Stata
Intermedio
Análisis Microeconométrico

Juan Carlos Abanto Orihuela

15 de noviembre de 2012

Índice general

Sesión 1
Modelos de Elección Discreta

1.1. Estimación y Análisis

Las estimaciones lineales clásicas permiten la modelización de variables dependientes cuantitativas para identificar relaciones estadísticas en las que se asume una serie de supuestos sobre la forma del error de la ecuación lineal (homocedasticidad, normalidad, etc.). Sin embargo, en muchos contextos, el fenómeno que se quiere modelizar no es continuo sino discreto, por ejemplo cuando se quiere modelar la elección de compra de un bien o servicio; o la decisión de participar o no en el mercado laboral. Estos son los modelos conocidos como modelos de respuesta cualitativa. Llamamos variables cualitativas a aquellas que no aparecen en forma numérica, sino como categorías o atributos como por ejemplo, el sexo o la profesión de una persona. En general, se dice que una variable es discreta cuando está formada por un número finito de alternativas que miden cualidades.

1.1.1. Interpretación Estructural

Existen tres enfoques para la interpretación estructural de los modelos de elección discreta. El primero hace referencia a la modelización de una variable latente a través de una función índice, que trata de modelizar una variable inobservable o latente. El segundo de los enfoques permite interpretar los modelos de elección discreta bajo la teoría de la utilidad aleatoria, de tal manera que la alternativa seleccionada en cada caso será aquella que maximice la utilidad esperada. El tercero pasa por plantear un modelo de probabilidad no lineal.

Bajo el primero de los enfoques se trata de modelizar una variable índice, inobservable o latente no limitada en su rango de variación y*. Cuando la variable latente supera un determinado nivel, la variable discreta toma el valor 1, y si no lo supera toma el valor 0. La variable latente depende de un conjunto de variables explicativas que generan las alternativas que se dan en la realidad y que permiten expresar el modelo dicotómico como:

$$Y = \begin{cases} 1, & \text{si } Y^* > 0, \\ 0, & \text{si } Y^* \leq 0. \end{cases}$$

Donde el supuesto sobre la distribución de error determina el tipo de modelo a estimar. Si se supone una función de distribución uniforme, se utiliza el Modelo Lineal de Probabilidad truncado; si se distribuye como una normal con media cero y varianza uno, el modelo generado será un Probit; mientras que si se supone que se distribuye como una curva logística, se trataría de un modelo Logit. La hipótesis de que el umbral a superar por la variable latente sea cero se puede modificar por cualquier otro valor sugiriéndose, en determinados estudios, que el valor crítico sea el definido por el término constante.

Bajo este enfoque, el modelo probabilistico quedaría:

$$Y^* = X\beta + \epsilon$$

$$\begin{aligned} Pr(Y = 1/X) &= Pr(Y^* > 0/X) \\ Pr(Y = 1/X) &= Pr(\epsilon > -(X\beta)/X) \\ Pr(Y = 1/X) &= F(X\beta) \end{aligned}$$

Con el modelo así definido, la variable endógena del modelo dicotómico representa la probabilidad de ocurrencia del fenómeno analizado, siendo la probabilidad de que ocurra la opción 1 más elevada cuando mayor sea el valor de Y^*.

El segundo de los enfoques para la interpretación de los modelos de respuesta dicotómica es el que hace referencia a la modelización a través de la formulación de una utilidad aleatoria. Bajo este enfoque un individuo debe adoptar una decisión que le permita elegir entre dos alternativas excluyentes, la 1 o la 0, lo que hará maximizando la utilidad esperada que le proporciona cada una de las alternativas posibles sobre las que tiene que decidir. Es decir, el individuo i-ésimo elegirá una de las dos alternativas dependiendo de que la utilidad que le proporciona dicha decisión sea superior a la que le proporciona su complementaria.

La formulación del modelo bajo esta teoría parte del supuesto de que la utilidad derivada de una elección, U_{i0} o U_{i1}, es función de las variables explicativas de dicha decisión, que son las características propias de cada una de las alternativas de elección y las características personales propias del individuo, de manera que suponiendo linealidad en las funciones, se tiene:

$$U_{i0} = \alpha_0 + X_{i0}\beta + \epsilon_{i0}$$
$$U_{i1} = \alpha_1 + X_{i1}\beta + \epsilon_{i1}$$

Donde los ϵ_{ij} recogen las desviaciones que los agentes tienen respecto a lo que sería el comportamiento del agente medio y que se debe a factores aleatorios. El agente i elegirá la opción 1 si la utilidad de esa decisión supera la de la opción 0 y viceversa, de manera:

$$Y_i = \begin{cases} 1, & \text{si } U_{i1} > U_{i0}, \\ 0, & \text{si } U_{i1} < U_{i0}. \end{cases}$$

Y el modelo dicotómico quedaría definido por:

$$Pr(Y = 1/X) = Pr(U_{i1} > U_{i0}/X) = Pr(\epsilon_{i1} - \epsilon_{i0} > -(X\theta)/X)$$
$$Pr(Y = 1/X) = F(X\theta)$$

Según que la función asociada a la perturbación aleatoria ϵ_{ij} (que será la función de distribución, $F(X\theta)$, que se suponga siga dicha probabilidad), sea una función de distribución uniforme, la función de distribución de la normal tipificada o la de la curva logística, se obtienen el Modelo Lineal de Probabilidad Truncado, el Probit o el Logit, respectivamente.

El tercer enfoque pasa por estructurar un modelo de probabilidad no lineal, como lo sugiere Theil - 1970, de tal manera que:

$$Pr(Y = 1/X) = M_i = \frac{exp(X\beta)}{1+exp(X\beta)}$$

$$\Omega(x) = \frac{Pr(Y=1/X)}{Pr(Y=0/X)} = \frac{Pr(Y=1/X)}{1-Pr(Y=1/X)}$$

$$Ln(\Omega(x)) = X\beta + \epsilon$$

Es decir medir que tan a menudo ocurre algo (Y=1), respecto a que tan a menudo no ocurre (Y=0).

1.1.2. Modelo de Probabilidad Lineal

La primera alternativa teórica desarrollada para estudiar modelos con variables dicótomas se planteó como una extensión del modelo lineal general:

$$Y_t = \alpha_t + X_{kt}\beta_k + \epsilon_t$$

$$Donde:$$
$$Y_t = \begin{cases} 1, & \text{si ocurre una alternativa,} \\ 0, & \text{en caso contrario.} \end{cases}$$
X_{kt}=Variables explicativas
ϵ_t=Variable aleatoria que se distribuye $N(0,\sigma^2)$

En general, la distribución de los modelos de elección binaria se caracteriza por configurar una nube de puntos de tal manera que las observaciones se dividen en dos subgrupos. Uno de ellos esta formado por las observaciones en las que ocurrió el acontecimiento objeto de estudio (Y_i =1), y el otro, por los puntos muéstrales en los que no ocurrió (Y_i =0).Para el desarrollo de los modelos de elección discreta se utilizará la base de datos "labora.dta".

```
use labora.dta, clear
```

Antes de desarrollar el modelo de probabilidad lineal, es posible obtener una descripción rápida de la base de datos a utilizar, el comando describe mostrará el tipo de información con la que se cuenta. Esta base de datos hipotética contiene 400 observaciones en las que se detalla si el postulante es admitido a un programa de Post Grado (admit), el puntaje obtenido en la prueba Graduate Record Exam (gre), el puntaje obtenidio en el pregrado (Grade Point Average, gpa) y finalmente se considera si el postulante proviene de una universidad de prestigio o no (topnotch). Seguidamente se procederá a estimar la regresión lineal en donde la variable dependiente admit esta explicada por el puntaje obtenido en el gpa.

```
regress admit gpa
```

Problemas con esta estimación

La interpretación de los coeficientes en los modelos de probabilidad es similar a la de los modelos de regresión lineal, en donde el valor de los parámetros recoge el efecto de una variación unitaria en cada una de las variables explicativas sobre la probabilidad de ocurrencia del acontecimiento objeto de estudio, sin embargo, el MPL presenta algunas inconsistencias.

Se puede apreciar en el modelo inicial que algunos de los valores estimados se encuentran fuera de rango, lo cual carece de lógica considerando que deben interpretarse como probabilidades.

```
tw sc y admit gpa
```

Solución: ¿Modelo de probabilidad truncada?

A través del gráfico de la densidad de Kernel para el modelo que incluye todas las variables, se observa que los residuos no se distribuyen de manera normal, por lo tanto no es eficiente, es decir, pueden presentarse problemas de minimización de la varianza a medida que la muestra aumenta.

```
kdensity r, normal
```

¿Invalida esto la estimación por MCO? ¿Los estimadores siguen siendo MELI (BLUE)?

Problemas de Heterocedasticidad. Aún en el caso de que se cumpliesen las hipótesis de media y correlación nula en la perturbación aleatoria E(ϵ_i) = 0 E(ϵ_i, ϵ_j) = 0 para todo $i \neq j$, no se cumple la hipótesis de varianza constante, es decir, la perturbación aleatoria no es homocedástica.

$$Var(\epsilon_t) = E[(\epsilon_i - E(\epsilon_i))(\epsilon_i - E(\epsilon_i))'] = E(\epsilon_i^2)$$

$$Var(\epsilon_t) = (1 - X\beta)^2 f_i(1) + (0 - X\beta)^2(1 - f_i(1))$$

$$Var(\epsilon_t) = (1 - f_i(1))^2 f_i(1) + (f_i(1))^2(1 - f_i(1))$$

$$Var(\epsilon_t) = (1 - f_i(1))f_i(1)$$

En STATA es posible realizar un análisis tanto gráfico como a través de números índice para verificar la presencia de heterocedasticidad.

```
rvfplot, yline (0)
hettest
```

Para el presente ejemplo la hipótesis nula de varianza constante (homocedasticidad) será rechazada debido a que el p value de la distribución del estadístico chicuadrado es muy pequeño, aceptándose la hipótesis alterna de varianza no homogénea.

Solución: ¿MCG o MCP?

1.1.3. Modelo de Probabilidad No Lineal

Los problemas en la interpretación y estimación de los parámetros del modelo de probabilidad lineal han llevado a la búsqueda de modelos alternativos que permitan estimaciones más fiables de las variables dicótomas. Es el caso de los modelos de probabilidad no lineal, donde la función de especificación utilizada garantiza un resultado en la estimación comprendido en el rango 0-1. Estos son los modelos logit y probit. Analizaremos a continuación los datos a través de una regresión logística, la cual se formula a continuación.

$$Pr(Y=1) = \frac{e^{X\beta}}{1+e^{X\beta}} = \Delta(X\beta)$$

```
logit admit gpa
predict l
tw sc l admit gpa
```

Pos-estimación

a. Test de efectos individuales

Si los supuestos bases del modelo se sostienen, los estimadores son distribuidos de manera asintótica y normal:

$$\hat{\beta}_k \xrightarrow{a} N(\beta_k, \sigma^2_{\hat{\beta}_k})$$

Donde la hipótesis nula de significancia del parámetro puede ser testeada a partir de:

$$z = \frac{\hat{\beta}_k - \beta^*}{\sigma^2_{\hat{\beta}_k}}$$

Si la hipótesis nula es verdadera entonces z se distribuirá aproximadamente como una normal con media cero y varianza unitaria para muestras grandes.

b. Test de Wald

Podemos analizar el modelo una vez estimado, mediante un testeo de hipótesis que validen una correcta especificación. Para esto el test de Wald calculado para hipótesis lineales sobre los parámetros de los modelos estimados nos será de mucha utilidad. También puede usarse el test bajo una estructura no lineal, la cual no abordaremos en esta sección.

```
logit admit gre gpa topnotch

test gpa=0
test gre=gpa, accumulate
```

c. Test LR

El estadístico de verosimilitud también nos será de gran utilidad para evaluar mediante hipótesis la significacia de modelos. Este estadístico compara modelos anidados.

```
logit admit gre gpa topnotch
lrtest, saving(0)
logit admit gre gpa
lrtest
```

Donde nuestra hipótesis nula es $H_0 = \beta_{topnotch} = 0$

```
logit admit gre gpa topnotch
lrtest, saving(M1)
logit admit gre gpa
lrtest, using(M1)

logit admit gre gpa topnotch
lrtest, saving(0)
logit admit gre gpa
lrtest, saving(1)
lrtest, using(1) model(0)
```

Muchas medidas escalares han sido desarrolladas para resumir las bondades de ajuste de modelos de regresión continuo o de variables categóricas. Sin embargo no hay evidencia convincente de selección de un modelo que maximice los valores de una medida comparada con la medida de otro modelo. Mientras las medidas de ajuste proveen información, esta es solo parcial, que debería ser sostenida con una teoría económica razonable, o investigaciones anteriores como referencia.

El comando Fifstat nos permite obtener una tabla con estadísticos que ayudaran a evaluar la bondad de ajuste del modelo. De los cuales analizaremos algunos.

d. Fitstat

A continuación proveeremos de una breve descripción de cada una de las medidas que computa el "fitstat". Mayores detalles de las medidas las podemos encontrar en Long(1997).

- **Medida basada en Log-Likehood**

Stata comienza su análisis maximizando iteracciones de verosimilitud y calculando sus logaritmos, para determinado modelo, con todos los parámetros excepto el intercepto en un nivel de cero L($M_{intercepto}$), mientras que cuando los parámetros son diferentes de cero, el logaritmo de verosimilitud calculado será L(M_{full})

- **Test Chi-Cuadrado de todos los coeficientes**
  Un test LR donde la hipótesis nula de que todos los coeficientes excepto el intercepto son ceros puede ser calculado comparando el logaritmo de verosimilitud LR=2[Ln(M_{full})-Ln($M_{intercepto}$)], a veces a este estadistico se le designa con el valor G2. El LR es reportado por Stata como chi2(gl), donde los gl son el número de parámetros restringidos.

- **Desviación**
  La desviación compara un modelo dado con un modelo que tienen un parámetro para cada observación, así el modelo reproduce perfectamente la data observada. La desviación es definida como D= -2Ln(M_{full}) con N-K gl. Notar que esta medida no es una chi2.

- **McFadden's R^2**
  R^2 en MRL
  Para una regresión lineal el "fitstat" reporta el coeficiente de determinación estándar:

$$R^2 = 1 - \frac{\sum_1^N (y_i - \hat{y}_i)^2}{\sum_1^N (y_i - \bar{y}_i)^2} = \frac{Var(\hat{y})}{Var(\hat{y}) + Var(\hat{e})} = 1 - \left[\frac{L(M_{intercepto})}{L(M_{full})}\right]^{\frac{2}{N}}$$

  Y el R^2 ajustado seria:

$$\tilde{R}^2 = \left(R^2 - \frac{K}{N-1}\right)\left(\frac{N-1}{N-K-1}\right)$$

  R^2 en MRNL
  En modelos no lineales la medida calculada por Stata son los pseudos R^2. El R^2 de McFadden, también conocido como el índice del ratio de verosimilitud, compara dos modelos:

$$R^2_{McF} = 1 - \left[\frac{LnL(M_{full})}{LnL(M_{intercepto})}\right]$$

  Y como el R^2 de McFadden siempre se incrementa con el numero nuevo de variables explicativas, se ajusta su versión con:

$$R^2_{McF} = 1 - \left[\frac{LnL(M_{full}) - K^*}{LnL(M_{intercepto})}\right]$$

Donde K^* es el numero de variables independientes, no el numero de parámetros.

- **R^2 de Máxima Verosimilitud**

  Es otra medida análoga al R2 en el MRL, sugerido por Maddala:

$$R^2_{ML} = 1 - \left[\frac{L(M_{intercepto})}{L(M_{full})}\right]^{\frac{2}{N}} = 1 - exp\left(\frac{-G^2}{N}\right)$$

- **Cragg & Uhler's R^2**

  La corrección a la estimación anterior R^2_{ML} propuesta por Cragg y Uhler fue:

$$R^2_{C\&U} = \frac{R^2_{ML}}{max(R^2_{ML})} = \frac{1 - \left[\frac{L(M_{intercepto})}{L(M_{full})}\right]^{\frac{2}{N}}}{1 - [L(M_{intercepto})]^{\frac{2}{N}}}$$

- **Efron's R^2** Para salidas binarias, el R^2 Efron define el "y" estimado como:

$$\hat{y} = \hat{\pi} = Pr(y = 1/x)$$

$$R^2_{Efron} = 1 - \frac{\sum_1^N (y_i - \hat{\pi}_i)^2}{\sum_1^N (y_i - \bar{y}_i)^2}$$

- **V(y*), V(e) - McKelvey & Zavoina's R^2**

  Algunos modelos pueden ser definidos en términos de una variable latente y^*. Esto compete a modelos de elección binaria u ordinal, como a algunos modelos censurados, así:

  Dado el modelo $Y^* = X\beta + \epsilon$ y usando $\hat{V}ar(\hat{Y}^*) = \hat{\beta}'\hat{V}ar(x)\hat{\beta}$, McKelvey y Zavoina proponen:

$$R^2_{M\&Z} = \frac{\hat{V}ar(\hat{y}^*)}{\hat{V}ar(\hat{y}^*)} = \frac{\hat{V}ar(y^*)}{Var(y^*) + Var(\epsilon)}$$

- **El R^2Count y el R^2Count Ajustado**

  De los valores observados y predichos, se calcula el R^2_{Count}. El comando "lstat" nos provee de valores de aciertos y desaciertos de los computados por el modelo.

  Definimos así el R^2_{Count} como:

$$R^2_{Count} = \frac{1}{N}\sum_j n_{jj}$$

donde n_{jj} es el numero de predicciones correctas en la tabla. Pero el R^2_{Count} puede darnos una interpretación fallida del poder de predicción del modelo. En un modelo binario sin previo conocimiento de las variables independientes es posible corregir las predicciones en al menos el 50 % de los casos eligiendo una categoría con el mayor porcentaje de casos observados. El ajuste se hace de la siguiente manera:

$$R^2_{Count} = \frac{\sum_j n_{jj} - max_r(n_{++})}{N - max_r(n_{++})}$$

Donde n_{++} es el mayor valor marginal de la ultima fila.

- **Medidas de Información**
  AIC
  Este criterio compara modelos de diferentes tamaños de muestra o también modelos no anidados. Akaike (1973) definió:

  $$AIC = \frac{-2Ln\hat{L}(M_k) + 2p}{N}$$

  Donde "p" es el número de parámetros en el modelo (K+1 en los modelos de regresión binaria donde K es el número de regresores)
  BIC
  El criterio de información Bayesiana fue propuesto por Raftery (1996) como una medida que compara modelos anidados como modelos no anidados. Definimos BIC de la siguiente manera:

  $$BIC_K = D(M_K) - gl_k Ln(N)$$

  Donde gl_k son los grados de libertad asociados con la desviación. La segunda versión de BIC es basada al ratio de verosimilitud del Chi2 con gl'_k definiendo dichos grados de libertad, como el numero de regresores (no parámetros) en le modelo.

  $$BIC'_K = -G^2(M_K) - gl'_k Ln(N)$$

```
quietly  logit admit gre gpa topnotch, nolog
quietly fitstat, saving(M1)
quietly  logit admit gpa , nolog
fitstat, using(M1)
```

Otra posible solución a las inconsistencias que presenta el modelo de probabilidad lineal para explicar el comportamiento de una variable dependiente binaria es el uso del modelo probit de la forma:

$$y = f(\beta_0 + \beta_1 x_1 + ... + \beta_k x_k) + \epsilon$$

Donde f es la función de distribución normal estándar

$$f(X\beta) = \int_{-\infty}^{X\beta} \frac{1}{\sqrt{2\pi}} e^{\frac{s^2}{2}} \, ds + \epsilon_i$$

```
probit admit gre gpa topnotch

quietly  probit admit gre gpa topnotch, nolog
estimates store A, title(Modelo Probit)
quietly  logit admit gre gpa topnotch, nolog
estimates store B, title(Modelo logit)
est table A B, stat(aic bic)
est table A B, stat(aic bic) star
est table A B, stat(aic bic) star b(%9.3f)
est table A B, stat(aic bic) b(%9.3f) se(%9.2f)  t(%9.2f) p(%7.2f)
est table A B, stat(aic bic) b(%9.3f) se(%9.2f)  t(%9.2f) p(%7.2f)
est table A B, stat(aic bic rank N ll chi2) b(%9.3f) se(%9.2f) ///
t(%9.2f) p(%7.2f)  label
```

ODD Ratios

Respecto a la interpretación de los parámetros estimados tanto en el modelo logit como en el modelo probit, la cuantía del parámetro no coincide con la magnitud de las variaciones en la probabilidad (como en el MLP), una interpretación más sencilla de los parámetros estimado es la que se obtiene a través de la linealización del modelo. En el caso del modelo logit.

$$E(y_i) = Prob(y_i = 1) = M_i = \frac{e^{\beta_0 + \beta_k x_{kt}}}{1 + e^{\beta_0 + \beta_k x_{kt}}}$$

Donde

$$M_i + M_i e^{\beta_0 + \beta_k x_{kt}} = e^{\beta_0 + \beta_k x_{kt}}$$

$$\frac{M_i}{1 - M_i} = e^{\beta_0 + \beta_k x_{kt}}$$

Al cociente de las probabilidades de que se elija la opción 1, frente a la posibilidad de que se elija la opción 0, se le denomina Odds:

$$Odds = \frac{Pr(y = 1)}{1 - Pr(y = 1)}$$

Mientras que el concepto de Odds Ratio se define como el cociente de dos odds asociados.

Veamos un ejemplo
Digamos que la probabilidad de éxito de un evento es 0.8, entonces p=0.8
Entonces la probabilidad de falla del evento será: q=0.2
El odds de éxito es definido como:
Odds(éxito)=p/q=0.8/0.2=4, es decir, el odds de éxito es de 4 a 1.
El odds de falla debe ser entonces:
Odds(falla)=q/p=0.2/0.8=0.25
Tanto el odds de éxito como el de falla son recíprocos.

Veamos otro ejemplo
Adaptado por Pedhazur(1997). Supongamos que siete de 10 hombres son admitidos a una escuela de ingenieros, mientras que tres de 10 mujeres también son admitidas. La probabilidad de ser admitido para los hombres es:
p=7/10=0.7; q=1-0.7=0.3
La probabilidad de ser admitido para las mujeres es:
p=3/10=0.3; q=1-0.3=0.7
Podemos usar las probabilidades y computar el odds de admisión para ambos sexos:
Odds(hombres)=0.7/0.3=2.333 Odds(mujeres)=0.3/0.7=0.42857
Finalmente calculamos el Odds ratio para admisión:
OR=2.333/0.42857=5.44
Así, para los hombres el odds de ser admitidos es 5.44 veces mayor que el odds para la admisión de mujeres. Es decir es mas probable o menos riesgoso el hecho de ser admitidos de un hombre que de una mujer.

El odds en la regresión logística:

$$Odd = \Omega = \frac{Pr(y=1)}{Pr(y=0)} = \frac{M_i}{(1-M_i)} = e^{\beta_0+\beta_k x_{kt}}$$

Entonces

$$Ln\Omega = \beta_0 + \beta_k x_{kt}$$

Lo cual nos indica que por cada unidad de cambio en X_k , esperamos que el logit cambie en β_k manteniendo las demás variables constantes.

El problema estriba en que un cambio en β_k en el ln del odds tenga un significado muy claro para muchas personas. Por ello tomaremos un modelo multiplicativo a partir del cual pasaremos al análisis.

$$\Omega(x, x_k) = e^{\beta_0} e^{\beta_1 x_1} e^{\beta_2 x_2} ... e^{\beta_k x_k}$$

Si nosotros hacemos que X_k cambie en una unidad, entonces:

$$\Omega(x, x_{k+1}) = e^{\beta_0} e^{\beta_1 x_1} e^{\beta_2 x_2} ... e^{\beta_k x_{k+1}}$$

Lo cual nos conduce al odd ratio:

$$\frac{\Omega(x, x_{k+1})}{\Omega(x, x_k)} = \frac{e^{\beta_0} e^{\beta_1 x_1} e^{\beta_2 x_2} ... e^{\beta_k x_{k+1}}}{e^{\beta_0} e^{\beta_1 x_1} e^{\beta_2 x_2} ... e^{\beta_k x_k}} = e^{\beta_k}$$

Por cada unidad de cambio en X_k esperamos que el odds cambie en el factor e^{β_k}, siendo lo demás constante.

Medidas superiores a uno de e^{β_k}, quiere decir que los odds son e^{β_k} veces mayores, mientras que medidas inferiores a uno de e^{β_k}, quieren decir que los odds son e^{β_k} veces menores.

Veamos un ejemplo

Aquí codificamos a admit como 1 para si, y 0 para no, gender es codificado como 1 para hombre y 0 para mujeres. El comando Logistic produce resultados en términos de odds ratios, mientras Logit produce resultados en términos de coeficientes.

```
clear
input admit gender freq
1 1 7
1 0 3
0 1 3
0 0 7
end
logistic admit gender [weight=freq]
logit admit gender [weight=freq]
```

Noten el Z-value en ambas regresiones.

Existe una relación entre los coeficientes que produce Logit y los odds ratios que produce Logistic. Primero un Logit es definido como un logaritmo base e (log) de un odds:

$$Logit(p) = Log(odds) = Log(p/q)$$

Una regresión Logística es una relación ordinal, usa el logit como la variable dependiente:

$$Logit(p) = \beta_0 + \beta_k X_k$$
$$log(p/q) = \beta_0 + \beta_k X_k$$

Esto significa que el coeficiente en una regresión logística esta en términos del Log(odds), es decir el valor de 1.694596, lo que implica que una unidad de cambio en el genero, altera en 1.694596 unidades al log(odds). Podemos decir entonces que:

$$p/q = e^{\beta_0 + \beta_k X_k}$$

$$OR = e^{\beta} = e^{1{,}694596} = 5{,}44$$

Veamos la base de datos labora y corramos un logit:

```
logit admit gre topnotch gpa
```

```
Logistic regression                               Number of obs   =        400
                                                  LR chi2(3)      =     130.22
                                                  Prob > chi2     =     0.0000
Log likelihood = -189.26319                       Pseudo R2       =     0.2560
```

admit	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
gre	.0067809	.0013239	5.12	0.000	.004186	.0093758
topnotch	.7623019	.3197932	2.38	0.017	.1355187	1.389085
gpa	1.597449	.2935155	5.44	0.000	1.022169	2.172728
_cons	-10.23092	1.182362	-8.65	0.000	-12.54831	-7.913537

```
listcoef, help
```

admit	b	z	P>|z|	e^b	e^bStdX	SDofX
gre	0.00678	5.122	0.000	1.0068	2.1926	115.7813
topnotch	0.76230	2.384	0.017	2.1432	1.3386	0.3826
gpa	1.59745	5.442	0.000	4.9404	2.7419	0.6314

```
       b = raw coefficient
       z = z-score for test of b=0
   P>|z| = p-value for z-test
     e^b = exp(b) = factor change in odds for unit increase in X
 e^bStdX = exp(b*SD of X) = change in odds for SD increase in X
   SDofX = standard deviation of X
```

¿Que podemos decir de los resultados en ambas tablas?

1.1.4. Análisis de Probabilidades y Cambios Marginales

Los efectos marginales suelen proporcionar una buena aproximación del cambio que la presencia o no de la variable binaria o continua, originaria sobre la probabilidad predicha de algún modelo.

Para analizar esto, veremos los comandos “prvalue”, “prtab”, “prgen”, “prchange”, “mfx”.

PRVALUE
Calcula valores predichos de la endógena, para especificaciones de las variables independientes, pudiendo calcular diferencias en predicciones para dos set de valores.

PRTAB
Crea una tabla de endógena predicha, para un cruce de clasificaciones por encima de cuatro categorías de variables independientes, mientras las restantes son mantenidas en valores específicos.

PRCHANGE
Calcula el cambio discreto o continuo de la variable endógena predichas.

PRGEN
Calcula valores predichos de la endógena, cuando una variable independiente cambia sobre un rango especificado, manteniendo las demás variables constantes.

RETO 1

a. Con la base de datos “highschool” genere una variable latente llamada “hiwrite” que marque el valor de la unidad si la nota de escritura supera al menos 52, luego estime un logit con las variables explicativas read, female y prog. Realice un análisis econométrico y responda:

- ¿Cuál es el efecto de un cambio en los parámetros de las variables explicativas sobre hiwrite?.
- ¿Como interpretarias los ODDs de la regresión logistica?
- ¿Realice un testeo de hipótesis LM sobre un modelo que incluya vs uno que no incluya los efectos de la variable “prog”, programa academico?.
- ¿Cómo cambian las probabilidades si prog_2=0 y prog_2=1?.

Sesión 2
Modelos de Elección Ordinal

Cuando la variable dependiente es discreta, pero sus valores indican un orden, no es correcto realizar la estimación de la misma a través de los modelos presentados en el apartado anterior, ya que la inclusión de la información que aporta el orden de las alternativas en la especificación del modelo permite obtener unos mejores resultados.

Las variables ordinales son a menudo codificadas como enteros consecutivas de 1 al número de categorías, no sería correcto el uso de un modelo de regresión clásico, ya que codificadas las posibles alternativas como 1, 2, ...(j+1), ..., J, se estaría considerando la diferencia entre (j+1) y (j+2) como la existente entre 1 y 2, lo cual no tiene porque ser así ya que los números utilizados en la codificación solo representan un orden dentro de una clasificación. Así, con modelos de salida ordinal es mejor usar modelos que eviten el supuesto de que las distancias entre las categorías sean iguales, ahora nos enfocaremos en un logit y probit que consideren esta ordenación, modelos introducidos por McKelvey y Zavoina (1975) en términos de una variable latente.

Cuando las salidas son ordinales o nominales la dificultad de explicar más de dos respuestas se incrementa. Una variable puede ser ordenada de cierta manera cuando consideramos un tema, y ordenada de otra manera cuando consideramos un tema diferente. Millar y Volker (1985) mostraron como diferentes supuestos sobre el ordenamiento de ocupaciones, proyectan diferentes resultados. Una variable podría reflejar ordenamiento sobre más de una dimensión tal como escalas de actitudes, que reflejen ambas la intensidad y dirección de opinión. Mas aún es muy común que encuestas incluyan la categoría "no sabe, no opina", lo cual probablemente no corresponda a la categoría intermedia en una escala, aun cuando en el análisis uno este tentado a colocarla como tal, sobretodo cuando la propuesta de ordenamiento es ambigua, el modelo de salidas nominales podría ser considerado.

2.1. Estimación y Análisis

Los MRO pueden ser desarrollados de diferentes maneras, cada una de ellas nos conduce al mismo resultado. El modelo de regresión binaria (MRB) pueden ser vistos como un caso especial de los MRO, en el cual la variable endógena solo tiene dos categorías.

2.1.1. Modelo de Variable Latente

El modelo de regresión ordinal es comúnmente presentado como un modelo de variable latente. Definida y^* como una variable latente cuyo rango va desde -∞ a ∞

$$y_i^* = x_i\beta + \epsilon$$

Donde la variable endógena toma los siguientes valores:

$$y_i = m, \text{ si } r_{m-1} \leq y_i^* < r_m \; \forall m = 1 ... J$$

O también de manera extendida:

$$y_i = \begin{cases} 1, & \text{si } -\infty = r_0 \leq y_i^* < r_1, \\ 2, & \text{si } r_1 \leq y_i^* < r_2, \\ 3, & \text{si } r_2 \leq y_i^* < r_3, \\ \vdots & \\ J, & \text{si } r_{J-1} \leq y_i^* < r_J = \infty. \end{cases}$$

Donde los puntos de corte r_j son estimados. Como ejemplo, podríamos tener la siguiente pregunta en una encuesta: ¿Una mujer trabajadora establece un fuerte y seguro vinculo con su hijo, así como una mujer que no trabaja?

