# HURDLE MODELS OF ALCOHOL AND TOBACCO EXPENDITURE IN SOUTH AFRICAN HOUSEHOLDS[1]

MARC GROUND and STEVEN F. KOCH

*Abstract*
Estimates of participation or expenditure elasticities depend upon the assumptions made regarding the observation of zero expenditure at the household level. This research examines two single-hurdle models across two commodities for which nearly two-thirds of the observations are zero. The research shows that one hurdle model consistently outperforms the other, and does so for intuitively appealing reasons.

*Keywords: Hurdle model, Hausman-type test*

## 1. INTRODUCTION

One underlying feature of microeconomic data sets is the plethora of zero responses. Observed zeroes in expenditure survey data occur for four primary reasons: *(i)* abstention, *(ii)* budgetary constraints, *(Hi)* infrequent purchase and *(iv)* under-reporting.[2] Abstention, for example, will occur when the marginal utility per rand of a commodity is less than its price, while budget constraints may stop an individual from purchasing a commodity that does provide felicity at the margin. In other words, expenditure data could be generated by two hurdle processes. One hurdle determines whether or not a household will abstain, while the other hurdle determines the quantity a participating household will consume, which could be zero if budgetary constraints result in non-purchase.

Double hurdle models have been applied in many areas of the literature: Moffatt (2005) considers credit scoring; while Blundell and Meghir (1986) and Blundell *et al.* (1987) consider labour supply decisions for males and females, respectively. Hurdle models applied to tobacco and alcohol consumption have an especially long history. Early applications include Deaton and Irish (1984), who considered alcohol, and Mullahy (1985), who considered tobacco. Further research examining alcohol and tobacco demand or expenditure through hurdle models includes, but is not limited to: Jones (1989, 1992), Fry and Pashardes (1994), Garcia and Labeaga (1996), Labeaga (1999), Shonkwiler and Yen (1999), Su and Yen (2000), Yen (2005 a,b).

---

[1] The research contained in this document was funded in part by NRF Grant 2053446; however, the views expressed in this research document are not necessarily the views of the National Research Foundation. The authors would like to thank seminar participants at the University of Pretoria and the University of Kwazulu Natal for their helpful comments. All other errors, however, remain the responsibility of the authors.

[2] The research presented here does not consider the effect of infrequent purchase (where items are purchased only occasionally, *e.g.* cars) or under-reporting (where small values might be reported as zeroes rather than as small values, *e.g.* a single piece of candy).

In applying double hurdle models, two broad concerns arise. The first concern is over the appropriateness of the assumption regarding the independence of the two hurdle processes. Although the independence assumption is, in principle, testable, the information requirements for suitable identification are non-negligible, which is reflected in the cigarette demand research literature.[3] Due to the fact that dependent double hurdle models are simultaneous equation models, identification may require a suitable exclusion restriction, although non-linearity can, in theory, be used for identification. Smith (2003), however, has shown that the identification of dependence is likely to be very weak without exclusion restrictions. Smith shows that a first-hurdle dominant population needs at least 50% zeroes to be well-suited to identifying the dependence based solely on model non-linearities. Smith (2003:591) further argues that it would be safer to ignore dependency to allow "the statistical information to reveal as much about the parameters as possible". For Smith's reasons, as well as the fact that the cross-sectional survey data used in this analysis does not offer any useful exclusion restrictions, dependent double hurdle models are not estimated for this paper.

The second concern arises over the appropriateness of the second hurdle assumption. Essentially, the researcher is assuming that households in the survey who want to purchase a product fail to do so during the survey period.[4] It is possible, however, to assume that those who want to participate do so, such that the second hurdle is cleared for all participants. This version of the model, which is characterized by first hurdle dominance, is often referred to as a single hurdle model and it is relevant in the case where the good is perishable. Due to the properties of the survey data used here, first hurdle dominance is likely to be relevant for household purchases of alcohol and tobacco in the past week.[5] It is, however, possible that households purchase other alcoholic items, especially wine, for later consumption. Therefore, in what follows, we assume that household inventories are in steady-state, such that purchases are made for stock replacement purposes.

Although not generally referred to as a single hurdle model, the censoring model proposed by Tobin (1958) is a special case; essentially, the dominant hurdle and behaviour, once the hurdle is cleared, follow a single data generating process (DGP). The first treatment allowing for potentially different DGPs was Cragg (1971), who extended Tobin's model in a number of ways. In this paper, we consider Cragg's independent single hurdle model, in which one DGP determines the dominant hurdle, while another DGP determines behaviour, once the hurdle has been cleared. Furthermore, Cragg's first hurdle dominance model closely resembles Tobin's model, such that the models can be easily compared.

---

[3] Research by Blaylock and Blisard (1992), Jones (1992) and Garcia and Labeaga (1996) could not identify separate hurdles, while later research by Yen (2005b) was able to identify separate hurdles.

