COMPARACIÓN DE MODELOS EN EL ANÁLISIS DEL NÚMERO DE TIPOS DE VINO CONSUMIDOS

Víctor J. Cano Fernández
Departamento de Economía de las Instituciones, Estadística Económica y Econometría
Universidad de La Laguna
vcano@ull.es

Ginés Guirao Pérez
Departamento de Economía de las Instituciones, Estadística Económica y Econometría
Universidad de La Laguna
gguirao@ull.es

María Carolina Rodríguez Donate
Departamento de Economía de las Instituciones, Estadística Económica y Econometría
Universidad de La Laguna
cdonate@ull.es

Margarita Esther Romero Rodríguez
Departamento de Economía de las Instituciones, Estadística Económica y Econometría
Universidad de La Laguna
mromero@ull.es

Resumen

En la literatura econométrica se han propuesto distintos modelos alternativos para el análisis de los denominados datos de recuento que permiten recoger de forma adecuada la naturaleza discreta y no negativa de algunos fenómenos de interés y pueden ser útiles para explicar ciertas estructuras de preferencias de los individuos.

En este trabajo, se lleva a cabo el análisis del número de tipos de vino consumidos por los residentes de Tenerife, con el fin de observar qué rasgos determinan la mayor o menor exclusividad en el consumo, en el contexto actual de creciente competencia en el sector. Las características específicas de la variable considerada, permiten que el análisis realizado cubra dos aspectos. Uno metodológico, implicado por la variedad de modelos que pueden plantearse, y que consiste en la comparación de las diferentes propuestas en su ajuste al tipo de recuento que se trata. Y otro, eminentemente empírico, que reside en describir no sólo el mecanismo de decisión más adecuado sino en identificar qué factores son los que explican la mayor o menor diversidad del consumo de vino.

Palabras clave: Modelos de regresión de Poisson, binomial negativo, clase latente, número tipos de vino.

Área temática: Economía Agraria y Recursos Naturales.

1. Introducción.

La creciente competencia en el mercado del vino, junto a los rasgos que definen la evolución reciente del consumo –la disminución del consumo per-cápita de vino, así como, el aumento de la demanda de vinos de calidad–, ha motivado la necesidad de estudios detallados sobre el comportamiento del consumidor. Éstos no sólo tratan de explicar la cantidad o la frecuencia del consumo, sino también las formas y modos en que éste se lleva a cabo. Las características que definen el lugar de compra o consumo, el origen o región de procedencia, la marca o los tipos de vino, pueden ser importantes para ayudar al desarrollo de estrategias de diferenciación adecuadas en las circunstancias actuales[1]. Algunas investigaciones apuntan que es difícil una prospectiva económica adecuada en el sector del vino sin un análisis detallado de la demanda. En cualquier caso, las posibilidades de estudio del comportamiento del consumidor son bastante amplias, habiendo ocupado un espacio importante el análisis de la fidelidad (a marcas, establecimientos,...)[2].

En este trabajo, dadas las limitaciones de la información disponible, no afrontamos el tema en su estricta dimensión, si bien adoptamos una óptica que al menos permite observar un comportamiento de exclusividad o diversidad en el consumo. Específicamente, se trata de describir los rasgos socioeconómicos que mejor definen el número de tipos de vino diferentes consumidos por los individuos, a partir de una encuesta realizada a residentes de la isla de Tenerife.

Dada la naturaleza de la variable que intenta explicarse, se hace necesario acudir a un tipo especial de modelos que recojan adecuadamente las características del fenómeno, esto es, su naturaleza discreta y no negativa. En la literatura econométrica se han propuesto distintas alternativas que pueden englobarse bajo la denominación de *modelos de regresión para datos de recuento*, aunque a veces se hace referencia a ellos como *modelos de regresión de Poisson*, que permiten afrontar este tipo de análisis, ofreciendo, además algunas alternativas para definir la estructura de preferencias de los consumidores.

[1] Véanse, por ejemplo, Sánchez y Gil (1998), Bello y Gómez (1996), Bernabéu et al (2001) y Martínez Carrasco, et al. (2004).
[2] Véase Flavián et al (1997).

Las características específicas de la variable considerada, permiten que el análisis realizado cubra dos aspectos. Uno metodológico, implicado por la variedad de modelos que pueden plantearse, y que consiste en la comparación de las diferentes propuestas en su ajuste al tipo de recuento que se trata. Y otro, eminentemente empírico, que reside en describir no sólo el mecanismo de decisión más adecuado sino en identificar qué factores son los que explican la mayor o menor diversidad del consumo de vino.

Con estos objetivos, el trabajo se estructura en tres apartados. En el primero, se desarrolla la metodología utilizada, en el que se presentan los distintos modelos, destacándose sus rasgos más sobresalientes. En el siguiente, se describen los datos y los resultados obtenidos en el proceso de estimación y contraste de los modelos. Y finalmente, se presentan las conclusiones más relevantes.

2. Metodología.

2.1. Los modelos estándar para datos de recuento.

El número de tipos de vino es una variable de naturaleza discreta y sólo toma valores enteros positivos, que oscilan entre el valor cero (no consumidores) y el valor siete (consumidores de todos los tipos de vino considerados). Para su análisis, en general, no es adecuado utilizar el modelo de regresión lineal, ya que ignora la especial naturaleza de esta variable y puede conducir a estimadores sesgados, ineficientes e inconsistentes.

Los modelos que se utilizan habitualmente en el análisis de este tipo de variables son los denominados modelos de regresión para datos de recuento. El modelo de recuento por excelencia es el *modelo de Poisson*, en el que la probabilidad de cada recuento se determina a partir de la distribución de Poisson, cuya media, λ_i, es función de un conjunto de variables explicativas, x_i, esto es:

$$P(Y_i = y_i / x_i) = \frac{e^{-\lambda_i}\ \lambda_i^{y_i}}{y_i!}, \quad y_i = 0,1,2,...$$

$$\lambda_i = E[y_i / x_i] = exp(x'_i\ \beta)^{3}.$$

Las variables explicativas del modelo recogen la denominada *heterogeneidad observada*, ya que la única fuente de diferencia entre los individuos se atribuye a los distintos valores observados de las variables explicativas, es decir, diferentes valores de

[3] Nótese que la formulación log-lineal del parámetro λ_i garantiza la naturaleza positiva de la media de la variable dependiente.

estas variables conducen a diferentes valores de λ, mientras que para todos los individuos con los mismos valores de las variables explicativas, el valor de λ es el mismo.

