



INSTITUT FÜR STATISTIK
SONDERFORSCHUNGSBEREICH 386



Smith:

On Dependency in Double-Hurdle Models

Sonderforschungsbereich 386, Paper 139 (1998)

Online unter: <http://epub.ub.uni-muenchen.de/>

Projektpartner



On Dependency in Double-Hurdle Models

Murray D. Smith¹

December, 1998

ABSTRACT

In microeconometrics, consumption data is typically zero-inflated due to many individuals recording, for one reason or another, no consumption. A mixture model can be appropriate for statistical analysis of such data, with the Dependent Double-Hurdle model (DDH hereafter) one specification that is frequently adopted in econometric practice. Essentially, the DDH model is designed to explain individual demand through a sequential two-step process: a market participation decision (first hurdle), followed by a consumption level decision (second hurdle) - a non-zero correlation/covariance parameter allows for dependency between the hurdles. A significant feature of the majority of empirical DDH studies has been the lack of support for the existence of dependency. This empirical phenomenon is studied from a theoretical perspective using examples based on the bivariate normal, bivariate logistic, and bivariate Poisson distributions. The Fisher Information matrix for the parameters of the model is considered, especially the component corresponding to the dependency parameter. The main finding is that the DDH model contains too little statistical information to support estimation of dependency, even when dependency is truly present. Consequently, the paper calls for the elimination of attempts to estimate dependency using the DDH framework. The advantage of this strategy is that it permits flexible modelling; some possibilities are proposed.

¹Department of Econometrics, University of Sydney, Sydney NSW, Australia 2006. E-mail: <Murray.Smith@econ.usyd.edu.au>. This paper was written while visiting the Sonderforschungsbereich 386, Institut für Statistik, Ludwig-Maximilians-Universität München, Deutschland. For discussions which proved instrumental in improving the content of the paper, I am indebted to Ludwig Fahrmeir, Tim Fry, Helmut Küchenhoff, Peter Moffatt, Hans Schneeweiss, Peter Schönfeld, and seminar participants at the universities of Bonn, Monash and München. Support from the Alexander von Humboldt-Stiftung and the Australian Research Council (through VISLAB and MEMLab) are acknowledged gratefully.

1 Introduction

The Double-Hurdle model (DH hereafter) has been used in economics to analyse a wide range of individual commodity demand and labour supply behaviour; important contributions include Blundell and Meghir [3] and Jones [13]. Other fields in which the DH model has been applied include finance (*e.g.*, Dionne *et al.* [6] examine credit-scoring) and sociology (*e.g.*, Zorn [20] examines legislative response to court rulings). In terms of commodity demand, the DH model is designed to explain the mechanism of individual demand through a sequential two-step process: a market participation decision (first hurdle), followed by a consumption level decision (second hurdle). The statistical origins of the model are due to Cragg [5], and its basis in consumer choice theory is due to Pudney [16, pp.160-162].

The generalisation of the DH model to allow for dependence between the participation and consumption decisions - the Dependent Double-Hurdle model (DDH hereafter) - has recently been the subject of empirical attention. Importantly, the arguments mounted for this generalisation have not been based on economic theory. Rather, justification has been based on intuitive behavioural grounds (*e.g.*, [8, p.491] and [17, p.216]), and on statistical grounds (*e.g.*, in [9] where it improves fit). In statistical terms, a parameter θ , representing dependency, is incorporated into the DH model. Typically, a DDH model nests its DH counterpart through the restriction $\theta = 0$. A summary of a number of published DDH studies appears in Table I:

Table I: Dependent Double-Hurdle Studies

	Application: Demand for	Sample size	% of 0's	DDH vs DH
Blaylock and Blisard [2]	Cigarettes (USA)	2962	60.7	insig
Burton <i>et al.</i> [4]	Meat (UK)	2144	6.3	insig
Garcia and Labeaga [8]	Cigarettes (Spain)	23669	41.2	insig
Gould [9]	Cheese (USA)	5017	59.0	sig
Jones [13]	Cigarettes (UK)	1573	na	insig
Jones [14]	Cigarettes (UK)	2321	48.5	insig
Yen and Jones [19]	Cheese (USA)	4245	18.1	insig

The important feature to note are the entries in the last column - "DDH vs DH" - which indicates whether the fitted DDH is either insignificant from (insig), or significantly different to (sig), its nested DH version. The relevant hypothesis test showed that the data in 6 of the 7 studies did not support the DDH model over the DH model at any conventional level of significance - the exception is Gould's study where an estimate of θ of 0.22 was reported to be significantly different from zero (estimated standard error 0.06). This persistent finding against the DDH model provides the motivation for this paper - an explanation is sought for why DDH models appear to be statistically indistinguishable from their DH counterparts.

The DDH model has as its statistical origin the specification of a bivariate distribution for two latent utility variables, denoted by Y_1^{**} and Y_2^{**} . From these variables, the two decision/hurdle variables are derived: denoted Y_1^* for the yes-no (1-0) decision of whether or not to participate in the market, and Y_2^* for how much to consume. Note that Y_1^* and Y_2^* are, in general, latent variables. In a survey, the variable that is observed is individual consumption Y ; its relation to the hurdle variables is given by $Y = Y_1^* \times Y_2^* \geq 0$. In this model there is a two-fold source of zero observations: the first occurring if $Y_1^* = 0$ (the individual does not participate - the first hurdle is failed); and should $Y_1^* = 1$, the second occurring if $Y_2^* = 0$ (the person participates, but elects not to consume - the first hurdle is passed, but the second is not). Such models lead to zero-inflated data. In this respect, the DDH studies referenced in Table I provide no exceptions. The entries in the column "% of 0's" gives the proportion of zero observations in the data set; Blaylock and Blisard, for example, report 60.7% of their sample of 2962 women as not smoking cigarettes. A detailed statistical construction of the DDH model is given in Section 2.

Many reasons can be given for the lack of significance of the dependency parameter, including: θ may truly be zero, the model parameters may not be identified, specification error may be inducing estimator inefficiency, the power and size properties of the test for $\theta = 0$ may be poor, θ is a component of the variance-covariance matrix and such parameters are typically harder to estimate precisely when compared to mean/regression parameters. However, recurring insignificance across a number of distinct studies is suggestive of a fundamental statistical weakness in the model, namely: that the generalisation of the DH model to the DDH model adds little to the informational content of the latter over the former. In such a situation, this would imply that the parameters of a DDH model, even if formally identified, may only be weakly identified. Section 3 provides an illustration of this for the bivariate normal DDH model.

Fisher's Information [7] (per observation) is tied intricately to statistical estimation theory and inference. In this paper, Fisher's measure is used for the purpose of assessing the quantity of statistical information contained on the model parameters (for extensive details on information theory in statistics see Kullback [15]). The greater the Fisher Information that is present on a parameter, the better the prospects of obtaining a reliable estimate of it and of performing inference about it. In Section 4, Fisher's Information on θ is examined in the context of the bivariate logistic DDH model. The findings there show the need for an abundance of zero observations if the DDH model is to contain statistical information over and above the DH model. Sections 5 and 6 examine Fisher's Information matrix in the context of the bivariate Poisson DDH model. The findings there reflect the informational trade-offs that can occur between the parameters of the DDH model, as well as the relative lack of statistical information that is present on θ - irrespective of whether θ is truly non-zero or not.

Closing remarks are given in Section 7. Chiefly, the suggestion is made that the introduction of the dependency parameter into DH models (thereby yielding a DDH model) is a statistically spurious generalisation. This seemingly negative outcome is, however, of considerable benefit to practitioners. In the absence of dependency, practitioners have the opportunity to explore more flexible distributional forms, for in DH models it is unnecessary to be concerned with the task of specifying a bivariate distribution with dependency for the underlying utility variables.

