

EFFECTOS DE COVARIANTES EN LA DISTRIBUCIÓN DE LA *DISPONIBILIDAD A PAGAR* POR EL MÉTODO DE TURNBULL

Covariate Effects on the *Willingness To Pay* by the Turnbull estimator method

Villa Cox, Gonzalo Gabriel
Facultad de Ciencias Humanísticas y Económicas
Escuela Superior Politécnica del Litoral
Guayaquil - Ecuador
gvilla@espol.edu.ec
(04) 2269110

Resumen

Este artículo pretende servir como una guía para la implementación del método de Turnbull aplicado a la investigación sobre valoración contingente. En él se discutirá la metodología y se implementarán algoritmos de solución para calcular los estimadores necesarios en SCILAB 4.1, un lenguaje de propósito general orientado a matrices. Como parte de la discusión, se tratarán varios problemas prácticos que surgen en el manejo de datos en este formato. Finalmente, se probará la implementación como parte de un estudio de valoración contingente, donde analizarán aspectos relevantes de la *disponibilidad a pagar* por una política pública para mitigar los efectos de externalidades negativas relacionadas con la enfermedad del VIH/SIDA. Se enfatizará el efecto que tiene sobre esta medida que tiene la *percepción del riesgo de infección* (definido como covariante).

Abstract

This article intends to aid as an implementation guide for Turnbull's estimator method in applied contingent valuation research. It will discuss the methodology and implement detailed algorithmic solutions to obtain the necessary estimators in SCILAB 4.1, a general purpose matrix-oriented language. As a part of the discussion, several practical issues involved in data management will be addressed as well. Finally, the implementation will be tested as part of a contingent valuation study, where it will analyze relevant aspects of the *willingness to pay* for public intervention to undermine the effects of negative externalities related with the HIV/AIDS disease. Special emphasis will be made on the effect that *infection risk perception* (defined as covariate) has on this measure.

1. Introducción

El método de *Valoración Contingente* se ha constituido en una herramienta indispensable para el estudio de las externalidades, especialmente si su fuente se distingue por tener características de bien público. De manera general, en dichas situaciones no es fácil asignar derechos de propiedad, por lo que suele ser común recurrir a la intervención pública, sea para mitigar su impacto (*externalidades negativas*) o para incentivarlo (*externalidades positivas*). Estos factores, entre otros, impiden establecer un sistema de precios por medio de la creación de mercados que permita medir los efectos que la política pública pueda tener en el nivel de bienestar de los agentes económicos (*p. ej.* Variaciones en el excedente del consumidor). El método en cuestión permite aproximar variaciones en el nivel de bienestar mediante la estimación de la *Disponibilidad a Pagar (DAP)* de los agentes. El enfoque del trabajo actual no radica en detallar las características de esta metodología, sin embargo, es necesario puntualizar que el análisis que se deriva de la misma es susceptible al comportamiento que tenga la disponibilidad a pagar.

El *método de Turnbull* es un método de estimación no paramétrica que aplicado a la valoración contingente permite aproximar la distribución de la disponibilidad a pagar de los agentes de tal manera que los estimadores que se obtengan exhiban propiedades asintóticas similares a las de la distribución normal. [1] Es aplicable en los casos en que el investigador utilice un formato dicotómico, muestreo aleatorio y una sola pregunta para la parte de valoración. Es decir, se puede utilizar cuando la pregunta de valoración se refiera a una sola política y a los encuestados se les asigne aleatoriamente un vector dado de precios. Sin embargo, esta característica no impide que el investigador recoja información adicional acerca de otros factores que puedan afectar la disponibilidad a pagar de los individuos que conforman la muestra. Estos pueden ser factores demográficos, psicográficos, sociales o u otros que serán referidos de manera general como *covariantes*.

El presente trabajo se enfoca en la descripción, implementación y aplicación de los estimadores de Turnbull en el contexto de la valoración contingente. Se detallarán los procedimientos de cómputo de los mismos, cuya justificación teórica se encuentra detallada en el apéndice, puntualizando ciertos aspectos empíricos que surgen al momento de su desarrollo. Finalmente, se presentará un caso de estudio en el que se mostrará el desempeño de la metodología y se analizarán las conclusiones que se pueden extraer a partir del método para el estudio en particular. Como parte de ello, se indagará el efecto que tienen covariantes específicos en el comportamiento de la *DAP* de los involucrados en el problema abordado.

2. Implementación de la metodología

El esquema en base al que se realizan la mayor parte de estudios de valoración contingente ofrece información limitada acerca de la *DAP* de los encuestados. Si la persona responde que *sí* a la pregunta *¿Esta dispuesto a pagar un precio \$ P_j por una política ... ?* únicamente revela que su disponibilidad a pagar (DAP_i) es mayor o igual al precio P_j que se le oferta en la encuesta. En caso contrario simplemente revela que esta es menor. Comúnmente a los individuos que conforman la muestra se les asigna aleatoriamente un vector de precios conocido con M elementos. Se puede dividir a los K_j sujetos que la conforman en M sub-muestras distintas conformadas por K_j elementos, dentro de las cuales todos son ofertados el mismo precio. Para cada uno de estas sub-muestras, se puede denotar al número de personas que responden *sí* como Y_j y al número que responden *no* como N_j . Esta claro entonces que se podría hablar de la DAP_i del individuo i dentro de cada sub-muestra j como una variable aleatoria con distribución acumulada $Pr(DAP_i < P_j)$. Esta es la probabilidad que una persona responda *no* al precio P_j . El apéndice detalla de forma extensa la derivación formal del estimador \tilde{F}_j que se utilizará para aproximar dicha probabilidad: $\tilde{F}_j = \frac{N_j}{K_j}$. Es decir, el estimador para la probabilidad de que un individuo al azar responda que *no* a pagar \$ P_j es igual a la proporción de encuestados que respondieron *no* a dicho precio.

Para completar el proceso anterior es necesario considerar varios problemas prácticos, ya que los datos muestrales obtenidos a raíz del proceso de encuesta por lo general están disponibles en formato tabular, sin orden específico y plagados de datos inválidos. Por ende, es necesario estandarizar un proceso de depuración y ordenación de los datos disponibles para poder separar las sub-muestras especificadas de manera correcta y poder obtener las estimaciones mencionadas anteriormente. Se escogió desarrollar algoritmos reutilizables que permitan esto en *SCILAB 4.1*¹, un lenguaje de propósito general especializado en métodos numéricos, completamente extensible y cuyo formato nativo de datos son las matrices. Si se disponen de los datos en bruto en cualquier software de hoja de cálculo o de está índole compatible con el formato *CSV*², no supone ninguna dificultad crear una variable en SCILAB que contenga una matriz que represente la tabla con los datos de la encuesta. En dicha matriz, cada una de sus filas representa un individuo de la muestra. Esta debe contar con al menos dos columnas, una para el precio ofertado a cada individuo y la otra para su respuesta a la pregunta de valoración. Cualquier información adicional sobre cada individuo (incluyendo las respuestas de este a cualquier covariante sobre el que se le pregunte) podrán ser incluidas como columnas adicionales a las anteriormente mencionadas.