La igualdad de la media y varianza condicionales constituye una de las principales características del modelo de Poisson e implica su naturaleza heterocedástica. Sin embargo, también se convierte en uno de sus principales inconvenientes, ya que impide captar la posible *sobredispersión*[4] que suele estar presente en muchos datos.

La búsqueda de mayor flexibilidad ha propiciado la aparición de otros modelos -algunos de los cuales también están basados en la distribución de Poisson- que han recogido mejor la sobredispersión y otras características, como el exceso de ceros o la existencia de grandes colas a la derecha, consideradas como implicaciones de la *heterogeneidad no observada*[5] (Mullahy, 1997).

Habitualmente, esta heterogeneidad no observada se recoge introduciendo un término de error multiplicativo, v_i, en la media condicional del modelo de Poisson, dando lugar, así, a los *modelos de Poisson mixtos o compuestos*, cuya media es una variable aleatoria:

$$\lambda_i^* = E[y_i / x_i, v_i] = \lambda_i \ v_i = e^{x'_i \beta} \ e^{\varepsilon_i} .$$

Uno de estos modelos es el modelo binomial negativo[6]. Su representación como modelo de Poisson compuesto se consigue bajo el supuesto de que el término de heterogeneidad no observada[7], v_i, se distribuye como gamma ($\Gamma(\delta, \delta)$) con $\sigma_{v_i}^2 = 1/\delta \equiv \alpha$ (parámetro de dispersión) y $E[v_i] = 1$, lo que conduce a la distribución de probabilidad binomial negativa[8]:

$$P(Y_i = y_i / x_i) = \frac{\Gamma(\alpha^{-1} + y_i)}{\Gamma(\alpha^{-1})\Gamma(y_i + 1)} \left(\frac{\alpha^{-1}}{\alpha^{-1} + \lambda_i} \right)^{\alpha^{-1}} \left(\frac{\lambda_i}{\alpha^{-1} + \lambda_i} \right)^{y_i} ,$$

con media y varianza de la forma:

---

4 La sobredispersión implica que la varianza condicional excede a la media condicional.

5 El problema de la heterogeneidad no medida surge en aplicaciones en las que las diferencias de comportamiento entre individuos no pueden ser adecuadamente capturadas por el conjunto de variables explicativas de la función media condicional del modelo.

6 Este tipo de modelo también puede ser motivado de diferentes formas. Véase Boswell y Patil (1970).

7 Este término puede recoger un error de especificación, como la omisión de alguna variable explicativa (Gourieroux et al, 1984a, b) o bien la aleatoriedad intrínseca del proceso (Hausmann et al, 1984).

8 Esta función de probabilidad corresponde concretamente al modelo denominado NEGBIN II.

$$E[y_i / x_i] = \lambda_i,$$

$$V[y_i / x_i] = \lambda_i (1 + \alpha \lambda_i).$$

Cameron y Trivedi (1986) sugieren un modelo más general, el NEGBIN k con varianza $V[y_i / x_i] = \lambda_i + \alpha \lambda_i^{2-k}$. Si $k = 0$, el modelo resultante es el NEGBIN II y $k = 1$ conduce al denominado modelo NEGBIN I.

Entre las principales limitaciones de los modelos estándar planteados se encuentran, por un lado, que en ellos se supone que el recuento nulo y los recuentos positivos presentan el mismo proceso generador de datos, esto es, en el caso que estamos analizando, suponer una misma estructura de preferencia para consumidores y no consumidores. Y por otro lado, muestran una estructura un tanto rígida en relación a los posibles supuestos que se realizan sobre la heterogeneidad no observada, y que se basa en distribuciones paramétricas conocidas. Conviene, por tanto, tomar en consideración otros modelos que proporcionen una mayor flexibilidad en relación a estas cuestiones.

2.2. Modelos con obstáculo o en dos partes.

Una de las modificaciones considerada habitualmente, y que ha llevado a un buen número de aplicaciones, la constituyen los denominados modelos con obstáculo o en dos partes. Estos fueron desarrollados por Cragg (1971) para datos continuos en el contexto de los modelos Tobit, sin embargo, resultan ser especialmente útiles para el análisis de datos de recuento, si en ellos se observa un número considerable de ceros e incluso, un número escaso de ceros en comparación con el modelo de Poisson (Winkelmann y Zimmermann, 1995)[9].

Los primeros desarrollos de este tipo de modelos para datos de recuento son descritos en Johnson y Kotz (1969). Sin embargo el primer tratamiento económico, explicando los efectos de la regresión, es de Mullahy (1986).

Una de las principales características de estos modelos es que tienen en cuenta el proceso de toma de decisiones del individuo. En un primer estadio, el individuo decide si quiere, por ejemplo, consumir o no un producto, y en un segundo estadio, una vez que ha tomado la decisión afirmativa, decide, qué cantidad consume. Como consecuencia de

[9] En la base de datos considerada en el presente trabajo el número de ceros representa un no despreciable 24% del total de observaciones.

ello, la probabilidad del resultado o respuesta cero se modeliza separadamente de la probabilidad de un resultado positivo, esto es:

$$P(Y=0)=f_1(0)$$
$$P(Y=y)=f_2(y)\,\frac{1-f_1(0)}{1-f_2(0)}=\phi\, f_2(y)\,, \qquad y=1,2,\ldots$$

donde $f_1(\cdot)$ y $f_2(\cdot)$ representan distribuciones de probabilidad, que pueden ser o no del mismo tipo, $1\text{-}f_1(0)$ recoge la probabilidad de cruzar el primer estadio de decisión y $1\text{-}f_2(0)$ se aplica para truncar en el cero la segunda distribución. Estos modelos pueden ser vistos como de mezcla finita frente al Binomial Negativo, que supone una mezcla de variables aleatorias continuas, por la combinación de los ceros generados por una distribución y los valores positivos generados por una distribución truncada (Cameron y Trivedi, 1998).

