Average partial effects in multivariate probit models with an application to immigrants' ethnic identity and economic performance

G. S. F. Bruno*and O. Dessy†

October 29, 2014

Abstract

We extend the univariate results in [Wooldridge, J. M. (2005): "Unobserved heterogeneity and estimation of average partial effects," in Identification And Inference For Econometric Models: Essays In Honor Of Thomas Rothenberg, ed. by D. W. K. Andrews, and J. H. Stock. Cambridge University Press, New York] to multivariate probit models, proving the following. 1) Average partial effects (APEs) based on joint or marginal response probabilities are identified in multivariate probit models with general conditionally independent latent heterogeneity (LH), provided that the error covariance matrix in the structural model is unconstrained beyond normalization. If this caveat is not met, identification requires that the covariance matrix for the LH components be restricted. This finding is substantial since in most coded routines for multivariate probit models it is not possible to adjust the form of the covariance matrix to the parametric structure of the latent regression model. Stata's `biprobit`, `mvprobit`, `mprobit` and `cmp` or Limdep's `BIVARIATE PROBIT` and `MPROBIT` are cases in point. 2) Conditionally independent LH does not assure identification of APEs based on conditional response probabilities, unless additional independence restrictions are maintained. 3) The dimensionality benefit observed by [Mullahy, J. (2011): "Marginal effects in multivariate probit and kindred discrete and count outcome models, with applications in

*Bocconi University, Milan
†Cà Foscari University, Venice and Catholic University of Milan

health economics," NBER WP SERIES 17588, NBER] in the estimation of partial effects extends to APEs. We exploit this feature to design a simple procedure for estimating APEs, which is both faster and more accurate than simulation-based codes, such as Stata's `mvprobit` and `cmp`. To demonstrate the finite-sample implications of our results, we carry out extensive Monte Carlo experiments with bivariate and trivariate probit models. Finally, we apply our procedure in (3) to Italian survey data of immigrants in order to estimate the APEs of a trivariate probit model of ethnic identity formation and economic performance.

September 2014

JEL classification: J24, C41, C33, C35, J6.

Keywords: latent heterogeneity, multivariate probit, multinomial probit, panel probit, recursive probit, response probability, normalization

1 Introduction

How much robust are estimators of average partial effects (APEs) in conventional multivariate probit models to latent heterogeneity (LH)? Wooldridge (2005) shows that univariate probit estimators of APEs are robust to LH if the latter is independent of the regressors, given a set a control variables (see also Wooldridge 2010). We find that in many interesting multivariate probit models the assumption of conditionally independent LH needs being strengthen in order to identify APEs.

Specifically, we find the following.

1. Identification of APEs based on joint or marginal probabilities in conventional probit models with conditionally independent LH requires a structural-error covariance matrix that is unconstrained beyond normalization, that is one with a minimal set of normalization restrictions given the other constraints in the model, if any. For example with cross-equation equality restrictions, as in the *multinomial probit model with alternative-specific characteristics* (implemented by Stata's `asmprobit`) or in the *panel probit model* examined by Bertschek and Lechner (1998) and Greene (2004), identification requires a structural-error covariance matrix that is arbitrary up to a fixed element. Should the latter be in correlation form, and as such constrained beyond normalization, identification would break down unless the LH components are truly homoskedastic. This finding is substantial since in most statistical packages there is no way to obtain a more general covariance matrix than that in correlation form. Stata's `biprobit`, `mvprobit` (Cappellari and Jenkins 2003) and `cmp` (Roodman 2011) or Limdep's `BIVARIATE PROBIT` and `MPROBIT` are cases in point. Similarly, the covariance matrix used by the *multinomial probit model with i.i.d. errors*, implemented by Stata's `mprobit`, is constrained beyond normalization and as such supports only a limited class of conditionally independent LH components. Stata's `asmprobit` ensures more flexibility in this respect, but restricts to a specific probit model.

2. Conditional independence of the LH components is not sufficient for identification of APEs based on conditional probabilities. Adding independence of the LH components in the conditioning subvector from all the observables in the model does not help either. Consistency can be proved only if we expand the latter set to include the remaining latent components. Fortunately, within either

sub-vector of LH components the covariance matrix can be arbitrary. We carry out extensive Monte Carlo experiments that reveal that the bias in the conditional probability estimates is severe when the additional independence assumption is violated. Importantly, this holds true regardless of the model being a conventional one, i. e. with only exogenous regressors, or recursive, as in Maddala (1983), Wooldridge (2010),Roodman (2011) and Greene (2012). We show that the restricted framework is still compatible with applications of the control function approach by Rivers and Vuong (1988) (see also Wooldridge 2010, pp. 585-594) to multivariate probit models with both binary and continuous endogenous explanatory variables (Wooldridge 2010, pp. 594-599).

The second contribution of this paper is computational. We show that the dimensionality benefit observed by Mullahy (2011) in the estimation of partial effects (PEs) of joint probabilities extends to APEs with conditionally independent heterogeneity. We exploit this feature in the construction of a simple estimation routine that, based on the marginalization property of the normal distribution, estimates coefficients and covariances from a combination of $m(m-1)/2$ bivariate probit models and then evaluates the APEs averaging Mullahy's formulas over the sample. In trivariate models this routine completely dispenses with numerical evaluation of cumulative normal distributions by time-consuming simulation methods, as otherwise required by `mvprobit` (Cappellari and Jenkins 2003) and `cmp` (Roodman 2011), two popular Stata commands implementing multivariate probit models and both based on the GHK simulator (see Geweke 1989, Hajivassiliou 1990, Keane 1994). Our routine is extremely simple, as it can be executed through existing commands in Stata, but is also suboptimal in large samples. A part of our Monte Carlo section evaluates the finite-sample performances of the routine. Not only is it from 7 to 14 times faster than both `mvprobit` and `cmp`, but also it proves much more accurate than the two GHK-based alternatives, both in terms of bias and root mean squared error and even when the number of draws required by the GHK simulator is increased beyond defaults.

To demonstrate our routine on real data we focus on an issue that has been attracting increasing attention among researchers and policy makers, that of ethnic identity formation among immigrants in western countries and its interaction with immigrants economic performance. Robust APEs of joint and conditional probabilities are estimated by applying multivariate probit models to the ISMU data base, a unique survey of immigrants in Italy.

1.1 Notation and conventions

Throughout, Φ_Σ indicates the zero-mean multivariate Normal distribution function with covariance matrix Σ; when no subscript is specified we mean the standard Normal distribution and ϕ indicates the standard Normal density function. A generic covariance matrix, Σ, is expressed in correlation form if

$$\Sigma \equiv \begin{pmatrix} 1 & \sigma_{12} & \cdots & \sigma_{1m} \\ \sigma_{12} & 1 & & \vdots \\ \vdots & & \ddots & \\ \sigma_{1m} & \cdots & & 1 \end{pmatrix}.$$

In this case, we will use sometimes the compact notation $\Sigma \equiv C(\sigma_{ij})$, with σ_{ij} indicating the covariance parameter, $i \neq j$.

A covariance matrix is said *unconstrained beyond normalization* (UBN) if it presents a minimal set of normalization-for-scale restrictions given the other constraints in the model. In other words, covariance restrictions are maintained only if the restrictions on the model coefficients, if any, are not sufficient to normalize for scale.

1.2 Paper structure

The next section derives a general non-parametric identification result for APEs based on joint response probability in multivariate models. Section 3 contains the main results for joint and marginal response probabilities in probit models with and without constraints beyond normalization. Section 4 focuses on conditional probabilities. In Section 5 we describe our estimation routine. The Monte Carlo analysis is carried out in Section 6. Section 7 contains the empirical application. Section 8 concludes.

2 Average partial effects in multivariate models

Let the indicator function $\mathbf{1}(A)$ be unity if event A occurs and zero if not. Then, define the random vector $\mathbf{y} = (y_1\ y_2...\ y_m)$ such that

$$y_i = \sum_{c=1}^{g} (c-1)\,\mathbf{1}\left(\lambda_{i,c-1} < y_i^* \leq \lambda_{i,c}\right), \tag{1}$$

$i = 1, ..., m$, where $-\infty \equiv \lambda_{i,0} < \lambda_{i,1} \equiv 0 < \lambda_{i,2}... < \lambda_{i,g-1} < \lambda_{i,g} \equiv +\infty$ are constant thresholds, $g \geq 2$ denotes the number of outcomes and y_i^* is a latent continuous random variable. Hence, $\mathbf{y}$ can be equal to any of the g^m possible $m \times 1$ vectors $\mathbf{k} = (k_1\ ...\ k_m)$, where $k_i \in \{0, 1, ..., g-1\}$ and $i = 1, ..., m$. If $g = 2$ and $m = 1$, the model is a standard binomial model (e. g. *probit* or *logit*), whereas if $g > 2$ and $m = 1$, it is a model of ordered outcomes (e. g. ordered probit or logit). If $m > 1$, we have multivariate models.