Las posibles respuestas podrían ser: 1=Desacuerdo Total, 2=Desacuerdo, 3=Acuerdo, 4=Acuerdo Total

La variable latente continua puede imaginarse como el grado de aceptación a favor de que las mujeres trabajadoras son buenas madres.

$$y_i = \begin{cases} 1 = DT, & \text{si } -\infty = r_0 \leq y_i^* < r_1, \\ 2 = D, & \text{si } r_1 \leq y_i^* < r_2, \\ 3 = A, & \text{si } r_2 \leq y_i^* < r_3, \\ 4 = AT, & \text{si } r_3 \leq y_i^* < r_4 = \infty. \end{cases}$$

La probabilidad de una variable observada dado el valor de x, corresponde a la región en la que la distribución de y^* cae entre r_{m-1} y r_m

$$Pr(y = m/x) = Pr(r_{m-1} \leq y^* < r_m/x)$$

Sustituyendo $x\beta + \epsilon$ por y^* y usando algo de algebra obtenemos la formula estándar que predice la probabilidad en el MRO

$$Pr(y = m/x) = F(r_m - x\beta) - F(r_{m-1} - x\beta)$$

Donde F es la función de probabilidad acumulada para ϵ. En el probit ordinal, F es una normal con Var(ϵ)=1, en el logit ordinal, F es una logistica con Var(ϵ)=$\frac{\pi^2}{3}$. Notar que cuando y=1 el termino F(-$\infty - x\beta$)=0 y cuando y=J el primer termino de F($\infty - x\beta$)=1.

Comparando estas ecuaciones con las de un MRB se observa que el MRO es idéntico a la regresión binaria, veamos:

```
use mroz
logit  inlf kidslt6 kidsge6 age wage hushrs faminc, nolog
outreg using salida,replace
ologit  inlf kidslt6 kidsge6 age hushrs faminc, nolog
outreg using salida, append
```

Los coeficientes y sus desviaciones estándar son los mismos pero el intercepto para el logit, es reportado, mientras que para el ologit ese intercepto es reemplazado por el punto de corte del mismo nivel pero de signo opuesto.

En Stata, la identificación del MRO asume que el intercepto es cero y así los valores de los puntos de corte son estimados.

El modelo de regresion ordinal puede tambien ser desarrollado como un modelo de probabilidad no lineal sin recurrir a la idea de variable latente. Para mostrar esto, primero definimos el odds de que la variable explicada es menor o igual a "m" vs que sea mayor que "m" dado las variables exogenas "x":

Por ejemplo, podriamos calcular el odds de desagrado o fuerte desagrado, versus el agrado o fuerte agrado. Asi el logaritmo del odds es igual a:

$$\Omega_{\leq m|>m} = \frac{Pr(y \leq m/x)}{Pr(y| > m/x)}$$

Para una simple variable independiente y tres categorías en la explicada, donde el intercepto fue fijado en "0", tendriamos:

$$Ln(\frac{Pr(y \leq 1/x)}{Pr(y > 1/x)}) = r_1 - \beta_1 x_1$$
$$Ln(\frac{Pr(y \leq 2/x)}{Pr(y > 2/x)}) = r_2 - \beta_1 x_1$$

Parece confuso que el modelo substraiga xb en lugar de añadirlo, esto es consecuencia del calculo del logit de $y \leq m$ vs $y > m$.

Aquí un ejemplo basado en la encuesta realizada entre 1977 y 1989 de General Social Survey, donde el tema y pregunta tratado fue: "¿Una madre trabajadora puede establecer una calida y segura relacion sentimental con su hijo como una madre que no trabaja?"

```
use warm, clear
describe
table warm
summarize
tab warm
```

Usando los datos, nosotros estimamos el siguiente modelo:

$$Pr(warm = m/x_i) = F(r_m - x\beta) - F(r_{m-1} - x\beta)$$

Donde

$$x\beta = \beta_{yr89} yr89 + \beta_{male} male + \beta_{white} white + \beta_{age} age + \beta_{prst} prst$$

Aquí las salidas sean con ologit, oprobit, pueden ser comparadas con el outreg:

```
ologit  warm yr89 male white age ed prst,nolog
outreg using ordenado,replace
oprobit  warm yr89 male white age ed prst,nolog
outreg using ordenado,append
```

Como en el análisis de los modelos de regresión binaria, la diferencia estriba en que los coeficientes tienen una razón de 1.7, es decir, solo hay diferencia en escala, sin embargo los z-test, son los mismos y no se ven afectados por la escala.

2.1.2. Testeo de Hipótesis

Para el testeo de hipótesis podremos usar el test de wald, máxima verosimilitud, o usar el fitstat para elegir el mejor modelo.

```
ologit  warm yr89 male white age ed prst,nolog
test male

ologit  warm yr89 male white age ed prst,nolog
test age  white male

ologit  warm yr89 male white age ed prst,nolog
lrtest, saving(0)
ologit  warm yr89 white age ed prst,nolog
lrtest

ologit  warm yr89 male white age ed prst,nolog
lrtest, saving(0)
ologit  warm yr89 ed prst,nolog
lrtest

version 8
ologit  warm yr89 male white age ed prst,nolog
fitstat
```

2.1.3. Supuesto de Paralelismo

Antes de discutir la interpretación, es importante entender un supuesto que esta implícito en el MRO, conocido como paralelismo de la regresión, y para el modelo ologit, el supuesto de odds proporcional.

$$
\begin{aligned}
&Pr(y = 1/x) = F(r_m - x\beta)\\
&Pr(y = m/x) = F(r_m - x\beta) - F(r_{m-1} - x\beta), cuando : m = 2...J - 1\\
&Pr(y = J/x) = 1 - F(r_{m-1} - x\beta)
\end{aligned}
$$

Las ecuaciones presentadas pueden ser usadas para calcular la probabilidad acumulada, lo cual tienen la siguiente forma:

$$Pr(y \leq m/x) = F(r_m - x\beta), cuando : m = 1...J - 1$$

En esta ecuación se muestra que el MRO es equivalente para J-1 regresiones binarias con el supuesto de que las pendientes o coeficientes son idénticos a lo largo de cada regresión.

Por ejemplo, si tenemos cuatro categorías en nuestra endógena y una variable independiente las ecuaciones serian:

$$
\begin{aligned}
Pr(y \leq 1/x) &= F(r_1 - \beta x_1) \\
Pr(y \leq 2/x) &= F(r_2 - \beta x_1) \\
Pr(y < 3/x) &= F(r_3 - \beta x_1)
\end{aligned}
$$

El intercepto no se encuentra en las ecuaciones dado que se ha asumido que $\beta_0 = 0$, cada curva de probabilidad diferirá únicamente en su inclinación hacia la derecha o izquierda, es decir, son paralelas como consecuencia de que el parámetro β es el mismo en cada ecuación. De esta manera el supuesto de paralelismo implica que, $\beta_1 = \beta_2 = ... = \beta_{J-1}$. El grado de paralelismo se asume con parámetros muy cercanos entre sí.

El comando "omodel", de Wolfe y Gould (1998) calcula una aproximación del test LR, en el que se compara el logaritmo de la verosimilitud del ologit (o oprobit) para la obtención de un set de J-1 modelos binarios estimados con ologit (o oprobit), haciendo un ajuste para la correlación entre las salidas binarias definidas por $y \leq m$.

La hipótesis nula será: Existencia del paralelismo en la regresión.

```
findit omodel
omodel logit  warm yr89 male white age ed prst
```

Uno no puede determinar si el coeficiente de algunas variables son idénticos a lo largo de las ecuaciones binarias, mientras que los coeficientes de otras variables difieren. Al final un test de wald elaborado por Brant (1990) es útil pues el test asume el paralelismo de la regresión para cada variable individual.

```
ologit  warm yr89 male white age ed prst,nolog
brant,detail
```

La chi-cuadrado de 49.18 de Brant es muy cercano al valor de 48.91 del test LR, sin embargo, Brant muestra que las mayores violaciones son por yr89 y male, las cuales producen el problema.

2.1.4. Análisis de Probabilidades y Cambios Marginales

El MRO es no lineal, entonces, no hay una sola aproximación que pueda describir totalmente la relación entre una variable y las probabilidades, por lo tanto, se debería considerar cada uno de estos métodos antes de decidir que aproximación es más efectiva en nuestra aplicación.

En el MRO, $y^* = x\beta + \epsilon$, el cambio marginal en y^* con respecto a x_k es: Siendo y^* una variable latente (cuya medida es desconocida), el cambio marginal no puede ser interpretado sin la estandarización, mediante la desviación estandar de y^*.

$$\hat{\sigma}^2_{y^*} = \hat{\beta}'\hat{V}ar(x)\hat{\beta} + Var(\epsilon)$$

Donde $\hat{V}ar(x)$ es la matriz de covarianza para las explicativas, $Var(\epsilon)$ es 1 para los probit ordenados, o $\pi^2/3$ para los logit ordenados. Entonces el estandarización y^* del coeficiente de x_k es:

$$\beta_k^{S_y^*} = \frac{\beta_k}{\sigma_{y^*}}$$

Por cada unidad en que se incremente x_k, se espera que y^* se incremente en $\beta_k^{S_y^*}$ desviaciones estándar, manteniendo las demás variables constantes.

El coeficiente con una total estandarización seria:

$$\beta_k^S = \frac{\sigma_k \beta_k}{\sigma_{y^*}} = \sigma_k \beta_k^{S_{y^*}}$$

Por cada desviación estándar en que se incremente x_k, se espera que y^* se incremente en β_k^S desviaciones estándar, manteniendo las demás variables constantes.

```
ologit  warm yr89 male white age ed prst,nolog
listcoef,std help
```

```
ologit (N=2293): Unstandardized and Standardized Estimates

 Observed SD: .9282156
   Latent SD: 1.9410634
```

warm	b	z	P>|z|	bStdX	bStdY	bStdXY	SDofX
yr89	0.52390	6.557	0.000	0.2566	0.2699	0.1322	0.4897
male	-0.73330	-9.343	0.000	-0.3658	-0.3778	-0.1885	0.4989
white	-0.39116	-3.304	0.001	-0.1287	-0.2015	-0.0663	0.3290
age	-0.02167	-8.778	0.000	-0.3635	-0.0112	-0.1873	16.7790
ed	0.06717	4.205	0.000	0.2123	0.0346	0.1094	3.1608
prst	0.00607	1.844	0.065	0.0880	0.0031	0.0453	14.4923

```
       b = raw coefficient
       z = z-score for test of b=0
   P>|z| = p-value for z-test
   bStdX = x-standardized coefficient
   bStdY = y-standardized coefficient
  bStdXY = fully standardized coefficient
   SDofX = standard deviation of X
```

Figura 2.1: Efecto Marginal en ologit

Podemos observar que en 1989 el apoyo hacia las madres que trabajan fue de 0.27 desviaciones estándar mayores que en 1977, manteniendo las demás variables constantes.

Por cada desviación estándar en que se incremente la educación, se incrementa el apoyo para las madres que trabajan en 0.11 desviaciones estándar, manteniendo las demás variables constantes.

Predicción de Probabilidades

Predecimos las probabilidades como:

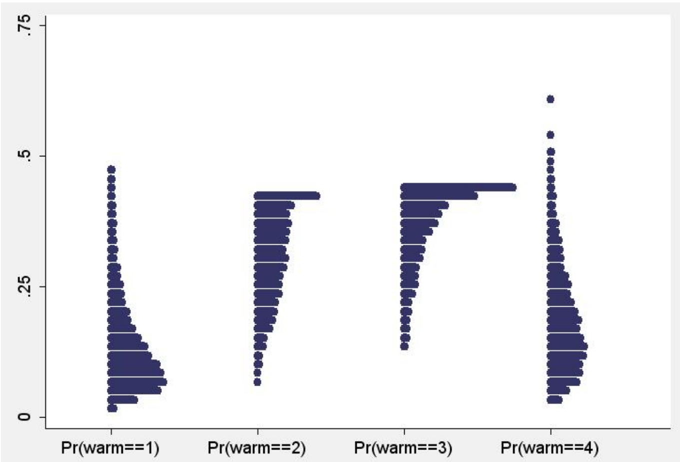
$$\hat{P}r(y = m/x) = F(\hat{r}_m - x\hat{\beta}) - F(\hat{r}_{m-1} - x\hat{\beta})$$

Con probabilidades acumuladas:

$$\hat{P}r(y \leq m/x) = F(r_m - x\hat{\beta})$$

Luego de estimar el modelo es útil calcular las probabilidades, indicando una variable nueva por cada categoría estimada

```
predict sdlogit dlogit alogit salogit
dotplot  sdlogit dlogit alogit salogit, ylabel(0(.25).75)
```



Las probabilidades predichas para las categorías extremas tienden a ser menos que 0.25, la mayor cantidad de las predicciones para las categorías intermedias caen entre 0.25 y 0.5, solo unas cuantas tienden a ser mayores que 0.5

Predicción de Probabilidades con prvalue

La predicción de probabilidades para individuos con un conjunto de características pueden ser calculadas mediante "prvalue", por ejemplo, nosotros podríamos desear, examinar las probabilidades predichas para individuos con las siguientes características:

- Hombres de la clase trabajadora en 1977 quienes están cerca de retirarse.
- Mujeres jóvenes con elevada educación y prestigiosos trabajos.
- Individuo promedio en 1977
- Individuo promedio en 1989

```
ologit  warm yr89 male white age ed prst,nolog
prvalue, x( yr89=0  male=1  prst=20 age=64 ed=16) rest(mean)
prvalue, x( yr89=1  male=0  prst=80  age=30 ed=24) rest(mean)
prvalue, x( yr89=0) rest(mean)
prvalue, x( yr89=1) rest(mean)
```

	Probabilidad Predicha			
Tipo de individuo	**SD**	**D**	**A**	**SA**
Hombres de la clase trabajadora en1997 quienes están cerca del retiro	0.23	0.42	0.27	0.07
Mujeres jóvenes con alta educación en 1989 con trabajos prestigiosos	0.02	0.08	0.32	0.59
Individuo promedio en 1977	0.13	0.36	0.37	0.14
Individuo promedio en 1989	0.08	0.28	0.43	0.21

Predicción de Probabilidades con prtab

En algunos casos nos puede ser de utilidad el calcular las probabilidades predichas para todas las combinaciones de un conjunto de variables independientes categóricas, por ejemplo, si estamos interesados en ver la importancia del género y de los años cuando las preguntas fueron realizadas:

```
prtab  yr89 male
```

Las salidas las podemos reorganizar en la siguiente tabla, donde se observa claramente como los hombres probablemente tienden a estar mas en desacuerdo, comparados con las mujeres, al hecho de que las madres trabajadoras tiendan a tener una calidad relación con sus hijos como una madre que no trabaja. También se observa que entre 1977 y 1989 hubo un cambio en la opinión, tanto para hombres como para mujeres, hacia una actitud mas positiva respecto a la pregunta:

1977	SD	D	A	SA
Hombres	0.19	0.4	0.32	0.1
Mujeres	0.1	0.31	0.41	0.18
Diferencia	0.09	0.09	-0.09	-0.08

1989	SD	D	A	SA
Hombres	0.12	0.34	0.39	0.15
Mujeres	0.06	0.23	0.44	0.27
Diferencia	0.06	0.11	-0.05	-0.12

Cambio de 1977 a 1989				
	SD	**D**	**A**	**SA**
Hombres	-0.07	-0.06	0.07	0.05
Mujeres	-0.04	-0.08	0.03	0.09

Predicción de Probabilidades con prgen

Grafiquemos las probabilidades lo cual nos será de mucha utilidad en los MRO, por ejemplo si consideramos una mujer en 1989 y mostramos como las predicciones de sus probabilidades son afectadas por la edad:

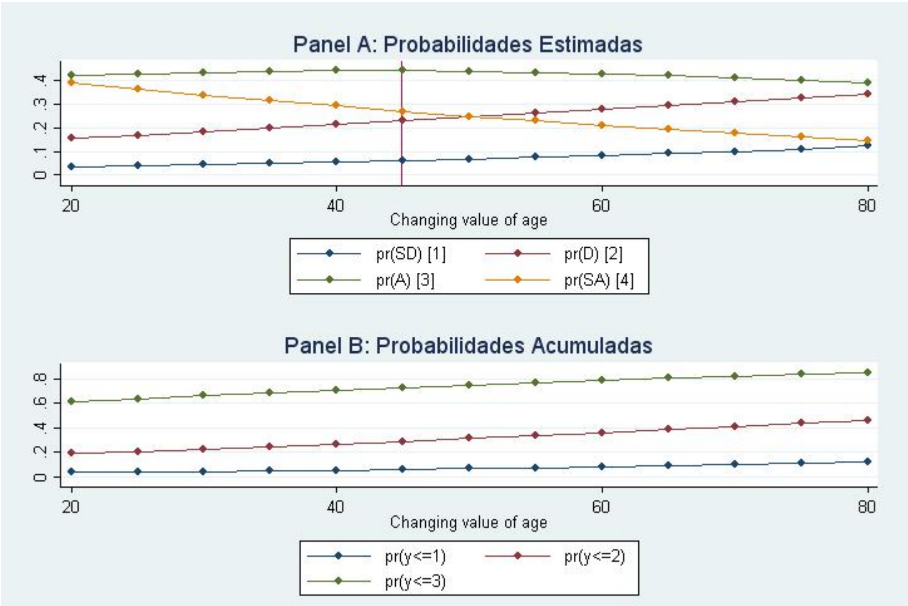
```
prgen age, from(20) to (80) gen(w89) x(male=0 yr89=1) ncases(13)
desc w89*
graph tw sc  w89p1 w89p2 w89p3 w89p4 w89x, connect(l l l l)
```

En este ejemplo "w98x" tendrá los valores de "age", para el rango de 20-80, la p# variable contiene la predicción de la probabilidad para la opción # de la endógena. Cuando el modelo es ordinal, "prgen" también calcula las probabilidades acumuladas, las que son indicadas por w89s#, la cual es la suma de probabilidades para las características 1 y 2.

En el grafico, la edad de 44.93 marca el promedio en el Panel A. Observamos que cuando la edad se incrementa, la probabilidad de "SA",decrece rápidamente mientras que la probabilidad de "D" y "SD" se incrementan, la grafica del Panel B muestra la probabilidad acumulada.

```
graph tw sc  w89p1 w89p2 w89p3 w89p4 w89x, connect(l l l l) ///
xline(44.93) title(Panel A: Probabilidades Estimadas)
graph save temp1,replace
```

```
graph tw sc  w89s1 w89s2 w89s3 w89x, connect(l l l l) ///
title(Panel B: Probabilidades Acumuladas)
graph save temp2,replace
graph combine temp1.gph temp2.gph ,col(1)
```



Cambios en las Probabilidades Predichas con prchange

Cuando existen múltiples variables en el modelo no es practico dibujarlas, para ello es útil usar prchange como resumen de los efectos de cada variable sobre la endógena.

El cambio marginal en la probabilidad es calculado como:

$$\frac{\partial Pr(y=m/x)}{\partial x_k} = \frac{F(r_m - x\beta)}{\partial x_k} - \frac{\partial F(r_{m-1} - x\beta)}{\partial x_k}$$

La cual es la pendiente de la curva que relaciona x_k a $Pr(y = m/x)$, manteniendo las otras variables constantes. En nuestro ejemplo, nosotros consideraremos el efecto marginal de la edad $\frac{\partial Pr(y=m/x)}{\partial age}$, para mujeres en 1989, manteniendo en su media a las demás variables. Esto corresponde a la pendiente de las curvas en el Panel A del grafico anterior evaluado sobre la línea vertical. Con prchange, el calculo seria el siguiente

```
ologit: Changes in Probabilities for warm

age
            Avg|Chg|          SD           D           A          SA
Min->Max   .16441458   .10941909   .21941006  -.05462247  -.27420671
   -+1/2   .00222661   .00124099   .00321223   -.0001803  -.00427291
  -+sd/2    .0373125    .0208976   .05372739  -.00300205  -.07162295
MargEfct   .00222662   .00124098   .00321226  -.00018032  -.00427292

                SD          D          A         SA
Pr(y|x)  .06099996  .22815652  .44057754  .27026597

          yr89      male     white       age        ed      prst
    x=       1         0   .876581   44.9355   12.2181   39.5853
sd(x)=  .489718   .498875   .328989    16.779   3.16083   14.4923

.
```

Figura 2.2: Cambio Marginal en ologit con prchange

Lo primero que debemos notar es la fila denotada por Pr(y/x), la cual es la probabilidad predicha para los valores fijados en x() y en rest(). En la fila de efectos marginales se listan las pendientes de las curvas de probabilidades en el punto de intersección con la línea vertical de la figura anterior. Por ejemplo, la pendiente de "SD" es de 0.00124, mientras que la pendiente de "A" es negativa y muy pequeña, pero no corresponde exactamente a la cantidad de cambio en probabilidad para el cambio en una unidad en la variable independiente. Sin embargo cuando la curva de probabilidad es aproximadamente lineal, el efecto marginal puede ser usado para resumir el efecto de una unidad de cambio en la variable exógena sobre la probabilidad de ocurrencia de un evento.

El cambio marginal también puede ser analizado con mfx, este comando no calcula los efectos del conjunto de variables independientes y solo estima el efecto marginal para una categoría por vez, la cual es especificada en la opción predict(outcome(#)). Veamos ésto con una estimación del ologit y considerando las mismas variables.

```
mfx compute, at(male=0 yr89=1) predict(outcome(1))
```

El impacto marginal de la edad es de 0.001241 como lo muestra la figura2.3 lo cual es comparable con el resultado obtenido con “prchange”. La ventaja de usar “mfx” es que podemos obtener las desviaciones estándar inherentes a cada cambio marginal.

```
warning: no value assigned in at() for variables white age ed prst;
   means used for white age ed prst

Marginal effects after ologit
      y  = Pr(warm==1) (predict, outcome(1))
         =  .06099996
```

variable	dy/dx	Std. Err.	z	P>|z|	[95% C.I.	]	X
yr89*	-.0378526	.00601	-6.30	0.000	-.049633	-.026072	1
male*	.0581355	.00731	7.95	0.000	.043803	.072468	0
white*	.0197511	.0055	3.59	0.000	.008972	.03053	.876581
age	.001241	.00016	7.69	0.000	.000925	.001557	44.9355
ed	-.0038476	.00097	-3.96	0.000	-.005754	-.001941	12.2181
prst	-.0003478	.00019	-1.83	0.068	-.000721	.000025	39.5853

(*) dy/dx is for discrete change of dummy variable from 0 to 1

Figura 2.3: Cambio Marginal en ologit con mfx

RETO 2

a. Se realizo una encuesta a 400 padres de familia, preguntandoles el grado de aceptación sobre la graduación de sus hijos, sus respuestas fueron categorizadas en tres niveles (desacuerdo, moderado acuerdo, muy de acuerdo), además se tomo información sobre el record academico de sus hijos, el tipo de universidad al que asistio su hijo, y el nivel de educación de los padres (si alguno logro algún grado universitario).

Con la información contenida en el archivo ologit, se le pide:

- Verificar las condiciones sobre los cuales el ologit será estimado.
- Estimar el modelo e interpretar las salidas de la regresión y ver a través de una prueba de Wald y LR si el parámetro asociado al tipo de escuela es significativo.
- Interpretar los odds ratios de las variables.
- Realice los test de paralelismo de la regresión. Interprete el resultado.
- ¿Cómo influye sobre la probabilidad de estar de acuerdo o no con la graduación, incremento en la educación de algun miembro de la familia?
- ¿Cómo varian las opiniones sobre la graduación de los estudiantes, a medida que se incrementa el record academico? ¿Y si evaluamos a los tipos de escuelas publicas?
- ¿Cómo varian las probabilidades sobre la graduación para estudiantes provenientes de escuelas publicas, con record promedio y padres con grado?
- ¿Cómo son las probabilidades sobre la graduación para estudiantes provenientes de escuelas privadas, con record promedio y padres con grado?
- ¿Cómo son las probabilidades sobre la graduación para estudiantes provenientes de escuelas publicas, con record promedio y padres sin grado?
- ¿Cómo son las probabilidades sobre la graduación para estudiantes provenientes de escuelas privadas, con record promedio y padres sin grado? ¿A que conclusiones puede llegar?
- ¿Analice los cambios marginales de las variables? ¿A que conclusiones se puede llegar?

Sesión 3
Modelos de Elección Nominal

Una elección o salida es nominal cuando se asumen categorías desordenadas. Por ejemplo, el estado marital puede ser agrupado nominalmente como divorciado, no casado, casado o viudo. Las ocupaciones puedes ser organizadas como profesional, empleador (trabajador de oficina o contratante), obrero (o trabajador tecnico), artesano y sirviente. En algunos casos se suele tratar las salidas nominales como ordenadas o parcialmente ordenadas, por ejemplo, si tu respuesta fuese Totalmente de acuerdo, de acuerdo, en desacuerdo, totalmente en desacuerdo y no sabe no opina, la categoría "no sabe no opina", invalida el modelo ordinal. Se podría decidir usar un modelo de regresión nominal cuando el supuesto del paralelismo de la regresión es rechazada. En general, si uno es consciente del ordenamiento de la variable dependiente, la perdida potencial de eficiencia en el uso de modelos de salida nominal es mayor que la ganancia por evitar el sesgo.

Ahora estimaremos un modelo logit multinomial que es uno de los más usados en los modelos de regresión nominal. El mayor reto de usar el multinomial es que este incluye un montón de parámetros, y fácilmente podría estar sobreestimado. La dificultad nace por el cálculo no lineal del modelo lo cual conduce a problemas de interpretación.

3.1. Estimación y Análisis

3.1.1. Modelo Logit Multinomial

El modelo puede ser imaginado como una estimación simultanea y binaria de logits, para todas las comparaciones posibles de categorías dependientes. Por ejemplo, dejemos que "ocupación" sea una salida nominal con la categoría "S" para trabajos manuales, "E" para trabajos de oficina o empleadores, y "P" para trabajos profesionales. Asumimos que hay una simple variable independiente que mide los años de educación "ed". Entonces podemos examinar los efectos de "ocupación" mediante la estimación de tres logits binarios:

$$
\begin{aligned}
ln\left[\frac{Pr(P/x)}{Pr(S/x)}\right] &= \beta_{0,P/S} + \beta_{1,P/S}ed \\
ln\left[\frac{Pr(E/x)}{Pr(S/x)}\right] &= \beta_{0,E/S} + \beta_{1,E/S}ed \\
ln\left[\frac{Pr(P/x)}{Pr(E/x)}\right] &= \beta_{0,P/E} + \beta_{1,P/E}ed
\end{aligned}
$$

Donde el subíndice de β indica que comparación esta siendo hecha.

Hay que señalar que los tres logits binarios incluyen información redundante, dado que $ln(a/b) = ln(a) - ln(b)$, obteniéndose la siguiente igualdad.

$$ln\left[\frac{Pr(P/x)}{Pr(S/x)}\right] - ln\left[\frac{Pr(E/x)}{Pr(S/x)}\right] = ln\left[\frac{Pr(P/x)}{Pr(E/x)}\right]$$

Esto implica que:

$$
\begin{aligned}
\beta_{0,P/S} - \beta_{0,E/S} = \beta_{0,P/E} \\
\beta_{1,P/S} - \beta_{1,E/S} = \beta_{1,P/E}
\end{aligned}
$$

En general, con J salidas, solamente J-1 logits binarios necesitarían ser estimados. El problema con la estimación de un modelo multinomial, es que cada logit binario es basado en diferentes muestras, por ejemplo, si comparamos P con S, borrariamos E.

Formalmente el modelo puede ser escrito como:

$$ln\Omega_{m/b}(x) = ln\frac{Pr(y = m/x)}{Pr(y = b/x)} = x\beta_{m/b}, \forall m = [1, J]$$

Donde "b" es la categoría base, la cual hace referencia al grupo de comparación. Las J ecuaciones pueden ser resueltas calculando las probabilidades predichas:

$$Pr(y = m/x) = \frac{exp(x\beta_{m/b})}{\sum_{j=1}^{J} exp(x\beta_{j/b})}$$

Mientras las probabilidades predichas serán obtenidas con la categoría b, cambiar la base de la categoría podría confundir a algunos, dado que los resultados de los parámetros tienden a ser algo diferentes. Solo habría un cambio en la parametrización mas no en la estimación de las probabilidades predichas, dado que estas serán las mismas, sea cual sea la categoría base. Las probabilidades para tres categorias podrian ser:

$$Pr(y=m/x)=\frac{exp(x\beta_{m/1})}{\sum_{j=1}^{J}exp(x\beta_{j/1})}$$

Obteniendo los estimadores $\hat{\beta}_{2/1}$ y $\hat{\beta}_{3/1}$, siendo $\hat{\beta}_{1/1}=0$. Si cambiaramos la categoria base, las probabilidades podrian ser:

$$Pr(y=m/x)=\frac{exp(x\beta_{m/2})}{\sum_{j=1}^{J}exp(x\beta_{j/2})}$$

Y obtendriamos los estimadores $\hat{\beta}_{1/2}$ y $\hat{\beta}_{3/2}$, siendo $\hat{\beta}_{2/2}=0$. Así los parámetros estimados serian diferentes, una diferente parametrización es obtenida pero no diferentes probabilidades.

