\hat{\theta}_n) = g_n(\theta_0) + G_n(\hat{\theta}_n, \theta_0, \lambda_n)(\hat{\theta}_n - \theta_0).$$

Ако помножимо са $W_n^{\frac{1}{2}}\sqrt{n}$ слева, добијамо да је

$$W_n^{\frac{1}{2}}\sqrt{n}g_n(\hat{\theta}_n) = W_n^{\frac{1}{2}}\sqrt{n}g_n(\theta_0) + W_n^{\frac{1}{2}}G_n(\hat{\theta}_n, \theta_0, \lambda_n)\sqrt{n}(\hat{\theta}_n - \theta_0).$$

Пошто је $\sqrt{n}(\hat{\theta}_n - \theta_0) = -\big(G_n(\hat{\theta}_n)^T W_n G_n(\hat{\theta}_n, \theta_0, \lambda_n)\big)^{-1} G_n(\hat{\theta}_n)^T W_n \sqrt{n} g_n(\theta_0)$, следи да је

$$W_n^{\frac{1}{2}}\sqrt{n}g_n(\hat{\theta}_n) = T_n W_n^{\frac{1}{2}}\sqrt{n}g_n(\theta_0)$$

где је $T_n = I_q - W_n^{\frac{1}{2}} G_n(\hat{\theta}_n, \theta_0, \lambda_n)\big(G_n(\hat{\theta}_n)^T W_n G_n(\hat{\theta}_n, \theta_0, \lambda_n)\big)^{-1} G_n(\hat{\theta}_n)^T (W_n^{\frac{1}{2}})^T$. Видимо да $W_n^{\frac{1}{2}}\sqrt{n}g_n(\hat{\theta}_n)$ заправо има исту структуру као $\sqrt{n}(\hat{\theta}_n - \theta_0)$ - случајна матрица пута случајни вектор $\sqrt{n}g_n(\theta_0)$ који конвергира нормалној расподели. Дакле, на исти начин можемо извести следећу теорему:

Теорема 3.3: Асимптотска нормалност оцене узорачког момента
Ако важе претпоставке 3.1-3.5, 3.7-3.13, онда
$W_n^{\frac{1}{2}}\sqrt{n}g_n(\hat{\theta}_n) \xrightarrow{D} \mathcal{N}(0, TW^{\frac{1}{2}}S(W^{\frac{1}{2}})^T T^T)$, *где је* $T = (I_q - P(\theta_0))$ *и* $P(\theta_0) = F(\theta_0)(F(\theta_0)^T F(\theta_0))^{-1} F(\theta_0)^T$.

Веза између оцене узорачког момента и $\mathcal{O}$ услова се манифестује у асимптотској расподели. Пошто је $W_n^{\frac{1}{2}}\sqrt{n}g_n(\hat{\theta}_n) = T_n W_n^{\frac{1}{2}}\sqrt{n}g_n(\theta_0)$, следи да је

$$W_n^{\frac{1}{2}}\sqrt{n}g_n(\hat{\theta}_n) = (I_q - P(\theta_0))W^{\frac{1}{2}}\sqrt{n}g_n(\theta_0) + o_p(1).$$

Дакле, асимптотско понашање оцене узорачког момента је одређено функцијом података која се појављује у $\mathcal{O}$ условима.

3.5 Оцењивање S

Да бисмо започели анализу и увели технике оцењивања S, за почетак погледајмо мало боље њену структуру. Дефинисали смо S са[25]

$$S = \lim_{n\to\infty} D(\sqrt{n} g_n(\theta_0)) = \lim_{n\to\infty} D(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} g_t)$$

$$= \lim_{n\to\infty} E\Big((\frac{1}{\sqrt{n}} \sum_{t=1}^{n} g_t - E(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} g_t)) (\frac{1}{\sqrt{n}} \sum_{t=1}^{n} g_t - E(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} g_t))^T \Big)$$

$$= \lim_{n\to\infty} E\Big((\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \big(g_t - E(g_t)\big)) (\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \big(g_t - E(g_t)\big))^T \Big)$$

$$= \lim_{n\to\infty} E\Big((\frac{1}{n} \sum_{t=1}^{n} \sum_{s=1}^{n} (g_t - E(g_t))(g_s - E(g_s))^T \Big).$$

Претпоставка о стационарности имплицира да је $E\big((g_t - E(g_t))(g_{t-j} - E(g_{t-j}))^T\big) = \Gamma_j$ за $\forall t$ и $\forall j \geq 0$, па је

$$S = \Gamma_0 + \lim_{n\to\infty} \Big(\sum_{j=1}^{n-1} (\frac{n-j}{n})(\Gamma_j + \Gamma_j^T) \Big) = \Gamma_0 + \sum_{j=1}^{\infty} (\Gamma_j + \Gamma_j^T).$$

Матрица Γ_j представља j-ту аутоковаријациону матрицу од g_t, па је јасно је да ће оцењивање S захтевати одређене претпоставке о овим аутоковаријационим матрицама. У овом поглављу размотрићемо оцене које су предложене при две различите претпоставке о природи g_t. Прво ћемо посматрати случај где $\{g_t\}$ формира серијски некорелисан процес, а затим се фокусирамо на општији случај који укључује серијску корелисаност. У оквиру ове дискусије, позабавићемо се класом HAC матрица оцена[26].

Серијски некорелисан процес

Ако је $\{g_t\}$ серијски некорелисан процес, онда је $\Gamma_j = 0$ за $j \neq 0$, па је $S = S_{SU} = \Gamma_0 = E(g_t g_t^T)$. Природно, ово нас води до оцене[27]

$$\hat{S}_{SU} = \frac{1}{n} \sum_{t=1}^{n} \hat{g}_t \hat{g}_t^T$$

и може се показати да $\hat{S}_{SU} \xrightarrow{P} S_{SU}$. Приметимо да је матрица $\hat{S}_{SU}$ по конструкцији позитивно семи-дефинитна јер је $\hat{S}_{SU} = \frac{1}{n} G^T G$, где је G $n \times q$ матрица са t-тим

[25] $g_t := g(v_t, \theta_0)$.

[26] Heteroskedasticity and autocorrelation covariance consistent estimators.

[27] $\hat{g}_t := g(v_t, \hat{\theta}_n)$.

редом $\hat{g}_t^T$.

НАС оцене

Сада желимо да конструишемо оцене које ће бити постојане за S у присуству серијске корелисаности. Вратимо се укратко на израз за S: $S = \Gamma_0 + \sum_{j=1}^{\infty}(\Gamma_j + \Gamma_j^T)$. Природно је да покушамо да оценимо S одсецањем бесконачне суме и коришћењем узорачких аутоковаријација за оцењивање њихових популационих аналогона. Ово нас води до оцене

$$\hat{S}_{TR} = \hat{\Gamma}_0 + \sum_{j=1}^{b_n}(\hat{\Gamma}_j + \hat{\Gamma}_j^T)$$

где је $\hat{\Gamma}_j = \frac{1}{n}\sum_{t=j+1}^{n} \hat{g}_t \hat{g}_{t-j}^T$. Показано је да под одређеним условима $\hat{S}_{TR}$ конвергира у вероватноћи позитивно дефинитној матрици, али може бити недефинтна на коначним узорцима, при чему проблем не потиче од одсецања, већ од тежина које су додељене узорачким аутоковаријацијама у $\hat{S}_{TR}$.

Идеја која се крије иза класе НАС матрица је конструисање оцене у којој је утицај узорачких аутоковаријација довољно смањен на коначним узорцима да би важила позитивна семи-дефинитност - ово постижемо додавањем одговарајаћих тежина аутоковаријацијама, које одражавају ово својство, али теже 1 кад $n \to \infty$ да би се осигурала постојаност. НАС класа се састоји од оцена облика

$$\hat{S}_{HAC} = \hat{\Gamma}_0 + \sum_{j=1}^{n-1} w_{jn}(\hat{\Gamma}_j + \hat{\Gamma}_j^T)$$

где је w_{jn} одговарајућа тежина (често се назива и језгром). Језгро се бира тако да обезбеди постојаност и семи-дефинитност, а најпознатија и најчешће коришћена језгра дата су у следећој табели:

Језгро	Подразумевано b_n	w_{jn}		
Језгро одсецања	$[4(\frac{n}{100})^{\frac{1}{5}}]$	1, за $a_j < 1$	0, за $a_j \geq 1$	
Бартлет	$[4(\frac{n}{100})^{\frac{1}{4}}]$	$1 - a_j$, за $a_j \leq 1$	0, за $a_j > 1$	
Парзен	$[4(\frac{n}{100})^{\frac{4}{25}}]$	$1 - 6a_j^2 + 6a_j^3$, за $0 \leq a_j \leq 0.5$	$2(1-a_j)^3$, за $0.5 \leq a_i \leq 1$	0, за $a_j > 1$
Квадратно спектрално	$[4(\frac{n}{100})^{\frac{4}{25}}]$	$\frac{25}{12\pi^2 d_j^2}\left(\frac{\sin(m_j)}{m_j} - \cos(m_j)\right)$		
$a_j = \frac{j}{b_n+1}$, $d_j = \frac{j}{b_n}$, $m_j = \frac{6\pi d_j}{5}$				

Параметар b_n мора бити ненегативан и он контролише колико је аутоковаријација укључено у НАС оцену када се користе језгро одсецања, Бартлет или Парзен језгро. У овим случајевима, b_n мора бити цео број, док код квадратног спектралног језгра не мора бити. Избор језгра и b_n одређује статистичка својства оцене $\hat{S}_{HAC}$.

Наравно, питање које се природно намеће је које би језгро требало користити. Пре свега, видимо да језгро одсецања одговара $\hat{S}_{TR}$, па је $\hat{S}_{TR}$ специјални случај HAC оцене. Међутим, оно се ређе користи, управо због наведених разлога. Вршене су разне симулационе студије у циљу одређивања најпогоднијег језгра, у којима су добијени различити резултати, али општи закључак је да заправо није могуће јасно рангирање. Међутим, закључено је да избор језгара и није нарочито важан, већ да b_n игра много значајнију улогу у понашању $\hat{S}_{HAC}$ на коначним узорцима. За постојаност, b_n мора тежити бесконачности кад $n \to \infty$. Показано је да је асимптотска средње квадратна грешка минимална за $b_n = O(n^{\frac{1}{3}})$ за Бартлет језгро и $O(n^{\frac{1}{5}})$ за Парзен и квадратно спектрално. Међутим, ови резултати и нису од нарочитог практичног значаја, јер само говоре да b_n за нпр. Бартлет језгро треба да буде облика $cn^{\frac{1}{3}}$, за било које коначно $c > 0$. Ово је подстакло развој непараметарске методе за аутоматски избор b_n и показано је да се за овако одабрано b_n минимизује асимптотска средње квадратна грешка[28].

3.6 Оптимални избор тежинске матрице

Видели смо да асимптотска дисперзија $\hat{\theta}_n$ зависи од тежинске матрице (осим у случају када је $p = q$), па закључци могу зависити од избора W. Оптимална тежинска матрица, у смислу да даје најпрецизнију оцену, односно да минимизује дисперзију у матричном смислу, добија се, као и код линеарних модела за $W = S^{-1}$. Сада је право време да покажемо зашто ово важи.

Теорема 3.4: Оптимални избор тежинске матрице
Ако важе претпоставке 3.1-3.5, 3.7-3.13, тада се минимална асимптотска дисперзија[29] за $\hat{\theta}_n$ добија за $W = S^{-1}$.