Now, consider the probability that $\mathbf{y} = \mathbf{k}$ conditional on a vector of observed random variables $\mathbf{x}$, $Pr(\mathbf{y} = \mathbf{k}|\mathbf{x})$. Greene (2012), in the context of the bivariate probit model, observes that $Pr(\mathbf{y} = \mathbf{k}|\mathbf{x})$ is not a conditional expectation function, intending that unlike the univariate case, where $Pr(y = 1|\mathbf{x}) = E(y|\mathbf{x})$, here $Pr(\mathbf{y} = \mathbf{k}|\mathbf{x})$ cannot be reformulated as the conditional expectation of any of the components in $\mathbf{y}$. In a different sense, however, $Pr(\mathbf{y} = \mathbf{k}|\mathbf{x})$ *is a* conditional expectation function since, given two generic random vectors $\mathbf{u} \in R^k$ and $\mathbf{v} \in R^l$ and the set $U \subset R^k$ one can always write

$$Pr(\mathbf{u} \in U|\mathbf{v}) = E_{\mathbf{u}|\mathbf{v}}\left[1(\mathbf{u} \in U)|\mathbf{v}\right]. \tag{2}$$

A more general result that we will use later on is the following (see, for example, Wooldridge 2010, p. 524): given three random vectors $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{z}$

$$Pr(\mathbf{u} \in U|\mathbf{v}) = E_{\mathbf{z}|\mathbf{v}}\left\{E_{\mathbf{u}|\mathbf{v},\mathbf{z}}\left[1(\mathbf{u} \in U)|\mathbf{v},\mathbf{z}\right]\right\}. \tag{3}$$

Equation (2) is our starting point for deriving APEs in multivariate models with conditionally independent LH. [1] We bring LH into the analysis in the form of a latent random vector $\mathbf{q} = (q_1\ q_2...\ q_m)$,

[1] Equation (2) is an immediate implication of univariate representations of multivariate discrete choice models as, for

which may be related to a vector of observed control variables $\mathbf{w}$. We make the following conditional independence assumptions

A.1 $Pr\left(\mathbf{y}=\mathbf{k}|\mathbf{x},\mathbf{w},\mathbf{q}\right)=Pr\left(\mathbf{y}=\mathbf{k}|\mathbf{x},\mathbf{q}\right)$ for any $\mathbf{k}$.

A.2 $D\left(\mathbf{q}|\mathbf{x},\mathbf{w}\right)=D\left(\mathbf{q}|\mathbf{w}\right)$, where $D\left(\cdot|\cdot\right)$ denotes conditional distributions.

Thus, conditioning on $\mathbf{x}$ and $\mathbf{q}$, gives the structural probability expressed in the form of Equation (2)

$$Pr\left(\mathbf{y}=\mathbf{k}|\mathbf{x},\mathbf{q}\right)=E_{\mathbf{y}|\mathbf{x},\mathbf{q}}\left[1\left(\mathbf{y}=\mathbf{k}\right)|\mathbf{x},\mathbf{q}\right]. \tag{4}$$

Now, adapting the derivations in Wooldridge (2005), we have

$$E_{\mathbf{q}}\left[Pr\left(\mathbf{y}=\mathbf{k}|\mathbf{x}^{0},\mathbf{q}\right)\right] \quad = \quad E_{\mathbf{w}}\left\{E_{\mathbf{q}|\mathbf{w}}\left[Pr\left(\mathbf{y}=\mathbf{k}|\mathbf{x}^{0},\mathbf{q}\right)|\mathbf{w}\right]\right\}.$$

That $E_{\mathbf{q}|\mathbf{w}}\left[Pr\left(\mathbf{y}=\mathbf{k}|\mathbf{x}^{0},\mathbf{q}\right)|\mathbf{w}\right]$ is identified follows from

$$\begin{aligned}
E_{\mathbf{q}|\mathbf{w}}\left[Pr\left(\mathbf{y}=\mathbf{k}|\mathbf{x},\mathbf{q}\right)|\mathbf{w}\right] &= E_{\mathbf{q}|\mathbf{x},\mathbf{w}}\left[Pr\left(\mathbf{y}=\mathbf{k}|\mathbf{x},\mathbf{q}\right)|\mathbf{x},\mathbf{w}\right] \\
&= E_{\mathbf{q}|\mathbf{x},\mathbf{w}}\left\{E_{\mathbf{y}|\mathbf{x},\mathbf{q}}\left[1\left(\mathbf{y}=\mathbf{k}\right)|\mathbf{x},\mathbf{q}\right]|\mathbf{x},\mathbf{w}\right\} \\
&= E_{\mathbf{q}|\mathbf{x},\mathbf{w}}\left\{E_{\mathbf{y}|\mathbf{x},\mathbf{w},\mathbf{q}}\left[1\left(\mathbf{y}=\mathbf{k}\right)|\mathbf{x},\mathbf{w},\mathbf{q}\right]|\mathbf{x},\mathbf{w}\right\} \\
&= E_{\mathbf{y}|\mathbf{x},\mathbf{w}}\left[1\left(\mathbf{y}=\mathbf{k}\right)|\mathbf{x},\mathbf{w}\right], \\
&= Pr\left(\mathbf{y}=\mathbf{k}|\mathbf{x},\mathbf{w}\right)
\end{aligned}$$

where the first equality follows from A.2, the second equality from Equation (4), the third equality from A.1, the fourth equality from the the law of iterated expectations and the last using Equation (2). Hence,

$$E_{\mathbf{q}|\mathbf{w}}\left[Pr\left(\mathbf{y}=\mathbf{k}|\mathbf{x}^{0},\mathbf{q}\right)|\mathbf{w}\right] \quad = \quad Pr\left(\mathbf{y}=\mathbf{k}|\mathbf{x}^{0},\mathbf{w}\right),$$

example, in Mullahy (2011).

implying in turn identification of $E_{\mathbf{q}}\left[Pr\left(\mathbf{y}=\mathbf{k}|\mathbf{x}^{0},\mathbf{q}\right)\right]$:

$$E_{\mathbf{q}}\left[Pr\left(\mathbf{y}=\mathbf{k}|\mathbf{x}^{0},\mathbf{q}\right)\right] \quad = \quad E_{\mathbf{w}}\left[Pr\left(\mathbf{y}=\mathbf{k}|\mathbf{x}^{0},\mathbf{w}\right)\right]. \tag{5}$$

The average partial effect of a generic component x of $\mathbf{x}$ on the joint probability $Pr\left(\mathbf{y}=\mathbf{k}|\mathbf{x}^{o}\,\mathbf{q}\right)$ is defined as

$$APE_{x}\left(\mathbf{k},\mathbf{x}^{o}\right) \equiv E_{\mathbf{q}}\left\{\partial_{x}Pr\left(\mathbf{y}=\mathbf{k}|\mathbf{x}^{o}\,\mathbf{q}\right)\right\}$$

and provided that regularity conditions enabling interchange of derivatives and integrals are satisfied, Equation (5) implies identification of APEs:

$$APE_{x}\left(\mathbf{k},\mathbf{x}^{o}\right) \quad = \quad E_{\mathbf{w}}\left[\partial_{x}Pr\left(\mathbf{y}=\mathbf{k}|\mathbf{x}=\mathbf{x}^{0},\mathbf{w}\right)\right]. \tag{6}$$

The following Lemma contains properties of conditional independence that will be useful in the analysis of subvectors.

