Problem Set #2

Profesor: Tomás Rau
Ayudantes: Valentina Andrade y Nicolás Valle

Fecha de Entrega: 10 de Mayo

El Problem Set 2 consta de 150 puntos e incluye una breve sección teórica y otra empírica. Ambas secciones se deben desarrollar de manera **individual**. Ud. puede intercambiar ideas, no respuestas ni códigos.

Las entregas deben realizarse exclusivamente a través de CANVAS. Cualquier entrega realizada después de la fecha límite será penalizada con 5 décimas por día de atraso.

En cuanto al formato, se requiere la entrega de un archivo comprimido que contenga: 1) documento pdf (escrito preferentemente en Latex) y 2) códigos para replicación (puede utilizar Matlab, R, STATA, Julia). Respuestas sin códigos tendrán la nota mínima. Los códigos entregados deben replicar los resultados de las respuestas.

Parte 1. Teórica (25 puntos)

GMM

Considere el modelo de regresión truncada como es presentado en Hayashi (Secc. 8.2). El modelo se puede estimar mediante GMM utilizando K+1 condiciones de momentos. Las primeras K vienen dadas por:

$$\underset{(K\times 1)}{\mathbf{g}_1}\left(\mathbf{w}_t;\boldsymbol{\delta},\gamma\right)=\left(y_t-\frac{1}{\gamma}\mathbf{x}_t'\boldsymbol{\delta}-\frac{\lambda\left(v_t\right)}{\gamma}\right)\cdot\mathbf{x}_t$$

donde $\mathbf{w_t}=(y_t,x_t)'$ y $v_1\equiv\gamma c-\mathbf{x}'\boldsymbol{\delta}$, la K+1 ésima condición de momento está dada por:

$$g_2\left(\mathbf{w}_t;\boldsymbol{\delta},\gamma\right)=y_t^2-\frac{1}{\gamma^2}\left[1+\lambda\left(v_t\right)\gamma c+\left(\mathbf{x}_t'\boldsymbol{\delta}\right)^2+\lambda\left(v_t\right)\mathbf{x}_t'\boldsymbol{\delta}\right]$$

a) Verifique que $E(\mathbf{g_1}(\mathbf{y}_t;\ \boldsymbol{\delta}),\gamma)=0$ y $E(\mathbf{g_2}(\mathbf{y}_t;\ \boldsymbol{\delta}),\gamma)=0$. **Hint:** Use la ley de expectativas iteradas. (10 puntos)

b) Muestre que ($\boldsymbol{\delta}^*,\gamma$) es el estimador ML si y solo si satisface las K+1 condiciones de momentos mencionadas. **Hint:** Sea $s_2\left(\mathbf{w}_t;\boldsymbol{\delta},\gamma\right)$ = el K+1 ésimo elemento del vector de score de likelihood, entonces:

$$s_2\left(\mathbf{w}_t;\boldsymbol{\delta},\gamma\right)+\gamma g_2\left(\mathbf{w}_t;\boldsymbol{\delta},\gamma\right)=\left(\mathbf{x}_t'\boldsymbol{\delta}\right)\left(y_t-\frac{1}{\gamma}\mathbf{x}_t'\boldsymbol{\delta}-\frac{\lambda\left(v_t\right)}{\gamma}\right)$$

(15 puntos)

Parte 2. Ejercicios Empíricos (125 puntos)

Modelos de conteo y media condicional (40 puntos)

Para la resolución de esta tarea se cuenta con un archivo con datos llamado **ofensas.dta** en el que se registra para una muestra de hombres entre 20 y 45 años los delitos cometidos en un período de tiempo determinado. Además dispone de la edad (y su cuadrado), un indicador de matrimonio y un indicador de residencia en zona urbana.

a) Discuta porque un modelo de conteo es apropiado para este contexto. Presenta una tabla con estadística descriptiva de las variables con las que cuenta y aventure una explicación de los determinantes del número de delitos. Discuta cual espera que sean los signos de las derivadas en un modelo que tome como variable dependiente el número de delitos y como explicativas el resto de variables. (10 puntos)

b) Estime el modelo suponiendo un proceso de Poisson, interprete cuantitativamente los efectos marginales de las variables. Sea cuidadoso en la interpretación de las variables dicotómicas. Presente tablas con estimaciones y significancia estadística. (10 puntos)

c) Discuta la presencia de *overdispersion* ¿Como la mediría? (10 puntos)

d) Estime un modelo Binomial Negativo. Cuantifique la sobredispersión a luz de sus resultados en el segundo ejercicio teórico. Presente una tabla con el resultado de sus estimaciones. (10 puntos)

Modelos de duración (40 puntos)

En el artículo ¿Cuánto Dura el Desempleo de la Población más Pobre en Chile? de Rodrigo Montero, el autor analiza la pregunta en cuestión, como los determinantes de dicha duración.

a) Lea el paper y escriba un resumen y las conclusiones más importantes. Lo puede leer en este link: **https://www.scielo.cl/scielo.php?script=sci_arttext&pid=S0717-68212007000200004**. (5 puntos)

b) Usando la encuesta **sample.dta** y las variables usadas en el artículo, replique la estimación del modelo de duración por máxima verosimilitud con censura por la derecha, asumiendo una distribución lognormal (cuadro 4 completo). Use el factor de expansión del Panel Casen. Nota, las duraciones iguales a cero son duraciones iguales a cero. Nota 2, los datos censurados puede observarlos por el hecho que hay personas con duraciones mayores que cero pero cuya situación ocupacional (al momento de la encuesta) es de desocupado, eso significa que el individuo entró al desempleo y no ha salido de dicho estado. Aquellos con duraciones mayores que cero y con situación "ocupado" son historias completas.