Доказ:
Нека је $\hat{\theta}_n(W)$ оцена УММ на основу тежинске матрице W_n. Желимо да покажемо да је $D_a(W) - D_a(S^{-1})$ позитивно семи-дефинитна матрица, где је $D_a(W)$ асимптотска дисперзија од $\sqrt{n}(\hat{\theta}_n(W) - \theta_0)$. Пошто је

$$\sqrt{n}(\hat{\theta}_n(W) - \theta_0) = \sqrt{n}(\hat{\theta}_n(S^{-1}) - \theta_0) + \sqrt{n}(\hat{\theta}_n(W) - \hat{\theta}_n(S^{-1}))$$

и

$$\sqrt{n}(\hat{\theta}_n(W) - \theta_0) = -(G_0^T W G_0)^{-1} G_0^T W \sqrt{n} g_n(\theta_0) + o_p(1)$$

следи да је

$$\sqrt{n}(\hat{\theta}_n(W) - \hat{\theta}_n(S^{-1})) = -(M(W) - M(S^{-1}))\sqrt{n} g_n(\theta_0) + o_p(1)$$

где је $M(W) = (G_0^T W G_0)^{-1} G_0^T W$. Када заменимо ово у прву једначину и израчунамо асимптотску дисперзију, добијамо да је

$$D_a(W) = D_a(S^{-1}) + V + C + C^T \implies D_a(W) - D_a(S^{-1}) = V + C + C^T$$

[28] Newey and West метод за избор b_n - детаље нећемо наводити.
[29] $(G_0^T S^{-1} G_0)^{-1}$.

где су

$$V = \lim_{n\to\infty} D((M(W) - M(S^{-1}))\sqrt{n}g_n(\theta_0))$$

и

$$C = \lim_{n\to\infty} cov((M(W) - M(S^{-1}))\sqrt{n}g_n(\theta_0), M(S^{-1})\sqrt{n}g_n(\theta_0)).$$

Треба да покажемо да је $V + C + C^T$ позитивно семи-дефинитна да бисмо имали важење теореме. V је по својој конструкцији позитивно семи-дефинитна, а

$$C = \lim_{n\to\infty} E\big((M(W) - M(S^{-1}))\sqrt{n}g_n(\theta_0)\sqrt{n}g_n(\theta_0)^T M(S^{-1})^T\big)$$

$$= (M(W) - M(S^{-1})) \lim_{n\to\infty} E\big(\sqrt{n}g_n(\theta_0)\sqrt{n}g_n(\theta_0)^T\big)M(S^{-1})^T$$

$$= M(W)SM(S^{-1})^T - M(S^{-1})SM(S^{-1})^T = 0.$$

Теорема нам говори да је оптимални избор за W_n матрица $\hat{S}_n^{-1}$, где је $\hat{S}_n$ постојана оцена за S. Дакле, као и код линеарних модела, да бисмо израчунали оптималну оцену параметра, потребна су нам бар два корака - у првом користимо било који валидан избор тежинске матрице W_n да добијемо $\hat{\theta}_n^{(1)}$, а онда њу користимо да добијемо постојану оцену за S, $\hat{S}_n^{(1)}$. У другом кораку користимо $W_n = (\hat{S}_n^{(1)})^{-1}$ да добијемо оцену $\hat{\theta}_n^{(2)}$, која ће, на основу претходне теореме, бити оптимална. Добијена оцена се назива *(оптималном) двостепеном* оценом. Међутим, пошто је оцена у првом кораку израчуната на основу произвољног избора тежинске матрице, очекујемо да бисмо можда могли имати бољу апроксимацију на коначним узорцима ако наставимо овај поступак - односно рачунамо $\hat{\theta}_n^{(3)}$ за $W_n = (\hat{S}_n^{(2)})^{-1}$, $\hat{\theta}_n^{(4)}$ за $W_n = (\hat{S}_n^{(3)})^{-1}$, итд. Све ове оцене ће имати исту асимптотску расподелу као $\hat{\theta}_n^{(2)}$, али очекујемо да буду ефикасније на коначним узорцима. Коначан резултат називамо *(оптималном)*[30] *итерираном* оценом. Дакле, итерирано оцењивање се спроводи на следећи начин. У i-том кораку:

Ако је $i = 1$: Оцењујемо θ_0 на основу произвољног избора тежинске матрице која задовољава тражене услове - добијену оцену $\hat{\theta}_n^{(1)}$ користимо да добијемо постојану оцену за S неком од описаних метода - $\hat{S}_n^{(1)}$.

Ако је $i > 1$: Оцењујемо θ_0 на основу $W_n = (\hat{S}_n^{(i-1)})^{-1}$ где је $\hat{S}_n^{(i-1)}$ постојана оцена за S израчуната на основу $\hat{\theta}_n^{(i-1)}$. Ако је $||\hat{\theta}_n^{(i)} - \hat{\theta}_n^{(i-1)}|| < \epsilon$, онда кажемо да је процедура исконвергирала и као оцену за θ_0 узимамо $\hat{\theta}_n^{(i)}$. Ако је $||\hat{\theta}_n^{(i)} - \hat{\theta}_n^{(i-1)}|| \geq \epsilon$ и $i < I_{max}$, где је I_{max} максималан број корака[31], онда се прелази на $(i+1)$-и корак.

[30] Оптималност се односи само на избор тежинске матрице, тј. ово су најпрецизније оцене које се могу конструисати на основу датог популационог услова момената $E(g(v_t, \theta_0)) = 0$. Оне не говоре ништа о оптималности самог популационог услова момената.

[31] Потребно је навести га, јер у пракси нема гаранције да ће метод конвергирати, па можемо упасти у бесконачну петљу. Међутим, без обзира да ли се конвергенција деси или не пре одабраног I_{max}, све $\{\hat{\theta}^{(i)}, i \geq 2\}$ имају исту асимптотску расподелу.

3.7 Закључак

На крају, присетимо се седам кључних аспеката уопштене методе момената које смо формулисали код линеарних модела. Основна својства преносе се са овог једноставнијег случаја на општи оквир, међутим, због присуства нелинеарности, анализа је доста сложенија.

1) **Идентификација:** Да би оцењивање било успешно, популациони услов момената мора бити тачан и мора обезбедити довољно информација да се јединствено одреди вектор параметара. Код нелинеарних модела, доста је сложеније идентификовати параметар него код линеарних модела, па је неопходно увођење појмова локалне и глобалне идентификације.

2) **$\mathcal{I}$ и $\mathcal{O}$ услови:** Оцењивање код преидентификованих модела укључује декомпозицију услова момената на $\mathcal{I}$ и $\mathcal{O}$ услове. $\mathcal{I}$ услови садрже информације које се користе у оцењивању, а $\mathcal{O}$ услови представљају остатак.

3) **Оцена параметра:** Обично не можемо добити експлицитни израз за оцену параметра због присуства нелинеарности, па се оцена добија нумеричким методама. Оцена параметра је постојана и има граничну нормалну расподелу[32].

4) **Оцена узорачког момента:** Оцена узорачког момента има граничну нормалну расподелу, чије особине директно зависе од функције података у $\mathcal{O}$ условима.

5) **Оцењивање асимптотске коваријационе матрице:** Да бисмо могли да изводимо практичне закључке на основу асимптотске нормалности, неопходно је да постојано оценимо асимптотску дисперзију узорачког момента. Да бисмо конструисали погодну оцену, морамо да направимо одређене претпоставке о структури зависности $\{g(v_t, \theta_0)\}$, функције података која се појављује у популационом услову момената. Разматрамо случај када $\{g(v_t, \theta_0)\}$ чини серијски некорелисан процес, као и општији случај када се користе такозване HAC оцене.

6) **Оптимални избор тежинске матрице:** Оптимални избор тежинске матрице зависи од асимптотске дисперзије узорачког момента, па је потребно спроводити двостепено или итерирано оцењивање.

7) **Дијагностика модела:** J-тест обезбеђује базу за тестирање коректности спецификације модела коришћењем оцене узорачког момента (ово ћемо показати касније).

[32]Када је скалирана на одговарајући начин.

4 УММ у погрешно спецификованим моделима

Сви закључци изведени до сад почивали су на претпоставци да је модел коректно спецификован. Међутим, у пракси је тешко знати да ли претпостављени модел одговара правом стању, па је важно да размотримо утицај погрешне спецификације на статистичка својства оцена.

Прво, дефинишимо погрешну спецификацију (ПС). Означимо статистички модел који се састоји од претпоставки о процесу генерисања података v_t са $\mathcal{M}$. Овај модел нам даје скуп популационих услова момената који се могу користити као база у оцењивању θ_0 уопштеном методом момената, тј.

$$\mathcal{M} \implies E(g(v_t, \theta_0)) = 0, \forall t, \text{ за неко јединствено } \theta_0 \in \Theta.$$

Ако посумњамо да $\mathcal{M}$ није тачан модел, намећу се два алтернативна сценарија. Прво, прави модел, $\mathcal{M}_1$, задовољава исти популациони услов момената у некој другој тачки параметарског простора, тј.

$$\mathcal{M}_1 \implies E(g(v_t, \theta_*)) = 0, \forall t, \text{ за неко јединствено } \theta_* \in \Theta.$$

Асимптотска својства оцене параметра и оцене узорачког момента су суштински иста при $\mathcal{M}$ и $\mathcal{M}_1$ - једина разлика је у тачки у којој је задовољен популациони услов момената. Зато се обично термин ПС користи да означи други случај, да је прави модел $\mathcal{M}_2$, који има особину

$$\mathcal{M}_2 \implies \nexists \theta \in \Theta \text{ такво да је } E(g(v_t, \theta)) = 0, \text{ за } \forall t.$$

$\mathcal{M}$ и $\mathcal{M}_2$ имају потпуно различите претпоставке о $E(g(v_t, \theta))$ и то се одражава на понашање оцене параметра и оцене узорачког момента. Задржавамо претпоставку да је v_t стационаран процес, па $E(g(v_t, \theta))$ не зависи од t, тј. ограничићемо се на следећу класу ПС модела:

Претпоставка 4.1: Природа ПС
$E(g(v_t, \theta)) = \mu(\theta)$, за $\forall t$, где је $||\mu(\theta)|| > 0$, за $\forall \theta \in \Theta$.[33]

У пракси се закључак обично изводи на основу двостепене или итериране оцене и то је приступ који ћемо пратити овде. Њихова главна особина је, присетимо се, то што се за оцењивање у i-том кораку[34] као тежинска матрица користи инверз оцене коваријационе матрице израчунате на основу оцене $\hat{\theta}_n^{(i-1)}$ из $(i-1)$-ог корака, па се утицај ПС преноси са сваког корака на следећи, те да бисмо испитали утицај ПС на итерирану оцену, неопходно је да испитамо сваки корак

[33] Дакле, избацујемо случајеве у којима је $E(g(v_t, \theta)) = \mu_t$. Ово свакако ограничава општост анализе, међутим, умногоме олакшава прелаз са тачно спецификованих на ПС моделе и омогућава нам да једноставније уочимо разлике између та два случаја.

[34] Код двостепене само за $i = 2$, код итериране за $2 \leq i \leq I_{max}$.

