On dependency in double-hurdle models

Smith, Murray D.

Published in:
Statistical Papers

DOI:
10.1007/BF02926011

Publication date:
2003

Citation for published version (APA):
Smith, M. D. (2003). On dependency in double-hurdle models. *Statistical Papers*, *44*(4), 581-595.
https://doi.org/10.1007/BF02926011

Statistical Papers **44**, 581–595 (2003)

Statistical Papers

On dependency in double-hurdle models

Murray D. Smith

Econometrics and Business Statistics, University of Sydney, Sydney, NSW 2006, Australia (e-mail: Murray.Smith@econ.usyd.edu.au)

Received: December 6, 2001

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 simultaneous two-step process: a market participation decision (first hurdle), and 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.

Key words: Dependency; Weak identification; Fisher's information

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 [10]. 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 two-step decomposition of the individual's decision process: (i) a market participation decision (whether to buy or not), and (ii) a consumption level decision (how much to buy). The statistical origins of the DH model are due to Cragg [5], and its basis in consumer choice theory is due to Pudney [14, 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, the latter silent on this issue. Rather, justification has been based on intuitive behavioural grounds (*e.g.* [7, p.364], [8, p.491], [15, 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

Reference:	Application:	Sample Size	% of 0's	DDH vs DH
[2]	Cigarettes (USA)	2962	60.7	insig
[4]	Meat (UK)	2144	6.3	insig
[7]	Rice (USA)	4273	67.0	insig & sig
[8]	Cigarettes (Spain)	23669	41.2	insig
[9]	Cheese (USA)	5017	59.0	sig
[10]	Cigarettes (UK)	1573	na	insig
[11]	Cigarettes (UK)	2321	48.5	insig
[12]	Beef (USA)	4150	14.7	sig
[17]	Donations (Canada)	13572	89.9	sig
[19]	Cheese (USA)	4245	18.1	insig

The entries in the last column "DDH vs DH" indicate 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 the majority of the studies did not support the DDH model over the DH model at any conventional level of significance. This feature of the DDH literature has also been noted in Gao *et al.* [7, p.364]. The 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 nested DH counterparts.

The DH and DDH models are members of the class of hierarchical limited and qualitative dependent variable models. This class of model often suffers parameter identification problems, although detection of this typically surfaces only when attempting to compute parameter estimates. However, by focusing on the distribution of the DDH model (as opposed to secondary issues such as the properties of estimators and test statistics used in the model), the problem of *weakly identified parameters* is shown to be present (Section 3).

In the paper's remaining sections, the extent of the identification problem is quantified using *Fisher's information* to measure the amount of statistical information on the parameters of the model. Section 4 focuses on Fisher's information on θ, while in Section 5, Fisher information matrices are inspected.

The main suggestion of the paper, is that the introduction of dependency into DH models (thereby yielding a DDH model) is a statistically spurious generalisation - the DDH model adds little to the informational content of its nested DH counterpart. This seemingly negative outcome is, however, of considerable benefit to practitioners. In the absence of dependency, there is greater opportunity to explore more flexible distributional forms. Section 6 concludes with a proposal on this theme.

2 Statistical Construction of the DDH Model

In this section, a brief description of the construction of the statistical distribution of DDH models is given. The derivation has general applicability.

For a given commodity, begin by defining Y_1^{**} as the utility derived by an individual from participation in the market for the commodity, and Y_2^{**} the utility derived by an individual from consumption of the commodity. Assume (for now) that these variables are continuous, and real-valued. Next, assume a parametric bivariate model for (Y_1^{**}, Y_2^{**}) is specified by assigning a joint cumulative distribution function (cdf), denoted by $F(y_1^{**}, y_2^{**})$, for real-valued pairs (y_1^{**}, y_2^{**}). The cdf depends upon unknown parameters, one in particular 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 the hurdle variables:

$$Y_1^* = 1\{Y_1^{**} > 0\} \qquad \text{and} \qquad Y_2^* = 1\{Y_2^{**} > 0\}\, Y_2^{**} \tag{1}$$

where $1\{A\}$ is the indicator function, taking value 1 if event A holds and 0 otherwise. 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 potentially a consumer), and Y_2^* represents the second hurdle consumption (if $Y_2^* = 0$ the hurdle is failed, the individual elects not to consume; if $Y_2^* > 0$ the hurdle is passed, the individual is potentially a consumer). In general, Y_1^* and Y_2^* are latent. Finally, to complete the construction of the DDH model, individual consumption