Aplicación

En 1982 General Social Survey, pregunto a 337 personas sobre su nivel de ocupación, categorizando de cinco maneras las respuestas: Trabajos Serviciales "S", trabajos obreril "O", trabajos artesanales "A", trabajos de oficina o contratador "E" y trabajos profesionales "P". Tres variables independientes son consideradas, "raza" que indica raza del encuestado, "ed" que indica años de educación del encuestado y "exper" que mide los años de experiencia laboral.

```
use ocupacion, clear
describe
sum
tab ocupacion, missing
```

```
  obs:           337                          1982 General Social Survey
 vars:             4                          18 Nov 2009 00:05
 size:         2,696 (99.9% of memory free)   (_dta has notes)
-------------------------------------------------------------------------------
              storage  display     value
variable name   type   format      label      variable label
-------------------------------------------------------------------------------
ocupacion       byte   %11.0g      ocupa      ocupación
raza            byte   %10.0g      raz        raza, 1=blanco
ed              byte   %10.0g                 años de educación
exper           byte   %10.0g                 años de experiencia
-------------------------------------------------------------------------------
Sorted by:  ocupacion
```

```
    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
   ocupacion |       337    3.397626    1.367913          1          5
        raza |       337    .9169139    .2764227          0          1
          ed |       337    13.09496    2.946427          3         20
       exper |       337    20.50148    13.95936          2         66
```

ocupación	Freq.	Percent	Cum.
sirviente	31	9.20	9.20
obrero	69	20.47	29.67
artesano	84	24.93	54.60
empleador	41	12.17	66.77
profesional	112	33.23	100.00
Total	337	100.00	

Usando estas variables el siguiente modelo fue estimado:

$$LnΩ_{S/P}(x_i) = \beta_{0,S/P} + \beta_{1,S/P}raza + \beta_{2,S/P}ed + \beta_{3,S/P}exper$$
$$LnΩ_{O/P}(x_i) = \beta_{0,O/P} + \beta_{1,O/P}raza + \beta_{2,O/P}ed + \beta_{3,O/P}exper$$
$$LnΩ_{A/P}(x_i) = \beta_{0,A/P} + \beta_{1,A/P}raza + \beta_{2,A/P}ed + \beta_{3,A/P}exper$$
$$LnΩ_{E/P}(x_i) = \beta_{0,E/P} + \beta_{1,E/P}raza + \beta_{2,E/P}ed + \beta_{3,E/P}exper$$

Especificando las cinco categorías y fijando la categoría base "P":

```
mlogit  ocupacion raza ed exper, b(5) nolog
```

```
Multinomial logistic regression                   Number of obs   =        337
                                                  LR chi2(12)     =     166.09
                                                  Prob > chi2     =     0.0000
Log likelihood = -426.80048                       Pseudo R2       =     0.1629
```

ocupacion	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
sirviente						
raza	-1.774306	.7550543	-2.35	0.019	-3.254186	-.2944273
ed	-.7788519	.1146293	-6.79	0.000	-1.003521	-.5541826
exper	-.0356509	.018037	-1.98	0.048	-.0710028	-.000299
_cons	11.51833	1.849356	6.23	0.000	7.893659	15.143
obrero						
raza	-.5378027	.7996033	-0.67	0.501	-2.104996	1.029391
ed	-.8782767	.1005446	-8.74	0.000	-1.07534	-.6812128
exper	-.0309296	.0144086	-2.15	0.032	-.05917	-.0026893
_cons	12.25956	1.668144	7.35	0.000	8.990061	15.52907
artesano						
raza	-1.301963	.647416	-2.01	0.044	-2.570875	-.0330509
ed	-.6850365	.0892996	-7.67	0.000	-.8600605	-.5100126
exper	-.0079671	.0127055	-0.63	0.531	-.0328693	.0169351
_cons	10.42698	1.517943	6.87	0.000	7.451864	13.40209
empleador						
raza	-.2029212	.8693072	-0.23	0.815	-1.906732	1.50089
ed	-.4256943	.0922192	-4.62	0.000	-.6064407	-.2449479
exper	-.001055	.0143582	-0.07	0.941	-.0291967	.0270866
_cons	5.279722	1.684006	3.14	0.002	1.979132	8.580313

(ocupacion==profesional is the base outcome)

Por defecto mlogit deja como categoría base a la salida con mayor cantidad de observaciones. Alternativamente, uno puede seleccionar la categoría base con "basecategory()". Uno podría estar interesado en saber como la raza afecta la ubicación de los trabajadores entre artesanos y sirvientes, lo cual no fue estimado en la salida anterior, pero podría ser calculado estimando el mlogit con una categoría diferente, sin embargo es mas fácil usar "listcoef", el cual presenta las estimaciones para todas las combinaciones de categorías.

```
listcoef raza, help
```

```
mlogit (N=337): Factor Change in the Odds of ocupacion

Variable: raza (sd=.27642268)

    Odds comparing|
Group 1 vs Group 2|        b         z     P>|z|      e^b    e^bStdX
------------------+---------------------------------------------------
sirvient-obrero   |  -1.23650    -1.707    0.088    0.2904    0.7105
sirvient-artesano |  -0.47234    -0.782    0.434    0.6235    0.8776
sirvient-empleado |  -1.57139    -1.741    0.082    0.2078    0.6477
sirvient-profesio |  -1.77431    -2.350    0.019    0.1696    0.6123
obrero  -sirvient |   1.23650     1.707    0.088    3.4436    1.4075
obrero  -artesano |   0.76416     1.208    0.227    2.1472    1.2352
obrero  -empleado |  -0.33488    -0.359    0.720    0.7154    0.9116
obrero  -profesio |  -0.53780    -0.673    0.501    0.5840    0.8619
artesano-sirvient |   0.47234     0.782    0.434    1.6037    1.1395
artesano-obrero   |  -0.76416    -1.208    0.227    0.4657    0.8096
artesano-empleado |  -1.09904    -1.343    0.179    0.3332    0.7380
artesano-profesio |  -1.30196    -2.011    0.044    0.2720    0.6978
empleado-sirvient |   1.57139     1.741    0.082    4.8133    1.5440
empleado-obrero   |   0.33488     0.359    0.720    1.3978    1.0970
empleado-artesano |   1.09904     1.343    0.179    3.0013    1.3550
empleado-profesio |  -0.20292    -0.233    0.815    0.8163    0.9455
profesio-sirvient |   1.77431     2.350    0.019    5.8962    1.6331
profesio-obrero   |   0.53780     0.673    0.501    1.7122    1.1603
profesio-artesano |   1.30196     2.011    0.044    3.6765    1.4332
profesio-empleado |   0.20292     0.233    0.815    1.2250    1.0577
----------------------------------------------------------------------
       b = raw coefficient
       z = z-score for test of b=0
   P>|z| = p-value for z-test
     e^b = exp(b) = factor change in odds for unit increase in X
 e^bStdX = exp(b*SD of X) = change in odds for SD increase in X
```

No solamente se puede controlar la variable a ser reportada por listcoef, sino también aquellos coeficientes significativos en cierto nivel:

```
listcoef raza, pvalue(0.05) help
```

```
mlogit (N=337): Factor Change in the Odds of ocupacion when P>|z| < 0.05

Variable: raza (sd=.27642268)

    Odds comparing|
Group 1 vs Group 2|        b         z     P>|z|      e^b    e^bStdX
------------------+---------------------------------------------------
sirvient-profesio |  -1.77431    -2.350    0.019    0.1696    0.6123
artesano-profesio |  -1.30196    -2.011    0.044    0.2720    0.6978
profesio-sirvient |   1.77431     2.350    0.019    5.8962    1.6331
profesio-artesano |   1.30196     2.011    0.044    3.6765    1.4332
----------------------------------------------------------------------
       b = raw coefficient
       z = z-score for test of b=0
   P>|z| = p-value for z-test
     e^b = exp(b) = factor change in odds for unit increase in X
 e^bStdX = exp(b*SD of X) = change in odds for SD increase in X
```

3.1.2. Testeo de Hipótesis

En el MNLM uno puede testear los coeficientes de manera individual y reportar los z-statistic, con un test de Wald o con un test LR. Hay buenas razones para testear los coeficientes de manera grupal. Al testear el hecho de que una variable no tenga efectos requiere que el test para J-1 coeficientes sean simultáneamente iguales a cero. Luego el testeo de que las variables independientes como un grupo sean diferentes entre dos estimaciones, requiere un test de K coeficientes.

Testeo de efectos de variables independientes

Con J categorías dependientes, hay J-1 coeficientes no redundantes asociados con cada variable independiente. Por ejemplo para nuestro logit de ocupación hay cuatro coeficientes asociados con educación "ed", $\beta_{2,M/P}$, $\beta_{2,B/P}$, $\beta_{2,C/P}$, $\beta_{2,W/P}$. La hipótesis de que x_k no tiene efectos sobre la variable dependiente puede ser escrita como:

$$H_0 : \beta_{k,1/b} = ... = \beta_{k,J/b} = 0$$

Donde "b" es la categoría base, y como $\beta_{k,b/b} = 0$, la hipótesis impone restricciones sobre J-1 parámetros, ésta puede ser testeada con Wald o con LR test.

Test del ratio de verosimilitud (LR)

El LR-test involucra:

- Estimar el modelo completo incluyendo todas las variables, resultantes en el estadístico del ratio de verosimilitud LR_{sr}.
- Estimar el modelo restricto excluyendo las variables x_k, y obtener el ratio de verosimilitud LR_r.
- Calcular la diferencia $LR = LR_{sr} - LR_r$, el cual es distribuido como una chi-cuadrado con J-1 grados de libertad.

El cálculo puede ser hecho con el comando "lrtest":

```
mlogit  ocupacion raza ed exper, b(5) nolog
lrtest, saving(0)
mlogit  ocupacion ed exper, b(5) nolog
lrtest
```

```
Likelihood-ratio test                                 LR chi2(4)  =      8.10
(Assumption: . nested in LRTEST_0)                    Prob > chi2 =    0.0881
```

El "mlogtest" puede calcular cada uno de estos pasos de manera integral para cada una de las variables explicativas en el modelo.

```
**** Likelihood-ratio tests for independent variables

 Ho: All coefficients associated with given variable(s) are 0.
```

ocupac~n	chi2	df	P>chi2
raza	8.095	4	0.088
ed	156.937	4	0.000
exper	8.561	4	0.073

```
mlogit  ocupacion raza ed exper, b(5) nolog
mlogtest, lr
```

El efecto de la raza en la ocupación es significativa al 0.1 de significancia, pero no lo es al 0.05. El efecto de la educación es significativa al 0.01. O mas formal, la hipótesis de que todos los coeficientes asociados con la educación son simultáneamente iguales a cero pueden ser rechazados al 0.01.

Test de Wald

Aunque el LR test es generalmente considerado superior, si el modelo es complejo, la muestra es muy grande, es muy costoso usar este test. Alternativamente, K test de Wald puede ser calculado usando "test", sin ninguna estimación adicional, por ejemplo:

```
mlogit  ocupacion raza ed exper, b(5) nolog
test raza
test ed
test exper
```

Una forma de resumir lo anterior es:

```
mlogtest, wald
```

La lógica del test de Wald o LR puede ser extendida para testear el efecto de que dos o mas variables independientes sean simultáneamente cero.

```
mlogit  ocupacion raza ed exper, b(5) nolog
lrtest, saving(0)
mlogit  ocupacion raza, b(5) nolog
lrtest
```

O también:

```
mlogit  ocupacion raza ed exper, b(5) nolog
mlogtest, lr set( ed exper)
```

```
**** Likelihood-ratio tests for independent variables

 Ho: All coefficients associated with given variable(s) are 0.

    ocupac~n |      chi2   df   P>chi2
-------------+-------------------------
        raza |     8.095    4    0.088
          ed |   156.937    4    0.000
       exper |     8.561    4    0.073
-------------+-------------------------
      set_1: |   160.773    8    0.000
          ed |
       exper |
---------------------------------------
```

Si ninguna de las variables independientes afectan significativamente al odds de la categoria m vs la categoria n, nosotros decimos que m y n son indistinguibles con respecto a las variables en el modelo [1].

Que las categorias m y n sean indistinguibles corresponde a probar la siguiente hipótesis:

$$H_0 : \beta_{1,m/n} = ... = \beta_{K,m/n} = 0$$

La cual será testeada con Wald o LR. Ambos test proveen resultados muy similares. Si las dos categorias son indistinguibles con respecto a las variables en el modelo, entonces podríamos obtener estimadores más eficientes, asociandolas. Para testear esto usamos “mlogtest”.

Test de Wald para categorias combinadas

El comando combina el calculo de Wald, para la hipótesis nula de que dos categorias pueden ser combinadas, para todas las combinaciones de categorias que existan. Por ejemplo:

```
mlogit  ocupacion raza ed exper, b(5) nolog
mlogtest, combine
```

Podemos rechazar la hipótesis de que la categoría para sirviente (menial) y profesional (prof) son indistinguibles, pero no podemos rechazar que sirviente (menial) y obrero (bluecol) sean inditinguibles.

Para testear que la categoría de sirviente (menial) sea inditinguible de la categoría base Prof:

```
test [Menial]
```

Lo cual es igual a la fila del mlogtest Menial-Prof

[1] Anderson 1984

```
**** Wald tests for combining outcome categories

 Ho: All coefficients except intercepts associated with given pair
     of outcomes are 0 (i.e., categories can be collapsed).

Categories tested |      chi2   df   P>chi2
------------------+-------------------------
sirvient-  obrero |     3.994    3    0.262
sirvient-artesano |     3.203    3    0.361
sirvient-empleado |    11.951    3    0.008
sirvient-profesio |    48.190    3    0.000
  obrero-artesano |     8.441    3    0.038
  obrero-empleado |    20.055    3    0.000
  obrero-profesio |    76.393    3    0.000
artesano-empleado |     8.892    3    0.031
artesano-profesio |    60.583    3    0.000
empleado-profesio |    22.203    3    0.000
--------------------------------------------

 ( 1)  [sirviente]raza = 0
 ( 2)  [sirviente]ed = 0
 ( 3)  [sirviente]exper = 0

           chi2(  3) =   48.19
         Prob > chi2 =    0.0000
```

El test es más complicado cuando ninguna categoría es la base, por ejemplo cuando testeamos que m y n son indistinguibles cuando la categoría base no es ni m ni n, la hipótesis nula seria:

$$H_0 : (\beta_{1,m/b} - \beta_{1,n/b}) = ... = (\beta_{K,m/b} - \beta_{K,n/b}) = 0$$

De ésta manera deseariamos testear la diferencia entre los dos conjuntos de coeficientes. Por ejemplo el testear si la categoría de sirviente y artesano pueden ser combinados, requeriría:

```
test [Sirviente=Artesano]
```

Una vez más, los resultados son idénticos a los reportados en mlogtest.

Test LR para categorias combinadas

El test LR que combina m y n se calcula estimando el modelo completo sin restricciones y obteniendo el estadístico LR^2_{sr}, y luego uno restricto en la cual la categoría m es usada como la base y todos los otros coeficientes excepto la constante en la ecuación de categoría n son ceros, obteniéndose el estadístico LR^2_r. El estadístico final es la diferencia $LR^2_{sr,r} = LR^2 sr - LR^2_r$, el cual se distribuye como una chi-cuadrada con K grados de libertad.

```
mlogit  ocupacion raza ed exper, b(5) nolog
mlogtest, lrcomb
```

Este comando puede usar restricciones, para ver esto, nosotros usamos el test que compara la categoria sirviente con obrero. Primero calculamos el modelo completo y guardamos los resultados:

```
**** LR tests for combining outcome categories

 Ho: All coefficients except intercepts associated with given pair
     of outcomes are 0 (i.e., categories can be collapsed).
```

Categories tested	chi2	df	P>chi2
sirvient- obrero	4.095	3	0.251
sirvient-artesano	3.376	3	0.337
sirvient-empleado	13.223	3	0.004
sirvient-profesio	64.607	3	0.000
obrero-artesano	9.176	3	0.027
obrero-empleado	22.803	3	0.000
obrero-profesio	125.699	3	0.000
artesano-empleado	9.992	3	0.019
artesano-profesio	95.889	3	0.000
empleado-profesio	26.736	3	0.000

```
mlogit  ocupacion raza ed exper, nolog
lrtest, saving(lrf)
```

Luego construimos la restricción

```
constraint define 999 [Sirviente]
```

Así la opción [Menial] indica que todos los coeficientes excepto la contante de las ecuaciones de la categoria sirvientes (Menial), serán cero. Finalmente re-estimamos el modelo con la restricción. La categoría base deberá ser obreros (Bluecol), así que los coeficientes indicados por "[Menial]" son comparados entre Bruecol y Menial.

```
mlogit  ocupacion raza ed exper, base(2) constraint(999) nolog
```

Donde observamos que restricción es impuesta y así calculamos el test de verosimilitud.

```
lrtest, using(lrf)
```

```
Likelihood-ratio test                                 LR chi2(3)  =      4.09
(Assumption: . nested in LRTEST_lrf)                  Prob > chi2 =    0.2514
```

3.1.3. Independencia de las Alternativas irrelevantes (IIA)

Tanto el MNLM y el condicional tienen como supuesto la independencia de alternativas irrelevantes, mostramos este supuesto en términos del modelo logit multinomial.

$$\frac{Pr(y=m/x)}{Pr(y=n/x)} = exp[x(\beta_{m/b} - \beta_{n/b})]$$

Donde el Odds no depende de otras categorías que sean viables. En este sentido, estas categorías alternativas son irrelevantes, lo que significa que al añadir o borrar una categoría esta no afectara la cantidad de Odds en las categorías principales. Este punto es explicado a menudo con un ejemplo de transporte en autobuses rojos/azules: Supongamos que se tiene que elegir entre un autobús rojo y un carro para ir a trabajar y que el Odds de tomar el autobús comparado con el carro es de 1:1. La IIA implica que el Odds deberá mantenerse 1:1 entre estas dos alternativas, aún si una nueva compañía de autobuses azules llega al pueblo, autobuses de características idénticas a la compañía de autobuses roja. Así, las probabilidades de manejar un carro pueden ser tan pequeñas aún al añadir diferentes colores de autobuses. Más razonable, seria esperar que el Odds de comparar un autobús rojo y un carro debería reducirse a 1:2 ya que la mitad de personas que subían al autobús rojo, ahora se esperaría que suban al azul.

Hay dos test que tratan el supuesto IIA. Hausman y McFadden (1984) propusieron un test tipo Hausman. Y McFadden, Tye y Train (1976) propusieron una aproximación al test de ratio de verosimilitud, que fue implantado por Small y Hsiao (1985). Ambos, asumían que el MNLM es estimado con la categoría base "b", y existían por tanto J-1 test a ser calculados excluyendo cada uno las principales categorías para formar un modelo restricto. Para cambiar la categoría base, el test puede ser calculado excluyendo b. El resultado del test difieren dependiendo de cual es la categoría base que fue usada para estimar el modelo.

<u>Test de Hausman</u>

El test de Hausman implica los siguientes pasos:

- Estimar el modelo completo con todas las J categorías incluidas y obtener el estimador $\hat{\beta}_{sr}$.
- Estimar un modelo restricto eliminando una a una las diferentes categorías y obtener el estimador $\hat{\beta}_r$.
- Dejar que $\hat{\beta}^*_{sr}$ sea una sub muestra de $\hat{\beta}_{sr}$ luego de eliminar los coeficientes no estimados en el modelo restricto. El test será:

$$H = (\hat{\beta}_r - \hat{\beta}^*_{sr})'[\hat{Var}(\hat{\beta}_r) - \hat{Var}(\hat{\beta}^*_{sr})]^{-1}(\hat{\beta}_r - \hat{\beta}^*_{sr})$$

  Donde H es distribuido asintóticamente como una chi-cuadrado con grados de libertad iguales a las filas de $\hat{\beta}_r$ si IIA es verdadero. Los valores signicativos de H indican que el supuesto de IIA ha sido violado.

```
mlogit  ocupacion raza ed exper, b(5) nolog
mlogtest, hausman base
```

```
**** Hausman tests of IIA assumption

 Ho: Odds(Outcome-J vs Outcome-K) are independent of other alternatives.

 Omitted |      chi2   df   P>chi2   evidence
---------+------------------------------------
sirvient |     7.324   12    0.835   for Ho
  obrero |     0.320   12    1.000   for Ho
artesano |   -14.436   12    1.000   for Ho
empleado |    -5.541   11    1.000   for Ho
profesio |    -0.119   12    1.000   for Ho
----------------------------------------------
```

Cinco test fueron reportados, los primeros cuatro corresponden a la exclusión de una de las cuatro categorías no base. La quinta es calculada reestimando el modelo usando la mas categoría mas grande como categoría base. Ninguna rechaza la H_0, de que la IIA sea verdadera. Los resultados difieren considerablemente al cambiar la categoría base. Tres de los test estadísticos son negativos, lo cual es común encontrar, un resultado asi, presenta evidencias de que la IIA no ha sido violada. Un mayor sentimiento de la variabilidad de los resultados puede ser visto corriendo el mlogit con una categoría diferente y volviendo a calcular el test.

Test Small y Hsiao

Para calcular el test, la muestra es dividida aleatoriamente en dos submuestras de igual medida. El MNLM irrestricto es estimado sobre ambas submuestras, donde $\hat{\beta}_{sr}^{S_1}$ contiene las estimaciones del modelo irrestricto para la primera submuestra y $\hat{\beta}_{sr}^{S_2}$ es la contraparte para la segunda submuestra. Una media ponderada de los coeficientes se calcula como:

$$\hat{\beta}_{sr}^{S_1,S_2} = \frac{1}{\sqrt{2}}\hat{\beta}_{sr}^{S_1} + [1 - \frac{1}{\sqrt{2}}]\hat{\beta}_{sr}^{S_2}$$

Luego una muestra restricta es creada de la segunda submuestra, al eliminar todos los valores de una categoria elegida en la variable dependiente. El MNLM es estimado usando la muestra restricta y obteniendo los estimadores $\hat{\beta}_{r}^{S_2}$ y el verosimilitud L($\hat{\beta}_{r}^{S_2}$). El estadístico de Small y Hsiao será:

$$SH = -2[L(\hat{\beta}_{sr}^{S_1,S_2}) - L(\hat{\beta}_{r}^{S_2})]$$

El cual es distribuido asintóticamente como una chi-cuadrada con grados de libertad igual a K+1 donde K es el número de variable independientes.

```
mlogtest, smhsiao
```

```
**** Small-Hsiao tests of IIA assumption

 Ho: Odds(Outcome-J vs Outcome-K) are independent of other alternatives.
```

Omitted	lnL(full)	lnL(omit)	chi2	df	P>chi2	evidence
sirvient	-173.287	-166.950	12.675	4	0.013	against Ho
obrero	-154.895	-150.543	8.705	4	0.069	for Ho
artesano	-133.658	-130.611	6.095	4	0.192	for Ho
empleado	-152.900	-148.357	9.086	4	0.059	for Ho

Los resultados varían respecto al test de Hausman, ahora vemos que un caso violo la IIA. Dado que el test de Small Hsiao, requiere una división aleatoria de la data en submuestras, el resultado puede diferir con sucesivas llamadas de comandos, dado la diferente división aleatoria por vez. Para obtener un test que replique los resultados deberíamos fijar un número aleatorio para las muestras.

```
set seed 8675309
mlogtest, smhsiao
```

Estos test a menudo dan resultados inconsistentes y proveen de violaciones al supuesto de IIA. Desafortunadamente no hay estudios que examinen las propiedades para una pequeña muestra. Quizás como un resultado de las limitaciones practicas de estos test, McFadden (1973) suguiere que las IIA implican que el logit multinomial y condicional deberían solo usarse en casos donde las categorías "puedan asumirse distintas y ponderadas independientemente a los ojos del que toma las decisiones". De manera similar Amemiya (1981) suguiere que el MNLM trabaja bien cuando las alternativas son disimiles. Hay que cuidar, que al especificar el modelo se considere distinguir las categorías que no sean sustitutas una de otras, pareciendo ser razonable aunque desafortunadamente ambiguo.

Para medir el ajuste, podemos usar el fitstat como medida de análisis.

```
fitstat
```

Measures of Fit for **mlogit** of **ocupacion**

Log-Lik Intercept Only:	-509.844	Log-Lik Full Model:	-426.800
D(321):	853.601	LR(12):	166.087
		Prob > LR:	0.000
McFadden's R2:	0.163	McFadden's Adj R2:	0.131
Maximum Likelihood R2:	0.389	Cragg & Uhler's R2:	0.409
Count R2:	0.501	Adj Count R2:	0.253
AIC:	2.628	AIC*n:	885.601
BIC:	-1014.646	BIC':	-96.246

3.1.4. Análisis de Probabilidades y Cambios Marginales

Mientras el MNLM es una simple extensión matemática del modelo binario, la interpretación se dificulta por la gran cantidad de posibles comparaciones que se pueden hacer. Aún en, nuestro ejemplo con cinco categorías, nosotros tendríamos muchas comparaciones por hacer. Pero existen comandos que nos proveen de herramientas muy potentes para llevar a cabo dicha tarea.

Predicción de las probabilidades con "predict"

Las probabilidades son obtenidas de la siguiente manera:

```
mlogit  ocupacion raza ed exper, b(5) nolog
predict probM probC probB probW probP
describe prob*
summarize prob*
```

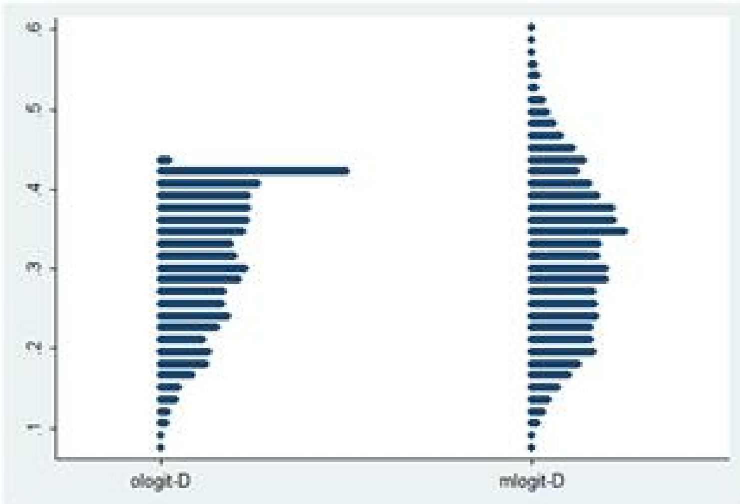
Variable	Obs	Mean	Std. Dev.	Min	Max
probS	337	.0919881	.059396	.0010737	.3281906
probO	337	.2047478	.1450568	.0012066	.6974148
probA	337	.2492582	.1161309	.0079713	.551609
probE	337	.1216617	.0452844	.0083857	.2300058
probP	337	.3323442	.2870992	.0001935	.9597512

Podemos realizar una comparación también entre un mlogit y un ologit, una manera de ver esta comparación es gráficamente. En la estimación del ologit, observamos que la predicción de probabilidades de las categorías intermedias eran sesgadas hacia abajo mientras que las categorías extrémales tenían una distribución mas gradual. Veamos esto:

```
use ordwarm2,clear
ologit warm  yr89 male white age ed prst, nolog
predict SDologit Dologit Aologit SAologit
label var Dologit "ologit-D"
mlogit  warm yr89 male white age ed prst,nolog
predict SDmlogit Dmlogit Amlogit SAmlogit
label var Dmlogit "mlogit-D"
dotplot Dologit Dmlogit
```

```
corr Dologit Dmlogit
```

La correlación entre ambos grupos de predicciones es de 0.92, pero el truncamiento de la distribución para el modelo logit ordinal es algo irreal.



Predicción de las probabilidades con "prvalue"

Predecir las probabilidades para un individuo con características especificas puede realizarse con "prvalue". Por ejemplo, podríamos desear calcular las probabilidades de cada categoría ocupacional comparando a los negros con los blancos, con educación y experiencia promedia.

```
use ocupacion,clear
mlogit  ocupacion raza ed exper, b(5) nolog
quietly prvalue, x( raza 0) rest(mean) save
prvalue, x(raza 1) rest(mean) dif
```

```
mlogit: Change in Predictions for ocupacion

Confidence intervals by delta method

                     Current     Saved    Change    95% CI for Change
  Pr(y=sirvient|x):   0.0860    0.2168   -0.1309   [-0.3056,   0.0439]
  Pr(y=obrero|x):     0.1862    0.1363    0.0498   [-0.0897,   0.1893]
  Pr(y=artesano|x):   0.2790    0.4387   -0.1597   [-0.3686,   0.0491]
  Pr(y=empleado|x):   0.1674    0.0877    0.0797   [-0.0477,   0.2071]
  Pr(y=profesio|x):   0.2814    0.1204    0.1611   [ 0.0277,   0.2944]

               raza         ed        exper
Current=          1  13.094955    20.501484
  Saved=          0  13.094955    20.501484
   Diff=          1          0            0
```

Predicción de probabilidades con "prtab"

Si se desea predecir las probabilidades para todas las combinaciones de un conjunto de variables categóricas independientes, "prtab" seria útil. Por ejemplo si deseáramos conocer como la respuesta de los blancos y negros difieren en sus probabilidades de tener un trabajo de sirvientes, conforme aumentan los años de educación.

```
label def lwhite 0 Negros 1 Blancos
label val raza lwhite
prtab ed raza, novarlbl outcome(1)
prtab ed raza, novarlbl
```

```
mlogit: Predicted probabilities of outcome 1 (sirviente) for ocupacion
```

años de educación	raza, 1=blanco no blanco	blanco
3	0.2847	0.1216
6	0.2987	0.1384
7	0.2988	0.1417
8	0.2963	0.1431
9	0.2906	0.1417
10	0.2814	0.1366
11	0.2675	0.1265
12	0.2476	0.1104
13	0.2199	0.0883
14	0.1832	0.0632
15	0.1393	0.0401
16	0.0944	0.0228
17	0.0569	0.0120
18	0.0310	0.0060
19	0.0158	0.0029
20	0.0077	0.0014

```
          raza          ed       exper
x=   .91691395   13.094955   20.501484
```

La tabla muestra una sustancial diferencia entre blancos y negros en la probabilidad de tener trabajos de sirvientes, y como estas probabilidades son afectadas según aumentan los años de educación. Sin embargo, dado el número de categorías para "ed", el graficar las probabilidades predichas con "prgen" será la manera más útil de examinar estos resultados.