посебно. Према томе, мора се почети испитивањем утицаја ПС на оцену у првом кораку. Испоставиће се да ПС значајно компликује анализу граничне расподеле, јер брзина конвергенције $\hat{\theta}_n^{(1)}$ ка својој граничној вредности $\theta_*^{(1)}$ зависи од брзине конвергенције W_n ка W, па у неким случајевима $\sqrt{n}(\hat{\theta}_n^{(1)} - \theta_*^{(1)})$ не конвергира у расподели, док $\sqrt{n}g_n(\hat{\theta}_n^{(i)})$ дивергира, без обзира на брзину конвергенције тежинске матрице.

У првом кораку, оцена се може рачунати за било који избор тежинске матрице која је позитивно семи-дефинитна и тежи у вероватноћи позитивно дефинитној матрици. Показали смо да у коректно спецификованим моделима таква оцена (при одређеним условима регуларности) конвергира у вероватноћи ка θ_0. У ПС моделима, $\hat{\theta}_n^{(1)}$ ће такође конвергирати у вероватноћи неком $\theta_*^{(1)} \in \Theta$. Међутим, ако рачунамо итерирану оцену, лимеси у вероватноћи се сада могу разликовати у сваком кораку.

Као што смо већ рекли, када се ради о граничној расподели оцене, за разлику од коректно спецификованих модела код којих $\sqrt{n}(\hat{\theta}_n - \theta_0)$ конвергира нормалној расподели, у случају ПС анализа је доста компликованија, јер тежинска матрица игра много значајнију улогу. Ово не важи само за оцену у првом кораку, већ је ова зависност присутна у сваком кораку оцењивања.

Да бисмо испитали утицај ПС на двостепену и итерирану оцену, морамо прво да видимо како ПС утиче на оцењивање асимптотске коваријационе матрице узорачког момента.

Код коректно спецификованих модела, навели смо постојане оцене које се могу користити у два различита случаја. Међутим, кључна ствар је била што је модел коректно спецификован. Ако је модел ПС, ниједна од предложених оцена неће бити постојана. Срећом, оне се могу лако модификовати тако да се осигура постојаност и у ПС моделима.

Посматрајмо, за почетак, једну аутоковаријациону матрицу. По дефиницији, j-та аутоковаријациона матрица од $g(v_t, \theta_*)$ је[35]

$$\Gamma_j = E\big((g(v_t,\theta_*) - \mu_*)(g(v_{t-j},\theta_*) - \mu_*)^T\big) = E\big(g(v_t,\theta_*)g(v_{t-j},\theta_*)^T\big) - \mu_*\mu_*^T.$$

Ако бисмо оценили Γ_j истом оценом као раније,

$$\hat{\Gamma}_j = \frac{1}{n}\sum_{t=j+1}^{n} g(v_t,\hat{\theta}_n)g(v_{t-j},\hat{\theta}_n)^T$$

не бисмо добили постојану оцену Γ_j, већ само првог члана разлике, јер је $\mu_* \neq 0$.

[35] $\mu_* := E(g(v_t,\theta_*))$.

Међутим, природно се намеће постојана оцена за Γ_j

$$\tilde{\Gamma}_j = \frac{1}{n}\sum_{t=j+1}^{n}(g(v_t,\hat{\theta}_n) - g_n(\hat{\theta}_n))(g(v_{t-j},\hat{\theta}_n) - g_n(\hat{\theta}_n))^T.$$

Кључна разлика између ове две оцене је што су подаци $\{g(v_t,\hat{\theta}_n)\}$ *центрирани* око средње вредности $g_n(\hat{\theta}_n)$ код $\tilde{\Gamma}_j$, а *нецентрирани* код $\hat{\Gamma}_j$. Ове придеве ћемо користити и да разликујемо оцене коваријационе матрице засноване на центрираним или нецентрираним аутоковаријацијама.

Ако је $\{g(v_t,\theta_*)\}$ серијски некорелисан процес, онда је $\Gamma_j = 0, \forall j \neq 0$, тј. $S_* = \Gamma_0$. Пошто је $\hat{S}_{SU} = \hat{\Gamma}_0$, следи да $\hat{S}_{SU} \xrightarrow{P} S_* + \mu_*\mu_*^T$. Лако добијамо постојану оцену за S_*

$$\hat{S}_{SU,\mu} = \frac{1}{n}\sum_{t=1}^{n}(g(v_t,\hat{\theta}_n) - g_n(\hat{\theta}_n))(g(v_t,\hat{\theta}_n) - g_n(\hat{\theta}_n))^T.$$

Сада размотримо утицај ПС на НАС оцене. Нецентрирана оцена је дата са

$$\hat{S}_{HAC} = \hat{\Gamma}_0 + \sum_{j=1}^{n-1} w_{jn}(\hat{\Gamma}_j + \hat{\Gamma}_j^T).$$

Ова оцена је, за погодно изабране w_{jn} и b_n, постојана код коректно спецификованих модела. Међутим, то је потицало од чињенице да је $\hat{\Gamma}_j$ постојана оцена за Γ_j, а пошто то више није случај, ни ова оцена више није постојана. Поново, лако ју је модификовати тако да добијемо постојану оцену - ово нас води до центриране НАС оцене

$$\hat{S}_{HAC,\mu} = \tilde{\Gamma}_0 + \sum_{j=1}^{n-1} w_{jn}(\tilde{\Gamma}_j + \tilde{\Gamma}_j^T).$$

Сада ћемо укратко навести резултате о утицају ПС на двостепену и итерирану оцену. Утицај зависи од оцене коваријационе матрице која се користи, па морамо размотрити два случаја - први, када користимо $\hat{S}_{SU}$ или $\hat{S}_{SU,\mu}$ да конструишемо тежинску матрицу и други - када користимо $\hat{S}_{HAC}$ или $\hat{S}_{HAC,\mu}$. Испоставља се да се понашање ових оцена веома разликује, као и да се разликује од понашања оцене у првом кораку.

Размотримо прво случај када је $W_n = \hat{S}_{SU}^{-1}$ или $W_n = \hat{S}_{SU,\mu}^{-1}$. Испоставља се да је у општем случају $\theta_*^{(2,n)} \neq \theta_*^{(2,c)}$, где $\theta_*^{(2,n)}$ означава лимес у вероватноћи оцене у другом кораку ако се користи инверз нецентриране оцене као тежинска матрица, а $\theta_*^{(2,c)}$ ако се користи инверз центриране оцене. Такође, $\theta_*^{(2,n)} \neq \theta_*^{(1)}$ и $\theta_*^{(2,c)} \neq \theta_*^{(1)}$, где је $\theta_*^{(1)}$ лимес у вероватноћи од $\hat{\theta}_n^{(1)}$.

Међутим, без обзира да ли се користи центрирана или нецентрирана оцена, могуће је показати да двостепена оцена има граничну нормалну расподелу при одговарајућим условима, али асимптотско понашање $\sqrt{n}(\hat{\theta}_n^{(2)} - \theta_*^{(2)})$ зависи од асимптотског понашања $\sqrt{n}(\hat{\theta}_n^{(1)} - \theta_*^{(1)})$, па ће у итерираном оцењивању асимптотско понашање $\sqrt{n}(\hat{\theta}_n^{(i)} - \theta_*^{(i)})$ зависити од $\{\sqrt{n}(\hat{\theta}_n^{(j)} - \theta_*^{(j)}),\ j = \overline{1, i-1}\}$ и ова рекурзивна структура се мора узети у обзир при извођењу закључака.

Ако се HAC оцене користе за конструисање тежинске матрице, испоставља се да се гранична расподела доста мења у односу на разматране случајеве, јер оцена више не конвергира брзином $\frac{1}{\sqrt{n}}$.

Ако је $W_n = \hat{S}_{HAC,\mu}^{-1}$, слично као за оцену у првом кораку, може се показати да $\hat{\theta}_n^{(2)}$ конвергира у вероватноћи некој вредности $\theta_*^{(2)}$, али је, у општем случају, $\theta_*^{(2)} \neq \theta_*^{(1)}$.

Рекли смо да гранична расподела зависи од брзине конвергенције W_n ка W. Испоставља се да $\sqrt{n}(\hat{S}_{HAC,\mu}^{-1} - (S_*^{(1)})^{-1})$ дивергира кад $n \to \infty$, па исто важи и за $\sqrt{n}(\hat{\theta}_n^{(2)} - \theta_*^{(2)})$. Дакле, да бисмо израчунали граничну расподелу, морамо скалирати $\hat{\theta}_n^{(2)} - \theta_*^{(2)}$ неком функцијом c_n, таквом да $c_n \to \infty$, кад $n \to \infty$ и $c_n = o(\sqrt{n})$. Избор c_n и гранично понашање $c_n(\hat{S}_{HAC,\mu}^{-1} - (S_*^{(1)})^{-1})$ (а стога и $c_n(\hat{\theta}_n^{(2)} - \theta_*^{(2)})$) зависе од избора b_n и језгара. У неким случајевима се може добити гранична нормална расподела, али прецизне детаље сада нећемо наводити.
Оно што је важно је да, за разлику од претходног случаја, асимптотска расподела двостепене оцене не зависи од оцене из првог корака, међутим, ово се не преноси и на итерирану оцену. Испоставља се да асимптотско понашање $c_n(\hat{\theta}_n^{(i)} - \theta_*^{(i)})$ зависи од $\{c_n(\hat{\theta}_n^{(j)} - \theta_*^{(j)})\ j = \overline{2, n-1}\}$ у општем случају (сада зависност иде уназад само до другог корака, због наведеног разлога).

Ако је $W_n = \hat{S}_{HAC}^{-1}$, оцена параметра конвергира истој вредности и у првом и у другом кораку оцењивања, а ово се проширује и на итериране оцене. Према томе, за овакав одабир W_n, лимеси у вероватноћи се понашају као код коректно спецификованих модела.
Брзина конвергенције је мања од $\sqrt{n}$, па се поново рачуна асимптотско понашање $c_n(\hat{\theta}_n - \theta_*)$, где избор c_n зависи од језгара и b_n, а гранична расподела од c_n.

Остало је још да испитамо утицај ПС на оцену узорачког момента. Анализа је веома једноставна и не зависи од тежинске матрице. Да бисмо спровели анализу, потребно је да дефинишемо СЗВБ, који ће нам бити потребан:

Лема 4.5: СЗВБ
Ако је $\theta \in \Theta$ такво да је $E(g(v_t, \theta)) = \mu(\theta)$ и важе претпоставке 3.1, 3.2, 3.8 и 3.10,
онда $\frac{1}{n}\sum_{t=1}^{n} g(v_t, \theta) \xrightarrow{P} \mu(\theta)$.

Теорема 4.3: Асимптотско понашање оцене узорачког момента
Ако: 1) *важе претпоставке 3.1, 3.2, 3.8-3.10, 4.1;*
2) $\frac{dg(v_t,\theta)}{d\theta^T}$ *задовољава услове регуларности у* $\theta = \theta_*$;
3) $\hat{\theta}_n \xrightarrow{P} \theta_* \in \Theta$, *тада:* $g_n(\hat{\theta}_n) \xrightarrow{P} \mu(\theta_*)$, *где је* $||\mu(\theta_*)|| > 0$.
Доказ:

$$g_n(\hat{\theta}_n) = g_n(\theta_*) + G_n(\hat{\theta}_n, \theta_*, \lambda_n)(\hat{\theta}_n - \theta_*)$$

При наведеним условима $G_n(\hat{\theta}_n, \theta_*, \lambda_n) \xrightarrow{P} G_* = O(1)$, $\hat{\theta}_n - \theta_* \xrightarrow{P} 0$, $g_n(\theta_*) \xrightarrow{P} \mu(\theta_*)$, *а* $||\mu(\theta)|| > 0$, *за* $\forall\theta \in \Theta$, *из чега следи тврђење теореме.*

Дакле, ако је модел коректно спецификован, $\sqrt{n}g_n(\hat{\theta}_n)$ конвергира нормалној расподели са очекивањем 0, али ако је ПС, онда дивергира.