Lemma 1. *(Dawid 1979) Given the random vectors* $\mathbf{x}$*,* $\mathbf{y}$ *and* $\mathbf{z}$*, the following properties of conditional independence hold: (i)* $D\left(\mathbf{y}|\mathbf{x},\mathbf{z}\right)=D\left(\mathbf{y}|\mathbf{z}\right)$ *implies* $D\left(\mathbf{x}|\mathbf{y},\mathbf{z}\right)=D\left(\mathbf{x}|\mathbf{z}\right)$*; ii) let* $\mathbf{u}=h\left(\mathbf{y}\right)$*,* $D\left(\mathbf{y}|\mathbf{x},\mathbf{z}\right)=D\left(\mathbf{y}|\mathbf{z}\right)$ *implies* $D\left(\mathbf{u}|\mathbf{x},\mathbf{z}\right)=D\left(\mathbf{u}|\mathbf{z}\right)$ *and* $D\left(\mathbf{y}|\mathbf{x},\mathbf{z},\mathbf{u}\right)=D\left(\mathbf{y}|\mathbf{z},\mathbf{u}\right)$*.*

Then, Lemma 1 assures that A.1 continue to hold with $\mathbf{y}$ replaced by any its subvector $\mathbf{y}_{a}$:

$$Pr\left(\mathbf{y}_{a}=\mathbf{k}_{a}|\mathbf{x},\mathbf{w},\mathbf{q}\right)=Pr\left(\mathbf{y}_{a}=\mathbf{k}_{a}|\mathbf{x},\mathbf{q}\right) \tag{7}$$

for any $\mathbf{k}_{a}$. This combined with A.2 assures identification of APEs of marginal probabilities

$$E_{\mathbf{q}}\left[Pr\left(\mathbf{y}_{a}=\mathbf{k}_{a}|\mathbf{x}^{0},\mathbf{q}\right)\right]=E_{\mathbf{w}}\left[Pr\left(\mathbf{y}_{a}=\mathbf{k}_{a}\theta|\mathbf{x}^{0},\mathbf{w}\right)\right] \tag{8}$$

and APEs

$$APE_{x}\left(\mathbf{k}_{a},\mathbf{x}^{o}\right) \equiv E_{\mathbf{q}}\left\{\partial_{x}Pr\left(\mathbf{y}_{a}=\mathbf{k}_{a}|\mathbf{x}^{o}\,\mathbf{q}\right)\right\}=E_{\mathbf{w}}\left[\partial_{x}Pr\left(\mathbf{y}_{a}=\mathbf{k}_{a}|\mathbf{x}=\mathbf{x}^{0},\mathbf{w}\right)\right]. \tag{9}$$

If, partitioning $\mathbf{x}$ as $\mathbf{x} = (\mathbf{x}_a\ \mathbf{x}_b)$, $Pr\left(\mathbf{y}_a = \mathbf{k}_a|\mathbf{x}, \mathbf{q}\right) = Pr\left(\mathbf{y}_a = \mathbf{k}_a|\mathbf{x}_a, \mathbf{q}\right)$, then by Lemma 1 Equation (7) and A.2 continue to hold with $\mathbf{x}$ replaced by $\mathbf{x}_a$ and consequently Equations (8) and (9) follow suit. Conditional response probabilities requires a specific analysis in Section 4.

3 The multivariate probit model

Multivariate probit models are constructed by supplementing the random vector $\mathbf{y}$ defined in (1) with the latent regression model

$$y_j^* = \alpha_j + \mathbf{x}'\boldsymbol{\beta}_j + \varepsilon_j \tag{10}$$

$j = 1, ..., m$, and α_j, $\boldsymbol{\beta}_j$, $\mathbf{x}$ and ε_j are, respectively, the constant term, the $p \times 1$ vectors of parameters and explanatory variables and the error term. Stacking all ε_j's into the vector $\boldsymbol{\varepsilon} \equiv (\varepsilon_1, \ldots, \varepsilon_m)'$, we assume $\boldsymbol{\varepsilon}|\mathbf{x} \sim N(\mathbf{0}, R)$. The covariance matrix R is subject to normalization restrictions that will be made explicit below. Equation specific regressors are accommodated by allowing $\boldsymbol{\beta}_j$ to have zeroes in the positions of the variables in $\mathbf{x}$ that are excluded from equation j. Indeed, we will see below that exclusion restrictions have no consequence as far as consistent estimation of APEs is concerned. Cross-equation restrictions on the $\boldsymbol{\beta}$'s are also permitted as we will see in Subsection 3.2.

Augmenting each of the equations in (10) with an additive LH component q_j yields the structural model

$$y_j^* = \alpha_j + \mathbf{x}'\boldsymbol{\beta}_j + q_j + \varepsilon_j \tag{11}$$

$j = 1, ..., m$.

Given the vector of control variables, $\mathbf{w}$, we maintain throughout an assumption of conditionally independent LH: $\boldsymbol{\varepsilon}|\mathbf{x}, \mathbf{w}, \mathbf{q} \sim N(\mathbf{0}, R)$ and $\mathbf{q}|\mathbf{x}, \mathbf{w} \sim N(\boldsymbol{\mu}, \Omega)$, where

$$\boldsymbol{\mu} \equiv \begin{pmatrix} \eta_1 + \mathbf{w}'\boldsymbol{\delta}_1 \\ \vdots \\ \eta_m + \mathbf{w}'\boldsymbol{\delta}_m \end{pmatrix},$$

R is normalized for scale and may or may not be UBN, Ω is either arbitrary or constrained, with ω_{ij}

denoting its (i, j) element.

Defining $\boldsymbol{\nu} \equiv \mathbf{q} - \boldsymbol{\mu}$, the foregoing distributional assumptions imply $\boldsymbol{\varepsilon}|\mathbf{x}, \mathbf{w}, \boldsymbol{\nu} \sim N(\mathbf{0}, R)$ and $\boldsymbol{\nu}|\mathbf{x}, \mathbf{w} \sim N(\mathbf{0}, \Omega)$ and, consequently, we end up with the following general multivariate regression model

$$y_j^* = (\alpha_j + \eta_j) + \mathbf{x}'\boldsymbol{\beta}_j + \mathbf{w}'\boldsymbol{\delta}_j + \varepsilon_j + \nu_j, \tag{12}$$

$j = 1, ..., m$ and $\boldsymbol{\varepsilon} + \boldsymbol{\nu}|\mathbf{x}, \mathbf{w} \sim N(\mathbf{0}, \Omega + R)$.

We begin with a multivariate probit model based on the latent regressions (11) and with no explicit parameter constraints. The number of outcomes is $g = 2$, so that $k_j \in \{0, 1\}$, and R is in correlation form, $R \equiv C(\rho_{ij})$. Let $diag(\Omega)$ indicate the $m \times m$ diagonal matrix with the same diagonal as Ω,

$$diag(\Omega) \equiv \begin{pmatrix} \omega_{11} & 0 & \dots & 0 \\ 0 & \omega_{22} & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & \omega_{mm} \end{pmatrix},$$

then the normalized covariance matrix

$$\Xi \equiv (I_m + diag(\Omega))^{-1/2} (\Omega + R) (I_m + diag(\Omega))^{-1/2} \tag{13}$$

is also in correlation form, with $\Xi = C(\xi_{ij})$ and

$$\xi_{ij} \equiv \frac{\omega_{ij} + \rho_{ij}}{\sqrt{(1 + \omega_{ii})(1 + \omega_{jj})}}. \tag{14}$$

It is evident from Equation (14) that the ξ_{ij}'s are free to take on any value. Hence, the reduced-form m latent regressions

$$\frac{y_j^*}{\sqrt{1 + \omega_{jj}}} = \frac{\alpha_j + \eta_j}{\sqrt{1 + \omega_{jj}}} + \mathbf{x}'\frac{\boldsymbol{\beta}_j}{\sqrt{1 + \omega_{jj}}} + \mathbf{w}'\frac{\boldsymbol{\delta}_j}{\sqrt{1 + \omega_{jj}}} + \frac{\varepsilon_j + \nu_j}{\sqrt{1 + \omega_{jj}}} \tag{15}$$

$j=1,...,m$, constitute a legitimate multivariate probit model with

$$(I_m + diag(\Omega))^{-1/2} (\boldsymbol{\varepsilon} + \boldsymbol{\nu}) | \mathbf{x}, \mathbf{w} \sim N(\mathbf{0}, \Xi)$$

and

$$Pr(\mathbf{y} = \mathbf{k} | \mathbf{x}\, \mathbf{w}) = \Phi_{C\left(s_{k_i} s_{k_j} \xi_{ij}\right)} (s_{k_1} h_1, ..., s_{k_m} h_m), \tag{16}$$

where $\Phi_{C\left(s_{k_i} s_{k_j} \xi_{ij}\right)}$ denotes the zero-mean multivariate normal distribution with covariance matrix $C\left(s_{k_i} s_{k_j} \xi_{ij}\right)$, $s_{k_j} \equiv 2k_j - 1$, $k_j \in \{0,1\}$,

$$h_j \equiv \frac{\alpha_j + \eta_j}{\sqrt{1+\omega_{jj}}} + \mathbf{x}' \frac{\boldsymbol{\beta}_j}{\sqrt{1+\omega_{jj}}} + \mathbf{w}' \frac{\boldsymbol{\delta}_j}{\sqrt{1+\omega_{jj}}},$$

and $j = 1, ..., m$.

