دانشگاه پیام نور
دانشکده علوم پایه و کشاورزی مرکز تهران

پایان نامه برای دریافت مدرک کارشناسی ارشد
رشته آمار ریاضی

مدلهای رگرسیونی سانسور شده نیمه پارامتری

نرگس دل آرام

استاد راهنما:
دکتر مسعود یارمحمدی
استاد مشاور:
دکتر علی شادرخ

مهر ۱۳۹۰

چکیده

برآوردگرهای پارامترها در مدلهای رگرسیونی سانسور شده از چپ و سانسور شده از راست مورد نظر می باشد. این شکل از برآوردگرها نیازی به شرط متقارن بودن توزیع خطاها ندارند. برآوردگرهای معرفی شده "برآوردگر سانسور از چپ"، "برآوردگر سانسور از راست"، "برآوردگر مد چارکی" و "برآوردگر میانگین وینزوری" می باشند. خاصیت سازگاری و بطور مجانبی نرمال بودن این برآوردگرها نشان داده می شود. در پایان ویژگی های این برآوردگرها در یک مطالعه شبیه سازی با استفاده از معیارهای میانگین خطا و اریبی مورد بحث و بررسی قرار می گیرد.

واژگان کلیدی : رگرسیون، سانسور شده، نیمه پارامتری، برآوردگر.

فهرست مطالب

عنوان صفحه

فصل سوم رگرسیون نیمه پارامتری

فصل چهارم برآورد پارامترها در مدلهای رگرسیون سانسور شده نیمه پارامتری

فصل پنجم شبیه سازی با استفاده از نرم افزار *R*

مقدمه

در تحلیل های آماری، زمانیکه داده های مورد بررسی در قسمتی از زمان وجود نداشته باشند با داده های سانسور شده مواجه هستیم. در اینگونه مواقع استفاده از روش های رگرسیون معمولی جهت برآورد پارامترهای مدل رگرسیونی، منجر به نتایج غیر واقعی می شوند و عملاً کاربردی ندارند. به علاوه گاهی اوقات در مدل رگرسیونی هم مؤلفه پارامتری وجود دارد و هم مؤلفه ناپارامتری که در این حالت می بایست از روشهای رگرسیونی نیمه پارامتری استفاده نمود.

تا کنون چند برآوردگر برای مدلهای رگرسیونی سانسور شده ارائه شده است. درسال ۱۹۹۴ زمینه سازی برای بررسیهای تئوری برآورد مدلهای نیمه پارامتری توسط پاول انجام گرفت. پس از آن لی[۱] و کیم[۲] (۱۹۹۸) برآوردگرهایی را برای مدلهای رگرسیونی اصلاح شده از چپ ارائه کردند و بررسی های خود را با شبیه سازی این مطالعات در حالتهای کلی تر ادامه دادند. این مطالعات، منجر به یافتن برآوردگر کمترین مربعات اصلاح شده متقارن ($STLS$ [۳]) (پاول ۱۹۸۶) ، برآوردگرمدل چارکی (QME [۴]) (لی۱۹۹۳) و برآوردگر کسینوس (COS [۵]) (لی و کیم ۱۹۹۸) شد.

هدف از نگارش این پایان نامه معرفی چند مورد از برآوردگرهای مدل رگرسیونی سانسور شده نیمه پارامتری و همچنین مقایسه این برآوردگرها با استفاده از علم شبیه سازی می باشد.

در فصل اول داده های سانسور شده و انواع آن را معرفی می کنیم در فصل دوم متغیرها و رگرسیون سانسور شده را مورد بحث و بررسی قرار داده و سپس به معرفی رگرسیون نیمه پارامتری در فصل سوم می پردازیم در فصل چهارم برآورد پارامترها در مدلهای رگرسیونی نیمه پارامتری بیان شده که در آن چند مورد از برآوردگرهایی را که در مورد داده های سانسور شده در حالت بررسی نیمه پارامتری کاربرد دارند را معرفی می کنیم در فصل پنجم ابتدا روشهای شبیه سازی را معرفی نموده و سپس بسته نرم افزاری مورد استفاده در مدلهای رگرسیونی سانسور شده نیمه پارامتری در نرم افزار R را معرفی و با استفاده از شبیه سازی داده، چند مورد برآوردگرهای عنوان شده در فصل چهارم را با هم مقایسه می کنیم.