[4] A durable good, from which flows are consumed, would be an example of such a product.

[5] Research by Miles (2000), using the Spanish Expenditure Survey, shows that 98-99% of all smokers will purchase cigarettes in a given week.

The research presented in this paper examines household tobacco and alcohol expenditure shares in the context of single hurdle models, as developed by Cragg and Tobin. These models can be nested and compared *via* a likelihood ratio test; however, the models are, instead, compared *via* a test devised by Lin and Schmidt (1984). The results show that Cragg's more general hurdle model is statistically preferred, which implies that there are differences between participation elasticities and expenditure share elasticities. These differences are also explored in the paper.

The remainder of the paper is organized as follows. The theoretical underpinnings and the empirical specification are discussed in section 2. The investigated data is discussed in section 3, and estimated results are available in section 4. Finally, concluding comments are provided in section 5.

## 2. THE MODEL

### *2.1 Introduction*

The unit of analysis presented in this research is the household. Each household may purchase any item from the broad spectrum of items available in their area. Actual household purchases will depend upon household circumstances, including household expenditure, composition and location. Furthermore, due to budget constraints, any decision to consume one product results in substitution; expenditure on at least one item must entail less expenditure on some other item. Therefore, expenditure estimates are often undertaken in system form, using household expenditure shares based upon Deaton and Muellbauer's (1980) Almost Ideal Demand System or, more recently, the Quadratic Almost Ideal Demand System developed by Banks *et al.* (1997). However, the focus of the analysis here is on alcohol and tobacco consumption, a few of the potential purchases made at the household level. Given the limited focus of the analysis, estimation of the entire system is not necessary and would, otherwise, obscure the desired estimates.[6]

Although the analysis here does not focus on systemic estimates, the Working (1943) and Leser (1963) share equations are a useful basis for proceeding.[7] Engel curves, based on household expenditure share behaviour, can be expressed by

$$w_i = \alpha + \beta \ln x_i + Z_i \Gamma + \nu_i \qquad (1)$$

In the preceding equation, *w* represents the share of expenditure by household *i* on any alcoholic beverage or tobacco product, *x* represents total household expenditure, $Z$ represents other household characteristics, and V represents white noise error.[8]

Econometrically although the assumed DGP is linear, the dependent variable may be censored. Even if the data is not actually censored *(i.e.* economics does not provide any intuition regarding the possibility of negative expenditure) a large number of households choose not to purchase any alcohol

---

[6] It is possible to allow for a composite commodity, which represents all other goods, but the mere size of the other goods component, representing well over 99% of household expenditures in most cases, does not lend itself to precise system estimation; see Koch (forthcoming) for further discussion on some of the problems within the system.

[7] Importantly, Deaton and Muellbauer (1980) developed the Almost Ideal Demand System, which yields the Working-Leser share equations under rather general assumptions.

[8] Further discussion of the household characteristics used in the analysis is presented in section 4.

or tobacco item. In the case of alcoholic beverages, for example, only 29-2% of households purchase positive quantities, while 35-2% purchase tobacco products.

*2.2 The Empirical Analysis*

Tobin developed a model to examine censored data. The model is a combination of probit, used to demarcate participation and non-participation, and OLS, used to examine the behavioural attributes of the participants. The key feature of Tobin's model, often referred to as the tobit, is the fact that the population parameters underlying participation and behaviour are the same, a result of the censoring assumption. On the other hand, if the data is not truly censored and is, instead, correctly observed, then the population parameters for participation might differ from the population behavioural parameters.[9] Formally, the analysis below will examine the prospect that population participation determinants differ from population behavioural determinants by comparing the tobit model to a less restrictive model developed by Cragg. Although Greene (2000) discusses a simple likelihood ratio test to compare these models, a Hausman-type test developed by Lin and Schmidt is actually used.[10]

*2.2.1 The Tobit* The standard formulation of Tobin's model considers a household for which we observe the censored variable *w* defined below.

$$w_i = \begin{cases} 0, & \text{if } w_i^* \leq 0 \\ w_i^*, & \text{if } w_i^* > 0 \end{cases} \qquad (2)$$

The latent function, $w_i^* = X_i\beta + \varepsilon_i$, where $\varepsilon_i : N(0,\sigma^{11})$, characterizes household participation, defined as the observation of positive purchases, and behaviour, defined as the household expenditure share mean given participation. The log-likelihood function for the tobit model is provided below; in terms of notation, $\Phi$ represents the normal CDF.

$$\ln L_T = \sum_{w_i = w_i^*} -\frac{1}{2}\left[\ln(2\pi) + \ln\sigma^2 + \frac{(w_i - X_i\beta)^2}{\sigma^2}\right] + \sum_{w_i = 0} \ln\left[1 - \Phi\left(\frac{X_i\beta}{\sigma}\right)\right] \qquad (3)$$

The tobit log-likelihood function verifies that the coefficients in the participation equation are assumed to be the same as the coefficients in the behavioural equation, which may not be an appropriate assumption.

