BANKWEST CURTIN ECONOMICS CENTRE

WORKING PAPER SERIES

13/3: MODELLING ILLEGAL DRUG PARTICIPATION IN AUSTRALIA

Sarah Brown, Mark N. Harris and Preety Srivastava

business.curtin.edu.au/bcec

Bankwest Curtin Economics Centre Working Paper Series
ISSN: 2202-2791
ISBN: 978-1-925083-02-6

Sarah Brown[1], Mark N. Harris[2] and Preety Srivastava[3]
[1]Department of Economics, University of Sheffield, United Kingdom
[2]Curtin Business School, Curtin University, Australia (Corresponding author)
[3]Department of Econometrics and Business Statistics, Monash University, Australia

Modelling Illegal Drug Participation in Australia*

Sarah Brown+, Mark N. Harris$ and Preety Srivastava^
+Department of Economics
University of Sheffield, United Kingdom
$Curtin Business School
Curtin University, Australia
(Corresponding author)
^Department of Econometrics and Business Statistics
Monash University, Australia

July 2013

Abstract

We contribute to the small, but important, literature exploring the incidence and implications of mis-reporting in survey data. Specifically, when modelling "social bads", such as illegal drug consumption, researchers are often faced with exceptionally low reported participation rates. We propose a modelling framework where firstly an individual decides whether to participate or not and, secondly for participants there is a subsequent decision to mis-report or not. We explore mis-reporting in the context of the consumption of a system of drugs and specify a *multivariate inflated probit model*. Compared to observed participation rates of 12, 3 and 1.3% (marijuana, speed and cocaine, respectively) true participation rates are estimated to be some 5 percentage points higher for marijuana, and nearly double for cocaine. There was an estimated 36% (18%) percent chance that a cocaine (marijuana) user would mis-report their participation. Less evidence of mis-reporting was found for speed users.

JEL Classification: C3, D1, I1

Keywords: Discrete data, illegal drug consumption, inflated responses, mis-reporting.

*The authors would like to thank the Bankwest-Curtin Economics Centre for their generous funding under the project *Socioeconomic Issues in Western Australia (Part 1): Drug Consumption.* All views expressed are those of the authors and do not necessarily represent those of Bankwest, or any of its associates. The authors would also like to thank both David Hille and Jake Prendergast, for excellent research assistance.

1 Introduction and Background

Over the past three decades, the increased availability of micro level data sets has enabled researchers to explore an extensive range of research themes at the individual and household level. Such micro level data is invariably collected using survey techniques with the result that the quality of the data gathered hinges critically on the respondents providing reliable and accurate information. It is apparent however, that the subject matter of some surveys may be such that respondents have an incentive to mis-report the true situation due to the sensitive nature of the questions. Individuals may have an incentive to under-report activities which are regarded as socially undesirable or which are associated with perceived social stigma or legal consequences, such as smoking, alcohol, illicit drug consumption and sexual behaviours (see, for example, Berg and Lien 2006, Pudney 2007).

In a similar vein, self-reported *versus* the true incidence of cheating behaviour has attracted some interest in the economics literature. For example, Caudill and Mixon (2005) study undergraduate cheating behaviour using a logit model applicable in the case of a mis-classified dependent variable, as developed by Hausman, Abrevaya, and Scott-Morton (1998), who show that ignoring such mis-classification can result in biased and inconsistent estimates. Their findings indicate that the incidence of cheating estimated at 70% is considerably higher than the self-reported incidence of 51%. Thus, it is apparent that mis-reporting is potentially prevalent across a wide range of economic (and other) areas.

Such mis-reporting will result in inaccurate estimates of the prevalence of such behaviours, which may lead one to question the validity of empirical conclusions drawn from such surveys. Moreover, any mis-reporting will likely lead to biased inferences in econometric analyses as well as to potentially inappropriate decision-making by policy-makers. Despite these extremely important implications, however, there is a shortage of relevant research exploring the incidence and likely effects of such mis-reporting in survey data.

Mis-reporting will often lead to the presence of "excess" zeros in empirical economics (and other areas), which has long been of interest to the applied researcher. To address such concerns, hurdle and double-hurdle models have been developed, and have found favour in areas ranging from a continuous dependent variable with a non-zero probability mass at (typically, but not exclusively) zero levels (Cragg 1971, Smith 2003); to the so-called zero-inflated (augmented) Poisson count data models (Mullahey 1986, Heilbron

1989, Lambert 1992, Greene 1994, Pohlmeier and Ulrich 1995, Mullahey 1997); and, more recently, to zero-inflated ordered probit (ZIOP) models (Harris and Zhao 2007). Typically, the issue that arises is that "zero" observations can result from two distinct processes and that ignoring this can lead to seriously mis-specified models.

In this paper, we explore the modelling of *sensitive* response variables: that is, variables where there is an associated loss-function (either perceived or actual) involved for the individual in terms of the responses he/she reports. Here, it is clear that the researcher must be aware of the potential for mis-reporting. Indeed, with regard to categorical ordered dependent variables, such mis-reporting has been approached by allowing the model's inherent boundary variables to vary by observed personal characteristics (see, for example, Maassen van den Brink and Groot 1999, Kristensen and Johansson 2008). However, here we suggest a more fundamental form of accounting for the potential mis-reporting which is likely to be present in data perceived to embody such a strong loss-function (social and/or legal) for the individual.

For consumption of goods with associated reporting loss-functions, the approach suggested here allows for these zero observations to correspond to both non-participants, but importantly also to those participants who, fearing repercussions, report zero-consumption when in fact this is not so. To be specific, we suggest a two-tiered sequencing of decision making. First, the individual makes a decision whether to participate or not; secondly, for participants only, there is the decision to mis-report or not. The second stage allows for our contention, that, especially regarding participation of "social bads" (licit, and in particular, illicit drugs, for example), some participants may intentionally mis-report their true consumption patterns. So within our econometric framework the probability of a zero observation is "inflated" as it is a combination of the probability of non-participation plus that from mis-reporting. In particular, we hypothesize that a potentially significantly large proportion of participants may actually report themselves as being non-participants, due to both moral and legal concerns about participation.

Our particular application lies in the important area of mis-reporting within the context of the consumption of illicit drugs. Given the considerable individual and social costs associated with the consumption of illegal drugs (including increased crime, health issues and difficulties at school or work) it is not surprising that an extensive body of research exists exploring issues related to the addictive nature of drugs as well as the relationship between the consumption of different types of drugs. However, as argued by MacDonald and Pudney (2000) and Pudney (2010), there is no consensus regarding policy prescriptions relating to drug abuse and, furthermore, analysis of survey data relating to drug use could potentially contribute to the policy debate in this area. It is apparent that the shortcomings of such data should be well understood, such as the likely possibility of under-reporting, which may mask the true extent of the problem, in order to make appropriate policy decisions. Indeed, in the context of survey response rates and response accuracy, Pudney (2010), p.26, comments that 'these problems cannot be overcome completely and their impact on research findings is not yet well understood.' Hence, we aim to contribute to the relatively small, but clearly important literature exploring the incidence and extent of mis-reporting (specifically with regard to drug consumption in our example) in individual level survey data. The results indicate that mis-reporting does has a significant effect on recorded drug participation rates. Compared to observed participation rates of 12, 3 and 1.3% (marijuana, speed and cocaine, respectively) true participation rates are estimated to be some 5 percentage points higher for cannabis, and nearly double for cocaine. There was an estimated 36% (18%) percent chance that a cocaine (cannabis) user would mis-report their participation. Less evidence of mis-reporting was found for speed users.

In summary then, our specific contributions to the literature are threefold. Firstly, we extend the general approach of Hausman, Abrevaya, and Scott-Morton (1998) to allow for covariates to influence the mis-reporting/misclassification decision; this will be very important for policy-makers in helping to identify those individuals with greater propensities to do so. Secondly, we acknowledge that many *sensitive* response variables of interest (illicit drug taking in our example) are likely to be consumed jointly as part of a *script*. Thus we once more extend the simpler univariate approach of Hausman, Abrevaya, and Scott-Morton (1998) to a multivariate one. Finally, we apply this new model to the consumption of illegal drugs (in Australia) and thereby give new evidence as to the likely extent of mis-reporting across three illegal drugs, but simulateneously also more likely rates of participation across these drugs than gleaned from a simple inspection of the raw observed participation rates.

2 The Econometric Framework

2.1 An Inflated Probit Model (IP)

We start by defining a discrete random variable y that is observable and assumes the binary outcomes of 0 and 1. A standard probit approach would map a single latent variable to the observed outcome $y = 1$ via an index function, and one would essentially model participation rates. However, it is our contention here that, especially regarding participation of "social bads", participants may intentionally mis-report their true consumption patterns. In particular, we hypothesize that a (potentially significantly large) proportion of participants will actually report themselves as being non-participants, due to both moral and legal concerns about participation.

Specifically, let r^* denote a binary variable indicating the split between Regime 0 (with $r = 0$ for non-participants) and Regime 1 (with $r = 1$ for participants). Although unobservable, r is related to a latent variable r^* via the mapping: $r = 1$ for $r^* > 0$ and $r = 0$ for $r^* \leq 0$. Thus r^* represents the propensity for participation and is related to a set of explanatory variables ($\mathbf{x}_r$) with unknown weights $\boldsymbol{\beta}_r$, and a standard-normally distributed error term (as is commonly assumed in the literature), ε_r such that

$$r^* = \mathbf{x}_r'\boldsymbol{\beta}_r + \varepsilon_r. \tag{1}$$

For participants ($r = 1$), a second latent variable, m^* represents the propensity to mis-report. Again this is related to a second unobserved variable m such that $m = 1$ for $m^* > 0$ and $m = 0$ for $m^* \leq 0$, where $m = 0$ represents a mis-reporter and $m = 1$, a true-reporter. Again, we can write this (linear) latent form as

$$m^* = \mathbf{x}_m'\boldsymbol{\beta}_m + \varepsilon_m. \tag{2}$$

Of course, neither r nor m is directly observed; the observability criterion for observed y is

$$y = r \times m. \tag{3}$$

Under the assumption that the stochastic terms $\boldsymbol{\varepsilon}$ $(\varepsilon_r, \varepsilon_m)$ are independent and follow standard Gaussian distributions, the full probabilities for $y = 0$ are given by

$$\Pr(y = 0\,|\mathbf{x}) = \Pr(r = 0|\mathbf{x}) + \Pr(r = 1|\mathbf{x})\Pr(m = 0|\mathbf{x}, r = 1) \tag{4}$$

and for $y = 1$ are

$$\Pr(y = 1 \,|\mathbf{x}) = \Pr(r = 1|\mathbf{x}) \Pr(m = 1|\mathbf{x}, r = 1). \tag{5}$$

where the first term on the right-hand side of equation (4) represents a genuine non-participant, the second term, a (participant) mis-reporter and the right-hand side term in equation (5), a (participant) true reporter.