The above implies that the reduced-form multivariate-probit ML estimator will provide accurate estimates of $(\alpha_j + \eta_j)/\sqrt{1+\omega_{jj}}$, $\boldsymbol{\beta}_j/\sqrt{1+\omega_{jj}}$ and ξ_{ji} as defined in (14). Hence, referring to $\hat{\alpha}_j$, $\hat{\boldsymbol{\beta}}_j$, $\hat{\boldsymbol{\delta}}_j$ and $\hat{\rho}_{ij}$ as the components of the foregoing estimator for, respectively, the constant terms, coefficients on $\mathbf{x}$ and $\mathbf{w}$ and the covariances, we have proved the following.

Result 1 $\hat{\alpha}_j, \hat{\boldsymbol{\beta}}_j$ and $\hat{\boldsymbol{\delta}}_j$ converge in probability to $(\alpha_j + \eta_j)/\sqrt{1+\omega_{jj}}$, $\boldsymbol{\beta}_j/\sqrt{1+\omega_{jj}}$ and $\boldsymbol{\delta}_j/\sqrt{1+\omega_{jj}}$, respectively.

Result 2 $\hat{\rho}_{ij}$ converges in probability to ξ_{ji} as defined in (14).

Within this framework $E_{\mathbf{w}}\left[Pr\left(\mathbf{y} = \mathbf{k} | \mathbf{x} = \mathbf{x}^0, \mathbf{w}\right)\right]$ is identified under weak regularity conditions. Indeed, let

$$\begin{aligned} h_j^0 &\equiv \frac{\alpha_j + \eta_j}{\sqrt{1+\omega_{jj}}} + \mathbf{x}^{0\prime} \frac{\boldsymbol{\beta}_j}{\sqrt{1+\omega_{jj}}} + \mathbf{w}' \frac{\boldsymbol{\delta}_j}{\sqrt{1+\omega_{jj}}}, \\ \hat{h}_{j,i}^0 &\equiv \hat{\alpha}_j + \mathbf{x}^{0\prime} \hat{\boldsymbol{\beta}}_j + \mathbf{w}_i' \hat{\boldsymbol{\delta}}_j, \end{aligned}$$

where $i = 1, ...n$ denote the i.th observation in the sample and $j = 1, ..., m$. Then

$$E_{\mathbf{w}}\left[Pr\left(\mathbf{y} = \mathbf{k} | \mathbf{x}^0, \mathbf{w}\right)\right] = E_{\mathbf{w}}\left[\Phi_{C(s_{ip} s_{jp} \xi_{ij})}\left(s_{k_1} h_1^0, ..., s_{k_m} h_m^0\right)\right]$$

Throughout we maintain that the regularity conditions of Lemma 12.1 in Wooldridge (2010) are met, so that sample averages evaluated at consistent estimates are consistent estimators of population means. Hence, in this case

Result 3 $(1/n)\sum_{i=1}^{n}\left[\Phi_{C\left(s_{k_1}s_{k_j}\hat{\rho}_{ij}\right)}\left(s_{k_1}\hat{h}^0_{1,i},...,s_{k_m}\hat{h}^0_{m,i}\right)\right]$ converges in probability to

$$E_{\mathbf{w}}\left[\Phi_{C\left(s_{k_i}s_{k_j}\xi_{ij}\right)}\left(s_{k_1}h^0_1,...,s_{k_m}h^0_m\right)\right].$$

From **Result 1** we find that $\hat{\alpha}_j$ estimates α_j with both location and rescaling biases, while $\hat{\boldsymbol{\beta}}_j$ estimates $\boldsymbol{\beta}_j$ with a rescaling bias, as already evidenced in Yatchew and Griliches (1985) in the univariate context. **Result 2** uncovers that $\hat{\rho}_{ij}$, as an estimator of ρ_{ij}, is affected by both rescaling and location biases. Given **Result 3** and Equation (6), it turns out that APEs of the generic component x of $\mathbf{x}$ at any point $\mathbf{x}^o$ are consistently estimated by:

$$\widehat{APE_x\left(\mathbf{k},\mathbf{x}^0\right)} \equiv (1/n)\sum_{i=1}^{n}\left[\partial_x\Phi_{C\left(s_{k_1}s_{k_j}\hat{\rho}_{ij}\right)}\left(s_{k_1}\hat{h}^0_{1,i},...,s_{k_m}\hat{h}^0_{m,i}\right)\right]. \tag{17}$$

Importantly, this holds with a covariance matrix of the LH components, Ω, that is fully general.

The multivariate normal distribution has the well-known property that any $m_1\times 1$ sub-vector of an $m\times 1$ random vector with a multivariate normal distribution, $m_1 < m$, has a marginal distribution that is a m_1-variate normal distribution (see Rao 1973, p. 522). This proves at once that **Result 3** carries over to all possible marginal response probabilities of the same multivariate probit model, which along with Equation (9) implies consistent estimation of the the corresponding APEs.

Another direct implication of the marginalization property is worth reporting here as a lemma for future reference.

Lemma 2. *Any subsystem of m_1 latent equations from an m-variate probit latent equation system, $m_1 < m$, constitutes a legitimate m_1-variate probit latent equation system.*

A first implication of Lemma 2 is that APEs on marginal response probabilities can be estimated using multivariate probit models of suitably reduced dimensionality.

Remark 3. This framework applies to clustered data with LH, such as a panel-data multivariate probit with correlated effects modeled a la Mundlak (or a la Chamberlain) where $\mathbf{q}$ is the vector of correlated effects for an individual in the population and the $\mathbf{w}$'s are the cluster means (the time values) of the $\mathbf{x}$'s. In principle, random effect estimation can be applied to the reduced-form model, identifying the reduced-form equation variances, ξ_{ii}, $i = 1, ...m$. As far as APEs are concerned, though, this is not necessary and a computationally easier partial-ML estimator is available, which simply applies the conventional reduced-form multivariate-probit ML estimator and corrects the standard-error estimates for cluster correlation.

3.1 Within-equation restrictions

The nonparametric analysis at the end of Section 2 has shown that exclusion restrictions are not a problem for the identification of APEs. In the Probit framework this straightforwardly extends to within-equation linear homogenous restrictions, such as exclusion or equality restrictions, since these are invariant to rescaling of coefficients: given $\omega_{jj} \neq 0$, a known full-row-rank $J \times k$ matrix A and some vector $\mathbf{b}$ such that $A\boldsymbol{\beta}_j = \mathbf{b}$, then $A\boldsymbol{\beta}_j/\sqrt{1+\omega_{jj}} = \mathbf{b}$ if and only if $\mathbf{b} = \mathbf{0}$. From a practical point of view, this implies that APEs can be consistently estimated by imposing the structural restrictions directly onto the the rescaled parameters of the reduced-form model.

Non-homogenous restrictions, i. e. $\mathbf{b} \neq \mathbf{0}$, apply in various interesting situations. Willingness-to-pay models comparing the observed cost and the latent benefit of a given option are an example (Wooldridge 2005). From the previous paragraph it is clear that if R is in correlation form, that is all J non-homogeneous restrictions are beyond normalization, the constrained ML reduced-form estimator is consistent if and only if $\omega_{jj} = 0$. This restrictive framework is what we would implement by using, for example, Stata's `biprobit` with all the non-homogenous constraints passed through the options `constraint` or `offset`. But unlike within-equation homogenous restrictions, non-homogenous restrictions have normalizing power, which can be exploited to allow a more general R, with an arbitrary ρ_{jj} replacing 1. This also allows an easy estimation procedure. In fact, A can be always partitioned as $A = (A_1\ A_2)$, where A_1 is of full column rank. Correspondingly let $\boldsymbol{\beta}_j = \left(\boldsymbol{\beta}_{1j}\ \boldsymbol{\beta}_{2j}\right)$ and $\mathbf{x} = (\mathbf{x}_1\ \mathbf{x}_2)$, then $\boldsymbol{\beta}_{1j} = \mathbf{c} - C\boldsymbol{\beta}_{2j}$ and $\mathbf{x}'\boldsymbol{\beta}_j = \mathbf{x}_1'\mathbf{c} + (\mathbf{x}_2 - C'\mathbf{x}_1)'\boldsymbol{\beta}_{2j}$, where $\mathbf{c} = (A_1'A_1)^{-1} A_1'\mathbf{b}$ and

$C=\left(A_1'A_1\right)^{-1}A_1'A_2$ are known. This proves that in the absence of LH, given the rescaled equation

$$\frac{y_j^*}{\rho_{jj}}=\frac{\alpha_j}{\rho_{jj}}+\mathbf{x}_1'\mathbf{c}\frac{1}{\rho_{jj}}+\left(\mathbf{x}_2-C'\mathbf{x}_1\right)'\frac{\boldsymbol{\beta}_{2j}}{\rho_{jj}}+\frac{\varepsilon_j}{\rho_{jj}}$$

ρ_{jj} would be identified along with $\boldsymbol{\beta}_j$. With LH the following equation

$$\frac{y_j^*}{\sqrt{\rho_{jj}+\omega_{jj}}}=\frac{\alpha_j+\eta_i}{\sqrt{\rho_{jj}+\omega_{jj}}}+\mathbf{x}_1'\mathbf{c}\frac{1}{\sqrt{\rho_{jj}+\omega_{jj}}}+\left(\mathbf{x}_2-C'\mathbf{x}_1\right)'\frac{\boldsymbol{\beta}_{2j}}{\sqrt{\rho_{jj}+\omega_{jj}}}+\mathbf{w}'\frac{\boldsymbol{\delta}_j}{\sqrt{\rho_{jj}+\omega_{jj}}}+\frac{\varepsilon_j+\nu_j}{\sqrt{\rho_{jj}+\omega_{jj}}}$$

allows to identify $\boldsymbol{\beta}_j$ and $\rho_{jj}+\omega_{jj}$ and, consequently, the APEs. This proves that APEs are identified in the presence of arbitrary Ω and $J-1$ non-homogeneous restrictions that are beyond normalization. All of the above considerations clearly hold regardless of the model being univariate or multivariate.