Predicción de probabilidades con "prgen"

Las probabilidades predichas pueden ser graficadas usando los mismos métodos considerados para los modelos de regresión ordinal. Luego de estimar el modelo, nosotros usaremos "prgen" para calcular las probabilidades predichas para blancos con experiencia laboral promedia e incrementos en los años de educación de seis a veinte años.

```
prgen ed, x(raza=1) from(6) to (20) gen(wht) ncases(15)
describe wht*
prgen ed, x(raza=0) from(6) to (20) gen(nwht) ncases(15)
```

```
mlogit: Predicted values as ed varies from 6 to 20.

         white          ed       exper
x=           1   13.094955   20.501484
```

```
                  storage  display     value
variable name       type   format      label      variable label
-------------------------------------------------------------------------------
blanx               float  %9.0g                  Changing value of ed
blanp1              float  %9.0g                  pr(sirvient)=Pr(1)
blanp2              float  %9.0g                  pr(obrero)=Pr(2)
blanp3              float  %9.0g                  pr(artesano)=Pr(3)
blanp4              float  %9.0g                  pr(empleado)=Pr(4)
blanp5              float  %9.0g                  pr(profesio)=Pr(5)
blans1              float  %9.0g                  pr(y<=1)
blans2              float  %9.0g                  pr(y<=2)
blans3              float  %9.0g                  pr(y<=3)
blans4              float  %9.0g                  pr(y<=4)
blans5              float  %9.0g                  pr(y<=5)

mlogit: Predicted values as ed varies from 6 to 20.

          white          ed       exper
x=            0   13.094955   20.501484
```

Las variables nwhtp1 whtp1 contienen las probabilidades predichas de tener trabajos de sirvientes para negros y blancos. La grafica de éstas, pueden proveer información más clara que los resultados de "prtab".

```
label var whtp1 "blancos"
label var nwhtp1 "Negros"

set textsize 125

tw sc whtp1 nwhtp1 nwhtx, connect(ss ss) ///
xtitle(Años de educación de los blancos) ///
ytitle(Pr(Trabajo de Sirviente))
```

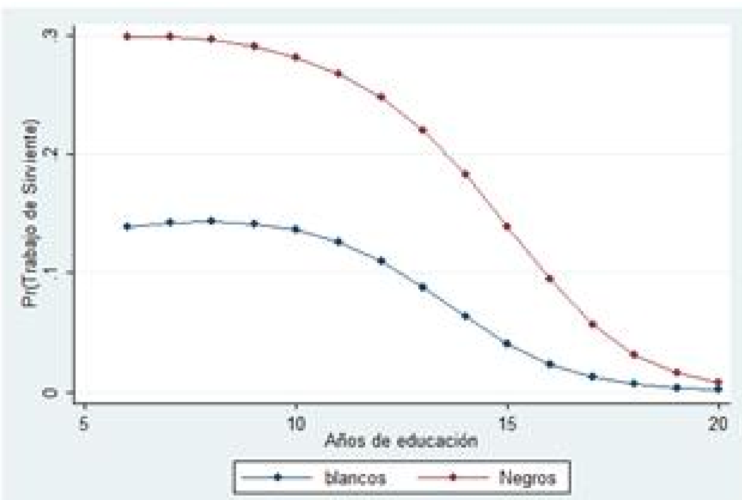


Figura 3.1: "Probabilidades vs Educación"

Aunque las categorías nominales no están ordenadas, el grafico que suma las probabilidades puede ser una útil manera de mostrar las probabilidades predichas para todas las categorías. Para esto construimos una grafica que muestra como la educación afecta la probabilidad de cada ocupación para blancos.

```
label var whts1 "sirvientes"
label var whts2 "artesano"
label var whts3 "obreros"
label var whts4 "empleador"

set textsize 125

tw sc  whts1 whts2 whts3 whts4 whtx, c(ss ss ss ss) ///
xtitle(Años de educación de los blancos) ///
ytitle(Probabilidades Acumuladas)
```

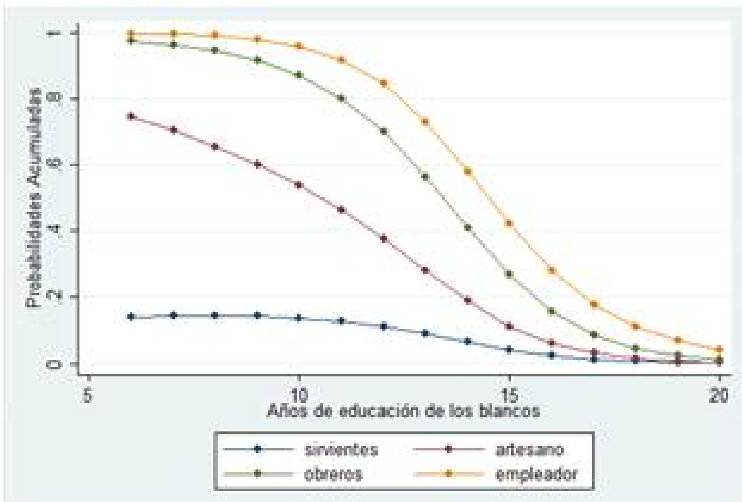


Figura 3.2: "Probabilidades Acumuladas vs Educación"

El grafico muestra las cuatro probabilidades acumuladas conforme aumenta los años de educación La línea mas baja etiquetada con sirvientes grafica las probabilidades de tener un trabajo de sirviente según varia los años de educación. Es la misma información que se presento en el gráfico anterior para las personas de raza blanca. La siguiente línea, etiquetada como artesano grafica la suma de probabilidades de tener trabajo un trabajo de sirviente o artesano. De esta manera el área entre la línea roja y azul, es la probabilidad de tener trabajo de artesano.

Cambio en las probabilidades predichas

Cambios marginales y discretos pueden ser usados de la misma manera que en modelos de salidas ordinales. Como antes, ambas pueden ser calculadas con "prchange".

- **Cambio Marginal** Podemos definir el cambio marginal como:

$$\frac{\partial Pr(y=m/x)}{\partial x_k} = Pr(y=m/x)[\beta_{k,m/J} - \sum_{j=1}^{J} \beta_{k,m/J} Pr(y=j/x)]$$

  Dado que esta ecuación combina todos los $\beta_{k,j/J}$, el valor de los cambios marginales dependen de los valores de todas las variables del modelo. Mas aun cuando el valor de x_k cambia, el signo del impacto marginal puede cambiar. Por ejemplo, en algún punto, el efecto marginal de la educación sobre tener una ocupación de sirvientes podría ser positivo, mientras que en otro punto dicho efecto podría ser negativo.

- **Cambio Discreto** Podemos definir el cambio discreto como:

$$\frac{\Delta Pr(y=m/x)}{\Delta x_k} = Pr(y=m/x, x_k = x_E) - Pr(y=m/x, x_k = x_S)$$

  Donde la magnitud de el cambio depende de los niveles de todas las variables y del tamaño del cambio que es realizado. Los J cambios discretos de los coeficientes para una variable (uno por cada categoria) pueden ser resumidos calculando un promedio del valor absoluto de los cambios a través de todas las categorías.

$$\bar{\Delta} = \frac{1}{J} \sum_{j=1}^{J} \left| \frac{\Delta Pr(y=j/x)}{\Delta x_k} \right|$$

  Donde el valor absoluto es tomado porque la suma de los cambios sin tomar el valor absoluto es necesariamente cero.

`prchange`

Los cambios marginales son listados en las filas del Efecto Marginal. Para variables que no son binarias, los cambios discretos son reportados sobre el rango completo de las variables (reportado como Min → Max), para cambios en una unidad centrada alrededor de los valores base (reportado como -+1/2) y para cambios en una desviación estándar centrada

```
mlogit: Changes in Predicted Probabilities for ocupacion

raza
          Avg|Chg|    sirvient      obrero    artesano    empleado    profesio
    0->1  .11623582  -.13085523   .04981799  -.15973434   .07971004    .1610615

ed
          Avg|Chg|    sirvient      obrero    artesano    empleado    profesio
Min->Max  .39242268  -.13017954  -.70077323  -.15010394   .02425591   .95680079
   -+1/2  .05855425  -.02559762  -.06831616  -.05247185   .01250795   .13387768
  -+sd/2   .1640657  -.07129153  -.19310513  -.14576758   .03064777   .37951647
MargEfct  .05894859  -.02579097  -.06870635  -.05287415   .01282041   .13455107

exper
          Avg|Chg|    sirvient      obrero    artesano    empleado    profesio
Min->Max  .12193559  -.11536534  -.18947365   .03115708   .09478889   .17889298
   -+1/2  .00233425  -.00226997  -.00356567   .00105992    .0016944   .00308132
  -+sd/2  .03253578  -.03167491  -.04966453   .01479983   .02360725   .04293236
MargEfct  .00233427  -.00226997  -.00356571   .00105992   .00169442   .00308134

          sirvient     obrero   artesano   empleado   profesio
Pr(y|x)  .09426806  .18419114  .29411051  .16112968  .26630062

            raza        ed     exper
    x=   .916914    13.095   20.5015
sd(x)=   .276423   2.94643   13.9594
```

alrededor de los valores base (reportado como -+sd/2). Si la opción "uncentered" es usada, los cambios comienza en la opción especificada por x() y rest() y se incrementan en una unidad o una desviación estándar desde allí. Para variables binarias, el cambio discreto de 0 a 1 es la cuantía apropiada y es la única cuantía presentada. Vemos en el resultado para White, que para aquellos que tienen educación y experiencia promedio, la probabilidad predicha de tener un trabajo profesional es de 0.16 veces mayor para blancos que para negros. El cambio promedio es listado en la primera columna. Por ejemplo, para White, el cambio absoluto promedio en la probabilidad de varias categorías laborales de ser blanco como oposición a ser negro es de 0.12.

El cambio marginal también puede ser calculado con "mfx", que al igual que "prchange", calcula el cambio manteniendo todo el conjunto de variables independientes en su media. Hay que notar que no solo nos permite calcular el efecto de un conjunto de variables en el modelo, sino que también estima los efectos marginales para una categoría a la vez:

```
mfx compute, predict(outcome(1))
```

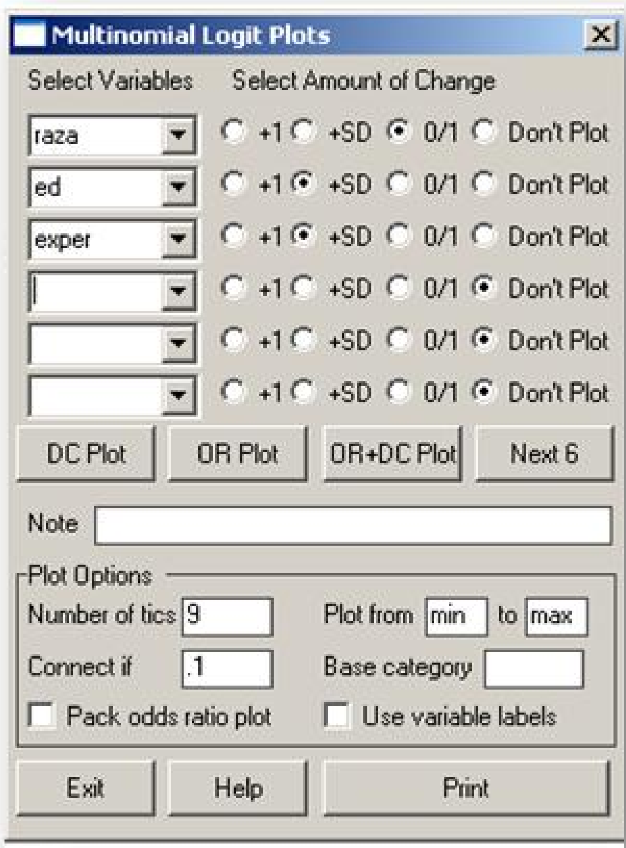
```
Marginal effects after mlogit
      y  = Pr(ocupacion==1) (predict, outcome(1))
         =  .09426806
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
   raza* |  -.1308552       .08914   -1.47   0.142  -.305562  .043852   .916914
      ed |   -.025791       .00688   -3.75   0.000  -.039269 -.012312    13.095
   exper |    -.00227       .00126   -1.80   0.071  -.004737  .000197   20.5015
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1
```

Estos resultados son para la categoría "sirvientes", obteniéndose los mismos resultados que "prchange" respecto a las variables continuas y discretas. La ventaja una vez mas es que podemos obtener los valores de las desviaciones estándar, la desventaja es que puede tomar un largo tiempo su estimación, luego de la estimación multinomial, si el número de observaciones y las variables independientes, son muchas.

3.1.5. Graficando los Cambios Discretos con "prchange" y "mlogview"

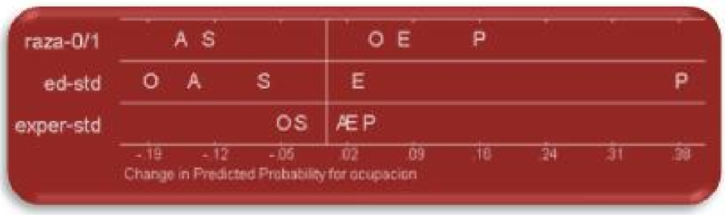
Una dificultad con las categorías nominales es la gran cantidad de coeficientes que necesitan ser considerados, para ayudar a que toda la información sea ordenada, los coeficientes de los cambios discretos pueden ser graficados con el programa "mlogview", el cual puede ser ejecutado, luego de estimar el mlogit y el prchange, y presentara la siguiente tabla:

```
mlogview
```



Que aceptando la opción DCplot nos presenta el siguiente grafico:

El grafico muestra como una unidad de incremento en cada variable afecta a la probabilidad de cada categoría. Parece que los efectos de ser blanco son los más grandes, cambios en una unidad en educación y mas aun, en experiencia son muy pequeños como para brindar alguna información. Esto podría sensibilizarse más si vemos los efectos de cambios en una desviación estándar en estas variables. Si optamos por dicha alternativa obtendríamos:



Ahora podemos observar que los efectos de la educación son los mayores y los efectos de la experiencia son los mas pequeños. Cada coeficiente también puede ser interpretado de manera individual:

- Los efectos de un cambio en una desviación estándar en la educación son los mas grandes, con un incremento que sobrepasa los 0.35 en la probabilidad de tener un trabajo profesional.

- Los efectos de la raza, también son importantes, en promedio los de raza negra son menos probables de conseguir trabajos como obreros, contratadores o profesionales, que los de raza blanca.

- Los cambios en una desviación estándar en la experiencia son mucho mas pequeños y muestran que los años de experiencia incrementan la probabilidad de tener ocupaciones con mayor grado de habilidad.

Al usar estos gráficos, debemos tener presente que los diferentes valores de los cambios discretos son obtenidos con diferentes niveles de las variables, lo cual es especificado con la opción x() o rest() en “prchange”.

3.1.6. Odds ratios usando “listcoef” y “mlogview”

Una reducción en la educación incrementa la probabilidad de elección de trabajo como obrero y como artesano, pero ¿cómo afecta al Odds de una persona que elige un trabajo de artesano en relación a un trabajo de obrero? Para responder a esta interrogante, los Odds ratios (también conocidos como factor de cambio en los coeficientes) pueden ser usados. Así manteniendo las demás variables constantes, el factor de cambio en el Odds de una categoría m vs una categoría n cuando x_k se incrementa en δ es igual a:

$$\frac{\Omega_{m/n}(x, x_k + \delta)}{\Omega_{m/n}(x, x_k)} = e^{\beta_{k,m/n}\delta}$$

De esta manera, si la cantidad de cambio es $\delta = 1$, entonces el Odds ratio puede ser interpretado como:

- Por una unidad de cambio en x_k, el Odds de m vs n se espera que cambie en el factor de $exp(\beta_{k,m/n})$, manteniendo las demás variables constantes.

O también, Si la cantidad de cambio es $\delta = s_{x_k}$, entonces el Odds ratio puede ser interpretado como:

- Por un cambio en una desviación estándar en x_k, se esperaria que el Odds de m vs n cambie en el factor de $exp(\delta = s_{x_k})$, manteniendo las demás variables constantes.

La dificultad de interpretar los Odds ratios para el MNLM, radica en que para entender el efecto de una variable, uno necesita examinar los coeficientes que resultan de la comparación en pares de todas las categorías. El reporte estándar del mlogit, incluye la comparación de J-1 categorías, considerando una categoría base. Mientras uno podría estimar los coeficientes de todas las posibles comparaciones, re-estimando el mlogit con categorías base diferentes. Usando el comando “listcoef” es mucho mas fácil, por ejemplo, para examinar el efecto de la raza:

El Odds ratio muestra que el efecto que tiene la raza sobre el tener un trabajo profesional vs un trabajo de sirviente es 5.9 lo cual puede ser interpretado como:

- El Odds de tener una ocupación profesional en relación a tener una ocupación de sirviente es 5.9 veces mayor para personas de raza blanca que para personas de raza negra, manteniendo las demás variables constantes.

mlogit (N=337): Factor Change in the Odds of **ocupacion**

Variable: **raza** (sd=.27642268)

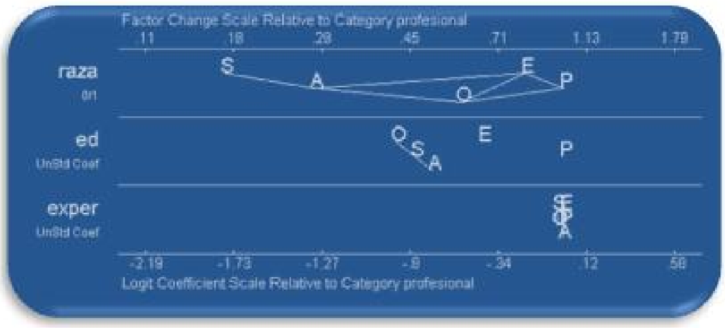
Odds comparing Group 1 vs Group 2	b	z	P>|z|	e^b	e^bStdX
sirvient-obrero	-1.23650	-1.707	0.088	0.2904	0.7105
sirvient-artesano	-0.47234	-0.782	0.434	0.6235	0.8776
sirvient-empleado	-1.57139	-1.741	0.082	0.2078	0.6477
sirvient-profesio	-1.77431	-2.350	0.019	0.1696	0.6123
obrero -sirvient	1.23650	1.707	0.088	3.4436	1.4075
obrero -artesano	0.76416	1.208	0.227	2.1472	1.2352
obrero -empleado	-0.33488	-0.359	0.720	0.7154	0.9116
obrero -profesio	-0.53780	-0.673	0.501	0.5840	0.8619
artesano-sirvient	0.47234	0.782	0.434	1.6037	1.1395
artesano-obrero	-0.76416	-1.208	0.227	0.4657	0.8096
artesano-empleado	-1.09904	-1.343	0.179	0.3332	0.7380
artesano-profesio	-1.30196	-2.011	0.044	0.2720	0.6978
empleado-sirvient	1.57139	1.741	0.082	4.8133	1.5440
empleado-obrero	0.33488	0.359	0.720	1.3978	1.0970
empleado-artesano	1.09904	1.343	0.179	3.0013	1.3550
empleado-profesio	-0.20292	-0.233	0.815	0.8163	0.9455
profesio-sirvient	1.77431	2.350	0.019	5.8962	1.6331
profesio-obrero	0.53780	0.673	0.501	1.7122	1.1603
profesio-artesano	1.30196	2.011	0.044	3.6765	1.4332
profesio-empleado	0.20292	0.233	0.815	1.2250	1.0577

Sin embargo, examinar todos los coeficientes aún para una simple variable, con solo cinco categorías, es complicado. El grafico del Odds ratio hace mas fácil de ver los resultados de un modelo complejo como el MNLM. Para explicar como interpretar un grafico de Odds ratio, debemos considerar cuatro aspectos al analizar la grafica:

- El signo de los coeficientes, si una letra esta a la derecha de otra letra, incrementos en la variable independiente hacen que la categoría de la derecha sea la más común.
- El efecto marginal, la distancia entre un par de letras indica la magnitud del efecto.
- La relación aditiva entre los coeficientes, lo cual se muestra acorde con las distancias entre las letras.
- La categoría base, hay una escala aditiva debajo del eje de medidas del valor de los coeficientes $\beta_{k,m/n}$. La escala multiplicativa se ubica encima de las medidas de $\exp(\beta_{k,m/n})$. La elección de la categoría base de comparación es arbitraria.

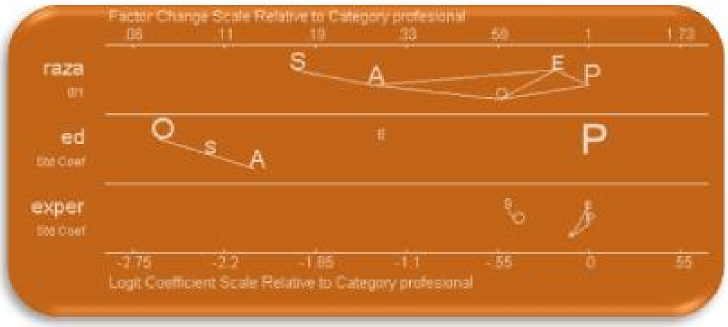
Algunas cosas son evidentes, por ejemplo, el efecto de la experiencia es el mas pequeño, aunque incrementos en la experiencia hacen mucho mas probable que uno este en un trabajo de artesano, empleador o profesional que en un trabajo de sirviente u obrero. Podemos observar también, que la educación es la variable que tiene el mayor efecto. Cuando las expectativas de educación se incrementan, hay un incremento en la probabilidad de tener un trabajo profesional comparado con cualquier otro tipo.

El grafico anterior no refleja la significancia estadística. La carencia de una significancia estadística es mostrada por la conexión mediante líneas, sugiriendo que estas dos categorías están atadas. Uno puede añadir el nivel de significancia en las salidas reportadas.



El grafico muestra que en la variable raza, se ordeno las categorías como sirviente, artesano, obrero, empleador y profesional, pero las conexiones muestran que ninguna de las categorías adyacentes son significativas, para raza. El ser blanco, incrementa el Odds de ser un trabajador artesanal con relación a tener un trabajo de sirviente, pero el efecto no es significativo. Sin embargo, el ser blanco incrementa significativamente el Odds de ser obrero, empleador o profesional con respecto a tener un trabajo de sirviente. El efecto para "ed" y "exper" puede ser interpretado de manera similar.

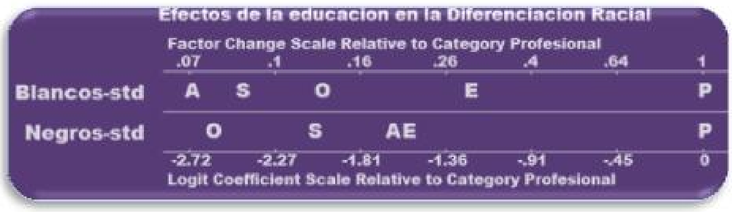
Nosotros sabemos que mientras el factor de cambio en los Odds es constante para todos los niveles de las variables, el cambio discreto puede ser mayor o menor, para diferentes valores de las variables. Por ejemplo, si el Odds se incrementa en el factor de 10, pero el Odds actual es 1:10000, entonces el impacto será pequeño. Pero si el Odds actual fuera 1:5, el impacto será mayor. La información de un cambio discreto en la probabilidad puede ser incorporado en el grafico, mediante el tamaño de las letras proporcionales al cambio discreto del Odds (específicamente, el área de las letras es proporcional al tamaño del cambio discreto). Esto puede añadirse a nuestro grafico de manera muy sencilla.



Si deseáramos comparar la estimación de los coeficientes de dos grupos, tales como el grupo racial, blancos y negros, debemos estimar los modelos para cada grupo, obtener los coeficientes de la variable que deseamos analizar introduciéndolas en una matriz, calcular la desviación estándar de dicha variable analizada y finalmente realizar la grafica.

```
mlogit  occ white ed exper if white==1, b(5) nolog
mlogit  occ white ed exper if white==0, b(5) nolog
matrix mnlbeta=( -.8307514, -.9225522, -.6876114,-.4196403\ ///
-.7012628, -.560695, -.882502,-.5311514 )
matrix list mnlbeta
matrix mnlbeta=mnlbeta'
matrix list mnlbeta
global mnlname="Blancos  Negros"
global mnlcatnm="Sirviente Artesano Obrero Empleador Profesional"
sum ed
matrix mnlsd=( 2.946427, 2.946427)
mlogplot, vars(Blancos Negros) packed or matrix std(ss) ///
note("Efectos de la educacion en la Diferenciacion Racial")
```

Dada las limitaciones de los datos (hubieron 28 casos en la estimación del logit del grupo de raza negra) y nuestro modelo, los resultados no muestran serios desarrollos en la diferenciación racial en las categorías laborales, pero si se ilustra la flexibilidad del comando.



3.1.7. Uso del Clarify

Clarify usa simulaciones estocásticas para ayudarnos a interpretar y presentar resultados estadísticos. Usa un modelo estadístico que el investigador elija el cual asuma cambios no estadísticos. Primero, el programa simula los principales parámetros de una distribución muestral asintótica, en muchos casos una normal multivariante, con media, igual al vector de parámetros estimados y varianza igual a la matriz de varianza-covarianza.

Por defecto el programa diseña 1000 simulaciones de parámetros, los cuales serán suficientes para múltiples aplicaciones. Luego Clarify convierte los parámetros simulados en cantidades interesantes con los cuales se predicen valores o valores esperados o primeras diferencias. Para conseguir este objetivo el usuario necesita escoger reales o hipotéticos valores de las variables explicativas (X) e indicar que cantidad debe ser calculada, condicional a dichas X. El programa permite a los investigadores calcular virtualmente una cantidad que debería resolver un problema particular y proveer de un numero de procedimientos fáciles de interpretar.

Sesión 4
Modelos Truncados y Censurados

Esta sección[1] estudia el conjunto de modelos con solución de esquina. Para el uso de los modelos, es importante el recordar el por qué se usan variables logit y probit en modelos de elección binaria, modelos tobit en modelos de respuesta de solucion de esquina o modelos tipo poison en modelos de recuento, y es por eso que se necesitan modelos que tomen en cuenta ciertas caracteristicas importantes de la distribución de y.

En e caso de la participación de la mujer en el mercado laboral, el problema es que una parte importante de las mujeres casadas decide no tener ningun trabajo asalariado.

En el caso de notas que se obtinene en una evaluación, las mismas que según el sistema de calificación pueden fluctuar solo entre 0 y 20. También se presenta cuando solo podemos observar el gasto efectivo de aquellas personas que adquieren un bien pero no su disponibilidad a pagar, más aún si es inferior al precio mínimo con el que es posible acceder al bien. Finalmente, también es el caso de los ingresos percibidos por el trabajo remunerado, dado que no es posible observar el ingreso potencial de una persona que no está laborando en el momento en que se recoge la información por analizar. En cualquiera de estas situaciones, las observaciones correspondientes son excluidas de la muestra (lo que se define como "truncamiento", ya sea incidental o no), o su incorporación en ella es distorcionada por un valor específico que no es el real (lo cual definimos como "censura").

Podemos tener tres tipos de variables dependientes continuas limitadas: las truncadas, las censuradas y las que poseen sesgo de selección (o truncamiento incidental).

[1] Basado en Introducción a la econometría de Jeffrey M Wooldridge y Modelos de panel y variables limitadas de Arlette Beltran y Juan Francisco Castro

4.1. Variables Dependientes con Truncamiento No Incidental

El truncamiento se produce cuando la variable dependiente (y_i) se observa, si y solo si esta toma un valor mayor que "a", donde "a" es una constante cualquiera. Lo mismo ocurre con toda la información referida a las posibles explicativas del modelo, el vector x_i, asociadas con estas observaciones truncadas.

Un ejemplo podria ser el análisis de la disponibilidad a pagar por un automovil nuevo, si es que es cierto que en el mercado el más barato que se puede encontrar tiene un precio de $7,000. De esta manera, cuando la persona está dispuesta a pagar dicho monto o más, es probable que compre el auto y que se registre su gasto efectivo y toda su información socioeconómica (x_i). Si la persona está dispuesta a pagar menos de $7,000, no realiza ninguna compra y no se cuenta con sus datos asociados; es decir está observación desaparece de la muestra.

4.1.1. Variable Aleatoria Truncada

Definamos el concepto de variable aleatoria truncada. Es aquella que tiene una tiene función de densidad de la forma:

$$f(y|y < a) = \frac{f(y)}{Pr(y > a)}$$

Dada la condicionalidad detrás de esta ecuación se justifica la necesidad de escalar la función de densidad original, f(y), de tal manera que su integral sea uno cuando solo se incluyan los valores no truncados, es decir, en este caso, los valores mayores a "a". Este procedimiento se conoce como "normalización de la densidad", donde el denominador de ésta ecuación es la constante normalizada que corresponde al integral del numerador en el rango entre $-\infty$ y "a".

La distribución de una variable truncada, tiene características especiales que pueden resumirse como sigue:

Si y $\rightarrow N(\mu, \mu^2)$ y "a" es una constante, entonces:

$$E(y|truncamiento) = \mu + \sigma\lambda(a)$$

$$Var(y|truncamiento) = \sigma^2[1 - \delta(\alpha)]$$

Donde $\alpha = \frac{a-\mu}{\sigma}$

La función $\lambda(*)$ es conocida como la "inversa del ratio de Mills", que en este caso, puede ser:

$$\lambda(\alpha) = \frac{f(\alpha)}{1-F(\alpha)}$$

si el truncamiento es hacia abajo ($y > a$)

$$\lambda(\alpha) = \frac{-f(\alpha)}{F(\alpha)}$$

si el truncamiento es hacia arriba ($y \leq a$)

La función $\delta(*)$, por su parte, viene dada por $\delta(\alpha) = \lambda(\alpha)[\lambda(\alpha) - \alpha]$, donde $0 < \delta(\alpha) < 1, \forall\alpha$

Notese que si se truncan los valores por debajo de una constante "a", la media de la variable truncada será mayor que la original, mientras que si se truncan hacia arriba, la primera será menor que la última. De otro lado, la varianza de la variable truncada será siempre menor que la de la variable original (dado que $\delta(\alpha)$ se encuentra entre 0 y 1).

4.1.2. Truncamiento en el Modelo de Regresión

Volviendo al ejemplo de la disponibilidad a pagar por un automóvil (y_i), definamos el siguiente modelo para explicarla a partir de un conjunto de variables explicativas (x_i):

$$y_i = x_i\beta + \mu_i$$

donde $\mu_i \rightarrow N(0, \sigma^2)$, por lo que $E(y_i|x_i = x_i\beta)$.

Recuérdese que solo es posible observar la variable dependiente y sus determinantes cuando esta supera el precio más bajo del mercado "a".Tomando el valor esperado de la disponibilidad pago, condicionado al truncamiento, se tiene:

$$E(y_i|y_i > a; x_i) = x_i\beta + E(\mu_i|y_i > a; x_i) = x_i\beta + E(\mu_i|\mu_i > a - x_i\beta; x_i)$$

Aplicando las caracteristicas antes descritas, se tiene:

$$E(y_i|y_i > a; x_i) = x_i\beta + \sigma\lambda(\alpha_i)$$

Donde $\lambda(\alpha_i) = \frac{f(\alpha_i)}{1-F(\alpha_i)}, \alpha_i = \frac{a - x_i\beta}{\sigma}$

De esta forma el modelo de variable dependiente truncada sería:

$$y_i|y_i > a = x_i\beta + \sigma\lambda(\alpha_i) + \mu_i$$

el mismo que solo es posible estimar para el conjunto de observaciones no truncadas.

4.1.3. Estimación del Modelo de Regresión con Variable Truncada

Si se estima linealmente y_i en función solo de x_i se estaría omitiendo la variable explicativa $\lambda(\alpha)$, la cual, debido a la perdida de información que implica el truncamiento, no es posible estimar de manera alguna. Por ello no es adecuado usar directamente MCO, y la alternativa es estimar el modelo por máxima verosimilitud utilizando la función de verosimilitud truncada:

$$L = \prod_{i=1}^{N} \frac{f(\mu_i)}{1 - F(\alpha_i)}$$

4.1.4. Impacto Marginal en el Modelo de Regresión

¿Qué resultado es el que interesa en el modelo de regresión truncada?¿El efecto impacto o los coeficientes estimados $\hat{\beta}$? Si es que solo se quiere analizar los efectos del cambio en una variable explicativa sobre la dependiente para aquellas observaciones no truncadas incluidas en la regresión, bastara con el efecto impacto correspondiente. El uso de los coeficientes β será de interés si se quiere generalizar los resultados a toda la población, esté truncado o no.

Consistente con nuestro ejemplo, mostremos a continuación cómo se deriva el efecto impacto correspondiente cuando la variable dependiente está truncada para valores menores que "a".

$$\frac{\partial E(y_i|y_i > a; x_i)}{\partial x_{ij}} = \beta_j + \sigma \frac{\partial \lambda(\alpha_i)}{\partial \alpha_i} \frac{\partial \alpha_i}{\partial x_{ij}}$$
$$\frac{\partial E(y_i|y_i > a; x_i)}{\partial x_{ij}} = \beta_j + \sigma \frac{\partial \lambda(\alpha_i)}{\partial \alpha_i} \frac{-\beta_j}{\sigma}$$
$$\frac{\partial E(y_i|y_i > a; x_i)}{\partial x_{ij}} = \beta_j \left[1 - \frac{\partial \lambda(\alpha_i)}{\partial \alpha_i}\right]$$

Para hallar el diferencial $\frac{\partial \lambda(\alpha_i)}{\partial \alpha_i}$ es necesario tomar en cuenta que $\frac{\partial F(\alpha_i)}{\partial \alpha_i} = f(\alpha_i)$ y que la función de densidad supuesta es la normal, por lo que $\frac{\partial f(\alpha_i)}{\partial \alpha_i} = -\alpha_i f(\alpha_i)$. Con esto, se tiene el siguiente resultado.

$$\frac{\partial E(y_i|y_i > a; x_j)}{\partial x_{ij}} = \beta[1 - \lambda(\alpha_i)(\lambda(\alpha_i) - \alpha_i)]$$

La expresión entre llaves, que se encuentra entre 0 y 1, es el factor de ajuste del coeficiente β_j (que corresponde el efecto impacto en un modelo lineal para toda la población), que da cuenta del efecto del truncamiento. Notese que σ afecta la magnitud de los efectos impacto (a través de α) mas no la dirección.