These expressions can be stated simply in terms of joint probabilities by writing conditional probabilities as joint over marginals. Moreover, by independence these joint probabilities are simply products of the marginals such that they are respectively, given by

$$\Pr(y = 0 \,|\mathbf{x}) = [1 - \Phi(\mathbf{x}_r'\boldsymbol{\beta}_r)] + \Phi(\mathbf{x}_r'\boldsymbol{\beta}_r)[1 - \Phi(\mathbf{x}_m'\boldsymbol{\beta}_m)]$$

and

$$\Pr(y = 1 \,|\mathbf{x}) = \Phi(\mathbf{x}_r'\boldsymbol{\beta}_r) \Phi(\mathbf{x}_m'\boldsymbol{\beta}_m), \tag{6}$$

where Φ is the standard normal cumulative distribution function (*c.d.f.*). So here the probability of a zero observation has been "inflated" as it is a combination of the probability of non-participation plus that from mis-reporting. Accordingly we term this an Inflated Probit (IP) approach. This approach thus models "mis-reporting" explicitly and as a function of a set of explanatory variables unlike the model developed by Hausman, Abrevaya, and Scott-Morton (1998) where mis-reporting is accounted for using constant terms; or by Dustmann and Soest (2001) who decompose mis-classification errors in panel data into time-persistent and time-varying components and where the probability of classification is independent of respondent characteristics or any other factor.[1]

Given the assumed form for the probabilities and an *i.i.d.* sample of size N from the population on $(y_i, \mathbf{x})$, $i = 1, \ldots, N$, the parameters of the full model $\boldsymbol{\theta} = (\boldsymbol{\beta}_r', \boldsymbol{\beta}_m')' = \boldsymbol{\beta}'$ can be consistently and efficiently estimated using maximum likelihood (ML) techniques; the log-likelihood function is

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{N} \sum_{j} h_{ij} \ln [\Pr(y_i = j \,|\mathbf{x}, \boldsymbol{\theta})], \tag{7}$$

[1] As mentioned in the Introduction and Background, when there are ordered levels of consumption, some studies have modelled mis-reporting or mis-classification using a generalised ordered probit model or other variants of the generalised ordered probit model where the cut points are functions of covariates (Kristensen and Johansson 2008, Gannon 2009). The presence of mis-reporting is then inferred by comparing the predicted probabilities from these generalised models with those from the standard ordered probit ones.

where the indicator function h_{ij} is

$$h_{ij} = \begin{cases} 1 \text{ if individual } i \text{ chooses outcome } j \\ 0 \text{ otherwise.} \end{cases} \quad (i = 1, ..., N; \; j = 0, 1). \tag{8}$$

2.2 Generalising the Model to Correlated Disturbances (IPC)

As described above, the observed realisation of the random variable y can be viewed as the result of two separate latent equations, equations (1) and (2), with uncorrelated error terms. However, these equations correspond to the same individual so it is likely that the vector of stochastic terms $\boldsymbol{\varepsilon}_i$ will be related across equations. So, we can now extend the model to have $(\varepsilon_r, \varepsilon_m)$ follow a bivariate normal distribution with covariance matrix Ω, whilst maintaining the identifying assumption of unit variances. Thus Ω will have the form

$$\Omega = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \tag{9}$$

and the relevant probabilities will have the form

$$\Pr(y) = \begin{cases} \Pr(y = 0 \,|\mathbf{x}) = [1 - \Phi(\mathbf{x}_r'\boldsymbol{\beta}_r)] + \Phi_2(\mathbf{x}_r'\boldsymbol{\beta}_r, -\mathbf{x}_m'\boldsymbol{\beta}_m; \Omega) \\ \Pr(y = 1 \,|\mathbf{x}) = \Phi_2(\mathbf{x}_r'\boldsymbol{\beta}_r, \mathbf{x}_m'\boldsymbol{\beta}_m; \Omega) \end{cases} \tag{10}$$

where Φ_2 denotes the *c.d.f.* of the standardised bivariate normal distribution. ML estimation would again involve maximisation of equation (7) replacing the probabilities of (6) with those of (10) and re-defining $\boldsymbol{\theta}$ as $\boldsymbol{\theta} = (\boldsymbol{\beta}', \rho)'$. A test of $\rho = 0$ is jointly a test for independence of the two error terms and also one of the more general model given by equation (10) against the null of a simpler nested model of equation (6).

3 Extending to a Multivariate System

Often social bads such as licit and illicit drugs are consumed in a consumption bundle (see, for example, Collins, Ellickson, and Bell 1998, Ives and Ghelani 2006), given that they are habit-forming. For example, Saffer and Chaloupka (1999) report evidence suggesting complementarity between marijuana, cocaine and heroin and, similarly, DeSimone and Farrelly (2003) find that marijuana and cocaine are complements. Instead of modelling the consumption of such social bads in isolation, the above set-up can be extended to a multivariate framework where participation decisions are considered to be taken jointly by the same individual (see, for example, Zhao and Harris 2003, Ramful and Zhao 2009). The IP approach described in Section 2 ignores the potential cross-product correlations across multiple commodities for the same individual. Due to unobservable characteristics (such as individual tastes, addictive traits and risk-taking attitudes) the decision to consume multiple drugs is very likely to be related through the error terms of the participation and mis-reporting equations: that is, via the unobservables. As a consequence, vital cross-drug information is lost when the IP model is estimated in a univariate framework.

For a set of k $(k = 1, \ldots, K)$ multivariate IP models, the propensity for participation will be:

$$r_k^* = \mathbf{x}_{rk}' \boldsymbol{\beta}_{rk} + \varepsilon_{rk};\ (k = 1, ..., K) \tag{11}$$

and the propensity to mis-report will be:

$$m_k^* = \mathbf{x}_{mk}' \boldsymbol{\beta}_{mk} + \varepsilon_{mk};\ (k = 1, ..., K). \tag{12}$$

There is no necessary restriction that $\mathbf{x}_{rk} = \mathbf{x}_{rh}$ or that $\mathbf{x}_{mk} = \mathbf{x}_{mh}, \forall k \neq h$, but we will assume so, both in the empirical application and also below, to simplify notation. That is, the same specification in terms of covariate selection will apply to all of the drug equations. Note that economic and mathematical identification here though, will require that $\mathbf{x}_{rk} \neq \mathbf{x}_{mk}, \forall k$. That is, for each drug equation we do require exclusion restrictions with regard to the participation and mis-reporting equations; but these are not necessary *across* the different drug equations. The most general specification is to assume that the ε_{rk}'s and the ε_{mk}'s are freely correlated both within and across equations. This results in $s = 2K$ latent equations where the error terms jointly follow a multivariate normal distribution of order s with covariance matrix Σ given by

$$\Sigma_s = \begin{pmatrix} 1 & & & & & \\ \rho_{r_1 m_1} & 1 & & & & \\ \rho_{r_1 r_2} & \rho_{m_1 r_2} & 1 & & & \\ \vdots & \vdots & \vdots & \ddots & & \\ \rho_{r_1 r_K} & \rho_{m_1 r_K} & \rho_{r_2 r_K} & \cdots & 1 & \\ \rho_{r_1 m_K} & \rho_{m_1 m_K} & \rho_{r_2 m_K} & \cdots & \rho_{r_K m_K} & 1 \end{pmatrix}, \tag{13}$$

where $\rho_{r_1 m_1} = \rho_{m_1 r_1}$ and so on; that is, Σ_s is symmetric.

Consider a system of IP models for three illicit drugs. Since $K = 3$, we have six latent equations with a variance covariance matrix defined as

$$\Sigma_6 = \begin{pmatrix} 1 & & & & & \\ \rho_{r_1 m_1} & 1 & & & & \\ \rho_{r_1 r_2} & \rho_{m_1 r_2} & 1 & & & \\ \rho_{r_1 m_2} & \rho_{m_1 m_2} & \rho_{r_2 m_2} & 1 & & \\ \rho_{r_1 r_3} & \rho_{m_1 r_3} & \rho_{r_2 r_3} & \rho_{m_2 r_3} & 1 & \\ \rho_{r_1 m_3} & \rho_{m_1 m_3} & \rho_{r_2 m_3} & \rho_{m_2 m_3} & \rho_{r_3 m_3} & 1 \end{pmatrix}, \tag{14}$$

where, for example, $\rho_{r_1m_1}$ relates to the correlation between ε_{r_1} and ε_{m_1}: the respective error terms from the participation equation and the mis-reporting equation relating to the first drug; $\rho_{r_1r_2}$ is the correlation between ε_{r_1} and ε_{r_2}: the respective error terms from the participation equation for the first drug and the participation equation for the second drug; and $\rho_{r_1m_2}$, the correlation between ε_{r_1} and ε_{m_2}: the respective error terms from the participation equation for the first drug and the mis-reporting equation for the second drug; and so on. This results into a range of joint probabilities of interest such as the polar cases of

$$\begin{aligned} Pr(y_1 = 1, y_2 = 1, y_3 = 1|\mathbf{x}) = \\ \Phi_6(\mathbf{x}'_{r_1}\boldsymbol{\beta}_{r_1}, \mathbf{x}'_{m_1}\boldsymbol{\beta}_{m_1}, \mathbf{x}'_{r_2}\boldsymbol{\beta}_{r_2}, \mathbf{x}'_{m_2}\boldsymbol{\beta}_{m_2}, \mathbf{x}'_{r_3}\boldsymbol{\beta}_{r_3}, \mathbf{x}'_{m_3}\boldsymbol{\beta}_{m_3}; \Sigma_6) \end{aligned} \tag{15}$$

and

$$\begin{aligned} Pr(y_1 = 0, y_2 = 0, y_3 = 0|\mathbf{x}) = \Phi_3(-\mathbf{x}'_{r_1}\boldsymbol{\beta}_{r_1}, -\mathbf{x}'_{r_2}\boldsymbol{\beta}_{r_2}, -\mathbf{x}'_{r_3}\boldsymbol{\beta}_{r_3}; \Sigma_3) \\ + \Phi_4(\mathbf{x}'_{r_1}\boldsymbol{\beta}_{r_1}, -\mathbf{x}'_{m_1}\boldsymbol{\beta}_{m_1}, -\mathbf{x}'_{r_2}\boldsymbol{\beta}_{r_2}, -\mathbf{x}'_{r_3}\boldsymbol{\beta}_{r_3}; \Sigma_4) \\ + \Phi_4(-\mathbf{x}'_{r_1}\boldsymbol{\beta}_{r_1}, \mathbf{x}'_{r_2}\boldsymbol{\beta}_{r_2}, -\mathbf{x}'_{m_2}\boldsymbol{\beta}_{m_2}, -\mathbf{x}'_{r_3}\boldsymbol{\beta}_{r_3}; \Sigma_4) \\ + \Phi_4(-\mathbf{x}'_{r_1}\boldsymbol{\beta}_{r_1}, -\mathbf{x}'_{r_2}\boldsymbol{\beta}_{r_2}, \mathbf{x}'_{r_3}\boldsymbol{\beta}_{r_3}, -\mathbf{x}'_{m_3}\boldsymbol{\beta}_{m_3}; \Sigma_4) \\ + \Phi_5(\mathbf{x}'_{r_1}\boldsymbol{\beta}_{r_1}, -\mathbf{x}'_{m_1}\boldsymbol{\beta}_{m_1}, \mathbf{x}'_{r_2}\boldsymbol{\beta}_{r_2}, -\mathbf{x}'_{m_2}\boldsymbol{\beta}_{m_2}, -\mathbf{x}'_{r_3}\boldsymbol{\beta}_{r_3}; \Sigma_5) \\ + \Phi_5(\mathbf{x}'_{r_1}\boldsymbol{\beta}_{r_1}, -\mathbf{x}'_{m_1}\boldsymbol{\beta}_{m_1}, -\mathbf{x}'_{r_2}\boldsymbol{\beta}_{r_2}, \mathbf{x}'_{r_3}\boldsymbol{\beta}_{r_3}, -\mathbf{x}'_{m_3}\boldsymbol{\beta}_{m_3}; \Sigma_5) \\ + \Phi_5(-\mathbf{x}'_{r_1}\boldsymbol{\beta}_{r_1}, \mathbf{x}'_{r_2}\boldsymbol{\beta}_{r_2}, -\mathbf{x}'_{m_2}\boldsymbol{\beta}_{m_2}, \mathbf{x}'_{r3}\boldsymbol{\beta}_{r_3}, -\mathbf{x}'_{m_3}\boldsymbol{\beta}_{m_3}; \Sigma_5) \\ + \Phi_6(\mathbf{x}'_{r_1}\boldsymbol{\beta}_{r_1}, -\mathbf{x}'_{m_1}\boldsymbol{\beta}_{m_1}, \mathbf{x}'_{r_2}\boldsymbol{\beta}_{r_2}, -\mathbf{x}'_{m_2}\boldsymbol{\beta}_{m_2}, \mathbf{x}'_{r_3}\boldsymbol{\beta}_{r_3}, -\mathbf{x}'_{m_3}\boldsymbol{\beta}_{m_3}; \Sigma_6), \end{aligned} \tag{16}$$

and also intermediate ones such as

$$\begin{aligned} Pr(y_1 &= 1, y_2 = 0, y_3 = 0|\mathbf{x}) = \Phi_4(\mathbf{x}'_{r_1}\boldsymbol{\beta}_{r_1}, \mathbf{x}'_{m_1}\boldsymbol{\beta}_{m_1}, -\mathbf{x}'_{r_2}\boldsymbol{\beta}_{r_2}, -\mathbf{x}'_{r_3}\boldsymbol{\beta}_{r_3}; \Sigma_4) \\ &+\Phi_5(\mathbf{x}'_{r_1}\boldsymbol{\beta}_{r_1}, \mathbf{x}'_{m_1}\boldsymbol{\beta}_{m_1}, \mathbf{x}'_{r_2}\boldsymbol{\beta}_{r_2}, -\mathbf{x}'_{m_2}\boldsymbol{\beta}_{m_2}, -\mathbf{x}'_{r_3}\boldsymbol{\beta}_{r_3}; \Sigma_5) \\ &+\Phi_5(\mathbf{x}'_{r_1}\boldsymbol{\beta}_{r_1}, \mathbf{x}'_{m_1}\boldsymbol{\beta}_{m_1}, -\mathbf{x}'_{r_2}\boldsymbol{\beta}_{r_2}, \mathbf{x}'_{r_3}\boldsymbol{\beta}_{r_3}, -\mathbf{x}'_{m_3}\boldsymbol{\beta}_{m_3}; \Sigma_5) \\ &+\Phi_6(\mathbf{x}'_{r_1}\boldsymbol{\beta}_{r_1}, \mathbf{x}'_{m_1}\boldsymbol{\beta}_{m_1}, \mathbf{x}'_{r_2}\boldsymbol{\beta}_{r_2}, -\mathbf{x}'_{m_2}\boldsymbol{\beta}_{m_2}, \mathbf{x}'_{r_3}\boldsymbol{\beta}_{r_3}, -\mathbf{x}'_{m_3}\boldsymbol{\beta}_{m_3}; \Sigma_6). \end{aligned} \tag{17}$$

where Φ_n denotes the $n-$dimensional multivariate normal *c.d.f.*.