Al pasar los datos a SCILAB, este interpreta cualquier dato inválido o perdido como un *dato no numérico*. En base a esto, la siguiente rutina libra la variable que recibe como argumento de cualquier dato inválido, simplemente eliminando por completo la observación que lo contiene:

¹Para mayores detalles sobre este software, incluyendo manuales completos de uso y desarrollo, visite www.scilab.org

²Este formato de archivo es relativamente simple y muy común para el manejo de tablas de datos. Para mayor información visite es.wikipedia.org/wiki/CSV

```

function [nan_free]=remove_nan(data)
    nan_free=data;

    wasnan = or(isnan(nan_free),2);
    havenans = or(wasnan);
    if havenans
        nan_free(wasnan,:) = [];
    end
endfunction

```

Aunque esto implique perder cierta cantidad de observaciones, las restantes serán válidas para aplicar la metodología siempre y cuando se considere que no existe ningún patrón sistemático con los datos no numéricos que puedan ser causa para posibles sesgos.

El siguiente paso consta en ordenar la matriz que contiene los datos depurados según la columna que contenga la información acerca del precio ofertado a cada individuo. Esto permitirá posteriormente separar esta matriz en sub-muestras de acuerdo al precio ofertado. Para ello, SCILAB dispone del procedimiento `M=lex_sort(M,col)` que ordena la matriz `M` según la columna o columnas `col` que se pasen como argumentos a la rutina. Posteriormente será posible extraer cada sub-muestra simplemente contando el número de elementos sucesivos repetidos en la columna indicada. El procedimiento siguiente permite efectuar esta tarea:

```

function [count_el, endpos]=count_rep_el(mat, col, startpos)
    n=size(mat,1);
    if (startpos<n)
        el=mat(startpos,col);
        for endpos=startpos+1:n
            if el~=mat(endpos,col)
                break;
            end
        end
        if (endpos==n)
            count_el=endpos-startpos+1;
        else
            count_el=endpos-startpos;
        end
    elseif (startpos==n)
        endpos=n;
        count_el=1;
    else
        error('Index out of bounds.');
```

La implementación de esta rutina toma en cuenta ciertas sutilezas que se pueden presentar al momento de obtener el número de elementos de cada sub-muestra. La columna en donde se busca en la matriz `mat` viene dada por el argumento `col`, y se permite su uso de manera iterativa mediante el argumento de entrada `startpos` que indica la posición de inicio de búsqueda y el argumento de salida `endpos` en el que se devuelve la posición de finalización de la búsqueda dentro de la matriz de datos. El número de elementos de la sub-muestra viene dado por el argumento de salida `count_el`.

El algoritmo de obtención de K_j , N_j , Y_j y \tilde{F}_j para cada sub-muestra aprovechará la rutina descrita anteriormente de manera iterativa a través de toda la matriz que contiene los datos de la muestra. Sin embargo, antes de mostrar su implementación conviene enfocarse en el problema central sobre el cual gira el método de Turnbull. Los estimadores \tilde{F}_j representan la distribución acumulada de la *DAP*, por lo que se debería esperar que estos presenten valores monotónicamente crecientes a medida que se incremente el precio ofertado a cada sub-muestra. No obstante, debido a que el muestreo se realiza de forma aleatoria no se puede garantizar esta condición simplemente calculando F_j de la manera descrita. El apéndice describe de manera formal como se deriva el proceso de *price pooling*, el cual consiste en agrupar la primera sub-muestra para la cual se viole condición de monotonicidad creciente con la sub-muestra anterior a ella, dado que estas se encuentran ordenadas de acuerdo al precio. Luego, se calcula el nuevo estimador \hat{F}_j en base a los valores de \hat{N}_j y \hat{K}_j de la nueva sub-muestra que se forma. Si se realiza esto de manera sucesiva en todas las sub-muestras, repitiendo el proceso cuantas veces sea necesario hasta garantizar que todos los estimadores \hat{F}_j de las nuevas \hat{M} sub-muestras resultantes cumplan con esta condición, se puede corregir el problema y garantizar la que estos estimadores son monotónicamente crecientes. El proceso descrito no es más que una suavización conservadora, que permite demostrar que el comportamiento asintótico de la distribución de la *DAP* es normal, siempre que sea aplicable para corregir el problema y

la muestra con la que se cuente pueda considerarse grande. Esto permite calcular medidas de tendencia central para la *DAP*, tal como se describe en el apéndice. Para ello es necesario apuntar que la varianza del estimador \widehat{F}_j viene dada por $\frac{\widehat{F}_j(1-\widehat{F}_j)}{\widehat{K}_j}$.

A partir del método de Turnbull se puede obtener dos medidas de tendencia central para la *DAP*. La mediana de la distribución se puede ubicar entre las sub-muestras que presenten los valores de \widehat{F}_j inmediatamente inferiores y superiores al 50%. La *cota inferior* de la *DAP* viene dada por $\sum_{j=0}^{\widehat{M}} \widehat{P}_j [\widehat{F}_{j+1} - \widehat{F}_j]$ con varianza $\sum_{j=1}^{\widehat{M}} \frac{\widehat{F}_j(1-\widehat{F}_j)}{\widehat{K}_j} (\widehat{P}_j - \widehat{P}_{j-1})^2$, en donde la notación $\widehat{}$ indica que todas las variables son el resultado del proceso de pooling, siempre y cuando no se consideren *DAP* negativas. Esta medida es conservadora en el sentido que no apunta a aproximar el valor puntual de la *DAP* del individuo, sino trata de estimar el límite inferior de la misma. Esto se debe, como señala el apéndice, a que el límite superior por el que está acotada la *DAP* es desconocido. La cota inferior es un estimador no sensible al valor que pueda tomar este límite, a diferencia de lo que sería el tratar de hallar un valor puntual para la *DAP*. Al ser insesgado y asintóticamente normal, se puede realizar pruebas de hipótesis interesantes acerca de esta medida que permitan realizar inferencia estadística.