На крају, корисно је да упоредимо закључке добијене код коректно спецификованих и ПС модела (претпостављамо да је $p < q$):
Особине УММ у коректно спецификованим моделима:
1) Популациони услов момената важи за $\forall t$.
2) $\hat{\theta}_n \xrightarrow{P} \theta_0$, за било који избор позитивно семи-дефинитне W_n, такве да $W_n \xrightarrow{P} W$, која је позитивно дефинитна.
3) $\sqrt{n}(\hat{\theta}_n - \theta_0)$ конвергира нормалној расподели и избор тежинске матрице утиче на дисперзију ове расподеле са W.
4) Двостепена и итерирана оцена имају иста асимптотска својства.
5) $\sqrt{n}g_n(\hat{\theta}_n)$ конвергира нормалној расподели са очекивањем 0.
6) Неопходно је да знамо само структуру зависности g_t да бисмо конструисали постојану оцену асимптотске коваријационе матрице узорачког момента.

Особине УММ у ПС моделима:
1) Популациони услов момената није задовољен.
2) У општем случају $\lim_{\mathrm{p}_{n\to\infty}} \hat{\theta}_n$ зависи од W.
3) Брзина конвергенције $\hat{\theta}_n$ ка свом лимесу θ_* зависи од брзине конвергенције W_n ка W, а гранична расподела $c_n(\hat{\theta}_n - \theta_0)$ зависи од граничне расподеле $c_n(W_n - W)$.
4) Двостепена и итерирана оцена имају различита асимптотска својства и асимптотска расподела $\hat{\theta}_n^{(i)}$ зависи од оцена из претходних корака (осим у случају кад је $W_n = \hat{S}_{HAC}^{-1}$).
5) $\sqrt{n}g_n(\hat{\theta}_n)$ дивергира.
6) При конструкцији постојане оцене асимптотске коваријационе матрице мора се узети у обзир и структура зависности g_t и не-нула очекивање.

Дакле, ако имамо ПС закључци у већини случајева нису добри, што даје мотивацију за конструисање тестова спецификације модела, које ћемо описати у наредном поглављу.

5 Тестирање хипотеза

Ако желимо да користимо УММ за закључивање, морамо прво размотрити два питања - да ли је модел коректно спецификован и да ли модел задовољава рестрикције које имплицира статистичка или економска теорија. У оквиру уопштене методе момената, овим питањима се приступа тестирањем хипотеза које се заснивају на популационом услову момената и/или вектору параметара. У пракси, закључци се обично изводе на основу двостепене или итериране оцене, па је то приступ који ћемо пратити овде. Такође, претпостављамо да је $p < q$.

5.1 Тест $\mathcal{O}$ услова (J-тест)

У претходном поглављу смо видели да ПС може учинити све раније изведене закључке неважећим, па је неопходно конструисати тест који ће нам помоћи да одредимо да ли је модел коректно спецификован. Код линеарних модела смо увели идеју коришћења $\mathcal{O}$ услова за испитивање ПС, а ова идеја се може пренети и на општији случај.

Код линеарних модела, идеја је била да, ако је $E(z_t e_t(\theta_0)) = 0$, онда би то требало да важи и за оцену узорачког момента, односно да $\frac{1}{n} Z^T e(\hat{\theta}_n)$ буде приближно нула. Исту идеју преносимо на нелинеарне моделе - ако је $E(g(v_t, \theta_0)) = 0$, онда би $g_n(\hat{\theta}_n)$ требало да буде приближно једнако нули. Предложено је да се нулта хипотеза од интереса

$$H_0 : E(g(v_t, \theta_0)) = 0$$

тестира коришћењем тест статистике

$$J_n = nQ_n(\hat{\theta}_n) = ng_n(\hat{\theta}_n)^T \hat{S}_n^{-1} g_n(\hat{\theta}_n)$$

где је $\hat{\theta}_n$ двостепена или итерирана оцена.

Важи следећа теорема о граничној расподели тест статистике при H_0:

Теорема 5.1: Асимптотска расподела J-статистике при H_0
Ако важе претпоставке 3.1-3.5, 3.8-3.13, $\hat{S}_n$ је позитивно семи-дефинитна и $\hat{S}_n \xrightarrow{P} S$, тада: $J_n \xrightarrow{D} \chi^2_{q-p}$.

5.2 Тестирање хипотеза о вектору параметара

У многим ситуацијама економска или статистичка теорија поставља одређене рестрикције на вектор параметара, па је могуће испитати тачност претпостављене теорије испитивањем да ли су ти услови задовољени на подацима.
Претпостављамо да су подаци генерисани моделом $\mathcal{M}$, таквим да

$$\mathcal{M} \implies E(g(v_t, \theta_0)) = 0, \text{ за неко јединствено } \theta_0 \in \Theta.$$

Оно што нас интересује је да ли су подаци генерисани моделом $\mathcal{M}_r \subset \mathcal{M}$, таквим да

$$\mathcal{M}_r \implies E(g(v_t, \theta_0)) = 0, \text{ за неко јединствено } \theta_0 \in \Theta_r \subset \Theta$$

где је $\Theta_r = \{\theta : r(\theta) = 0\}$, а $r(\theta_0)$ је $s \times 1$ вектор нелинеарних функција по θ_0. Према томе, питање од интереса је да ли је θ_0 у Θ_r или Θ_r^c. Вектор $r(.)$ мора задовољавати следећа својства:

Претпоставка 5.3: Услови регуларности за $r(.)$
$r : \mathbb{R}^p \to \mathbb{R}^s$ је $s \times 1$ вектор функција такав да:
1) $r(.)$ је вектор непрекидно диференцијабилних функција;
2) $rank(R(\theta_0)) = s$, где је $R(\theta) = \frac{dr(\theta)}{d\theta^T}$.

За тестирање

$$H_0 : r(\theta_0) = 0 \text{ против } H_1 : r(\theta_0) \neq 0$$

предложене су три статистике засноване на рестрикованој оцени од θ_0, у ознаци $\hat{\theta}_n^{(r)}$, што је вредност која минимизује $Q_n(\theta)$ уз услов $r(\theta) = 0$ и ,,комплетној" оцени од θ_0, $\hat{\theta}_n$, дефинисаној раније[36]. Први предлог је Валдова статистика

$$W_n = nr(\hat{\theta}_n)^T \big(R(\hat{\theta}_n)(G_n(\hat{\theta}_n)^T \hat{S_n}^{-1} G_n(\hat{\theta}_n))^{-1} R(\hat{\theta}_n)^T\big)^{-1} r(\hat{\theta}_n).$$

Други предлог је скор тест

$$S_n = ng_n(\hat{\theta}_n^{(r)})^T \hat{S}_n^{-1} G_n(\hat{\theta}_n^{(r)}) \big(G_n(\hat{\theta}_n^{(r)})^T \hat{S_n}^{-1} G_n(\hat{\theta}_n^{(r)})\big)^{-1} G_n(\hat{\theta}_n^{(r)})^T \hat{S}_n^{-1} g_n(\hat{\theta}_n^{(r)}).$$

И на крају, D-тест, са статистиком

$$D_n = n(Q_n(\hat{\theta}_n^{(r)}) - Q_n(\hat{\theta}_n)).$$

Показано је да су све три статистике асимптотски еквивалентне при H_0, односно да важи следећа теорема:

Теорема 5.6: Асомптотска еквивалентност W_n***,*** S_n ***и*** D_n ***при*** H_0
Ако: 1) *важе претпоставке 3.1-3.5, 3.7-3.13 и 5.3;*
2) $\hat{S}_n^{-1} \xrightarrow{P} S^{-1}$, *тада при* H_0:[37]
$W_n = T_n + o_p(1)$, $S_n = T_n + o_p(1)$ *и* $D_n = T_n + o_p(1)$
где је $T_n = \nu_n^T V^{-1} \nu_n$, $\nu_n = R_0(G_0^T S^{-1} G_0)^{-1} G_0^T S^{-1} \sqrt{n} g_n(\theta_0)$, *а* $V = R_0(G_0^T S^{-1} G_0)^{-1} R_0^T$.

[36] Претпостављамо да се у оба случаја користи тежинска матрица $\hat{S}_n^{-1}$.
[37] $R_0 := R(\theta_0)$.

Ова теорема нам говори да се све три статистике у лимесу понашају као T_n при H_0. Под овим условима важи и да $\nu_n \xrightarrow{D} \mathcal{N}(0, V)$, па добијамо следећи резултат о расподели при H_0:

Теорема 5.7: Гранично понашање W_n, S_n и D_n при H_0
Ако важе услови претходне теореме, $W_n \xrightarrow{D} \chi^2_s$, $S_n \xrightarrow{D} \chi^2_s$ и $D_n \xrightarrow{D} \chi^2_s$ при H_0 кад $n \to \infty$.

С обзиром да су ове статистике асимптотски еквивалентне, природно се намеће питање коју одабрати. D-тест је мало рачунски компликованији, јер захтева два оцењивања, док је за Валдов и скор тест довољно по једно. Валдов тест има две мане у односу на остала два, јер није инваријантан на репараметризацију модела или рестрикција параметара, док остала два јесу, а такође, испоставља се да χ^2_s код њега лошије апроксимира расподелу на коначним узорцима, него код преостала два. Ипак, не постоји униформно правило које говори који је тест најбољи, све зависи од конкретног проблема.

Важно је напоменути још једну ствар у вези са овим тестовима. Током анализе у овом поглављу, претпоставили смо да важи да је $E(g(v_t, \theta_0)) = 0$. Међутим, ако ово не важи, то може такође довести до одбацивања нулте хипотезе, тј. хипотеза може бити одбачена јер важи популациони услов момената, али је $r(\theta_0) \neq 0$, или због ПС модела. Показано је да Валдов, скор и D-тест не конвергирају χ^2_s расподели у ПС моделима, чак и ако су рестрикције на параметре задовољене. Овај резултат наглашава важност коришћења теста спецификације модела пре закључивања о параметрима.

6 Понашање на коначним узорцима

До сада се сва анализа заснивала на асимптотској теорији, међутим, иако нам асимптотска теорија даје моћан оквир закључивања, знамо да су добијени резултати тачни само у лимесу, односно када $n \to \infty$, па представљају апроксимацију понашања на коначним узорцима. Питање које се природно намеће је колико је апроксимација добра. Испоставља се да се ово разликује од случаја до случаја.

Овом испитивању се приступило кроз симулационе студије које одговарају моделима који се срећу у пракси. Велики број студија испитује последице повећања броја услова момената и показано је да коначан пораст степена преидентификације никада нема штетан утицај на асимптотско понашање оцене. Међутим, у неким случајевима, додавање нових услова не доводи ни до каквог побољшања, па се ови услови момената називају *сувишнима*.

Након симулационих студија, добијени су различити резултати. У неким моделима од интереса, апроксимација је добра и за $n = 100$, а у некима не ваља чак ни за $n = 10000$. Такође, симулациони докази говоре да укључење сувишних елемената понекад може водити погоршању квалитета апроксимације на коначним узорцима. Из симулација је заправо изведен само један јасан заључак - понашање на коначним узорцима је много комплексније него оно које предвиђа асимптотска теорија.