2 The Statistical Construction of the DDH Model

2.1 Distribution and Density

Let (Y_1^{**}, Y_2^{**}) denote the two utility variables underlying an individual's decision process, and assume (for now) that they are continuous and take values throughout \mathbb{R}^2 , the two-dimensional real plane. Next, assume a parametric bivariate model for (Y_1^{**}, Y_2^{**}) is specified by assigning a joint cumulative distribution function (cdf) to the variables; denote the cdf by $F(y_1^{**}, y_2^{**})$, for values $(y_1^{**}, y_2^{**}) \in \mathbb{R}^2$. The cdf depends upon unknown parameters collected in $\gamma \in \Gamma$, one of its components being the dependency parameter θ . Importantly, variables Y_1^{**} and Y_2^{**} are not observed. The observed variable is individual consumption $Y \geq 0$. The relationship between (Y_1^{**}, Y_2^{**}) and Y is established by defining two intermediate random variables:

$$Y_1^* = \begin{cases} 1 & \text{if } Y_1^{**} > 0 \\ 0 & \text{if } Y_1^{**} \leq 0, \end{cases} \quad \text{and} \quad Y_2^* = \begin{cases} Y_2^{**} & \text{if } Y_2^{**} > 0 \\ 0 & \text{if } Y_2^{**} \leq 0. \end{cases} \quad (1)$$

Variables Y_1^* and Y_2^* are the hurdle variables: the discrete variable Y_1^* represents the first hurdle decision (if $Y_1^* = 0$ the hurdle is failed, the individual does not participate; if $Y_1^* = 1$ the hurdle is passed, the individual is a potential consumer), and the continuous-discrete mixture variable Y_2^* represents the second hurdle consumption (if $Y_2^* = 0$ the hurdle is failed, the individual does not consume; if $Y_2^* > 0$ the hurdle is passed, the individual consumes). In general, Y_1^* and Y_2^* are latent. Finally, to complete the construction of the DDH model, individual consumption

$$Y = Y_1^* \times Y_2^* \geq 0. \quad (2)$$

Due to the sequential decomposition of the decision, a zero observation on Y can occur in two ways: (i) if the first hurdle is not passed ($Y_1^* = 0$); or, (ii) if the first hurdle is passed and yet the second hurdle is not ($Y_1^* = 1$ and $Y_2^* = 0$). Any positive-valued observation occurs only when both hurdles are passed ($Y_1^* = 1$ and $Y_2^* > 0$).²

Under the continuity assumption, the probability density function (pdf) of Y is a (univariate) continuous-discrete mixture, with functional form depending upon the specification assumed for F . Denote it by

$$f(y) = \begin{cases} f_1(y) & \text{if } y \in \mathbb{R}_+ \\ f_0 & \text{if } y = 0, \end{cases} \quad (3)$$

where \mathbb{R}_+ denotes the positive part of the real line. When $y \in \mathbb{R}_+$, the $f_1(y)$ component of the pdf of Y is continuous. Its functional form may be derived by differentiating, with respect to y , the following probability:

$$\begin{aligned} P(Y < y) &= 1 - P(Y > y) \\ &= 1 - P(\{Y_1^* = 1\} \cap \{Y_2^* > y\}) \\ &= 1 - P(\{Y_1^{**} > 0\} \cap \{Y_2^{**} > y\}) \\ &= F_1(0) + F_2(y) - F(0, y), \end{aligned} \quad (4)$$

where $F_i(\cdot)$ denotes the marginal cdf of Y_i^{**} ($i = 1, 2$). Thus,

$$\begin{aligned} f_1(y) &= \frac{\partial}{\partial y} P(Y < y) \\ &= \frac{\partial}{\partial y} (F_2(y) - F(0, y)). \end{aligned} \quad (5)$$

When $y = 0$, the f_0 component of the pdf is the probability mass at the origin:

$$\begin{aligned} f_0 &= P(Y = 0) \\ &= 1 - P(Y > 0) \\ &= F_1(0) + F_2(0) - F(0, 0). \end{aligned} \quad (6)$$

2.2 Fisher Information

In the DDH models given in the remaining sections of the paper, interest lies with Fisher's Information on the parameters contained in γ - especially the information on the dependency parameter θ . Given the general form for the pdf of Y , it is easy to write down, for example, the definition of Fisher's Information on θ :

$$\begin{aligned} E \left(\frac{\partial}{\partial \theta} \log f(y) \right)^2 &= P(Y = 0) E \left(\left(\frac{\partial}{\partial \theta} \log f(y) \right)^2 \middle| Y = 0 \right) + P(Y > 0) E \left(\left(\frac{\partial}{\partial \theta} \log f(y) \right)^2 \middle| Y > 0 \right) \\ &= f_0 \times \left(\frac{\partial}{\partial \theta} \log f_0 \right)^2 + \int_0^\infty \left(\frac{\partial}{\partial \theta} \log f_1(y) \right)^2 f_1(y) dy \\ &= \frac{1}{f_0} \left(\frac{\partial f_0}{\partial \theta} \right)^2 + \int_0^\infty \frac{1}{f_1(y)} \left(\frac{\partial f_1(y)}{\partial \theta} \right)^2 dy. \end{aligned} \quad (7)$$

It is because the pdf of Y is a continuous-discrete mixture that there are two components in (7): the form of the first term arises because the density at the origin represents a probability mass, the second because the density at a given value in \mathbb{R}_+ represents a sliver of mass.

²It is useful to view the sample space of Y on the (Y_1^{**}, Y_2^{**}) plane - any (y_1^*, y_2^*) pairs generated in the L-shaped region $\{\{Y_1^{**} \leq 0\} \cup \{\{Y_1^{**} > 0\} \cap \{Y_2^{**} \leq 0\}\}\}$ map to the origin, and any (y_1^*, y_2^*) pairs generated in the upper-right quadrant map onto the line $Y_1^{**} = 1$ at the value y_2^* .

3 The Bivariate Normal DDH

In this first example, assume (Y_1^{**}, Y_2^{**}) is distributed according to the following bivariate normal:

$$\begin{bmatrix} Y_1^{**} \\ Y_2^{**} \end{bmatrix} \sim N \left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} 1 & \sigma\theta \\ \sigma\theta & \sigma^2 \end{bmatrix} \right).$$

Without loss of generality, $\text{Var}(Y_1^{**})$ is normalised to unity because in the construction of the DDH model all scale information on Y_1^{**} is lost due to the transformation of Y_1^{**} to Y_1^* . As is well known, the dependency parameter θ is equivalent to the correlation coefficient between Y_1^{**} and Y_2^{**} ; thus $\theta \in \Theta = \{\theta : |\theta| < 1\}$. When $\theta = 0$, the random variables Y_1^{**} and Y_2^{**} are independent. The parameters of the model are collected in the list $\gamma = (\mu_1, \mu_2, \sigma^2, \theta) \in \Gamma = \mathbb{R}^2 \times \mathbb{R}_+ \times \Theta$.

For the bivariate normal DDH model, the joint cdf of (Y_1^{**}, Y_2^{**}) is given by:

$$F(y_1^{**}, y_2^{**}) = \Omega \left(y_1^{**} - \mu_1, \frac{y_2^{**} - \mu_2}{\sigma} \right),$$

where $\Omega(\cdot)$ denotes the cdf of a standardised bivariate normal distribution with correlation coefficient θ . Substituting this expression into (4), the continuous component of the cdf of Y is given by:

$$P(Y < y) = \Phi(-\mu_1) + \Phi(z) - \Omega(-\mu_1, z),$$

for $y \in \mathbb{R}_+$. The function $\Phi(\cdot)$ denotes the cdf of a standardised univariate normal distribution, and

$$z = \frac{y - \mu_2}{\sigma}.$$

The probability mass at the origin is given by:

$$P(Y = 0) = \Phi(-\mu_1) + \Phi(-\sigma^{-1}\mu_2) - \Omega(-\mu_1, -\sigma^{-1}\mu_2).$$

In Figure I, distributions of Y are plotted for three points in the parameter space Γ , namely: $\gamma_1 = (0, 2, 1, 0)$, $\gamma_2 = (0, 1.6, 1, 0.5)$, and $\gamma_3 = (0, 2.4, 1, -0.5)$. These points can be regarded as distant from each other in Γ .