It is intuitive to take a closer look at these probabilistic expressions. Take, for example, equation (15), which corresponds to the probability of *observing* participation in all three drugs. Here all the six elements in parentheses on the right-hand side (RHS) relate to participation and true-reporting in all of the three respective drugs. On the other hand, $Pr(y_1 = 0, y_2 = 0, y_3 = 0|\mathbf{x})$, or equation (16), has a more complex form. This probability corresponds to an *observed* zero in each of the three drugs. This can occur in eight distinct ways; the individual can be:

1. a true non-participant in each drug; line 1.
2. a mis-reporting participant in drug 1, with the relevant (upper) integration limits being $\mathbf{x}'_{r_1}\boldsymbol{\beta}_{r_1}$ and $-\mathbf{x}'_{m_1}\boldsymbol{\beta}_{m_1}$, but a true non-participant in drugs 2 and 3 $\left(-\mathbf{x}'_{r_2}\boldsymbol{\beta}_{r_2}, -\mathbf{x}'_{r_3}\boldsymbol{\beta}_{r_3}\right)$; line 2.
3. $\cdots$

   $\vdots$

8. a mis-reporting participant in all drugs; line 8.

Note that Σ_j defines the relevant sub-matrices of Σ with appropriate signs in the correlations. For example, the relevant lower sub-matrix of Σ_4 in the second RHS term of equation (16) is defined as

$$\Sigma_4 = \begin{pmatrix} 1 & & & \\ -\rho_{r_1 m_1} & 1 & & \\ -\rho_{r_1 r_2} & \rho_{m_1 r_2} & 1 & \\ -\rho_{r_1 r_3} & \rho_{m_1 r_3} & \rho_{r_2 r_3} & \end{pmatrix}.$$

Given an *i.i.d.* sample of N individuals making an observed choice of $(0, 1)$ across all three drugs with associated probability P_i^* of the form defined above, the log-likelihood function is

$$\log L = \sum_{i=1}^{N} \log P_i^* \tag{18}$$

where P_i^* is the joint probability across all K drugs corresponding to the observed choice of individual i.

The multivariate Inflated Probit model can then be estimated by maximising this log-likelihood function. Because the probabilities that enter the likelihood are functions of high dimensional multivariate normal distributions, these are simulated using the GHK algorithm. We follow the recent literature and use Halton sequences to generate the uniform variates required to evaluate the GHK probability simulator.[2] In addition, since

[2]500 such draws were used in estimation: using more draws did not affect results.

the joint and conditional probabilities are highly non-linear functions of $\mathbf{x}$, analytical solutions of partial effects are difficult to obtain. Thus, the partial effects are calculated using numerical gradients. As is standard in the literature, the standard errors of the partial effects are then estimated using the delta method (using the estimated Hessian) which provides an approximation to the asymptotic distributions of the partial effects. We return to such partial effects, and other summary measures, in the results sections below.

4 An Application to Drug Consumption

As mentioned in Section 1, an extensive body of research exists exploring issues related to the addictive nature of drugs as well as the relationship between the consumption of different types of drugs. This is not surprising given the considerable individual and social costs associated with the consumption of illegal drugs. Consequently, large amounts of public funds are spent worldwide on educational programs and promotional campaigns to reduce the consumption of drugs. Empirical studies play a crucial role in helping to identify the socioeconomic and demographic factors associated with the consumption of illicit drugs, providing invaluable information to facilitate well-targeted public health policies.

One strand of the existing literature in this area focuses on exploring the determinants of the decision to take illegal drugs. For example, in one of the early contributions Sickles and Taubman (1991) use data drawn from the US National Longitudinal Survey of Youth (NLSY) to explore the question as to who uses illegal drugs. The findings, which are based on a logistic model, suggest that socioeconomic variables such as religious preference have a statistically significant influence on drug use. Similarly, Gill and Michaels (1991) also analyse the NLSY in order to explore the determinants of an individual's decision to use illegal drugs using a probit model. Their findings suggest that personal attributes rather than economic factors (including marital status, ethnicity and family background) play a dominant role in such decisions. The authors, however, acknowledge that due to the absence of information on prices, their findings should be regarded as "tentative". Saffer and Chaloupka (1999) analyse the U.S. Household Survey on Drug Abuse using a variety of probit specifications to explore the effects of drug prices on participation for a range of illegal drugs. Their findings suggest that drug participation responds to prices and there is evidence of complementarity across some drug types, namely marijuana, cocaine and heroin.

More recently, Duarte, Escario, and Molina (2005) find that illegal drug use among Spanish adolescents is determined by economic factors such as income as well as socio-demographic characteristics such as personal habits, family environment and receiving information relating to the adverse effects of drug use. Using data from the Australian National Drug Strategy Household Survey, Ramful and Zhao (2009) find marijuana and cocaine to be more popular among young adults, males, and single and unemployed individuals.

However, one of the key issues in the empirical literature on drug addiction and the demand for illicit drugs relates to the accuracy of self-reported data and the incentive to mis- and under-report illicit drug use. The extent of such mis- and under-reporting is likely to be influenced by a variety of factors. In terms of differences across socioeconomic groups, Mensch and Kandel (1988) find that females and ethnic minorities have a tendency to under-report drug consumption. Similarly, Fendrich and Vaughn (1994) find that ethnicity has an important influence on the under-reporting of substance abuse.

Mis-reporting of drugs use may also be influenced by how the survey is conducted. In particular, traditional paper and pencil self-administration 'interview' methods by post, or handing out paper questionnaires in person and asking participants to complete them by hand and return them to the researcher, has been associated with lower under-reporting of sensitive information (Bowling 2005). This is due to the greater anonymity, more privacy and confidentiality of the method. For instance, comparing the mail survey method to computer-assisted telephone interviews (CATI), Kraus and Augustin (2001) found that a lower number of respondents would admit alcohol consumption if questioned by telephone compared to self-reports from questionnaires. In a similar vein, Hoyt and Chaloupka (1994) and Fendrich and Vaughn (1994) find that lower reported drugs use is associated with telephone interviews. The increased use of computer assisted self-interviewing in the gathering of information has arguably improved the accuracy of such data although it is not clear to what extent the accuracy has been improved (Morrison-Beedy, Carey, and Tu 2006).

In addition, given the apparent complex interrelationships between the demand for different types of illicit drugs, it is apparent that the extent of mis-reporting may vary across different types of drugs, arguably being particularly serious in the case of "hard" drugs (such as heroin and cocaine). Pudney (2007) analyses the consequences of mis-reporting of illicit drugs use for statistical inference, using UK panel data containing repeated information of self-reported lifetime drugs use, *i.e.*, repeated questions relating to whether individuals have ever taken particular drugs. The findings indicate serious under-reporting of the use of cannabis and cocaine, which in turn leads to bias in statistical modelling. For example, for one of the data sets analysed, under-reporting rates for cannabis (cocaine) with bounds averaging from 23 to 60% (31 to 95%) for all individuals were found. Such findings are supported by the evidence from surveys which check self-reported data via drug tests (usually for prisoners or arrestees), which indicate serious mis-reporting problems in the case of hard drugs (see, for example, MacDonald and Pudney 2003). For example, in an early contribution, Wish (1987) analysed a sample of men arrested in New York City in 1984. For cocaine, the interview data indicated a drug use rate of 43% as compared to a drug use rate of 82% elicited from urine specimens. More recently Lu, Taylor, and Riley (2001) compare under-reporting of crack cocaine use with that of other drugs by validating information obtained via interviews with urinalysis for a sample of adult arrestees. The findings indicate significant levels of under-reporting for all drugs with the least amount of mis-reporting for the use of marijuana, followed by methamphetamine, crack, and opiate, and with truthful reporting declining from 64% in the case of marijuana to 46% in the case of opium.

It should be acknowledged, however, that the extent to which such findings from studies where such cross validation is possible can be generalised, is not apparent and is arguably limited given that such samples are based on somewhat atypical circumstances. The modelling strategy outlined above, in contrast, relies on a single source of cross-section survey data without recourse to validation from other sources such as drug tests or historical information on lifetime drugs consumption.

4.1 The Data

The data we use for the model are drawn from the Australian National Drug Strategy Household Survey (NDSHS), which is a nationally representative survey of the non-institutionalized Australian civilian population aged over 14 providing information on drug use patterns, attitudes and behaviour (NDSHS 2010). A multi-stage, stratified area sample design ensured a random sample of households in each geographical stratum. As mentioned above, there has been some discussion in the existing literature regarding the potential for mis-reporting to be influenced by how the survey is conducted. The earlier waves of the NDSHS used face-to-face and drop-and-collect methods to collect data. The computer-assisted telephone interview (CATI) method of data collection was introduced in the 2001 survey. In that particular survey, all three methods were employed to collect data. The 2004 and 2007 surveys, on the other hand, were administered using only drop and collect and CATI. Note that this survey consists of independent cross-sectional surveys over time (that is, it is not a panel data set, and we pool the years into a single data set).

In this data set, neither the monetary expenditures nor the physical quantities of the illicit drugs consumed are reported. The information on individuals' drugs consumption is given via a discrete variable measuring whether they have consumed the drug in question over the last 12 months. The three questions of interest were: have you used marijuana or cannabis in the last 12 months? Have you used speed in the last 12 months? And have you used cocaine in the last 12 months? There have been seven NDSHSs conducted since 1985. In this paper, due to consistency with respect to the key variables of interest, we use data from the three most recent surveys (2001, 2004 and 2007). A sample of 50,153 individuals is thus available for estimation. This data has been used in several previous studies (see, for example, Cameron and Williams 2001, Williams 2003, Zhao and Harris 2004, Harris and Zhao 2007). We focus on three illicit drugs, namely, marijuana, speed and cocaine since information on state level prices is available for these three drugs. Information on the price of heroin is also available; however, the extremely low recorded prevalence rate of heroin consumption (at 0.2%) essentially precludes us from modelling consumption of this drug.

As mentioned above, the absence of such price data has been problematic in some of the existing studies in this area (see, for example, Gill and Michaels 1991). Hence, the inclusion of such information is an important feature of our empirical analysis. The prices of the three illicit drugs are obtained at state level from the Illicit Drug Reporting System (IDRS). The IDRS collects such data predominantly from interviewing injecting drug users and key informants who have regular contact with illicit drug users but which may potentially exhibit coverage error (NDARC 2009). In occasional cases where a price report is missing, it is constructed using information from the Australian Bureau of Criminal Intelligence (ABCI), recently replaced by the Australian Crime Commission (ACC). The ABCI/ACC is an alternative source for drug prices, which collects information on drugs through covert police units and police informants (ACC 2010). The advantage of using price data from the IDRS is that they are provided with unified measures and fewer missing observations. To be specific, the price of marijuana is measured in dollars per ounce and the respective price of speed and cocaine is measured in dollars per gram.