$$Y = Y_1^*\, Y_2^* \tag{2}$$

Due to the decomposition into separate decisions (jointly taken), a zero observation on Y can occur in either of two ways: (i) when the first hurdle is failed, or the first hurdle is passed and the second failed $(Y_1^* = 0) \cup (Y_1^* = 1 \cap Y_2^* = 0)$, and (ii) when the second hurdle is failed, or the first hurdle is failed but the second hurdle passed $(Y_2^* = 0) \cup (Y_1^* = 0 \cap Y_2^* > 0)$. A positive-valued observation on Y occurs only when both hurdles are passed $(Y_1^* = 1) \cap (Y_2^* > 0)$.

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

$$f(y) = \begin{cases} f_+(y) & \text{if } y > 0 \\ f_0 & \text{if } y = 0 \end{cases} \tag{3}$$

When $y > 0$, the $f_+(y)$ component may be derived by differentiating, with respect to y, the following probability:

$$\begin{aligned} \Pr(Y \leq y) &= 1 - \Pr(Y > y) \\ &= F_1(0) + F_2(y) - F(0, y) \end{aligned} \tag{4}$$

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

$$f_+(y) = \frac{\partial}{\partial y}\left(F_2(y) - F(0, y)\right) \tag{5}$$

which can be simplified once a specific functional form is assumed for F. When $y = 0$, the f_0 component is the probability mass at the origin:

$$\begin{aligned} f_0 &= \Pr(Y = 0) \\ &= F_1(0) + F_2(0) - F(0, 0) \end{aligned} \tag{6}$$

With the pdf, (5) and (6), it is, for example, easy to see how log-likelihood functions are constructed: $L = \sum_0 \log f_0 + \sum_+ \log f_+(y)$, where $\sum_0$ indicates summation over zero-valued observations on Y, and $\sum_+$ summation over positive valued observations.

3 The Bivariate Normal DDH Model

As a 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, $\mathrm{Var}(Y_1^{**})$ is normalised to unity because 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^{**}. The parameters of the model are $(\mu_1, \mu_2, \sigma^2, \theta)$.

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

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

where $\Omega(\cdot, \cdot; \theta)$ denotes the cdf of a standardised bivariate normal distribution with correlation coefficient θ. Substituting this expression into (4) and (6), the cdf of Y is given by:

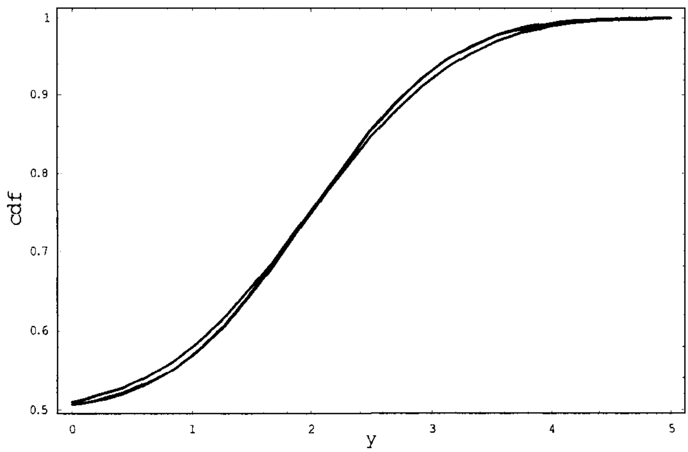
$$\Pr(Y \leq y) = \begin{cases} \Phi(-\mu_1) + \Phi(z) - \Omega(-\mu_1, z; \theta) & \text{if } y > 0 \\ \Phi(-\mu_1) + \Phi(-\sigma^{-1}\mu_2) - \Omega(-\mu_1, -\sigma^{-1}\mu_2; \theta) & \text{if } y = 0 \end{cases}$$

where $\Phi(\cdot)$ denotes the cdf of a standardised univariate normal distribution, and $z = (y - \mu_2)/\sigma$.

In Figure I, distributions of Y are plotted for three far-apart points in the parameter space; namely, at $(\mu_1, \mu_2, \sigma^2, \theta) = (0, 2, 1, 0), (0, 1.6, 1, 0.5), (0, 2.4, 1, -0.5)$ The significant feature to notice is that the three distributions are almost indistinguishable across the support of Y. Certainly, the fact that the distributions are not identical, 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 the parameter space. Of course, there are an infinity of parameter configurations 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 $\Pr(Y = 0)$ is a little over 50% for each parameter choice. 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 - as reported in the Burton *et al.* DDH study, see [4, p.205].