[9] It might also be true that households select into certain types of purchases, such as those related to smoking or drinking; however, selection models (or dependent double hurdle models) would be identified only on the non-linearity of the model, and, therefore, selection issues are not considered in this analysis.

[10] Households in the same primary sampling units are likely to be similar and these correlations are not easily included in the likelihood function. A reviewer suggested that dummies be used to control for the cluster effects; unfortunately, there are approximately 10 households in each cluster, such that many clusters contain only smokers or non-smokers, and thus the probit models cannot be estimated for a large percentage of the clusters. Although dummies for each cluster are not used, the standard errors are corrected for data clustering.

*2.2.2 An Extension to the Tobit* Consider a generalization of Tobin's model. In this model, first proposed by Cragg, the behavioural equation and the participation equation are allowed to have different coefficients. Furthermore, the behavioural equation, rather than being estimated *via* OLS, is estimated as a truncated regression, due to the fact that behaviour is only observed for participants.[11]

Formally, define a participation dummy, as below.

$$d_i = \begin{cases} 0, & \text{if } w_i^* \leq 0 \\ 1, & \text{if } w_i^* > 0 \end{cases} \tag{4}$$

A probit model, where $prob[w_i^* > 0] = \Phi(X_i\gamma)$ and $prob[w_i^* \leq 0] = 1 - \Phi(X_i\gamma)$, can be created using the previously defined dummy variable. The expected value of the expenditure share, given participation, follows; importantly, it includes a sample truncation correction referred to as the inverse Mills ratio.

$$E[w_i | d_i = 1] = X_i\beta + \sigma\lambda_i \tag{5}$$

The truncation correction accounts for the fact that only a portion of the distribution is observed, and, therefore, the mean is only calculated based upon what is observed.[12] The preceding expected value is much like the expected value from a model that incorporates sample selection (Heckman, 1976); however, the sample selection model assumes that the probit and the truncated regression are correlated, which requires exclusion restrictions for identification. In this version of Cragg's model, see below, the probit and truncated regressions are assumed to be uncorrelated.

The log-likelihood function for the version of Cragg's model subsumed by equation (5) and the discussion surrounding equation (4) is given in equation (6).

$$\begin{aligned} \ln L_C = & \sum_{d_i=0} \ln[1 - \Phi(X_i\gamma)] + \sum_{d_i=1} \ln[\Phi(X_i\gamma)] \\ & + \sum_{d_i=1} -\frac{1}{2}\left[\ln(2\pi) + \ln\sigma^2 + \frac{(w_i - X_i\beta)^2}{\sigma^2}\right] - \sum_{d_i=1} \ln\left[1 - \Phi\left(\frac{-X_i\beta}{\sigma}\right)\right] \end{aligned} \tag{6}$$

The first portion (top line) is the log-likelihood for a probit, while the second portion (bottom line) is the log-likelihood for a truncated regression, with truncation at zero. Therefore, the log-likelihood from the Cragg model is the sum of the log-likelihood from a probit and a truncated regression, $L_C = L_P + L_{TR}$. More useful, however, is the fact that these two component pieces are entirely separable, such that the probit and truncated regression can be estimated separately[13]

---

[11] See footnote 8.

[12] For a more detailed discussion of the moments of a distribution in the face of truncation, refer to Greene (2000).

[13] The probit parameters are not included anywhere in the truncated regression, while the truncated regression parameters are not included anywhere in the probit regression.

*2.2.3 Empirical Tests* A careful comparison of equations (3) and (6) provides a very simple test of the Tobit model relative to the Cragg model.[14] Since the version of Cragg's model discussed here allows for population participation parameters to differ from the population behavioural parameters, that hypothesis can be tested. It is appropriate to test whether or not the restriction of equal parameterization is supported by the data, *i.e.* $\beta/\sigma = \gamma$ . As with all empirical models, there are many ways to test the restrictions. A LaGrange multiplier test, such as was devised by Lin and Schmidt (1984) is one useful avenue; if the restriction is true, there is no need to estimate the additional models. The test statistic, though, does require the creation of additional data, and, therefore is not necessarily easy to implement. On the other hand, a likelihood ratio test, as discussed by Greene (2000) and is reviewed in Appendix B, is not difficult to implement, given current computing power. Unfortunately, the test requires exact likelihood values, which are not available when data is clustered.[15] Similar to the LM test, it is possible to develop a conditional moment test, such as was envisioned by Pagan and Vella (1989) and discussed in Cameron and Trivedi (2005). However, such a test, when including multiple moments, can be interpreted as a test of functional form. Therefore, we employ a Hausman-type test, which can also be interpreted as a test of functional form. The test is easy to employ and is based on the assumption that one set of parameters is consistent and efficient under the null, while the other set of parameters are consistent, but inefficient under the null. Defining *V* as the variance of the estimators, the test statistic is the following.