(Figure I about here)

The significant feature to notice about the plotted distributions is that all three are extremely close to each other across all values of $Y = y \geq 0$. In fact, numerical computation reveals that the greatest difference between the cdf of Y evaluated at γ_1 , and the cdf evaluated at γ_2 , occurs when $y = 2.879$ - the difference being only 0.0107. This is smaller again³ if the cdf of Y is compared at γ_2 and γ_3 ; the largest difference now is only 0.0028 at $y = 2$. Certainly, the fact that the cdf of Y is distinct at each point in Γ is sufficient to identify the parameters of the bivariate normal DDH model; however, what Figure I reveals is that identification is weak in the selected neighbourhoods of Γ . Of course, there are an infinity of parameter configurations within Γ at which the distribution of Y may be examined, many of which may lead to differing conclusions about the strength with which the parameters of the model are identified. However, it is interesting to observe in Figure I that $P(Y = 0)$ is a little over 50% for each parameter point. This fact, taken alongside similar percentages for the data statistics reported in Table I, suggests that the distributions depicted in Figure I may be of relevance in practice. Estimation of demonstrably weakly-identified parameters must be of considerable concern, for it can lead to computational problems such as lack of convergence - this in fact occurred in the Burton *et al.* DDH study, see [4, p.205].

In the examples which follow, attempts to quantify the implications of “weak identification” in the DDH model are undertaken using Fisher’s Information, the latter serves to measure the amount of statistical information that is contained on the parameters of the model.

³In fact, for the vertical scale used in Figure I the difference between the distributions is barely discernable to the eye.

4 The Bivariate Logistic DDH

4.1 Distribution and Density

For this example, assume (Y_1^{**}, Y_2^{**}) is distributed according to Gumbel's Type II bivariate logistic distribution (see [10]) with cdf:

$$F(y_1^{**}, y_2^{**}) = F_1(y_1^{**})F_2(y_2^{**})(1 + \theta(1 - F_1(y_1^{**}))(1 - F_2(y_2^{**}))), \quad (8)$$

for $(y_1^{**}, y_2^{**}) \in \mathbb{R}^2$, and where the dependency parameter $\theta \in \Theta = \{\theta : |\theta| < 1\}$. The marginal cdf

$$F_i(y_i^{**}) = (1 + \exp(-(y_i^{**} - \mu_i)))^{-1}$$

($i = 1, 2$), corresponds to the cdf of a logistic random variable with mean μ_i and variance $\pi^2/3$. Using integration by parts, it can be shown that $\text{Cov}(Y_1^{**}, Y_2^{**}) = \theta$. When $\theta = 0$, the random variables Y_1^{**} and Y_2^{**} are independent.

For this specification, the continuous component of the pdf of Y , according to (5), can be shown to equal:

$$f_1(y) = (1 - F_1(0))(1 - \theta F_1(0)(1 - 2F_2(y)))F_2'(y),$$

for $y \in \mathbb{R}_+$, and where $F_i'(y) = \partial F_i(y)/\partial y = F_i(y)(1 - F_i(y))$ for $i = 1, 2$. From (6), the probability mass at the origin is

$$f_0 = F_1(0) + F_2(0) - F(0, 0).$$

The parameters may be collected in the list $\gamma = (\mu_1, \mu_2, \theta) \in \Gamma = \mathbb{R}^2 \times \Theta$.

4.2 Fisher's Information on the Dependency Parameter

The advantage of examining the bivariate logistic DDH model is that Fisher's Information on θ , denote it by i , can be derived in closed form, a fact which will prove useful when it comes to analysis and interpretation, as well as computation. Even so, in the analysis which follows it is convenient to work with an intermediate form:⁴

$$i = \frac{1}{2}F_1(0)F_1'(0) \left(\Psi(-\theta F_1(0); 1; 3) + (1 - 2F_2(0))^3 \Psi(\theta F_1(0)(1 - 2F_2(0)); 1; 3) \right) + \frac{F_1'(0)^2 F_2'(0)^2}{F_1(0) + F_2(0) - F(0, 0)}, \quad (9)$$

where

$$\Psi(x; y; z) = \sum_{j=0}^{\infty} \frac{x^j}{(z + j)^y}$$

is known as Lerch's function - a generalisation of Riemann's zeta function and the polylogarithm function (see Spanier and Oldham [18] for further details).

Further simplification of (9) is possible:

$$i = \theta^{-3} e^{\mu_1} \left(-\theta F_1(0)(1 - F_2(0))(1 - \theta F_1(0)F_2(0)) + \frac{1}{2} \log \frac{1 + \theta F_1(0)}{1 - \theta F_1(0)(1 - 2F_2(0))} \right) + \frac{F_1'(0)^2 F_2'(0)^2}{F_1(0) + F_2(0) - F(0, 0)}, \quad (10)$$

which is valid for all $\theta \in \Theta$ except at $\theta = 0$. When $\theta = 0$, $\Psi(0; 1; 3) = \frac{1}{3}$ and

$$i = \frac{1}{6}F_1(0)F_1'(0) \left(1 + (1 - 2F_2(0))^3 \right) + \frac{F_1'(0)^2 F_2'(0)^2}{F_1(0) + F_2(0) - F_1(0)F_2(0)}. \quad (11)$$

Although it is a complicated function of θ , μ_1 , and μ_2 , Fisher's Information on θ , represented by (10) and (11), is in closed form.

⁴See the Appendix for a derivation of (9) and (10).

4.3 Analysis

An interesting analytic result may be deduced in the following scenario: fix θ and μ_1 , and allow μ_2 to increase, implying that should the first hurdle be passed, the second hurdle is increasingly passed. In the limit, this scenario corresponds to so-called *first-hurdle dominance* because the source of zeros on Y is due only to failing the first hurdle; for further details see [13, pp.25-26].⁵ As $\mu_2 \rightarrow \infty$, finds $F_2(0) \rightarrow 0$ and $i \rightarrow i_D$, where, from (9),

$$\begin{aligned} i_D &= \lim_{\mu_2 \rightarrow \infty} i \\ &= \frac{1}{2} F_1(0) F_1'(0) (\Psi(-\theta F_1(0); 1; 3) + \Psi(\theta F_1(0); 1; 3)) \\ &= \frac{1}{2} F_1(0) F_1'(0) \Psi(\theta^2 F_1(0)^2; 1; \frac{3}{2}) \end{aligned} \quad (12)$$

denotes Fisher's Information on θ under dominance.⁶ Clearly $i_D = i_D(\mu_1, \theta)$ is a function of μ_1 and θ . Furthermore, i_D is invariant to the sign of the dependency parameter θ ; *i.e.*, $i_D(\mu_1, \theta) = i_D(\mu_1, -\theta)$. Also, given that $\Psi(x_1; y; z) > \Psi(x_2; y; z)$ if $x_1 > x_2 \geq 0$, then i_D increases monotonically with the absolute value of θ . For any given value of $\gamma \in \Gamma$, these facts imply that:

$$i < i_D < \overline{i_D},$$

where

$$\begin{aligned} \overline{i_D} &= i_D(\mu_1, 1) \\ &= \frac{1}{2} F_1(0) F_1'(0) \Psi(F_1(0)^2; 1; \frac{3}{2}). \end{aligned}$$

In other words, for any given value of $\mu_1 \in \mathcal{R}$ and $\theta \in \Theta$, Fisher's Information on θ under dominance (i_D) exceeds Fisher's Information on θ (i) irrespective of the value of $\mu_2 \in \mathcal{R}$. And for the next part of the inequality, $\overline{i_D}$ represents the upper bound on Fisher's Information under dominance irrespective of the value of $\theta \in \Theta$, and therefore, by the previous point, must bound Fisher's Information on θ for all $\mu_2 \in \mathcal{R}$ and $\theta \in \Theta$ too.