In the following examples, attempts to quantify the implications of weak identification in the DDH model are undertaken using *Fisher's information*, a well-known measure of statistical information.

Figure I: Distribution of Y



4 The Bivariate Logistic DDH Model

4.1 Distribution and Density

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

$$F(y_1^{**}, y_2^{**}) = F_1(y_1^{**})F_2(y_2^{**})(1 + \theta(1 - F_1(y_1^{**}))(1 - F_2(y_2^{**}))) \qquad (7)$$

for real-valued pairs (y_1^{**}, y_2^{**}). The notation

$$F_i(y_i^{**}) = (1 + \exp(-(y_i^{**} - \mu_i)))^{-1} \quad (i = 1, 2)$$

corresponds to the cdf of a logistic random variable with mean μ_i and variance $\pi^2/3$. Also, the dependency parameter θ is such that $-1 < \theta < 1$, moreover, it is equivalent to the covariance between Y_1^{**} and Y_2^{**}.

For this specification, the pdf of Y, applying (5) and (6), can be shown to equal:

$$f(y) = \begin{cases} (1 - F_1(0))(1 - \theta F_1(0)(1 - 2F_2(y)))F_2'(y) & \text{if } y > 0 \\ F_1(0) + F_2(0) - F(0,0) & \text{if } y = 0 \end{cases} \qquad (8)$$

where $F_i'(y) = \partial F_i(y)/\partial y = F_i(y)(1 - F_i(y))$. The parameters of this DDH model are (μ_1, μ_2, θ).

In the bivariate logistic DDH model, Fisher's information on θ, denote it by i, can be derived in closed form. This said, in what follows it is actually more convenient to work with an intermediate form:

$$\begin{aligned} i &= E\left(\frac{\partial}{\partial\theta}\log f(Y)\right)^2 \\ &= \frac{1}{2}F_1(0)F_1'(0)\left[\Psi(-\theta F_1(0); 1; 3) + \right. \\ &\qquad \left. +(1 - 2F_2(0))^3\ \Psi(\theta F_1(0)(1 - 2F_2(0)); 1; 3)\right] \\ &\qquad + \frac{F_1'(0)^2 F_2'(0)^2}{F_1(0) + F_2(0) - F(0,0)} \end{aligned} \qquad (9)$$

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 [16]).

4.2 Analysis

Consider 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 *first-hurdle dominance* because the source of zeros on Y is due only to failing the first hurdle - zero consumption results only from market non-participation. Allowing $\mu_2 \to \infty$, finds $F_2(0) \to 0$ and $i \to i_{\mathrm{D}}$, where, from (9),

$$\begin{aligned} i_{\mathrm{D}} &= \lim_{\mu_2 \to \infty} i \\ &= \frac{1}{2} F_1(0) F_1'(0)\ \Psi(\theta^2 F_1(0)^2; 1; \frac{3}{2}) \end{aligned} \tag{10}$$

denotes Fisher's information on θ under first-hurdle dominance. Clearly, i_{D} is a function of μ_1 and θ, moreover, it is invariant to the sign of the θ, and increases monotonically with θ^2. For any given set of values for the parameters, these facts imply that $i < i_{\mathrm{D}}$.

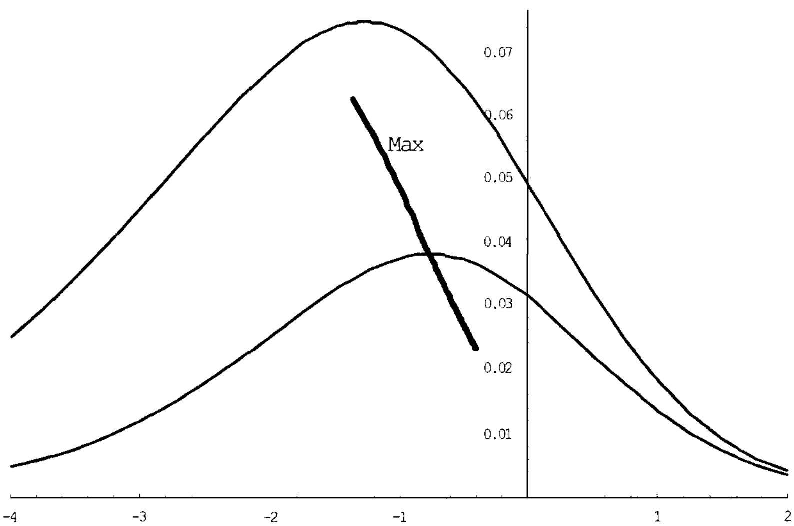
Now it can be shown that:

$$\left. \frac{\partial i_{\mathrm{D}}}{\partial \mu_1} \right|_{\mu_1 = 0} < 0 \quad \forall\, \theta \tag{11}$$

which implies that i_{D} is larger for some $\mu_1 < 0$, than it is when $\mu_1 = 0$. In turn, this means that statistical information on θ will be maximised only when more than half of the population do not participate in the market. In other words, when sampling from a first-hurdle dominant population, if more than 50% of respondents announce zero consumption, then it is under these conditions that we are best-placed to perform inference on θ, for this is precisely the situation when the amount of statistical information present on θ can be at its greatest.

Figure II plots i_{D} and i against values of μ_1 in the interval $[-4, 2]$. The upper curve is i_{D}, evaluated at $\theta = 1$. This curve bounds Fisher's information i for all μ_2 and θ in the parameter space. The lower curve depicts Fisher's information i, setting $\mu_2 = 1$ and $\theta = -0.5$. Significantly, the plot of i exhibits the same behaviour as was deduced algebraically for i_{D}, providing evidence for the *claim*:

Figure II: Fisher's information



$$\left.\frac{\partial i}{\partial \mu_1}\right|_{\mu_1=0} < 0 \quad \forall\, \theta \text{ and } \mu_2 \tag{12}$$

implying that i is maximised for some $\mu_1 < 0$, whatever the value of θ and μ_2. The claim is further evidenced by the thicker line, labelled Max, which traverses through all maximums of i for every θ such that $-1 < \theta < 1$, where μ_2 is fixed at unity; this line is situated entirely over negative values of μ_1. If the claim (12) is true, then the earlier remark in regard to excess zeros in first-hurdle dominant populations (11) now applies in all populations - a sample containing over 50% zero observations will always indicate the best conditions under which the dependency parameter θ can be reliably estimated. This feature is seen empirically, with Gao *et al.* [7] reporting 67% of their sample with zero consumption, for Gould [9] the proportion was 59%, and for Yen *et al.* [17] the proportion was just below 90% - all three DDH studies report significant dependency parameter estimates.

5 The Bivariate Poisson DDH Model

In the previous 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. Accordingly, in this example the impact of dependency on all DDH parameters is examined by inspecting elements of Fisher's information matrix.

5.1 Distribution and Density

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

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

for y_1^{**} and y_2^{**} each taking non-negative integer values, 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. The marginal pdf are $Y_1^{**} \sim \text{Poisson}(\mu_1)$ and $Y_2^{**} \sim \text{Poisson}(\mu_2)$. The dependency parameter θ is equivalent to the covariance between Y_1^{**} and Y_2^{**}, and it satisfies $0 \le \theta < \min(\mu_1, \mu_2)$.

Here (Y_1^{**}, Y_2^{**}) is discrete, so modification to the statistical theory presented in Section 2 is required. 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:

$$\begin{aligned} f_+(y) &= \Pr(Y_2^{**} = y) - \Pr(Y_1^{**} = 0, Y_2^{**} = y) \\ &= \frac{e^{-\mu_2}\mu_2^y}{y!} - \frac{\delta(\mu_2 - \theta)^y}{y!} \end{aligned}$$

for positive integers y, and at the origin

$$\begin{aligned} f_0 &= \Pr(Y_1^{**} = 0) + \Pr(Y_2^{**} = 0) - \Pr(Y_1^{**} = 0, Y_2^{**} = 0) \\ &= \exp(-\mu_1) + \exp(-\mu_2) - \delta \end{aligned}$$