El procedimiento mostrado a continuación es la implementación completa del método de Turnbull según se ha descrito hasta ahora. El código cuenta con comentarios generales acerca de su funcionamiento y especifica la sección que calcula cada uno de los estimadores descritos:

```
function [tblull_table, exp_wtp, var_exp_wtp]=turnbull(svy_data)

// Validar que la matriz tenga al menos 2 columnas
if (size(svy_data,2)<2)
    error('Invalid data format.');
```

end

```
// Eliminar datos inválidos de la muestra
data=remove_nan(svy_data);

// Marcador para las columnas que contienen la oferta
// y la respuesta a la pregunta de valoración
global COL_BID COL_ANS;
```

```
// Si no esta definida globalmente, asumir un valor
if (COL_BID==[])
    col_bid=1;
else
    col_bid=COL_BID;
end

// Si no esta definida globalmente, asumir un valor
if (COL_ANS==[])
    col_ans=2;
else
    col_bid=COL_ANS;
end

// Organizar la muestra por sub-muestras de acuerdo a cada precio
// ofertado. Dentro de cada sub-muestra ordenar las respuestas
// positivas (1) y negativas (0).
data=lex_sort(data, [col_bid col_ans]);

// Construir la tabla para el problema de maxima verosimilitud no
// restringida. Calcular las variables "Kj", "Yj", "Nj" y "Pj" para cada
// una de las sub-muestras resultantes de la ordenación.
n_data=size(data,1);
i1_data=1; i_bid=0;
while (i1_data<n_data)
    [n_bid, i2_data]=count_rep_el(data,col_bid, i1_data);
    i_bid=i_bid+1;
    Pj(i_bid)=data(i1_data, col_bid);
    Kj(i_bid)=n_bid; Yj(i_bid)=0; Nj(i_bid)=0;
    i1_ans=i1_data;
    while (i1_ans<i1_data+n_bid-1)
        [n_ans, i2_ans]=count_rep_el(data,col_ans, i1_ans);
        if (i2_ans-i1_data-n_bid>0)
            n_ans=i1_data+n_bid-i1_ans;
```

```

        end
        if (data(i1_ans,col_ans)==0)
            Nj(i_bid)=n_ans;
        else
            Yj(i_bid)=n_ans;
        end
        i1_ans=i2_ans;
    end
    i1_data=i2_data;
end

// Número de sub-muestras sin aplicar restricción de monotonicidad
M=i_bid;

// Estimador de máxima verosimilitud para "Fj" (sin restricción)
Fj=Nj./Kj;

// Aplicar el proceso de "pooling" sobre todas las variables para
// corregir errores de monotonicidad.
j=1;
while (j<M)
    if (Fj(j+1)>Fj(j))
        j=j+1;
    else
        Kj(j)=Kj(j)+Kj(j+1);
        Yj(j)=Yj(j)+Yj(j+1);
        Nj(j)=Nj(j)+Nj(j+1);
        Fj(j)=Nj(j)/Kj(j);

        Pj(j+1)=[];
        Kj(j+1)=[];
        Yj(j+1)=[];
        Nj(j+1)=[];
        Fj(j+1)=[];

        j=1;
        M=M-1;
    end
end

// Calcular la varianza de los estimadores Fj
Vj=(Fj-Fj.^2)./Kj;

// Reportar la tabla de Turnbull
tbull_table=[Pj Nj Kj Fj Vj];

// Calcular el estimador de la cota inferior de la DAP y su varianza
// Este calculo es sensible al supuesto F_0=0 y F_M+1=1
exp_wtp=0;
var_exp_wtp=0;
for j=1:M
    if (j<M)
        exp_wtp=exp_wtp+Pj(j)*(Fj(j+1)-Fj(j));
    else
        exp_wtp=exp_wtp+Pj(M)*(1-Fj(M));
    end

    if (j==1)
        var_exp_wtp=var_exp_wtp+Vj(j)*Pj(j)^2;
    else
        var_exp_wtp=var_exp_wtp+Vj(j)*(Pj(j)-Pj(j-1))^2;
    end
end
endfunction

```

El argumento de entrada `svy_data` le indica a la rutina que los datos muestrales en bruto están en forma matricial. Las variables globales `COL_BID` y `COL_ANS` permiten personalizar el índice de las columnas que contienen el precio ofertado y la respuesta dicotómica respectivamente. En caso de no definir un valor para estas variables, se asume por defecto que esta información se ubica en la primera y segunda columna de la matriz `svy_data`. La función realiza los procesos de depuración, ordenación y cálculo descritos en esta sección, y reporta una tabla de resumen en el argumento de salida `tbull_table`. La tabla contiene columnas con la información del precio ofertado P_j , el número de respuestas negativas al precio N_j , el

número de encuestados K_j , el estimador F_j y la varianza de este estimador V_j para cada sub-muestra. La mediana puede ser obtenida directamente de esta tabla por simple inspección. Finalmente, el procedimiento estima el valor esperado de la cota inferior de la *DAP* y su varianza y los reporta en los argumentos de salida `exp_wtp` y `var_exp_wtp` respectivamente.

Además de código mostrado para la obtención de los estimadores, se utilizan funciones adicionales para otras tareas. La siguiente función implementa una prueba de hipótesis t de igualdad de medias con tamaños muestrales distintos y varianzas desiguales basados en los estimadores de Turnbull calculados por el procedimiento anterior:

```
function [h, t, df]=tbul_ttest(wtp1, v_wtp1, n1, wtp2, v_wtp2, n2, conf)
    df=(v_wtp1/n1+v_wtp2/n2)^2/((v_wtp1/n1)^2/(n1-1)+(v_wtp2/n2)^2/(n2-1));
    crit=cdf('T',df,conf,1-conf);
    t=abs(wtp1-wtp2)/sqrt(v_wtp1+v_wtp2);
    h=(t>crit);
endfunction
```

El significado de los argumentos de entrada para esta rutina son obvios. El argumento de entrada `conf` permite especificar el nivel de confianza de la prueba y los argumentos de salida `t` y `df` permiten evaluar el estadístico t y los grados de libertad con los que se realiza la prueba. El argumento `h` es un marcador con valor 1 si no se rechaza la prueba de hipótesis y 0 si se la rechaza. Por último, la función detallada a continuación imprime en cualquier destino con identificador válido (*standard output*) la tabla Turnbull resultante en formato \LaTeX :

```
function []=tbul_to_latex(fid, table_data)
    fprintf(fid,'\n\begin{tabular}{|c c c c|}\n');
    fprintf(fid, '\t\\hline\n');
    fprintf(fid,'\t$P_j$ & $N_j$ & $T_j$ & $F_j$ & $V_j$ \\\n');
    fprintf(fid, '\t\\hline\n');
    n=size(table_data,1);
    for i=1:n
        fprintf(fid,'\t$%4.3f$ & $%4.3f$ & $%4.3f$ & $%4.3f$ & $%4.3f$ \\\n',table_data(i,:));
    end
    fprintf(fid, '\t\\hline\n');
    fprintf(fid, '\\end{tabular}\n');
    fprintf(fid,'\n');
endfunction
```

Para analizar el efecto que tienen los covariantes sobre la distribución, es necesario dividir cada sub-muestra en sub-clases de acuerdo a uno o más covariantes que estén disponibles en la muestra. Claro está que las sub-clases deben ser las misma para todas las sub-muestras, lo que permite que se separe la muestra original en dos o más sub-clases según el número de categorías que se obtengan de la conjunción de los covariantes. Posteriormente, se puede aplicar el método de Turnbull a cada sub-clase, obteniendo así las sub-muestras y sus estimadores agrupados por cada sub-clase. No obstante, se debe tener cuidado con ciertos aspectos al momento de dividir a la muestra de esta forma:

- Se debe delimitar claramente las sub-clases en las que se puede clasificar a los individuos encuestados. Esto implica que no se puede crear sub-clases a partir de covariantes continuos, a menos que estos se definan en intervalos no ambiguos.
- Se puede tener problemas si las sub-clases en las sub-muestras no cuentan con suficientes observaciones. Específicamente, puede que el método de Turnbull no sea aplicable para corregir problemas de monotonicidad o que los estimadores provenientes del mismo presenten varianzas relativamente grandes.