$$(\gamma - \beta/\sigma)'[V(\gamma) - V(\beta/\sigma)]^{-1}(\gamma - \beta/\sigma) : \chi^2_k \tag{7}$$

## 3. THE DATA

The data used for the analysis was taken from the 2000 Income and Expenditure Survey (IES) of South Africa, data that is collected by Statistics South Africa primarily for the purpose of indexing prices *via* the CPI and CPI-X to measure inflation in South Africa. In 2000, data from 104,153 people in 26,264 households was collected *via* the household head or another adult within the household. Other than for CPI and CPI-X, the data has been widely used to examine poverty and inequality in South Africa. For example, Hoogeveen and Ozler (2006) and Leibbrandt *et al.* (2005) use the 2000 IES and other data sets to examine income changes between 1995 and 2000. Summary statistics of the data used in the analysis are presented in Table l.[16]

In addition to poverty and inequality analysis based on income and expenditure, the focus of the IES allows for the examination of categorical expenditure across households, and has been used to compare household alcohol and tobacco expenditure patterns from 1995 to 2000.[17] The analysis presented here, which seeks to analyse the empirical effects of observed

---

[14] The derivation, based on a statement in Greene (2000), is available from the authors upon request.

[15] True likelihood values can only be calculated from the correct likelihood function. When data is clustered, the data is assumed to be block-diagonally correlated. Inclusion of that assumption in the likelihood function is not obvious or easily estimated.

[16] In an effort to save space, no formal discussion of these statistics is included in the text.

[17] See, for example, Ground *et al.* (forthcoming) and Koch (2006).

*Table 1. Descriptive household statistics across alcohol and tobacco expenditure categories: categories are zero expenditure or positive expenditure*

| | Categorical household Tobacco expenditure | | | | Categorical household Alcohol expenditure | | | |
|---|---|---|---|---|---|---|---|---|
| | Exp = 0 (n = 15,941) | | Exp > 0 (n = 8,693) | | Exp = 0 (n = 17,356) | | Exp > 0 (n = 7,278) | |
| **Independent variable** | **Mean** | **Std. error#** | **Mean** | **Std. error#** | **Mean** | **Std. error#** | **Mean** | **Std. error#** |
| African | 0.842*** | 0.006 | 0.705 | 0.010 | 0.827*** | 0.006 | 0.714 | 0.009 |
| White | 0.075 | 0.004 | 0.089*** | 0.005 | 0.052 | 0.003 | 0.149*** | 0.007 |
| Coloured | 0.062 | 0.004 | 0.181*** | 0.009 | 0.096 | 0.005 | 0.122*** | 0.007 |
| Asian | 0.020 | 0.002 | 0.022 | 0.003 | 0.024*** | 0.003 | 0.013 | 0.002 |
| Male | 0.521 | 0.005 | 0.759*** | 0.005 | 0.519 | 0.005 | 0.811*** | 0.005 |
| Log total expenditure | 9.732 | 0.014 | 9.851*** | 0.017 | 9.634 | 0.013 | 10.113*** | 0.020 |
| Head employed | 0.584 | 0.006 | 0.651*** | 0.006 | 0.561 | 0.005 | 0.721*** | 0.006 |
| Welfare recipient | 0.190 | 0.004 | 0.184 | 0.005 | 0.206*** | 0.004 | 0.143 | 0.005 |
| Age of household head | 46.3 | 0.174 | 46.8*** | 0.198 | 46.8*** | 0.167 | 45.6 | 0.208 |
| Household members | 3.944 | 0.033 | 4.006** | 0.036 | 4.081*** | 0.030 | 3.687 | 0.039 |
| Household proportion of females | 0.530*** | 0.003 | 0.426 | 0.004 | 0.530*** | 0.003 | 0.405 | 0.004 |
| HH average adult equivalency | 0.881 | 0.001 | 0.898*** | 0.001 | 0.878 | 0.001 | 0.908*** | 0.002 |
| Children aged 4 and under | 0.491*** | 0.006 | 0.441 | 0.008 | 0.509*** | 0.006 | 0.388 | 0.008 |
| Children aged 5 to 16 | 1.087*** | 0.010 | 0.931 | 0.013 | 1.123*** | 0.010 | 0.811 | 0.014 |
| Urban residence | 0.598 | 0.010 | 0.633*** | 0.010 | 0.590 | 0.010 | 0.659*** | 0.010 |
| Rent dwellling | 0.272 | 0.007 | 0.338*** | 0.008 | 0.271 | 0.006 | 0.354*** | 0.009 |
| Western cape | 0.068 | 0.004 | 0.154*** | 0.009 | 0.085 | 0.005 | 0.130*** | 0.009 |
| Eastern cape | 0.143*** | 0.007 | 0.115 | 0.007 | 0.148*** | 0.007 | 0.098 | 0.009 |
| Northern cape | 0.036 | 0.003 | 0.074*** | 0.006 | 0.045 | 0.004 | 0.060** | 0.007 |
| Free state | 0.075 | 0.005 | 0.112*** | 0.007 | 0.081 | 0.005 | 0.105*** | 0.004 |
| Kwazulu Natal | 0.195*** | 0.009 | 0.117 | 0.007 | 0.184*** | 0.008 | 0.129 | 0.006 |
| Northwest | 0.101 | 0.006 | 0.119* | 0.007 | 0.103 | 0.006 | 0.118 | 0.010 |
| Gauteng | 0.156 | 0.007 | 0.148 | 0.008 | 0.146 | 0.007 | 0.170** | 0.010 |
| Mpumalanga | 0.088 | 0.006 | 0.077 | 0.005 | 0.081 | 0.005 | 0.092 | 0.007 |