Пошто је избор популационих услова момената кључан за перформансе методе, развијени су критеријуми одабира момената који омогућавају да се одреди најбољи члан из скупа кандидата. Међутим, иако селекција момената може водити побољшању, квалитет асимптотске апроксимације може и даље бити лош. За побољшање апроксимације понашања оцене параметра и повезаних статистика на коначним узорцима развијене су одређене алтернативне методе које дају прецизније закључке, а такође, користи се и Бутстреп.

7 Примери у R-у

Пример 1: Нормална расподела

Претпоставимо да желимо да оценимо непознате параметре $\mathcal{N}(m, \sigma^2)$ расподеле - m и σ. Иако ово није пример у коме бисмо користили УММ (јер знамо расподелу, па можемо користити ММВ[38]), корисно је да размотримо овај пример у овом оквиру, јер је јако једноставан и омогућава нам да лако уочимо основне идеје методе. Размотрићемо оцењивање на основу прва три момента расподеле, па ће ово уједно бити и пример који ће нам приказати слабост методе када услови момената нису довољно информативни. Видећемо да ММВ даје ефикасније оцене, што се и могло очекивати, јер функција веродостојности пружа више информација од неколико услова о моментима расподеле. Дефинишемо следећи скуп популационих услова момената

$$E(g(v_t, \theta)) = 0$$

где је $\theta = (m, \sigma)^T$, а

$$g(v_t, \theta) = \begin{pmatrix} m - v_t \\ \sigma^2 - (v_t - m)^2 \\ v_t^3 - m(m^2 + 3\sigma^2) \end{pmatrix}.$$

За рад са уопштеном методом момената у R-у користи се пакет *gmm*. Најважнија функција је *gmm*() која креира објекат класе *gmm*, а чији су главни аргументи g и x, где је g функција која се јавља у популационом услову момената, а x представља податке (у линеарном случају, g је формула, а x су инструменти). Остале опције и аргументе увешћемо у примерима, како нам буду били потребни. Дакле, прво формирамо функцију g:

```
library(gmm)

g = function(theta, v) {
  # theta = (m, sigma)^T
  # v = (v_t)

  m1 = theta[1]-v
  m2 = theta[2]^2 - (v-theta[1])^2
  m3 = v^3 - theta[1]*(theta[1]^2 + 3*theta[2]^2)

  m = cbind(m1, m2, m3)
  return(m)
}
```

Сада ћемо генерисати бројеве из $\mathcal{N}(3, 2)$ расподеле, који ће нам играти улогу података:

```
set.seed(11)
n = 100
v = rnorm(n, 3, sqrt(2))
```

[38] Што је, са друге стране, опет УММ, али сада ћемо користити друге услове момената, а не оне који одговарају ММВ, ради илустрације.

Прелазимо на оцењивање. Параметре оцењујемо уопштеном методом момената, користећи почетне вредности $(0,1)$[39]. Детаљне информације о моделу нам пружа функција *summary*():

```
t0 = c(0, 1)
gmm.fit = gmm(g, v, t0)
summary(gmm.fit)
```

На излазу добијамо следеће:

```
Call:
gmm(g = g, x = v, t0 = t0)


Method:  twoStep

Kernel:  Quadratic Spectral(with bw =  0.92956 )

Coefficients:
           Estimate     Std. Error    t value      Pr(>|t|)
Theta[1]   2.7555e+00   1.0021e-01   2.7496e+01   1.9386e-166
Theta[2]   1.2355e+00   6.4971e-02   1.9017e+01    1.2320e-80

J-Test: degrees of freedom is 1
                 J-test   P-value
Test E(g)=0:     1.72204  0.18943

Initial values of the coefficients
Theta[1] Theta[2]
2.847150 1.289139

#############
Information related to the numerical optimization
Convergence code =  0
Function eval. =  55
Gradian eval. =  NA
```

Прокоментаришимо укратко добијене резултате. Видимо да је подразумевано да се рачуна двостепена оцена са квадратним спектралним језгром. На излазу добијамо оцене параметара, као и резултате тестирања спецификације модела J-тестом. Мала вредност тест статистике говори нам да је модел одговарајући. Такође, видимо да оцене параметара не одступају много од популационих вредности.
Основне информације можемо добити позивањем функције *print*():

```
print(gmm.fit)
```

```
Method
 twoStep

```

[39]Почетне вредности је неопходно навести само ако је g функција јер се онда користе нумеричке методе за минимизацију и добијање оцена.

```
Objective function value:  0.0172204

Theta[1]  Theta[2]
  2.7555    1.2355

Convergence code =  0
```

Такође, можемо да издвојимо неке важне компоненте, попут оцена параметара и резултата J-теста:

```
coef(gmm.fit)
```

```
 Theta[1] Theta[2]
2.755459 1.235548
```

```
specTest(gmm.fit)
```

```
##  J-Test: degrees of freedom is 1  ##

                    J-test    P-value
Test E(g)=0:      1.72204   0.18943
```

Број степени слободе је 1, јер имамо 3 услова, а 2 непозната параметра. Можемо израчунати и инетрвале поверења за параметре:

```
confint(gmm.fit)
```

```
Wald type confidence interval
#############################################
             0.025     0.975
Theta[1]   2.5590   2.9519
Theta[2]   1.1082   1.3629
```

Видимо да нису успели да покрију праве вредности параметара. Пошто се оцењивање подразумевано врши коришћењем HAC оцене са квадратним спектралним језгром, можемо да погледамо шта се дешава ако покушамо са другим језгрима:

```
gmm.fit.Tr = gmm(g, v, t0, kernel = "Truncated")
gmm.fit.Bar = gmm(g, v, t0, kernel = "Bartlett")
gmm.fit.Pz = gmm(g, v, t0, kernel = "Parzen")
confint(gmm.fit.Tr)
```

```
Wald type confidence interval
#############################################
             0.025     0.975
Theta[1]   2.5554   2.9480
Theta[2]   1.1105   1.3629
```

```
confint(gmm.fit.Bar)
```

```
Wald type confidence interval
#############################################
             0.025     0.975
Theta[1]   2.5554   2.9480
Theta[2]   1.1105   1.3629
```

```
confint(gmm.fit.Pz)
```

```
Wald type confidence interval
#########################################
             0.025    0.975
Theta[1]   2.5557   2.9449
Theta[2]   1.1111   1.3651
```

Ситуација се готово не мења, али то се и могло очекивати, јер смо раније рекли да избор језгара нема нарочито велики утицај на резултате. Ситуација је слична и за оцене параметара и њихова стандардна одступања:

```
coef(gmm.fit)
```

```
 Theta[1]  Theta[2]
2.755459 1.235548
```

```
coef(gmm.fit.Tr)
```

```
Theta[1]  Theta[2]
2.751722 1.236670
```

```
coef(gmm.fit.Bar)
```

```
Theta[1]  Theta[2]
2.751722 1.236670
```

```
coef(gmm.fit.Pz)
```

```
Theta[1]  Theta[2]
2.750310 1.238066
```

```
diag(vcov(gmm.fit))^0.5
```

```
Theta[1]     Theta[2]
0.10021163 0.06497052
```

```
diag(vcov(gmm.fit.Tr))^0.5
```

```
Theta[1]    Theta[2]
0.1001498 0.0643959
```

```
diag(vcov(gmm.fit.Bar))^0.5
```

```
Theta[1]    Theta[2]
0.1001498 0.0643959
```

```
diag(vcov(gmm.fit.Pz))^0.5
```

```
Theta[1]     Theta[2]
0.09926828 0.06479061
```

Можемо покушати да променимо b_n, који има више утицаја од избора језгара. Подразумевано је $bwAndrews$, а можемо нпр. поставити и $bw = bwNeweyWest$, да користимо аутоматски метод селекције b_n, који смо поменули раније.

```
gmm.fit.nw = gmm(g, v, t0, bw = bwNeweyWest)
summary(gmm.fit.nw)
```

```
Call:
gmm(g = g, x = v, t0 = t0, bw = bwNeweyWest)


Method:  twoStep

Kernel:  Quadratic Spectral(with bw =  2.12697 )

Coefficients:
           Estimate     Std. Error   t value      Pr(>|t|)
Theta[1]   2.7736e+00   9.8191e-02   2.8247e+01   1.5569e-175
Theta[2]   1.2347e+00   6.9373e-02   1.7798e+01    7.3798e-71

J-Test: degrees of freedom is 1
                J-test   P-value
Test E(g)=0:    1.82063  0.17724

Initial values of the coefficients
Theta[1] Theta[2]
2.847150 1.289139

##############
Information related to the numerical optimization
Convergence code =  0
Function eval. =  47
Gradian eval. =  NA
```

```
confint(gmm.fit.nw)
```

```
Wald type confidence interval
#########################################
           0.025    0.975
Theta[1]   2.5811   2.9660
Theta[2]   1.0987   1.3706
```

Ситуација се није нарочито изменила. Оцене параметара су сличне, вредност J-теста је поново мала, али интервали поверења не успевају да ухвате праве вредности.

Пошто знамо облик расподеле, рекли смо да је боље да користимо ММВ јер води до ефикаснијих оцена, што ћемо оправдати у малој симулационој студији[40].

[40]Наравно, када кажемо да је ММВ ефикаснија од УММ, мислимо на овај конкретан пример,

Присетимо се, оцене ММВ су $\hat{m}_{MMV} = \frac{1}{n}\sum_{t=1}^{n} v_t$ и $\hat{\sigma}^2_{MMV} = \frac{1}{n}\sum_{t=1}^{n}(v_t - \overline{v}_n)^2$.

```
sim = function(n, i_max) {
  theta_UMM = matrix(0, i_max, 2)
  theta_MMV = matrix(0, i_max, 2)
  for(i in 1:i_max) {
    v = rnorm(n, 3, sqrt(2))
    theta_UMM[i, ] = gmm(g, v, t0)$coefficients
    theta_MMV[i,1] = mean(v)
    theta_MMV[i,2] = sqrt(var(v)*(n-1)/n)
  }
  b = cbind(rowMeans(t(theta_UMM) - c(3,sqrt(2))),
            rowMeans(t(theta_MMV) - c(3,sqrt(2))))
  dimnames(b) = list(c("m", "sigma"), c("UMM", "MMV"))

  D = cbind(diag(var(theta_UMM)), diag(var(theta_MMV)))
  dimnames(D) = list(c("m", "sigma"), c("UMM", "MMV"))

  MSE = cbind(rowMeans((t(theta_UMM) - c(3,sqrt(2)))^2),
              rowMeans((t(theta_MMV) - c(3,sqrt(2)))^2))
  dimnames(MSE) = list(c("m", "sigma"), c("UMM", "MMV"))

  return(list(Pristrasnost = b, Disperzija = D, SK_greska = MSE))
}

sim(20,1000)
```

Добијамо следеће резултате:

	Пристрасност		Дисперзија		СК грешка	
	УММ	ММВ	УММ	ММВ	УММ	ММВ
m	-0.009429885	-0.003867692	0.12037397	0.10191886	0.12034252	0.10183190
σ	-0.113198065	-0.051117671	0.05841906	0.04824317	0.07117445	0.05080794

Размотримо шта се дешава са повећањем броја симулација и повећањем обима узорка:

```
sim(20,10000)
```

	Пристрасност		Дисперзија		СК грешка	
	УММ	ММВ	УММ	ММВ	УММ	ММВ
m	-0.004438992	0.001047054	0.11994790	0.09742277	0.11995561	0.09741412
σ	-0.113787619	-0.053467606	0.06194184	0.04915903	0.07488327	0.05201290

односно на оцењивање на основу овог скупа популационих услова момената. Није сама процедура неефикасна, већ услови на основу којих се изводе закључци.

```
sim(50,10000)
```

	Пристрасност		Дисперзија		СК грешка	
	УММ	ММВ	УММ	ММВ	УММ	ММВ
m	-0.002648086	-0.002564774	0.05105047	0.03976518	0.05105238	0.03976778
σ	-0.053501457	-0.019605194	0.02307663	0.02030941	0.02593673	0.02069174

У свим ситуацијама бољи резултати се добијају за ММВ, па на основу ове мини симулационе студије закључујемо да су оцене ММВ ефикасније него оцене добијене уопштеном методом момената на основу овог одабира услова момената.