---

۱ Lee
۲ Kim
۳ Symmetrically Trimmed Least Squares
۴ quadratic model estimator
۵ Cosine estimator

فصل اول

معرفی داده های سانسور شده

۱-۱ طبقه بندی داده ها [۶]

بیشترین نوع از داده های غیر حیاتی[۷]، به خوبی داده های حیاتی هستند که آنها را داده های کامل می نامیم. داده ی کامل به مقداری از هر واحد نمونه گفته می شود که مشاهده شدنی و قابل اندازه گیری است. در بسیاری از موارد وقتی فرآیندی را بررسی می کنیم، داده های حیاتی قابل اطمینان نیستند. در چنین مواقعی با داده های سانسور شده روبرو هستیم. به طور کلی داده ها، به داده های کامل [۸]و داده های سانسور شده[۹] طبقه بندی می شوند.

۱-۲ داده های کامل

داده های کامل ، داده هایی هستند که مقدار هر واحد نمونه معلوم و مشاهده شده است. برای مثال فرض می کنیم میانگین نمرات ۱۰ دانش آموز را در امتحانی محاسبه کرده باشیم. بنابراین داده های کامل همان نمرات دانش آموزان بوده که مقدارش معلوم است. همچنین برای تحلیل داده های حیاتی، مجموعه داده ها در صورت کامل بودن، می بایست بر اساس زمان به پایان رسیدن رخداد پدیده ی مورد نظر هر واحد نمونه، ثبت شود. به عنوان مثال، فرض کنیم از ۵ دانش آموز آزمونی به عمل آمده باشد. همچنین زمان به پایان رسیدن امتحان توسط هر دانش آموز ثبت شده باشد، در اینجا اطلاعات کاملی در مورد نمره دانش آموز به همراه زمان به پایان رساندن امتحان برای تمام افراد در دسترس است.

۱-۳ داده های سانسور شده

این مجموعه داده ها آنهایی هستند که ممکن است در طول مطالعه به طور کامل شرکت نداشته و یا تا پایان مطالعه فعال یا سالم باشند. باید توجه کرد که داده سانسور شده را باید با استفاده از آخرین اطلاع از واحدها دقیقاً ثبت کرد، تا در تحلیل داده ها مورد استفاده قرار گیرند.

برای معرفی دقیق تر انواع سانسور ابتدا به معرفی چند مفهوم کلی می پردازیم:

۱-۴ شکست

---

[۶] Data Classification
[۷] non life
[۸] Complete Data
[۹] Censored Data

منظور از شکست[۱۰] در تحلیل بقا، یعنی رخداد همان حادثه ای که منتظر وقوعش بودیم. برای مثال در یک بررسی بالینی درباره نحوه اثر بخشی رژیم غذایی در درمان بیماری سرطان ریه، شکست مرگ به دلیل بیماری سرطان ریه است.

۱-۵ زمان مبدأ

زمانی که بقا از آن به بعد اندازه گیری می شود، زمان مبدأ است. سن فرد هنگام ورود به مطالعه، فشار اولیه ای که به قطعه تولیدی وارد می شود، و در بسیاری از موارد زمان رخداد برخی حوادث نظیر ورود فرد به مطالعه یا زمان تشخیص یک بیماری خاص می توانند به عنوان زمان مبدأ در نظر گرفته شوند. البته زمان مبدأ برای همه افرادی که مورد بررسی قرار می گیرند لزوماً یکسان نیست و در اندازه گیری زمان بقا باید به این مسئله توجه شود.

۱-۶ زمان شکست

همان زمانی است که برای هر فرد شکست رخ می دهد. که آن را با $i=1,...,n$ ، T_i نشان می دهند. زمان شکست از زمان مبدأ تا لحظه ای که شکست رخ دهد به عنوان T_i در نظر گرفته می شود. البته همواره امکان مشاهده زمان شکست برای تک تک افراد مورد بررسی نیست. در چنین مواردی سانسور رخ می دهد.