*Source:* Author's Calculations from 2000 SAIES. Calculations from STATA SE 9.2.
# Standard errors corrected for 2955 clusters in the data.
*** Significantly larger at 1% (t-test conducted across expenditure category). ** Significantly larger at 5% (t-test conducted across expenditure category). * Significantly larger at 10% (t-test conducted across expenditure category).

zeroes, also focuses on household tobacco and alcohol purchases, primarily due to the large number of reported zeroes in the data.

One major concern in using the IES data is whether or not the data can be trusted. As pointed out by Hoogeveen and Ozler, the 2001 census, from which the 2000 IES sample is drawn, over-represents Africans and under-represents whites, while Simkins (2004) argues that property income in the 2000 IES is poorly measured, thus yielding understated measures of household income. On the other hand, Leibbrandt *et al.* (2005), show that the 1995 IES and 2000 IES are fairly consistent with each other, such that although there are likely to be some problems in the data, those problems are not extensive enough to make the 1995 and 2000 data sets incomparable, although their argument can not be used to suggest that either data set on its own is correct, as it is possible that each of the surveys suffers from the same set of over or under-representation of households.

For this research, there is a further data reporting concern. Due to the fact that only the household head or some other adult in the household completes the survey, it is likely that youth alcohol and tobacco expenditure is underreported, if it is reported at all. In an effort to consider the depth of underreporting, additional analysis using single-person households was also undertaken; due to the qualitative similarity of the results for single-person and multiple-person households, those results are not reported here.

## 4. THE RESULTS

### *4.1 Tobit Models*

Selected results from Tobit models, using tobacco and alcohol expenditure shares are presented in the third and sixth columns of numbers in Table 2. The tobacco coefficients are located in column 3, while the alcohol estimates are in column 6. The results show that as total household expenditure increases, the average tobacco expenditure share falls, while the average alcohol expenditure rises. For both goods, the expenditure share increases for non-African households, whose head is male, employed and older. Furthermore, older and larger households who rent their dwellings devote a larger proportion of their expenditure to tobacco and alcohol. On the other hand, the share of expenditure devoted to either alcohol or tobacco is smaller for households composed of relatively young children and women and are located in urban areas.

### *4.2 Probit Participation Models*

In this subsection, results from the probit models of positive expenditure, where the binary outcome is either positive expenditure or zero expenditure, are discussed. The results, in the first column of numbers in Table 2, focus only on aggregate tobacco expenditure shares, while the fourth column of numbers in Table 2 provide information on alcohol expenditure shares.[18] The results show that household composition is strongly related to the probability of purchasing either tobacco products or alcoholic beverages. Notably, larger non-African households headed by older men are more likely to purchase tobacco products, while urban households with children and a higher proportion of females are less likely to purchase tobacco products. In addition to composition, economic factors also affect participation. The results show that households whose head is employed, which receive welfare payments or rent their dwelling are more likely to participate in smoking-related purchases. On the other hand, higher household level expenditures are associated with less participation; since the results are presented as marginal effects, a 1% increase in total expenditure results in a 2.6% decrease in the probability of participation. Household participation in the consumption of alcohol is similarly related to household composition, although Asian households are less likely to purchase alcohol than any other racially composed household, while higher levels of total expenditure increase the probability of participation; for every 1% increase in total expenditure, the probability of purchasing an alcoholic beverage increases by about 5%.

### *4.3 Truncated Regression Model Results*

The results from the truncated regression model are reported in columns 2 and 5 of Table 2.[19] One interesting feature of the results is the fact that few of the examined determinants significantly influence the household share of expenditure, given that the household has chosen to participate. The smaller number of significant determinants does imply that the tobit assumptions are inappropriate (see below), despite the fact that the qualitative implications of most of the coefficients are not generally different than those previously discussed. Four results, however, stand out as being different than the rest.