Пример 2: Ендогеност у линеарним моделима

Претпоставимо да желимо да оценимо β из

$$y_t = \beta x_t + e_t,\ t = \overline{1,n}$$

$$x_t = e^{-r_t^2} + u_t,\ r_t \overset{iid}{\sim} \mathcal{N}(0,1),\ E(r_t e_t) = 0$$

и $(e_t, u_t)^T \overset{iid}{\sim} \mathcal{N}(\mathbf{0}_2, \Sigma)$, где су $\mathbf{0}_2 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ и $\Sigma = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$.

Ако бисмо покушали да оценимо параметар β МНК, добили бисмо оцену која је пристрасна и највероватније не води до добрих закључака. Проблем потиче од ендогености, што је проблем који се често јавља у линеарним моделима и најчешће се решава оцењивањем инструментима, што је, присетимо се, опет специјални случај уопштене методе момената.

За почетак, генеришемо податке (претпоставићемо да је $\beta = 0.3$ и вадимо узорак обима 200):

```
library(mvtnorm)

set.seed(16)
Sigma = matrix(c(1, 0.5, 0.5, 1), nrow = 2)
n = 200
e_u = rmvnorm(n, rep(0,2), Sigma)
e = e_u[, 1]
u = e_u[, 2]
r = rnorm(n)
x = exp(-r^2) + u
y = 0.3*x + e
```

Сада прелазимо на оцењивање. За почетак, морамо одабрати инструменте које ћемо користити у оцењивању. Видимо да се било која функција од r_t може користити као инструмент, јер је ортогонална на e_t и корелисана са x_t, па постоји бесконачно много избора могућих инструмената. Претпоставимо да смо одабрали вектор инструмената (r_t, r_t^2, r_t^3). Пошто је у питању линеарни модел, у функцији $gmm()$ аргумент g је формула, а x је матрица инструмената.

```
instr = cbind(r, r^2, r^3)
g = y~x
gmm.fit = gmm(g, x = instr)
summary(gmm.fit)
```

```
Call:
gmm(g = g, x = instr)


Method:  twoStep

Kernel:  Quadratic Spectral(with bw =  0.49468 )

Coefficients:
             Estimate   Std. Error   t value    Pr(>|t|)
(Intercept)  0.055302   0.161190     0.343082   0.731537
x            0.325545   0.201056     1.619182   0.105408

J-Test: degrees of freedom is 2
                 J-test   P-value
Test E(g)=0:     1.4468   0.4851

Initial values of the coefficients
(Intercept)           x
 0.04490423   0.33391084
```

Издвајамо резултате J-теста:

```
specTest(gmm.fit)
```

```
 ##  J-Test: degrees of freedom is 2  ##

                 J-test   P-value
Test E(g)=0:     1.4468   0.4851
```

Мала вредност тест статистике говори у прилог нултој хипотези.

Функција *gmm*() подразумевано рачуна двостепену оцену, али променом вредности аргумента *type* можемо променити ово подешавање и рачунати итерирану оцену:

```
gmm.fit.iter = gmm(g, instr, type = "iterative", itermax = 100, crit = 10
    e-7)
summary(gmm.fit.iter)
```

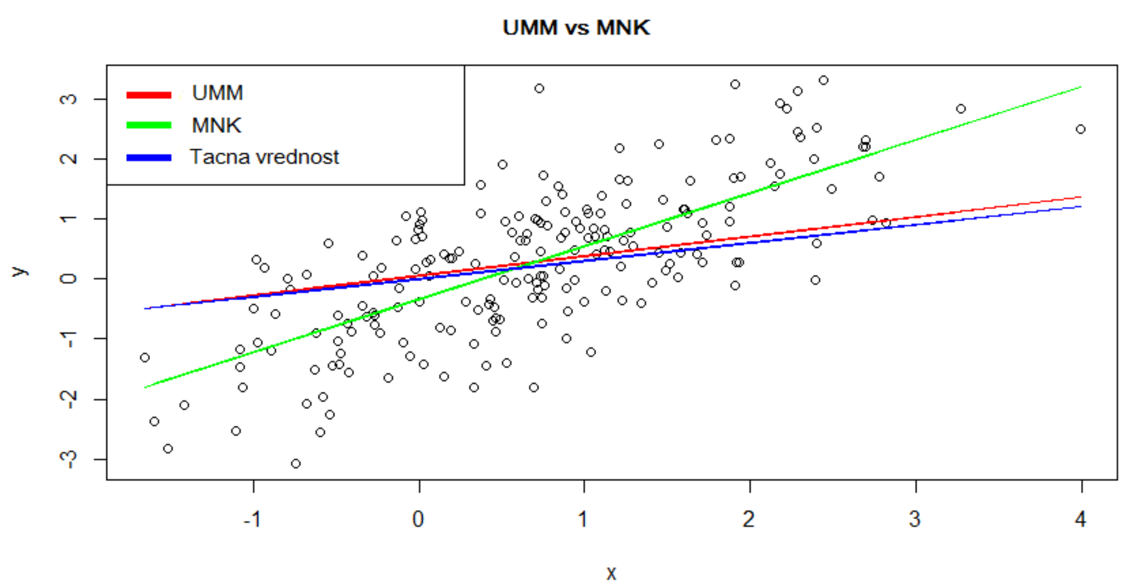
```
Call:
gmm(g = g, x = instr, type = "iterative", crit = 1e-06, itermax = 100)


Method:  iterative

Kernel:  Quadratic Spectral(with bw =  0.49679 )

```

```
Coefficients:
              Estimate   Std. Error   t value    Pr(>|t|)
(Intercept)   0.056708   0.161263     0.351649   0.725102
x             0.323755   0.201129     1.609683   0.107467

J-Test: degrees of freedom is 2
                 J-test    P-value
Test E(g)=0:     1.45914   0.48212

Initial values of the coefficients
(Intercept)            x
 0.04490423   0.33391084

```

Добијају се слични резултати.

За рад са линеарним моделима, пакет *gmm* обезбеђује методе *fitted* и *residuals*. Ово можемо да искористимо да бисмо графички представили резултате оцењивања методом најмањих квадрата и уопштеном методом момената, да стекнемо увид зашто не треба користити МНК.

```
plot(x, y, main = "UMM vs MNK")
lines(x, fitted(gmm.fit), col = "red")
lines(x, fitted(lm(y ~ x)), col = "green")
lines(x, 0.3*x, col = "blue")
legend("topleft", legend=c("UMM", "MNK", "Tacna vrednost"),  col=c("red",
    "green", "blue"))
```



Иако делује да се МНК боље уклапа у податке, ипак, није тако, јер график крије проблем ендогености - МНК прецењује везу између y_t и x_t, јер не узима у обзир чињеницу да су неке од корелација узорковане чињеницом да је x_t корелисано са e_t.

Пример 3: Оцењивање AR коефицијената ARMA процеса

У овом примеру ћемо размотрити случај где су подаци серијски корелисани. Рецимо да желимо да оценимо коефицијенте AR процеса

$$X_t = a_1 X_{t-1} + a_2 X_{t-2} + u_t$$

где је $u_t = 0.4e_{t-1} + 0.2e_{t-2} + e_t$ и $e_t \overset{iid}{\sim} \mathcal{N}(0,1)$.
За почетак, генеришемо податке (претпоставимо да је $a_1 = 0.6$ и $a_2 = 0.1$ и вадимо узорак обима 200):

```
n = 200
set.seed(28)
x = arima.sim(n, model = list(ar = c(0.6, 0.1), ma = c(0.4, 0.2)))
```

Параметре можемо оценити инструментима, па је аргумент g поново формула, а x вектор инструмената. Пошто су X_{t-j}, за $j > 2$ некорелисане са u_t, а корелисане са X_{t-1} и X_{t-2}, представљају одговарајуће кандидате за инструменте. Претпоставимо да смо одабрали вектор инструмената $(X_{t-3}, X_{t-4}, X_{t-5})$:

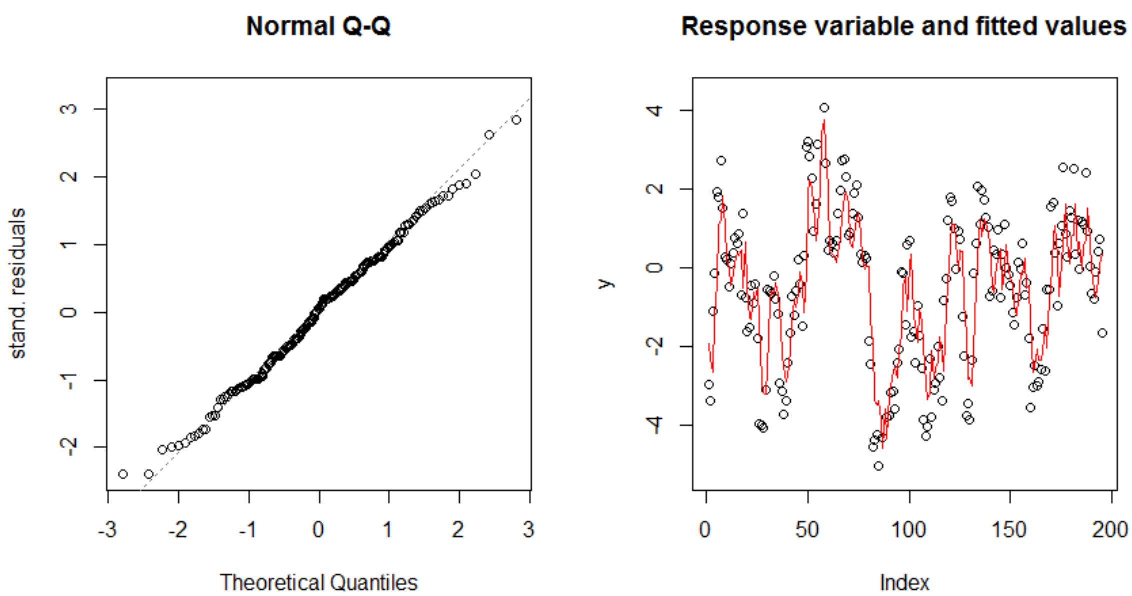
```
xn = x
for (i in 1:5) {
  xn = cbind(xn, lag(x, -i))
}
xn = na.omit(xn)

g = xn[, 1] ~ xn[, 2] + xn[, 3]
instr = xn[, 4:6]

gmm.fit = gmm(g, x = instr)
summary(gmm.fit)
```

```
Call:
gmm(g = g, x = instr)


Method:  twoStep

Kernel:  Quadratic Spectral(with bw =  1.4454 )

Coefficients:
              Estimate    Std. Error   t value     Pr(>|t|)
(Intercept)  -0.154165    0.098058    -1.572193    0.115906
xn[, 2]       0.644758    0.302133     2.134020    0.032841
xn[, 3]       0.108245    0.272076     0.397847    0.690743

J-Test: degrees of freedom is 1
                J-test    P-value
Test E(g)=0:    0.55012   0.45827

```

```
Initial values of the coefficients
(Intercept)      xn[, 2]      xn[, 3]
 −0.1240697    0.5665695    0.1794077
```

Оцене коефицијената су блиске правим вредностима, а резултати теста иду у прилог нултој хипотези. Пошто је у питању линеарни модел, можемо нацртати график резидуала, као и график опсервација и фитованих вредности:

```
par(mfrow = c(2,1))
plot(gmm.fit , which = 2)
plot(gmm.fit , which = 3)
```



Резидуали не показују значајна одступања од нормалности, а видимо да је модел успео добро да се уклопи у податке, што је у складу са резултатима J-теста.