۱-۷ سانسور نوع I

این نوع سانسورحالتی را توصیف می کند که آزمایش در زمان معینی تمام می شود. بنابراین برای واحدهایی که باقی مانده اند تنها می دانیم شکست (حادثه مورد نظر) تا لحظه ی پایان مطالعه رخ نداده است. در چنین حالتی زمان سانسور ثابت و تعداد واحدهایی که شکست خورده اند تصادفی است. برای مثال آزمایش را با ۱۰۰ لامپ آغاز کرده و تا زمان مشخص و از پیش تعیین شده ادامه می دهیم. اگر زمان شروع آزمایش را زمان صفر و زمان پایان را T_0 بنامیم برای سانسور نوع **I** سه نوع سانسور (سانسور از چپ، سانسور از راست، سانسور بر روی فواصل) وجود دارد که در زیر تعریف می شود.

---

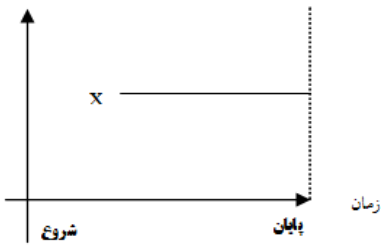
[۱۰] failure

۱-۸ سانسور از چپ[۱۱]

این نوع سانسور به این مفهوم است که اطلاعات مورد بررسی در ابتدای مطالعه وجود ندارد. یعنی از زمانی به بعد وارد مطالعه شده است و تا پایان آزمایش شکست نخورده است. به این نوع سانسور از چپ یا چپ سانسور می گوییم. شکل (۱-۱) را ببینید.

۱- فرد به دلایلی نتواند از ابتدای آزمایش در حوزه ی مطالعه قرار گیرد و مثلاً از زمان T ($T < T_0$) وارد آزمایش شده و تا پایان آزمون برایش شکست اتفاق نیفتد.

۲- فرد ابتدا از شرکت در مطالعه کناره گیری کرده ولی از زمان T ($T < T_0$) ، وارد آزمایش شود.



شکل (۱- ۱) سانسور از چپ در سانسور نوع I

۱-۹ سانسور از راست[۱۲]

برخی از داده های بقا شامل مواردی است که زمان شکست افراد در طول دوره مطالعه مشاهده نمی شود. به زمان مشاهده شده مربوط به چنین افرادی زمان سانسور از راست گفته می شود. این زمان کمتر یا برابر زمان پایان مطالعه است. در موارد زیر امکان مشاهده کامل افراد تا پایان مطالعه وجود ندارد، یعنی زمان سانسور کمتر از زمان پایان مطالعه است. شکل (۱-۲) را ببینید.

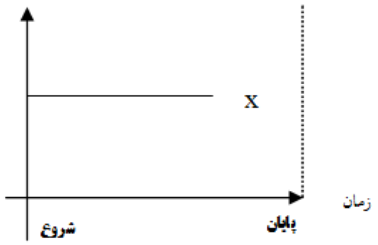
۱- فرد نقل مکان کرده و از حوزه ی مطالعه خارج شده باشد و تا پایان مطالعه باز نگردد.

۲- فرد از شرکت در مطالعه کناره گیری کرده و یا تصمیم به کناره گیری کرده باشد، که می تواند به دلیل تغییر در وضعیت فرد مثل بهبود یافتن و یا وخیم شدن علایم بالینی باشد.

این در حالی است که در بیش تر موارد افراد تا پایان مطالعه مشاهده شده اند ولی هنوز شکست برای آن ها رخ نداده است. در چنین مواردی زمان سانسور برابر زمان پایان مطالعه در نظر گرفته می شود.