---

[18] Additional disaggregated results are available from the authors, upon request.

[19] Analyses were also considered for single-person households, due to the fact that single-person households are devoid of any problems associated with intrahousehold bargaining, which may affect the unitary nature of the household. However, the results were not qualitatively different, and, therefore, they are not included here. Interested readers can contact the authors for the results.

*Table 2. Results for alcohol and tobacco shares for all South African households*

| | Tobacco expenditure shares | | | Alcohol expenditure shares | | |
|---|---|---|---|---|---|---|
| **Independent variable** | **Probit marginal effects** | **Truncated regression estimates** | **Tobit regression estimates** | **Probit marginal effects** | **Truncated regression estimates** | **Tobit regression estimates** |
| Log expenditure | −0.0264*** | −1.6925*** | −0.0894*** | 0.0518*** | −2.6830*** | 0.0353*** |
| | (0.004) | (0.419) | (0.007) | (0.004) | (0.612) | (0.009) |
| HH head employed | 0.0475*** | 0.3915** | 0.0759*** | 0.0693*** | 1.0596*** | 0.1877*** |
| | (0.008) | (0.170) | (0.012) | (0.007) | (0.381) | (0.019) |
| HH welfare receipt | 0.0414*** | −0.2632 | 0.0418*** | 0.0309*** | 1.2713** | 0.0871*** |
| | (0.011) | (0.272) | (0.015) | (0.011) | (0.516) | (0.024) |
| Male HH head | 0.1707*** | 0.2694 | 0.2409*** | 0.1721*** | −0.1953 | 0.4106*** |
| | (0.008) | (0.203) | (0.012) | (0.007) | (0.414) | (0.021) |
| Caucasian HH | 0.08267*** | 0.0916 | 0.1284*** | 0.1259*** | −16.8295*** | 0.0746*** |
| | (0.016) | (0.335) | (0.020) | (0.016) | (5.142) | (0.026) |
| Coloured HH | 0.2351*** | 0.7916*** | 0.2827*** | 0.0243* | −1.1163* | 0.0493* |
| | (0.017) | (0.298) | (0.020) | (0.014) | (0.639) | (0.030) |
| Asian HH | 0.1319*** | −0.0493 | 0.1725*** | −0.1385*** | −6.2275** | −0.3829*** |
| | (0.029) | (0.481) | (0.034) | (0.017) | (2.696) | (0.059) |
| HH head age | 0.0008*** | −0.0269*** | 0.0005 | 0.0004 | −0.01852 | 0.0012** |
| | (0.000) | (0.009) | (0.000) | (0.000) | (0.012) | (0.001) |
| HH members | 0.0388*** | 0.0514 | 0.0468*** | 0.0129*** | −0.3819** | 0.0183*** |
| | (0.003) | (0.063) | (0.004) | (0.003) | (0.163) | (0.006) |
| Proportion of females | −0.1575*** | −1.4311*** | −0.2801*** | −0.1221*** | −2.4286*** | −0.3568*** |
| | (0.013) | (0.454) | (0.021) | (0.013) | (0.731) | (0.031) |
| Avg adult equivalence | 0.0022 | 3.2806** | 0.1496*** | 0.1170** | 7.5012*** | 0.5001*** |
| | (0.049) | (1.539) | (0.069) | (0.048) | (2.862) | (0.111) |
| # Children under 5 | −0.0470*** | −0.0141 | −0.0501*** | −0.0171** | 1.3641*** | −0.0029 |
| | (0.008) | (0.194) | (0.011) | (0.008) | (0.509) | (0.018) |
| # Children 5 to 16 | −0.0526*** | −0.3141** | −0.0706*** | −0.0327*** | −0.2997 | −0.0729*** |
| | (0.005) | (0.138) | (0.007) | (0.005) | (0.282) | (0.011) |
| Rented dwelling | 0.0189** | 0.2616 | 0.0189 | 0.0238*** | 0.3937 | 0.0522*** |
| | (0.008) | (0.169) | (0.012) | (0.008) | (0.285) | (0.018) |
| Urban locale | −0.0288*** | 0.5336*** | −0.0112 | −0.0299*** | 0.1595 | −0.0480** |
| | (0.009) | (0.183) | (0.012) | (0.008) | (0.283) | (0.019) |
| Observations | 25,353 | 8,922 | 25,353 | 25,353 | 7,413 | 25,353 |
| Clusters | 2,955 | 2,689 | 2,955 | 2,955 | 2,592 | 2,955 |
| Pseudo LL | −14,868 | 925 | −15,272 | −13,643 | −702 | −16,471 |

Cluster level standard errors in parenthesis. Provincial dummies also used in the analysis, but not reported. The marginal effects for dummy variables measures discrete change from 0 to 1.

HH: household.