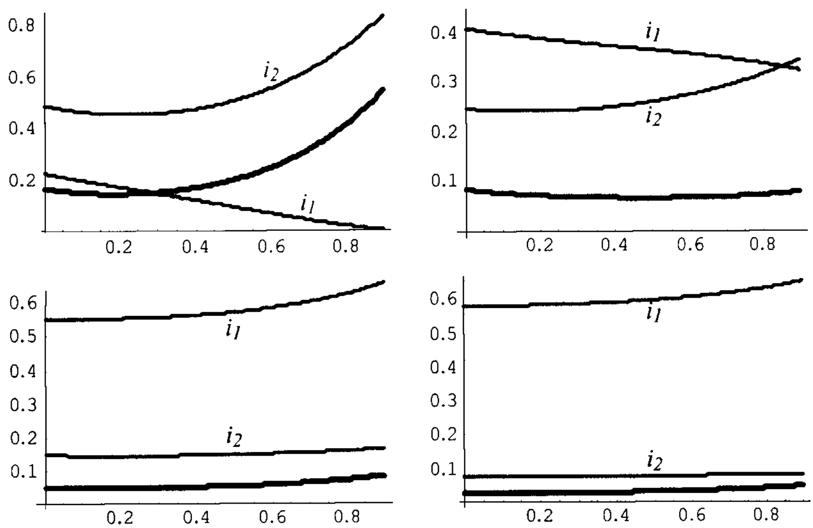
5.2 Analysis

Figure III gives four plots, against values of θ, of Fisher's information on parameters μ_1, μ_2, and θ, denoted respectively by i_1, i_2, and i_θ (the thicker line). These measures correspond to the elements on the leading diagonal of Fisher's information matrix:

$$i = E\left(\frac{\partial}{\partial\alpha}\log f(Y) \times \frac{\partial}{\partial\alpha'}\log f(Y)\right) = \begin{array}{cc} \begin{array}{ccc} \mu_1 & \mu_2 & \theta \end{array} & \\ \begin{pmatrix} i_1 & i_{12} & i_{13} \\ & i_2 & i_{23} \\ & & i_\theta \end{pmatrix} & \begin{array}{c} \mu_1 \\ \mu_2 \\ \theta \end{array} \end{array}$$

where the column vector $\alpha = (\mu_1, \mu_2, \theta)'$. In each plot, the true value of μ_1 is fixed at unity, whereas μ_2 is assigned values 1, 2, 4 and finally 8. Note the differing vertical scales.

Figure III: Fisher's information



For the north-west plot ($\mu_1 = \mu_2 = 1$), it is apparent that the statistical information associated with θ increases with the true value of θ. This is a positive finding, and one that accords with intuition. However, also evident in this plot is the trade-off between Fisher information on μ_1 and θ, the former virtually disappearing as θ increases. There appears here to be a "competition" amongst the parameters for statistical information.

In the north-east plot ($\mu_1 = 1$ and $\mu_2 = 2$), the trade-off in Fisher information is still in evidence, however, the magnitudes of i_1 and i_2 are such that neither vanishes as the true value of θ increases. Nevertheless, the situation in respect of Fisher Information on θ has worsened considerably: i_θ remains fairly constant and fairly small. There is little statistical information on θ present in the model, irrespective of its true value.

In the remaining plots, there is evidence of a vast disparity between the Fisher information on μ_1, and that of μ_2 and θ. These plots, especially the south-east plot for which $\mu_1 = 1$ and $\mu_2 = 8$ (implying that the decision process is *close* to first-hurdle dominant), clearly demonstrate that there is barely no Fisher information on θ, irrespective of the true value of θ. Given the scarcity of statistical information on θ, there seems little chance of data (except perhaps if it is collected in very 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.*

The trade-off in statistical information evidenced in this example, suggests that DDH models are over-parameterised. Incorporating a dependency parameter into a DH model (yielding a DDH model), while perhaps justified on behavioural grounds, manages only to expose a statistical weakness in the DDH model.

There is an alternative viewpoint here, one which may be seen in the approach of Gao *et al.* [7]. In that study, estimation of a bivariate normal DDH model returned an insignificant estimate of θ. The authors then proceeded to specify a second DDH model, based on an inverse hyperbolic sine transformation of Y (see [19] for a further empirical example of this model), which ultimately returned a significant estimate of θ. Electing, in light of the results of this paper, to ascribe insignificance in the first DDH model to the difficulties caused by weak identification, then this can be "overcome" by inducing sufficient non-linearity into the likelihood function. Unfortunately, non-linear transformations (such as the one used by Gao *et al.*) may manage to hide weak identification in a shower of parameters, but it typically comes at a cost of violating the principle of parsimony. Indeed, θ is only identified through the parametric form assumed for F, for otherwise, the (nonparametric) joint distribution of (Y_1^{**}, Y_2^{**}) is not identified. To see this, take the distribution of Y_2^{**} as that of Y when $Y = y > 0$, and anything for negative values that leads to $\int_{-\infty}^{0} dF_2(y) = \Pr(Y = 0)$, combined with $Y_1^{**} = Y_2^{**}$. Such a model, with perfect correlation between Y_1^{**} and Y_2^{**}, is always valid.