---

[۱۱] Left censoring
[۱۲] Right censoring



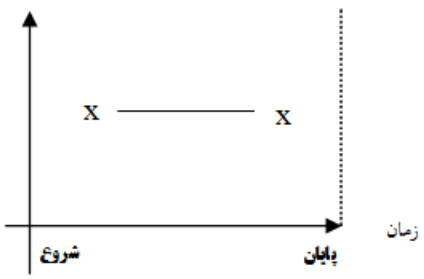
شکل (۱-۲) سانسور از راست در سانسور نوع I

۱-۱۰ سانسور بر روی فواصل (سانسور بازه ای)[۱۳]

این نوع سانسور به این مفهوم است که قسمتی از اطلاعات در دوره جمع آوری اطلاعات در یک یا چند فاصله زمانی جدا از هم از دست رفته باشد. به نوعی می توان گفت در این نوع سانسور فرد بعد از زمان صفر در حیطه مطالعه قرار می گیرد و آزمایش را به انتها نمی رساند یعنی قبل از زمان تعیین شده پایانی شکست می خورد. شکل (۱-۳) را ببینید.

۱- واحدهایی که بین دو زمان مشخص به از کار افتادگی خود رسیده اند.

۲- فرد ابتدا از شرکت در مطالعه کناره گیری کرده ولی از جایی وارد آزمایش شود و بعد از زمانی از شرکت در مطالعه کناره گیری کرده و یا تصمیم به کناره گیری می گیرد که می تواند به دلیل تغییر در وضعیت فرد مثل بهبود یافتن و یا وخیم شدن علایم بالینی باشد.



شکل (۱-۳): سانسور بر روی فواصل در سانسور نوع I

[۱۳] Interval censoring

۱-۸ سانسور نوع II

در این حالت مشاهده (آزمایش) تا زمان رخداد i امین شکست ادامه می یابد. فرض کنید هدف برآورد میانگین طول عمر یک قطعه تولید شده باشد. آزمایش را تا زمان خرابی قطعه ۵۰ ام ادامه می دهیم. بعد از این که ۵۰ امین قطعه خراب شد، زمان از کار افتادن سایر قطعه ها مشخص نیست. در این صورت تعداد واحدهایی که شکست می خورند از قبل مشخص و برابر تعداد ثابتی است که توسط محقق تعیین می شود (در این جا $i = ۵۰$) و زمان پایان آزمایش (T_0) تصادفی است.

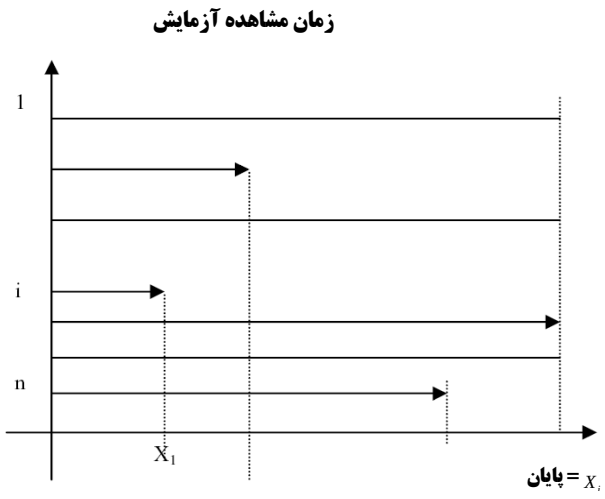
به بیان دیگر می توان سانسور نوع II را به زبان آماری به شکل زیر بیان کرد:

فرض کنید، $T_{(۱)} < T_{(۲)} < ... < T_{(n)}$ آماره های مرتب $T_۱, T_۲, ..., T_n$ باشند،وآزمایش را بعد از i امین از کار افتادگی مشاهدات خاتمه دهیم، دراین صورت در مورد مشاهدات سانسور شده داریم:

$$y_۱ = T_{(۱)}, y_۲ = T_{(۲)}, ..., y_n = T_{(n)} \qquad (i > 0 \text{ و ثابت})$$

به عنوان مثال، فرض کنید، n قطعه را در زمان $t = ۰$وارد آزمایش طول عمر کنیم و هنگامی آزمایش را خاتمه دهیم که i امین زمان شکست $(r = 1,2,...,i)$ را مشاهده کنیم. بنابراین اگر $T_1, T_2, ..., T_i$ زمان از کارافتادگی این i قطعه باشد، اطلاع دقیق در مورد زمان از کارافتادگی آنها را داریم و در مورد $n-i$ قطعه ی باقی مانده تنها این اطلاع را داریم که طول عمر آنها از $T_{(i)}$ بیشتر است. اصطلاحاً می گوییم $n-i$ قطعه سانسور شده اند. این نوع سانسور، سانسور نوع II است i ثابت و $T_{(i)}$متغیر تصادفی است.