In terms of explanatory variables, we have two sets, one to determine participation and the other, for mis-reporting (recalling the same sets are used for each drug). While many of these variables overlap, to facilitate identification we ensure that both $\mathbf{x}_r$ and $\mathbf{x}_m$ have exclusion restrictions. We discuss the set of common variables first. In line with several past studies on drug consumption (see, for example, Gill and Michaels 1991, Hoyt and Chaloupka 1994, Saffer and Chaloupka 1999, Cameron and Williams 2001), we include a wide range of personal and demographic characteristics in $\mathbf{x}_r$ (the variables determining participation), namely: gender; marital status; a quadratic in the individual's age (standardised; mean subtracted and scaled by sample standard deviation); a dummy variable for whether the individual migrated to Australia in the last 10 years; a dummy variable for whether the respondent is of Aboriginal or Torres Strait Islander origin; a dummy variable for whether there are preschool children in the household; and, finally, whether the individual comes from a single parent household. We also control for educational attainment distinguishing between four categories of highest educational attainment: a tertiary degree; a non-tertiary diploma or trade certificate; year 12 education; and less than year 12 education, which is the omitted category. In terms of regional controls, we include dummy variables for whether the individual resides in a capital city and for whether the individual resides in a state where possession of small amounts of marijuana is decriminalised. With regard to the individual's economic situation, we control for the natural logarithm of real personal annual income before tax measured in Australian Dollars and the individual's main labour market status, *i.e.*, employed, studying, unemployed and other activities such as retired, on a pension or performing home duties, which form the omitted category.

As noted, we do not rely uniquely on functional form for identification but also include a range of identifying variables in the participation equation: variables that affect participation, but do not so mis-reporting propensities. In particular, drug culture, or peer drug use, has been identified as an important risk factor for drug consumption (see, for example, Delaney, Harmon, and Wall 2008, Kenkel, Reed III, and Wang 2002, Pudney 2004). We therefore include a dummy in $\mathbf{x}_r$ that takes value 1 if most or all of the individual's friends and acquaintances used any of the three drugs.

The inclusion of the prices of the three illicit drugs is an important feature of our data set. Hence we control for the natural logarithm of the real price of: marijuana (measured in dollars per ounce); speed (dollars per gram); and cocaine (dollars per gram). We also control for the prices of a range of complementary/substitute drugs, namely, the natural logarithm of the real price index of: alcohol; tobacco; and, finally, heroin (dollars per gram). The data on alcohol and tobacco prices are obtained in the form of indices from the Australian Bureau of Statistics (ABS 2010). The price of heroin is obtained from similar sources as those of the other illegal drugs and is measured in dollars per gram and converted into natural logarithm. As these prices will be orthogonal to any mis-reporting propensities, they provide us with further identifying identifying variables.

A number of studies have identified long-term implications of risk-taking by youths with evidence of an inter-temporal correlation of risk-taking as a youth and as an adult (see, for example, Gruber 2001, Hanna, Yi, Dufour, and Whitmore 2001). Tattooing and body piercing are yet another set of potential identifying variables available to us for drug use; these reflect an individual's attitude towards risk and some studies have revealed that there is an association between tattoo and body piercing procedures and illicit drug use (Deschesnes, Finès, and Demers 2006, Stephens et al. 2003). We thus include as further identifying variables in the participation equation dummy variables for whether the individual has ever undergone a tattoo procedure; and whether the individual has ever undergone a body piercing procedure. Finally, although we include year dummies to allow for the fact that participation rates may follow different trends over time, we have no prior concerning any similar trends in mis-reporting such that these variables are omitted from the mis-reporting equation, thus providing further identifying variables for the participation equation.[3]

Arguably of more interest are the variables we include in $\mathbf{x}_m$: the variables determining mis-reporting. Other than the standard demographic and socioeconomic characteristics noted above, we include additional variables to capture mis-reporting and further enhance identification. Our model fundamentally relies on these identifying variables and it is therefore important to identify factors that influence the mis-reporting decision but not

[3]It is important to acknowledge that some of the variables are potentially endogenous (such as having tattoos, piercings or peer group effects. We have experimented with different specifications whereby we omit such variables and our findings are generally unchanged.

the participation one. In this study we choose identifying variables that mostly relate to the conditions under which the survey was administered, which as mentioned above, may potentially influence the extent to which individuals mis-report but will clearly be independent from any participation propensities.

Specifically, we control for: if anyone else was present when the respondent was completing the survey questionnaire; if anyone helped the respondent complete the survey questionnaire; and the survey mode, *i.e.*, a dummy variable which takes a value of 0 if the drop-and-collect method was used and takes the value of 1 if the CATI or face-to-face method is used. These variables conform with the factors that have been associated with mis-reporting or mis-classification in prior studies (see, for example, Mensch and Kandel 1988, Hoyt and Chaloupka 1994, O'Muircheartaigh and Campanelli 1998, Lu, Taylor, and Riley 2001, Kraus and Augustin 2001, Berg and Lien 2006) although none of these studies have modelled "mis-reporting" explicitly. There is a general consensus among most of these studies that the presence of interviewers - either on telephone or face to face - leads to a reluctance to reveal characteristics believed to be socially negative or unacceptable. For instance, Kraus and Augustin (2001) who examined mode differential effects on alcohol consumption and alcohol-related problems found that patterns of drinking and alcohol-related problems are more easily reported in self-administration questionnaires compared to telephone interviews. O'Muircheartaigh and Campanelli (1998) found some effect on survey precision when children were present during the interview.

Finally, we also include as an instrument a variable indicating a general lack of trust in the survey which we measure using the percentage of compulsory questions left unanswered in the survey. This is based on the significant amount of literature suggesting that the longer a respondent spends with the interviewer, the more trusting they are of both him/her and the survey in general (see, for example, Corbin and Morse 2003). For each respondent it is possible to calculate the total number of compulsory (asked to everyone) left unanswered (as a percentage); this is clearly both a strong proxy for length of time spent completing the survey, and as such an increasing proxy for trust, and also (arguably) a direct measure of trust.[4]

Table 1 presents summary statistics relating to the variables used in our econometric

[4]Given that the selection of identifying variables is always open to debate, we have explored various sets of identifying variables and our broad findings remain generally unchanged.

analysis for the pooled cross-section data set.[5] It is apparent that out of the three illicit drugs, the (recorded) consumption of marijuana is the most prevalent at 12%, followed by speed at 3% and cocaine at 1%. In terms of personal characteristics, 47% of the sample are male and 60% of the sample are married, with 63% of the sample being employed and the most populated highest equational attainment category is diploma level education at 35%. In terms of the variables employed to capture the possibility of mis-reporting, 29% of the sample were interviewed with another individual present, 23% of the sample indicated that they were given help to complete the questionnaire, 18% were interviewed using the CATI or face-to-face method and on average 4% of compulsory questions were left unanswered in the survey.

5 Results

In Tables 2 and 3 we present the estimated coefficients for the participation and mis-reporting equations for the three drugs (respectively, for variables common to both equations and identifying variables), and in Table 4 the estimated correlation coefficients across the implicit six equations. We note in general, that the systems-model does a very good job, in terms of statistical significance, of modelling such difficult data with such low observed (recorded) participation rates. We turn first to the results determining participation.

In accordance with the existing literature, being male has a strong statistically significant positive effect on the probability of participation in the case of all three drugs, which ties in with existing evidence suggesting that males are less risk averse than females (see, for example, Borghans, Golsteyn, Heckman, and Meijers 2009). Age also appears to have a statistically significant effect on the participation probabilities of all three drugs indicating that illicit drug consumption varies over the life cycle. The issue of considering the future consequences of current consumption behaviour is an important feature of the rational addiction model proposed by Becker and Murphy (1988). Current consumption of an addictive good by a young adult raises his or her marginal utility of future consumption but also lowers overall utility in the future due to a detrimental health effect. We thus allow for the likelihood of an "n-shaped" age-participation profile by specifying

[5]Note that the full variable definitions are given in the Appendix, as well as defining any acronyms used.

a quadratic relationship for age. The significant negative age-squared terms in the marijuana and cocaine equations indicate an $n-$shaped profile of participation with respect to age for these two drugs, while in the case of speed, participation declines exponentially with age. In accordance with intuition, both being married and having preschool children have a strong inverse effect on the probability of participation across all three drugs, but single parent status has no significant effect on any. The effect of ethnicity clearly varies by the type of illicit drug. For example, being of Aboriginal or Torres Strait background is associated with a higher probability of marijuana participation but lower probabilities of speed and cocaine participation. Living in capital cities is associated with a higher probability of speed and cocaine participation but has no significant effect on marijuana participation. This is an interesting finding and may well be related to the differing distribution networks for these respective drugs. While being a new migrant has no significant effect on speed participation, it is significantly associated with that for both marijuana and cocaine.

In accordance with previous literature (see, for example, Farrelly, Bray, Zarkin, and Wendling 2001, Saffer and Chaloupka 1998), whether the individual resides in a state where possession of small amounts of marijuana is decriminalised, is positively associated with participating in marijuana consumption; but statistically insignificantly associated with the other two drugs, suggesting that individuals do respond to specific changes in the legal environment. With respect to labour market status, relative to being retired or performing home duties, being employed or currently studying are insignificantly associated with the probability of participating in any of the three drugs. In accordance with previous studies, such as Ramful and Zhao (2009), being unemployed relative to being retired or performing home duties, on the other hand, is positively associated with participating in all three drugs, but statistically insignificant for speed, which may reflect the social environment and/or norms of the unemployed. Consistent with the standard demand schedule, income is also positively and significantly associated with the probability of participation in all of the three drugs. The relationship between education and illicit drug consumption appears to be somewhat complex with the effect of education varying significantly across the three drugs: education has no significant effect on cocaine use; in the case of marijuana, higher levels of education are associated with a higher probability of participation; while for speed the more highly educated is the individual, the lower is his/her probability of participating. Such findings may reflect the different social norms, recreational activities and/or preferences across educational groups, with a particularly marked difference across the educational profile of the consumers of marijuana and speed.

As mentioned above, the identification of our model relies to a large extent on the exclusion variables. Our estimates in the participation equation indicate that our identifying variables are all, in general, statistically significant. Having a tattoo or body piercing is positively correlated with participation in the case of all three drugs. As expected, peer effect has a strong and positive relationship on participation for all three drugs, *i.e.*, if most, or all, of the individual's friends and acquaintances used drugs then he/she has a higher probability of participation. This is consistent with previous literature which indicates a positive effect of the wetness of the environment or peer effect on drug use. As for the price effects, our results suggest complex interrelationships between the demand for different types of illicit drugs. Whilst all three drugs' own prices have a negative relationship with their probability of consumption only the price effect of speed is statistically significant. The cross-price effects of all three drugs are statistically insignificant except (somewhat weakly) for the price of cocaine on marijuana participation. In terms of the effects of the prices of other complementary or substitute drugs, the price of tobacco has a significant negative effect only on speed participation, which accords with a complementary relationship between these two drugs. Alcohol price has a positive and significant effect on speed and cocaine participation which suggests that alcohol is a substitute for the two drugs. In terms of the price of heroin, positive effects on the consumption of speed and marijuana suggest substitution effects whilst a negative effect in the case of cocaine, albeit at 10% level of significance, is in accordance with a complementary relationship between heroin and cocaine. This ties in with the findings of Jofre-Bonet and Petry (2008) who explore polydrug use patterns for heroin and cocaine addicts in the US based on experiments measuring drug elasticities following changes in heroin and cocaine prices in the context of an Almost Ideal Demand System, and find the two drugs to be economic complements.