3.2 Cross-equation equality restrictions

Cross-equation equality restrictions are often part of the conventional specifications. The *multinomial probit with alternative-specific regressors* (Train 2002), henceforth ASMP, implemented for example by Stata's `asmprobit`, and the *panel probit model* (Bertschek and Lechner 1998 and Greene 2004) are two notable examples.

The equality restrictions permit an error covariance matrix that is more than general than in the unconstrained model. Partition $\mathbf{x}$ as $\mathbf{x}=\left(\mathbf{x}_1',\mathbf{x}_2'\right)'$, with each component of the $k_2\times 1$ vector $\mathbf{x}_2$ having a common coefficient across equations. A non-empty $\mathbf{x}_2$ allows identification of the $\boldsymbol{\beta}'s$ with a general covariance matrix for $\boldsymbol{\varepsilon}$:

$$R=\begin{pmatrix} 1 & \rho_{12} & \cdots & \rho_{1m} \\ \rho_{12} & \rho_{22} & & \vdots \\ \vdots & & \ddots & \\ \rho_{1m} & \cdots & & \rho_{mm} \end{pmatrix}. \tag{18}$$

It is not hard to see that the normalized covariance matrix $\Xi=\left(1+\omega_{11}\right)^{-1}(\Omega+R)$ satisfies the same restrictions as R defined in Equation (18). Hence, the m latent regressions

$$\frac{y_j^*}{\sqrt{1+\omega_{11}}} = \frac{\alpha_j + \eta_j}{\sqrt{1+\omega_{11}}} + \mathbf{x}_1' \frac{\boldsymbol{\beta}_{1j}}{\sqrt{1+\omega_{11}}} + \mathbf{x}_2' \frac{\boldsymbol{\beta}_2}{\sqrt{1+\omega_{11}}} + \mathbf{w}' \frac{\boldsymbol{\delta}_j}{\sqrt{1+\omega_{11}}} + \frac{\varepsilon_j + \nu_j}{\sqrt{1+\omega_{11}}}, \tag{19}$$

$j = 1, ..., m$, constitute a legitimate multivariate probit model with common coefficients for $\mathbf{x}_2$ and

$$(1+\omega_{11})^{-1/2} (\boldsymbol{\varepsilon} + \boldsymbol{\nu}) \, | \mathbf{x}, \mathbf{w} \sim N(\mathbf{0}, \Xi),$$

$\Xi \equiv (1+\omega_{11})^{-1} (\Omega + R)$. This proves that results analogous to **Results 1-3** hold true with $(k_2 - 1)(m-1)$ cross-equation equality restrictions beyond normalization, if $k_2 > 1$, and arbitrary Ω. This result exactly parallels what found in the case of non-homogenous restrictions.

Remark 4. If the variables $\mathbf{x}_2$ and $\varepsilon_j + \nu_j$ are differences with respect to an alternative 0, Model (19) can be thought of as the part of an ASMP specification with $m+1$ alternatives that determines whether or not alternative 0 is chosen (see Mullahy 2011).

Remark 5. If $\mathbf{x}_1$ is empty and m is the number of occasions, Model (19) is the the panel probit model. This model is not to be confused with that of Remark 3. Here we have a univariate model that is replicated m times and cluster correlation is fully taken into account with no necessity of robust standard-error estimates.

With cross-equation equality restrictions, homoskedasticity of $\boldsymbol{\varepsilon}$ across equations implies that R is in correlation form, with $m-1$ restrictions beyond normalization and as such not UBN. Estimating the reduced-form model with the same constraints and in correlation form, then, will provide consistent estimates of APEs if and only if $\boldsymbol{\varepsilon} + \boldsymbol{\nu}$ is homoskedastic across equations, or equivalently if and only if $\omega_{11} = \omega_{22} = ... = \omega_{mm}$. Indeed, if an homoskedastic idiosyncratic error may be a reasonable assumption, as observed by Greene (2004) in the context of the panel probit model, it may be overly restrictive for the LH components, especially if these capture the impacts of omitted variables.

The limitation documented in the previous paragraph is substantial, since for most statistical softwares it is not possible to get an error covariance matrix that is more general than the correlation form. Within Stata this occurs not only with `biprobit` and `mvprobit`, but also with the more flexible `cmp` (Roodman 2011), where the most general covariance option, `unstructured`, when applied to a multivariate probit model imposes a covariance matrix in correlation form. Limdep's `BIVARIATE PROBIT` and `MPROBIT` also impose the correlation form. Maintaining cross-equation equality restrictions in all of those cases boils down to maintaining homoskedasticity in the LH covariance matrix. Stata's `asmprobit` allows much more flexibility in this respect, but it is specialized to ASMP.

3.3 Covariance restrictions

Covariance restrictions beyond normalization are not immaterial to consistent estimation of APEs in the presence of LH. For example, by direct inspection of Equation (14) it follows that the restriction $\rho_{ij} = 0$ for some i and j, do not translate into a zero-covariance restrictions for the composite error, unless $\omega_{ij} = 0$ and as such is not compatible with an arbitrary Ω.

3.3.1 The multinomial probit model implemented by Stata's `mprobit`

The model implemented by Stata's `mprobit` is a multinomial probit model with no cross-equation equality constraints, maintaining the assumption of *independence from irrelevant alternatives*. As in the ASMP model of Subsection 3.2, the latent regression system is to be thought of as the part of the multinomial probit model that identifies the probability of choosing alternative 0. But there is the additional assumption that the idiosyncratic errors in the utility equation, say ϵ_i, $i = 0, 1, ..., m$, are jointly distributed according to a $m+1$-variate normal distribution with zero means and covariance matrix equal to I_{m+1}. Hence, $\varepsilon_j \equiv \epsilon_j - \epsilon_0$, $j = 1, ..., m$, are zero-mean jointly normal with covariance matrix

$$R \equiv \begin{pmatrix} 2 & 1 & \cdots & 1 \\ 1 & 2 & & \vdots \\ \vdots & & \ddots & \\ 1 & \cdots & & 2 \end{pmatrix}.$$

Unlike the ASMP, here consistent estimation of APEs occurs only under a specific class of conditional independent LH. This is shown by proving that a positive, diagonal matrix A such that $R = A(R + \Omega)A$ exists if and only if Ω is suitably restricted. Indeed, $R = A(R + \Omega)A$ is equivalent to the system of $m(m+1)/2$ equations in m unknowns

$$a_i a_j = \frac{\rho_{ij}}{\omega_{ij} + \rho_{ij}}$$

$i, j = 1, ..., m$, which is overdetermined when $m \geq 2$. It is thereby evident that the ω_{ij}'s must be restricted to make the system consistent. For example if $m = 2$,

$$\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} a_1 & 0 \\ 0 & a_2 \end{pmatrix} \begin{pmatrix} 2+\omega_{11} & 1+\omega_{12} \\ 1+\omega_{12} & 2+\omega_{22} \end{pmatrix} \begin{pmatrix} a_1 & 0 \\ 0 & a_2 \end{pmatrix}$$

and so

$$\begin{aligned} a_1 &= \sqrt{\frac{2}{2+\omega_{11}}} \\ a_2 &= \sqrt{\frac{2}{2+\omega_{22}}} \\ \omega_{12} &= \frac{1}{a_1 a_2} - 1 \end{aligned}$$

with $0 < a_1 < 1$ and $0 < a_2 < 1$, so that eventually ω_{12} depends on the values of ω_{11} and ω_{22} and is positive. Therefore, the APE estimates here are robust to (conditional) independent heterogeneity such that

$$\Omega = \begin{pmatrix} \omega_{11} & -1+\frac{1}{2}\sqrt{(2+\omega_{11})(2+\omega_{22})} \\ -1+\frac{1}{2}\sqrt{(2+\omega_{11})(2+\omega_{22})} & \omega_{22} \end{pmatrix}. \tag{20}$$

A matrix Ω^o that belongs to the class of matrices in Equation (20) emerges in the case of unobserved random components in the non-differenced utility equations, $\tilde{q}_j$, $j = 0, 1, ..., m$ that are independent and with unknown common variance, $\omega > 0$. In this case, $\Omega^o = \omega R$, which satisfies Equation (20).