Las rutinas que realizan este análisis no serán mostradas debido a que son demasiado específicas del problema que se analizará, lo que las vuelve difícilmente reutilizables. Sin embargo, se puede aprovechar los procedimientos ya provistos para implementar análisis personalizados de este tipo sin mucha dificultad.

3. Análisis empírico de la metodología

Para mostrar la implementación empírica del modelo, se utilizará una base de datos proveniente de un estudio de valoración contingente para políticas públicas enfocadas a disminuir el riesgo de infección del VIH/SIDA, aplicada a una muestra representativa de la población general de la ciudad de Guayaquil. [2] Específicamente, se pregunta acerca de contribuciones hipotéticas a manera de impuestos anuales para políticas de acceso a medios de prevención (PPV) y tratamiento anti-retroviral para individuos

seropositivos (ARV). Cada individuo de la muestra recibe aleatoriamente una encuesta enfocada en una de ellas, en la cuál se le informa de manera breve sobre aspectos relevantes y los objetivos en los que se enmarca la política referida; se le pregunta por factores socioeconómicos y de comportamiento relevantes y finalmente, acerca de su valoración individual sobre la política en cuestión. El formato en el que se administra la encuesta es compatible con la metodología Turnbull, por lo que la implementación de la misma ayudará a determinar si la disponibilidad a pagar en ambos casos exhibe un comportamiento adecuado para la inferencia.

Uno de los factores de comportamiento por los que se pregunta en el estudio mencionado es la *percepción individual del riesgo de infección*. Este es el covariante cuyo efecto se analizará. Para ello se divide a las muestras generales tanto de ARV como PPV en dos sub-clases definidas por individuos que indiquen que su percepción de riesgo es nula (sub-clase *base*) y por individuos que se identifiquen con algún grado de riesgo percibido (sub-clase *otro*). Posteriormente, es posible aplicar los algoritmos descritos anteriormente (basados en el método de Turnbull) a estas sub-clases para poder determinar la distribución de la *DAP* de los individuos que las integran. En base a ello y a los diferentes estimadores que se obtengan se podrán efectuar conjeturas interesantes.

3.1. Resultados

La muestra de ARV cuenta con 584 observaciones válidas, mientras que la de PPV cuenta con 526. Al dividir la muestra en las dos sub-clases mencionadas, la sub-clase *base* tiene 295 observaciones válidas y la sub-clase *otro* tiene 201. El tamaño de cada sub-clase en la muestra PPV son 297 y 168 respectivamente. Estos detalles son importantes ya que se requieren muestras grandes para poder realizar cualquier tipo de inferencia válida, sea a nivel de muestra general o cualquiera de las sub-clases que se estudien.

Si se observa el comportamiento de la proporción muestral de respuestas negativas en cada una de las sub-muestras que se conforman para cada precio ofertado P_j , se puede apreciar claramente que existen problemas con el comportamiento de esta variable para ambas muestras:

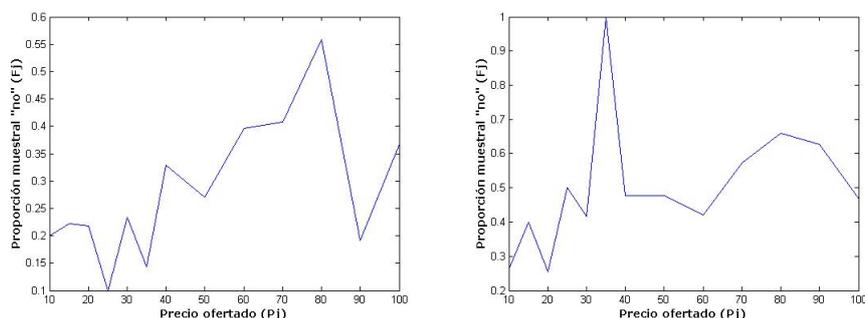


Figura 1: F_j para las muestras ARV y PPV (sin restricción)

Después de aplicar el método de Turnbull en ambas muestras el comportamiento de los estimadores F_j en ambos casos se corrige hasta resultar monótonicamente creciente:

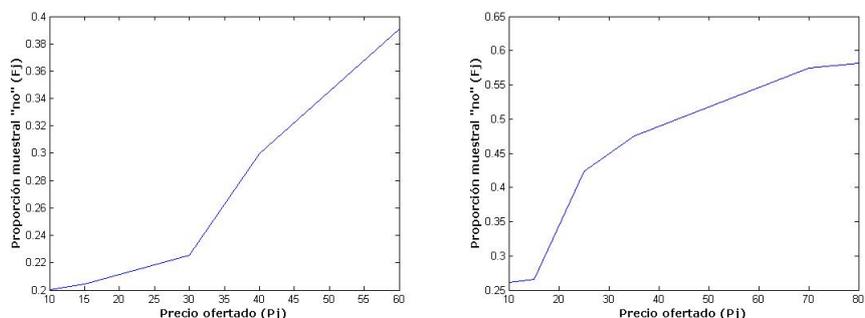


Figura 2: F_j para las muestras ARV y PPV (restringido)

Los resultados específicos de la aplicación del método de Turnbull para cada una de las muestras se muestran en los siguientes cuadros:

P_j	N_j	T_j	F_j	V_j
10,000	14,000	70,000	0,200	0,002
15,000	18,000	88,000	0,205	0,002
30,000	16,000	71,000	0,225	0,002
40,000	39,000	130,000	0,300	0,002
60,000	88,000	225,000	0,391	0,001

Cuadro 1: Tabla *Turnbull* para la muestra ARV

P_j	N_j	T_j	F_j	V_j
10,000	18,000	69,000	0,261	0,003
15,000	17,000	64,000	0,266	0,003
25,000	28,000	66,000	0,424	0,004
35,000	66,000	139,000	0,475	0,002
70,000	31,000	54,000	0,574	0,005
80,000	78,000	134,000	0,582	0,002

Cuadro 2: Tabla *Turnbull* para la muestra PPV

En dichas tablas se puede observar, de manera general, que la aplicación de este método no solo resuelve problemas de monotonicidad, sino que también ocasiona que las varianzas de los estimadores resultantes sea relativamente pequeña. En base a esto, se puede obtener estimadores de las medidas de tendencia central para la *DAP*. El cuadro (3) muestra los valores de estos estimadores y el resultado de la aplicación de una prueba *t* para igualdad de medias con tamaños muestrales distintos. Es evidente que el rango en el que se ubica la mediana de la muestra ARV no es observable. Esto señala que el vector de precios ofertados no recoge de manera adecuada las disponibilidades más altas en este caso. Sin embargo, aún es posible comparar ambas muestras en términos de la cota inferior de la *DAP* ya que esta medida no es sensible al problema mencionado. La prueba *t* indica que ambas muestras tienen cotas inferiores estadísticamente iguales. La implicación de este resultado es que existe evidencia para decir que la valoración de los individuos de la población para ambas políticas es por lo menos igual, siempre y cuando se los considere homogéneos. Claro está que no es posible afirmar nada acerca de si la valoración en sí es igual para ambas muestras a partir de este resultado, pero si es posible sugerir que ambas políticas son al menos igual de relevantes en términos de valoración.