Пример 4: Оцењивање параметара стабилне расподеле

У првом примеру смо видели да је најбоље користити ММВ ако знамо расподелу података. Међутим, обично или не знамо тачну расподелу, или је знамо, али је израчунавање прекомпликовано, па морамо покушати на друге начине. Сада ћемо представити један пример оваквих околности и показати како нам УММ може помоћи.

У теорији вероватноће, стабилне расподеле представљају класу расподела за које важи да линеарна комбинација две независне случајне величине са истом расподелом поново има исту расподелу до на параметре положаја и размере. Стабилне расподеле су индексиране са четири параметра: $\alpha \in (0,2]$ - параметар стабилности, $\beta \in [-1,1]$ - параметар асиметрије, $c > 0$ - параметар размере и $\mu \in \mathbb{R}$ - параметар положаја. Ова фамилија расподела обухвата неке познате расподеле, које се добијају фиксирањем једног или више параметара (нпр. нормална за $\alpha = 2$, Кошијева за $\alpha = 1$ и $\beta = 0$, итд.). Међутим, стабилне расподеле су један од примера где немамо аналитички израз за густину у општем случају, па не можемо користити ММВ.

Покушаћемо да искористимо УММ да оценимо параметре. За почетак, потребно је да одредимо услове момената које ћемо користити. Оно што је згодно код ових расподела је то што се, за разлику од густине, карактеристична функција може изразити аналитички. Ако је v случајна величина са стабилном расподелом са параметрима $\theta = (\alpha, \beta, c, \mu)^T$, онда је

$$\varphi_v(t) = E(e^{itv}) = e^{it\mu - |ct|^\alpha (1 - i\beta sgn(t)\phi)},\ t \in \mathbb{R}$$

где је

$$\phi = \left\{ \begin{array}{ll} tan(\frac{\pi\alpha}{2}), & \text{за } \alpha \neq 1 \\ -\frac{2}{\pi}\log|t|, & \text{за } \alpha = 1 \end{array} \right\}$$

Наравно, функција $gmm()$ не може радити са бесконачним бројем услова момената, али можемо одбрати $\{t_1, ..., t_q\}$ из датог интервала и оценити параметре коришћењем услова момената

$$E(e^{it_k v_j} - \varphi_{v_j}(t_k; \theta)) = 0, \text{ за } k = \overline{1, q},\ j \in \overline{1, n}.$$

Функција која рачуна карактеристичну функцију стабилне расподеле је укључена у пакет *gmm* и може се користити за конструисање g[41]. Дакле, услове можемо писати на следећи начин:

```
g = function(theta, v) {
  # theta = (alpha, beta, c, mu)^T
  # v = (v_j)
  t = seq(1, 5, length.out = 10)

  m1 = matrix(complex(imaginary = v %*% t(t)), ncol = length(t))
  m1 = exp(m1)

  m2 = charStable(theta, t, 1)
  # poslednji argument se odnosi na parametrizaciju
  # stabilna raspodela ima vise izbora parametrizacije
  # ova koju smo odbrali je oznacena brojem 1 u funkciji charStable()

  m = t(t(m1) - m2)
  m = cbind(Im(m), Re(m))
  return(m)
}
```

Генеришемо податке из стабилне расподеле са параметрима $\alpha = 1.5$, $\beta = 0.5$, $c = 1$ и $\mu = 0$ коришћењем функције $rStable()$ из пакета *portes*. Затим оцењујемо параметре уопштеном методом момената.

[41] Функција $charStable()$ која враћа имагинарне и реалне делове у засебним колонама.

```r
library(portes)

set.seed(29)
n = 500
v = rStable(n, 1.5, 0.5, 1, 0)
t0 = c(2, 0, sd(v)/sqrt(2), 0)
gmm.fit = gmm(g, v, t0)
summary(gmm.fit)
```

```
Call:
gmm(g = g, x = v, t0 = t0)


Method:  twoStep

Kernel:  Quadratic Spectral(with bw =  0.98161 )

Coefficients:
          Estimate    Std. Error   t value     Pr(>|t|)
Theta[1]  1.00005110        Inf   0.00000000  1.00000000
Theta[2]  0.00081978        Inf   0.00000000  1.00000000
Theta[3]  1.93596632        Inf   0.00000000  1.00000000
Theta[4]  1.16885938        Inf   0.00000000  1.00000000

J-Test: degrees of freedom is 16
                 J-test      P-value
Test E(g)=0:     1.2263e+02  1.7151e-18

Initial values of the coefficients
 Theta[1]  Theta[2]  Theta[3]  Theta[4]
1.1096786 0.2053601 1.0597467 1.2091574

#############
Information related to the numerical optimization
Convergence code =  1
Function eval. =  501
Gradian eval. =  NA
```

```r
coef(gmm.fit)
```

```
    Theta[1]      Theta[2]      Theta[3]      Theta[4]
1.0000511043 0.0008197832 1.9359663175 1.1688593780
```

```r
specTest(gmm.fit)
```

```
 ##  J-Test: degrees of freedom is 16  ##

                 J-test      P-value
Test E(g)=0:     1.2263e+02  1.7151e-18
```

Видимо да су добијене оцене лоше и да нису блиске правим вредностима. Ово се слаже са резултатима J-теста, који даје велику вредност тест статистике, те одбацује нулту хипотезу да су услови момената задовољени. Међутим, овакве резултате смо могли и да очекујемо, јер нисмо навели никаква ограничења за параметре, која треба да важе. Да бисмо то урадили, постављамо аргумент $optfct$ на "$nlminb$", што нам омогућава ограничимо параметарски постор у складу са претпоставкама.

```
Call:
gmm(g = g, x = v, t0 = t0, optfct = "nlminb", lower = c(0, -1, 0, -Inf),
    upper = c(2, 1, Inf, Inf))


Method:  twoStep

Kernel:  Quadratic Spectral(with bw =  0.98161 )

Coefficients:
           Estimate     Std. Error   t value      Pr(>|t|)
Theta[1]   1.2857e+00   1.3644e-01   9.4232e+00   4.3764e-21
Theta[2]   4.0814e-01   2.5813e-01   1.5811e+00   1.1385e-01
Theta[3]   9.4589e-01   5.4401e-02   1.7387e+01   1.0283e-67
Theta[4]   7.5401e-01   6.5812e-01   1.1457e+00   2.5192e-01

J-Test: degrees of freedom is 16
                 J-test    P-value
Test E(g)=0:     23.3899   0.1037

Initial values of the coefficients
  Theta[1]    Theta[2]    Theta[3]    Theta[4]
 1.2472310  -0.1234856   1.0060237  -0.4255512

###############
Information related to the numerical optimization
Convergence code =  0
Function eval. =  96
Gradian eval. =  289
Message:  relative convergence (4)
```

Сада се добијају приближније оцене. Такође, на основу резултата J-теста, сада можемо прихватити нулту хипотезу.

С обзиром да смо произвољно одабрали тачке у којима рачунамо карактеристичну функцију, можемо да погледамо шта се дешава ако, на пример, повећамо број тачака, односно број популационих услова момената:

```
g2 = function(theta, v) {
  t = seq(1, 5, length.out = 15)

  m1 = matrix(complex(imaginary = v %*% t(t)), ncol = length(t))
  m1 = exp(m1)

  m2 = charStable(theta, t, 1)
```

```

  m = t(t(m1) - m2)
  m = cbind(Im(m), Re(m))
   return(m)
}

gmm.fit2 = gmm(g2, v, t0, optfct = "nlminb", lower = c(0, -1 , 0, -Inf),
    upper = c(2, 1, Inf, Inf))
summary(gmm.fit2)
```

```
Call:
gmm(g = g2, x = v, t0 = t0, optfct = "nlminb", lower = c(0, -1,
    0, -Inf), upper = c(2, 1, Inf, Inf))


Method:  twoStep

Kernel:  Quadratic Spectral(with bw =  0.97445 )

Coefficients:
           Estimate     Std. Error   t value      Pr(>|t|)
Theta[1]   1.6488e+00   8.9935e-02   1.8333e+01   4.5274e-75
Theta[2]   6.7852e-01   3.0053e-01   2.2577e+00   2.3961e-02
Theta[3]   8.9690e-01   3.7515e-02   2.3908e+01   2.5390e-126
Theta[4]   2.8785e-01   1.4339e-01   2.0075e+00   4.4700e-02

J-Test: degrees of freedom is 26
                J-test      P-value
Test E(g)=0:    50.4991197   0.0027397

Initial values of the coefficients
  Theta[1]   Theta[2]   Theta[3]   Theta[4]
 1.2301386 -0.1453023  1.0120517 -0.5317970

##############
Information related to the numerical optimization
Convergence code =  0
Function eval. =  29
Gradian eval. =  103
Message:  relative convergence (4)
```

Повећање степена преидентификације је довело до одбацивања нулте хипотезе. Вероватно је да су неки од укључених услова момената сувишни за дате преостале услове.[42]

Погледајмо сада шта се дешава на конкретним подацима. Рекли смо већ да је уопштена метода момената имала највећи утицај у области економетрије, па ћемо се позабавити финансијским подацима. Претпоставимо да желимо да

[42] У овом примеру, сигурно би нам били од помоћи поменути критеријуми за селекцију момената.

испитамо да ли приноси неке компаније прате стабилну расподелу. У оквиру пакета *gmm* налази се велика база *Finance* која садржи 24 временске серије, које представљају податке о дневним приносима за 20 различитих компанија, као и колоне rf и rm, које представљају безризичну каматну стопу и принос тржишног портфолија, редом[43]. Ако нам, на пример, податке представљају приноси компаније Mattel Inc.:

```
data("Finance")
v = Finance[1:500, "MAT"]
t0 = c(2, 0, sd(v)/sqrt(2), 0)
gmm.fit.F = gmm(g, v, t0, optfct = "nlminb", lower = c(0, -1 , 0, -Inf),
    upper = c(2, 1, Inf, Inf))
summary(gmm.fit.F)
```

```
Call:
gmm(g = g, x = v, t0 = t0, optfct = "nlminb", lower = c(0, -1,
    0, -Inf), upper = c(2, 1, Inf, Inf))


Method: twoStep

Kernel: Quadratic Spectral(with bw = 0.81207 )

Coefficients:
           Estimate    Std. Error  t value     Pr(>|t|)
Theta[1]  1.5054e+00  1.7781e-01  8.4664e+00  2.5307e-17
Theta[2]  7.1052e-01  3.7386e-01  1.9005e+00  5.7371e-02
Theta[3]  1.0820e+00  5.9719e-02  1.8119e+01  2.2698e-73
Theta[4]  7.0282e-01  5.2156e-01  1.3475e+00  1.7781e-01

J-Test: degrees of freedom is 16
                J-test    P-value
Test E(g)=0:    18.62373  0.28866

Initial values of the coefficients
 Theta[1]  Theta[2]  Theta[3]  Theta[4]
1.0809333 0.2249989 1.2578663 2.1242856

##############
Information related to the numerical optimization
Convergence code = 0
Function eval. = 32
Gradian eval. = 117
Message: relative convergence (4)
```

Видимо да J-тест не одбацује хипотезу да приноси прате стабилну расподелу.