At this point, it is convenient to rewrite (12) as:⁷

$$i_D = \frac{1}{2} \theta^{-3} e^{\mu_1} B\left(\frac{3}{2}; 0; \theta^2 F_1(0)^2\right),$$

for the derivative of this expression with respect to μ_1 can be written

$$\frac{\partial i_D}{\partial \mu_1} = i_D - \{2(1 - F_1(0))\} \times \left[\frac{F_1(0) F_1'(0)}{2(1 - \theta^2 F_1(0)^2)} \right].$$

Now the term within curly braces takes values between 0 and 2, and the term within square braces is larger than i_D ; hence, for some μ_1 the derivative may be negative. In particular, this is true at $\mu_1 = 0$ for all $\theta \in \Theta$, implying that i_D is larger for some $\mu_1 < 0$ than it is when $\mu_1 = 0$.

(Figure II about here)

Figure II plots four curves against values of μ_1 in the interval $[-4, 2]$. The uppermost curve is $\overline{i_D}$, the upper bound on Fisher's Information for all permissible choices of μ_2 and θ , as too it is the upper bound on Fisher's Information under dominance for all permissible θ . The next curve depicts Fisher's

⁵In statistical terms, the first hurdle dominant DDH model arises when there is no probability mass in the quadrant $(\{Y_1^{**} > 0\} \cap \{Y_2^{**} \leq 0\})$.

⁶Note that the relation $\Psi(x; y; z) + \Psi(-x; y; z) = 2^{1-y} \Psi(x^2; y; z/2)$ has been used.

⁷The relation given in the Appendix between the Lerch function and the incomplete beta function has been used.

Information under dominance for $\theta = \pm 0.9$, it bounds Fisher's Information i for all μ_2 and $|\theta| < 0.9$. The remaining two curves depict Fisher's Information i when $\mu_2 = 1$, with $\theta = -0.5$ and 0.9 .

Significantly, both i curves, “ i for $\theta = -0.5$ ” and “ i for $\theta = 0.9$ ”, exhibit the same behaviour as was deduced algebraically for i_D ; namely, that i_D is maximised for some $\mu_1 < 0$, for all $\theta \in \Theta$. Also, these curves provide evidence for the claim

$$\left. \frac{\partial i}{\partial \mu_1} \right|_{\mu_1=0} < 0,$$

implying that i is maximised for some $\mu_1 < 0$, for all $\theta \in \Theta$ and $\mu_2 \in \mathbb{R}$. Importantly, when random sampling from Y , under a bivariate logistic DDH model in which $\mu_1 < 0$, it must be the case that more than half of the sample would be expected to fail the first hurdle - in which case at least half of the observations in the sample would be recorded as zero. In other words, the situation in which more than half of the sample data are recorded as zero, is precisely the situation in which that data can yield greater Fisher Information on the dependency parameter θ . To illustrate the impact of this finding, consider DDH models for populations indexed by μ_1 , with set values $\mu_2 = 1$ and $\theta = -0.5$ (the Fisher Information on θ for these populations is the curve labelled “ i for $\theta = -0.5$ ”). In this setting, Fisher's Information on θ is maximised when $\mu_1 = -0.764$. Now at this value of μ_1 , the probability of failing the first hurdle is $F_1(0) = 0.682$, and the probability that a zero is recorded is $P(Y = 0) = 0.789$. Thus, when random sampling from the particular population in which $\mu_1 = -0.764$, $\mu_2 = 1$, and $\theta = -0.5$, although it is true that Fisher's Information on θ is maximised relative to all other populations indexed by μ_1 (θ and μ_2 held fixed), there is the harsh reality that approximately 80% of a (representative) sample will be observed as zero! In practice, a sample of data containing such a large mass of zeros might in all likelihood simply be set aside and not analysed, and yet it is precisely this type of sample which might yield enough statistical information on θ to support estimation and inference on θ .

Jones [14] gives an empirical example of a DDH model in which as many as 48.5% of the sample data is recorded with zero expenditure on tobacco (see Table I). It is an interesting exercise to derive the amount of Fisher Information on θ that results from a bivariate logistic DDH model with this proportion of zeros - taking, as before, the example of populations indexed by values of μ_1 , and for which $\theta = -0.5$ and $\mu_2 = 1$. To begin, the equation $P(Y = 0) = 0.485$ needs to be solved for μ_1 , yielding $\mu_1 \simeq 1$. Examining the relevant curve in Figure I at this value of μ_1 , finds Fisher's Information on θ approximately equal to 0.014. Contrasting this with the maximum Fisher Information on θ that is possible (approximately 0.038 when $\mu_1 = -0.764$), over 60% of the value of the latter information has been lost. It is therefore not surprising to see reports of unreliable estimates of dependency parameters in empirical applications of the DDH as well as insignificance of the DDH over its DH counterpart, for rarely will reported sample data consist of the large mass of zeros necessary to yield sufficient amounts of statistical information on the dependency parameter.

In this example, attention focused solely on Fisher's Information on the dependency parameter. Of course, DDH models will, in general, contain parameters in addition to the dependency parameter. In the following examples, consideration is given to the impact of dependency on all parameters in DDH models. In order to undertake this task, all elements of Fisher's Information matrix are required.

5 The Bivariate Poisson DDH

5.1 Distribution and Density

For this example, assume (Y_1^{**}, Y_2^{**}) is distributed according to Holgate's bivariate Poisson distribution (see [11]) with pdf:

$$P(\{Y_1^{**} = y_1^{**}\} \cap \{Y_2^{**} = y_2^{**}\}) = \delta \sum_{j=0}^{\min(y_1^{**}, y_2^{**})} \frac{\theta^j (\mu_1 - \theta)^{y_1^{**}-j} (\mu_2 - \theta)^{y_2^{**}-j}}{j! (y_1^{**} - j)! (y_2^{**} - j)!},$$

for $(y_1^{**}, y_2^{**}) \in \mathcal{N}^2$, where $\mathcal{N} = \{0, 1, 2, \dots\}$ is the set of non-negative integers, and

$$\delta = \exp(-\mu_1 - \mu_2 + \theta).$$

Holgate's distribution arises from the decomposition of the random variables into

$$Y_1^{**} = U + V,$$

and

$$Y_2^{**} = U + W,$$

where U , V , and W , are independent Poisson variables with parameters θ , $\mu_1 - \theta$, and $\mu_2 - \theta$, respectively. Clearly, the marginal pdf are $Y_1^{**} \sim Po(\mu_1)$, $Y_2^{**} \sim Po(\mu_2)$, with $\text{Cov}(Y_1^{**}, Y_2^{**}) = \theta$ the dependency parameter. In contrast to the previous examples, the values which θ may take depend on the values of the Poisson parameters μ_1 and μ_2 . In particular, assuming $(\mu_1, \mu_2) \in \mathbb{R}_+^2$, satisfying $\mu_1 > \theta$ and $\mu_2 > \theta$, where real $\theta \geq 0$, then $\theta \in \Theta = \{\theta : 0 \leq \theta < \min(\mu_1, \mu_2)\}$. The parameters $\gamma = (\mu_1, \mu_2, \theta) \in \Gamma = \mathbb{R}_+^2 \times \Theta$. When $\theta = 0$, Y_1^{**} and Y_2^{**} are independent.