	ARV	PPV
Mediana	-	35 - 70
$E_{LB}(DAP_i)$	42,775	41,159
$V[E_{LB}(DAP_i)]$	1,413	6,634
Grad. libertad	721,83	
Estad. <i>t</i>	0,567	

Cuadro 3: Medidas de tendencia central para la *DAP*

Un enfoque un poco más útil se deriva de considerar cierta heterogeneidad observable entre los individuos mediante el uso de covariantes. Si se separa las muestras de ARV y PPV por la percepción individual del riesgo de infección (en dos sub-clases según se definió anteriormente) es factible aplicar el método de Turnbull en cada una de ellas y efectuar una prueba *t* similar a la utilizada en el cuadro (3) para determinar si existe afectación en la *DAP* individual debido a la percepción de riesgo de infección. Los resultados para la muestra ARV son los siguientes:

ARV clase base					ARV clase otro				
P_j	N_j	T_j	F_j	V_j	P_j	N_j	T_j	F_j	V_j
10,000	7,000	37,000	0,189	0,004	10,000	11,000	68,000	0,162	0,002
15,000	9,000	47,000	0,191	0,003	40,000	7,000	27,000	0,259	0,007
30,000	11,000	39,000	0,282	0,005	50,000	8,000	30,000	0,267	0,007
40,000	18,000	56,000	0,321	0,004	60,000	20,000	71,000	0,282	0,003
60,000	13,000	31,000	0,419	0,008	100,000	2,000	5,000	0,400	0,048
70,000	37,000	85,000	0,435	0,003	-	-	-	-	-

Cuadro 4: Tabla *Turnbull* para covariantes de la muestra ARV

	ARV clase base	ARV clase otro
Mediana	-	-
$E_{LB}(DAP_i)$	46,966	69,121
$V[E_{LB}(DAP_i)]$	5,486	84,338
Grad. libertad	217,83	
Estad. t	2,338	

Cuadro 5: Medidas de tendencia central de la *DAP* para covariantes de la muestra ARV

Por otro lado, la muestra de PPV separada por las sub-clases mencionadas arroja los siguientes resultados:

PPV clase base					PPV clase otro				
P_j	N_j	T_j	F_j	V_j	P_j	N_j	T_j	F_j	V_j
10,000	24,000	74,000	0,324	0,003	10,000	5,000	39,000	0,128	0,003
25,000	45,000	108,000	0,417	0,002	25,000	10,000	28,000	0,357	0,008
70,000	69,000	115,000	0,600	0,002	40,000	55,000	101,000	0,545	0,002

Cuadro 6: Tabla *Turnbull* para la muestra PPV separada por subclases

	PPV clase base	PPV clase otro
Mediana	25 - 70	25 - 40
$E_{LB}(DAP_i)$	33,507	25,192
$V[E_{LB}(DAP_i)]$	5,029	2,684
Grad. libertad	433,71	
Estad. t	2,994	

Cuadro 7: Medidas de tendencia central de la *DAP* para covariantes de la muestra PPV

En general, se puede decir que las varianzas para los estimadores F_j se mantienen relativamente pequeñas en todos los casos. Sin embargo, el cuadro (5) indica que el problema con la muestra ARV es recurrente, ya que tampoco se puede observar el rango por el que está acotada la mediana de la distribución para ninguna de las sub-clases. Además, se puede observar que la varianza de la cota inferior de la *DAP* es significativamente mayor para la sub-clase con algún grado de percepción. Intuitivamente, este hecho se puede justificar dado que este estrato incluye individuos que tienen distintos grados de percepción, lo que puede causar el comportamiento mencionado en la muestra ARV. La razón para agrupar de este modo las sub-clases del covariante se debe a que no se cuenta con observaciones suficientes para hilar más fino en las mismas y tener tamaños muestrales satisfactoriamente grandes para sostener el análisis. Por último, las pruebas t para ambas muestras revelan que la cota inferior de la *DAP* es estadísticamente distinta entre las sub-clases. Esto quiere decir que se puede sugerir que la percepción individual de riesgo de infección afecta la valoración individual de ambas políticas.

4. Conclusiones

El objetivo principal del presente trabajo es brindar una implementación consistente y fácil de extender del método de *Turnbull*, aplicado en el contexto de la valoración contingente. Además, las rutinas

mostradas intentan resolver problemas comunes de tratamiento de datos en el trabajo aplicado. Se escogió desarrollar dicha implementación en SCILAB 4.1, debido a las facilidades que presta este entorno de desarrollo. Al ser de arquitectura abierta y orientado a matrices, este permite la creación de herramientas personalizadas de análisis, muchas veces no disponibles en los paquetes informáticos tradicionales. Por último, las ventajas de tratar la información de manera matricial son obvias, ya que no todos los paquetes de análisis permiten el tratamiento directo de la información muestral compatible con el formato necesario para el análisis de valoración en general.

Respecto a la aplicación del método de Turnbull en el estudio sobre valoración de políticas, se observó que este permite realizar conjeturas interesantes. Permite corregir los problemas de monotonicidad que se presentaron en las muestras sin imponer problemas aparentes, mostrando que no es necesario adoptar supuestos específicos sobre el comportamiento de la *DAP* para poder realizar inferencia. Adicionalmente, la metodología permite diagnosticar ciertos problemas que pueden surgir al momento obtener dicha medida, como se mencionó en la discusión de los resultados. Finalmente, se mostró que los covariantes observables pueden tener efectos sobre el comportamiento de la *DAP* de los agentes. Sin embargo, este enfoque no permite explorar en profundidad dicha influencia, ya que para ello sería necesaria la construcción de modelos de comportamiento más elaborados. No obstante, el marco de trabajo no deja de ser un buen punto de partida para el análisis del comportamiento de la *DAP*.

Referencias

- [1] Haab, T.C. and McConnell, K.E. (1997), "*Referendum Models and Negative Willingness To Pay: Alternative Solutions*", American Journal of Agricultural Economics, **78**, 251-270.
- [2] Fernandez Cadena, M. y Lemus Sares, D. (2007), "*Evaluación de la Calidad de Vida Personas Viviendo con VIH/SIDA (PVVS) y Valoración Contingente de los Beneficios de las Intervenciones Públicas en la Epidemia del VIH/SIDA: Prevención de la Infección de VIH y Entrega Gratuita de Terapia Antirretroviral (ARV)*", Universidad de Guayaquil - Escuela Superior Politécnica del Litoral - Fundación Kimirina, Working Paper.
- [3] Turnbull, B. (1976), "*The empirical Distribution Function with Arbitrary Grouped, Censored and Truncated Data*", Journal of the Royal Statistical Society, **38B**, 290-295.