   Compare sus resultados con los de Montero (2008). (10 puntos)

c) Estime el mismo modelo de la letra b) usando una distribución Weibull. Interprete los resultados y compárelos con los de la letra b). Sea cuidadosa. (10 puntos)

d) Demuestre que la distribución lognormal proporciona un mejor fit para los datos que la Weibull usando algún criterio de información. ¿Ocurre lo mismo si se compara con una Gamma generalizada? (10 puntos)

e) Grafique la función de supervivencia (paramétrica y no paramétricamente) y la hazard function. (5 puntos)

Ayuda: use el comando **streg** de STATA para la estimación. Sea cuidadosa en setear sus datos, para ello use **stset**. Lea la ayuda.

GMM (45 puntos)

En este ejercicio veremos cómo explotar información de un censo, mediante GMM. Esta idea se basa en el artículo de Hellerstein e Imbens (1999). El objetivo de esta pregunta es que aprendamos sobre las ganancias de eficiencia de GMM y sean capaces de programarlo en STATA usando el comando gmm.

a) Resuma brevemente (máximo una plana) el paper de Hellerstein e Imbens (1999) y señale las principales contribuciones. (10 puntos)

   Se dispone de una muestra similar a la usada por los autores (no es exactamente igual), llamada nls.dta, la que contiene cuatro variables: educ, earn, iq y edad. Considere las siguientes condiciones de momento:

$$E[\psi(\ln(earn), educ, expr, iq, \beta)] = 0$$

   donde

$$\psi = \begin{pmatrix} \ln(\text{earn}) - \beta_0 - \beta_1 * educ - \beta_2 * \text{expr} - \beta_3 * \text{expr}^2 - \beta_4 * iq \\ \text{educ} * \left(\ln(\text{ earn }) - \beta_0 - \beta_1 * \text{educ} - \beta_2 * \text{expr} - \beta_3 * \text{expr}^2 - \beta_4 * iq\right) \\ \text{expr} * \left(\ln(\text{ earn }) - \beta_0 - \beta_1 * \text{educ} - \beta_2 * \text{expr} - \beta_3 * \text{expr}^2 - \beta_4 * iq\right) \\ \text{expr} * \left(\ln(\text{ earn }) - \beta_0 - \beta_1 * educ - \beta_2 * \text{expr} - \beta_3 * \text{expr}^2 - \beta_4 * iq\right) \\ iq * \left(\ln(\text{ earn }) - \beta_0 - \beta_1 * educ - \beta_2 * \text{expr} - \beta_3 * \text{expr}^2 - \beta_4 * iq\right) \end{pmatrix}$$

   Note que la condición de momento se da naturalmente en un modelo lineal del tipo:

$$\ln(\text{ earn }) = \beta_0 + \beta_1 * educ + \beta_2 * expr + \beta_3 * \text{expr}^2 + \beta_4 * iq + v$$

   donde el supuesto de identificación $E(x'v) = 0$, con $x_i = \left[1, educ_i, \text{expr}_i, \text{expr}_i^2, iq_i\right]$, es equivalente a la condición de momento antes descrita.

b) Usando los datos entregados (**nls.dta**) estime los parámetros por MCO con la corrección Eicker-Huber-White para los errores estándar. Luego, compárelos con la estimación por GMM. Hint: Puede implementar el estimar en 2 etapas usando el comando gmm de STATA con la opción weighting robusta. (10 puntos)

c) Considere que se dispone de información censal que nos provee 4 nuevas condiciones de momento, así

$$\psi = \begin{pmatrix} ln(earn) - \beta_0 - \beta_1 * educ - \beta_2 * expr - \beta_3 * expr^2 - \beta_4 * iq \\ educ * (ln(earn) - \beta_0 - \beta_1 * educ - \beta_2 * expr - \beta_3 * expr^2 - \beta_4 * iq) \\ expr * (ln(earn) - \beta_0 - \beta_1 * educ - \beta_2 * expr - \beta_3 * expr^2 - \beta_4 * iq) \\ expr^2 * (ln(earn) - \beta_0 - \beta_1 * educ - \beta_2 * expr - \beta_3 * expr^2 - \beta_4 * iq) \\ iq * (ln(earn) - \beta_0 - \beta_1 * educ - \beta_2 * expr - \beta_3 * expr^2 - \beta_4 * iq) \\ ln(earn) - 2{,}0647 \\ ln(earn)^2 - 4{,}5258 \\ educ - 13{,}97 \\ ln(earn) * educ - 29{,}099 \end{pmatrix}$$

   donde las últimas **4** se obtienen del censo. Por ejemplo, en el censo de 1980 se encuentra que la media poblacional del logaritmo natural del ingreso (earn) es 2.0647. Compute el estimador de GMM en 2 etapas con weighting matrix robusta. (10 puntos)

d) Compare los resultados con los obtenidos en b) con c). En particular, refiérase a la eficiencia. ¿Cómo se comparan los errores estándar de GMM con las 9 condiciones de momentos con los de GMM con 5 condiciones de momento? (10 puntos)

e) Haga un test para las restricciones de sobreidentificación y concluya si rechaza o no la hipótesis nula. (5 puntos)

Suponga ahora que $E[\ \text{vi} \mid x_i] \neq 0$ y por lo tanto, utiliza un vector $l \times 1$ de instrumentos z_i con $l = k$, tal que $E[z_i \mid v] = E[z_i]$.