Here (Y_1^{**}, Y_2^{**}) is discrete, so modification to the statistical theory presented in Section 2 is required. Fortunately, this task is straightforward, with a suitable DDH model constructed according to the same basic sequence of steps as has already been outlined. Zero observations on Y result from any pair recorded on either Y_1^{**} or Y_2^{**} axis, otherwise a positive observation, equivalent to the value of Y_2^{**} , is recorded on Y . Discreteness of (Y_1^{**}, Y_2^{**}) implies that Y is also discrete, with its pdf given by:

$$\begin{aligned} f_1(y) &= P(Y = y) \\ &= P(Y_2^{**} = y) - P(\{Y_1^{**} = 0\} \cap \{Y_2^{**} = y\}) \\ &= \frac{e^{-\mu_2} \mu_2^y}{y!} - \frac{\delta (\mu_2 - \theta)^y}{y!}, \end{aligned} \tag{13}$$

for $y \in \mathcal{N}_+ = \{1, 2, 3, \dots\}$ the set of positive integers, and

$$\begin{aligned} f_0 &= P(Y = 0) \\ &= P(Y_1^{**} = 0) + P(Y_2^{**} = 0) - P(\{Y_1^{**} = 0\} \cap \{Y_2^{**} = 0\}) \\ &= \exp(-\mu_1) + \exp(-\mu_2) - \delta. \end{aligned}$$

In the following analysis it is convenient to define, for any $y \in \mathcal{N}$, the following functions:

$$g(y) = \frac{e^{-\mu_2} \mu_2^y}{y!},$$

and

$$h(y) = \frac{e^{-(\mu_2 - \theta)} (\mu_2 - \theta)^y}{y!}.$$

Clearly g is the pdf of Y_2^{**} , a $Po(\mu_2)$ variable, and h is the pdf of W , a $Po(\mu_2 - \theta)$ variable. With these definitions (13), for example, may be re-written:

$$f_1(y) = g(y) - e^{-\mu_1} h(y).$$

5.2 Fisher's Information Matrix

Let $i = \{i_{jk}\}$ denote the (3×3) Fisher Information matrix, where $j, k \in \{1, 2, 3\}$. The j, k element of i is given by:

$$\begin{aligned} i_{\gamma_j, \gamma_k} &= E \left(\frac{\partial}{\partial \gamma_j} \log f(y) \times \frac{\partial}{\partial \gamma_k} \log f(y) \right) \\ &= \sum_{y=0}^{\infty} \frac{1}{f(y)} \frac{\partial f(y)}{\partial \gamma_j} \frac{\partial f(y)}{\partial \gamma_k}, \end{aligned}$$

where γ_j denotes the j th component of γ , and $f(y) = P(Y = y)$, $y \in \mathbb{N}$, denotes the pdf of Y . The derivatives required to determine the components of Fisher's Information matrix are relatively straightforward to derive; *e.g.*, those with respect to θ are:

$$\frac{\partial f_0}{\partial \theta} = -\delta,$$

and

$$\frac{\partial f_1(y)}{\partial \theta} = \begin{cases} -\delta(\mu_2 - \theta - 1) & \text{if } y = 1 \\ -e^{-\mu_1} (h(y) - h(y - 1)) & \text{if } y \in \{2, 3, \dots\}. \end{cases}$$

Thus, for example, the element of Fisher's Information matrix corresponding to Fisher's Information on θ , is equal to

$$i_{\theta, \theta} = \frac{\delta^2}{f_0} + \frac{\delta^2(\mu_2 - \theta - 1)^2}{f_1(1)} + e^{-2\mu_1} \sum_{y=2}^{\infty} \frac{1}{f_1(y)} (h(y) - h(y - 1))^2.$$

Although this expression cannot be analytically closed, it can, for a given value of γ , be computed to convergence according to a suitable degree of numerical tolerance.

5.3 Analysis

Table II gives Fisher's Information matrices i at selected values of $\gamma \in \Gamma$. In particular, μ_1 is fixed at unity, μ_2 takes values 1, 2, and 8, and θ is varied from 0 to 0.9 in increments of 0.1. Only the upper diagonal of the matrix is given for Fisher Information matrices are symmetric. The order of the entries matches the sequence in $\gamma = (\mu_1, \mu_2, \theta)$. Numbers are reported rounded to 4 decimal places.

(Table II about here)

Consider the bivariate Poisson DDH model for which the population is described by $\gamma = (1, 1, 0.9)$. As the true value of $\theta = 0.9$, it would be hoped that inference in a (representative) random sample drawn from this population would support the dependency hypothesis. This may be plausible, for the magnitude of Fisher's Information on θ appears to be quite strong, $i_{\theta, \theta} = 0.5488$.⁸ Certainly, for example, $i_{\theta, \theta}$ for the population described by $\gamma = (1, 1, 0.1)$ is far lower at 0.1489, and so this is in accordance with the intuition that there should be more Fisher Information on θ the further from zero the value of θ truly is. The impact of this positive finding does, however, need to be modified once account is taken of Fisher Information on all parameters. Examining the leading-diagonal entries of Fisher's Information matrix (*i.e.*, i_{μ_1, μ_1} , i_{μ_2, μ_2} , and $i_{\theta, \theta}$) throughout the first column of the table, there is clearly a trade-off between Fisher Information on μ_1 and θ .⁹ Indeed, at $\gamma = (1, 1, 0.9)$, there is almost no information on μ_1 , and so what has been gained in terms of information on θ at this point, is countered by a severe loss in information on μ_1 .

In the second column of Table II, $(\mu_1, \mu_2) = (1, 2)$. The trade-off in Fisher Information on the parameters is still in evidence, however, the magnitudes of i_{μ_1, μ_1} and i_{μ_2, μ_2} across the different values of θ does not find either disappearing, as almost occurred previously. Nevertheless, the situation in respect of Fisher Information on θ has worsened. Across all values of θ in this column, $i_{\theta, \theta}$ remains fairly constant and fairly small. There is little information on θ , irrespective of its true value.

In the third column of Table II, $(\mu_1, \mu_2) = (1, 8)$. In this case $P(Y_2^{**} = 0) = e^{-8}$. Accordingly, because there is almost no probability mass along the line $Y_2^{**} = 0$, each bivariate Poisson DDH population is *close* to first hurdle dominant.¹⁰ This time the trade-off in Fisher Information on the

⁸ "Strong" is meant in the sense that its numerical value is larger than all other values of $i_{\theta, \theta}$ given in the table.

⁹ i_{μ_2, μ_2} tends to move in line with $i_{\theta, \theta}$; hence, as statistical information on μ_1 trades-off against statistical information on θ , so too does μ_2 trade-off against μ_1 in respect of statistical information.

¹⁰ In fact, $P(Y = 0) \simeq P(Y_1^{**} = 0) = e^{-1}$; the former probability varying from 0.3681 when $\theta = 0$, to 0.3679 when $\theta = 0.9$.

parameters is barely detectable, with (at the tabulated points) i_{μ_1, μ_1} and $i_{\theta, \theta}$ increasing in θ , while i_{μ_2, μ_2} drops slightly at first, before increasing in θ . Scanning down the column, there is little Fisher Information on μ_2 whatever the value of θ , and barely any Fisher Information on θ itself, irrespective of the value of θ . Given the scarcity of Fisher Information on θ , there seems little chance of data (except perhaps if it is collected in large quantity) being able to reliably estimate θ , much less it being able to support the dependency hypothesis *even when θ truly is non-zero*.

6 A Non-Identically Distributed Population

Presented so far has been theory and analysis for iid populations. There may be some improvement in statistical information on the parameters of the DDH model (especially in respect of the dependency parameter) if these assumptions are weakened. Although there are many differing directions in which iid-ness may be relaxed, for the purposes of this illustration independence will be maintained and non-identity will be introduced through the mean parameters - the context remains that of the bivariate Poisson distribution.

Let $(Y_{1p}^{**}, Y_{2p}^{**})$ denote the relevant utility random variables of the p th individual in some (large) population of interest, assume they have joint pdf:

$$P(\{Y_{1p}^{**} = y_{1p}^{**}\} \cap \{Y_{2p}^{**} = y_{2p}^{**}\}) = \delta_p \sum_{j=0}^{\min(y_{1p}^{**}, y_{2p}^{**})} \frac{\theta^j (\mu_{1p} - \theta)^{y_{1p}^{**} - j} (\mu_{2p} - \theta)^{y_{2p}^{**} - j}}{j! (\mu_{1p}^{**} - j)! (\mu_{2p}^{**} - j)!}, \quad (14)$$

for $(y_{1p}^{**}, y_{2p}^{**}) \in \mathbb{N}^2$ and $p \in \mathbb{N}_+$; note that μ_{1p} and μ_{2p} are indexed by p , whereas the dependency parameter θ is not. Assume (14) holds across all individuals' utility variables; moreover, assume that all pairs are mutually independent.