Mullahy (1986) propuso dos distribuciones de Poisson con parámetros $\lambda_1=e^{x'_i\,\beta_1}$ y $\lambda_2=e^{x'_i\,\beta_2}$ para $f_1(\cdot)$ y $f_2(\cdot)$, respectivamente[10], dando lugar, así, al *modelo de Poisson con obstáculo*, también conocido como modelo de Poisson con obstáculo sencillo, que colapsa al modelo de Poisson estándar si ambos conjuntos de parámetros coinciden[11]:

$$P(y_i=0)=e^{-\lambda_{1,i}}$$
$$P(y_i=j)=\frac{(1-e^{-\lambda_{1,i}})\,e^{-\lambda_{2,i}}\;\lambda_{2,i}^{j}}{j!\,(1-e^{-\lambda_{2,i}})}\,, \qquad j=1,2,\ldots$$

La estimación de los parámetros por máxima verosimilitud de este tipo de modelos puede llevarse a cabo mediante la optimización separada de los dos componentes, esto es, del proceso binario y del truncado. En muchas ocasiones, la primera parte es representada por un modelo logia o probit. Como señalan Deb y Trivedi (2002), esta elección no tiene efectos significativos sobre las probabilidades estimadas.

2.3. Modelos de mezcla finita.

Si bien, como ya se indicó, los modelos anteriores pueden ser vistos como modelos de mezcla finita, suponen también ciertas restricciones que pueden ser superadas. En estos, se considera, exclusivamente, la muestra dividida en dos subconjuntos representados

[10] Aunque en este caso el conjunto de variables explicativas es el mismo, podría darse el caso en el que difiriesen.
[11] Otros autores han usado versiones distintas de este modelo como la binomial negativa (Gurmu y Trivedi, 1996).

por los recuentos ceros y positivos, que pueden ser generados por distintos procesos y son un caso especial de mezcla finita con un componente degenerado. Una alternativa a esta propuesta, que muestran una mayor flexibilidad e importantes ventajas y que permiten la mezcla respecto a los valores ceros y positivos, la constituye los modelos de mezcla finita pertenecientes a la familia de los denominados modelos de clase latente (Lindsay, 1995). Éstos tienen una larga tradición pero ha sido en los últimos años donde han alcanzado una mayor popularidad extendiéndose su aplicación a muchas áreas de investigación (Leisch, 2004).

En este tipo de modelos se parte de una variable aleatoria que se supone extraída de una población que es una mezcla aditiva de C subpoblaciones distintas en proporciones $\pi_1,\dots,\pi_c$, donde:

$$\sum_{j=1}^{C}\pi_j = 1, \quad \pi_j \geq 0, \quad j = 1,\dots,C.$$

La función de probabilidad de la mezcla viene dada por:

$$f(y_i / \theta) = \sum_{j=1}^{C}\pi_j f(y_i / \theta_j) \quad i = 1,\dots,n,$$

siendo $f(y_i / \theta_j)$ la función de probabilidad de cada subpoblación $j\ (j = 1,\dots,C)$. Estas distribuciones podrían pertenecer a familias paramétricas distintas, aunque normalmente se supone que pertenecen a la misma (por ejemplo, Poisson, Binomial negativa, etc). Por lo general, los π_j son parámetros desconocidos que deben ser estimados conjuntamente con el resto de parámetros del modelo, θ, además $\pi_C = 1 - \sum_{j=1}^{C-1}\pi_j$. Estos π_j podrían ser parametrizados como función de un conjunto de variables explicativas, usando por ejemplo una función logit (véase por ejemplo Wang et al, 1998), sin embargo, podrían surgir problemas de identificación, de ahí que en muchos casos éstos sean estimados como simples constantes (Deb y Trivedi, 1997, entre otros).

La mezcla finita proporciona una representación natural e intuitivamente atractiva de la heterogeneidad en un número finito, usualmente pequeño, de clases latentes, cada una de las cuales podría ser considerada como un tipo o grupo. La elección del número de componentes de la mezcla determina el número de tipos. Si estos componentes tienen una interpretación natural[12], una caracterización de mezcla finita es atractiva, sin embargo esto no es esencial, ya que la mezcla finita podría ser simplemente una manera flexible y parsimoniosa de modelizar los datos (Cameron y Trivedi, 1998).

2.4. Modelos para una submuestra.

En algunas situaciones, el rango de valores de la variable dependiente es parcialmente observable, o bien, interesa representar el comportamiento de un subconjunto de la población. Por ejemplo, sólo los individuos que han experimentado al menos un cierto número de sucesos son considerados. En este contexto, se trabajaría con una muestra truncada a la izquierda, siendo la más habitual la truncación en cero, en el caso que nos ocupa, sólo consumidores.

La forma habitual de recoger este comportamiento es mediante los modelos de Poisson o Binomial Negativo truncados[13], que se corresponderían con la segunda parte del modelo con obstáculo planteado con anterioridad. En este caso, el supuesto es que el comportamiento de la variable dependiente sigue alguna de esas distribuciones y que los recuentos observados están restringidos a tomar únicamente valores positivos.

Una forma alternativa de plantear la truncación en cero, diferente a la obtenida a través de la distribución condicionada, es cambiando el espacio muestral. De esta forma, la función de probabilidad vendría dada por:

$$P(Y_i = y_i / x_i) = \frac{e^{-\lambda_{ii}} \lambda_i^{y_i - 1}}{(y_i - 1)!}, \quad y_i = 1, 2, 3...$$

siendo ahora

$$E[y_i / x_i] = \lambda_i + 1$$
$$V[y_i / x_i] = \lambda_i$$

Esta distribución es conocida como *Poisson desplazada* (Jonson y Kotz, 1969) y coincide con la obtenida por Shaw (1988) para el tratamiento del denominado *on-site sampling*[14]. Aunque el fundamento para la distribución truncada y la obtenida de esta

---

[12] Por ejemplo, si se consideran dos componentes en el análisis del número de visitas médicas, uno de ellos podría interpretarse en relación al grupo de los usuarios poco frecuentes y el otro a los usuarios frecuentes. (Véase, por ejemplo, Deb y Trivedi, 2002).