*** Significant at 1%. ** Significant at 5%. * Significant at 10%.

Caucasian households, households with larger total budget outlays and larger households, who purchase alcohol, expend a lesser share of their budgets on alcohol products, once accounting for participation. On the other hand, urban households, after controlling for participation, expend a larger proportion of their budgets on tobacco products. Each of these four results disagrees in sign with their values in both the probit and tobit estimates, suggesting the benefits to be derived from generalized hurdle models.

*4.4 Statistical Tests*

The results presented so far suggest that certain variables affect participation differently than behaviour, which is the reason this research considered the generalization of Tobin's model developed by Cragg. Intuitively, considering the case of alcohol and tobacco expenditure, the potential difference between one margin, participation, and another margin, behaviour given participation, should not be surprising. When it comes to the consumption of alcoholic beverages and tobacco products, it is very reasonable that once participation has been determined, expenditure could, for example, stay rather consistent, because individuals will smoke the same number of packs per day. If that were the case, a negative relationship between expenditure and the expenditure share would exist, while participation could be either negatively or positively related to expenditure.

A formal statistical test of the Cragg generalization, based on an LM test developed by Lin and Schmidt, is used to formally test the results; however, given the differences in sign discussed above, the test results have already been informally confirmed. As the analysis has only focused on aggregate alcohol and tobacco expenditure, the LM test is only calculated and discussed for aggregated tobacco and alcohol expenditure shares. In the case of tobacco, the calculated test statistic leads to rejection of the null hypothesis that the Tobin restriction is valid.[20] In the case of alcohol expenditure, the statistical result is the same; the Tobin restriction is rejected.[21] Therefore, the model developed by Cragg, which allows for different estimates of the participation determinants relative to the behavioural effects of various determinants, given participation, is statistically preferred to the tobit model, which does not allow for any differences. In other words, participation and behavioural elasticities, as already shown, will be different.

## 5. CONCLUSIONS, REMARKS AND EXTENSIONS

The research presented in this paper has examined the expenditure behaviours of South African households, using data from the 2000 South African Income and Expenditure Survey. The analysis considered probit, tobit and a more general version of the tobit, which was developed by Cragg. The tobit is a restricted version of Cragg's model, and Hausman-type tests reject the tobit restrictions in favour of the more general version. The generalization treats participation determinants differently than behavioural determinants, given that participation has occurred. Participation is examined within the context of a probit model, while behaviour, given participation, is examined with truncated regression models.

The results show that expenditure reduces the likelihood of participation in tobacco consumption, but raises the likelihood of positive alcohol consumption. Furthermore,

---

[20] For tobacco expenditure, the calculated value is 364.81, which is well beyond the critical value of 9.88 at the 0.005% level for 24 degrees of freedom (including the constant).

[21] For alcohol expenditure the calculated likelihood ratio is 442.50.

positive household purchases of alcohol and tobacco products are more likely in White or Coloured households with an older male household head, who is employed. Higher household average ages, as measured by the average adult equivalence in the household, and receipt of welfare payments, raise both likelihoods. On the other hand, positive purchases of either tobacco or alcohol, is less likely for urban households, which have a higher proportion of female members and live in dwellings that they own. Given participation, the actual shares of expenditure devoted to tobacco or alcohol consumption are lower for larger and richer households and lower for households with a larger proportion of female members. Importantly, the share of the budget devoted to alcoholic beverages is lower for wealthier households, an effect not picked up with a tobit model, which assumes that participation and behaviour are governed by the exact same process.

The analysis has shown that tobit models are not the preferred models for examining expenditure behaviour for households, when there are a large portion of zeroes among the dependent variable observations. Given the large number of people in South Africa who have limited spending power, and the number of analyses, in which a significant number of zeroes exist, this analysis suggests careful consideration of the treatment of participation relative to behaviour, once the participation hurdle has been crossed. For example, in examining returns to education, where nearly 40% of the population is unemployed, Cragg's model might be a useful model to consider. Although estimates of returns to education often attempt to control for sample selection, it is often the case that appropriate exclusion restrictions do not exist in the dataset. For that reason, the model analysed here, which does not depend upon exclusion restrictions, may provide very reasonable estimates of returns to education, once employment has been obtained.

Although not discussed in this research, Cragg's model is also a restricted version of Heckman's sample selection model, and, therefore, it is possible to consider whether or not the sample selection model provides better estimates than non-selected model estimates. Future research will consider this line of thinking, especially as it relates to the types of expenditure shares, where selection might be more easily envisioned. For example, only households with members suffering from an illness are likely to expend resources on health care, such that selection can be controlled.

## REFERENCES


BANKS, J., BLUNDELL, R. and LEWBEL, A. (1997). Quadratic engel curves and consumer demand. *Review of Economics and Statistics*, 79(4): 527-539.

BLAYLOCK, J. R. and BLISARD, W. N. (1992). U.S. cigarette consumption: The case of low-income women. *American Journal of Agricultural Economics*, 74(3): 698-705.