Нормална расподела је специјалан случај стабилне расподеле за $\alpha = 2$, па сада можемо да тестирамо хипотезу да приноси прате нормалну расподелу.

[43]Садржи и колоне *hml* и *smb*, али оне нам тренутно нису од значаја.

То можемо учинити тестирањем хипотезе да вектор параметара задовољава ограничења потребна да би расподела била баш нормална, односно тестирањем

$$H_0 : R\theta = r$$

где је $R = (1, 0, 0, 0)$, $\theta = (\alpha, \beta, c, \mu)^T$, а $r = 2$, односно $H_0 : \alpha = 2$.

```
library(car)

R = c(1,0,0,0)
r = 2
linearHypothesis(gmm.fit.F, R, r)
```

```
Linear hypothesis test

Hypothesis:
Theta[1] = 2

Model 1: restricted model
Model 2: gmm.fit.F

  Df  Chisq Pr(>Chisq)
1
2  1 7.7366   0.005411 **
---
Signif. codes:  0   ***   0.001   **   0.01   *   0.05   .   0.1
      1
```

На основу резултата, јасно је да се нулта хипотеза о нормалности расподеле одбацује.

Пример 5: Оцењивање параметара CAPM модела

На крају, размотримо један значајан пример из економије - CAPM[44] модел. Капитални модел за процењивање вредности акција (односно CAPM) доводи у везу стопу добити за један период за одређени вредносни папир i, R_i, са стопом добити целе берзе - R_m, односно са стопом добити тржишног портфолија.

Нека је r_f безризична каматна стопа. Једначина CAPM модела је дата са

$$R_i = r_f + \beta_i(R_m - r_f) + e_i$$

где је β_i константа, e_i представља грешку, за коју се претпоставља да је независна од R_m и $E(e_i) = 0$. R_i и R_m су случајне величине, $E(R_i) = r_i$, $E(R_m) = r_m$, па из једначине модела добијамо да је

$$r_i - r_f = \beta_i(r_m - r_f).$$

Дакле, претпоставља се да је разлика између очекиване стопе добити од i-те акције (или неког другог вредносног папира) r_i и безризичне каматне стопе r_f

[44] Capital Asset Pricing Model.

једнака производу константе β_i и разлике очекиване стопе добити берзе r_m и безризичне каматне стопе r_f.

Исправност ове теорије можемо тестирати са[45]

$$R_t - r_f = \alpha + \beta(R_{mt} - r_f) + e_t.$$

Ако важи CAPM модел, α треба да буде 0. Према томе, важење модела можемо испитати оцењивањем параметара модела уопштеном методом момената и тестирањем хипотезе да је $\alpha = 0$. У питању је линеарни модел, па се УММ своди на оцењивање инструментима. Претпоставимо да је вектор инструмената $z_t = R_{mt} - r_f$[46]. Податке поново добијамо из базе *Finance*.

```
data(Finance)
n = 200

r = Finance[1:n, 1:5]
rm = Finance[1:n, "rm"]
rf = Finance[1:n, "rf"]

y = as.matrix(r - rf)
instr = as.matrix(rm - rf)
g = y ~ instr
gmm.fit = gmm(g, instr)
summary(gmm.fit)
```

```
Call:
gmm(g = g, x = instr)


Method:  twoStep

Kernel:  Quadratic Spectral

Coefficients:
                       Estimate     Std. Error   t value      Pr(>|t|)
WMK_(Intercept)        1.1899e-02   7.5548e-02   1.5750e-01   8.7485e-01
UIS_(Intercept)       -1.8255e-02   1.4371e-01  -1.2703e-01   8.9892e-01
ORB_(Intercept)        1.7573e-01   2.2835e-01   7.6956e-01   4.4156e-01
MAT_(Intercept)        2.0692e-02   1.4147e-01   1.4626e-01   8.8371e-01
ABAX_(Intercept)       5.9418e-02   3.3767e-01   1.7597e-01   8.6032e-01
WMK_instr              2.9195e-01   1.7152e-01   1.7021e+00   8.8741e-02
UIS_instr              1.2872e+00   2.6048e-01   4.9419e+00   7.7358e-07
ORB_instr              1.3501e+00   5.6962e-01   2.3702e+00   1.7777e-02
MAT_instr              1.0878e+00   3.0196e-01   3.6025e+00   3.1522e-04
ABAX_instr             4.9595e-01   7.1598e-01   6.9269e-01   4.8850e-01

```

[45]Изостављамо индекс i ради једноставнијег записа.

[46]У овом случају је $p = q$, односно димензија вектора параметара је једнака димензији вектора инструмената.

```
J-Test: degrees of freedom is 0
                    J-test                  P-value
Test E(g)=0:        1.96514599031991e-30    *******
```

Сада можемо да тестирамо хипотезу

$$H_0 : R\theta = \mathbf{0}_5$$

где је $R = (I_5, \mathbf{0}_{5\times5})$,$\theta = (\alpha_1, ..., \alpha_5, \beta_1, ..., \beta_5)^T$ и $\mathbf{0}_5 = (0, 0, 0, 0, 0)^T$.

```
R = cbind(diag(5), matrix(0, 5, 5))
I5 = rep(0, 5)
linearHypothesis(gmm.fit, R, I5)
```

```
Linear hypothesis test

Hypothesis:
WMK_((Intercept) = 0
UIS_((Intercept) = 0
ORB_((Intercept) = 0
MAT_((Intercept) = 0
ABAX_((Intercept) = 0

Model 1: restricted model
Model 2: y ~ instr

  Res.Df Df  Chisq Pr(>Chisq)
1    203
2    198  5 0.6391     0.9861
```

Нулта хипотеза се прихвата, односно сматрамо да CAPM модел важи.

Важење модела можемо тестирати и на други начин - оценимо рестриковани модел за $\alpha = 0$, а онда спроведемо J-тест. Приметимо да је сада могуће да спроведемо тест, јер је број инструмената већи од броја параметара, односно имамо преидентификацију.

```
g.r = y ~ instr - 1
instr.r = cbind(1, instr)
gmm.fit.r = gmm(g.r, instr.r)
specTest(gmm.fit.r)
```

```
##  J-Test: degrees of freedom is 5  ##

                    J-test   P-value
Test E(g)=0:        0.64011  0.98609
```

Долазимо до истог закључка - прихватамо нулту хипотезу да је $\alpha = 0$, односно да је претпостављена теорија одговарајућа.

8 Закључак

У овом раду описали смо уопштени метод момената и приказали његову примену на временским серијама. Увели смо УММ као уопштење методе момената, које се може користити када имамо више популационих услова момената него непознатих параметара. Међутим, показали смо да УММ обухвата и многе друге методе закључивања попут методе максималне веродостојности, методе најмањих квадрата, као и оцењивања инструментима у линеарним регресионим моделима, које користимо када се јавља проблем ендогености.

Анализу смо почели специјалним случајем оцењивања инструментима у линеарним регресионим моделима, описали оцењивање, асимптотске особине оцене параметра и оцене узорачког момента и навели проблеме који могу настати. Затим смо уопштили дискусију и показали како се метода примењује на нелинеарне моделе.

У оквиру ове дискусије, дефинисали смо популациони услов момената на коме се заснива оцењивање и показали како се може разложити на идентификационе услове и услове преидентификације, део који се користи у оцењивању и остатак који игра важну улогу при формирању тестова. Показали смо да је оцена параметра добијена овом методом постојана и да има асимптотски нормалну расподелу, као и да оцена узорачког момента има асимптотски нормалну расподелу.

Даље, показали смо да је оптимални избор тежинске матрице инверз постојане оцене асимптотске коваријационе матрице узорачког момента, те смо испитали и начине на које се она може постојано оценити под различитим претпоставкама о структури зависности процеса.

Све добијене закључке смо извели под претпоставком да је модел коректно спецификован. Међутим, ако је модел погрешно спецификован, испоставило се да добијени закључци не важе, па смо увели тест спецификације модела. Описали смо тест статистику и приказали њену граничну расподелу при нултој хипотези. Такође, некада претпостављена теорија поставља рестрикције на параметре, па смо описали тестове који испитују да ли су рестрикције задовољене на подацима.

На крају, показали смо практичну примену уопштене методе момената на конкретним примерима и показали како нам некада може помоћи у добијању бољих закључака.

ЛИТЕРАТУРА

[1] A. R.Hall: *Generalized Method of Moments*, Oxford University Press, 2005.

[2] C. F. Baum, M. E. Schaffer, S. Stillman: *Instrumental variables and GMM: Estimation and Testing*, Boston College Department of Economics, 2003.

[3] E. Zivot: *Generalized Method of Moments*, Faculty of Washington, 2005.

[4] H. B. Nielsen: *Generalized Method of Moments Estimation*, 2016.

[5] J. P. Nolan: *Stable Distribution: Models for Heavy Tailed Data*, 2018.

[6] S. Majoros: *Multivariate stable distributions and their application in modeling returns*, Corvinus University of Budapest, 2018.

[7] W. J. Den Haan, A.Lewin: *A Practioner's guide to robust covariance matrix estimation*, 1997.

[8] G. V. Farnsworth: *Econometrics in R*, 2008.

[9] С. Јанковић, Б. Милошевић: *Елементи финансијске математике*, Математички факултет, Београд, 2017.

[10] V. Martin, S. Hurn, D. Harris: *Econometric Modelling with Time Series: Specification, Estimation and Testing*, Cambridge University Press, 2013.

[11] L. P. Hansen: *Generalized Method of Moments Estimation*, Department of Economics University of Chicago, 2007.

[12] D. W. K. Andrews: *Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation*, Econometrica, 1991.

[13] R. Garcia, E. Renault, D. Veredas: *Estimation of Stable Distributions by Indirect Inference*, 2004.

[14] P. Chausse: *Generalized Method of Moments and Generalized Empirical Likelihood*, 2013.

[15] M. Carasco, J. P. Florens *Efficient GMM Estimation Using the Empirical Characteristic Function*, 2002.

[16] K.P. Lin: *Econometric Analysis by Examples*, Xiamen University, 2003.

Биографија

Мирјана Вељовић рођена је 16. маја 1996. у Краљеву. Основну школу ,,Попински борци" завршила је у Врњачкој Бањи. Након основне школе, школовање је наставила у Гимназији у Врњачкој Бањи, коју је завршила као носилац дипломе ,,Вук Караџић". Школске 2014./2015. године, уписала је Математички факултет у Београду, смер Статистика, финансијска и актуарска математика. Дипломирала је у јулу 2018., а у октобру школске 2018./2019. године уписала је мастер студије, смер Статистика, финансијска и актуарска математика, на истом факултету. Запослена је на Математичком факултету у Београду као сарадник у настави на катедри за Вероватноћу и статистику од октобра 2018. године.