[13] Estos modelos han sido discutidos, entre otros, por Gurmu (1991), Grogger y Carson (1991), Gurmu y Trivedi (1992).

[14] Véase Englin y Shonkwiler (1995) para una extensión al binomial negativo.

última forma es el mismo, las dos distribuciones son diferentes (Cameron y Trivedi, 1998, p. 331).

Finalmente, esta aproximación puede extenderse a las diferentes versiones binomial negativa y al de mezcla finita planteados en los apartados anteriores. Para este último, se requiere simplemente hacer explícitas las nuevas distribuciones obtenidas para las distintas clases que puedan considerarse. El procedimiento de optimización sigue las mismas pautas que en el original.

3. Datos y Resultados.

3.1. Datos.

Los datos utilizados en este trabajo proceden de una encuesta exhaustiva sobre consumo de vino en Tenerife, realizada durante los meses de abril y mayo de 2001, en la que se incluyeron preguntas relativas a los hábitos, actitudes y preferencias de los residentes en la isla sobre el consumo de vino en general y el consumo de vino de Tenerife[15]. Las variables que se utilizan en nuestro análisis son la variedad de consumo de vino, la frecuencia de consumo y algunas de carácter socioeconómico -cuya descripción se detalla en la tabla 1 del anexo.

Con el propósito de proporcionar una primera idea relativa a la “diversificación” del consumo se ha considerado interesante realizar un análisis descriptivo del número de tipos de vino. Además se comentan algunos resultados significativos obtenidos a partir de las tablas de contingencia entre esta variable y la frecuencia de consumo, así como otras de tipo socioeconómico.

En primer lugar, se observa que la mayor parte de los individuos que consumen vino suele elegir entre uno y tres tipos distintos. Concretamente, un 35.73% de los individuos suele tomar dos tipos distintos, siendo las combinaciones de vinos más consumidas la del embotellado del país con D.O. y embotellado de fuera con D.O. (29.2%), seguida de la del granel del país y embotellado del país con D.O. (25.8%). Por otro lado, en el grupo de individuos que sólo toman un tipo de vino (27.64%), los que se consumen con mayor frecuencia son el granel del país (35.8%), el embotellado del país con D.O. (33.7%), el embotellado de fuera con D.O. (16.3%) y el granel del país cosecha propia (12.6%). Asimismo, un 20.67% consume tres tipos distintos, y dentro de éstos cabe

[15] Consultar Guirao et al (2001) para una descripción detallada de la encuesta.

destacar la combinación formada por los vinos de granel del país, embotellado del país con D.O. y embotellado de fuera con D.O., como la más demandada (38%).

Tabla 1. ***Análisis descriptivo del número de tipos de vino***

Número de tipos	**Frecuencia**	**Porcentaje sobre el total**	**Porcentaje sobre los que consumen**
0	282	24.06	------
1	246	20.99	27.64
2	318	27.13	35.73
3	184	15.70	20.67
4	73	6.23	8.20
5	33	2.82	3.71
6	26	2.22	2.92
7	10	0.85	1.13

Por otro lado, a través del análisis de las tablas de contingencia se puede observar que, por término medio, el hombre consume un 25% más de tipos distintos de vino que la mujer. Además, el número medio de tipos de vino consumidos en los cuatro primeros intervalos de edad es prácticamente el mismo y cercano a 2, mientras que entre los 60 y 69 años esta media se reduce a 1.5, y finalmente, a partir de los 70 años pasa a ser 1. Por otra parte, en las zonas sur y metropolitana se consume una mayor variedad de vinos, especialmente entre los solteros y casados. Esta variedad también es mayor entre los profesionales, empresarios y funcionarios, mientras que las amas de casa consumen un menor número de tipos de vino. Por otro lado, se registra un aumento en la variedad a medida que crece el nivel de ingresos, aunque no ocurre lo mismo para los distintos niveles de estudio, ya que los universitarios y los individuos con estudios primarios son los que consumen más tipos distintos. En relación a la frecuencia de consumo se observa que, por lo general, a medida que el consumo de vino se hace más frecuente, el número medio de tipos de vino consumidos se incrementa. En concreto, los que consumen de manera ocasional suelen tomar, por término medio, aproximadamente 2 tipos de vino distintos, mientras que esta cantidad se incrementa hasta 2.5 para los que consumen algunas veces y llega a ser de 2.8 para los que tienen una alta frecuencia en el consumo. Finalmente, el número medio de tipos de vino de los que realizan un consumo diario es 2.5, aunque la dispersión de éstos respecto al número medio de tipos de vino es la más alta de todas las categorías.

Esta relación que se observa entre el número de tipos de vino y la frecuencia de consumo sugiere que la frecuencia, junto con algunas características socioeconómicas, puede ser un factor determinante de la variedad de tipos de vino que consume el individuo, por lo que se ha incluido como variable explicativa en los modelos utilizados.

3.2. Resultados.

Con el fin de aproximarnos al análisis de la mayor o menor exclusividad en el consumo de vino se han considerado los diferentes modelos planteados para el conjunto de los individuos y para la submuestra de consumidores. En ambos casos, se utiliza inicialmente las aproximaciones de Poisson y binomial negativa. En relación a esta última, se hacen dos supuestos: se considera que la varianza es proporcional a la media, lo que da lugar al modelo NEGBIN I, y se define una varianza que es función cuadrática de la media, resultando así el modelo NEGBIN II.

Para el conjunto de la información muestral, tanto en el modelo NEGBIN I como en el NEGBIN II, el parámetro de dispersión (*α)* no resulta significativo, sugiriendo que la sobredispersión presente en los datos no es muy elevada (véase tabla 2 del anexo). Este resultado parece razonable dada la relativa proximidad existente entre la media y la varianza de la variable dependiente, 1,8063 y 2,365, respectivamente.