3.3.2 Zero-covariance-factorizing restrictions

We say that a covariance matrix presents zero-covariance-factorizing restrictions if it can be transformed into a block diagonal matrix through identical permutation of rows and columns. If R is such a matrix, then $\boldsymbol{\varepsilon}$ that can be partitioned into independent subvectors and, in the absence of LH, the likelihood function would factorize. Two examples are 1) $R = I$ and 2) any R such that, for fixed i, $\rho_{ij} = 0$ for all $j \neq i$. The former requires separate estimation of m univariate probit models, while the latter involve separate estimations of an $(m-1)$-variate probit model and a univariate probit model.

What happen if we apply such restrictions to the rescaled models? It is easy to prove that **Results 1** and **2** hold true, limited to the unconstrained parameters (all $\boldsymbol{\beta}'s$ and all $\rho_{ij} \neq 0$). In fact, the patterns in R do not carry over into Ξ and so the true likelihood function does not factorize: for some i and j there is the true constraint $\rho_{ij} = 0$, but LH may bring about a non-zero covariance between the scaled latent components. Nonetheless, in the light of Lemma 2 the constrained ML estimator remains consistent for the scaled coefficients, as well as the unconstrained elements of Ξ (Wooldridge 2010, p. 595, makes similar considerations in the context of the bivariate probit model). Clearly, **Result 3** fails for the joint response probability, implying inconsistent estimates of the corresponding APEs, unless the Ω matrix replicates the same patterns as R. It does go through with arbitrary Ω, limited to the APEs of the subsystem marginal probabilities.

Example 6. *If all $\rho_{ij} = 0$, $i \neq j$, running m separate probit regressions yield consistent estimates of the marginal probabilities $Pr\left(y_j = k_j|\mathbf{x}\,\mathbf{w}\right)$ and so of $\prod_{j=1}^{m} Pr\left(y_j = k_j|\mathbf{x}\,\mathbf{w}\right)$, but the latter equals $Pr\left(\mathbf{y} = \mathbf{k}|\mathbf{x}\,\mathbf{w}\right)$ if and only if also $\omega_{ij} = 0$, $i \neq j$. Only in this case, in fact, $\Xi = I_m$. Since the scaled coefficients are consistently estimated, the resulting inconsistency on the APEs is only due to non-zero covariances and is bound to be small if the latter are small enough. In fact, focussing on $m = 2$, the APE of x, a generic component of the $\mathbf{x}$ vector, is*

$$APE_x\left(\mathbf{k},\mathbf{x}^0\right) = E_{\mathbf{w}}\left[\phi\left(\gamma_1^0\right)\Phi\left(\frac{\gamma_2^0 - \xi_{12}\left(\mathbf{k}\right)\gamma_1^0}{\sqrt{1-\xi_{12}^2}}\right)\frac{\beta_1}{\sqrt{1+\omega_{11}}} + \phi\left(\gamma_2^0\right)\Phi\left(\frac{\gamma_1^0 - \xi_{12}\left(\mathbf{k}\right)\gamma_2^0}{\sqrt{1-\xi_{12}^2}}\right)\frac{\beta_2}{\sqrt{1+\omega_{22}}}\right].$$

where β_j is the coefficient on x in equation $j = 1, 2$, $\gamma_j^0 \equiv s_{k_j} h_j^0$, $j = 1, 2$, $\xi_{12}\left(\mathbf{k}\right) = s_{k_1} s_{k_2} \xi_{12}$ and ξ_{12}

is defined as in (14). If $\rho = 0$ and ML is incorrectly carried out subject to $\xi_{12} = 0$, the estimated APE converges to

$$APE_x^{plim}\left(\mathbf{k}, \mathbf{x}^0\right) = E_{\mathbf{w}}\left[\phi\left(\gamma_1^0\right)\Phi\left(\gamma_2^0\right)\frac{\beta_1}{\sqrt{1+\omega_{11}}} + \phi\left(\gamma_2^0\right)\Phi\left(\gamma_1^0\right)\frac{\beta_2}{\sqrt{1+\omega_{22}}}\right].$$

with a bias that can be approximated through a first-order Taylor expansion around zero as

$$bias\left(\mathbf{k}, \mathbf{x}^0\right) = E_{\mathbf{w}}\left[\xi_{12}\left(\mathbf{k}\right)\phi\left(\gamma_1^0\right)\phi\left(\gamma_2^0\right)\left(\frac{\beta_1}{\sqrt{1+\omega_{11}}}\gamma_1^0 + \frac{\beta_2}{\sqrt{1+\omega_{22}}}\gamma_2^0\right)\right].$$

From the above it is clear that the bias gets smaller the smaller ξ_{12}. Clearly, APEs evaluated for the marginal probabilities, $Pr\left(y_j = k_j|\mathbf{x}\,\mathbf{w}\right)$, are instead identified. If, more in general, R is exchangeable, i.e. $\rho_{ij} = \rho$, $i \neq j$, then by direct inspection of Equation (14) one proves that assuming $\omega_{ii} = \omega_{jj}$, $i \neq j = 1, ..., m$, Ξ is exchangeable if and only if Ω is exchangeable too. With $\omega_{ii} \neq \omega_{jj}$, for some $i \neq j = 1, ..., m$, exchangeability of Ξ is ensured by more complex non-linear relationships across the ω_{ij}'s and ρ. In either case, consistent constrained estimation of APEs no longer supports a general Ω matrix.

Remark 7. Zero-covariance-factorizing restrictions are not problematic since they can be easily tested in most statistical packages. They are mainly implemented for computational ease, therefore in the case of non-rejection it would make sense to apply them to the rescaled systems, independently by what it is believed about the structural covariances (see Greene 1998, for example).

3.4 The recursive multivariate probit model

The recursive multivariate probit model is a notable case of within equation exclusion restrictions. It includes the $\mathbf{y}$ in the right-hand side of the latent system with an $m \times m$ matrix of coefficients that is restricted to be triangular (Roodman 2011)[2]. Starting with the contributions of Evans and Schwab (1995) and Greene (1998), there are by now many econometric applications of this model, including the recent articles by Fichera and Sutton (2011) and Entorf (2012).

[2]Wooldridge (2010) argues that substantial identification in recursive models may also require exclusion restrictions involving the $\mathbf{x}$

The feature that makes the recursive multivariate probit model appealing is that it accommodates endogenous, binary explanatory variables without special provisions for endogeneity, simply maximizing the log-likelihood function as if the explanatory variables were all ordinary exogenous variables (see Maddala 1983, Wooldridge 2010,Greene 2012 and, for a general proof, Roodman 2011). This can be easily seen here in the simple case of a recursive bivariate probit model with no LH

$$\begin{aligned}
Pr\left(y_1=1,y_2=1|\mathbf{x}\right) &= Pr\left(y_1=1|y_2=1,\mathbf{x}\right)P\left(y_2=1|\mathbf{x}\right)\\
&= Pr\left[\varepsilon_1>-\alpha_1-\mathbf{x}'\boldsymbol{\beta}_1-\lambda|y_2=1,\mathbf{x}\right]P\left[y_2=1|\mathbf{x}\right]\\
&= Pr\left[\varepsilon_1>-\alpha_1-\mathbf{x}'\boldsymbol{\beta}_1-\lambda|\varepsilon_2>-\alpha_2-\mathbf{x}'\boldsymbol{\beta}_2,\mathbf{x}\right]P\left[\varepsilon_2>-\alpha_2-\mathbf{x}'\boldsymbol{\beta}_2|\mathbf{x}\right]\\
&= Pr\left[\varepsilon_1>-\alpha_1-\mathbf{x}'\boldsymbol{\beta}_1-\lambda,\varepsilon_2>-\alpha_2-\mathbf{x}'\boldsymbol{\beta}_2|\mathbf{x}\right]\\
&= \Phi_{C(\rho_{12})}\left(\alpha_1+\mathbf{x}'\boldsymbol{\beta}_1+\lambda,\alpha_2+\mathbf{x}'\boldsymbol{\beta}_2\right)
\end{aligned}$$

The crux of the above derivations is that, given

$$y_1=1\left(\varepsilon_1>-\alpha_1-\mathbf{x}'\boldsymbol{\beta}_1-\lambda y_2\right)\text{ and }y_2=1\left(\varepsilon_2>-\alpha_2-\mathbf{x}'\boldsymbol{\beta}_2\right),$$

ε_1 is independent of the lower limit of integration conditional on $\varepsilon_2>-\alpha_2-\mathbf{x}'\boldsymbol{\beta}_2$ and so no endogeneity issue emerges when working out the joint probability as a joint normal distribution. The other three joint probabilities are similarly derived, so that eventually the likelihood function is assembled exactly as in a conventional multivariate probit model. But this implies, crucially, that a separate analysis for the recursive multivariate probit model with conditionally independent LH is not needed and so **Results 1-3** carry over smoothly.