4.1.5. Variable Aleatoria Censurada

Retomando el ejemplo de la disponibilidad a pagar por el automóvil y supongamos que aún si la persona no compra el auto, si se registran sus datos x_i como cliente potencial. En este caso la variable y_i, tomara el valor pagado por la persona si ésta compra el auto, y el de 0 si no lo compra. En cualquiera de los dos casos, se habrá recogido información sobre el cliente. De esta manera, podemos decir que la variable y_i ha sido censurada en 0 para disponibilidades a pagar menores que $7,000 valor que es el precio mínimo de mercado.

El modelo conceptual utilizado para el caso de variables discretas, donde asumimos la existencia de una variable latente continua e ilimitada y cuya media condicional puede ser modelada como una combinación lineal de un conjunto de explicativas, también puede ser modelada como una combinación lineal de un conjunto de explicativas, también puede ser aplicado en este contexto. En el ejemplo anterior, la variable latente es la disponibilidad de pago la cual puede adoptar cualquier valor. La variable observada, en este caso, corresponde a la latente pero solo cuando esta última supera el precio mínimo de mercado.

Otro ejemplo nos ayudara en la formalización de este modelo. Supongamos que la variable latente y_i^* es el puntaje en una prueba de aptitud que incluye puntos en contra, mientras que y_i^* es el puntaje en una prueba de aptitud que incluye puntos en contra, mientras que y_i, se define de tal forma que:

$$y_i = \begin{cases} y_i^*, & \text{si } y_i^* > 0, \\ 0, & \text{si } y_i^* \leq 0. \end{cases}$$

En cualquiera de los dos casos se conocen los potenciales factores explicativas del puntaje x_i.

De esta manera la distribución de la variable y_i tiene dos componentes claramente diferenciados: la parte continua, para las observaciones no censuradas, y la discreta, para aquellas a las que se asigna el puntaje de corte. En este caso, no hay necesidad de escalar la distribución (como lo fue en el de las variables truncadas) ya que la probabilidad acumulada es de 100 % si se considera que a las observaciones censuradas se les asigna la probabilidad de estarlo.

4.1.6. Censura en el Modelo de Regresión

Si trabajamos en el ámbito del modelo de regresión, tenemos que la variable latente puede ser representada como:

$$y_i^* = x_i\beta + \mu_i$$

Para establecer el valor esperado de la variable observada (y), que considera también las observaciones censuradas, es necesario diferenciar entre dos situaciones alternativas. Al igual que en el ejemplo anterior, en lo que sigue suponemos que el valor de corte es igual a cero (a=0).

- Para una observación tomada al azar:

$$\begin{aligned} E(y_i|x_i) &= (0)Pr(y_i = 0) + E(y_i|y_i > 0; x_i)Pr(y_i > 0) \\ &= E(y_i|y_i > 0; x_i)Pr(y_i > 0) \\ &= (x_i\beta + \sigma\lambda(\alpha_i))(1 - F(\alpha_i)) \end{aligned}$$

  Nótese que ahora, como la censura es en 0, se tiene que:

$$\alpha_i = \frac{-x_i\beta}{\sigma}$$

$$\lambda(\alpha_i) = \frac{f(-x_i\beta/\sigma)}{1 - F(-x_i\beta/\sigma)} = \frac{f(x_i\beta/\sigma)}{F(x_i\beta/\sigma)}$$

  Y su varianza sería, en cambio:

$$Var(y_i|x_i) = \sigma^2 F(\alpha_i)[1 - \delta(\alpha_i) + (\alpha_i - \lambda(\alpha_i))^2(1 - F(\alpha_i))]$$

- Para una observación no censurada

  Como es la situación similar a la de las observaciones no truncadas, el modelo sería el mismo que el de la ecuación:

$$E(y_i|y_i > 0; x) = x_i\beta + \sigma\lambda(\alpha_i)$$

Para este modelo aplica todo lo dicho anteriormente, la pregunta seria ahora como estimar los modelos que contiene variables dependientes censuradas y específicamente aquellos planteados antes.

4.1.7. Estimación del Modelo de Regresión Censurada

a. Estimación por MCO en 2 Etapas

La estimación MCO se realiza mediante un procedimiento en dos etapas, que consiste en modelar el proceso de censura previamente a la estimación de la ecuación principal.

- Primera Etapa

  Se utiliza una variable auxiliar "z_i" de la forma:

$$z_i = \begin{cases} 1, & \text{si } y_i^* > 0\text{: No hay censura} \\ 0, & \text{si } y_i^* \leq 0\text{: Hay censura} \end{cases}$$

  A partir de ella y de un conjunto de explicativas que den cuenta de la censura, se estima un modelo probit para obtener el vector β/σ de estimados y construir $\hat{\alpha}$ y $\lambda(\hat{\alpha})$, según están definidos en las ecuaciones previas

- Segunda Etapa

  Se utiliza $\hat{\alpha}$ para estimar por MCO cualquiera de los dos modelos de las ecuaciones para una observación tomada al azar o para observaciones no censuradas:

  Modelo con todas las observaciones

$$y_i = (x_i\beta + \sigma\lambda(\alpha_i))F(-\alpha_i) + \mu_i = F(-\alpha_i)x\beta + \sigma f(\alpha_i) + \mu_i$$

  Modelo con todas las observaciones no censuradas

$$y_i|y_i > 0; y_i = x_i\beta + \sigma\lambda(\alpha_i) + \mu_i$$

El uso de uno u otro modelo dependerá del objetivo de la investigación. El primero permitirá predecir el valor promedio del total de observaciones. En el ejemplo de la disponibilidad a pagar por un automóvil, seria el pago promedio realizado por una persona cualquiera de la muestra total, haya comprado el auto o no (el valor promedio de compra, consirando que aquellos que no realizaron la compra pagaron un monto igual a cero). El segundo modelo, en cambio, servirá para calcular el valor promedio pagado por aquellas observaciones no censuradas y haría posible predecir el valor promedio de las ventas efectivas.

b. Máxima Verosimilitud: El Modelo Tobit[2]

Para estimar un modelo con variable dependiente censurada mediante el método de máxima verosimilitud (MV), es necesario considerar que se tiene dos tipos de información. Aquella referida a las observaciones no censuradas, para las que se conoce la esperanza condicional de y_i, y aquella referida a las observaciones censuradas, para las que se conoce la probabilidad de estar censurada.

La función de verosimilitud se construye considerando ambos componentes. Así:

$$L = \prod_{y_i>0} Pr(y_i > 0) f(y_i|y_i > 0) \prod_{y_i=0} Pr(y_i = 0)$$

Si recordamos que la función de densidad truncada viene dada por:

$$f(y_i|y_i > 0) = \frac{f(y_i)}{Pr(y_i > 0)}$$

Por lo tanto podemos establecer la función de verosimilitud como:

$$L = \prod_{y_i>0} f(y_i) \prod_{y_i=0} Pr(y_i = 0)$$

Note que el modelo tobit implica que los coeficientes estimados promedian dos tipos de efectos de las variables explicativas, aquel sobre la probabilidad de estar censurado y dado que no lo está, el efecto sobre el valor esperado de y_i.

Si no es posible garantizar que las mismas variables explicativas den cuenta de la censura, así como del fenómeno económico que se quiere analizar condicionado a dicha censura, el tobit puede no ser el

[2] Tobin (1956) fue el primero en vincular el problema de censura con el análisis de regresión. Relacionó este problema con el modelo probit el el sentido de que hay dos tipos de observaciones: sobre las que se tiene el valor de la dependiente y las que tienen un valor de cero asignado. Por dicha razón se le conoce como el modelo probit de tobin o tobit.

modelo más adecuado para realizar la estimación, ya que el procedimiento que involucra implica restringir ambos modelos a un mismo set de variables explicativas. Por ejemplo, saber conducir un automóvil, puede ser una explicativa importante para adquirir o no uno, pero podría no tener mayor impacto sobre la cantidad que se paga por él una vez que se ha decidido comprarlo. En este caso es mejor usar el método de estimación en dos etapas visto previamente, en el que se da libertad para incorporar variables explicativas distintas en cada una de ellas.

Las estimaciones por MCO sobre toda la muestra que desconocen el problema de censura, son inconsistentes y suelen ser menores en valor absoluto a los del modelo Tobit.

4.1.8. Efectos Marginales y Bondad de Ajuste

Si se analiza cuál es la medida de bondad de ajuste más apropiada en el caso de un modelo censurado, podría elegirse el cuadrado del coeficiente de correlación entre y_i e $\hat{y}_i$, donde esta última se construye a partir del modelo dado con todas las observaciones. El estadístico es distinto al R-cuadrado del MCO.

La definición basada en el coeficiente de correlación es preferida a la del R-cuadrado debido a que tiene la ventaja de fluctuar entre 0 y 1, cosa que no ocurre con el segundo, el que puede ser negativo en regresiones sin intercepto. De todas formas, es necesario tener en cuenta que el R-cuadrado no es tan importante en modelos censurados , especialmente en el caso del tobit, q que a diferencias de MCO, no maximiza este estadístico sino la función log-verosimil.

En cuanto a los efectos impacto, puede ser interesante estimarlos tanto para la muestra completa, como para las observaciones no censuradas. En este segundo caso, el efecto impacto será similar al de variables truncadas, aún cuando se observa un cambio de signo (si tomamos en cuenta que se está trabajando con una censura hacia abajo, con un corte igual a cero y suponemos una distribución simétrica). Así:

$$\frac{\partial E(y_i|y_i > 0; x_i)}{\partial x_{ij}} = \beta_j[1 - \lambda(\alpha_i)(\alpha_i + \lambda(\alpha_i))]$$

Este resultado, sin embargo, tiene las mismas consecuencias vistas previamente respecto del problema de truncamiento.

En el caso para el modelo de la muestra completa, se tiene el siguiente efecto marginal:

$$\frac{\partial E(y_i|x_i)}{\partial x_{ij}} = \beta_j F(-\alpha_j) + x_i\beta f(\alpha_i)\frac{\beta_j}{\sigma} - \sigma\frac{x_i\beta}{\sigma}f(\alpha_i)\frac{\beta_j}{\sigma}$$

$$\frac{\partial E(y_i|x_i)}{\partial x_{ij}} = \beta_j F(-\alpha_j)$$

De esta manera, en el caso de trabajar con la muestra completa, para que el coeficiente β_j refleje el efecto impacto de la variable explicativa j sobre el valor esperado de y, es necesario multiplicarlo por la probabilidad de la no censura, $F(-\alpha_i)$. Si comparamos este efecto impacto con aquel asociado al de toda la población (β), notaremos que ambos se asemejarán en la medida en que $F(-\alpha_i)$ tiende a 1. Como es de esperarse, los resultados que toman en cuenta una potencial censura en la muestra y aquellos referidos a la data sin censurar serán equivalentes en la medida en que la mayoría de observaciones se concentren en la parte no censurada. Bajo estas circunstancias, las estimaciones que toman en cuenta la especificación para la medida condicional dada en el modelo de censura, serán equivalentes a aquellas que se obtenían si se regresiona y_i sobre x_i mediante MCO. Es decir, $E(y_i|x_i) = (x_i\beta + \sigma\lambda(\alpha_i))(1 - F(\alpha_i)) \rightarrow x_i\beta$, en la medida en que $a \rightarrow \infty$.

4.2. Variable de Truncamiento Incidental, Sesgo de Selección

El problema de sesgo de selección se produce cuando la inclusión de una unidad económica en la muestra depende de una decisión previa que no es exógena, por lo que resulta ser una muestra no aleatoria, solo se presenta sesgo de selección cuando la muestra no es aleatoria o la selección muestral no es exógena. Es decir, si por ejemplo se separan observaciones de una muestra de manera aleatoria, o se utiliza algún criterio exógeno como la edad, el sexo, la raza, no se producirá un problema de sesgo de selección. En particular, y tal como veremos más adelante, el sesgo ocurre cuando el componente no observable de la decisión de pertenecer a la muestra está correlacionado con el componente no observable del fenómeno bajo análisis.

Por ejemplo, supongamos que se quiere analizar el rendimiento estudiantil pero solo se cuenta con información suficiente sobre dicho rendimiento y sus determinantes para el caso de escuelas privadas. Como veremos, el hecho de trabajar solo con aquellos niños jóvenes cuya familias decidieron matricularlos en un colegio particular puede tener un efecto sobre el modelo que se busca estimar y en especial, sobre su media.

4.2.1. El modelo de Truncamiento Incidental

Analicemos primero la decisión de asistir a determinado tipo de colegio (ecuación de selección). Para esto, y de acuerdo con la formulación desarrollada para los modelos de elección binaria, supongamos que la utilidad de asistir a un colegio privado (z_i^*) puede representarse como:

$$z_i^* = w_i\gamma + \epsilon_i$$

Dicha ecuación seria la ecuación de selección, la variable z_i^* no es directamente observable. Lo que si se observa es si el estudiante está matriculado en un colegio privado o no, resultado que depende de que la utilidad de hacerlo supere determinado umbral (a). De ésta manera, si $z_i^* > a$, el alumno se matricula en un colegio privado y, por lo mismo, pertenece a la muestra de trabajo.

En lo que respecta al rendimiento, supongamos que, en general, este puede ser representado como:

$$Y_i^* = x_i\beta + \mu_i$$

Que es la ecuación de rendimiento, donde Y_i^* es la nota final obtenida en determinado año de estudios escolares. Es necesario notar que en la muestra de trabajo no se tienen observaciones de la distribución completa de Y_i^*, sino solo de aquellas observaciones provenientes de estudiantes matriculados en una escuela privada. Es decir, la variable dependiente observada Y_i viene dada según: $Y_i = Y_i^*$ si $z_i^* > a$. Esto implica que si bien $E[y_i^*|x_iw_i] = x_i\beta$, lo mismo no ocurre para $E[y_i|x_iw_i]$. En particular, la esperanza condicional de interés viene dada por: $E[y_i|x_iw_i] = E[y_i^*|z_i^* > a; x_iw_i]$.

En este caso será necesario definir la densidad condicional de y_i^* dado z_i^* de la siguiente manera:

$$f(y_i^*, z_i^*|z_i^* > a) = \frac{f(y_i^*, z_i^*|z_i^* > a)}{Pr(z_i^* > a)}$$

y verificar sus propiedades a partir de lo siguiente.

Distribución truncada conjunta

Si dos variables (y, z) tienen una distribución normal bivariada, con medias μ_y y μ_z, varianzas σ_y^2 y σ_z^2 y correlación ρ_{yz} (distinta de cero), entonces:

$$E[y|truncamientosobrez] = \mu_y + \rho_{yz}\sigma_y\lambda(\alpha_z)$$

$$Var[y|truncamientosobrez] = \sigma_y^2[1 - \rho_{yz}^2\delta(\alpha_z)]$$

Donde $\alpha_z = \frac{(a-\mu_z)}{\sigma_z}$, y $\lambda(.)$, la inversa del ratio de Mills, viene dada según:

$\lambda(\alpha_z) = \frac{f(\alpha_z)}{1-F(\alpha_z)}$ si el truncamiento es hacia abajo ($z > a$)
$\lambda(\alpha_z) = \frac{-f(\alpha_z)}{F(\alpha_z)}$ si el truncamiento es hacia arriba ($z \leq a$)
La función $\delta(.)$, por su parte, viene dada por $\delta(\alpha_z) = \lambda(\alpha_z)[\lambda(\alpha_z) - \alpha_z]$, donde $0 < \delta(\alpha_z) < 1$.

Nótese que la media de la variable truncada incidentalmente se desplaza en igual dirección que ρ_{yz} cuando el truncamiento es hacia abajo y en dirección opuesta cuando ($z \leq a$). La varianza se reduce cualquiera sea el caso ya que $\delta(.)$ y ρ_{yz}^2 están entre 0 y 1.

Si volvemos al ejemplo planteado y tomamos en cuenta los resultados así como las especificaciones para z_i^* e y_i^*, repectivamente tenemos que:

$$E[y_i|z_i^* > a; x_i w_i] = E[y_i^*|z_i^* > a; x_i w_i] = x_i\beta + \rho_{\mu\epsilon}\sigma_\mu\lambda(\alpha_z)$$

Donde: $\alpha_z = \frac{a - w_i\gamma}{\sigma_\epsilon}$ y $\lambda(\alpha_z) = \frac{\alpha_z}{1-F(\alpha_z)}$.

Vale la pena destacar varios elementos de la expresión anterior. En primer lugar, es claro que $E[y_i|z_i^* > a; x_i w_i] \neq x_i\beta$, excepto cuanto $\rho_{\mu\epsilon} = 0$ o cuando $a \to -\infty$. Es decir, no bastará con modelar la esperanza de nuestra variable dependiente como una combinación lineal de sus determinantes si es que solo es posible observarla efectivamente cuando el agente cumple con una característica especial (no es cierto que $a \to -\infty$) y dicha característica influye sobre el resultado que estoy modelando ($\rho_{\mu\epsilon} \neq 0$).

Para el ejemplo considerado, preguntarse si $\rho_{\mu\epsilon} \neq 0$ equivale a preguntarse si es que el hecho de estar matriculado en un colegio privado (la característica especial que hace que una unidad sea parte de la muestra) influye sobre el rendimiento del estudiante (el fenómeno que se está modelando). Al respecto, nuestra respuesta será afirmativa en la medida en que creamos que, además de las características socieconómicas tipicamente observables (como la importancia que da el hogar a la acumulación del capital humano) que afecta tanto a la decisión de qué tipo de colegio elegir como al rendimiento del niño en el colegio. Estos no observables serán capturados en ϵ_i y μ_i y el grado de dirección en el que afecten ambos fenómenos (selección y rendimiento) vendrá dado, precisamente por la correlación entre los dos términos de error ($\rho_{\mu\epsilon}$) y su signo.

De considerar un sistema educativo como el peruano, donde la calidad de educación básica privada es superior a la pública, cabria esperar una correlación positiva: más importancia asignada a la acumulación de capital humano por parte del hogar impactará positivamente tanto en la decisión de matricula en una escuela privada (la posibilidad de observar al agente en la muestra considerada) como en el rendimiento en la misma. En este sentido, lo que se plantea es corregir al alza, la esperanza del rendimiento para tomar en cuenta que se está trabajando con aquellos individuos que pertenecen a hogares especialmente preocupados por la educación de sus hijos.

Tan o más importantes que entender la corrección introducida sobre la esperanza de la variable de interés, es entender el riesgo que corremos de omitirla. Es claro que la corrección propuesta no es otra cosa que una variable relevante más, cuya inclusión es necesaria para lograr una correcta especificación de la media condicional de la variable dependiente. No incluirla, por tanto, conduciría a los conocidos problemas asociados a la omisión de variables. En particular, tendríamos estimadores sesgados o para el caso de muestras grandes, un estimador no consistente.

4.2.2. Estimación del Modelo de Truncamiento Incidental

La estimación del modelo de una variable dependiente con sesgo de selección puede hacerse a través de dos alternativas:

a. MCO: Modelo de Heckit[3] En este caso se usa también un procedimiento de dos etapas, en la primera se estima la ecuación de selección, que caracteriza la forma en que las observaciones son incluidas en la ecuación principal. La segunda etapa consiste en estimar el modelo principal con la muestra no truncada incidentalmente.

- Primera Etapa Se estima la ecuación de selección utilizando una variable auxiliar (z_i) de la forma:

$$z_i = \begin{cases} 1, & \text{si } z_i^* > 0\text{: Matriculado en colegio privado} \\ 0, & \text{si } z_i^* \leq 0\text{: Matriculado en colegio público} \end{cases}$$

[3]Heckman (1979)

Para ello se estima un probit que permitirá obtener los parámetros γ/σ_ϵ, con los cuales se construyen $\hat{\alpha}_z$ y $\lambda(\hat{\alpha}_z)$.

- Segunda Etapa

  En la segunda etapa se utiliza $\lambda(\hat{\alpha}_z)$ para estimar por MCO el siguiente modelo:

$$y_i = x_i\beta + \rho_{\mu\epsilon}\sigma_\mu\lambda(\hat{\alpha}_z)$$

  Es decir, regresionar y_i sobre x_i y $\lambda(\hat{\alpha}_z)$.

  Es necesario considerar que en la ecuación de selección se debe incluir, por lo menos, una variable explicativa adicional que no esté en la ecuación de interés. Si bien la inversa del ratio de Mills es una función no lineal de las explicativas de la ecuación de selección, frecuentemente se puede aproximar a través de una función lineal. Por lo mismo, no incluir dicho regresor adicional podría llevar a que la inversa del ratio de Mills esté altamente correlacionada con las otras explicativas de la ecuación de interés.

b. Máxima Verosimilitud

Para estimar un modelo con sesgo de selección a través del método MV es necesario considerar que se tiene dos tipos de información. Aquella referida a las observaciones no truncadas, para las que se conoce la esperanza condicional y aquella referida a las observaciones truncadas, para las que se cuenta con la probabilidad de estarlo.

Entonces, la función de verosimilitud se construye considerando ambos tipos de información:

$$L = \prod_{z_i^*>0} Pr(z_i^* > 0) f(y_i|z_i^* > 0) \prod_{z_i^*=0} Pr(z_i^* \leq 0)$$

Si tenemos en cuenta que:

$$f(y_i|z_i^* > 0) = \frac{f(y_i)}{Pr(z_i^* > 0)}$$

Por lo tanto podemos establecer la función de verosimilitud como:

$$L = \prod_{z_i^*>0} f(y_i) \prod_{z_i^*=0} Pr(z_i^* \leq 0)$$

4.2.3. Efectos Marginales

Finalmente el efecto impacto de una variable explicativa que se encuentra tanto en la ecuación de selección como en la de interés, sobre una dependiente con truncamiento incidental, teniendo:

$$E[y_i|z_i^* > a; x_i, w_i] = x_i\beta + \rho_{\mu\epsilon}\sigma_\mu\lambda(\hat{\alpha}_z)$$

donde

$$\lambda(\alpha_z) = \frac{f(\alpha_z)}{1 - F(\alpha_z)}$$

Si suponemos que a=0 tenemos además que $\alpha_z = \frac{-w_i\gamma}{\sigma_\epsilon}$ y $\lambda(\alpha_z) = \frac{f(\alpha_z)}{F(-\alpha_z)}$. Entonces el impacto de un cambio en una variable explicativa x_i sobre la media de y_i truncada incidentalmente seria:

$$\begin{aligned}
\frac{\partial E(y_i|z_i^* > 0; x_i, w_i)}{\partial x_{ij}} &= \beta_j + \rho_{\mu\epsilon}\sigma_\mu \frac{\partial\lambda(\alpha_z)}{\partial\alpha_z}\frac{\partial\alpha_z}{\partial_{ij}} \\
&= \beta_j + \rho_{\mu\epsilon}\sigma_\mu[-\alpha_z\lambda(\alpha_z) + \lambda(\alpha_z)^2]\left[-\frac{\gamma_j}{\sigma_\epsilon}\right] \\
&= \beta_j - \frac{\rho_{\mu\epsilon}\sigma_\mu\gamma_j}{\sigma_\epsilon}[\lambda(\alpha_z)^2 - \alpha_z\lambda(\alpha_z)]
\end{aligned}$$

Donde el último corchete es igual a $\delta(\alpha_z)$.

Recordando que la variable x_j se encuentra en ambas ecuaciones, la de selección y la de interés. Si $\rho_{\mu\epsilon}$ es positivo y la esperanza de y_i es mayor para valores positivos de z_i^*, como $\delta(\alpha_z)$ se encuentra entre 0 y 1, el segundo término que aparece restando a β_j reduce el efecto impacto. El cambio en la probabilidad de que $z_i = 1$ ante un cambio en x_j afecta a la media de y_i, ya que en el grupo donde $z_i = 1$ la media es más alta. Así el termino que resta a β_j compensa este efecto, dejando solo el efecto marginal de un cambio en x_j sobre la media de y_i, dado que $z_i^* > 0$.

Al igual que en el caso del truncamiento, β_j se refiere al efecto impacto del j-esimo regresor sobre la media de la variable dependiente en toda la población. En otras palabras β_j se refiere al efecto impacto sobre el rendimiento estudiantil, mientras que el resultado anterior se refiere al efecto impacto sobre el rendimiento en escuelas privadas de una variable que afecta tanto al rendimiento como a la probabilidad de estudiar en una escuela de este tipo.

La corrección por truncamiento incidental no es solo relevante cuando nos interesa conocer los efectos marginales para la muestra truncada. En muchos casos el interés se concentra en determinar el valor del vector β y su estimación requiere considerar la corrección por la inversa del ratio de Mills.

Por último, es necesario mencionar que en el caso de que las ecuaciones de interés tengan especificaciones diferentes para ambos grupos. Esto equivale a que el rendimiento de la escuela privada responda a un modelo distinto al de la pública. En algunos casos será necesario estimar dos regresiones separadas para cada uno, evaluando en ambas la corrección por el sesgo de selección correspondiente.

```
*simulacion del modelo censurado
********************************
clear
clear all
set obs 1000
gen u = invnorm(uniform())
gen x = invnorm(uniform())
gen y = x + u
su
replace y=0 if y < 0
su
hist y
tw (sc y x, m(Oh) msize(small) ) (lfit y x, lw(thick))
reg y x
tobit y x, ll(0)   /*censura a la izquierda*/

*** Proceso Generador de Datos ***
clear
set obs 1000
gen u = invnorm(uniform())
gen x = invnorm(uniform())
gen y = x + u
su
replace y=-2 if y < -2
replace y=2 if y > 2
su
hist y

*** Estimation ***
/*censura a la izquierda y derecha*/
```

```
reg y x
tobit y x, ll(-2) ul(2)

*Simulacion modelo truncado
***************************
clear
set obs 1000
gen u = invnorm(uniform())
gen x = invnorm(uniform())
gen y = x + u
replace y = . if y > 0  /*truncamiento hacia arriba*/
reg y x
truncreg y x , ul(0)

clear
set obs 1000
gen u = invnorm(uniform())
gen x = invnorm(uniform())
gen y = x + u
replace y = . if y < 0  /*truncamiento hacia abajo*/
reg y x
truncreg y x , ll(0)


*Simulacion modelo con truncamiento incidental
***********************************************
clear
set obs 1000
gen u = invnormal(uniform())   /*error de la ec de medicion*/
gen v = 1 + u + invnormal(uniform())  /*error de la ec de seleccion*/
gen x = invnormal(uniform())
gen z = invnormal(uniform())
gen d = (1 + x + z + v > 0)  /*ec de seleccion*/
gen ystar = 1 + x + u   /*ec de medicion*/
gen y = ystar if d==1
tw (kdensity y ) (kdensity ystar )

heckman y x, select(d = z x)  /*ec Heckit*/

Ahora apliquemos a un modelo de trabajo:

clear
clear all
```

```
set mem 50m
set matsize 800

use womenwk, clear
d
su age educ married children wage

* generando una dummy:
generate d = 1
replace  d = 0 if wage == .

*Ecuacion de selección
probit d married educ age children
predict xb,xb

*Construcción del ratio de Mills
gen Mills = normalden(xb)/(normal(xb) )

*comparacion entre el modelo mco y el corregido
regress wage educ age
regress wage educ age  Mills

*Modelación alternativa MC2E
heckman wage educ age, select (married children educ age) twostep
heckman wage educ age, select (d = married children educ age) twostep

predict cndwage, ycond

*Modelación por MV
heckman wage educ age,select(d=educ age married children)


***************************************************************
use mroz, clear
d
su lwage   educ exper expersq
reg lwage educ exper expersq
tab inlf
count
tabstat lwage,stat(N mean  sd) by( inlf)

heckman lwage educ exper expersq, \\\
select(inlf=educ exper expersq nwifeinc age kidslt6 kidsge6) twostep
```

```
heckman lwage educ exper expersq nwifeinc age kidslt6 kidsge6, \\\
select(inlf=educ exper expersq nwifeinc  age kidslt6 kidsge6) twostep

heckman lwage educ exper expersq, \\\
select(inlf=educ exper expersq nwifeinc age kidslt6 kidsge6)
```

Sesión 5
Variables Instrumentales

Tres problemas a considerar:

- Sesgo por omisión de variables (OV) no observadas (y, por tanto, no incluidas en la regresión) que están correlacionadas con X;
- Sesgo por causalidad simultánea (CS); es decir, X causa a Y e Y causa a X;
- Sesgo por errores en las variables (EV); es decir, medimos X con error.

La regresión VI puede eliminar los anteriores sesgos.

$$Y_i = \beta_0 + \beta_1 * X_i + \mu_i$$

- La regresión VI divide X en dos partes: una que puede estar correlacionada con μ, y la otra que no. Aislando esta última, podremos estimar β_1. Para ello, utilizaremos una variable instrumental, Z_i, no correlacionada con μ_i.
- Para estimar β_1, la VI detecta aquellos movimientos en X_i que no están correlacionados con μ_i.

5.0.4. Selección de los Instrumentos

Para que un "instrumento" Z sea válido, debe satisfacer las dos siguientes condiciones:

- Relevante: $corr(Z_i, X_i) \neq 0$
- Exógeno: $corr(Z_i, u_i) = 0$

5.1. Estimación por MC2E

Este método consta de dos etapas - dos regresiones:

a. Primero se aísla la parte de X que no está correlacionada con u: regresión de X sobre Z por MCO:

$$X_i = \pi_0 + \pi_1 Z_i + v_i$$

Como Z_i no está correlacionada con μ_i, $\pi_0 + \pi_1 Z_i$, tampoco lo estará con μ_i. No conocemos π_0 ó π_1 pero sabemos estimarlos. Hallar las estimaciones de X_i, $\hat{X}_i$, donde $\hat{X}_i = \hat{\pi}_0 + \hat{\pi}_1 Z_i$, para i = 1,...,n.

b. Reemplazar X_i por $\hat{X}_i$ en la regresión de interés, y estimar Y sobre $\hat{X}_i$ por MCO:

$$Y_i = \beta_0 + \beta_1 \hat{X}_i + \mu_i....(2)$$

Como $\hat{X}_i$ no está correlacionada con μ_i en muestras grandes, el primero de los supuestos MCO se cumple. Por tanto, β_1 puede estimarse por MCO en (2).
Éste es un argumento de muestras grandes (es decir π_0 y π_1 estarán bien estimadas en (1)) El estimador resultante es el MC2E, $\hat{\beta}_1^{MC2E}$.

Si disponemos de un instrumento válido, Z_i,

Etapa 1^{ra}: Regresión de X_i sobre Z_i, para obtener $\hat{X}_i$

Etapa 2^{da}: Regresión de Y_i sobre $\hat{X}_i$; el coeficiente de $\hat{X}_i$ es el MC2E, $\hat{\beta}_1^{MC2E}$.

Entonces, $\hat{\beta}_1^{MC2E}$ es consistente de β_1.