شکل (۱-۴) را ببینید.



شکل (۱-۴): سانسور نوع II

پس از معرفی انواع سانسور، در فصل دوم متغیرها و رگرسیون سانسور شده را مورد بررسی قرار می دهیم.

فصل دوم

متغیرها و رگرسیون سانسور شده

۲-۱ مقدمه

در این فصل متغیرها و رگرسیون سانسور شده را مورد بررسی قرار می دهیم. متغیر سانسور شده متغیری است که کسر بزرگی از مشاهدات را از ابتدا یا انتها و یا میان داده ها در بر دارد. برآوردگرهای حداقل مربعات معمولی (OLS)[۱۴] برای چنین متغیرهایی اریب و پایا نمی باشند، بعبارت دیگر اریبی این برآوردها با افزایش حجم نمونه کاهش نمی یابد. این مطلب نشان می دهد که چگونه برآوردهای حداکثر درستنمایی پایا بوده و اریبی آنها با افزایش حجم نمونه کم می شود.

۲-۲ متغیرهای سانسور شده نرمال

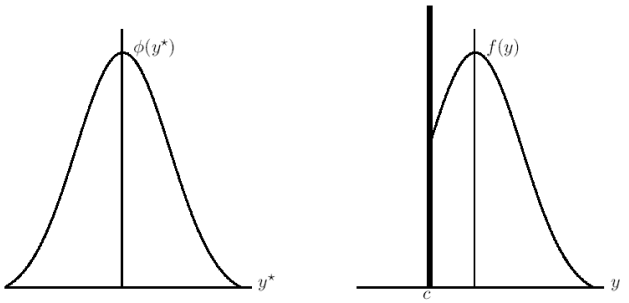
یک متغیر سانسور شده را می توان به این صورت تعریف کرد:

فرض کنید y^* دارای توزیع نرمال با میانگین μ و واریانس σ^2 باشد. یک متغیر مشاهده شده y از پایین سانسور شده است اگر

$y = c$ اگر $y^* \le c$

$= y^*$ در غیر اینصورت

که c ثابت در نظر گرفته می شود. این مطلب در شکل (۲-۱) نشان داده شده است.



شکل (۲-۱): y^* متغیر نرمال و y متغیر سانسور شده است.

فرض کنید ϕ و Φ تابع چگالی و تابع توزیع نرمال استاندارد باشند. تابع چگالی y به شکل زیر است.

$$f(y)=\left[\Phi\left(\frac{c-\mu}{\sigma}\right)\right]^{j}\left[\frac{1}{\sqrt{2\pi\sigma}}e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2}\right]^{1-j},$$

[۱۴] Ordinary least squares

که

$$\begin{cases} j=1 & \text{اگر} \quad y=c \\ j=0 & \text{در غیر اینصورت} \end{cases}$$

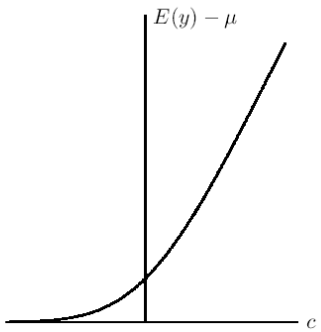
میانگین و واریانس y به شکل زیر است.

$$E(y)=\pi c+(1-\pi)(\mu+\lambda\sigma),$$
$$Var(y)=(1-\pi)\left[(1-\delta)+(\alpha-\lambda)^2\pi\right]\sigma^2,$$

که

$$\alpha=\frac{c-\mu}{\sigma},$$
$$\pi=\Phi(\alpha),$$
$$\lambda=\frac{\phi(\alpha)}{1-\Phi(\alpha)},$$
$$\delta=\lambda^2-\lambda\alpha.$$

میانگین و واریانس y برآوردهای پایایی از μ و σ^2 نیستند. اریبی میانگین $E(y)-\mu$ تابعی از c است که در شکل (۲-۲) برای $\mu=0$ و $\sigma=1$ نشان داده شده است.