Para comparar el modelo de Poisson con las aproximaciones binomiales negativas, en primer lugar se ha utilizado el *test óptimo basado en la regresión* propuesto por Cameron y Trivedi (1990) para contrastar la sobredispersión o subdispersión en el modelo de Poisson[16]. Este test se basa en la regresión auxiliar MCO de $z_i = \left[(y_i - \mu_i)^2 - y_i\right] / \sqrt{2}\,\mu_i$ sobre $w_i = g(\mu_i) / \sqrt{2}\,\mu_i$, donde $g(\mu_i)$ es igual μ o μ_i^2 y en el posterior análisis de la significación del coeficiente de la misma. En ambos casos, dicho coeficiente resultó casi nulo y no significativo, indicando una escasa presencia o ausencia de sobredispersión en los datos. En segundo lugar, utilizando el contraste habitual de razón de verosimilitudes (RV)[17] para la comparación de modelos anidados no se encuentra evidencia para rechazar el modelo de Poisson frente a los modelos NEGBIN I y II[18].

---

[16] $\begin{cases} H_0 : \operatorname{var}(y_i) = \mu_i \\ H_1 : \operatorname{var}(y_i) = \mu_i + \alpha\, g(\mu_i) \end{cases}$

[17] $RV = -2(Ln\ L_0 - Ln\ L_1) \sim \chi_q^2$, siendo L_0 la función de verosimilitud del modelo restringido y L_1 la del modelo no restringido.

[18] En un trabajo anterior (véase Guirao et al (2002)) analizamos también la influencia de algunas características socioeconómicas del individuo sobre el número de tipos de vino que consume, pero sin considerar el efecto de la frecuencia de consumo. En dicho trabajo el modelo de Poisson fue rechazado

Junto a las aproximaciones anteriores, se han estimado un modelo de Poisson con obstáculo (MPO) y un modelo de Poisson de mezcla finita (MPMF) (véase tabla 2 del anexo). En el primero, la probabilidad de que un individuo decida no consumir se ha modelizado separadamente de la probabilidad de que decida consumir un número determinado de tipos de vino, mientras que en el segundo, se han considerado dos componentes (clases latentes) que podrían interpretarse, a priori, y de manera natural en relación con una mayor o menor exclusividad en el consumo. La dicotomía entre estos dos grupos estaría caracterizada por una media y varianza del número de tipos consumidos inferior en el primero que en el segundo. Para nuestro caso, las medias son, respectivamente, 1,70 y 2,55, siendo las varianzas 0,974 y 6,15. Asimismo, la significación de π_j refuerza la hipótesis de dos subpoblaciones, cuyas proporciones estimadas serían 0,847 y 0.153.

Dado que los modelos MPO y MPMF no están anidados, para su comparación utilizamos dos criterios tradicionales de selección, el Criterio de Información de Akaike (AIC) y el Criterio de Información Bayesiano (BIC)[19].

Tabla 2. ***Comparación de modelos***

	Criterio/Modelo	**Poisson**	**MPO**	**MPMF**
Toda la Muestra	**AIC**	3547.2	3857.7	3535.4
(N=1172)	**BIC**	3704.9	4167.1	3848.3
Consumidores	**AIC**	2701,9	---	2665,3
(N=890)	**BIC**	2878,5	---	3032,0

En base a los resultados de la tabla anterior puede concluirse que el modelo con obstáculo debe ser rechazado en favor de las otras alternativas, siendo el modelo preferido el MPMF, según el criterio AIC y el de Poisson, de acuerdo con el criterio BIC.

frente a las dos aproximaciones binomiales negativas utilizando los tests de sobredispersión habituales, lo que induce a pensar que las diferencias entre individuos no sólo se deben a la heterogeneidad observada sino también a la heterogeneidad no observada. El término de heterogeneidad no observada que se incluye en la media condicional de las aproximaciones binomiales negativas puede estar recogiendo el efecto de una variable explicativa que ha sido omitida en el modelo. En el trabajo actual la inclusión de la frecuencia de consumo como variable explicativa no ha conducido al rechazo del modelo de Poisson frente a los dos binomiales negativos, por lo que se puede pensar que esta variable ha ayudado a explicar las diferencias individuales que inicialmente fueron atribuidas a la heterogeneidad no observada y en concreto, a la sobredispersión.

[19] $AIC = -2\ln L + k$; $BIC = -2\ln L + k\ln(N)$. Donde L es la función de verosimilitud maximizada, k es el número de parámetros que se estiman y N es el número total de observaciones muestrales.

Para comparar estos dos últimos modelos puede emplearse el contraste de razón de verosimilitudes, dado que están anidados, en este caso, el estadístico de contraste toma el valor 39,766 que, a un nivel de significación del 5%, conduce al rechazo del modelo de Poisson a favor del modelo de clase latente.

De acuerdo con el modelo MPMF, la edad, en general, influye en el número de tipos de vino que el individuo decide consumir, pudiendo afirmarse que cuanto mayor es la edad del individuo menor es la variedad de tipos de vino que consume. Esta característica también está presente en el resto de modelos. En éstos, otro de los factores determinantes es el área, registrándose un consumo más variado en la zona sur y, sobre todo, en la zona metropolitana, respecto a la zona norte de la isla; sin embargo, en el modelo de mezcla finita, este factor influye de manera diferenciada en las dos clases consideradas, mostrando un impacto significativo la pertenencia a la zona sur para el grupo de consumidores menos exclusivos y la metropolitana, para los más exclusivos. La frecuencia de consumo también constituye un factor relevante en todos los modelos, siendo su efecto similar en las dos clases recogidas en el modelo de mezcla finita. Este hecho puede interpretarse como que dicho factor no es discriminante entre los grupos obtenidos.

La variable ingreso, exceptuando el modelo de mezcla finita, no afecta al número de tipos de vino, aunque es destacable que esta variable sólo sea significativa en el segundo grupo de individuos. En este caso, su efecto es negativo, mostrando una mayor exclusividad con el nivel de renta. Asimismo, la ocupación, que también podría captar el efecto del ingreso, únicamente resulta significativa también para esta parte del modelo en algunas de sus categorías (empresarios, profesionales y otros), siendo su efecto positivo. Cabe destacar que la variable género no es un factor explicativo en ninguno de los modelos, aunque resulta interesante mencionar que existe una alta correlación entre la variable género y la frecuencia de consumo, que sí es significativa. Asimismo, el nivel educativo, por lo general, tampoco resulta significativo, siendo los individuos que poseen estudios universitarios los que muestran una mayor variabilidad.