Usemos la base de datos MROZ y empecemos a trabajar.

```
use MROZ, clear
describe

drop if inlf==0

regress lwage educ exper expersq

regress lwage  educ exper expersq motheduc fatheduc huseduc

regress educ  exper expersq motheduc fatheduc huseduc
```

Test de Hausman

Veamos el problema de endogeneidad

```
regress educ  exper expersq motheduc fatheduc huseduc
predict edu_res, res
regress lwage educ exper expersq edu_res
test edu_res
ivreg  lwage (educ =  motheduc fatheduc huseduc) exper expersq
ivreg  lwage (educ =  motheduc fatheduc huseduc) exper expersq, first
```

MC2E Primera etapa: Ecuación reducida para EDUC:

```
regress educ  exper expersq motheduc fatheduc huseduc
predict edu_pre, xb
```

Segunda etapa: Ecuación estructural usando edu_pre en lugar de educ

```
regress lwage edu_pre exper expersq

ivreg  lwage (educ =  motheduc fatheduc huseduc) exper expersq
ivendog
```

Veamos el problema con una simulación:

```
/*Modelo que replica mc2e vs mco*/
*Exacta Identificacion
clear all
clear
set obs 100

local b0=1.3
local b1=1.5
local b2=2.7

gen x1=invnorm(uniform())
gen x3=invnorm(uniform())
gen e1=invnorm(uniform())

gen x2=0.5+1.8*x3+1.5*e1

gen y=‘b0’+‘b1’*x1+‘b2’*x2+e1

reg y x2 x1
```

```
ivreg y (x2=x3) x1
ivendog
overid
ivreg2 y (x2=x3) x1


***************************************************
*Sobre Identificacion
clear all
clear
set obs 100

local b0=1.3
local b1=1.5
local b2=2.7

gen x1=invnorm(uniform())
gen x3=invnorm(uniform())
gen x4=invnorm(uniform())
gen e1=invnorm(uniform())

gen x2=0.5+0.8*x3+0.7*x4+1.5*e1

gen y=`b0'+`b1'*x1+`b2'*x2+e1

reg y x2 x1

ivreg y (x2=x3 x4) x1

ivendog
overid
ivreg2 y (x2=x3 x4) x1

*******************************************************
*Simulacion Montecarlo (Exacta Identificación)

clear all
clear
global nobs=1000
set obs $nobs

local b0=1.3
local b1=1.5
```

```
local b2=2.7

gen x1=invnorm(uniform())
gen x3=invnorm(uniform())
gen bmco=.
gen bvi=.
gen e1=invnorm(uniform())
gen x2=0.5+1.8*x3+1.5*e1
gen y=‘b0’+‘b1’*x1+‘b2’*x2+e1
forvalue  k=1/$nobs {
replace e1=invnorm(uniform())
replace x2=0.5+1.8*x3+1.5*e1
replace y=‘b0’+‘b1’*x1+‘b2’*x2+e1

quietly reg y x2 x1
replace bmco=_b[x2] in ‘k’/‘k’
quietly ivreg y (x2=x3) x1
replace bvi=_b[x2] in ‘k’/‘k’
}

tw (kdensity bmco) (kdensity bvi), xline(‘b2’) \\\
legend(label(1 "Estimacion MCO") label(2 "Estimacion VI")) \\\
scheme(vg_teal) title("Estimador MCO vs VI") ytitle("Parametro - b2")

ivendog
overid
ivreg2 y (x2=x3) x1


*******************************************************
*Simulacion Montecarlo (Exacta Identificación - Segundo Método)

capture program drop visim
program visim, rclass
version 11
syntax [, obs(integer 1) beta0(real 0) beta1(real 0) ]
drop _all
set obs ‘obs’
generate e = invnormal(uniform())
generate x1 = uniform()*10
generate x =0.5+0.9*x1 + 1.5*e
generate y = ‘beta0’ + ‘beta1’*x + e
ivreg y (x=x1)
```

```
return scalar beta1 = _b[x]
regress y x
return scalar beta2 = _b[x]
end

simulate bvi1 = r(beta1)  bmco1 = r(beta2), reps(10000): visim, \\\
obs(100) beta0(1.5) beta1(2.5)
sum bvi1 bmco1

tw (kdensity bvi1) (kdensity bmco1), xline(2.5) \\\
legend(label(1 "Dist_beta_vi") label(2 "Dist_beta_mco")) \\\
scheme(vg_blue)

tw (histogram bvi1, color(red)) (histogram bmco1, color(blue)), \\\
xline(2.5) legend(label(1 "Dist_beta_vi") label(2 "Dist_beta_mco"))
```

Sesión 6
Método Generalizado de Momentos

6.1. Técnica de Estimación

El modelo estadistico basa su analisis en la especificación de las condiciones de momentos. La notación usual es:

$$E[m(Y_i;\theta_0)] = 0$$

donde:

- $Y_1, ..., Y_n$ son variables aleatorias (o vectores) de una muestra observada.
- $m(Y;\theta)$ es una función (o vector de funciones) que especifica la forma funcional del modelo.
- θ_0 es el verdadero valor del parametro (vector) θ.

Un ejemplo simple es un modelo con una media constante. En este caso $Y_1, ..., Y_n$ son variables aleatorias unidimensionales y:

$$m(Y,\theta) = Y - \theta$$

De esta manera la condición de momento $E[m(Y_i,\theta_0)] = 0$ implica que $E[Y_i] = \theta_i$ para todo i.

Una manera alternativa de expresar el modelo seria: $Y_i = \theta_0 + U_i$ lo cual implica $E[U_i] = 0$, bastante usada en los modelos de regresión.

Un modelo GMM es correctamente especificado si $E[m(Y_i;\theta_0)]$ existe y se cumple la condición de momentos para algun θ_0.

Un modelo estadístico para una media constante y una varianza de $Y_1, ...Y_n$ puede ser escrita como:

$$m(Y;\theta) = \begin{bmatrix} Y-\mu \\ (Y-\mu)^2 - \sigma^2 \end{bmatrix}$$

Donde $\theta = (\mu, \sigma^2)'$ es un vector de parametros de 2x1 y dado $E[m(Y_i;\theta_0)] = 0$, implica que $E[Y_i] = \mu_0$ y $var[Y_i] = E[(Y_i - \mu_0)^2] = \sigma_0^2$, para todo i, donde $\theta_0 = (\mu_0, \sigma_0^2)'$

Ahora bien, otro ejemplo podria ser suponer que Y_i es una variable aleatoria bidimensional $Y_i = (Y_{1,i}, Y_{2,1})'$. Un modelo estadistico para las medias constantes de estas variables aleatorias puede ser escrito como:

$$m(Y;\theta) = \begin{bmatrix} Y_1 - \theta_1 \\ Y_2 - \theta_2 \end{bmatrix}$$

El cual es una forma simple de un sistema de ecuaciones que puede ser escrito como:

$$Y_{1,i} = \theta_{1,0} + U_{1,i}$$

$$Y_{2,i} = \theta_{2,0} + U_{2,i}$$

Este modelo no supone autocorrelación o heterocedasticidad

6.2. El modelo de regresión

El modelo de regresión propuesto para la expectativa condicional $E[Y_i|X_i]$ es:

$$E[Y_i|X_i] = \beta_{1,0} + \beta_{2,0}X_i$$

Notemos que $\beta_0 = (\beta_{1,0}, \beta_{2,0})'$ hace referencia a valores verdadeors de los parametros del vector $\beta = (\beta_1, \beta_2)'$. Equivalentemente, el modelo puede ser escrito de la siguiente forma:

$$Y_i = \beta_{1,0} + \beta_{2,0}X_i + U_i, E[U_i|X_i] = 0$$

Esta especificación no requiere un supuesto de homocedasticidad o no autocorrelación, estos supuestos son tipicamente añadidos a un modelo para justificar estimadores particulares.//

El modelo es expresado usando expectativas incondicionales. La ley de expectativas iteradas puede usarse para obtener momentos no condicionales:

$$E[U_i] = E[E[U_i|X_i]] = 0$$

Así los errores de U_i tienen una media no condicional igual a cero como resultado de la especificacion de la regresión, y no por tener un supuesto extra.

$$E[X_iU_i] = E(E[X_iU_i|X_i]) = E(X_iE[U_i|X_i]) = 0$$

La expresión anterior muestra que X_i y U_i estan no correlacionados. Usando $U_i = Y_i - \beta_{1,0} - \beta_{2,0}X_i$, obtenemos dos condiciones de momentos incondicionales.

$$E[Y_i - \beta_{1,0} - \beta_{2,0}X_i] = 0$$

$$E[X_i(Y_i - \beta_{1,0} - \beta_{2,0}X_i)] = 0$$

La expresión anterior también puede representarse $E[m(Y_i, X_i; \beta_0)] = 0$.

6.3. Identificación Exacta de la Estimación GMM

Suponemos que tenemos un vector de momentos kx1 $m(Y_i;\theta)$ y θ es un vector de parámetros px1, cuyos valores verdaderos son θ_0. El modelo estadístico:

$$E[m(Y_i;\theta_0)] = 0$$

La idea de GMM es estimar la expectativa $E[m(Y_i;\theta)]$ por los momentos muestrales:

$$\bar{m}(\theta) = \frac{1}{n}\sum_{i=1}^{n} m(Y_i;\theta)$$

y la elección $\hat{\theta}$ para el valor de θ que hace que $\hat{m}(\theta)$ sea igual a cero o cercana a cero. Es decir la condición de momentos poblacional $E[m(Y_i, \theta_0)] = 0$ son estimados por la condición de momentos muestrales $\frac{1}{n}\sum_{i=1}^{n} m(Y_i;\hat{\theta}) \approx 0$.

Si $k = p$ (el número de condición de momentos es igual al número de parámetros) entonces:

$$\bar{m}(\hat{\theta}) = 0$$

representa un set de k ecuaciones que en conjunto daran solución a $\hat{\theta}$. Además $\hat{\theta}$ puede ser elegido para hacer que $\bar(m)(\hat{\theta})$ sea igual a cero. Cuando k=p, θ_0 es identificado exactamente por las condiciones de momento.

En el caso del modelo de regresión lineal estimado por GMM:

$$\bar{m}(\beta) = \begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n}(Y_i - \beta_1 - \beta_2 X_i) \\ \frac{1}{n}\sum_{i=1}^{n} X_i(Y_i - \beta_1 - \beta_2 X_i) \end{bmatrix}$$

Las ecuaciones normales son las que corresponden a la estimación de MCO de regresionar Y_i sobre 1 y X_i.

6.4. Sobre-Especificación de la Estimación GMM

Si hay mas condiciones de momentos que parámetros a estimar entonces $\bar{m}(\hat{\theta}) = 0$ representa un sistema de ecuaciones con mas ecuaciones desconocidas. En general, de no haber valores $\hat{\theta}$ tales que $\bar{m}(\hat{\theta}) = 0$. Entonces, $\hat{\theta}$ deben ser elegidos para hacer que $\bar{m}(\hat{\theta})$ sea lo mas cercano a cero. La pregunta es, qué hacemos para que el vector $\bar{m}(\hat{\theta})$ sea tan cercano a cero como sea posible. Necesitamos una media de tamaño del vector $\bar{m}(\hat{\theta})$. El punto de inicio a usar será:

$$q(\theta) = \bar{m}(\theta)'\bar{m}(\theta)$$

Entonces elegiremos un $\hat{\theta}$ para minimizar $q(\theta)$. Así

$$\bar{m}(\theta)'\bar{m}(\theta) = \sum_{j=1}^{k} \bar{m}_j(\theta)^2$$

(Consideremos que $\bar{m}_j(\theta)$ es el j-esimo elemento del vector $\bar{m}(\theta)$), de esta manera elegir un $\hat{(\theta)}$ que minimice $q(\theta)$ implicaria elegir un $\hat{\theta}$ para minimizar la suma de cuadrados de la condición de momentos. Esto es la misma idea de minimización de la suma de cuadrados de residuos en un modelo de regresión cuando queriamos que los residuos sean los mínimos posibles.

Continuando con la analogía de una regresión, sabiamos que si las perturbaciones de una regresion tenian varianzas diferentes (heterocedasticas) o estaban correlacionadas (autocorrelación), entonces los parámetros estimados no eran eficientes, por lo que podríamos usar MCG en lugar de MCO. Específicamente, en una regresión $y = X\beta + \mu$ donde el vector de perturbación μ tiene matriz de varianza covarianza Ω, el estimador más eficiente de β se obtenia mediante la minimización del criterio MCG $\hat{\mu}(\beta)'\Omega^{-1}\hat{\mu}(\beta)$ con respecto a β, donde $\hat{\mu}(\beta) = y - X\beta$; más que el criterio MCO.

La misma consideración se aplica en un contexto más general de GMM. Éste se torna asintoticamente eficiente al elegir un $\hat{\theta}$ que minimice:

$$q(\theta) = \bar{m}(\theta)'V^{-1}\bar{m}(\theta)$$

donde V es la matriz de varianza covarianza asintotica de $\bar{m}(\theta)$. Obviamente la matriz V es desconocida en la practica y debe ser estimada con los datos.

Un ejemplo que nos ayudara a complender lo antes dicho:

$$\bar{m}(\theta) = \begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n}(Y_{1,i}-\theta) \\ \frac{1}{n}\sum_{i=1}^{n}(Y_{2,i}-\theta) \end{bmatrix} = \begin{bmatrix} \bar{Y}_1-\theta \\ \bar{Y}_2-\theta \end{bmatrix}$$

en este ejemplo existen dos condiciones de momentos para estimar un solo parametro θ.

La condición a minimizar será:

$$q(\theta) = \bar{m}(\theta)'\bar{m}(\theta) = (\bar{Y}_1-\theta)^2 + (\bar{Y}_2-\theta)^2$$

La condición de primer orden para la minimización de $q(\theta)$ es:

$$\left.\frac{dq(\theta)}{d\theta}\right|_{\theta=\hat{\theta}} = -2(\bar{Y}_1-\hat{\theta}) - 2(\bar{Y}_2-\hat{\theta}) = 0$$

dando $\hat{\theta} = (\bar{Y}_1+\bar{Y}_2)/2$. Así $\hat{\theta}$ es la media de medias grupales. La construcción de este estimador requiere realizar ciertos supuestos sobre la varianza y covarianza de $Y_{1,i}$ y $Y_{2,i}$.

Ahora supongamos que deseamos minimizar con un criterio eficiente GMM $q(\theta) = \bar{m}(\theta)'V^{-1}\bar{m}(\theta)$. Supongamos que tanto $Y_{1,i}$ y $Y_{2,i}$ son i.i.d., con $var(Y_{1,i}) = \sigma_1^2$, $var(Y_{2,i}) = \sigma_2^2$ y $cov(Y_{1,i}, Y_{2,i}) = 0$, donde σ_1^2 y σ_2^2 son tratados como conocidos

$$\begin{aligned} V &= \lim_{x\to+\infty} var(\sqrt{n}\bar{m}(\theta)) \\ &= \lim_{x\to+\infty} \begin{bmatrix} var(\sqrt{n}\bar{m}_1(\theta_0)) & cov(\sqrt{n}\bar{m}_1(\theta_0), \sqrt{n}\bar{m}_2(\theta_0)) \\ cov(\sqrt{n}\bar{m}_1(\theta_0), \sqrt{n}\bar{m}_2(\theta_0)) & var(\sqrt{n}\bar{m}_2(\theta_0)) \end{bmatrix} \\ &= \lim_{x\to+\infty} \begin{bmatrix} var(\frac{1}{n}(\sum_{i=1}^{n}(Y_{1,i}-\theta_0))) & cov(\frac{1}{n}(\sum_{i=1}^{n}(Y_{1,i}-\theta_0)), \frac{1}{n}(\sum_{i=1}^{n}(Y_{2,i}- \\ cov(\frac{1}{n}(\sum_{i=1}^{n}(Y_{1,i}-\theta_0)), \frac{1}{n}(\sum_{i=1}^{n}(Y_{2,i}-\theta_0))) & var(\frac{1}{n}(\sum_{i=1}^{n}(Y_{2,i}-\theta_0))) \end{bmatrix} \\ &= \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix} \end{aligned}$$

así:

$$q(\theta) = \bar{m}(\theta)'V^{-1}\bar{m}(\theta) = [(\bar{Y}_1-\theta)(\bar{Y}_2-\theta)] \begin{bmatrix} 1/\sigma_1^2 & 0 \\ 0 & 1/\sigma_2^2 \end{bmatrix} \begin{bmatrix} (\bar{Y}_1-\theta) \\ (\bar{Y}_2-\theta) \end{bmatrix}$$

$$q(\theta) = \frac{(\bar{Y}_1-\theta)^2}{\sigma_1^2} + \frac{(\bar{Y}_2-\theta)^2}{\sigma_2^2}$$

minimizando $q(\theta)$ respecto a θ nos da:

$$\hat{\theta} = \left(\frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}\right)\bar{Y}_1 + \left(\frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}\right)\bar{Y}_2$$

Ahora $\hat{\theta}$ es un promedio ponderado de $\bar{Y}_1$ y $\bar{Y}_2$ con ponderadores que son dados de acuerdo a las variaciones dentro de la muestra, de esta manera la muestra con mayor información sobre el comportamiento de la media θ_0 tendrá mayor ponderación.

Considerando las condición de primer orden para minimizar:

$$\left.\frac{\partial q(\theta)}{\partial \theta}\right|_{\theta=\hat{\theta}} = 2\left(\frac{\partial \bar{m}(\theta)'}{\partial \theta'}\right)' V^{-1}\bar{m}(\theta)\Big|_{\theta=\hat{\theta}} = 0$$

Ésta condición de primer orden presenta p ecuaciones para resolver y estimar p parámetros (θ), entonces la ecuación quedaría registrada como:

$$\left(\frac{\partial \bar{m}(\theta)'}{\partial \theta'}\right)' V^{-1}\bar{m}(\theta) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\partial \bar{(m)}(\theta)}{\partial \theta'}\right)' V^{-1}m(Y_i;\theta) = \frac{1}{n}\sum_{i=1}^{n} m^*(Y_i;\theta)$$

donde $m^*(Y_i;\theta) = \left(\frac{\partial \bar{m}(\theta)}{\partial \theta'}\right)' V^{-1}m(Y_i;\theta)$, una interpretación alternativa de las condiciones de primer orden es que ellas muestran como construir el óptimo reduciendo un conjunto de condiciones de momentos $m^*(Y_i;\theta)$ a una exacta identificación de θ_0.

6.5. Test de Sobreidentificación

El test para la especificación del modelo GMM es viable en el caso de sobreidentificación. El estimador $\hat{\theta}$ es definido de tal manera que minimizan:

$$q(\theta) = \bar{m}(\theta)'V^{-1}\bar{m}(\theta)$$

El estimador a evaluar será el estadístico J.

$$J = nq(\hat{\theta})$$

la cual se distribuye asintoticamente como una χ^2_{k-p} si la condición de momento es correctamente especificada. La hipótesis que contrasta es:

$$H_0 : E[m(Y_i;\theta_0)] = 0$$

(momento correctamente especificado)

$$H_0 : E[m(Y_i; \theta_0)] \neq 0$$

(momento incorrectamente especificado)

Si se rechaza H_0 el estadístico J es más grande que el valor crítico de la distribución χ^2_{k-p}. Note que este test no puede ser usado en el caso de exacta identificación.

Veamos una aplicación mediante simulación:

```
clear all
set memo 400m
set matsize 500
set seed 1245

*SE GENERAN 1000 OBSERVACIONES PARA 4 V.A. xaster, u, v1 y v2
*NO OBSERVABLES, NORMALES CON DISTINTAS MEDIAS Y DESVIOS

drawnorm xaster, n(1000) means(10) sds(2)
drawnorm u, n(1000) means(0) sds(1)
drawnorm v1, n(1000) means(0) sds(1)
drawnorm v2, n(1000) means(0) sds(2)

sum

g y = 3*xaster+u
g x1=xaster+v1
g x2=xaster+v2

sum y
* EJEMPLO 1 MEDIA DE Y
global u "y-{ymedia}"
gmm ($u) , winitial(identity)

* EJEMPLO 2 REGRESION MCO Y SOBRE XASTER
reg y xaster
reg y xaster, r

global u "y-{b1}*xaster-{b0}"
gmm ($u) , winitial(identity)  instruments(xaster)

* EJEMPLO 3 REGRESION MCO Y SOBRE X1 y X2,
```

```
reg y x1
reg y x2

* VARIABLES INSTRUMENTALES

* VI USANDO X1 COMO REGRESOR Y X2 COMO INSTRUMENTO
ivreg y ( x1 = x2)
ivendog

global u1 "y-{b1}*x1-{b0}"
gmm ( $u1 ) , instruments (x2) winitial(identity) one
gmm ( $u1 ) , instruments (x2) winitial(identity) two
gmm ( $u1 ) , instruments (x2) winitial(identity) i

* VI USANDO X2 COMO REGRESOR Y X1 COMO INSTRUMENTO
ivreg y ( x2 = x1)

global u2 "y-{b1}*x2-{b0}"
gmm ( $u2 ), instruments (x1) winitial(identity) one
global u2 "y-{b1}*x2-{b0}"
gmm ( $u2 ), instruments (x1) winitial(identity) two
global u2 "y-{b1}*x2-{b0}"
gmm ( $u2 ), instruments (x1) winitial(identity) i
```

Sesión 7
Modelos Panel

7.1. Introducción a la Estimación de los Modelos de Datos Panel

7.1.1. Preparando la base de datos

```
use nlswork1.dta, clear
```

Un aspecto inicial que se debe tener en cuenta para estimar modelos de datos de panel con Stata, es la forma en la que se encuentra ordenada la base de datos. Por ejemplo, si se cuenta con una base de datos que contenga información ordenada de la siguiente manera (forma ancha o wide form):

i		x_{ij}					
id	sexo	ing2001	ing2002	ing2003	egre2001	egre2002	egre2003
1	1	250	300	450	240	286	358
2	0	215	256	352	200	230	341
3	1	456	460	465	300	310	330

Figura 7.1:

Para estimar modelos de datos de panel en Stata, es necesario tener la información ordenada de forma larga (long form) tal como se muestra en el cuadro siguiente:

Esto es posible haciendo uso del comando reshape en nuestro ejercicio:

```
reshape wide birth_yr age race msp nev_mar grade collgrad not_smsa \\\
city south ind_code occ_code union wks_ue ttl_exp tenure hours \\\
wks_work ln_wage, i(idcode) j(year)

reshape long birth_yr age race msp nev_mar grade collgrad not_smsa \\\
c_city south ind_code occ_code union wks_ue ttl_exp tenure hours \\\
wks_work ln_wage, i(idcode) j(year 68 69 70 71 72 73 74 75 76 77 \\\
78 79 80 81 82 83 84 85 86 87 88)
```

i		j	x_{ij}	
id	sexo	año	ing	egre
1	1	2001	250	240
1	1	2002	300	286
1	1	2003	450	358
2	0	2001	215	200
2	0	2002	256	230
2	0	2003	352	341
3	1	2001	456	300
3	1	2002	460	310
3	1	2003	465	330

Figura 7.2:

La base de datos a usar es nlswork1.dta , la cual contiene información de una muestra de datos de panel para 4,711 mujeres empleadas, que han completado su educación y con salarios mayores a US$1 por hora pero menores a $700, para un período de 20 años (1968-1988) en los Estados Unidos.

A través del comando describe podemos observar todas las variables que contiene la base de datos nlswork1.dta Antes de estimar un modelo de datos de panel, se deben identificar las variables que representan a los individuos y a las observaciones.

```
iis idcode
tis year

generate age2=age^2
generate ttl_exp2=ttl_exp^2
generate tenure2=tenure^2
generate byte black=race==2
```

7.1.2. Estimando mi Primer Panel

Veamos unos cuantos comandos de estimación.

```
xtreg  ln_wage grade  age* ttl_exp* tenure* black \\\
not_smsa south, be
```

```
xtreg  ln_wage grade  age* ttl_exp* tenure* black \\\
not_smsa south, fe

xtreg  ln_wage grade  age* ttl_exp* tenure* black \\\
not_smsa south, re theta

xtreg  ln_wage grade  age* ttl_exp* tenure* black \\\
not_smsa south, mle

xthausman
```

7.2. Diagnostico y Especificación de los Modelos Panel

Utilizaremos la base panelusa50-89.dta para estimar el impacto de las variables políticas y sociodemográficas en el nivel de gasto estatal (spend) en los Estados Unidos durante el periodo 1950-1989.

7.2.1. Controlando la Heterogeneidad dentro de un Panel

Regresión Agrupada - Pool

El enfoque más simple de analizar datos tipo panel es omitir las dimensiones del espacio y el tiempo de los datos agrupados y sólo calcular la regresión MCO usual. Este modelo se expresa como:

$$Y_{it} = \alpha + \beta * X_{it} + e_{it}$$

Donde i significa la i-ésima unidad transversal (estado) y t el tiempo t (año). Si tratamos de explicar la variable spend con las variables independientes de la clase pasada, basta con que indiquemos en la ventana de comandos de Stata:

```
reg spend dem* divgov dis1 persinc* aper* popul*
```

Efectos Aleatorios

La ecuación (1) supone que el intercepto de la regresión es la misma para todas las unidades transversales. Sin embargo, es muy probable que necesitemos controlar el carácter "individual" de cada estado. El modelo de efectos aleatorios permite suponer que cada unidad transversal tiene un intercepto diferente. Este modelo se expresa como:

$$Y_{it} = \alpha_i + \beta * X_{it} + e_{it}$$

Donde $\alpha_i = \alpha + \mu_i$. Es decir, en vez de considerar a α como fija, suponemos que es una variable aleatoria con un valor medio α y una desviación aleatoria μ_i de este valor medio. Sustituyendo $\alpha_i = \alpha + \mu_i$ en la ecuación anterior obtenemos:

$$Y_{it} = \alpha + \beta_1 X_{1it} + \mu_i + e_{it}$$

Stata estima el modelo de efectos aleatorios con el comando xtreg, re. En nuestro ejemplo, indicamos en la ventana de comandos:

```
xtreg spend dem* divgov dis1 persinc* aper* popul*, re
```

Analicemos la ecuación anterior, y observaremos que si la varianza de μ_i es igual a cero, es decir $\sigma_u^2 = 0$, entonces no existe ninguna diferencia relevante entre un Pool y EA. ¿Cómo podemos saber si es necesario usar el modelo de efectos aleatorios o el de datos agrupados? Breusch y Pagan formularon la prueba conocida como Prueba del Multiplicador de Lagrange para Efectos Aleatorios. La hipótesis nula de esta prueba es que $\sigma_u^2 = 0$. Si la prueba se rechaza, sí existe diferencia entre un Pool y un modelo de EA, y es preferible usar el método de efectos aleatorios. La prueba de Breusch y Pagan se implementa en Stata con el comando "xttest0" después de la estimación de efectos aleatorios.

```
xtreg spend dem* divgov dis1 persinc* aper* popul*, re
xttest0
```

```
Breusch and Pagan Lagrangian multiplier test for random effects:

        spend[stcode,t] = Xb + u[stcode] + e[stcode,t]

        Estimated results:
                         |       Var     sd = sqrt(Var)
                ---------+-----------------------------
                   spend |   395200.4       628.6496
                       e |   33364.55       182.6597
                       u |   116856.9       341.8434

        Test:   Var(u) = 0
                             chi2(1) =  6960.99
                         Prob > chi2 =     0.0000
```

Figura 7.3:

El p-value nos indica que podemos rechazar la Ho; por lo tanto, los efectos aleatorios μ_i son relevantes y es preferible usar la estimación de efectos aleatorios en vez de la agrupada.

Efectos Fijos

Otra manera de modelar el carácter "individual" de cada estado es a través del modelo de efectos fijos. Este modelo no supone que las diferencias entre estados sean aleatorias, sino constantes o "fijas", y por ello debemos estimar cada intercepto μ_i . ¿Cómo podemos permitir que el intercepto varíe con respecto a cada estado? Una manera es la técnica de "las variables dicotómicas de intersección diferencial", que se expresa de la siguiente manera:

$$Y_{it} = v_i + \beta * X_{it} + e_{it}$$

Donde v_i es un vector de variables dicotómicas para cada estado. El modelo de efectos fijos puede ejecutarse en Stata con el comando:

```
xi: reg spend dem* divgov dis1 persinc* aper* popul* i.stcode
```

El cual estima una dummy para cada estado. Una opción más sencilla es el comando xtreg:

```
xtreg spend dem* divgov dis1 persinc* aper* popul*, fe
```

¿Cuál de los modelos, el Pool o el de EF es mejor? En relación con el modelo de EF, el Pool es un modelo restringido, pues asume un intercepto común para todos los estados (es decir, no incluye variables dicotómicas estatales). Por lo tanto, podemos utilizar una prueba F restrictiva para contestar la cuestión. La hipótesis nula es que $v_1 = v_2 = v_3 = ... = v_i = 0$ (o sea, que todas las variables dicotómicas estatales son iguales cero). Si la prueba se rechaza, significa que al menos algunas variables dicotómicas sí pertenecen al modelo, y por lo tanto es necesario utilizar el método de efectos fijos. La prueba F de significancia de los efectos fijos se reporta automáticamente con el comando xtreg, fe. (La prueba aparece al final de la estimación de EF)

El p-value nos indica que podemos rechazar la Ho, por lo que es preferible usar el método de efectos fijos al modelo agrupado.

Efectos Fijos vs. Efectos Aleatorios

Las pruebas de Breusch y Pagan para efectos aleatorios, y la prueba F de significancia de los efectos fijos nos indican que tanto el modelo de efectos aleatorios como el de efectos fijos son mejores que el modelo agrupado. ¿Pero cómo decidir cuál de los dos usar? La respuesta depende de la posible correlación entre el componente de error individual μ_i y las variables X.

El modelo de efectos aleatorios supone que esta correlación es igual a cero. Pero supongamos que en nuestro ejemplo, μ_i representa las reglas electorales estatales que favorecen a cierto partido (por ejemplo, gerrymandering); entonces es muy probable que μ_i se correlacione con las variables partidarias de nuestro modelo. Si las μ_i y las variables X están correlacionadas, entonces no incluir μ_i en el modelo producirá un sesgo de variable omitida en los coeficientes de X. Hausman demostró que la diferencia entre los coeficientes de efectos fijos y aleatorios $(\beta_{ef} - \beta_{ea})$ pude ser usada para probar la hipótesis nula de que μ_i y las variables X no están correlacionadas. Así pues, la Ho de la prueba de Hausman es que los estimadores de efectos aleatorios y de efectos fijos no difieren sustancialmente. Si se rechaza la Ho, los estimadores sí difieren, y la conclusión es; que efectos fijos es más conveniente que efectos aleatorios. Si no podemos rechazar Ho, no hay sesgo de qué preocuparnos y preferimos efectos aleatorios que, al no estimar tantas dummies, es un modelo más eficiente. La prueba de Hausman se implementa en Stata después de la regresión con efectos aleatorios con el comando xthausman:

```
xtreg spend dem* divgov dis1 persinc* aper* popul*, re
xthausman
```

Hausman specification test

spend	Coefficients Fixed Effects	Random Effects	Difference
dem1	126.4938	126.2369	.2569082
demmaj1	-3.50604	-.828879	-2.677161
demgov	34.40021	33.38822	1.011994
divgov	-20.52315	-22.1351	1.611952
dis1	-79.9025	-84.36639	4.463887
persinc	.1685387	.1727579	-.0042192
persinc2	-1.00e-06	-9.72e-07	-3.04e-08
aper5_17	-105.8086	-232.866	127.0574
aper65	3140.738	1936.944	1203.794
popul	-113.9934	-107.8277	-6.165682
popul2	3.038663	2.898781	.139882

```
Test:  Ho:  difference in coefficients not systematic

            chi2( 11) = (b-B)'[S^(-1)](b-B), S = (S_fe - S_re)
                      =     268.64
            Prob>chi2 =       0.0000
```

Figura 7.4:

En nuestro ejemplo, la Ho se rechaza; es decir, la diferencia entre los coeficientes de efectos aleatorios y fijos sí es sistemática. Por lo tanto, conviene usar el método de efectos fijos.