We next turn to the mis-reporting equations. Focusing only on the statistically significant effects and noting that a positive coefficient indicates a lower probability of mis-reporting, being male is associated with a higher probability of mis-reporting speed but with lower chances of mis-reporting marijuana consumption, which may reflect this particular drug being generally regarded as more socially acceptable. Surprisingly, age has no significant effect on the mis-reporting probability for any of the drugs, suggesting that the tendency to mis-report does not vary over the life cycle. In terms of labour market status, being unemployed is associated with a higher probability of mis-reporting cocaine but not the other two drugs. Interestingly, income is positively associated with mis-reporting marijuana but negatively related to mis-reporting cocaine. As expected, the more educated individuals have a higher probability of mis-reporting marijuana but education does not seem to affect reporting behaviour of the other two drugs. The prevalence of marijuana consumption amongst the more highly educated may lead to increased awareness of the consequences of consuming such an illicit drug amongst this group thereby influencing mis-reporting for this particular drug type. New migrants are more likely to mis-report participation in speed and cocaine, which may reflect a particular concern regarding the risks involved with consumption of the two relatively hard drugs. Surprisingly, decriminalisation is associated with a higher probability of drug mis-reporting but the effect is only significant in the case of marijuana and cocaine. A possible explanation is that the introduction of such decriminalisation is often associated with increased debate and discussion of illicit drug use as well as campaigns such as the Australian National Campaign Against Drug Abuse (NCADA), which may lead to increased awareness of the potential consequences associated with consuming illicit drugs thereby impacting on the tendency to mis-report consumption.

With respect to the effects of the identifying set of variables in the mis-reporting equation, it is apparent that the presence of anyone else when the respondent was completing the questionnaire is associated with a higher probability of mis-reporting across all three drugs, as found by Hoyt and Chaloupka (1994). Seeking help from someone to complete the questionnaire also increases the chances of mis-reporting marijuana but does not have a significant effect in the case of the other two drugs. Clearly, survey type, *i.e.*, the CATI method/face to face interview (relative to drop-and-collect), is associated with a higher probability of mis-reporting across all three drugs. Finally, if the respondent had a general lack of trust in the survey then he/she has a higher chance of mis-reporting drug use although the effect in the case of speed is statistically insignificant. In summary, these identifying variables exhibit high levels of significance and in the expected direction, such that along with similarly strong identifying variables in the participation equations, we are confident in our identification strategy and, consequently, our results overall.

In terms of the correlation coefficients presented in Table 4, strong statistically significant correlations between the mis-reporting equations are found for only marijuana and speed. Similarly, in terms of participation, statistically significant correlation coefficients are found for: speed and marijuana; cocaine and marijuana; and speed and cocaine. Interestingly, a strong correlation coefficient is found between the speed participation equation and the marijuana mis-reporting equation. A statistically significant correlation coefficient, albeit at 10%, is also found between the cocaine participation equation and the marijuana mis-reporting equation. The estimated correlation coefficients therefore suggest the existence of complex interrelationships between participation in the consumption of the three illicit drugs as well as in the propensity to mis-report participation across the three drugs. We note that this set of correlations is jointly significant; and also that they tend to be all positive - positive unobservables driving (in part) one equation are associated with similarly positive ones driving (in part) all others.

5.1 Predicted Probabilities

There is a range of probabilities one may be interested in predicting with the current model. For each drug in isolation, one may be interested in probabilities such as: the marginal probability of participation; the joint probability of participation and mis-reporting; or the probability of truthful reporting, conditional on participation. Moreover, considering the full system of demand equations, as is done in the current approach, one may be interested in any of numerous probabilities such as: the joint probability of participating in marijuana, speed and cocaine; the joint probability of mis-reporting marijuana, speed and cocaine; the conditional probability of mis-reporting cocaine conditional on marijuana participation; the participation in cocaine probability conditional on speed and marijuana participation; and so the list goes on.

In Table 5 we present (for obvious reasons) only a selection of such probabilities: the marginal and conditional probabilities associated with single drugs. Note that all of these expressions are evaluated at individually observed characteristics, and then averaged over the sample (as opposed to at mean values of observed characteristics). As expected, across all three drugs, the predicted marginal probabilities of participation are higher, in particular for marijuana and cocaine, than the sample rates of participation as indicated by the survey responses. Specifically, based on the survey responses, one would estimate participation rates in marijuana, speed and cocaine, respectively, to be 11.8, 3.0 and 1.3%. However, we estimate, once mis-reporting has been taken into account, that these are significantly higher at 16.7, 3.6 and 2.3%, respectively. Moreover, given the small standard errors of these, they appear to be quite precisely estimated.

What then of the estimated mis-reported probabilities? Conditional on an individual participating, we see that for cocaine there is staggering 36% chance of mis-reporting. For cannabis, this is still quite large, at 18%, although much less than with the harder drug of cocaine (as would be expected due to increased adverse relative perception of the two drugs). Somewhat surprising, is the relatively small conditional probability of mis-reporting for speed, at 5%. This is still not negligible, but clearly significantly less than the other two drugs.[6] Overall, these findings suggest that mis-reporting in survey data may lead to considerable underestimation of participation rates in the case of consumption of illicit drugs, especially with regard to both cannabis and cocaine in the current study.

To gain more insights into the source of the observed zeros, we also present in Table 5 the predicted probability of the zeros for each of the three drugs broken down into two respective components: non-participation and mis-reporting. For instance, the overall predicted probability of 86.7% of zero consumption in the case of marijuana is made up of the respective probability of, non-participation (83.3%), and mis-reporting (3.6%). For speed, of the total probability of a recorded zero of 97%, some 0.73 percentage points can be attributed to mis-reporting; and finally for cocaine, of the overall 99%, some 0.80 percentage points can be attributed to mis-reporting. In view of the low rates of participation, the mis-reporting components here may appear to be small. However, when translated to the Australian population aged 14 and above, they represent nearly 600,000, 130,000 and 145,000 cases of unreported cases of marijuana, speed and cocaine use, respectively. Such under-reporting can thus have important implications for drug policies.

Such probabilities can be thought of as *prior* probabilities. That is, they apply to a randomly selected individual from the population, about whom we know nothing except for their characteristics. However, to provide further insights into the extent of mis-reporting, it is possible to estimate *posterior* probabilities, analogous to those considered in latent class models (Greene 2008), that are conditional on us knowing what outcome the individual chose. Specifically, here this allows us to make a prediction on what percentage

[6] Possibly this finding is related to the demographic of speed users: typically being much more focussed in the young adult age group with low education. For young adults, the desire to appear more "socially acceptable" may be much less and speed consumption may be considered much more of a norm. Also, there is arguably a lower chance of dishonest reporting among the less educated groups: educated individuals being generally more concerned about the perceived stigma associated with drug use.

of these zeros come from non-participation and mis-reporting respectively, using all the information we have on the individual: this attempts to answer the question, *given that an individual recorded a zero, what is the probability that they are a true non-participant versus a mis-reporting participant (given their observed characteristics)?* The posterior probabilities for the two types of zeros are given as (Greene 2008),

$$\begin{aligned}
\Pr\left(r=0|\mathbf{x},y=0\right) &= \frac{f(r=0|\mathbf{x})}{f(y=0)} \\
&= \frac{1-\Phi\left(\mathbf{x}_r'\boldsymbol{\beta}_r\right)}{\left[1-\Phi\left(\mathbf{x}_r'\boldsymbol{\beta}_r\right)\right]+\left[\Phi_2(\mathbf{x}_r'\boldsymbol{\beta}_r,-\mathbf{x}_m'\boldsymbol{\beta}_m,-\rho_{rm})\right]}
\end{aligned} \tag{19}$$

and

$$\begin{aligned}
\Pr\left(r=1,m=0|\mathbf{x},y=0\right) &= \frac{f(r=1,m=0|\mathbf{x})}{f(y=0)} \\
&= \frac{\Phi_2(\mathbf{x}_r'\boldsymbol{\beta}_r,-\mathbf{x}_m'\boldsymbol{\beta}_m,-\rho_{rm})}{\left[1-\Phi\left(\mathbf{x}_r'\boldsymbol{\beta}_r\right)\right]+\left[\Phi_2(\mathbf{x}_r'\boldsymbol{\beta}_r,-\mathbf{x}_m'\boldsymbol{\beta}_m,-\rho_{rm})\right]}.
\end{aligned} \tag{20}$$

From Table 5, we find that just over 77% of the reported zeros for marijuana are estimated to come from genuine non-participation while almost 23% come from those who have mis-reported their participation. For speed and cocaine, about 95% of the zero participation is accounted by non-participation while about 5% comes from mis-reporting. Note that as with the prior probabilities presented earlier, these posterior ones for mis-reporting might appear, superficially, rather low. However, it is important to remember that the probabilities for mis-reporting here are not marginal, but joint of participation *and* mis-reporting. Thus given participation probabilities are so low for all of these drugs (estimated at some 17, 4 and 2%, respectively, for cannabis, speed and cocaine, see second row of Table 5), it is not surprising that these joint probabilities are also so small.

We can also estimate partial effects on all the different marginal, joint and conditional probabilities. For brevity, we present partial effects for two joint and one conditional probability, which we discuss briefly. In particular, Tables 6 and 7 present partial effects on the probabilities of the two extreme cases (estimated at sample means): the probability of zero consumption of all three drugs and the *recorded* probability of positive consumption of all three drugs, *i.e.*, $Pr(y_{mar}=1,y_{spd}=1,y_{coc}=1|\mathbf{x})$ which can alternately be written as $Pr(r_{mar}=1,m_{mar}=1,r_{spd}=1,m_{spd}=1,r_{coc}=1,m_{coc}=1|\mathbf{x})$. Firstly, with regard to zero reported consumption of all three drugs, it appears that being male, having a tattoo and body piercing, and peer effects are all inversely associated with this probability, with the non-participation and mis-reporting reporting effects serving to operate in the same direction. For instance, males are 3.5 percentage points (pp) less likely to abstain from all three drugs and they have a 1.2pp lower chance of truthfully reporting such zero consumption. This results in an overall 4.7pp lower probability of recording zero consumption for males compared to females. With regard to peer effects, individuals who had most or all of their friends and acquaintances using drugs, have a 25.2pp lower chance of abstaining from all three drugs, once again endorsing the importance of peer effects in the consumption of illicit drugs.

The effects of higher education and marijuana decriminalisation are interesting with negative effects on the probability of reporting non-participation across all three drugs with the mis-reporting effects operating in the opposite direction thereby serving to moderate the participation effects. In states which have decriminalised marijuana use, there is a 1.8pp lower probability of non-participation in all three drugs, counteracted by a higher 0.7pp chance of a truthful reporting of non-participation into all three, resulting in a 1.1pp lower chance overall of observing zero recorded consumption. From a policy-making perspective, a lower chance of observing zero consumption of all three drugs may be associated with a higher probability of consumption of softer drugs accompanied by lower consumption of harder drugs. Hence, the finding of decriminalisation being associated with a lower overall probability of observing zero recorded consumption of all three drugs does not in itself signal issues with the policy. Turning to education, degree holders have a 2.4pp lower chance of abstaining from all three drugs but a 1pp higher chance of truthfully reporting such non-participation resulting into an overall 1.4pp lower probability of recording joint zero consumption across all three drugs relative to those with less than year 12 qualifications. The only price effect that attains statistical significance is the (negative) effect of heroin price. In terms of the additional variables in the mis-reporting equation, positive statistically significant partial effects are apparent for all four survey-related variables, again highlighting the important role of survey conditions in the collection of accurate (or otherwise) information.

Secondly, in terms of the partial effects related to the probability of reporting consumption of all three drugs, since this overall probability of participating in all three drugs is very small, the partial effects are also very small. We thus scale them by $1,000$ here for the sake of presentation. While none of the variables achieved significance in the mis-reporting equation, we find that being male, being unemployed or a student, having a tattoo and body piercing, and peer effects are positively associated with joint participation in all three drugs. For example, males, and those with the majority of their peers taking drugs, have a 0.018pp and 0.077pp higher respective probability to jointly participate in all three drugs, which may reflect the lower levels of risk aversion typically observed amongst males as well as the importance of the social environment and networks.

Finally, in order to highlight the flexibility of our statistical framework, Table 8 presents partial effects on the probability of reporting zero consumption of speed and cocaine, conditional on *estimated* participation in marijuana, *i.e.*, $Pr(y_{spd} = 0, y_{coc} = 0|r_{mar} = 1, \mathbf{x})$. Bringing an analogy with the gateway effect where there is a progression from soft drugs to hard drugs, this probability allows us to examine zero reporting in the case of the harder drugs, cocaine and speed, in a subpopulation of marijuana users. We find a significant association of factors such as marital status, presence of young children, race and education with the non-participation of speed and cocaine in the subpopulation of marijuana participants. For example, among users of marijuana, degree holders have a 2.4pp higher chance of non-participation in speed and cocaine than those with less than year 12 education. Put differently, the more educated individuals are less likely to be hard core drug users (*i.e.*, jointly consume speed and cocaine) if they are already marijuana users, which does not suggest progression from soft to hard drugs (in this case).