En relación a la submuestra de consumidores, hemos seguido el procedimiento descrito en el apartado 2.4. Como se observa en la tabla 2, los resultados obtenidos para la selección de los modelos son paralelos a los de la muestra completa. Esto es, el criterio AIC hace preferible el modelo de MPMF, mientras que el BIC, al de Poisson. El valor del estadístico de contraste de razón de verosimilitudes entre estos modelos anidados es

64,64 rechazando, al 5% de significación, el modelo de Poisson frente al de mezcla finita. Para éste último, las proporciones para cada clase son significativas, estimándose en 0.758 y 0,242, respectivamente. Como señalamos con anterioridad, esto puede ser un indicio de la presencia de dos subpoblaciones que muestran una respuesta diferenciada en relación a la mayor o menor exclusividad en el consumo.

Los resultados obtenidos son, en algunos aspectos, similares a los obtenidos para toda la muestra (véase tabla 3 del anexo). Esto es, para el modelo de Poisson estándar, edad, área y frecuencia de consumo se muestran como factores relevantes, no así el nivel educativo que para el conjunto de los individuos, implicaba una mayor variabilidad en los de estudios universitarios. Esto último se reproduce para ambos grupos del modelo de mezcla finita. En este, además, la ocupación tampoco es relevante para la clase de menos exclusivos, debiendo destacarse ahora, el efecto positivo de la renta en los más exclusivos, aun cuando se mantiene su impacto negativo para el primer grupo. Por último, el resto de factores, edad, área y frecuencia de consumo, muestran un comportamiento similar al obtenido para toda la muestra, aunque en este caso, la frecuencia presenta un menor impacto para ambos grupos.

4. Conclusiones.

En este trabajo se han presentado varias aproximaciones para analizar el número de tipos de vino consumidos por los residentes en Tenerife. En concreto, se han comparado algunas aproximaciones estándar (Poisson, NEGBIN I y NEGBIN II) con otras aproximaciones más flexibles (modelo de Poisson con obstáculo y modelo de Poisson de mezcla finita con dos componentes). Estas últimas, han sido propuestas alternativas en el contexto de datos de recuento que muestran habitualmente medias relativamente elevadas y amplias colas, encontrándose mejoras marginales respecto a las versiones paramétricas habituales (Guo y Trivedi, (2002)). En nuestro caso, el tipo de recuento considerado muestra rasgos diferentes a los apuntados, de ahí el interés de explorar estas alternativas desde un punto de vista empírico.

A partir de algunos criterios habitualmente utilizados para la selección de modelos para datos de recuento se puede concluir que los modelos binomiales negativos no son preferidos al modelo de Poisson. Asimismo, al comparar el modelo de Poisson con los otros dos modelos que no suponen una misma estructura de preferencia para consumidores y no consumidores (MPO), o bien, suponen la existencia de subpoblaciones con respuesta diferente (MPMF), resulta favorecido, según los criterios empleados, el modelo de Poisson o el modelo de mezcla finita. No obstante, los resultados apuntan a que éstos últimos pueden ser, marginalmente, más apropiados incluso para este tipo particular de recuento. De cualquier forma, quedan abiertas otras alternativas semiparamétricas y no paramétricas que pueden ser investigadas.

Como resumen de los resultados obtenidos sobre los determinantes de la mayor o menor exclusividad en el consumo de vino se tiene que, los factores relevantes son, en general, la edad del individuo -registrándose una menor variedad de tipos de vino a medida que ésta se incrementa-, el área de residencia –siendo el consumo más exclusivo en la zona norte de la isla- y la frecuencia de consumo, siendo excepcionalmente relevantes para algunos modelos el ingreso y el nivel educativo. Asimismo, estos factores muestran pautas de comportamiento diferenciadas según los grupos de individuos considerados en el modelo de mezcla finita, dotando a éste de una aproximación quizás más realista y útil mediante la que representar la heterogeneidad del fenómeno observado.

5. Bibliografía.

Bello, L. y Gómez, J.T. (1996): "Las denominaciones de origen y otras señales de calidad en las estrategias de diferenciación de los productos agroalimentarios. Una propuesta metodológica", *Cuadernos Aragoneses de Economía*, 6 (2), 365-387.

Bernabéu, R., Tendero, A., Olmeda, M. y Castillo, S. (2001): "Actitud del consumidor de vino con denominación de origen en la provincia de Albacete", *IV Congreso Nacional de Economía Agraria*, Navarra, septiembre de 2001.

Boswell, M.T. y Patil, G.P. (1970): "Chance Mechanisms Generating the Negative Binomial Distributions", en Patil, G.P., *Random Counts in Models and Structures,* vol. 1-3, University Park, PA, and London, Pennsylvania State University Press.

Cameron, A.C. y Trivedi, P.K. (1990): "Regression-based Tests for overdispersion in the Poisson Model", *Journal of Econometrics*, 46, 347-364.

Cameron, A.C. y Trivedi, P.K. (1998): "Regression Analysis of Count Data", *Cambridge University Press.*

Cragg, J.C. (1971): "Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods", *Econometrica*, vol. 39, 829-844.

Deb, P. y Trivedi, P.K. (1997): "Demand for Medical Care by the Elderly in the United States: a Finite Mixture Approach", *Journal of Applied Econometrics,* 12, 313-336.

Deb, P. y Trivedi, P.K. (2002): “The Structure of Demand for Health Care: Latent Class versus two-part models”, *Journal of Health Economics*, 21, 601-625.

Englin, J. y Shonkwiler (1995): “Estimating Social Welfare using Count Data Models: An Application to Long-run Recreation Demand under Conditions of Endogenous Stratification and Truncation”, *The Review Economics and Statistics*, 77, 104-112.

Flavián, C., Martínez, E. y Polo, Y. (1997): “La fidelidad en la adquisición de artículos de compra frecuente”, *Revista Europea de Dirección y Economía de la Empresa*, 6 (4), 63-76.