Efectos Temporales (Two-Way Fixed Effects)

La incorporación de variables dicotómicas estatales permite modelar características de las unidades transversales (estados) que no cambian en el tiempo pero que sí afectan el resultado de interés. Ahora bien, también es posible agregar variables dicotómicas temporales a nuestro modelo, es decir, una para cada año en la muestra, que capturen eventos comunes a todos los estados durante un período u otro, como una gran depresión o guerra mundial. Agregando efectos temporales, la ecuación de EF se transforma en:

$$Y_{it} = v_i + \eta_t + \beta * X_{it} + e_{it}$$

Donde η_t representa un vector de variables dicotómicas para cada año. Estas variables dicotómicas permitirán controlar por aquellos eventos a los que fueron sujetos todos los estados en un año dado y, al igual que los efectos fijos, pueden reducir sesgos importantes. En Stata podemos incorporar efectos temporales a nuestro modelo de efectos fijos con el comando xi.

```
xi: xtreg spend dem1 demmaj1 demgov divgov dis1 persinc* \\\
aper* popul* i.year, fe
```

O bien, generando tanto las dummies de estado como de año,

```
xi: reg spend dem1 demmaj1 demgov divgov dis1 persinc* aper* \\
popul* i.stcode i.year
```

Al igual que con los efectos estatales, podemos realizar una prueba F para conocer la significancia conjunta de las variables dicotómicas temporales en nuestro modelo. La hipótesis nula es que $\eta_1 = \eta_2 = ... = \eta_t = 0$. En nuestro ejemplo, luego de estimar un modelo con efectos fijos estatales y temporales, indicamos en la ventana de comando:

```
testparm  _Iyear_1951 - _Iyear_1989
```

El p-value de la prueba F nos indica que rechazamos la Ho, por lo que es posible afirmar que las variables dicotómicas temporales son conjuntamente significativas y pertenecen al modelo.

Autocorrelación

Es importante señalar que aún cuando hemos modelado la heterogeneidad temporal y espacial en nuestro modelo, la ecuación anterior puede estar mal especificada en otros aspectos. Recordemos que de acuerdo con los supuestos de Gauss-Markov, los estimadores de OLS son los Mejores Estimadores Lineales Insesgados (MELI) siempre y cuando los errores e_{it} sean independientes entre si y se distribuyan idénticamente con varianza constante σ^2. Desafortunadamente, con frecuencia estas condiciones son violadas en datos panel: la independencia se viola cuando los errores de diferentes unidades están correlacionados (correlación contemporánea), o cuando los errores dentro de cada unidad se correlacionan temporalmente (correlación serial), o ambos. A su vez, la distribución "idéntica" de los errores es violada cuando la varianza no es constante (heterocedasticidad). En esta sección abordaremos al problema de la correlación serial o "autocorrelación"; es decir, cuando los errores e_{it} no son independientes con respecto al tiempo. En nuestro ejemplo, es muy probable que el nivel de gasto en t esté asociado con el nivel de gasto en t-1.

Existen muchas maneras de diagnosticar problemas de autocorrelación. Sin embargo, cada una de estas pruebas funciona bajos ciertos supuestos sobre la naturaleza de los efectos individuales. Wooldridge desarrolló una prueba muy flexible basada en supuestos mínimos que puede ejecutarse en Stata con el comando xtserial. La hipótesis nula de esta prueba es que no existe autocorrelación; naturalmente, si se rechaza, podemos concluir que ésta sí existe. El comando xtserial requiere que se especifiquen la variable dependiente e independientes de nuestro modelo. En nuestro ejemplo, indicamos:

```
xtserial spend dem1 demmaj1 demgov divgov dis1 persinc* \\\
aper* popul*, output
```

```
                                                  F( 11,    48) =    18.49
                                                  Prob > F      =   0.0000
                                                  R-squared     =   0.0920
Number of clusters (stcode) = 49                  Root MSE      =   91.779

------------------------------------------------------------------------------
             |               Robust
     D.spend |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        dem1 |
         D1. |    32.6159   26.11535     1.25   0.218    -19.89252    85.12433
     demmaj1 |
         D1. |   -1.61453   9.966516    -0.16   0.872    -21.65355    18.42449
      demgov |
         D1. |  -2.627346   6.331616    -0.41   0.680    -15.35791    10.10322
      divgov |
         D1. |  -.9898086   10.24538    -0.10   0.923    -21.58953    19.60991
        dis1 |
         D1. |  -53.24134   52.62154    -1.01   0.317     -159.044    52.56137
     persinc |
         D1. |   -.046158   .0512333    -0.90   0.372    -.1491694    .0568534
    persinc2 |
         D1. |   4.05e-06   2.73e-06     1.48   0.145    -1.45e-06    9.55e-06
    aper5_17 |
         D1. |  -952.4689   469.8396    -2.03   0.048    -1897.145   -7.793091
      aper65 |
         D1. |   1944.257   594.4707     3.27   0.002     748.9937     3139.52
       popul |
         D1. |   202.7253   43.83776     4.62   0.000     114.5836     290.867
      popul2 |
         D1. |  -4.075361   1.101919    -3.70   0.001    -6.290918   -1.859803
------------------------------------------------------------------------------

Wooldridge test for autocorrelation in panel data
H0: no first-order autocorrelation
    F(  1,      48) =     87.734
           Prob > F =      0.0000
```

Figura 7.5:

La prueba nos indica que tenemos un problema de autocorrelación que es necesario corregir. Una manera de hacerlo es a través de un modelo de efectos fijos con término (ρ) autorregresivo de grado 1 (AR1) que controla por la dependencia de t con respecto a t-1. El modelo AR1 con efectos fijos se especifica de la manera:

$$Y_{it} = v_i + \beta * X_{1it} + e_{it}$$

donde $e_{it} = \rho e_{i,t-1} + \eta_{it}$, es decir, los errores tienen una correlación de primer grado, ρ. El modelo AR1 es fácilmente ejecutable en Stata con el comando xtregar:

```
xtregar spend dem* divgov dis1 persinc* aper* popul*, fe
```

Heterocedasticidad

Cuando la varianza de los errores de cada unidad transversal no es constante, nos encontramos con una violación de los supuestos Gauss-Markov. Una forma de saber si nuestra estimación tiene problemas de heterocedastidad es a través de la prueba del Multiplicador de Lagrange de Breusch y Pagan. Sin embargo, de acuerdo con Greene, ésta y otras pruebas son sensibles al supuesto sobre la normalidad de los errores; afortunadamente, la prueba Modificada de Wald para Heterocedasticidad funciona aún cuando dicho supuesto es violado. La hipótesis nula de esta prueba es que no existe problema de heteroscedasticidad, es decir, $\sigma_i^2 = \sigma^2$ para toda $i = 1...N$, donde N es el número de unidades transversales ("estados" en nuestro ejemplo). Naturalmente, cuando la Ho se rechaza, tenemos un problema de heteroscedasticidad. Esta prueba puede implementarse en Stata con el comando xttest3 después de estimar el modelo de efectos fijos:

```
xtreg spend dem* divgov dis1 persinc* aper* popul*, fe
xttest3
```

La prueba nos indica que rechazamos la Ho de varianza constante y aceptamos la Ha de heteroscedasticidad. Antes de abordar cómo solucionar nuestro problema de heteroscedastidad, resulta conveniente analizar otro problema que surge de la estimación con datos tipo panel.

Correlación Contemporánea

Las estimaciones en datos panel pueden tener problemas de correlación contemporánea si las observaciones de ciertas unidades están correlacionadas con las observaciones de otras unidades en el mismo periodo de tiempo. Como discutimos en la sección sobre heterogeneidad, las variables dicotómicas de efectos temporales se incorporan al modelo para controlar por los eventos que afectan por igual a todas las unidades (estados) en un año dado. La correlación contemporánea es similar, pero con la posibilidad de algunas unidades estén más o menos correlacionadas que otras. El problema de correlación contemporánea se refiere a la correlación de los errores de al menos dos o más unidades en el mismo tiempo t. En otras palabras, tenemos errores contemporáneamente correlacionados si existen características inobservables de ciertas unidades que se relacionan con las características inobservables de otras unidades. Por ejemplo, los errores de dos estados pueden relacionarse pero mantenerse independientes de los errores de los demás estados. En nuestro ejemplo, una fuerte helada podría afectar a los estados agrícolas, disminuyendo la producción y por tanto el ingreso (que se asocia con nuestra variable dependiente spend). Pero este efecto probablemente no se manifieste en los estados no agrícolas.

El comando xttest2 de Stata ejecuta la prueba de Breusch y Pagan para identificar problemas de correlación contemporánea en los residuales de un modelo de efectos fijos. La hipótesis nula es que existe "independencia transversal" (cross-sectional independence); es decir, que los errores entre las unidades son independientes entre sí. Si la Ho se rechaza, entonces existe un problema de correlación contemporánea. El comando xttest2 se implementa después de un modelo de efectos fijos. En nuestro ejemplo:

```
xtreg spend dem* divgov dis1 persinc* aper* popul*, fe
xttest2
```

Si el p-value del estadístico χ^2 indica que podemos rechazar la Ho; entonces, también sera necesario corregir el problema de correlación contemporánea.

Solución a la Autocorrelación, Heterocedasticidad y Autocorrelación Contemporanea

Los problemas de correlación contemporánea, heteroscedasticidad y autocorrelación que hemos examinado pueden solucionarse conjuntamente con estimadores de Mínimos Cuadrados Generalizados Factibles (Feasible Generalizad Least Squares ó FGLS), o bien con Errores Estándar Corregidos para Panel (Panel Corrected Standard Errors ó PCSE). Beck y Katz (What to do (and not to do) with time-series cross-section data, 1995) demostraron que los errores estándar de PCSE son más precisos que los de FGLS. Desde entonces, muchos trabajos en la disciplina han utilizado PCSE en sus estimaciones para panel.

Stata ejecuta FGLS y PCSE con los comandos xtgls y xtpcse. Las opciones que ofrecen estos comandos dependen de los problemas detectados en las pruebas que hemos revisado. La Tabla "Resumen" presenta los comandos que puedes ejecutar cuando te enfrentes con problemas de correlación contemporánea, heteroscedasticidad y autocorrelación, y sus combinaciones. La Tabla "Resumen" se lee de la siguiente manera: si tus pruebas sólo detectaron problemas de heteroscedasticidad, entonces el comando es $xtgls[...]$, $p(h)xtpcse[...], het$. Si tienes problemas de heterocedasticidad y correlación contemporánea, los comandos son $xtgls[...], p(c)$ ó $xtpcse[...]$, etc. Estos comandos no calculan automáticamente efectos fijos, por lo que en caso de ser necesario, tendrás que introducir variables dicotómicas con el comando xi.

En nuestro ejemplo sobre el gasto estatal, hemos detectado problemas de heterogeneidad, heteroscedasticidad, correlación contemporánea y autocorrelación. Para corregir estos problemas ejecutamos el comando:

```
xi: xtgls spend dem1 demmaj1 demgov divgov dis1 persinc* aper* \\\
popul* i.stcode i.year, panels (correlated) correlation(ar1)
```

ó el comando:

```
xi: xtpcse spend dem1 demmaj1 demgov divgov dis1 persinc* aper* \\\
popul*  i.stcode i.year, correlation(ar1)
```

RESUMEN

	Heterocedasticidad	Correlación contemporánea	Autocorrelación
Heterocedasticidad	xtgls (VAR DEP)(VAR IND), p(h) xtpcse (VAR DEP)(VAR IND), het	xtgls (VAR DEP)(VAR IND),p(c) xtpcse (VAR DEP)(VAR IND)	xtgls (VAR DEP)(VAR IND), p(h)c(ar1) xtpcse (VAR DEP)(VAR IND), het c(ar1)
Correlación contemporánea	xtgls (VAR DEP)(VAR IND), p(c) xtpcse (VAR DEP)(VAR IND)	-	-
Autocorrelación	xtgls (VAR DEP)(VAR IND), p(h) c(ar1) xtpcse (VAR DEP)(VAR IND), het c(ar1)	-	xtregar (VAR DEP)(VAR IND),fe ó re

Heterocedasticidad, Correlación contemporánea y Autocorrelación:	xtgls (VAR DEP)(VAR IND),p(c) c(ar1) xtpcse (VAR DEP)(VAR IND), c(ar1)

Figura 7.6: Cuadro Resumen

RETO 3

a. En el modelo de crecimiento aleatorio:

$$Y_{it} = \alpha_i + g_i t + x_{it} * \beta + \epsilon_{it}$$

A este modelo también se le conoce como el modelo de tendencia aleatoria, es decir, cada firma, ciudad o persona tiene su propia tendencia. Esta tendencia es una fuente adicional de heterogeneidad. Si y_{it} es el logaritmo natural de una variable, entonces g_i sería la tasa de crecimiento promedio en el horizonte temporal evaluado, de ahí que se le conoce también como el modelo de crecimiento aleatorio. En muchas aplicaciones se asume que $(\alpha_i * g_i)$ están altamente correlacionados con las variables explicativas x_{it}.

Aplicar este modelo de crecimiento aleatorio a la base de datos JTRAIN1.DTA El objetivo es estimar el efecto de otorgar becas de estudio al trabajador en relación a su nivel de productividad medido inversamente por el ratio de desperdicios (scrap). [1]

$$Ln(scrap_{it}) = \alpha_i + g_i t + \beta_1 * grant_{it} + \beta_2 * grant_{it-1} + \epsilon_{it}$$

- Estime el modelo propuesto por efectos fijos, sustente su elección ante un modelo de efectos aleatorios y ante un modelo de efectos agregados y analice los resultados.
- Analizar la autocorrelación, heterocedasticidad y autocorrelación contemporanea del panel.
- Aplicar el modelo de efectos fijos al modelo en primeras diferencias. Discuta los resultados.

[1] La variable t es compuesta por una variable dummy d88 y d89.

Sesión 8
Panel Dinámico

Los datos de panel dinámicos constituyen hoy en día un terreno econométrico de basta extensión pero ampliamente explorado; los desarrollos en los últimos años han sido muchos y han permitido sistematizar los procesos de estimación e inferencia. En especial, se dedicará mayor detalle a la técnica de estimación del Método Generalizado de Momentos, aproximación que, puede considerarse como la más completa de las disponibles hasta a fecha.

8.1. Heterogeneidad de los paneles de datos

Desde los trabajos iniciales de Balestra y Nerlove (1966), los modelos dinámicos han jugado un importante papel en el análisis empírico con datos de panel en economía. Dada la escasa dimensión temporal exhibida por la mayor parte de paneles tradicionalmente disponibles, el énfasis se ha puesto en modelos con dinámica homogénea, dejándose relativamente al margen, hasta hace bien poco, al análisis de paneles dinámicos heterogéneos. Sin embargo, desde hace una década, han ido apareciendo un buen número de conjuntos de datos de panel con amplia cobertura de empresas, regiones y países y un número relativamente largo de observaciones temporales. La disponibilidad de estos seudo - paneles ha elevado el interés por analizar la conveniencia de esa homogeneidad en la dinámica supuesta en el análisis tradicional de datos de panel, al tiempo que ha permitido centrar los esfuerzos de análisis en la dimensión temporal de los paneles y su tratamiento.

8.2. Estimación intragrupo de modelos dinámicos de datos de panel

Las regresiones dinámicas de datos de panel presentan dos fuentes de persistencia a lo largo del tiempo: la autocorrelación debida a la presencia de la endógena retardada entre los regresores y la debida a los efectos individuales que resumen la heterogeneidad entre individuos.

La especificación más sencilla con la que representar un modelo dinámico de datos de panel es la siguiente:

$$Y_{it} = \mu_i + \alpha * Y_{it-1} + \epsilon_{it}$$

En el análisis convencional dinámico de datos de panel micro, en el que se cuenta con observaciones de un número elevado de empresas, sectores o individuos a lo largo de un breve espacio temporal son bien conocidos los problemas derivados de la utilización de procedimientos clásicos para paneles estáticos como el estimador intragrupos (IG): inconsistencia y sesgo asintótico.

Efectivamente, Nickell (1981) derivó la expresión exacta de ese sesgo para el caso general de el modelo autorregresivo de orden uno sin exógenas representado más arriba confirmando los resultados experimentales de Monte Carlo obtenidos previamente por Nerlove (1967) para el caso de un modelo sin exógenas, y por Maddala (1971) para el caso de un modelo completo. El sesgo, en el caso en que utilizásemos sólo una sección transversal para la estimación por MCO en desviaciones a la media, es una compleja función del tamaño muestral T y el verdadero valor del parámetro autorregresivo α:

$$plim(\hat{\alpha} - \alpha) = \frac{-(1-\alpha)}{T-1}\left[1 - \alpha^{t-1} - \alpha^{T-t} + \frac{(1-\alpha^T)}{T(1-\alpha)}\right] * \Phi^{-1}$$

$$\Phi = \left[1 - \frac{2\alpha}{(T-1)(1-\alpha)}\left(1 - \alpha^{t-1} - \alpha^{T-t} + \frac{(1-\alpha^T)}{T(1-\alpha)}\right)\right]$$

Esta expresión permite observar, en primer lugar, que para todo $alpha > 0$ el sesgo es negativo, en segundo lugar, que ese sesgo depende (y por tanto varía) con el corte transversal t elegido, siendo menor para los cortes situados en los extremos del intervalo muestral que para aquellos situados en el medio de la muestra.

En el caso en que utilizásemos la muestra completa para la estimación por MCO del modelo en diferencias con respecto a la media, la expresión de este sesgo toma la forma:

$$plim(\hat{\alpha}-\alpha)=\frac{-(1-\alpha)}{T-1}\left[1-\frac{(1-\alpha^T)}{T(1-\alpha)}\right]\left[1-\frac{2\alpha}{(T-1)(1-\alpha)}\left(1-\frac{(1-\alpha^T)}{T(1-\alpha)}\right)\right]^{-1}$$

expresión que, para valores extremos de T, puede simplificarse de forma notable:

$$plim(\hat{\alpha}-\alpha)=\frac{-(1-\alpha)}{T-1}$$

para valores de T razonablemente elevados.

La única ventaja del procedimiento de estimación intragrupo es que el sesgo no depende del componente de heterogeneidad transversal μ_i mientras que, en el caso de la estimación del modelo por MCO en niveles, el sesgo sería:

$$plim(\hat{\alpha}-\alpha)=\frac{\lambda}{\lambda(1-\alpha)+(1+\alpha)}$$

donde $\lambda=\sigma_u^2/\sigma^2$, , haciendo evidente que la estimación depende de σ_u^2 (la dispersión de los efectos μ_i en la población).

Gracias a estas expresiones puede observarse con sencillez cómo, en la mayor parte de los paneles micro ampliamente utilizados en la realidad, el tratamiento de la heterogeneidad inobservable por la vía tradicional provocará un sesgo muy importante que impedirá cualquier tipo de inferencia sobre el parámetro autorregresivo α. Por ejemplo, para el caso en que $\alpha=0{,}25$ y T=10 (y nótese que T=10 es un tamaño aceptable en un panel micro), el sesgo alcanzaría un valor en torno a 0.21, esto es, más de un 80 % del verdadero valor del parámetro.

Por último, si se considera además la presencia de variables exógenas Xijt, los resultados señalan que el sesgo en valor absoluto en la estimación del parámetro autorregresivo α será todavía mayor que en el caso en que las variables X_{ijt} se omitan. Así mismo, la estimación del parámetro o vector de parámetros b será también sesgada, siendo ese sesgo tanto más amplio cuanto mayor sea la relación entre las variables exógenas X_{ijt} y el retardo de la endógena y_{t-1}.

Recientemente, analizando las propiedades del estimador tradicional de efectos fijos en el contexto de un modelo dinámico simple, Kiviet (1995 y 1999), consciente de que las propiedades asintóticas de primer orden no conducían a una inferencia correcta en muestras pequeñas, examinó contextos asintóticos de orden superior con la esperanza de que pudieran aproximar mejor las propiedades en muestras pequeñas conduciendo a una mejora en los ejercicios de inferencia. Kiviet consideró un modelo dinámico simple sin autocorrelación residual y con exogeneidad estricta en los regresores y derivó el tamaño del sesgo para el estimador de efectos fijos. Una vez que se sustrae de este estimador de efectos fijos otro estimador consistente del sesgo, resulta un nuevo estimador corregido que parece funcionar relativamente bien si se compara con algunas de las alternativas más tradicionales, como el Método Generalizado de Momentos, que será resumido más adelante.

Por otro lado, otros estudios también recientes, como los de Judson y Owen (1999), invitan a seguir utilizando el estimador de efectos fijos en paneles en los que la dimensión temporal no sea muy pequeña con relación a la transversal, argumentando que el sesgo, en este caso, no habría de ser considerable. Los experimentos de Monte Carlo en este sentido sugieren que, incluso con un número aproximado de 30 observaciones temporales, el sesgo del estimador de efectos fijos no superaría, en el peor de los casos (es decir, en presencia de un parámetro autorregresivo elevado) el 20 % del verdadero valor del parámetro. Los experimentos de estos dos autores sirvieron para recomendar, como mejor alternativa, la modificación de Kiviet seguido del estimador de Método Generalizado de Momentos y, por último, el estimador simple de Anderson - Hsiao.

8.3. Alternativas de estimación de modelos dinámicos con datos de panel

8.3.1. Enfoque simple de máxima verosimilitud

Los problemas descritos, muy similares a los problemas clásicos de parámetros incidentales encontrados por Neyman y Scott (1948) y revisados en otros contextos econométricos por Lancaster (1998), han sido afrontados desde distintos puntos de vista aparentemente distintos pero que, en realidad, pueden conectarse con cierta sencillez.

Una primera alternativa consiste en tratar los parámetros relativos a los efectos fijos μ_i como variables aleatorias cuyas distribuciones pertenezcan a una familia de parámetros de dimensión finita. Dependiendo de las distintas especificaciones de la distribución conjunta de los parámetros μ_i e y_{i0} (observaciones iniciales del proceso autorregresivo), se podrían plantear distintas funciones de verosimilitud para las que los correspondientes estimadores máximo verosímiles (MV) se muestran consistentes en términos generales.

Este procedimiento de estimación por máxima verosimilitud normal presenta, como principal problema, el requerir fuertes requisitos en torno a las distribuciones de los efectos fijos μ_i y, sobre todo, de las condiciones iniciales y_{i0}. Concretamente, las propiedades de los estimadores resultantes son muy sensibles a estas condiciones iniciales, condiciones establecidas sin que, como señalan Arellano y Bover (1990), normalmente el inicio del período muestral coincida con el inicio del proceso dinámico, ni usualmente pueda disponerse de información a priori sobre el punto de partida. Por otro lado, como segundo inconveniente, la aplicación del método requiere frecuentemente cálculos complejos.

8.3.2. Enfoque de variables instrumentales: estimador simple de Anderson - Hsiao

Un método alternativo para evitar los problemas de sesgo en la estimación de modelos dinámicos consiste en utilizar una aproximación de variables instrumentales. Uno de los estimadores más utilizados y que con mayor sencillez ilustran el procedimiento de variables instrumentales en este contexto es el denominado estimador AH (Anderson - Hsiao). Para exponer su morfología supongamos un panel de datos con T=3 que permita reducir el sistema dinámico a 2 ecuaciones en niveles:

$$y_{i2} = \mu_i + \alpha y_{i1} + \epsilon_{i2}$$
$$y_{i3} = \mu_i + \alpha y_{i2} + \epsilon_{i3}$$

A partir de esta especificación en niveles se plantea, para eludir la presencia de la heterogeneidad transversal, la forma en diferencias para la que, en este caso, el sistema quedaría reducido ahora a una sola ecuación:

$$\triangle y_{i3} = \alpha \triangle y_{i2} + \triangle \epsilon_{i3}$$

Como queda dicho, el problema básico de cara a la estimación del parámetro autorregresivo α por MCO en esta ecuación, es la existencia de correlación entre $\triangle y_{i2}$ y $\triangle y_{i3}$. La estimación de variables instrumentales, exigirá encontrar un instrumento incorrelacionado con $\triangle \epsilon_{i2}$ y, sin embargo, correlacionado con la variable a la que deberá sustituir ($\triangle y_{i2}$). El instrumento seleccionado, en este caso, será el valor del nivel y_{i1}, ya que, guardando relación por construcción con $\triangle y_{i2} = y_{i2} - y_{i1}$, no estará sin embargo correlacionado con ϵy_{i3} En este caso (para T=3) el modelo estaría exactamente identificado al contar con una variable instrumental (una condición de ortogonalidad) para la estimación de un solo parámetro.

$$\hat{\alpha}_{AH} = \frac{\sum_{i=1}^{N} y_{i1}[y_{i3} - y_{i2}]}{\sum_{i=1}^{N} y_{i1}[y_{i2} - y_{i1}]}$$

Este estimador es, en realidad, observacionalmente equivalente al estimador de máxima verosimilitud que considerase la función de densidad condicional de las observaciones tomando como condición inicial para el proceso, la primera observación disponible (y_{i1}).

En general, sin embargo, la utilización de un enfoque de variables instrumentales implicará una pérdida de eficiencia respecto al caso de máxima verosimilitud. Una reciente ilustración de los términos de este intercambio puede encontrarse en Wansbeek y Bekker (1996). Los autores consideraron un modelo dinámico simple sin regresores exógenos y con perturbaciones y efectos fijos independientes y normalmente distribuidos. Sobre la base de este modelo derivaron la expresión para el estimador de variables instrumentales óptimo, es decir, aquel que presentaba una varianza asintótica mínima. Los resultado revelaron las importantes diferencias en eficiencia entre el enfoque de variables instrumentales y el de máxima verosimilitud: los autores encontraron que, para regiones del parámetro autorregresivo que son verosímiles en la práctica, el estimador máximo verosímil es superior. Bien es cierto que la diferencia en eficiencia puede ser reducida siempre que se consideren restricciones no lineales de momentos similares a las propuestas por Ahn y Schmidt (1995).

8.3.3. Método generalizado de momentos

A principios de los 80, y como generalización del método de variables instrumentales, se propone el método generalizado de momentos (MGM ó GMM). Siendo Arellano y Bond (1991), quienes proponen inicialmente el procedimiento del Método Generalizado de Momentos como alternativa más eficiente a la aproximación simple de Anderson - Hsiao.

La idea consiste en afrontar la estimación combinando diversos instrumentos en torno a un único vector numérico de coeficientes, que logre que correlaciones muestrales mínimas entre el término de error y cada uno de los instrumentos. Para la selección de instrumentos, GMM utiliza la información que las teorías económicas o el proceso generador de datos subyacente determinan sobre las condiciones los momentos poblacionales. Así, partiendo de determinadas asunciones sobre el proceso generador de datos del modelo dinámico de datos de panel, pueden encontrarse condiciones relativas a los momentos poblacionales sobre los que construir un estimador GMM eficiente que sea además consistente y asintóticamente normal. Dependiendo de las definición de las condiciones relativas a los momentos, son factibles varias formas del estimador GMM (Arellano y Bond (1991), Chamberlain (1992), Arellano y Bover (1995), Ahn y Schmidt (1995 y 1997), Blundell y Bond (1998)).

La estimación por variables instrumentales ofrece una interpretación intuitiva y sencilla del Método Generalizado de Momentos. Efectivamente, el estimador GMM vendría a ser un caso especial de estimación por variables instrumentales en el que el sistema de ecuaciones e instrumentos estuviese sobre - identificado. En ese caso, dado que para la estimación de un parámetro contaríamos con más de una restricción de momentos (condiciones de ortogonalidad), el estimador GMM puede entenderse como una combinación lineal de todos los estimadores obtenidos con cada una de esas condiciones, debidamente ponderados por la precisión de cada una de ellos.

Supongamos el anterior sistema utilizado como ejemplo para ilustrar el estimador AH. Conforme al argumento utilizado en aquel caso pero para T>3, la selección de instrumentos puede ampliarse sin más que asociar, para cada valor de t, las ecuaciones en diferencias y los correspondientes instrumentos (Arellano y Bover (1990)) obteniéndose la expresión genérica:

$$\hat{\alpha} = \frac{\sum_{i=1}^{N}\sum_{t=3}^{T} y_{i(t-2)}[y_{it} - y_{it-1}]}{\sum_{i=1}^{N}\sum_{t=3}^{T} y_{i(t-2)}[y_{it-1} - y_{it-2}]}$$

Definidos así instrumentos y ecuaciones, pueden plantearse conjuntamente las diversas condiciones de ortogonalidad asociadas a cada uno de los instrumentos disponibles mediante la expresión matricial:

$$E[Z_t'\bar{\epsilon}_t] = 0 \rightarrow \begin{bmatrix} y_{i1} & 0 & \dots & 0 \\ 0 & y_{i1}, y_{i2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & y_{i1}, y_{i2}, \dots, y_{i(T-2)} \end{bmatrix}' \begin{bmatrix} \triangle\epsilon_{i3} \\ \triangle\epsilon_{i4} \\ \vdots \\ \triangle\epsilon_{iT} \end{bmatrix} = 0$$

derivándose la correspondiente expresión del estimador óptimo de α. Para ello, debe considerarse que, si ϵ_{it} es una perturbación "ruido blanco", sus diferencias $\triangle\epsilon_{it}$ presentarán la matriz de varianzas y covarianzas simétrica $\sigma^2 H$ siguiente:

$$E[\bar{\epsilon}_{it}\bar{\epsilon}_{it}'] = \sigma^2 H = \sigma^2 \begin{bmatrix} 2 & -1 & 0 & \dots & 0 \\ -1 & 2 & -1 & \dots & 0 \\ 0 & -1 & 2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 2 \end{bmatrix}$$

con lo que la expresión generalizada del estimador sería entonces:

$$\hat{\alpha}_{AH} = \frac{\sum_t \bar{y}_{t(-1)} Z_t [\sum_t Z_t' H Z_t]^{-1} \sum_t Z_t' \bar{y}_t}{\sum_t \bar{y}_{t(-1)} Z_t [\sum_t Z_t' H Z_t]^{-1} \sum_t Z_t' \bar{y}_{t(-1)}}$$

Para afrontar la selección de instrumentos y la posterior construcción del estimador MGM no es imprescindible, como en el ejemplo expuesto más arriba, plantear el modelo en diferencias. Una transformación alternativa muy útil es la propuesta por Arellano (1988) que considera las variables expresadas en desviaciones ortogonales, esto es, cada valor de la variable menos todos sus adelantos. Puede demostrarse que la estimación MCO aplicada sobre los datos en desviaciones ortogonales utilizando los mismos instrumentos que en el caso anterior, conduce al mismo resultado que la estimación MCG del modelo en primeras diferencias expuesta más arriba. Sólo en el caso en que algunos de los instrumentos sean suprimidos, los estimadores no serán equivalentes. Siendo indiferente una u otra transformación según lo visto, las desviaciones ortogonales son más recomendables, ya que pueden servir para suavizar los efectos de algunos problemas adicionales en la estimación. Así, por ejemplo, es bien conocido que el sesgo sobre el parámetro estimado derivado de un eventual error de medida en las variables, queda amplificado por cualquier transformación del modelo, pero de forma más grave si se utilizan diferencias en lugar de desviaciones ortogonales.