A. Apéndice: Derivación de los estimadores de Turnbull

Considerando que todas las M sub-muestras fueron creadas aleatoriamente, se podría tratar a cada una de ellas como independiente y estimar la probabilidad de observar una respuesta negativa (F_j). En principio, es posible asumir una distribución para F_j , pero hacerlo implica imponer supuestos que no son necesarios para estimarlo. Si se define una variable dicotómica R_{ij} igual a 1 cuando el individuo i responda *sí* al precio P_j y 0 cuando responda *no*, la probabilidad condicional de observar R_{ij} dado F_j sigue una distribución *Bernoulli* con probabilidad de éxito $(1 - F_j)$:

$$Pr(R_{ij}|F_j) = F_j^{1-R_{ij}}(1 - F_j)^{R_{ij}} \quad (1)$$

Assumiendo que todos los individuos de que conforman cada sub-muestra son idénticos,³ entonces la probabilidad conjunta de observar $R_j = \{R_{1j}, R_{2j}, \dots, R_{K_jj}\}$ para cada sub-muestra es:

$$Pr(R_j|F_j, K_j) = \prod_{i=1}^{K_j} F_j^{1-R_{ij}}(1 - F_j)^{R_{ij}} \quad (2)$$

$$Pr(R_j|F_j, K_j) = F_j^{\sum_{i=1}^{K_j} (1-R_{ij})} (1 - F_j)^{\sum_{i=1}^{K_j} R_{ij}}$$

Dado que $\sum_{i=1}^{K_j} (1 - R_{ij}) = N_j$ y que $\sum_{i=1}^{K_j} R_{ij} = Y_j$, se puede reescribir la expresión anterior de la siguiente manera:

$$Pr(R_j|F_j, Y_j, N_j) = F_j^{N_j} (1 - F_j)^{Y_j} \quad (3)$$

Se puede estimar F_j en base a (3) para cada sub-muestra por el método de *máxima verosimilitud*. La idea detrás es estimar el valor de F_j que maximice la probabilidad de ocurrencia de R_j dado que se conocen Y_j y N_j . Si se asume que cada uno de estos estratos son independientes debido a como fueron estructurados, la *función de verosimilitud* para la muestra entera es:

$$L(F_j|Y_1, \dots, Y_M; N_1, \dots, N_M) = \prod_{j=1}^M Pr(R_j|F_j, Y_j, N_j)$$

$$L(F_j|Y_1, \dots, Y_M; N_1, \dots, N_M) = \prod_{j=1}^M F_j^{N_j} (1 - F_j)^{Y_j} \quad (4)$$

Las propiedades de este método permiten expresar el problema de maximización una manera más sencilla, en donde las condiciones de primer orden permiten obtener el estimador de máxima verosimilitud para F_j :

$$\max \ln L = \sum_{j=1}^M \left[N_j \ln(F_j) + Y_j \ln(1 - F_j) \right] \quad (5)$$

$$\frac{\partial \ln L}{\partial F_j} = \frac{N_j}{F_j} - \frac{Y_j}{1 - F_j} = 0; \forall j = [1, \dots, M] \quad (6)$$

$$\widetilde{F}_j = \frac{N_j}{N_j + Y_j} = \frac{N_j}{K_j} \quad (7)$$

Si la muestra es suficientemente grande, es de esperar que F_j incremente monotónicamente a medida que el precio P_j al que se enfrenten los encuestados sea mayor. Sin embargo, no hay nada en el procedimiento anterior que garantice que esto ocurra.

Si se impone al proceso de maximización la restricción $F_j \leq F_{j+1}$ para todo j , entonces ya no se puede tratar a cada sub-muestra como independiente, por lo que es conveniente re-exresar el problema de máxima verosimilitud en términos de la forma discreta de la función de densidad de probabilidad $f = \{f_1, f_2, \dots, f_j, \dots, f_{M+1}\}$, en donde $f_j = F_j - F_{j-1}$ puede ser interpretado como respuesta al

³Este supuesto se levantará posteriormente para analizar el efecto que tienen los *covariantes* en el análisis.

incremento de precios entre el precio P_{j-1} y P_j . Dado que $F_j = \sum_{i=1}^j f_i$, $F_0 = 0$ (no se consideran *DAP* negativas) y $F_{M+1} = 1$, el problema de optimización restringida queda expresado de la siguiente manera:

$$\max \ln L(f|Y_1, \dots, Y_M; N_1, \dots, N_M) = \sum_{j=1}^M \left[N_j \ln \left(\sum_{i=1}^j f_i \right) + Y_j \ln \left(1 - \sum_{i=1}^j f_i \right) \right] \quad (8)$$

sujeito a $f_i \geq 0 \forall j$

Las condiciones de Kuhn-Tucker para el máximo son:

$$\frac{\partial \ln L}{\partial f_i} = \sum_{j=i}^M \left(\frac{N_j}{\sum_{i=1}^j f_i} + \frac{Y_j}{1 - \sum_{i=1}^j f_i} \right) \leq 0 ; \forall j = \{1, \dots, M\} \quad (9)$$

$$f_i \frac{\partial \ln L}{\partial f_i} = 0 \quad (10)$$

$$f_i \geq 0 \quad (11)$$

Para hallar los estimadores, este sistema de inecuaciones debe ser resuelto de manera recursiva. Por construcción, se puede asegurar que $f_1 > 0$ siempre y cuando $N_1 \neq 0$. En este caso la condición (10) garantiza que la condición (9) se cumpla con igualdad para f_1 . Suponiendo por un momento que $f_2 > 0$, la condición (9) también se cumple con igualdad para f_2 , por lo que:

$$\frac{\partial \ln L}{\partial f_1} - \frac{\partial \ln L}{\partial f_2} = \frac{N_1}{f_1} - \frac{Y_1}{1 - f_1} = 0 \quad (12)$$

$$f_1 = \frac{N_1}{N_1 + Y_1} = \frac{N_1}{K_1} \quad (13)$$

De igual manera, si $f_3 > 0$, se puede restar $\frac{\partial \ln L}{\partial f_2}$ de $\frac{\partial \ln L}{\partial f_3}$ para obtener:

$$f_2 = \frac{N_2}{K_2} - f_1 \quad (14)$$

Por tanto, para asegurar que f_2 sea efectivamente mayor que cero, tiene que cumplirse que:

$$\frac{N_2}{K_2} > \frac{N_1}{K_1} \quad (15)$$

Si se continúa de manera iterativa para la M sub-muestras, y se puede garantizar que se cumpla una condición similar a (15) que asegure que la proporción de *no* incremente monotónicamente para cada una de ellas, entonces los estimadores de máxima verosimilitud para f_j y F_j son:

$$\tilde{f}_j = \frac{N_j}{K_j} - \sum_{i=1}^{j-1} f_i = \frac{N_j}{K_j} - \frac{N_{j-1}}{K_{j-1}} \quad (16)$$

$$\tilde{F}_j = \sum_{i=1}^j f_i = \frac{N_j}{K_j} \quad (17)$$

Los resultados (16) y (17) corroboran el resultado obtenido en (7) e indican simplemente que el mejor estimador para la probabilidad de observar un *no* es la proporción muestral de individuos que no están dispuestos a pagar un precio determinado, siempre y cuando dichas proporciones sean monotónicamente crecientes. Dado esto, la probabilidad que la *DAP* se encuentre entre dos precios adyacentes es la diferencia entre las proporciones muestrales de gente que *no* esta dispuesta a pagar ambos precios.