BLUNDELL, R. and MEGHIR, C. (1986). Selection criteria for a microeconometric model of labour supply. *Journal of Applied Econometrics*, 1(1): 55-80.

———, HAM, J. and MEGHIR, C. (1987). Unemployment and female labour supply. *Economic Journal*, 97: 44-64.

CAMERON, A. C. and TRIVEDI, P. K. (2005). *Microeconometrics: Methods and Applications*. Cambridge: Cambridge University Press.

CRAGG, J. G. (1971). Some statistical models for limited dependent variables with application to the demand for durable goods. *Econometrica*, 39(5): 829-844.

DEATON, A. S. and MUELLBAUER, J. (1980). An almost ideal demand system. *American Economic Review*, 70(3): 312-326.

——— and IRISH, M. (1984). Statistical models for zero expenditures in household budgets. *Journal of Public Economics*, 23(1-2): 59-80.

FRY, V. and PASHARDES, P. (1994). Abstention and aggregation in consumer demand: Zero tobacco expenditures. *Oxford Economic Papers*, 46(3): 502-518.

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(3): 489-506.
GREENE, W. H. (2000). *Econometric Analysis*, 4th ed. Upper Saddle River, NJ, USA: Prentice Hall.
GROUND, M., KOCH, S. F. and VAN WYK, D. (forthcoming). South African household expenditure patterns: Alcohol and tobacco in 1995 and 2000. *Acta Acadmica*.
HECKMAN, J. J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Annals of Economic and Social Measurement*, 5(4): 475-492.
HOOGEVEEN, Ö. (2006). Not separate, not equal: Poverty and inequality. In H. Bhorat and R. Kanbur (eds), *Poverty and Policy in Post-apartheid South Africa*. Pretoria, SA: HSRC Press.
JONES, A. M. (1989). A double-hurdle model of cigarette consumption. *Journal of Applied Econometrics*, 4(1): 23-39.
——— (1992). A note on computation of the double-hurdle model with dependence with an application to tobacco expenditure. *Bulletin of Economic Research*, 44(1): 67-74.
KOCH, S. F. (2006). Trends in South African household alcohol consumption risk factors. *University of Pretoria Working Paper*.
——— (forthcoming). South African household expenditure shares: Empirical issues surrounding the application of the ideal demand system to South African data. *Studies in Economics and Econometrics*.
LABEAGA, J. M. (1999). A double-hurdle rational addiction model with heterogeneity: Estimating the demand for tobacco. *Journal of Econometrics*, 93(1): 49-72.
LEIBBRANDT, M., LEVINSOHN, J. and MCCRARY, J. (2005). *Incomes in South Africa Since the Fall of Apartheid*. National Bureau of Economic Research, Inc, NBER Working Papers.
LESER, C. E. V. (1963). Forms of engel functions. *Econometrica*, 31: 694-703.
LIN, T. and SCHMIDT, P. (1984). A test of the Tobit specification against an alternative suggested by Cragg. *Review of Economics and Statistics*, 66(1): 174-177.
MILES, D. (2000). The probability that a smoker does not purchase tobacco: A note. *Oxford Bulletin of Economics and Statistics*, 62(5): 647-656.
MOFFATT, P. G. (2005). Hurdle models of loan default. *Journal of the Operations Research Society*, 56: 1063-1071.
MULLAHY, J. (1985). Cigarette smoking, habits, health concerns, and heterogeneous unobservables in a microeconometric analysis of consumer demand. PhD Dissertation, University of Virginia.
PAGAN, A. R. and VELLA, F. (1989). Diagnostic tests for models based on individual data: A survey. *Journal of Applied Econometrics*, 3(S): 29-60.
SHONKWILER, J. S. and YEN, S. T. (1999). Two-step estimation of a censored system of equations. *American Journal of Agricultural Economics*, 81(4): 972-982.
SIMKINS, C. E. W. (2004). What Happened to the Distribution of Income in South Africa Between 1995 and 2001? Unpublished Draft, accessed online, January 2007, http://www.sarpn.org.za/documents/d0001062/index.php.
SMITH, M. D. (2003). On dependency in double-hurdle models. *Statistical Papers*, 44(4): 581-595.
SU, S. B. and YEN, S. T. (2000). A censored system of cigarette and alcohol consumption. *Applied Economics*, 32(6): 729-737.
TOBIN, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica*, 26(1): 24-36.
WORKING, H. (1943). Statistical laws of family expenditure. *Journal of the American Statistical Association*, 38: 43-56.
YEN, S. T. (2005a). A multivariate sample-selection model: Estimating cigarette and alcohol demands with zero observations. *American Journal of Agricultural Economics*, 87(2): 453-466.
——— (2005b). Zero observations and gender differences in cigarette consumption. *Applied Economics*, 37(16): 1839-1849.