En términos generales puede afirmarse que el GMM resulta sorprendentemente flexible para eludir con relativa sencillez las eventualidades que aparecen frecuentemente en cualquier ejercicio de especificación. Sin embargo, como contrapartida, debe indicarse que la fortaleza relativa del método descansa críticamente en la adecuada selección de instrumentos, selección que deberá realizarse atendiendo escrupulosamente a las propiedades observadas de las variables con las que tratamos. Esta selección no podrá realizarse de forma automática sino que, muy al contrario, requerirá la plena implicación del investigador, que, de modo crucial, deberá definir detalladamente el modelo teórico considerado, incluyendo la posible existencia de errores de medida, autocorrelación residual, heterogeneidad inobservable, variabilidad exclusivamente temporal, etc. Sólo en ese caso, será posible una adecuada selección de instrumentos para cada parámetro a estimar; debe tenerse en cuenta que, en un panel con 10 observaciones temporales y 5 variables exógenas estrictas, existen 500 condiciones ó momentos que podrían incorporarse a la estimación GMM y que con T=15 y K=10, el número de condiciones alcanza las 1040. Por ello, Manuel Arellano prefiere utilizar siempre la expresión de Método de Variables Instrumentales y no meramente de Estimador de variables instrumentales.

8.4. Aplicación a una base de datos de empleo

El sistema de diferencias del estimador GMM, puede ser visto como parte de una tendencia econométrica que orienta su practica hacia estimadores que realizan menos supuestos sobre el proceso generador de datos y realizan técnicas mas complejas para aprovechar la información obtenida. El estimador esta diseñado para análisis panel, y conlleva los siguientes supuestos sobre el proceso que genero los datos.

- El proceso debería ser dinámico, con realizaciones de la variable dependendiente influenciada por su pasado.
- Los efectos fijos en la dinámica deberían estar arbitrariamente distribuidos, así la variable dependiente cambiaria consistentemente mas rápido para algunas observaciones que para otras. Esto argumenta contra las regresiones de corte transversal, el cual debería asumir efectos fijos, y favorece al panel estático donde la variación sobre el tiempo puede ser usada para identificar los parámetros.
- Algunos regresores podrian ser endogenos.
- Los cambios idiosincraticos (a parte de los efectos fijos) podrian tener heterocedasticidad individual y correlación serial.
- Los choques idiosincrativos no están correlacionados entre individuos.

Además algunos problemas secundarios de diseño:

- Algunos regresores podrían estar predeterminados pero no ser estrictamente exógenas, aun si la independencia de los errores actuales, están influenciados por el pasado. Los rezagos de la variable dependiente es un ejemplo.
- El número de periodos temporales debe ser pequeña T pequeño y el número de observaciones N debe ser grande.

Además como los estimadores son diseñados para uso general, no se asume que los instrumentos sean viables fuera del conjunto de datos, se asume:

- Los instrumentos viables son internos (basados en rezagos de las variables instrumentadas)

Sin embargo los estimadores permiten la inclusión de instrumentos externos.

$$y_{it} = \alpha y_{it-1} + x'_{it}\beta + \epsilon_{it}$$
$$\epsilon_{it} = \mu_i + v_{it}$$
$$E(\mu_i) = E(v_{it}) = E(\mu_i v_{it}) = 0$$

El error tendrá dos componentes ortogonales, el efecto fijo, y los shocks idiosincráticos.

$$\triangle y_{it} = (\alpha - 1)y_{it-1} + x'_{it}\beta + \epsilon_{it}$$

De esta manera el modelo se analiza, para el nivel de crecimiento de y.

Comenzaremos con el clásico estimador MCO aplicado a la ecuación inicial, y modificaremos paso a paso hasta el estimador de interés.

Trabajaremos con la aplicación del nivel de empleo en las firmas del modelo de Arellano & Bond (1991). Su panel data, es basado en una muestra de 140 firmas UK, encuestadas anualmente en 1976 - 1984. El panel es desbalanceado con algunas firmas teniendo mas información que otras. Dado que el despido y la huelga de trabajadores es costoso, nosotros esperamos que el empleo se ajuste con cambios en los factores como stock de capital, salarios y demanda de productos de las firmas. El proceso de ajuste a los cambios de los factores puede depender del tiempo, lo cual argumenta la inclusión de varios rezagos de estos factores así como la diferencia entre el nivel de empleo de equilibrio actual y del año pasado, lo cual argumenta el modelo dinámico con rezagos de la variable dependiente como regresores.

$$n_{it} = \alpha_1 n_{it-1} + \alpha_2 n_{it-2} + \beta'(L)x_{it} + \lambda_t + \mu_i + v_{it}$$

Las variables son; n: es el empleo de la firma; w: es el nivel de salario de la firma; k: es el nivel de capital de la firma; ys: es el producto agregado de la firma en el sector, como Proxy de la demanda. Todas las variables están en logaritmos. Las variables cuyos nombres terminan en L1 y L2 indican el primer y segundo rezago, respectivamente.

```
regress n nL1 nL2 w wL1 k kL1 kL2 ys ysL1 ysL2 yr*
```

Un problema de aplicar MCO en lo empirico, es que el rezago de y_t es endógena para el efecto fijo del termino error, lo cual nos da un sesgo de panel dinámico. Para verlo consideraremos la posibilidad de que una firma experimenta un largo y negativo choque de empleo por alguna razón no modelada, digamos en 1980, de esta manera el impacto se introduciría al error. Manteniendo todo lo demás fijo, el aparente efecto fijo para el periodo completo de la firma será subestimvado. En 1981, el empleo rezagado y el efecto fijo serán bajos. La correlación positiva entre éste regresor y el error viola el supuesto necesario de consistencia del MCO. En particular, se sobrevalua el coeficiente estimado para el rezago del empleo, por atribuir poder de predicción a este perteneciente al efecto fijo de la firma. Note que aquí T=9, si T fuera mayor, un choque en 1980 impactara, aparentemente, sobre el efecto fijo de la firma reduciéndola, y así manifestando el problema de endogeneidad.

Hay dos maneras de trabajar el problema de endogeneidad. La primera, es usar la transformación de la data removiendo los efectos fijos, la otra es intrumentar y_{it-1} y alguna otra variable endógena similar con variables no correlacionadas con el efecto fijo. El sistema GMM incorpora esa estrategia.

Una primera intuición para mostrar el efecto fijo abstrayéndolo del error es usando variables dummys para cada individuo, y así correr un estimador Mínimo Cuadrado de Variables Dummy:

```
xi: regress n nL1 nL2 w wL1 k kL1 kL2 ys ysL1 ysL2 yr* i.id
```

O corriendo lo mismo de manera resumida:

```
xtreg n nL1 nL2 w wL1 k kL1 kL2 ys ysL1 ysL2 yr*, fe
```

Una tercera manera de conseguir un resultado similar es mediante una regresión en dos etapas, en la primera etapa parcial, se obtienen dummys de otras variables y luego se corre la regresión final con los residuos estimados. La regresión parcial aplica una transformación de desviación promedio para cada variable, donde la media es calculada para cada firma. Un MCO sobre la data así transformada es el estimador Within. Este genera el mismo coeficiente estimado, pero los errores estándar que son casi corregidos pues no toman en cuenta la pre-transformación.

```
xtdata n nL1 nL2 w wL1 k kL1 kL2 ys ysL1 ysL2 yr*, fe
regress n nL1 nL2 w wL1 k kL1 kL2 ys ysL1 ysL2 yr*
```

Pero el estimador within no elimina el sesgo del panel dinámico. Bajo esta transformación, la variable dependiente rezagada, llegaría a ser $y^*_{it-1} = y_{it-1} - \frac{1}{T-1}(y_{i2} + ... + y_{iT})$ mientras que el error seria $v^*_{it} = v_{it} - \frac{1}{T-1}(v_{i2} + ... + v_{iT})$. El problema es que el termino y_{it-1} en y^*_{it-1} esta correlacionado negativamente con el $-\frac{1}{T-1}(v_{it-1})$ en v^*_{it} mientras que por simetria, $-\frac{1}{T-1}(y_{it-1})$ y v_{it} también se mueven juntos.

Peor aún uno no puede atacar la endogeneidad continua, instrumentando y_{it-1}^{*} con rezagos de y_{it-1} porque estos estan incrustados em eñ error transformado v_{it}^{*}. Por el contrario, si T fuera grande entonces los terminos $-\frac{1}{T-1}(v_{it-1})$ y $-\frac{1}{T-1}(y_{it-1})$ serian insignificativos y el problema desapareceria. Judson y Owen (1999) encontraron en una simulación que el sesgo era de 20 % del coeficiente de interés, aún con T=30.

Es interesante observar que desde nuestra primera estimación con MCO con variable endógena rezagada, existía una correlación positiva con el error, sesgando el coeficiente estimado hacia arriba, ahora tenemos el caso opuesto. Note que la estimación del coeficiente del rezago del empleo cae de 1.045 a 0.733. Una buena estimación del verdadero parámetro debería estar cayendo entre estos valores, o por lo menos cerca a éstos, dado que estos números son puntos estimados asociados a intervalos de confianza. Bond (2002) señala que provee de una útil inspección sobre los resultados de los parámetros superiores teóricamente.

Kiviet (1995) argumenta que la mejor manera de abordar el sesgo del panel dinámico es a través del estimador de variables dummys, el cuál corrige el sesgo, encontrando la posibilidad de hacer predicciones sorprendentemente precisas. Sin embargo la aproximación solo trabaja con paneles balanceados y no consideran la endogeneidad potencial de otros regresores.

Como resultado, la estrategia mas práctica ha sido desarrollar estimadores que teóricamente no necesita corrección. Lo necesario para remover el sesgo es transformar la data en diferencias, primero porque cancela los efectos fijos mientras evita la transformación within hecha para cada observación de y^{*} sobre cada individuo. Si las observaciones son ordenadas individualmente con datos de matrices X y Y entonces los efectos fijos pueden ser removidos multiplicando hacia la izquierda a las matrices por una matriz diagonal en bloques que poseen una amplitud de T y cuyas filas suman cero. Para realizar la elección, la transformación deberá de tener un rango completo y así no perderíamos información. Esto debería transformar las variables dependientes sobre los rezagos observados de la variable original, así se validarían los instrumentos. En otras palabras la matriz en bloques debería ser triangular superior o cercana a ella. Un sutil y tercer criterio es que la transformación debería ser resistente a la perdida de datos.

Dos transformaciones son comúnmente usadas, ambas son relativamente canónicas. La primera es la transformación en primeras diferencias, al cual se le llama GMM- en diferencias. Este es afectado por $I_N \otimes M_{\triangle}$ donde

I_N es una matriz identidad de orden N y $M_\triangle$ consiste en una diagonal de -1 con 1 solo hacia la derecha. Aplicando la transformación nos da:

$$\triangle y_{it} = \alpha \triangle y_{it-1} + \triangle x'_{it}\beta + \triangle v_{it}$$

Aunque los efectos fijos se fueron, la variable rezagada es aún endógena, dada que y_{it-1} en $\triangle y_{it-1} = y_{it-1} - y_{it-2}$ está correlacionado con v_{it-1} en $\triangle v_{it-1} = v_{it} - v_{it-1}$. Asimismo, alguna variable predeterminada en X que no es estrictamente exógena podría ser potencialmente endógena pues puede que este relacionada con el error. Pero con la transformación de desviaciones respecto a la media, amplios rezagos de los regresores se mantendran ortogonales con el error y serán viables como instrumentos.

La transformación en primeras diferencias tiene una debilidad. Esta se manifiesta en los rezagos de paneles desbalanceados. Si alguna y_{it} es perdida, entonces ambas $\triangle y_{it}$ y $\triangle yit - 1$ estara perdida de la data transformada y uno podria construir una base de datos que desaparesca completamente en primeras diferencias. Esto motiva la segunda transformación conocida, llamada "desviación ortogonal futura" o "desviación ortogonal" (Arellano y Bover 1995). En lugar de substraer la observación previa, de la contemporánea, esta transformación substrae la media de toda observación futura viable de la variable. No importa cuantos rezagos se considere, es calculable para todas las observaciones excepto para la ultima de cada individuo, asi se minimiza la perdida de la data. Y dado que los rezagos de las observaciones no entran en la formula, estas son instrumentos validos. Para ser precisos, si w es una variable entonces la transformación es:

$$w^{\perp}_{it+1} \equiv c_{it}\left[w_{it} - \frac{1}{T_{it}}\sum_{s>t} w_{is}\right]$$

Donde la suma es tomada sobre las observaciones futuras viables y el factor $c_{it} = \sqrt{\frac{T_{it}}{T_{it}+1}}$. En un panel valanceado la transformación podria ser escrita como $I_N \otimes M_\perp$.

$$M_\perp = \begin{bmatrix} \sqrt{\frac{T-1}{T}} & -\frac{1}{\sqrt{T(T-1)}} & -\frac{1}{\sqrt{T(T-1)}} & \cdots \\ & \sqrt{\frac{T-2}{T}} & -\frac{1}{\sqrt{(T-1)(T-2)}} & \cdots \\ & & \sqrt{\frac{T-3}{T}} & \cdots \\ & & & \ddots \end{bmatrix}$$

Una propiedad de esta transformación es que si w_{it} son independientemente distribuidas antes de la transformación, ellas lo serán después de ella (las filas de M son ortogonales entre ellas). La elección de c_{it} asegura que si w_{it} no es independiente pero si idénticamente distribuida, la propiedad aun persiste. En otras palabras $M_{\perp}M'_{\perp} = I$. Esto no es el caso con la diferenciación, en la cual la tendencia hace que sucesivos errores estén correlacionados aún si ellas están no correlacionadas antes de la transformación $\triangle v_{it} = v_{it} - v_{it-1}$ es matemáticamnte relacionada con $\triangle v_{it-1} = v_{it-1} - v_{it-2}$ por el termino v_{it-1}. Arellano y Bover, muestran que en paneles balanceados, algunas de las dos transformaciones de rango de filas completas podrían conducirnos a estimadores idénticos, manteniendo el set de instrumentos fijos.

Se usara un superíndice * para indicar que la data se transformo por difenciacion o por desviaciones ortogonales. El que aparesca t+1 en lugar de t al lado izquierdo refleja que los computadores prácticamente guardan las desviaciones estándar de las variables transformadas un periodo después, por consistencia, con la transformación en primeras diferencias. Con esta definición, ambas transformaciones eliminan la primera observación para cada individuo y para ambas, las observaciones w_{it-2} y las anteriores son las únicas ausencias de la formula para w^*_{it}, haciéndolas instrumentos validos.

Instrumentando con Rezagos

Se construye un estimador para aplicaciones generales, en el cual no asumimos que el investigador tenga excelentes instrumentos fuera de su set de datos, esperando ser utilizados. Así deberíamos abstraer instrumentos dentro del set de datos. Naturalmente los candidatos a instrumentos para y^*_{it-1}, son y_{it-2} y si la data es transformada por diferencias, $\triangle y_{it-2}$. En el caso de diferenciar, ambos, y_{it-2} y $\triangle y_{it-2}$ están matemáticamente relacionadas a $\triangle y_{it-1} = y_{it-1} - y_{it-2}$ pero no al termino error $\triangle v_{it} = v_{it} - v_{it-1}$. Tanto que v_{it} no estará serialmente correlacionado. La manera más simple de incorporar un instrumento es con MC2E, el cual nos conduce al estimador en diferencias y en niveles de Anderson - Hsiao (1981). Así el estimador en niveles, instrumentado con y_{it-2} en lugar de $\triangle y_{it-2}$, pareciera preferible para maximizar el tamaño muestral. $\triangle y_{it-2}$ no es viable hasta t=4, sin embargo y_{it-2} es viable con t=3, y un periodo adicional de la data es significativo en paneles cortos. Retornando al ejemplo anterior de empleo, podemos implementar el estimador de Anderson - Hsiao en niveles usando el comando ivreg:

```
ivreg D.n (D.nL1= nL2) D.(nL2 w wL1 k kL1 kL2 ys ysL1 \\\
ysL2 yr1979 yr1980 yr1981 yr1982 yr1983)
```

Este es el primer estimador consistente del modelo de empleo, dado nuestros supuestos. Su desempeño es mas pobre con una estimación puntual para la variable dependiente rezagada, el valor de 2.308 salio del rango de credibilidad de 0.733 - 1.045 y sus errores estándar son mas grandes de lo anteriormente calculados.

Para incrementar la eficiencia, nosotros podemos tomar mayores precisiones del estimador AH, usando más instrumentos rezagados de las variables. Extendiendo esta introducción de información, deberíamos incrementar la eficiencia, pero para el estimador MC2E estándar, el uso de mayores rezagos, hará la muestra mas pequeña, dado que los rezagos borraran observaciones inevitablemente.

Trabajando con el esquema de GMM, Holtz, Eakin, Newey y Rosen (1988) mostraron este trade off. Sin embargo, los autores llegaron a reemplazar los missing values por ceros y crear una matriz instrumental que cumplia con una condición de ortogonalidad.

Retomando el ejemplo anterior de empleo, con el siguiente comando, expandimos el estimador de Anderson - Hsiao, generando, al estilo de GMM, instrumentos rezagados de n, entonces usamos en la tecnica MC2E regresiones en diferencias. Esto trata a todos los otros regresores como exógenas, los instrumentos de ellos mismos, aparecen en ambas, en la matriz de regresores de X y en la matriz de instrumentos de Z. Así Z contienen Instrumentos al estilo de GMM y una columna de unos al estilo de VI.

```
forvalues yr=1978/1984 {
forvalues lag = 2 / '= 'yr' - 1976' {
 quietly generate z'yr'L'lag' = L'lag'.n if year == 'yr'
 }
}

quietly recode z* (. = 0)
ivreg D.n D.(nL2 w wL1 k kL1 kL2 ys ysL1 ysL2 yr1979 yr1980 \\\
yr1981 yr1982 yr1983) (D.(nL1 nL2) = z*), nocons
```

Aunque este estimador no es teóricamente consistente, pero es mas eficiente que el de AH, pero aún parece pobre. Ahora el coeficiente estimado para el rezago del empleo tienen un valor de 0.292, son tres desviaciones estándar por debajo del rango de 0.73 - 1.045.

GMM factible, evalua este problema, modelando la estructura de errores de manera mas realista, haciendolos, mas eficientes en teoría y mejor comportados en la práctica.

APLICANDO GMM

La única manera de que los errores puedan ser razonablemente esperados y esféricos, en GMM en diferencias, es si:

- La transformación de errores es iid, lo cual es usualmente no asumido.
- Las desviaciones ortogonales transformados sean usadas, asi los errores mantendrán la esfericidad.

Sabiendo que FEGMM es asintoticamente superior, su implementación requiere de que tengamos que estimar la matriz de varianza covarianza Ω^*, la matriz de covarianzas de los errores transformados (GMM en dos etapas). Para la primera etapa, elegimos arbitrariamente un H, como estimación previa de Ω^*, que es basado en el supuesto de que v_{it} sean i.i.d. Usando esto y dejando que v_i se refiera a los errores idiosincráticos de los i individuos, nosotros dejamos H para $I_N \otimes Var[v_i^*|Z]$ donde:

$$Var(v_i^*|Z) = Var(M_* v_i|Z) = M_* var(v_i v_i'|Z) M_*' = M_* M_*'$$

Que para desviaciones ortogonales es una I, y para diferencias toma la forma:

$$\begin{bmatrix} 2 & -1 & 0 & \dots & 0 \\ -1 & 2 & -1 & \dots & 0 \\ 0 & -1 & 2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 2 \end{bmatrix}$$

Para la segunda etapa FEGMM, se aproxima Ω^* con errores robustos, y estimación conglomerada, la cual se construye con el supuesto de errores que están correlacionadas dentro de los individuos, pero no entre ellos. Por esta razón es casi siempre prudente incluir variables dummys temporales para remover de manera general, choques que relacionan el tiempo con el error.

Con esta elección nosotros investigaremos el clásico estimador en diferencias GMM de Arellano - Bond (1991) para panel dinámico.

Como su nombre lo suguiere, Arellano-Bond propusieron originalmente una transformación en diferencias. Cuando se usan desviaciones ortogonales, en lugar del anterior, quizá el estimador pueda llamarse GMM en desviación, pero no es usual.

```
xtabond2 n L.n L2.n w L.w L(0/2).(k ys) yr*, gmmstyle(L.n) \\\
ivstyle(L2.n w L.w L(0/2).(k ys) yr*) h(1) noleveleq nocons small
```

Donde h(1) especifica que H=I, lo cual envuelve la incorrecta suposición de homocedasticidad, si borramos eso, por defecto H tendrá la forma de la matriz en diferencias, y los resultados serán grandemente incrementados.

```
xtabond2 n L.n L2.n w L.w L(0/2).(k ys) yr*, gmmstyle(L.n) \\\
ivstyle(L2.n w L.w L(0/2).(k ys) yr*) noleveleq nocons
```

Para obtener las estimaciones en dos etapas, debemos de cambiar "robust" a "twostep", el cual realiza la primera y segunda etapa de los resultados de Arellano-Bond (1991). Aún así , el coeficiente de la primera etapa en el rezago del empleo de 0.386 y el de la segunda etapa 0.629 no son los esperados en el rango, lo cual indica problemas de especificación. Blundell y Bond (1998) indicaron que ellos "no esperaban que las expectativas salariales y de capital sean estrictamente exógenas en la aplicación de empleo", pero la regresión asume que si lo son. Si nosotros los instrumentamos con la técnica GMM, entonces los coeficientes sobre empleo rezagado se moverán en un rango creible.

```
xtabond2 n L.n L2.n w L.w L(0/2).(k ys) yr*, gmmstyle(L.(n w k)) \\\
ivstyle(L(0/2).ys yr*) noleveleq nocons robust small
```

Instrumentando variables ortogonales a los efectos fijos

Arellano y Bond, compararon el desenvolvimiento del estimador en diferencias en una o dos etapas para MCO entre grupos y el estimador de Anderson - Hsiao en niveles y diferencias. Usaron simulaciones monte carlo de 7*100 paneles. El estimador GMM en diferencias exhibe el menor sesgo y varianza en la estimación del parámetro de interés, aunque en sus test, el estimador de AH, en niveles, sea muy buen estimador para la mayoría de parámetros elegidos. Pero hay muchos grados de libertad en el diseño de esas pruebas. Como Blundell y Bond (1998) demostraron en simulaciones separadas, si y es cercano a un randon walk, el comportamiento del estimador GMM en diferencias es pobre, ya que los niveles pasados expresan poca información sobre cambios futuros, asi la no transformación de rezagos son instrumentos débiles para las variables transformadas.

Para incrementar eficiencia (bajo supuestos adicionales), Blundell y Bond desarrollaron una mejora sobre Arellano y Bover (1995), esquematizando una segunda estrategia contra el sesgo del panel dinámico. Se transformaron los regresores para abstraer los efectos fijos, esta transformación diferencia los instrumentos, haciéndolos exógenos de los efectos fijos. Esto es válido asumiendo que cambios en las variables instrumentales w no están correlacionadas con los efectos fijos. Es decir $E(\triangle w_{it}, \mu_i) = 0$ para todo i y t. Si esto se mantienen, entonces $\triangle w_{it-1}$ es un instrumento valido para las variables en niveles.

En pocas palabras, donde Arellano y Bond instrumentaron diferencias con niveles, Blundell y Bond instrumentaron niveles con diferencias. Para paseos aleatorios como variables, cambios pasados deberían en efecto ser mas predictivos para niveles corrientes que niveles pasados para cambios corrientes. Así los nuevos instrumentos serán mas relevantes. Una vez mas, la validez depende del supuesto de que v_{it} no este serialmente correlacionado con w_{it-1} y w_{it-2}, lo cual podría correlacionarse con los errores pasados y contemporáneos y también con errores futuros.

En general, si w es endógena, $\triangle w_{it-1}$ es viable como un instrumento dado que $\triangle w_{it-1} = w_{it-1} - w_{it-2}$ no deberá estar correlacionado con v_{it} y tempranas realizaciones de $\triangle w$ podrían instrumentar muy bien. Y si w es predeterminada, la contemporaneidad $\triangle w_{it} = w_{it} - w_{it-1}$ es también valida dado que $E(w_{it}, v_{it}) = 0$.

Para aplicar Blundell y Bond en la ecuación de empleo, esta vez, se borra los dos rezagos de los dos periodos del empleo y del capital en el modelo, y precindimos del sector de demanda del producto. También se trata al salario y al capital como potencialmente endógeno, generando instrumentos GMM para ellos:

```
xtabond2 n L.n L(0/1).(w k) yr*, gmmstyle(L.(n w k)) \\\
ivstyle(yr*, equation(level)) robust small
```

Estos resultados no se publicaron, Blundell y Bond dejaron H=I en lugar de usar una forma como la matriz de errores de desviaciones ortogonales. La estimación puntual del coeficiente del empleo rezagado es mayor que la estimada antes, aunque no difiere estadísticamente en las desviaciones estándar del error. Mas aun, está dentro del rango deseado, y los errores estándar reportado son la mitad de los que fueron antes.

Aunque supuestos adicionales son requeridos para validar esta estimación no trivial, es costoso testearlo. El test de Sargant en diferencias, en la salida anterior para los instrumentos del GMM, es muy tranquilizador, con un p-value, de 1, quizá demasiado tranquilizador dada la debilidad del test de Hansen cuando los instrumentos son numerosos.

Testeando la autocorrelación

El test de Sargan/Hansen para validad los instrumentos es típico luego de una estimación GMM. Adicionalmente Arellano y Bond desarrollaron un test de un caso especial que debería hacer a algunos rezagos no validos como instrumentos, llamada autocorrelación en las perturbaciones idiosincráticas ϵ_{it}. Por supuesto, las perturbaciones v_{it} están presumidas de autocorrelación porque contienen efectos fijos y los estimadores son diseñados para eliminar ese problema, pero las ϵ_{it} están serialmente correlacionadas en orden 1, dado que y_{it-2} es endógena para v_{it-1} y dado que el termino error en diferencias $\triangle\epsilon_{it} = v_{it} - v_{it-1}$, hace de esta un instrumento invalido. Los investigadores tendrían que verse en la necesidad de restringir los instrumentos a tres rezagos o menores de y a menores que se encontrara correlación de segundo orden, en cuyo caso se debería necesitar empezar con rezagos mucho mas menores aún.

Por ello el test de autocorrelación a parte del de efectos fijos, el test de Arellano Bond es aplicado sobre los residuos en diferencias, dado que dvit matemáticamente esta relacionado con $\triangle v_{it-1}$ pues comparten a v_{it-1}, correlación serial de primer orden negativa es esperada en diferencias y es evidente. En general nosotros inspeccionamos la autocorrelación serial de orden l, en niveles para revisar la autocorrelacion de orden l+1 en diferencias. Tal aproximación no trabaja con desviaciones ortogonales porque todos los residuos en desviaciones están matematicamente interrelacionados dependiendo de su relación con rezagos adelantados. Siempre, luego de estimaciones en desviaciones, el test es evaluado sobre los residuos en diferencias.

El test de autocorrelacion de Arellano - Bond es valido para regresiones GMM, sobre panel data, incluyendo MCO y MC2E, tanto como ninguno de los regresores es "post determinado" dependiendo del futuro de los errores (una regresión de efectos fijos o Within puede violar el supuesto si T es pequeño). Tambien, se pues ver esto de manera resumida, asumiendo que los errores no están correlacionados entre individuos. Escribiendo el comando "abar" para hacer el test viable luego de "regress, ivreg, ivreg2, newey, newey2".

Debemos aprender dos lecciones, la primera es recordar la importancia de las variables dummys temporales para prevenir la mas común correlación entre individuos, correlacion contemporánea. El test asume no correlacion entre individuos. Segundo, es que el test depende del supuesto de que N es amplio. La amplitud no define precisión pero aplicarlo a un panel con N=20, por ejemplo, parece ser preocupante.

A diferencia de la regresión GMM que simula 7*100 paneles con AR(1), Arellano y Bond encuentran que su test tiene mayor poder que los test de Sargan y Hansen para detectar la valides de instrumentos, siendo estos invalidos cuando existe autocorrelación. El test incumple, sin embargo, cuando la autocorrelación cae a 0.2, donde se suele rechazar la hipótesis nula de no autocorrelacion serial la mayoría de las veces.

RETO 4

a. Para ilustrar el mercado de crédito en moneda extranjera, se considera que los bancos, a parte de los depósitos, pueden obtener todos los fondos que necesiten del exterior, además de que puede existir más de un instrumento de política monetaria, como la tasa de interés en MN y los encajes en MN y ME así como los encajes a las líneas del exterior. Así, pues se incorpora a este modelo, la variable que indica al producto y al nivel de precios, y se abandona el supuesto de que las tasas de encaje se mantienen constantes. El supuesto básico es que los bancos evaden el riesgo cambiario "casando" activos y pasivos por tipo de moneda; de esta manera, es como si hubiera dos sistema bancarios, uno que opera en moneda nacional (MN), y otro que opera en moneda extranjera (ME). Además de que se supone fija la tasa de interés en MN y el tipo de cambio, pasando a ser determinados los bonos en el mercado de crédito en ME y la oferta monetaria en el mercado de dinero. De esta manera, se puede modelar el equilibrio en el mercado de crédito en ME, lo cual es relevante en un análisis del crédito en la economía peruana, que está caracterizada por una alta dolarización y porque además se toma en cuenta fuentes alternativas de financiamiento distintas a los depósitos como las fuentes de financiamiento externo.

Con la información contenida en el archivo "finanzas.dta", se le pide:

- Estimar un modelo de panel dinámico explicando la cantidad de credito en equilibrio en función de la tasa de interés activa en moneda nacional, la tasa de interés activa en moneda extranjera, la tasa de interés de bonos en ME, la devaluación esperada, el volumen de lineas del exterior y los depósitos en ME. ¿Cuáles serian los signos esperados?
- ¿Cómo varia su análisis inicial, si se sabe que la tasa de interés activa en ME, depende de la tasa de interes internacional (libor a 3 meses),y de un ratio de morosidad de los prestamos en ME[1]? Realice los testeos respectivos.

[1]Se sabe que el ratio de morosidad=atrasados/(atrasados+refinanciados+vigentes)

Bibliografía

[1] Moya, Rufino - Estadística Descriptiva.

[2] Moya, Rufino; Saravia, Gregorio. -Probabilidad e Inferencia Estadística.

[3] Wooldridge, Jeffrey M. - Introducción a la Econometría.

[4] Gujarati, Damodar - Fundamentos de Econometría.

[5] Badi H. Baltagi - Econometric Analysis of Panel Data, 3rd Edition.

[6] Manuel Arellano and Stephen Bond - Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations.

[7] J. Scott Long, Jeremy Freese - Regression Models for Categorical Dependent Variables Using Stata, 2nd Edition.