4 Conditioning on response subvectors

Greene (1996, 1998, 2012), , Christofides, Stengos, and Swidinsky (1997) and Mullahy 2011 present the formulas of PEs based on response probabilities conditional on response subvectors and Entorf (2012) is an interesting recent application where these parameters are estimated. We notice, though, a key identification issue that seems to have received scarce attention in the literature so far, and that arises when estimating conditional-probability-based APEs in the presence of conditional independent LH. It regards the partition of the LH components between the conditional and the conditioning processes.

To elaborate, let $\mathbf{y} = (\mathbf{y}_a\, \mathbf{y}_b)$, $\mathbf{k} = (\mathbf{k}_a\, \mathbf{k}_b)$ and $\mathbf{q} = (\mathbf{q}_a\, \mathbf{q}_b)$ and let $Pr\left(\mathbf{y}_a = \mathbf{k}_a | \mathbf{y}_b = \mathbf{k}_b, \mathbf{x}, \mathbf{q}_a\right)$ be the conditional probability of interest. In addition to A.1 and A.2, we consider the following assumptions

A.3 $Pr\left(\mathbf{y}_b = \mathbf{k}_b | \mathbf{x}, \mathbf{q}\right) = Pr\left(\mathbf{y}_b = \mathbf{k}_b | \mathbf{x}, \mathbf{q}_b\right)$ for any $\mathbf{k}_b$.

A.4 $D\left(\mathbf{q}_b | \mathbf{w}\right) = D\left(\mathbf{q}_b\right).$

A.3 is met in multivariate probit models when $\mathbf{q}_b$ are the only LH components that enters the subsystem peculiar to $\mathbf{y}_b$. It is also met in recursive models if, in addition, $\mathbf{y}_b$ is the subsystem that does not have $\mathbf{y}_a$ as right-hand variables. We will elaborate further on this assumption in Remark 12 below. A.4 is more substantial, maintaining independence of $\mathbf{w}$ and $\mathbf{q}_b$.

In the following it is understood that $Pr\left(\mathbf{y}_a = \mathbf{k}_a | \mathbf{y}_b, \mathbf{x}, \mathbf{w}\right)$ is identified, which is the case not only in models with all exogenous variables, but also in recursive models, as observed in Subsection 3.4. Our question, then, is under what circumstances APEs computed on $E_{\mathbf{w}}\left[Pr\left(\mathbf{y}_a = \mathbf{k}_a | \mathbf{y}_b, \mathbf{x}, \mathbf{w}\right)\right]$ are equal to APEs computed on $E_{\mathbf{q}_a}\left[Pr\left(\mathbf{y}_a = \mathbf{k}_a | \mathbf{y}_b = \mathbf{k}_b, \mathbf{x}^0, \mathbf{q}_a\right)\right]$. We start with the following result

Proposition 8. *Given A.1-A.4*

$$E_{\mathbf{w}}\left[Pr\left(\mathbf{y}_a = \mathbf{k}_a | \mathbf{y}_b = \mathbf{k}_b, \mathbf{x}^0, \mathbf{w}\right)\right] = \frac{E_{\mathbf{q}}\left[Pr\left(\mathbf{y} = \mathbf{k} | \mathbf{x}^0, \mathbf{q}\right)\right]}{E_{\mathbf{q}_b}\left[Pr\left(\mathbf{y}_b = \mathbf{k}_b | \mathbf{x}^0, \mathbf{q}_b\right)\right]}$$

Proof. Given A.1, Lemma 1(ii) assures $Pr\left(\mathbf{y}_b = \mathbf{k}_b | \mathbf{x}, \mathbf{w}, \mathbf{q}\right) = Pr\left(\mathbf{y}_b = \mathbf{k}_b | \mathbf{x}, \mathbf{q}\right)$ for any $\mathbf{k}_b$, which along with A.2 yields

$$E_{\mathbf{q}}\left[Pr\left(\mathbf{y}_b = \mathbf{k}_b | \mathbf{x}^0, \mathbf{q}\right)\right] = E_{\mathbf{w}}\left[Pr\left(\mathbf{y}_b = \mathbf{k}_b | \mathbf{x}^0, \mathbf{w}\right)\right] \tag{21}$$

for any $\mathbf{k}_b$. Therefore, within the borders of A.1 and A.2, we can identify the ratio of the average of

$Pr\left(\mathbf{y}=\mathbf{k}|\mathbf{x}^0,\mathbf{q}\right)$ to the average of $Pr\left(\mathbf{y}_b=\mathbf{k}_b|\mathbf{x}=\mathbf{x}^0,\mathbf{q}\right)$:

$$\frac{E_{\mathbf{q}}\left[Pr\left(\mathbf{y}=\mathbf{k}|\mathbf{x}^0,\mathbf{q}\right)\right]}{E_{\mathbf{q}}\left[Pr\left(\mathbf{y}_b=\mathbf{k}_b|\mathbf{x}^0,\mathbf{q}\right)\right]}=\frac{E_{\mathbf{w}}\left[Pr\left(\mathbf{y}=\mathbf{k}|\mathbf{x}^0,\mathbf{w}\right)\right]}{E_{\mathbf{w}}\left[Pr\left(\mathbf{y}_b=\mathbf{k}_b|\mathbf{x}^0,\mathbf{w}\right)\right]}. \tag{22}$$

A.3 just specializes the left-hand side of Equation (22) to

$$\frac{E_{\mathbf{q}}\left[Pr\left(\mathbf{y}=\mathbf{k}|\mathbf{x},\mathbf{q}\right)\right]}{E_{\mathbf{q}}\left[Pr\left(\mathbf{y}_b=\mathbf{k}_b|\mathbf{x},\mathbf{q}\right)\right]} \quad = \quad \frac{E_{\mathbf{q}}\left[Pr\left(\mathbf{y}=\mathbf{k}|\mathbf{x},\mathbf{q}\right)\right]}{E_{\mathbf{q}_b}\left[Pr\left(\mathbf{y}_b=\mathbf{k}_b|\mathbf{x},\mathbf{q}_b\right)\right]}. \tag{23}$$

Then, Lemma 1(ii) and A.2 yield $D\left(\mathbf{q}_b|\mathbf{x},\mathbf{w}\right)=D\left(\mathbf{q}_b|\mathbf{w}\right)$ and hence, given A.4,

$$D\left(\mathbf{q}_b|\mathbf{x},\mathbf{w}\right)=D\left(\mathbf{q}_b\right). \tag{24}$$

This leads to the following redundancy result

$$\begin{aligned}
Pr\left(\mathbf{y}_b=\mathbf{k}_b|\mathbf{x},\mathbf{w}\right) &= E_{\mathbf{q}|\mathbf{x},\mathbf{w}}\left[Pr\left(\mathbf{y}_b=\mathbf{k}_b|\mathbf{x},\mathbf{w},\mathbf{q}\right)|\mathbf{x},\mathbf{w}\right] \\
&= E_{\mathbf{q}|\mathbf{x},\mathbf{w}}\left[Pr\left(\mathbf{y}_b=\mathbf{k}_b|\mathbf{x},\mathbf{q}\right)|\mathbf{x},\mathbf{w}\right] \\
&= E_{\mathbf{q_b}|\mathbf{x},\mathbf{w}}\left[Pr\left(\mathbf{y}_b=\mathbf{k}_b|\mathbf{x},\mathbf{q}_b\right)|\mathbf{x},\mathbf{w}\right] \\
&= E_{\mathbf{q_b}|\mathbf{x}}\left[Pr\left(\mathbf{y}_b=\mathbf{k}_b|\mathbf{x},\mathbf{q}_b\right)|\mathbf{x}\right] \\
&= Pr\left(\mathbf{y}_b=\mathbf{k}_b|\mathbf{x}\right),
\end{aligned} \tag{25}$$

where equalities follow from: 1) Equation (3); 2) Lemma 1(ii) and A.2; 3) A.3; 4) Equation (24). Therefore,