شکل (۲-۲): اریبی میانگین $E(y)-\mu$ برای $\mu=0$ و $\sigma=1$

یک متغیر مشاهده شده y از بالا سانسور شده است اگر:

$$\begin{aligned} y &= c \quad \text{اگر} \quad y^* \geq c \\ &= y^* \quad \text{در غیر اینصورت} \end{aligned}$$

یک متغیر می تواند همزمان هم از بالا و هم از پایین سانسور شود. در همه این موارد میانگین μ و واریانس σ^2 می تواند بوسیله روش ماکسیمم درستنمایی برآورد شوند.

۲-۳ رگرسیون سانسور شده نرمال

معادله رگرسیون برآورد شده زیر را در نظر بگیرید

$y^* = \alpha + \gamma X + z$,

α عرض از مبدأ است و γ بردار ضرایب رگرسیونی روی متغیرهای توضیحی X است. فرض می شود که جمله خطای z دارای توزیع نرمال با میانگین صفر و واریانس ψ^2 است. اگر y^* به مانند متغیر y که در تمامی محدوده موجود مشاهده شده است، وجود داشته باشد برآورد معادله فوق صحیح است بعبارت دیگر اگر متغیر y مانند آنچه در فصل اول گفتیم کامل باشد می توان با استفاده از روش برآورد حداقل مربعات معمولی (OLS) پارامترها را برآورد نمود، اما اگر متغیر y از بالا یا پایین سانسور شده باشد برآورد حداقل مربعات معمولی y روی X اریب است و α و γ را می توان با استفاده از روش ماکسیمم درستنمایی که در نمونه های بزرگ نا اریب اند برآورد کرد.

۲-۴ برآورد پارامترهای مدل رگرسیون سانسور شده به روش ماکسیمم درستنمایی

معادله رگرسیونی به شکل زیر را در نظر می گیریم

$$y^* = \alpha^* + \gamma^{*\prime} X + z \,, \qquad (۱-۲)$$

α عرض از مبدأ است و γ^* بردار ضرایب رگرسیونی و X متغیرهای توضیحی هستند. فرض کنید z دارای توزیع نرمال با میانگین صفر و واریانس ψ^{*2} است. اگر هیچ متغیر توضیحی وجود نداشته باشد جمله دوم در معادله بالا وجود نخواهد داشت.

متغیر مشاهده شده y را به صورت زیر در نظر می گیریم:

$$\begin{aligned} y &= c_1 \quad \text{اگر} \quad y^* \le c_1 \\ &= y^* \quad \text{اگر} \quad c_1 < y^* < c_2 \\ &= c_2 \quad \text{اگر} \quad y^* \ge c_2 \,, \end{aligned}$$

که c_1 و c_2 اعداد ثابتی هستند. اگر y از پایین سانسور شده باشد $c_2 = +\infty$ ، اگر y از بالا سانسور شده باشد $c_1 = -\infty$ ، و اگر y سانسور نشده باشد $c_1 = -\infty$ و $c_2 = +\infty$ قرار می دهیم.

فرض کنید (y_i, X_i) مقادیر مشاهده شده y و X در یک نمونه تصادفی از N مشاهده مستقل باشند. تابع درستنمایی (y_i, X_i) به صورت زیر است:

$$L_i = \left[\Phi\left(\frac{c_1 - \alpha^* - \gamma^{*\prime} X_i}{\psi^*}\right)\right]^{j_{1i}} \left[\frac{1}{\sqrt{2\pi}\psi^*} e^{-\frac{1}{2}\left(\frac{y_i - \alpha^* - \gamma^{*\prime} X_i}{\psi^*}\right)^2}\right]^{1-j_{1i}-j_{2i}} \left[1 - \Phi\left(\frac{c_2 - \alpha^* - \gamma^{*\prime} X_i}{\psi^*}\right)\right]^{j_{2i}} ,$$

که

$$\begin{cases} j_{i1}=1 & \text{اگر} \quad y=c_1 \\ j_{i1}=0 & \text{در غیر اینصورت} \end{cases} \qquad \begin{cases} j_{i2}=1 & \text{اگر} \quad y=c_2 \\ j_{i2}=0 & \text{در غیر اینصورت} \end{cases}$$

توجه می کنیم که j_{i1} و j_{i2} همزمان نمی توانند برابر ۱ باشند.