Gourieroux, C., Monfort, A. y Trognon, A. (1984a): “Pseudo Maximum Likelihood Methods: Theory”, *Econometrica*, 52, pp. 681-700.

Gourieroux, C., Monfort, A. y Trognon, A. (1984b): “Pseudo Maximum Likelihood Methods: Applications to Poisson Models”, *Econometrica*, 52, pp. 701-720.

Grogger, J.T. y Carson, R.T. (1991): “Models for Truncated Counts”, *Journal of Applied Econometrics*, 6, 225-238.

Guirao, G., Cano, V., López, M.I., Rodríguez, M.C. y Romero, M.E. (2002): “Relación entre la frecuencia de consumo de vino y algunas características socioeconómicas”. Comunicación presentada en las *IV Jornadas Técnicas Vitivinícolas Canarias*, Santa Cruz de Tenerife.

Guirao, G., Cáceres, J.J., Cano, V., Hernández, M., López, M.I., Martín, F.J. y Rodríguez, M.C. (2001): *El consumo de vino en Tenerife.* Servicio Técnico de Desarrollo Rural y Pesquero, Cabildo Insular de Tenerife.

Guo, J. y Trivedi, P. (2002) “Flexible Parametric Models for Long-tailed Patent Count Distributions”, *Oxford Bulletin of Economics and Statistics*, 64 (1), pp. 63-82.

Gurmu, S.(1991): “Test for Detecting Overdispersion in the Positive Regression Model”, *Journal of Business and Economics Statistics*, 9, 215-222.

Gurmu, S. y Trivedi, P.K. (1992): “Overdispersion Test for Truncated Poisson Regression Models”, *Journal of Econometrics*, 54, 347-370.

Gurmu, S. y Trivedi, P.K. (1996): “Excess of Zeros in Count Models for Recreational Trips”, *Journal of Business and Economics Statistics*, 14, 469-477.

Hausman, J.A., Hall, B.H. y Griliches, Z. (1984): "Econometric Models for Count Data with an Application to the Patents-R and D Relationship", *Econometrica,* 52, pp. 909-938.

Martínez-Carrasco, L. et al. (2004) "El Efecto de la Preocupación por la Salud, el Interés por la Gastronomía y la Actividad Social sobre la Intensidad de Consumo de Vino de Calidad", *Economía Agraria y Recursos Naturales*, Vol. 4, 7, pp. 27-42.

Mullahy, J. (1986): "Specification and Testing of Some Modified Count Data Models", *Journal of Econometrics,* 33, 341-365.

Mullahy, J. (1997): "Heterogeneity, Excess Zeros and the Structure of Count Data Models", *Journal of Applied Econometrics*, 12, pp. 337-350.

Johnson, N.L. y Kotz, S. (1969): *Discrete Distributions*. Boston. Houghton Mifflin.

Leisch, F. (2004): "FlexMix: A General Framework for Finite Mixture Models and Latent Class Regression in R", Journal of Statistical Software, 11, 1-18.

Lindsay, B.G. (1995): "Mixture Models: Theory, Geometry and Applications, *NSF-CBMS Regional Conference Series in Probability and Statistics*, vol. 5, IMS-ASA. Hayward, CA: Institute of Mathematical Statistics.

Sánchez, M. y Gil, J.M. (1998): "Consumer Preferences for Wine Attributes in Different Retail Stores: A Conjoint Approach", *International Journal of Wine Marketing*, 10 (1), 25-38.

Shaw, D. (1988): "On-site Samples Regression", *Journal of Econometrics*, 37, 211-223.

Wang, P., Cockburn, I.M. y Puterman, M.L. (1998): "Analysis of Patent Data –A Miexed Poisson Regression Model Approach", *Journal of Business and Economics Statistics*, 16 (1), 27-41.

Winkelmann, R. y Zimmermann, K.F. (1995): "Recent Development in Count Data Modelling: Theory and Applications", *Journal of Economics Surveys*, 9, 1-24.

Anexo

Tabla 1. Variables incluidas en los modelos

Frecuencia de consumo de vino FC	Nunca = 0 Ocasionalmente = 1 Alguna vez durante las comidas = 2 Con alta frecuencia en las comidas = 3 Diariamente durante las comidas = 4
Género	Variable dummy, hombre = 1 , mujer = 2
Edad: E1 E2 E3 E4 E5 E6	Variables dummy que recogen: 18-29 años 30-39 años 40-49 años 50-59 años 60-69 años ≥70 años
Area: A1 A2 A3	Variables dummy que recogen la pertenencia a cada una de las tres áreas siguientes: Zona Norte Zona Sur Zona Metropolitana
Situación Familiar: SF1 SF2 SF3	Variables dummy que recogen la situación familiar del encuestado Casado Soltero Viudo/Separado
Número de miembros de la unidad familiar MUF	Variable que recoge el número de miembros de la unidad familiar (1,2,3,.....)
Ocupación O1 O2 O3 O4 O5 O6 O7	Variables dummy que recogen la ocupación del encuestado Empleado Funcionario Estudiante Ama de casa Empresario Profesional Otros
Educación ED1 ED2 ED3 ED4	Variables dummy que recogen el nivel educativo del encuestado Sin estudios Estudios primarios Estudios secundarios Estudios universitarios
Ingresos I1 I2 I3 I4 I5	Variables dummy que recogen el nivel de ingresos (euros) <600 600-1200 1200-1800 1800-2400 >2400

Tabla 2. Estimaciones de los modelos para toda la muestra (N=1172)