6 Remarks

6.1 Summary

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.

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 statistical information in the data can then reveal as much about these parameters as is possible.

6.2 Other Modelling Strategies

The 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 a DH model is to be fitted. By construction, the underlying decision 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:

$$f(y) = \begin{cases} (1 - F_1(0))\frac{\partial F_2(y)}{\partial y} & \text{if } y > 0 \\ F_1(0) + F_2(0) - F_1(0)F_2(0) & \text{if } y = 0 \end{cases}$$

Now only univariate distributions F_1 and F_2 are required. The normal-normal combination (a normal distribution for F_1 and a normal distribution for F_2) is standard; however, it is conceivable that other distributional pairings may be suitable. This approach to DH modelling is well-known.

Now the previous construction focused on the relationship between Y and (Y_1^{**}, Y_2^{**}). Evidently there is a second, less-explored possibility - one that emphasizes specification of distributions for the hurdle variables Y_1^* and Y_2^* (*cf.* Dionne *et al.* [6]). Of course, specifying a distribution for Y_1^* is easy, it must be Bernoulli distributed:

$$\Pr(Y_1^* = y_1^*) = (1 - r)^{1-y_1^*} r^{y_1^*}$$

where y_1^* takes values 0 and 1, and real-valued 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.

To illustrate the second approach, suppose that the pdf of Y_2^* is given by:

$$g(y_2^*) = \begin{cases} g_+(y_2^*) & \text{if } y_2^* > 0 \\ g_0 & \text{if } y_2^* = 0 \end{cases}$$

for suitable functions g_+ and g_0, both of which may depend on parameters and covariates. Substituting into the second line of (4), exploiting independence, and then applying (5) and (6), yields the pdf of Y as:

$$f(y) = \begin{cases} rg_+(y) & \text{if } y > 0 \\ 1 - r + rg_0 & \text{if } y = 0 \end{cases}$$

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

7 Technical Appendix

Derivations of results given in Section 4 for the bivariate logistic DDH model.

7.1 Fisher's Information: Proof of (9)

The definition of Fisher's information on θ is:

$$\begin{aligned} E\left(\frac{\partial}{\partial\theta}\log f(Y)\right)^2 &= \Pr(Y=0)\; E\left(\left(\frac{\partial}{\partial\theta}\log f(Y)\right)^2 \middle| Y=0\right) \\ &\quad + \Pr(Y>0)\; E\left(\left(\frac{\partial}{\partial\theta}\log f(Y)\right)^2 \middle| Y>0\right) \\ &= f_0 \times \left(\frac{\partial}{\partial\theta}\log f_0\right)^2 + \int_0^\infty \left(\frac{\partial}{\partial\theta}\log f_+(y)\right)^2 f_+(y)dy \\ &= \frac{1}{f_0}\left(\frac{\partial f_0}{\partial\theta}\right)^2 + \int_0^\infty \frac{1}{f_+(y)}\left(\frac{\partial f_+(y)}{\partial\theta}\right)^2 dy \qquad (13) \end{aligned}$$

It is because the pdf of Y is a continuous-discrete mixture (see (8)) that there are two components in (13): the first term $f_0 = \Pr(Y=0)$ arises because the density at the origin represents a probability mass, the second $f_+(y)$ because the density at a given positive value represents a sliver of mass.

Focusing on the first term on the right-hand-side of (13), it is a simple matter to show that the contribution to Fisher's information on θ is:

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

The other contribution is given by the second term on the right-hand-side of (13). 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 for all allowable values of μ_1, μ_2, θ and $y > 0$, $|\theta F_1(0)(1 - 2F_2(y))| < 1$. Therefore, it is permissible to integrate term-by-term. The contribution is solved as:

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

where $\Psi(\cdot)$ denotes Lerch's function. Adding together the contributions gives (9).