6 Conclusions

In this paper we have explored the potential implications of mis-reporting in survey data in the context of reporting consumption of three illicit drugs (namely marijuana, cocaine and speed). The widespread use of data collected from individual and household level surveys by researchers and policy-makers is clearly reliant on respondents supplying accurate and reliable information. Indeed, estimated participation rates of illegal drugs are invariably inferred from such sample based data. It is apparent, however, that in the context of gathering sensitive information individuals may mis-report their true situation, leading (here) to an excess amount of zero observations in the context of questions relating to activities such as illicit drug consumption: individuals are likely to deny their participation due to a variety of reasons, such as fear of being caught, stigma and/or moral concerns.

The modelling framework proposed in this paper is based on a two stage decision-making process whereby an individual firstly decides whether to participate in the activity in question, and then decides whether or not to mis-report their behaviour. Furthermore, given that we apply this framework to survey data relating to the consumption of several illicit drugs, as decisions regarding these are likely to be taken inter-dependently, we expand the modelling approach to a multivariate framework whereby all of the participation and mis-reporting decisions are modelled jointly. indeed, the estimated correlation coefficients across these equations do suggest the existence of complex interrelationships in illicit drug behaviour.

Overall, we find that mis-reporting has a significant effect on observed participation rates such that, across all three drugs, the predicted marginal probabilities of participation are substantially higher than that in the sample rate of participation as indicated by the raw survey data. This is caused by some quite high propensities to mis-report. Interestingly, our findings suggest that the extent of mis-reporting is influenced by how the survey was administered as well as factors such as the presence of other individuals when the survey was completed. Such findings suggest that the conditions under which survey data is collected serve to influence the accuracy of the information obtained. Our findings suggest that accounting for mis-reporting is important in the context of using survey data related to sensitive activities, especially where such data is used to inform public policy.

References

ABS (2010): Consumer Price Index 14th Series: By Region by Group, Sub-Group and Expenditure Class - Alcohol and Tobacco, Cat. No. 6455.0.40.001, Australian Bureau of Statistics.

ACC (2010): Australian Illicit Drug Report, Australian Crime Commission, Commonwealth of Australia, Canberra.

Becker, G., and K. Murphy (1988): "A Theory of Rational Addiction," *Journal of Political Economy*, 96(4), 675–700.

Berg, N., and D. Lien (2006): "Same-Sex Sexual Behaviour: US Frequency Estimates from Survey Data with Simultaneous Misreporting and Non-Response," *Applied Economics*, 38(7), 757–769.

Borghans, L., B. Golsteyn, J. Heckman, and H. Meijers (2009): "Gender Differences in Risk Aversion and Ambiguity Aversion," *Journal of the European Economic Association*, 7(2-3), 649–658.

Bowling, A. (2005): "Mode of Questionnaire Administration can have Serious Effects on Data Quality," *Journal of Public Health*, 27(3), 281–291.

Cameron, L., and J. Williams (2001): "Cannabis, Alcohol and Cigarettes: Substitutes or Compliments," *The Economic Record*, 77(236), 19–34.

Caudill, S., and F. Mixon (2005): "Analysing Misleading Discrete Responses: A Logit Model Based on Misclassified Data," *Oxford Bulletin of Economics and Statistics*, 67(1), 105–113.

Collins, R., P. Ellickson, and R. Bell (1998): "Simultaneous Polydrug Use Among Teens: Prevalence and Predictors," *Journal of Substance Abuse*, 10(3), 233–253.

Corbin, J., and J. Morse (2003): "The Unstructured Interactive Interview: Issues of Reciprocity and Risks when Dealing with Sensitive Topics," *Qualitative Enquiry*, 9(3), 335–354.

Cragg, J. (1971): "Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods," *Econometrica*, 39, 829–44.

Delaney, L., C. Harmon, and P. Wall (2008): "Behavioral Economics and Drinking Behavior: Preliminary Results from an Irish College Study," *Economic Inquiry*, 46(1), 29–36.

Deschesnes, M., P. Finès, and S. Demers (2006): "Are Tattooing and Body Piercing Indicators of Risk-Taking Behaviours Among High School Students?," *Journal of Adolescence*, 29(3), 379–393.

DeSimone, J., and M. Farrelly (2003): "Price and Enforcement Effects on Cocaine and Marijuana Demand," *Economic Inquiry*, 41(1), 98–115.

Duarte, R., J. Escario, and J. Molina (2005): "Participation and Consumption of Illegal Drugs among Adolescents," *International Advances in Economic Research*, 11(4), 399–415.

Dustmann, C., and A. Soest (2001): "Language Fluency and Earnings: Estimation with Misclassified Language Indicators," *Review of Economics and Statistics*, 83(4), 663–674.

Farrelly, M., J. Bray, G. Zarkin, and B. Wendling (2001): "The joint demand for cigarettes and marijuana: Evidence from the National Household Surveys on Drug Abuse," *Journal of Health Economics*, 20(1), 51–68.

Fendrich, M., and C. Vaughn (1994): "Diminished Lifetime Substance Use Over Time: An Inquiry into Differential Under-reporting," *Public Opinion Quarterly*, 58(1), 96.

Gannon, B. (2009): "The Influence of Economic Incentives on Reported Disability Status," *Health Economics*, 18(7), 743–759.

Gill, A., and R. Michaels (1991): "The Determinants of Illegal Drug Use," *Contemporary Economic Policy*, 9(3), 93–105.

Greene, W. (1994): "Accounting for Excess Zeros and Sample Selection in Poisson and Negative Binomial Regression Models," Working Paper EC-94-10, Stern School of Business, New York University.

Greene, W. (2008): *Econometric Analysis*. Prentice Hall, New Jersey, USA, sixth edn.

Gruber, J. (2001): *Risky Behavior Among Youths: An Economic Analysis*. University of Chicago Press, Chicago, IL.

Hanna, E., H. Yi, M. Dufour, and C. Whitmore (2001): "The Relationship of Early-Onset Regular Smoking to Alcohol Use, Depression, Illicit Drug Use, and Other Risky Behaviors During Early Adolescence: Results from the Youth Supplement to the Third National Health and Nutrition Examination Survey," *Journal of Substance Abuse*, 13(3), 265–282.

Harris, M., and X. Zhao (2007): "A Zero-Inflated Ordered Probit Model, with an Application to Modelling Tobacco Consumption," *Journal of Econometrics*, 141(2), 1073–1099.

HAUSMAN, J., J. ABREVAYA, AND F. SCOTT-MORTON (1998): "Misclassification of the Dependent Variable in a Discrete-Response Setting," *Journal of Econometrics*, 87(2), 239–269.

HEILBRON, D. (1989): "Generalized Linear Models for Altered Zero Probabilities and Overdispersion in Count Data," Discussion paper, University of California, San Francisco.

HOYT, G., AND F. CHALOUPKA (1994): "Effect of Survey Conditions on Self-Reported Substance Use," *Contemporary Economic Policy*, 12(3), 109–121.

IVES, R., AND P. GHELANI (2006): "Polydrug Use (The Use of Drugs in Combination): A Brief Review," *Drugs: Education, Prevention, and Policy*, 13(3), 225–232.

JOFRE-BONET, M., AND N. PETRY (2008): "Trading Apples for Oranges?: Results of an Experiment on the Effects of Heroin and Cocaine Price Changes on Addicts' Polydrug Use," *Journal of Economic Behavior and Organization*, 66(2), 281–311.

KENKEL, D., R. REED III, AND P. WANG (2002): "Rational addiction, Peer Externalities and Long Run Effects of Public Policy," Discussion paper, National Bureau of Economic Research.

KRAUS, L., AND R. AUGUSTIN (2001): "Measuring Alcohol Consumption and Alcohol-Related Problems: Comparison of Responses from Self-Administered Questionnaires and Telephone Interviews," *Addiction*, 96(3), 459–471.

KRISTENSEN, N., AND E. JOHANSSON (2008): "New Evidence on Cross-Country Differences in Job Satisfaction Using Anchoring Vignettes," *Labour Economics*, 15(1), 96–117.

LAMBERT, D. (1992): "Zero Inflated Poisson Regression with an Application to Defects in Manufacturing," *Technometrics*, 34, 1–14.

LU, N., B. TAYLOR, AND K. RILEY (2001): "The Validity of Adult Arrestee Self-Reports of Crack Cocaine Use," *American Journal of Alcohol Abuse*, 27, 399–419.

MAASSEN VAN DEN BRINK, H., AND W. GROOT (1999): "Job Satisfaction and Preference Drift," *Economics Letters*, 63(3), 363–367.

MacDonald, Z., and S. Pudney (2000): "Analysing Drug Abuse with British Crime Survey Data: Modelling and Questionnaire Design Issues," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 49(1), 95–117.

——— (2003): "The Use of Self-Report and Drugs Tests in the Measurement of Illicit Drug Consumption," Discussion Papers in Economics, Department of Economics, University of Leicester.

Mensch, B., and D. Kandel (1988): "Underreporting of Substance Use in a National Longitudinal Youth Cohort," *Public Opinion Quarterly*, 52(1), 100.

Morrison-Beedy, D., M. Carey, and X. Tu (2006): "Accuracy of Audio Computer-Assisted Self-Interviewing (ACASI) and Self-Administered Questionnaires for the Assessment of Sexual Behavior," *AIDS and Behavior*, 10(5), 541–552.

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

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

NDARC (2009): "Australian Drug Trends 2007: Findings from the Illicit Drug Reporting System (IDRS)," Various Reports, National Drug and Alcohol Research Centre, University of New South Wales, Sydney.

NDSHS (2010): "Computer Files for the Unit Record Data from the National Drug Strategy Household Surveys," Social Science Data Archives, Australian National University, Canberra.

O'Muircheartaigh, C., and P. Campanelli (1998): "The Relative Impact of Interviewer Effects and Sample Design Effects on Survey Precision," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 161(1), 63–77.

Pohlmeier, W., and V. Ulrich (1995): "An Econometric Model of the Two-Part Decision-Making Process in the Demand for Health Care," *Journal of Human Resources*, 30, 339–361.

PUDNEY, S. (2004): "Keeping Off the Grass? An Econometric Model of Cannabis Consumption in Britain," *Journal of Applied Econometrics*, 19(4), 435–453.

——— (2007): "Rarely Pure and Never Simple: Extracting the Trust from Self-Reported Data on Substance Use," Cemmap Working Paper 11/07, Institute for Fiscal Studies and Institute for Social and Economic Research.

——— (2010): "Drugs Policy: What Should We Do About Cannabis?," *Economic Policy*, 25(61), 165–211.

RAMFUL, P., AND X. ZHAO (2009): "Participation in Marijuana, Cocaine and Heroin Consumption in Australia: A Multivariate Probit Approach," *Applied Economics*, 41(4), 481–496.

SAFFER, H., AND F. CHALOUPKA (1998): "Demographics Differentials in the Demand for Alcohol and Illicit Drugs," in *Economic Analysis of Substance Use and Abuse: An Integration of Econometric and Behavioral*, pp. 187–211. Chaloupka F., Bickel, W., Saffer, H., Grossman, M. (Eds.), Economic Research, University of Chicago Press, Chicago.

——— (1999): "The Demand for Illicit Drugs," *Economic Inquiry*, 37(3), 401–411.

SICKLES, R., AND P. TAUBMAN (1991): "Who Uses Illegal Drugs?," *The American Economic Review*, 81(2), 248–251.

SMITH, M. (2003): "On Dependency in Double-Hurdle Models," *Statistical Papers*, 44, 581–595.

STEPHENS, M., ET AL. (2003): "Behavioral Risks Associated with Tattooing," *Family Medicine*, 35(1), 52–54.

WILLIAMS, J. (2003): "The Effects of Price and Policy on Marijuana Use: What Can Be Learned from the Australian Experience?," *Health Economics*, 13(2), 123–137.