For purposes of this illustration, a particularly simple parameterisation is adopted, namely: $\mu_1 \in \mathbb{R}_+$ is assumed constant for all p ; and, to incorporate non-identity, $\mu_2 = \beta x$, where parameter $\beta \in \mathbb{R}_+$ and covariate $x \in \{1, 2, 8\}$ for all p .¹¹ In this case, the dependency parameter $\theta \in \Theta = \{\theta : 0 \leq \theta < \min(\mu_1, \beta)\}$. The list of parameters is therefore $\gamma = (\mu_1, \beta, \theta) \in \Gamma = \mathbb{R}_+^2 \times \Theta$.

For any given individual, their Fisher Information matrix depends on their particular value of x , and as there are only three levels of x in this example, let these matrices be i_1 , i_2 , and i_8 , in an obvious notation.¹² To avoid further unnecessary complications, assume that a representative random sample¹³ is drawn from the population, with sample proportions corresponding to true proportions t_1 , t_2 , and t_8 ($t_1, t_2, t_8 \geq 0$, $t_1 + t_2 + t_8 = 1$) of each level of x in the population; collect the list of weights in the list $t = (t_1, t_2, t_8)$. With every assumption in place, Fisher's Information matrix is given by

$$i = t_1 i_1 + t_2 i_2 + t_8 i_8.$$

The difference between the iid and non-iid settings is now apparent: iid examples are generated only by combinations of t of $(1, 0, 0)$, $(0, 1, 0)$, or $(0, 0, 1)$, whereas non-iid examples arise from weighting combinations with at least two positive components. Therefore, the problems observed under iid-ness may, in non-iid settings, be alleviated due to averaging - for i is a weighted average across individual Fisher Information matrices.

¹¹More appropriate in practice would be the parameterisations $\exp(\alpha'w)$ and $\exp(\beta'x)$, for μ_1 and μ_2 , respectively. Covariate vectors w and x may have common components, while parameters α and β are finite-dimensional real vectors.

¹²For example, if $\gamma = (1, 1, 0.5)$ then i_8 may be obtained from the $(1, 8, 0.5)$ cell in Table II as per

$$\begin{matrix} \mu_1 \\ \beta \\ \theta \end{matrix} \begin{bmatrix} 0.5990 & -0.0357 \times 8 & -0.0228 \\ & 0.0783 \times 64 & 0.0453 \times 8 \\ & & 0.0307 \end{bmatrix},$$

where the multiplicative factors in the second row and column are due to the reparameterisation of μ_2 to β .

¹³An alternative device that achieves the same end is to assume n is large, as then sample proportions tend towards true proportions.

To keep matters simple, set $\mu_1 = 1$ and $\beta = 1$, and allow θ to vary. Table III gives Fisher Information i on $\gamma = (\mu_1, \beta, \theta)$ for just two weighting combinations, namely, $t = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ and $t = (\frac{1}{6}, \frac{1}{6}, \frac{2}{3})$: the entries are computed by taking the weighted average across the appropriate row in Table II, not forgetting to account for the effect of the reparameterisation from μ_2 to β .

(Table III about here)

In the first column the weighting scheme is $t = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. Evident once again is a trade-off in information between parameters μ_1 and β , but unlike the iid case, the trade-off is of lesser consequence; *i.e.*, as information on μ_1 increases, the information on β is not wiped out (and *vice versa*). The statistical information on θ , while at first decreasing, improves as the true value of θ becomes larger, especially in the last jump from $\theta = 0.7$ to $\theta = 0.9$. Although relaxing iid-ness has improved the situation, there nevertheless persists a relative lack of statistical information on θ irrespective of its true value.

In the first column of Table III, only one-third of the population is (close to) first-hurdle dominant; *i.e.*, one-third of the population have a covariate value of $x = 8$. This is reversed in the second column of Table III, with the first-hurdle dominant proportion now two-thirds - reflected in the weighting scheme $t = (\frac{1}{6}, \frac{1}{6}, \frac{2}{3})$ that is chosen. As this weighting scheme emphasizes the i_8 component of i , and as that component adds little to the information on θ , irrespective of its true value, the consequent lack of support by data for the dependency hypothesis will again be manifest. Interestingly, for a fixed θ , the magnitude of information on the mean parameter μ_1 and on the regression parameter β are greater in the second column than in the first. This was certainly not the case under iid-ness when information on μ_2 tended to disappear the closer to first-hurdle dominant the population became (see the third column of Table II). So while the dependency parameter θ appears of little statistical consequence in double-hurdle models (for iid and non-iid populations), the situation in regard to the mean/regression parameters in non-iid populations appears to be the opposite - they remain important for statistical information is always present upon them - which is vital when it comes to applying these models in practice.

7 Remarks

Taken as a whole, the results of this paper demonstrate that the DDH model represents a spurious statistical generalisation of the DH model. The economic underpinnings of the model are not affected by this conclusion; nor does it invalidate the behavioural arguments mounted to justify the DDH model over the DH model. It is the statistical nature of the DDH model which is deficient. This has manifested itself in the empirical literature, with most studies being unable to support the existence of the dependency parameter, and it has been studied in this paper under ideal theoretical circumstances through means of Fisher's Information, a measure of the amount of statistical information that is contained on model parameters - the larger the information, the better.

Beginning with iid populations, a four-parameter bivariate normal DDH was shown to be weakly identified. With the bivariate logistic DDH model it was shown that populations which generate an excessive proportion of zero observations were those which produce the greatest Fisher Information on the dependency parameter. Finally, in a bivariate Poisson DDH it was shown that there exist trade-offs in Fisher Information across all parameters in the model; for a population in which Fisher Information on the dependency parameter is relatively large, the cost incurred may be relatively little provision of Fisher Information on other parameters. Also, it was shown that Fisher's Information on the dependency parameter can remain relatively constant irrespective of its true value. Under non-iid bivariate Poisson DDH populations, the extremes seen in the iid setting were alleviated slightly, but not enough to warrant confidence in the DDH model.

In practice, knowing the true DDH model is no longer the luxury it has been here. The indicator - an excessive proportion of zeros in the data - may provide favourable evidence to justify fitting

a DDH model, but taken in the broader perspective of all parameters in the model, it may be a costly strategy. To the extent that mean/regression parameters are usually of greater importance in estimation, it would appear safer to ignore dependency altogether and specify a DH model, the data can then reveal as much about these parameters as is possible.

This seemingly negative conclusion to the paper is, however, a boon to practitioners for it allows far more flexible distributional structures to be employed for the models' random variables. To see this, suppose (against statistical advice) that a practitioner wishes to allow for dependency, and attempts to specify a DDH model. In this event, they are obliged to determine a suitable bivariate distribution F , for as has been shown in (5) and (6), its functional form enters the pdf of Y . The options available are few, the bivariate normal being the typical choice. Reduced too, is the ability to employ flexible distributional forms.¹⁴ Now suppose the practitioner desires to fit a DH model instead. By construction the underlying utility variables are independent, in which case $F(y_1^{**}, y_2^{**}) = F_1(y_1^{**})F_2(y_2^{**})$, and the pdf of Y becomes:

$$f_1(y) = (1 - F_1(0)) \frac{\partial F_2(y)}{\partial y},$$

for $y \in \mathbb{R}_+$, and at the origin

$$f_0 = F_1(0) + F_2(0) - F_1(0)F_2(0).$$

Now only univariate distributions F_1 and F_2 are required, consequently there are many more options available. The normal-normal combination (a normal distribution for F_1 and a normal distribution for F_2) is standard; however, it is conceivable that another mixture may be more suitable in some situations, a logistic-Laplace mixture, for example.