Esta claro que todo el análisis anterior depende de la condición:

$$\frac{N_j}{K_j} \leq \frac{N_{j+1}}{K_{j+1}} \forall j = [1, \dots, M] \quad (18)$$

Para mostrar una manera de garantizarla, se puede suponer por ejemplo que $\frac{N_4}{K_4} > \frac{N_5}{K_5}$. Esto causaría que $f_4 = 0$ debido a la condición de primer orden (11), por lo que $\frac{\partial \ln L}{\partial f_4}$ se volvería irrelevante según (10). Asumiendo momentáneamente que $f_5 > 0$ se podría decir que:

$$\frac{\partial \ln L}{\partial f_3} - \frac{\partial \ln L}{\partial f_5} = \frac{N_3 + N_4}{\sum_{i=1}^2 f_i + f_3} - \frac{Y_3 + Y_4}{1 - \sum_{i=1}^2 f_i - f_3} = 0$$

$$\widehat{f}_3 = \frac{N_3 + N_4}{K_3 + K_4} - \sum_{i=1}^2 f_i \quad (19)$$

Si por el contrario, \widehat{f}_3 continúa siendo negativo, se puede repetir el mismo procedimiento recursivamente hasta garantizar que este sea estrictamente positivo. El proceso general aplicado a todo $j = [2, \dots, M]$ es conocido como *price pooling* y asegura que se cumpla la condición (18), aún cuando $\frac{N_{j+1}}{K_{j+1}} > \frac{N_{j+2}}{K_{j+2}}$, simplemente descartando el precio P_{j+1} y apilando los elementos de la sub-muestra $j+1$ en la sub-muestra j . Si se define a:

$$\widehat{N}_j = \begin{cases} \frac{N_j}{K_j} & \text{si } \frac{N_{j+1}}{K_{j+1}} \leq \frac{N_{j+2}}{K_{j+2}} \\ \frac{N_j + N_{j+1}}{K_j + K_{j+1}} & \text{si } \frac{N_{j+1}}{K_{j+1}} > \frac{N_{j+2}}{K_{j+2}} \end{cases} \quad (20)$$

$$\widehat{K}_j = \begin{cases} Y_j + N_j & \text{si } \frac{N_{j+1}}{K_{j+1}} \leq \frac{N_{j+2}}{K_{j+2}} \\ Y_j + Y_{j+1} + N_j + N_{j+1} & \text{si } \frac{N_{j+1}}{K_{j+1}} > \frac{N_{j+2}}{K_{j+2}} \end{cases} \quad (21)$$

Entonces, los *estimadores de Turnbull* para todo $j = [1, 2, \dots, \widehat{M}]$ resultante del procedimiento descrito son:

$$\widehat{f}_j = \frac{\widehat{N}_j}{\widehat{K}_j} - \sum_{i=1}^{j-1} \widehat{f}_i \quad (22)$$

$$\widehat{F}_j = \frac{\widehat{N}_j}{\widehat{K}_j} \quad (23)$$

Una vez completado el proceso de *pooling* de manera que se pueda garantizar que no existan violaciones a la condición de monotonidad, se puede tratar a los estimadores de Turnbull como provenientes del problema de optimización no restringida original (5) y por tanto cada uno de los \widehat{M} estimadores \widehat{F}_j para la distribución acumulada considerado independiente. Es posible obtener estimadores para la cota mínima de las varianzas y covarianzas de los estimadores Turnbull a partir de la *matriz de información* que se obtiene de las condiciones de segundo orden del problema. La matriz de varianzas y covarianzas de \widehat{F}_j viene dada por:

$$V(\widehat{F}_j) = \left[-\frac{\partial^2 \ln L}{\partial F_j^2} \right]^{-1} = \left[\frac{\widehat{N}_j}{\widehat{F}_j^2} - \frac{\widehat{Y}_j}{(1 - \widehat{F}_j)^2} \right]^{-1} = \frac{\widehat{F}_j(1 - \widehat{F}_j)}{\widehat{K}_j} ; \forall j = [1, 2, \dots, \widehat{M}] \quad (24)$$

$$\text{cov}(\widehat{F}_i, \widehat{F}_j) = 0 ; \forall i \neq j \quad (25)$$

Recordando que $\widehat{f}_j = \widehat{F}_j - \widehat{F}_{j-1}$, la varianza para el estimador de densidad de probabilidad es:

$$V(\widehat{f}_j) = V(\widehat{F}_j) - V(\widehat{F}_{j-1}) = \frac{\widehat{F}_j(1 - \widehat{F}_j)}{\widehat{K}_j} - \frac{\widehat{F}_{j-1}(1 - \widehat{F}_{j-1})}{\widehat{K}_{j-1}} \quad (26)$$

La covarianza de este viene dada por:

$$\text{cov}(\widehat{f}_i, \widehat{f}_j) = \text{cov}(\widehat{F}_i - \widehat{F}_{i-1}, \widehat{F}_j - \widehat{F}_{j-1})$$

$$\text{cov}(\widehat{f}_i, \widehat{f}_j) = \text{cov}(\widehat{F}_i, \widehat{F}_j) - \text{cov}(\widehat{F}_i, \widehat{F}_{j-1}) - \text{cov}(\widehat{F}_{i-1}, \widehat{F}_j) + \text{cov}(\widehat{F}_{i-1}, \widehat{F}_{j-1})$$

$$\text{cov}(\widehat{f}_i, \widehat{f}_j) = \begin{cases} -V(\widehat{F}_j) & j - 1 = i \\ -V(\widehat{F}_i) & i - 1 = j \\ 0 & \text{cualquier otro caso} \end{cases} \quad (27)$$

Los estimadores resultantes de este proceso muestran propiedades estadísticas interesantes, siempre y cuando se cuente con muestras suficientemente grandes. En tal caso, los estimadores para la distribución acumulada \widehat{F}_j son estrictamente crecientes e independientes, de modo que se puede tratar al set de respuestas a cada precio j como una serie de experimentos independientes que siguen una distribución binomial⁴. Esto permite que los estimadores \widehat{f}_j sean variables aleatorias con distribución asintóticamente normal.