WISH, E. (1987): *Drug Use Forecasting: New York 1984 to 1986*. US Department of Justice, National Institute of Justice Washington, DC.

Zhao, X., and M. Harris (2003): "Demand for Marijuana, Alcohol and Tobacco: Participation, Frequency and Cross-Equation Correlations," paper presented at the Australasian Meeting of the Econometrics Society, July, Sydney.

——— (2004): "Demand for Marijuana, Alcohol and Tobacco: Participation, Frequency and Cross-Equation Correlations," *Economic Record*, 80(251), 394–410.

Appendix: Definition of Variables

- **Stage**: standardised age (mean subtracted and scaled by sample standard deviation).
- **Stagesq**: age-squared, standardised (mean subtracted and scaled by sample standard deviation).
- **Male:** = 1 for male; and = 0 for female.
- **Married:** = 1 if married or *de facto*; and = 0 otherwise.
- **Preschool:** = 1 if the respondent has pre-school aged child/children, and = 0 otherwise.
- **Singpar:** 1 if respondent comes from a single parent household, and = 0 otherwise.
- **Capital**: = 1 if the respondent resides in a capital city, and = 0 otherwise.
- **ATSI:** = 1 if respondent is of Aboriginal or Torres Strait Islander origin, and = 0 otherwise.
- **Work:** = 1 if mainly employed; and = 0 otherwise.
- **Study:** = 1 if mainly study; and = 0 otherwise.
- **Unemp** = 1 if unemployed; and = 0 otherwise.
- **Other** = 1 if retired, home duty, or volunteer work; and = 0 otherwise. This variable is used as the base of comparison for work status dummies and is dropped in the estimation.
- **Degree**: = 1 if the highest qualification is a tertiary degree, and = 0 otherwise.
- **Diploma:** = 1 if the highest qualification is a non-tertiary diploma or trade certificate, and = 0 otherwise.
- **Yr12**: = 1 if the highest qualification is Year 12, and = 0 otherwise.
- **Less than Year 12**: = 1 if the highest qualification is below Year 12, and = 0 otherwise. This variable is used as the base of comparison for education dummies and is dropped in the estimation.

- **Lrpinc:** Logarithm of real personal annual income before tax measured in thousands of Australian dollars.
- **Decrim:** = 1 if respondent resides in a state where small possession is decriminalised and = 0 otherwise.
- **Migr10:** = 1 if migrated to Australia in the last 10 years, and = 0 otherwise.
- **Tattoo:** = 1 if undergone any tattoo procedure, and = 0 otherwise.
- **Piercing:** = 1 if undergone any body piercing procedure, and = 0 otherwise.
- **Peer**: = 1 most or all of respondent's friends and acquaintances use marijuana, speed or cocaine.
- **Lrpmar:** Logarithm of real price for marijuana measured in dollars per ounce.
- **Lrpspd:** Logarithm of real price of speed measured in dollars per gram.
- **Lrpcoc:** Logarithm of real price of cocaine measured in dollars per gram.
- **Lrpher:** Logarithm of real price of heroin measured in dollars per gram.
- **Lrptob:** Logarithm of real price index for tobacco.
- **Lrpalc:** Logarithm of real price index for alcohol.
- **Present**: = 1 if anyone else was present when the respondent was completing the survey questionnaire; and = 0 otherwise.
- **Help**: = 1 if anyone helped the respondent complete the survey questionnaire; and = 0 otherwise.
- **Survtype**: = 1 if the computer-assisted telephone interview (CATI) or face-to-face method was used to collect data; and = 0 if drop and collect method was used.
- **Trust**: percentage of compulsory questions left unanswered in the survey.

Table 1: Summary Statistics, Sample Size 50,345

Variable	Mean	Std Dev	Minimum	Maximum
MALE	0.4733	0.4993	0	1
STAGE	-0.0374	0.9314	-1.7157	2.9028
STAGESQ	-0.06	0.928	-1.2437	4.1373
MARRIED	0.5951	0.4909	0	1
PRESCHOOL	0.123	0.3285	0	1
SINGPAR	0.0701	0.2553	0	1
CAPITAL	0.6467	0.478	0	1
ATSI	0.0128	0.1125	0	1
WORK	0.6326	0.4821	0	1
STUDY	0.0634	0.2436	0	1
UNEMP	0.0221	0.1469	0	1
DEGREE	0.2712	0.4446	0	1
DIPLOMA	0.3486	0.4765	0	1
YR12	0.1306	0.337	0	1
LRPINC	9.7955	0.9338	6.64	11.2708
DECRIM	0.2561	0.4365	0	1
MIGR10	0.0442	0.2055	0	1
TATTOO	0.1081	0.3105	0	1
PIERCING	0.0827	0.2755	0	1
PEER	0.0483	0.2145	0	1
LRPMAR	5.2385	0.1557	4.8091	5.4743
LRPCOC	5.1819	0.2241	4.8175	5.8237
LRPHER	5.5266	0.3351	4.8314	6.3479
LRPSPD	4.6599	0.4766	3.5144	5.3459
LRPTOB	5.5593	0.0511	5.4505	5.646
LRPALC	4.7116	0.0362	4.63	4.7657
YR04	0.3506	0.4772	0	1
YR07	0.3183	0.4658	0	1
PRESENT	0.289	0.4533	0	1
HELP	0.2293	0.4204	0	1
SURVTYPE	0.1802	0.3843	0	1
TRUST	0.0359	0.0532	0	0.6371

Table 2: Marijuana, Speed and Cocaine Consumption: Estimated Coefficients (Common Variables)[a]

	Marijuana		Speed		Cocaine	
	Participation	Mis-reporting	Participation	Mis-reporting	Participation	Mis-reporting
CONSTANT	-2.652	2.195	-3.144	4.370	-5.959	-4.476
	(3.637)	(0.540)***	(5.865)	(9.637)	(9.846)	(1.808)**
MALE	0.259	0.518	0.302	-0.323	0.259	-0.135
	(0.036)***	(0.075)***	(0.035)***	(0.156)**	(0.059)***	(0.212)
STAGE	0.705	0.484	-0.399	6.765	2.869	-3.208
	(0.122)***	(0.440)	(0.186)**	(6.885)	(0.484)***	(2.080)
STAGESQ	-1.548	-0.404	-0.489	-5.377	-4.473	3.550
	(0.146)***	(0.645)	(0.224)**	(16.862)	(0.652)***	(3.016)
MARRIED	-0.427	0.142	-0.356	0.860	-0.441	0.398
	(0.031)***	(0.093)	(0.036)***	(0.635)	(0.068)***	(0.292)
PRESCHOOL	-0.133	-0.127	-0.271	0.594	-0.296	0.284
	(0.042)***	(0.108)	(0.047)***	(0.527)	(0.094)***	(0.453)
SINGPAR	0.041	-0.105	-0.047	0.014	-0.039	-0.399
	(0.051)	(0.103)	(0.053)	(0.223)	(0.136)	(0.339)
CAPITAL	-0.023	0.152	0.138	0.212	0.320	0.316
	(0.029)	(0.072)**	(0.034)***	(0.168)	(0.076)***	(0.282)
ATSI	0.315	-0.480	-0.296	0.164	-0.608	0.995
	(0.123)**	(0.156)***	(0.118)**	(0.523)	(0.286)**	(1.938)
MIGR10	-0.235	0.054	-0.064	-1.323	0.273	-0.614
	(0.062)***	(0.193)	(0.083)	(0.313)**	(0.128)**	(0.344)*
WORK	-0.070	0.043	-0.081	-0.233	0.127	0.099
	(0.046)	(0.120)	(0.059)	(0.534)	(0.157)	(0.457)
STUDY	-0.069	0.091	0.127	-0.666	0.248	0.262
	(0.062)	(0.149)	(0.083)	(0.541)	(0.178)	(0.573)
UNEMP	0.206	0.251	0.031	0.138	1.123	-1.528
	(0.073)***	(0.172)	(0.094)	(0.702)	(0.410)***	(0.551)***
DEGREE	0.193	-0.446	-0.173	0.083	0.195	0.202
	(0.045)***	(0.112)***	(0.051)***	(0.340)	(0.120)	(0.364)
DIPLOMA	0.120	-0.239	-0.084	0.108	-0.047	0.682
	(0.036)***	(0.093)***	(0.044)*	(0.197)	(0.119)	(0.349)*
YR12	0.113	-0.146	-0.056	0.020	0.048	0.538
	(0.042)***	(0.109)	(0.054)	(0.210)	(0.124)	(0.394)
LRPINC	0.052	-0.111	0.129	-0.014	0.121	0.484
	(0.020)**	(0.055)***	(0.028)***	(0.127)	(0.057)**	(0.162)***
DECRIM	0.137	-0.310	-0.006	-0.082	-0.008	-0.760
	(0.038)***	(0.073)***	(0.044)	(0.162)	(0.096)	(0.268)***

Standard errors are given in parentheses. *Significant at 10% level; **Significant at 5% level. [a] A positive coefficient for participation indicates an increase in participation probability while a negative coefficient for mis-reporting indicates an increase in mis-reporting probability.

Table 3: Marijuana, Speed and Cocaine Consumption: Estimated Coefficients (Idenitfying Variables)[a]

	Marijuana		Speed		Cocaine	
	Participation	Mis-reporting	Participation	Mis-reporting	Participation	Mis-reporting
TATTOO	0.408	-	0.365	-	0.237	-
	(0.029)***		(0.034)***		(0.054)***	
PIERCING	0.539	-	0.453	-	0.435	-
	(0.034)***		(0.036)***		(0.056)***	
PEER	1.911	-	1.182	-	1.159	-
	(0.082)***		(0.040)***		(0.075)***	
YR04	0.036	-	0.233	-	-0.060	-
	(0.039)		(0.057)***		(0.094)	
YR07	-0.049	-	0.200	-	0.271	-
	(0.059)		(0.089)**		(0.143)*	
LRPMAR	-0.024	-	0.046	-	0.095	-
	(0.083)		(0.124)		(0.225)	
LRPCOC	0.099	-	-0.097	-	-0.092	-
	(0.060)*		(0.091)		(0.149)	
LRPSPD	-0.022	-	-0.124	-	0.014	-
	(0.037)		(0.053)**		(0.096)	
LRPTOB	-0.483	-	-1.536	-	-1.149	-
	(0.383)		(0.614)**		(1.045)	
LRPALC	0.473	-	1.459	-	1.833	-
	(0.360)		(0.548)***		(0.854)**	
LRPHER	0.140	-	0.289	-	-0.245	-
	(0.063)**		(0.092)***		(0.146)*	
PRESENT	-	-0.212	-	-0.552	-	-0.586
		(0.068)***		(0.185)***		(0.248)**
HELP	-	-0.222	-	0.096	-	-0.209
		(0.088)**		(0.211)		(0.272)
SURVTYPE	-	-0.344	-	-0.326	-	-0.888
		(0.100)***		(0.225)		(0.336)***
TRUST	-	-3.068	-	-1.422	-	-5.280
		(0.583)***		(1.706)		(2.068)**

Standard errors are given in parentheses. *Significant at 10% level; **Significant at 5% level. [a] A positive coefficient for participation indicates an increase in participation probability while a negative coefficient for mis-reporting indicates an increase in mis-reporting probability.

Table 4: Correlation Coefficients

	M_{mar}	R_{mar}	M_{spd}	R_{spd}	M_{coc}	R_{coc}
M_{mar}	-					
R_{mar}	0.026	-				
	(0.144)					
M_{spd}	0.529	0.224	-			
	(0.143)***	(0.209)				
R_{spd}	0.198	0.639	0.098	-		
	(0.053)***	(0.026)***	(0.185)			
M_{coc}	0.339	0.275	0.368	0.479	-	
	(0.211)	(0.253)	(0.459)	(0.148)***		
R_{coc}	0.144	0.598	0.265	0.644	0.313	-
	(0.087)*	(0.047)***	(0.173)	(0.032)***	(0.259)	

Standard errors are given in parentheses. *Significant at 10% level; **Significant at 5% level.