Attention in this paper has focused mainly on the relationship between (Y_1^{**}, Y_2^{**}) and Y , through the specification of a suitable distribution for the former. However, once dependency is eliminated from consideration and efforts are devoted to specifying a DH model, the independency that also exists between Y_1^* and Y_2^* suggests a second approach to modelling - one that emphasizes specification of distributions for the hurdle variables Y_1^* and Y_2^* .¹⁵ Of course, specifying a distribution for Y_1^* is easy, it must be Bernoulli distributed:

$$P(Y_1^* = y_1^*) = (1 - r)^{1 - y_1^*} r^{y_1^*},$$

where $y_1^* \in \{0, 1\}$, and real r is such that $0 \leq r \leq 1$. The success probability r may depend on parameters and covariates, and can be parameterised with any function whose range is $(0, 1)$; *e.g.*, the cdf of the normal distribution yields the familiar probit, but possibly more flexible would be the pdf of a beta distribution. For the second hurdle variable Y_2^* , assume, for the moment, that it is observable. Those observations would give Y_2^* the appearance of being zero-inflated, hence it would be natural to specify a distribution for it from amongst this class. There are a number of possibilities; *e.g.*, Aalen's [1] distribution is a flexible compound Poisson-gamma mixture, while simpler perhaps is Castoldi's (see [12, p.289]) continuous analogue of the Poisson distribution with a mass at the origin. To derive the pdf of Y , suppose that the pdf of Y_2^* is given by:

$$g(y_2^*) = \begin{cases} g_1(y_2^*) & \text{if } y_2^* \in \mathbb{R}_+ \\ g_0 & \text{if } y_2^* = 0, \end{cases}$$

for suitable functions g_1 and g_0 , both of which may depend on parameters and covariates. Substituting into the second line of (4) and exploiting independence, yields the distribution of Y as:

$$\begin{aligned} P(Y < y) &= 1 - P(Y_1^* = 1)P(Y_2^* > y) \\ &= 1 - r \int_y^\infty g_1(y_2^*) dy_2^*, \end{aligned}$$

¹⁴The distinction between flexibility in distributional choice and flexibility in parameterisation may be highlighted with Yen and Jones' [19] DDH study. There a parametric transformation of Y was employed to induce a more flexible functional form to the model, and yet bivariate normality of the underlying utility variables was maintained.

¹⁵Dionne *et al.* [6] is one such study which took this route. In their case $Y \in \mathbf{N}$ was discrete, and the hurdle variables, Y_1^* and Y_2^* , were modelled respectively by logit and negative binomial distributions.

where $y \in \mathbb{R}_+$. As per (5), the derivative of the cdf yields the pdf of Y when $y \in \mathbb{R}_+$:

$$f_1(y) = rg_1(y).$$

And from (6), at the origin

$$\begin{aligned} f_0 &= P(Y_1^* = 0) + P(Y_1^* = 1)P(Y_2^* = 0) \\ &= 1 - r + rg_0. \end{aligned}$$

It is a subject of future research to contrast the performance of DH models based on these two approaches to modelling.

References

- [1] Aalen, O. O. (1992), "Modelling heterogeneity in survival analysis by the compound Poisson distribution", *The Annals of Applied Probability*, **2**, 951-972.
- [2] Blaylock, J. R., and Blisard, W. N. (1992), "U.S. cigarette consumption: the case of low-income women", *American Journal of Agricultural Economics*, **74**, 698-705.
- [3] Blundell, R., and Meghir, C. (1987), "Bivariate alternatives to the Tobit model", *Journal of Econometrics*, **34**, 179-200.
- [4] Burton, M., Tomlinson, M., and Young, T. (1994), "Consumers' decisions whether or not to purchase meat: a double hurdle analysis of single adult households", *Journal of Agricultural Economics*, **45**, 202-212.
- [5] Cragg, J. G. (1971), "Some statistical models for limited dependent variables with applications to the demand for durable goods", *Econometrica*, **39**, 829-844.
- [6] Dionne, G., Artis, M., and Guillen, M. (1996), "Count data models for a credit scoring system", *Journal of Empirical Finance*, **3**, 303-325.
- [7] Fisher, R. A. (1925), "Theory of statistical estimation", *Proceedings of the Cambridge Philosophical Society*, **22**, 700-725. (Also (1950), *Contributions to Mathematical Statistics*, Paper 11, New York: Wiley.)
- [8] Garcia, J., and Labeaga, J. M. (1996), "Alternative approaches to modelling zero expenditure: an application to Spanish demand for tobacco", *Oxford Bulletin of Economics and Statistics*, **58**, 489-506.
- [9] Gould, B. W. (1992), "At-home consumption of cheese: a purchase-infrequency model", *American Journal of Agricultural Economics*, **72**, 453-459.
- [10] Gumbel, E. J. (1961), "Bivariate logistic distributions", *Journal of the American Statistical Association*, **56**, 335-349.
- [11] Holgate, P. (1964), "Estimation for the bivariate Poisson distribution", *Biometrika*, **51**, 241-245.
- [12] Johnson, N. L., and Kotz, S. (1970), *Distributions in Statistics: Continuous Univariate Distributions-2*, New York: Wiley.
- [13] Jones, A. M. (1989), "A double-hurdle model of cigarette consumption", *Journal of Applied Econometrics*, **4**, 23-39.
- [14] Jones, A. M. (1992), "A note on computation of the double-hurdle model with dependence with an application to tobacco expenditure", *Bulletin of Economic Research*, **44**, 67-74.

- [15] Kullback, S. (1959), *Information Theory and Statistics*, New York: Wiley.
- [16] Pudney, S. (1989), *Modelling Individual Choice: the Econometrics of Corners, Kinks, and Holes*, London: Basil Blackwell.
- [17] Shonkwiler, J. S., and Shaw, W. D. (1996), "Hurdle count-data models in recreation demand analysis", *Journal of Agricultural and Resource Economics*, **21**, 210-219.
- [18] Spanier, J., and Oldham, K. B. (1987), *An Atlas of Functions*, Washington: Hemisphere.
- [19] Yen, S. T., and Jones, A. M. (1997), "Household consumption of cheese: an inverse hyperbolic sine double-hurdle model with dependent errors", *American Journal of Agricultural Economics*, **79**, 246-251.
- [20] Zorn, C. J. W. (1998), "An analytic and experimental examination of zero-inflated and hurdle Poisson specifications", *Sociological Methods and Research*, **26**, 368-400.

8 Appendix

For the continuous component of the pdf, the contribution to Fisher's Information on θ is given by the second term on the right-hand-side of (7). Substitution into this expression yields

$$F_1(0)F_1'(0) \int_0^\infty \frac{F_2'(y)(1 - 2F_2(y))^2}{1 - \theta F_1(0)(1 - 2F_2(y))} dy.$$

Now the denominator of the integrand may be expanded binomially; that is,

$$(1 - \theta F_1(0)(1 - 2F_2(y)))^{-1} = \sum_{j=0}^{\infty} \theta^j F_1(0)^j (1 - 2F_2(y))^j.$$

This infinite series representation is uniformly convergent because $|\theta F_1(0)(1 - 2F_2(y))| < 1$ for all $(y, \gamma) \in \mathbb{R}_+ \times \Gamma$, therefore it is permissible to integrate term-by-term. The contribution is solved as:

$$\begin{aligned} & F_1(0)F_1'(0) \sum_{j=0}^{\infty} (\theta F_1(0))^j \int_0^\infty F_2'(y)(1 - 2F_2(y))^{j+2} dy \\ &= F_1(0)F_1'(0) \sum_{j=0}^{\infty} (\theta F_1(0))^j \frac{-1}{2(j+3)} \left((-1)^{j+3} - (1 - 2F_2(0))^{j+3} \right) \\ &= \frac{1}{2} F_1(0)F_1'(0) \left(\Psi(-\theta F_1(0); 1; 3) + (1 - 2F_2(0))^3 \Psi(\theta F_1(0)(1 - 2F_2(0)); 1; 3) \right), \end{aligned}$$

where $\Psi(\cdot)$ denotes Lerch's function.