							Modelo con obstáculo				*Modelo de Mezcla Finita (Clase Latente)*			
Variables	*Poisson*		*NEGBIN I*		*NEGBIN II*		*Logia*		*P- Truncado*		*Clase I*		*Clase II*	
	Coef.	**p-value**	**Coef.**	**p-value**	**Coef.**	**p-value**	**Coef.**	**p-value**	**Coef.**	**p-value**	**Coef.**	**p-value**	**Coef.**	**p-value**
Constante	-0.947	0.000	-0.946	0.000	-0.947	0.000	1.268	0.005	-0.303	0.184	-0.832	0.000	-2.088	0.020
S2	-0.054	0.301	-0.052	0.361	-0.053	0.342	-0.864	0.000	-0.071	0.252	-0.012	0.859	-0.228	0.284
E1	0.646	0.000	0.646	0.001	0.647	0.000	0.960	0.026	0.502	0.008	0.592	0.002	1.149	0.058
E2	0.586	0.000	0.584	0.001	0.586	0.000	0.747	0.047	0.535	0.002	0.459	0.006	1.389	0.009
E3	0.425	0.001	0.423	0.032	0.426	0.002	0.537	0.125	0.473	0.005	0.234	0.159	1.254	0.007
E4	0.432	0.001	0.429	0.036	0.432	0.003	0.602	0.087	0.500	0.003	0.320	0.054	0.942	0.050
E5	0.166	0.182	0.163	0.248	0.166	0.228	0.369	0.229	0.225	0.177	-0.031	0.854	0.858	0.052
A2	0.129	0.018	0.128	0.027	0.130	0.022	0.021	0.905	0.159	0.016	0.045	0.537	0.431	0.030
A3	0.220	0.000	0.220	0.001	0.220	0.001	0.128	0.511	0.166	0.017	0.233	0.002	0.240	0.340
SF2	0.002	0.979	0.001	0.928	0.002	0.979	-1.138	0.000	0.093	0.290	-0.047	0.606	0.198	0.496
SF3	-0.086	0.292	-0.085	0.348	-0.085	0.336	-0.540	0.028	-0.057	0.567	-0.072	0.481	-0.068	0.818
MUF	0.024	0.130	0.024	0.166	0.024	0.152	0.0005	0.993	0.030	0.108	0.008	0.719	0.078	0.248
O2	-0.104	0.306	-0.105	0.397	-0.104	0.390	0.648	0.211	-0.080	0.510	-0.228	0.071	0.373	0.297
O3	-0.053	0.538	-0.054	0.568	-0.053	0.566	-0.682	0.017	0.052	0.604	-0.138	0.213	0.295	0.432
O4	0.019	0.829	0.017	0.863	0.019	0.842	-0.669	0.012	0.010	0.993	-0.133	0.244	0.706	0.105
O5	-0.0002	0.998	-0.002	0.979	-0.001	0.999	-0.008	0.980	0.097	0.282	-0.161	0.166	0.662	0.016
O6	0.074	0.406	0.072	0.423	0.074	0.400	0.759	0.127	0.111	0.281	-0.089	0.435	0.958	0.015
O7	-0.025	0.781	-0.027	0.778	-0.025	0.788	-0.115	0.705	-0.011	0.923	-0.194	0.095	0.803	0.015
ED2	0.198	0.053	0.199	0.054	0.198	0.051	0.795	0.002	0.078	0.540	0.200	0.160	0.396	0.272
ED3	0.166	0.132	0.166	0.123	0.166	0.114	0.817	0.008	-0.033	0.809	0.065	0.668	0.630	0.113
ED4	0.295	0.013	0.295	0.014	0.295	0.012	1.257	0.000	0.107	0.461	0.174	0.277	0.986	0.035
ING2	-0.102	0.228	-0.102	0.239	-0.102	0.231	-0.580	0.024	-0.014	0.891	-0.062	0.601	-0.263	0.371
ING3	-0.078	0.376	-0.076	0.418	-0.078	0.402	-0.360	0.200	-0.024	0.820	0.068	0.572	-0.690	0.036
ING4	-0.132	0.189	-0.129	0.231	-0.132	0.215	-0.306	0.363	-0.132	0.277	0.118	0.394	-1.297	0.004
ING5	-0.118	0.259	-0.116	0.289	-0.118	0.273	0.065	0.863	-0.080	0.525	0.152	0.317	-1.345	0.018
FC	0.409	0.000	0.409	0.000	0.409	0.000			0.153	0.000	0.430	0.000	0.471	0.003
α			0.019	0.683	0.0003	0.999								
Ln L	-1760.601		-1760.499		-1760.602		-565.248		-1338.331		-1740.718			

Tabla 3. Estimaciones de los modelos consumidores (N=890)

Variables	*Poisson*		*Modelo de Mezcla Finita (Clase Latente)* *Clase I*		*Clase II*	
	Coef.	**p-value**	**Coef.**	**p-value**	**Coef.**	**p-value**
Constante	-0.924	0.000	-0.969	0.017	-0.903	0.257
S2	-0.086	0.207	-0.052	0.604	-0.129	0.512
E1	0.598	0.004	0.425	0.174	1.371	0.158
E2	0.638	0.001	0.566	0.052	0.889	0.069
E3	0.564	0.002	0.349	0.229	0.874	0.058
E4	0.596	0.001	0.493	0.103	0.762	0.114
E5	0.265	0.144	-0.079	0.786	0.827	0.057
A2	0.193	0.008	-0.085	0.533	0.713	0.000
A3	0.201	0.009	0.287	0.013	0.063	0.796
SF2	0.113	0.242	0.256	0.078	-0.607	0.199
SF3	-0.068	0.534	-0.016	0.922	-0.285	0.345
MUF	0.036	0.130	-0.010	0.759	0.063	0.301
O2	-0.097	0.468	-0.209	0.248	-0.038	0.932
O3	-0.063	0.569	-0.032	0.854	0.252	0.436
O4	0.003	0.978	-0.229	0.219	0.455	0.192
O5	0.119	0.233	-0.167	0.367	0.563	0.084
O6	0.136	0.233	-0.082	0.657	0.352	0.261
O7	-0.014	0.906	-0.237	0.222	0.470	0.194
ED2	0.093	0.502	0.203	0.452	-0.041	0.908
ED3	-0.041	0.784	-0.276	0.341	0.585	0.145
ED4	0.129	0.418	0.044	0.884	0.412	0.353
ING2	-0.018	0.874	0.295	0.127	-0.696	0.018
ING3	-0.030	0.796	0.415	0.040	-0.866	0.012
ING4	-0.161	0.228	0.332	0.155	-0.988	0.012
ING5	-0.098	0.480	0.461	0.051	-1.075	0.006
FC	0.185	0.000	0.180	0.000	0.242	0.007
α						
Ln L	-1337.968		-1305.646			