تابع درستنمایی به صورت زیر بیان می شود.

$$\ln L=\sum_{i=1}^{N}\ln L_i$$

به طوری که نسبت به بردار پارامتر $\theta^{*'}=(\alpha^*,\gamma^{*'},\psi^*)$ ماکسیمم شود. اولین و دومین مشتق $\ln L$ نسبت به θ^* بسیار پیچیده است لذا برای حل این مشکل پارامترها را ساده تر می کنیم بدین منظور از بردار پارامتری $\theta'=(\alpha,\gamma',\psi)$ به جای بردار θ^* استفاده می کنیم که در آن $\alpha=\alpha^*/\psi^*$ ، $\gamma=\gamma^*/\psi^*$ و $\psi=1/\psi^*$ می باشد.

با ضرب جمله $\psi=1/\psi^*$ در معادله رگرسیونی (۲-۱) داریم:

$$\psi y^*=\alpha+\gamma' X+\upsilon\ ,$$

که $\upsilon=\psi z=z/\psi^*$ و دارای توزیع $N(0,1)$ است. پس داریم

$$y=c_1 \leftrightarrow y^*\le c_1 \leftrightarrow \psi y^*\le \psi c_1 \ \leftrightarrow\ \upsilon\le \psi c_1-\alpha-\gamma' X\ ,$$

$$y=c_2 \leftrightarrow y^*\ge c_2 \leftrightarrow \psi y^*\ge \psi c_2 \ \leftrightarrow\ \upsilon\ge \psi c_2-\alpha-\gamma' X.$$

بنابراین تابع درستنمایی L_i به صورت زیر می شود

$$L_i=\left[\Phi(\psi c_1-\alpha-\gamma' X_i)\right]^{j_{1i}}\left[\frac{1}{\sqrt{2\pi}}\psi\, e^{-\frac{1}{2}(\psi y_i-\alpha-\gamma' X_i)^2}\right]^{1-j_{1i}-j_{2i}}\left[1-\Phi(\psi c_2-\alpha-\gamma' X_i)\right]^{j_{2i}}$$

قرار می دهیم

$$\delta_i=\psi y_i-\alpha-\gamma' X_i\ .$$

پس $\ln L_i$ می شود

$$\ln L_i=-\ln\sqrt{2\pi}+(1-j_{1i}-j_{2i})(\ln\psi-\frac{1}{2}\delta_i^{\,2})+j_{1i}\ln\Phi(\delta_i)+j_{2i}\ln\left[1-\Phi(\delta_i)\right]\ .$$

مشتق های اول و دوم $\ln L_i$ را در ادامه آورده ایم.

$\partial\delta_i/\partial\alpha=-1$ ، $\partial\delta_i/\partial\gamma=-X_i$ و $\partial\delta_i/\partial\psi=y_i$. بعلاوه $\Phi'(t)=\phi(t)$ ، $\phi'(t)=-t\phi(t)$

و اگر $A(t)=\phi(t)/\Phi(t)$ ، در نتیجه

$$A'(t)=-A(t)\left[t+A(t)\right]=B(t),$$

با حذف اندیس i، مشتقات مورد نیاز به این شکل هستند

$$\partial \ln L / \partial \alpha = (1 - j_1 - j_2)\delta - j_1 A(\delta) + j_2 A(-\delta)$$

$$\partial \ln L / \partial \gamma = (1 - j_1 - j_2)X - j_1 A(\delta)X + j_2 A(-\delta)X$$

$$\partial \ln L / \partial \psi = (1 - j_1 - j_2)(1/\psi - \delta y) + j_1 A(\delta)y - j_2 A(-\delta)y$$

$$\partial^2 \ln L / \partial \alpha \partial \alpha = -(1 - j_1 - j_2) + j_1 B(\delta) + j_2 B(-\delta)$$