A.1. Medidas de tendencia central para la DAP

Una medida de tendencia central en base a los estimadores resultantes de la aplicación el método de Turnbull es la *mediana* de la distribución de la DAP. Por construcción, este método ordena cada una de las sub-muestras de manera ascendente según el orden del \widehat{P}_j que le corresponde y asegura que las proporciones \widehat{F}_j (distribución acumulada) de cada sub-muestra sean monótonicamente crecientes. Por ende, la mediana de la distribución se encontrará entre los precios \widehat{P}_j y \widehat{P}_{j+1} para los cuales el valor de \widehat{F}_j sea inmediatamente inferior a 0,5 y el de \widehat{F}_{j+1} sea inmediatamente superior a dicho valor. Claro está que no es posible hallar un estimador puntual para la mediana, debido al formato de datos con el que trabaja el método en cuestión. Sin embargo, este define el rango por el que esta medida se encuentra acotada.

Además de la mediana, es posible encontrar otro tipo de medida de tendencia central para la DAP. Su *esperanza matemática* se puede obtener asumiendo que la DAP de cualquier individuo es una variable aleatoria definida en el intervalo $[0, U]$. El valor esperado de la misma en términos su distribución acumulada de probabilidad es:

$$E(DAP_i) = \int_0^U X \partial\Pr(DAP_i \leq X) \quad (28)$$

donde el límite superior U no es conocido. Se puede particionar (30) en $\widehat{M} + 1$ sub-rangos $[0 - \widehat{P}_1, \widehat{P}_2 - \widehat{P}_1, \dots, \widehat{P}_{\widehat{M}} - U]$ de acuerdo al precio ofertado para cada sub-muestra. La expresión resultante es:

$$E(DAP_i) = \sum_{j=0}^{\widehat{M}} \left[\int_{\widehat{P}_j}^{\widehat{P}_{j+1}} X \partial\Pr(DAP_i \leq X) \right] \quad (29)$$

en donde $\widehat{P}_0 = 0$ y $\widehat{P}_{\widehat{M}+1} = U$. Dado que $\partial\Pr(DAP_i \leq X)$ es una función estrictamente creciente, se puede reemplazar la variable de integración X por el límite de integración inferior para obtener:

$$E(DAP_i) \geq \sum_{j=0}^{\widehat{M}} \left[\int_{\widehat{P}_j}^{\widehat{P}_{j+1}} \widehat{P}_j \partial\Pr(DAP_i \leq \widehat{P}_j) \right] \quad (30)$$

Es decir, no se puede calcular el valor esperado de la DAP sin conocer el valor de U . Sin embargo, se puede obtener un estimador para la *cota inferior* de la DAP si en (30) se sabe que $\int_{\widehat{P}_j}^{\widehat{P}_{j+1}} \widehat{P}_j \partial\Pr(DAP_i \leq \widehat{P}_j) = \widehat{P}_j [\Pr(DAP_i \leq \widehat{P}_{j+1}) - \Pr(DAP_i \leq \widehat{P}_j)]$. Utilizando esta condición y reemplazando los estimadores de Turnbull \widehat{F}_j de cada sub-muestra se obtiene:

$$E_{LB}(DAP_i) = \sum_{j=0}^{\widehat{M}} \widehat{P}_j [\widehat{F}_{j+1} - \widehat{F}_j] \quad (31)$$

Para calcular el valor de este estimador, es necesario recordar que $\widehat{F}_0 = 0$ y $\widehat{F}_{\widehat{M}+1} = 1$. Adicionalmente, el resultado (31) permite expresar $E_{LB}(DAP_i)$ en función de los estimadores \widehat{f}_j :

$$E_{LB}(DAP_i) = \sum_{j=0}^{\widehat{M}} \widehat{P}_j \widehat{f}_{j+1} \quad (32)$$

⁴Se puede demostrar que la expresión (2) puede escribirse como una distribución binomial sin afectar el análisis del proceso de máxima verosimilitud.

por lo que la varianza para este estimador esta dada por:

$$\begin{aligned}
V[E_{LB}(DAP_i)] &= \sum_{i=1}^{\widehat{M}} \sum_{j=1}^{\widehat{M}} \widehat{P}_i \widehat{P}_j \text{cov}(\widehat{f}_{i+1}, \widehat{f}_{j+1}) = \sum_{j=1}^{\widehat{M}} (\widehat{P}_j - \widehat{P}_{j-1})^2 V(\widehat{F}_j) \\
V[E_{LB}(DAP_i)] &= \sum_{j=1}^{\widehat{M}} \frac{\widehat{F}_j(1 - \widehat{F}_j)}{\widehat{K}_j} (\widehat{P}_j - \widehat{P}_{j-1})^2 \tag{33}
\end{aligned}$$

A pesar de ser una medida conservadora, el estimador para la cota mínima de la DAP es interesante ya que no es necesario asumir nada acerca de la distribución de probabilidad de la DAP , salvo no considerar disponibilidades negativas. Además, muestra propiedades útiles que permiten realizar pruebas de hipótesis estándar e intervalos de confianza respecto a la DAP . Dado que el vector de precios esta predefinido, la expresión (32) define a $E_{LB}(DAP_i)$ como una combinación lineal de \widehat{f}_j . Por tanto, las propiedades de normalidad asintótica de \widehat{f}_j permiten establecer la distribución de este estimador:

$$E_{LB}(DAP_i) \sim N \left(\sum_{j=0}^{\widehat{M}} \widehat{P}_j [\widehat{F}_{j+1} - \widehat{F}_j], \sum_{j=1}^{\widehat{M}} \frac{\widehat{F}_j(1 - \widehat{F}_j)}{\widehat{K}_j} (\widehat{P}_j - \widehat{P}_{j-1})^2 \right) \tag{34}$$

A.2. Efectos de covariantes

Hasta el momento, se ha mantenido el supuesto que todos los individuos en la muestra son idénticos. Esto ha permitido el planteamiento y desarrollo del método de Turnbull. Sin embargo, este supuesto no es muy práctico ya que resulta interesante observar el efecto que diversos factores observables puedan tener sobre el comportamiento de la DAP . Como se mencionó anteriormente, estos factores serán referidos como *covariantes*.

El problema original consistía en dividir la muestra general en M sub-muestras, en donde cada una recibe el mismo precio ofertado P_j . Si se puede separar cada una de ellas en *subclases* adicionales $h = 1, \dots, H$, según las características individuales observadas de cada encuestado, el problema (5) puede ser re-escrito como:

$$\begin{aligned}
&\max \ln(F_{jh} | N_{jh}, Y_{jh}, K_{jh}) \quad \forall j, k \\
&\sum_{h=1}^M \sum_{j=1}^M \left[N_{jh} \ln(F_{jh}) + Y_{jh} \ln(1 - F_{jh}) \right] \tag{35}
\end{aligned}$$

En donde las H subclases formadas a partir de uno o más covariantes son las mismas e independientes a través de cada una de las M sub-muestras. Debido a esto, el problema (35) puede ser tratado como H problemas de maximización separados y aplicar el método descrito en las dos primeras secciones para cada uno de ellos.