When $y = 0$, substituting into the first term on the right-hand-side of (7) yields the contribution to Fisher's Information on θ as:

$$\frac{F_1'(0)^2 F_2'(0)^2}{F_1(0) + F_2(0) - F(0,0)}.$$

Combining the contributions yields Fisher's Information on θ in the bivariate logistic DDH model given in (9).

The derivation of (10) is achieved by replacing the Lerch functions in (9) according to the relation (see [18, eq.64:12:13])

$$\Psi(x; 1; z) = x^{-z} B(z; 0; x),$$

where $B(a; b; x) = \int_0^x t^{a-1}(1-t)^{b-1} dt$ denotes the incomplete beta function. The particular result

$$B(3; 0; x) = - \left(x + \frac{1}{2}x^2 + \log(1-x) \right),$$

valid for all real x such that $|x| < 1$, is then applied. Finally, observe that $F_1(0)^{-2} F_1'(0) = \exp(\mu_1)$.

TABLE II: Fisher's Information Matrix (bivariate Poisson)

θ	$\mu_1 = 1, \mu_2 = 1$	$\mu_1 = 1, \mu_2 = 2$	$\mu_1 = 1, \mu_2 = 8$
0	$\begin{bmatrix} \mu_1 & 0.2254 & 0.2254 & 0.1312 \\ \mu_2 & & 0.4896 & 0.2850 \\ \theta & & & 0.1658 \end{bmatrix}$	$\begin{bmatrix} 0.4083 & 0.1098 & 0.0639 \\ & 0.2467 & 0.1435 \\ & & 0.0835 \end{bmatrix}$	$\begin{bmatrix} 0.5814 & 0.0003 & 0.0002 \\ & 0.0788 & 0.0459 \\ & & 0.0267 \end{bmatrix}$
0.1	$\begin{bmatrix} 0.1981 & 0.1755 & 0.1249 \\ & 0.4657 & 0.2560 \\ & & 0.1489 \end{bmatrix}$	$\begin{bmatrix} 0.3974 & 0.0827 & 0.0593 \\ & 0.2426 & 0.1343 \\ & & 0.0760 \end{bmatrix}$	$\begin{bmatrix} 0.5821 & -0.0069 & -0.0040 \\ & 0.0783 & 0.0456 \\ & & 0.0267 \end{bmatrix}$
0.2	$\begin{bmatrix} 0.1724 & 0.1305 & 0.1239 \\ & 0.4557 & 0.2288 \\ & & 0.1452 \end{bmatrix}$	$\begin{bmatrix} 0.3883 & 0.0567 & 0.0572 \\ & 0.2421 & 0.1256 \\ & & 0.0713 \end{bmatrix}$	$\begin{bmatrix} 0.5841 & -0.0141 & -0.0083 \\ & 0.0781 & 0.0454 \\ & & 0.0270 \end{bmatrix}$
0.3	$\begin{bmatrix} 0.1470 & 0.0896 & 0.1255 \\ & 0.4592 & 0.1995 \\ & & 0.1521 \end{bmatrix}$	$\begin{bmatrix} 0.3804 & 0.0314 & 0.0572 \\ & 0.2450 & 0.1169 \\ & & 0.0684 \end{bmatrix}$	$\begin{bmatrix} 0.5874 & -0.0212 & -0.0127 \\ & 0.0780 & 0.0453 \\ & & 0.0278 \end{bmatrix}$
0.4	$\begin{bmatrix} 0.1211 & 0.0526 & 0.1277 \\ & 0.4763 & 0.1652 \\ & & 0.1702 \end{bmatrix}$	$\begin{bmatrix} 0.3729 & 0.0065 & 0.0590 \\ & 0.2513 & 0.1078 \\ & & 0.0669 \end{bmatrix}$	$\begin{bmatrix} 0.5923 & -0.0284 & -0.0175 \\ & 0.0780 & 0.0452 \\ & & 0.0290 \end{bmatrix}$
0.5	$\begin{bmatrix} 0.0948 & 0.0203 & 0.1286 \\ & 0.5086 & 0.1229 \\ & & 0.2013 \end{bmatrix}$	$\begin{bmatrix} 0.3656 & -0.0184 & 0.0627 \\ & 0.2613 & 0.0978 \\ & & 0.0666 \end{bmatrix}$	$\begin{bmatrix} 0.5990 & -0.0357 & -0.0228 \\ & 0.0783 & 0.0453 \\ & & 0.0307 \end{bmatrix}$
0.6	$\begin{bmatrix} 0.0687 & -0.0059 & 0.1256 \\ & 0.5578 & 0.0690 \\ & & 0.2489 \end{bmatrix}$	$\begin{bmatrix} 0.3578 & -0.0434 & 0.0685 \\ & 0.2753 & 0.0865 \\ & & 0.0674 \end{bmatrix}$	$\begin{bmatrix} 0.6078 & -0.0431 & -0.0290 \\ & 0.0787 & 0.0455 \\ & & 0.0330 \end{bmatrix}$
0.7	$\begin{bmatrix} 0.0439 & -0.0241 & 0.1158 \\ & 0.6267 & -0.0006 \\ & & 0.3178 \end{bmatrix}$	$\begin{bmatrix} 0.3489 & -0.0686 & 0.0765 \\ & 0.2936 & 0.0734 \\ & & 0.0694 \end{bmatrix}$	$\begin{bmatrix} 0.6192 & -0.0507 & -0.0363 \\ & 0.0794 & 0.0458 \\ & & 0.0363 \end{bmatrix}$
0.8	$\begin{bmatrix} 0.0222 & -0.0317 & 0.0953 \\ & 0.7190 & -0.0910 \\ & & 0.4145 \end{bmatrix}$	$\begin{bmatrix} 0.3383 & -0.0940 & 0.0869 \\ & 0.3168 & 0.0576 \\ & & 0.0732 \end{bmatrix}$	$\begin{bmatrix} 0.6344 & -0.0586 & -0.0456 \\ & 0.0802 & 0.0462 \\ & & 0.0410 \end{bmatrix}$
0.9	$\begin{bmatrix} 0.0064 & -0.0253 & 0.0590 \\ & 0.8402 & -0.2095 \\ & & 0.5488 \end{bmatrix}$	$\begin{bmatrix} 0.3251 & -0.1194 & 0.1001 \\ & 0.3458 & 0.0381 \\ & & 0.0794 \end{bmatrix}$	$\begin{bmatrix} 0.6551 & -0.0669 & -0.0580 \\ & 0.0812 & 0.0468 \\ & & 0.0480 \end{bmatrix}$

TABLE III: Fisher's Information Matrix (non-iid bivariate Poisson)

θ		$t = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$	$t = (\frac{1}{6}, \frac{1}{6}, \frac{2}{3})$
0.1	μ_1	0.3925	0.4873
	β	0.0952	0.0200
	θ	0.0601	0.0280
0.3		2.1498	3.5816
		0.2964	0.3305
		0.0838	0.0553
0.5		0.3716	0.4795
		-0.0058	-0.0878
		0.0567	0.0220
0.7		2.1426	3.5658
		0.2651	0.3136
		0.0828	0.0553
0.9		0.3531	0.4761
		-0.1007	-0.1931
		0.0562	0.0167
		2.1882	3.5994
		0.2270	0.2947
		0.0995	0.0651
		0.3373	0.4783
		-0.1890	-0.2973
		0.0520	0.0078
		2.2935	3.6865
		0.1708	0.2684
		0.1412	0.0887
		0.3289	0.4920
		-0.2663	-0.4007
		0.0337	-0.0122
		2.4743	3.8368
		0.0805	0.2276
		0.2254	0.1367