$$\partial^2 \ln L / \partial \gamma \partial \alpha = -(1 - j_1 - j_2)X + j_1 B(\delta)X + j_2 B(-\delta)X$$

$$\partial^2 \ln L / \partial \gamma \partial \gamma' = -(1 - j_1 - j_2)XX' + j_1 B(\delta)XX' + j_2 B(-\delta)XX'$$

$$\partial^2 \ln L / \partial \psi \partial \alpha = (1 - j_1 - j_2)y - j_1 B(\delta)y - j_2 B(-\delta)y$$

$$\partial^2 \ln L / \partial \psi \partial \gamma' = (1 - j_1 - j_2)yX' - j_1 B(\delta)yX' - j_2 B(-\delta)yX'$$

$$\partial^2 \ln L / \partial \psi \partial \psi = -(1 - j_1 - j_2)(1/\psi^2 + y^2) + j_1 B(\delta)y^2 + j_2 B(-\delta)y^2$$

ماکسیمم سازی $\ln L$ معادل مینیمم سازی تابع برازش $F(\theta) = -\ln L$ می باشد.
فرض کنید $g(\theta) = \partial F / \partial \theta$ بردار گرادیان[۱۵] و $H(\theta) = \partial^2 F / \partial \theta \partial \theta'$ ماتریس هسی[۱۶] باشد. آممیا[۱۷] (۱۹۷۳) ثابت نمود که $H(\theta)$ همه جا معین مثبت[۱۸] است.

تابع برازش $F(\theta)$ با استفاده از روش نیوتن–رافسون[۱۹] (یک روش الگوریتمی برای ساختن یک دنباله از تقریبات برای ریشه یک معادله است.) که بسیار سریع همگرا می شود، مینیمم می شود. مقدار شروع θ_0 با استفاده از روش OLS بدست می آید. برآوردهای متوالی با فرمول زیر داده می شود.

$$\theta_{s+1} = \theta_s - H_s^{-1} g_s \ ,$$

که در آن $g_s = g(\theta_s)$ و $H_s = H(\theta_s)$.

---

[۱۵] Gradient vector
[۱۶] Hessian matrix
[۱۷] Amemia
[۱۸] Positive definite
[۱۹] Newton-Rafson procedure

فرض کنید $\hat{\theta}=(\hat{\alpha},\hat{\gamma},\hat{\psi})$ برآوردهای ماکسیمم درستنمایی θ باشند. ماتریس کوواریانس مجانبی $\hat{\theta}$ برابر $E=H^{-1}(\theta)$ است که با پارامتر صحیح θ ارزیابی می شود. از تبدیل یک به یک θ به θ^*، برآوردهای ماکسیمم درستنمایی θ^* برابر $\hat{\theta}^*=(\hat{\alpha}^*,\hat{\gamma}^*,\hat{\psi}^*)$ است که $\hat{\alpha}^*=\hat{\alpha}/\hat{\psi}$ ، $\hat{\gamma}^*=\gamma/\hat{\psi}$ و $\hat{\psi}^*=1/\hat{\psi}$.

برای بدست آوردن ماتریس کوواریانس مجانبی $\hat{\theta}^*$، ماتریس $\frac{\partial\theta^*}{\partial\theta'}$ را بدست می آوریم

$$\frac{\partial\theta^*}{\partial\theta'}=(1/\psi^2)=\begin{pmatrix}\psi & 0' & -\alpha\\ 0 & \psi 1 & \gamma\\ 0 & 0' & -1\end{pmatrix}=A(\theta)\,,$$

که **0** و **1** در ماتریس بالا نشان دهنده بردارهای ستونی صفر و یک هستند. ماتریس کوواریانس مجانبی $\hat{\theta}^*$، AEA' است که A و E از روی مقادیر واقعی پارامتر محاسبه می شود. یک برآورد از ماتریس کوواریانس مجانبی $\hat{\theta}^*$ برابر AEA' می باشد که در آن از A و E برآورد شده استفاده کنیم. برآورد خطای استاندارد مجانبی پارامترهای برآورد شده با استفاده از ریشه دوم عناصر روی قطر اصلی ماتریس بدست می آید.