As the text suggests, it is possible to simplify (9) even further. This is achieved by applying the following property of the Lerch function:

$$\Psi(x; 1; z) = x^{-z} B(z; 0; x) \tag{14}$$

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 to yield Fisher's information on θ as:

$$\begin{aligned} i \; = \; & \theta^{-3} e^{\mu_1} \Big[- \theta F_1(0)(1 - F_2(0))(1 - \theta F_1(0)F_2(0)) + \\ & + \frac{1}{2} \log \frac{1 + \theta F_1(0)}{1 - \theta F_1(0)(1 - 2F_2(0))} \Big] + \frac{F_1'(0)^2 F_2'(0)^2}{F_1(0) + F_2(0) - F(0,0)} \end{aligned} \tag{15}$$

which is valid for all permissible $\theta \neq 0$. When $\theta = 0$, $\Psi(0;1;3) = \frac{1}{3}$ and

$$i = \frac{1}{6} F_1(0) F_1'(0) \left(1 + (1 - 2F_2(0))^3\right) + \frac{F_1'(0)^2 F_2'(0)^2}{F_1(0) + F_2(0) - F_1(0)F_2(0)} \tag{16}$$

Although it is a complicated function of θ, μ_1 and μ_2, Fisher's information on θ, represented by (15) and (16), is in closed form.

7.2 First-Hurdle Dominance: Proof of (11)

Substitute (14) into (10) to yield:

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

The derivative of this expression with respect to μ_1 can be written

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

For all θ and μ_1, the term within curly braces takes values between 0 and 2, and the term within square braces is larger than i_{D} (to see this rewrite the square-bracketed term as $\frac{1}{2} F_1(0) F_1'(0) \; \Psi(\theta^2 F_1(0)^2; 0; \frac{3}{2})$, and note $\Psi(x;0;z) > \Psi(x;1;z)$ for any x such that $-1 < x < 1$). Hence, for some μ_1 and θ, the derivative may be negative. In particular, this is true at $\mu_1 = 0$ for all θ, because the term within curly braces simplifies to unity.

References

[1] Aalen OO (1992), Modelling heterogeneity in survival analysis by the compound Poisson distribution. *The Annals of Applied Probability* **2**, 951-972

[2] Blaylock JR, Blisard WN (1992), U.S. cigarette consumption: the case of low-income women. *American Journal of Agricultural Economics* **74**, 698-705

[3] Blundell R, Meghir C (1987), Bivariate alternatives to the Tobit model. *Journal of Econometrics* **34**, 179-200

[4] Burton M, Tomlinson M, 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 JG (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, Guillen M (1996), Count data models for a credit scoring system. *Journal of Empirical Finance* **3**, 303-325

[7] Gao XM, Wailes EJ, Cramer GL (1995), Double-hurdle model with bivariate normal errors: an application to U.S. rice demand. *Journal of Agricultural and Applied Economics* **27**, 363-376

[8] Garcia J, Labeaga JM (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 BW (1992), At-home consumption of cheese: a purchase-infrequency model. *American Journal of Agricultural Economics* **72**, 453-459

[10] Jones AM (1989), A double-hurdle model of cigarette consumption. *Journal of Applied Econometrics* **4**, 23-39

[11] Jones AM (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

[12] Jones AM (2000), A Box-Cox double-hurdle model. *The Manchester School of Economic and Social Studies* **68**, 203-221

[13] Lambert D (1992), Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics* **34**, 1-14

[14] Pudney S (1989), *Modelling Individual Choice: the Econometrics of Corners, Kinks, and Holes*, London: Basil Blackwell

[15] Shonkwiler JS, Shaw WD (1996), Hurdle count-data models in recreation demand analysis. *Journal of Agricultural and Resource Economics* **21**, 210-219

[16] Spanier J, Oldham KB (1987), *An Atlas of Functions*, Washington: Hemisphere

[17] Yen ST, Boxall PC, Adamowicz WL (1997), An econometric analysis of donations for environmental conservation in Canada. *Journal of Agricultural and Resource Economics* **22**, 246-263

[18] Yen ST, Jensen HH, Wang Q (1996), Cholesterol information and egg consumption in the US: a nonnormal and heteroscedastic double-hurdle model. *European Review of Agricultural Economics* **23**, 343-356

[19] Yen ST, Jones AM (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 CJW (1998), An analytic and experimental examination of zero-inflated and hurdle Poisson specifications. *Sociological Methods and Research* **26**, 368-400