Table 5: Sample and Predicted Probabilities

	Marijuana	Speed	Cocaine
Sample Rate of Participation	0.1183	0.0308	0.0132
Marginal Probability of Participation	0.1668	0.0357	0.0226
	(0.0080)***	(0.0017)***	(0.0039)***
Probability of Mis-reporting Conditional on Participation	0.1780	0.0461	0.3552
	(0.0651)***	(0.0096)***	(0.0039)***
Components of the zeros:			
Non-participation	0.8332	0.9643	0.9774
	(0.0080)***	(0.0017)***	(0.0039)***
Mis-reporting	0.0335	0.0073	0.0080
	(0.0078)***	(0.0016)***	(0.0038)***
Total	0.8667	0.9716	0.9854
	(0.0014)***	(0.0007)***	(0.0006)***
Posterior Probabilities:			
Non-Participation	0.7719	0.9473	0.9506
	(0.0259)***	(0.0121)***	(0.0217)***
Mis-reporting	0.2281	0.0527	0.0494
	(0.0259)***	(0.0121)***	(0.0217)**

Standard errors are given in parentheses. *Significant at 10% level; **Significant at 5% level.

Table 6: Partial Effects on Selected Joint Probabilities[a]

	$Pr(y_{mar}=0, y_{spd}=0, y_{coc}=0\|\mathbf{x})$			$Pr(y_{mar}=1, y_{spd}=1, y_{coc}=1\|\mathbf{x})^a$		
	Participation	Mis-reporting	Overall	Participation	Mis-reporting	Overall
MALE	-0.035	-0.012	-0.047	0.176	0.007	0.183
	(0.004)***	(0.003)***	(0.003)***	(0.000)***	(0.000)	(0.000)***
STAGE	-0.090	-0.011	-0.101	1.119	-0.028	1.091
	(0.015)***	(0.012)	(0.065)	(0.000)***	(0.000)	(0.063)
STAGESQ	0.203	0.009	0.212	-1.983	0.033	-1.950
	(0.031)***	(0.016)	(0.026)***	(0.022)	(0.000)	(0.022)
MARRIED	0.057	-0.003	0.053	-0.266	0.007	-0.259
	(0.021)***	(0.004)	(0.021)**	(0.020)	(0.004)	(0.020)
PRESCHOOL	0.018	0.003	0.021	-0.181	0.001	-0.180
	(0.005)***	(0.003)	(0.004)***	(0.000)**	(0.000)	(0.000)**
SINGPAR	-0.005	0.002	-0.003	-0.024	-0.006	-0.031
	(0.007)	(0.003)	(0.006)	(0.000)	(0.000)	(0.000)
CAPITAL	0.002	-0.003	-0.001	0.159	0.006	0.165
	(0.004)	(0.004)	(0.003)	(0.002)	(0.004)	(0.000)***
ATSI	-0.039	0.011	-0.028	-0.301	0.003	-0.298
	(0.018)**	(0.007)	(0.016)*	(0.000)*	(0.000)	(0.000)*
MIGR10	0.030	-0.001	0.029	0.092	-0.006	0.086
	(0.008)***	(0.005)	(0.007)***	(0.000)	(0.000)	(0.000)
WORK	0.009	-0.001	0.008	0.033	0.002	0.035
	(0.006)	(0.003)	(0.005)*	(0.000)	(0.000)	(0.000)
STUDY	0.008	-0.002	0.006	0.126	0.004	0.130
	(0.008)	(0.003)	(0.007)	(0.000)*	(0.000)	(0.000)*
UNEMP	-0.027	-0.006	-0.033	0.474	-0.013	0.461
	(0.010)***	(0.005)	(0.009)***	(0.000)**	(0.000)	(0.000)**
DEGREE	-0.024	0.010	-0.014	0.051	-0.005	0.045
	(0.007)***	(0.003)***	(0.006)**	(0.000)	(0.000)	(0.000)
YR12	-0.014	0.003	-0.011	0.012	0.004	0.015
	(0.006)**	(0.003)	(0.005)**	(0.000)	(0.000)	(0.000)
DIPLOMA	-0.015	0.005	-0.010	-0.033	0.004	-0.029
	(0.005)***	(0.002)**	(0.008)	(0.000)	(0.000)	(0.007)
LRPINC	-0.007	0.003	-0.005	0.078	0.004	0.081
	(0.003)**	(0.002)	(0.002)**	(0.001)	(0.002)	(0.000)
DECRIM	-0.018	0.007	-0.011	-0.001	-0.014	-0.014
	(0.005)***	(0.002)***	(0.005)**	(0.000)	(0.000)	(0.000)

[a] Multiplied by 1000 for presentation. Standard errors are given in parentheses. *Significant at 10% level; **Significant at 5% level. [a] A positive marginal effect for participation represents an increase in participation probability while a negative marginal effect for mis-reporting represents an increase in mis-reporting probability.

Table 7: Partial Effects on Selected Joint Probabilities (contd)[a]

	$Pr(y_{mar}=0, y_{spd}=0, y_{coc}=0\|\mathbf{x})$			$Pr(y_{mar}=1, y_{spd}=1, y_{coc}=1\|\mathbf{x})^a$		
	Participation	Mis-reporting	Overall	Participation	Mis-reporting	Overall
TATTOO	-0.054	-	-0.054	0.183	-	0.183
	(0.005)***		(0.005)***	(0.000)***		(0.000)***
PIERCING	-0.072	-	-0.072	0.287	-	0.287
	(0.006)***		(0.006)***	(0.000)***		(0.000)***
PEER	-0.252	-	-0.252	0.772	-	0.772
	(0.018)***		(0.018)***	(0.000)***		(0.000)***
YR04	-0.006	-	-0.006	0.024	-	0.024
	(0.005)		(0.005)	(0.000)		(0.000)
YR07	0.005	-	0.005	0.151	-	0.151
	(0.008)		(0.008)	(0.000)*		(0.000)*
LRPMAR	0.003	-	0.003	0.047	-	0.047
	(0.011)		(0.011)	(0.001)		(0.001)
LRPCOC	-0.012	-	-0.012	-0.055	-	-0.055
	(0.009)		(0.009)	(0.003)		(0.003)
LRPSPD	0.003	-	0.003	-0.020	-	-0.020
	(0.006)		(0.006)	(0.003)		(0.003)
LRPTOB	0.069	-	0.069	-0.798	-	-0.798
	(0.051)		(0.051)	(0.002)		(0.002)
LRPALC	-0.068	-	-0.068	1.063	-	1.063
	(0.049)		(0.049)	(0.003)		(0.003)
LRPHER	-0.019	-	-0.019	-0.038	-	-0.038
	(0.008)**		(0.008)**	(0.002)		(0.002)
PRESENT	-	0.005	0.005	-	-0.010	-0.010
		(0.002)***	(0.002)***		(0.000)	(0.000)
HELP	-	0.005	0.005	-	-0.006	-0.006
		(0.002)**	(0.002)**		(0.000)	(0.000)
SURVTYPE	-	0.008	0.008	-	-0.016	-0.016
		(0.003)***	(0.003)***		(0.000)	(0.000)
TRUST	-	0.070	0.070	-	-0.110	-0.110
		(0.018)***	(0.018)***		(0.000)	(0.000)

[a] Multiplied by 1000 for presentation. Standard errors are given in parentheses. *Significant at 10% level; **Significant at 5% level. [a] A positive marginal effect for participation represents an increase in participation probability while a negative marginal effect for mis-reporting represents an increase in mis-reporting probability.

Table 8: Partial Effects on Selected Conditional Probability [a]

	$Pr(y_{spd}=0, y_{coc}=0 \mid r_{mar}=1, \mathbf{x})$						
	Part'ng	Mis-rep'ng	Overall		Part'ng	Mis-rep'ng	Overall
MALE	-0.0186	0.0003	-0.0183	PIERCING	-0.0207	0.0003	-0.0203
	(0.004)	(0.001)	(0.004)		(0.004)	(0.001)	(0.004)
STAGE	0.0407	0.0060	0.0468	PEER	-0.0278	0.0006	-0.0271
	(0.029)	(0.010)	(0.027)		(0.007)	(0.001)	(0.007)
STAGESQ	0.0199	-0.0080	0.0119	YR04	-0.0198	-0.0001	-0.0199
	(0.032)	(0.013)	(0.028)		(0.006)	(0.000)	(0.006)
MARRIED	0.0172	-0.0007	0.0165	YR07	-0.0250	0.0004	-0.0247
	(0.004)***	(0.001)	(0.004)***		(0.009)	(0.001)	(0.009)
PRESCHOOL	0.0227	-0.0005	0.0222	LRPMAR	-0.0068	0.0001	-0.0067
	(0.006)***	(0.001)	(0.006)***		(0.013)	(0.000)	(0.013)
SINGPAR	0.0072	0.0002	0.0074	LRPCOC	0.0156	-0.0002	0.0154
	(0.006)	(0.000)	(0.006)		(0.010)	(0.000)	(0.010)
CAPITAL	-0.0183	0.0002	-0.0181	LRPSPD	0.0107	0.0000	0.0107
	(0.004)	(0.001)	(0.004)		(0.006)*	(0.000)	(0.006)*
ATSI	0.0522	-0.0016	0.0506	LRPTOB	0.1364	-0.0012	0.1351
	(0.020)***	(0.003)	(0.020)*		(0.064)**	(0.003)	(0.064)**
MIGR10	-0.0094	0.0009	-0.0085	LRPALC	-0.1377	0.0022	-0.1354
	(0.010)	(0.002)	(0.009)		(0.059)	(0.004)	(0.058)
WORK	0.0027	0.0002	0.0028	LRPHER	-0.0176	-0.0005	-0.0180
	(0.007)	(0.001)	(0.006)		(0.010)	(0.001)	(0.010)
STUDY	-0.0188	0.0002	-0.0186	PRESENT		0.0004	0.0004
	(0.009)	(0.001)	(0.009)			(0.001)	(0.001)
UNEMP	-0.0057	0.0025	-0.0032	HELP		0.0001	0.0001
	(0.022)	(0.005)	(0.022)			(0.000)	(0.000)
DEGREE	0.0244	0.0001	0.0245	SURVTYPE		0.0006	0.0006
	(0.006)***	(0.000)	(0.006)***			(0.001)	(0.001)
YR12	0.0107	-0.0003	0.0104	TRUST		0.0035	0.0035
	(0.006)*	(0.001)	(0.006)*			(0.005)	(0.006)
DIPLOMA	0.0149	-0.0006	0.0143				
	(0.005)***	(0.001)	(0.005)***				
LRPINC	-0.0112	-0.0002	-0.0114				
	(0.003)	(0.000)	(0.003)				
DECRIM	0.0079	0.0004	0.0083				
	(0.005)	(0.001)	(0.005)*				

Standard errors are given in parentheses. *Significant at 10% level; **Significant at 5% level. [a] A positive marginal effect for participation represents an increase in participation probability while a negative marginal effect for mis-reporting represents an increase in mis-reporting probability.

The Bankwest Curtin Economics Centre is an independent economic and social research organisation located within the Curtin Business School at Curtin University. The Centre was established in 2012 through the generous support from Bankwest (a division of the Commonwealth Bank of Australia), with a core mission to undertake high quality, objective research on the key economic and social issues of relevance to Western Australia.

The Centre's research and engagement activities are designed to influence economic and social policy debates in state and Federal Parliament, regional and national media, and the wider Australian community. Through high quality, evidence-based research and analysis, our research outcomes inform policy makers and commentators of the economic challenges to achieving sustainable and equitable growth and prosperity both in Western Australia and nationally.

The Centre capitalises on Curtin University's reputation for excellence in economic modelling, forecasting, public policy research, trade and industrial economics and spatial sciences. Centre researchers have specific expertise in economic forecasting, quantitative modelling, microdata analysis and economic and social policy evaluation.

A suite of tailored and national economic models and methods are maintained within the Centre to facilitate advanced economic policy analysis: these include macroeconomic and time series models, micro(simulation) models, computable general equilibrium (CGE) models, spatial modelling methods, economic index analysis, and behavioural modelling methods.

CONTACT

Bankwest Curtin Economics Centre
Curtin University
Kent Street Bentley WA 6102
GPO Box U1987 Perth WA 6845

Tel: +61 8 9266 2873

bankwesteconomicscentre@curtin.edu.au
business.curtin.edu.au/bcec