$$Pr\left(\mathbf{y}_a=\mathbf{k}_a|\mathbf{y}_b=\mathbf{k}_b,\mathbf{x},\mathbf{w}\right) \quad = \quad \frac{Pr\left(\mathbf{y}=\mathbf{k}|\mathbf{x},\mathbf{w}\right)}{Pr\left(\mathbf{y}_b=\mathbf{k}_b|\mathbf{x}\right)}$$

and so

$$\frac{E_{\mathbf{w}}\left[Pr\left(\mathbf{y}=\mathbf{k}|\mathbf{x}^0,\mathbf{w}\right)\right]}{E_{\mathbf{w}}\left[Pr\left(\mathbf{y}_b=\mathbf{k}_b|\mathbf{x}^0,\mathbf{w}\right)\right]} \quad = \quad E_{\mathbf{w}}\left[Pr\left(\mathbf{y}_a=\mathbf{k}_a|\mathbf{y}_b=\mathbf{k}_b,\mathbf{x}^0,\mathbf{w}\right)\right], \tag{26}$$

which, given Equations (22) and (23), completes the proof. □

Proposition 8 establishes that A.1 and A.2, even if supplemented by A.3 and A.4, are not able to identify averages of conditional probabilities through $E_{\mathbf{w}}\left[Pr\left(\mathbf{y}_a = \mathbf{k}_a | \mathbf{y}_b = \mathbf{k}_b, \mathbf{x}^0, \mathbf{w}\right)\right]$. This issue, which we did not find reported elsewhere in the literature, is general since the discrepancy between what is of interest and what is estimated holds irrespectively of the model for $\mathbf{y}$.

As strong as A.1-A.4 may be, forcing the LH in the conditioning process to be independent of regressors and controls, they leave the statistical dependence between $\mathbf{q}_a$ and $\mathbf{q}_b$ unrestricted. One may expect, therefore, that separating the two LH components can go a long way towards identification of APEs. Indeed, this turns out to be the case.

A.5 $D\left(\mathbf{q}_b | \mathbf{q}_a, \mathbf{w}\right) = D\left(\mathbf{q}_b | \mathbf{w}\right)$.

Proposition 9. *Assume A.1-A.5. Then*

$$E_{\mathbf{w}}\left[Pr\left(\mathbf{y}_a = \mathbf{k}_a | \mathbf{y}_b = \mathbf{k}_b, \mathbf{x}^0, \mathbf{w}\right)\right] = E_{\mathbf{q}_a}\left[Pr\left(\mathbf{y}_a = \mathbf{k}_a | \mathbf{y}_b = \mathbf{k}_b, \mathbf{x}^0, \mathbf{q}_a\right)\right]. \tag{27}$$

Proof. Repeatedly applying Lemma 1(ii) and A.2 gives first $D\left(\mathbf{q} | \mathbf{x}, \mathbf{w}, \mathbf{q}_a\right) = D\left(\mathbf{q} | \mathbf{w}, \mathbf{q}_a\right)$ and then $D\left(\mathbf{q}_b | \mathbf{x}, \mathbf{w}, \mathbf{q}_a\right) = D\left(\mathbf{q}_b | \mathbf{w}, \mathbf{q}_a\right)$. Hence, given A.4, A.5 assures

$$D\left(\mathbf{q}_b | \mathbf{x}, \mathbf{w}, \mathbf{q}_a\right) = D\left(\mathbf{q}_b\right). \tag{28}$$

Equation (28) has three implications. First,

$$\begin{aligned} E_{\mathbf{q}_b}\left[Pr\left(\mathbf{y}_b = \mathbf{k}_b | \mathbf{x}, \mathbf{q}_b\right)\right] &= E_{\mathbf{q}_b|\mathbf{x}}\left[Pr\left(\mathbf{y}_b = \mathbf{k}_b | \mathbf{x}, \mathbf{q}_b\right)\right] \\ &= Pr\left(\mathbf{y}_b = \mathbf{k}_b | \mathbf{x}\right), \end{aligned}$$

where the first equality stems from Equation (28) and the second from Equation (3). Second,

$$
\begin{aligned}
E_{\mathbf{q}}\left[Pr\left(\mathbf{y}=\mathbf{k}|\mathbf{x},\mathbf{q}\right)\right] &= E_{\mathbf{q}_a}\left\{E_{\mathbf{q}_b}\left[Pr\left(\mathbf{y}=\mathbf{k}|\mathbf{x},\mathbf{q}\right)\right]\right\} \\
&= E_{\mathbf{q}_a}\left\{E_{\mathbf{q}_b|\mathbf{x},\mathbf{q}_a}\left[Pr\left(\mathbf{y}=\mathbf{k}|\mathbf{x},\mathbf{q}\right)\right]\right\} \\
&= E_{\mathbf{q}_a}\left[Pr\left(\mathbf{y}=\mathbf{k}|\mathbf{x},\mathbf{q}_a\right)\right]
\end{aligned}
$$

where the first equality stems from Fubini's Theorem, the second from Equation (28) and the last from Equation (3). Finally,

$$
\begin{aligned}
Pr\left(\mathbf{y}_b=\mathbf{k}_b|\mathbf{x},\mathbf{q}_a\right) &= E_{\mathbf{q}_b|\mathbf{x},\mathbf{q}_a}\left[Pr\left(\mathbf{y}_b=\mathbf{k}_b|\mathbf{x},\mathbf{q}\right)\right] \\
&= E_{\mathbf{q}_b|\mathbf{x},\mathbf{q}_a}\left[Pr\left(\mathbf{y}_b=\mathbf{k}_b|\mathbf{x},\mathbf{q}_b\right)\right] \\
&= E_{\mathbf{q}_b|\mathbf{x}}\left[Pr\left(\mathbf{y}_b=\mathbf{k}_b|\mathbf{x},\mathbf{q}_b\right)\right] \\
&= Pr\left(\mathbf{y}_b=\mathbf{k}_b|\mathbf{x}\right),
\end{aligned}
$$

where the first equality stems from Equation (3), the second from A.1, Lemma 1(ii) and A.3, the third from Equation (28) and the last, again, from Equation (3). Therefore,

$$
\begin{aligned}
\frac{E_{\mathbf{q}}\left[Pr\left(\mathbf{y}=\mathbf{k}|\mathbf{x}^0,\mathbf{q}\right)\right]}{E_{\mathbf{q}_b}\left[Pr\left(\mathbf{y}_b=\mathbf{k}_b|\mathbf{x}^0,\mathbf{q}_b\right)\right]} &= E_{\mathbf{q}}\left[\frac{Pr\left(\mathbf{y}=\mathbf{k}|\mathbf{x}^0,\mathbf{q}\right)}{Pr\left(\mathbf{y}_b=\mathbf{k}_b|\mathbf{x}^0\right)}\right] \\
&= E_{\mathbf{q}_a}\left[\frac{Pr\left(\mathbf{y}=\mathbf{k}|\mathbf{x}^0,\mathbf{q}_a\right)}{Pr\left(\mathbf{y}_b=\mathbf{k}_b|\mathbf{x}^0\right)}\right] \\
&= E_{\mathbf{q}_a}\left[\frac{Pr\left(\mathbf{y}=\mathbf{k}|\mathbf{x}^0,\mathbf{q}_a\right)}{Pr\left(\mathbf{y}_b=\mathbf{k}_b|\mathbf{x}^0,\mathbf{q}_a\right)}\right] \\
&= E_{\mathbf{q}_a}\left[Pr\left(\mathbf{y}_a=\mathbf{k}_a|\mathbf{y}_b=\mathbf{k}_b,\mathbf{x},\mathbf{q}_a\right)\right],
\end{aligned}
$$

which along with Proposition 8 proves the result. □

Example 10. *Consider a trivariate probit model where we maintain A.1-A.5. We may want to estimate the APE of* y_3 *on the conditional probability of any the four combined outcomes for* $(y_1, y_2) =$

$\mathbf{y}_a$:

$$APE_{y_3}\left(\mathbf{k}_a, \mathbf{x}^0\right) = E_{\mathbf{q}_a}\left[Pr\left(\mathbf{y}_a = \mathbf{k}_a | y_3 = 1, \mathbf{x}^0, \mathbf{q}_a\right) - Pr\left(\mathbf{y}_a = \mathbf{k}_a | y_3 = 0, \mathbf{x}^0, \mathbf{q}_a\right)\right].$$

where $\mathbf{k}_a \in \{(1,0),(1,1),(0,1),(0,0)\}$. *This makes sense both in a model with all strictly exogenous variables and in a recursive model. To estimate* $APE_{y_3}\left(\mathbf{k}_a, \mathbf{x}^0\right)$, *we first invoke Equation (27), which establishes*

$$APE_{y_3}\left(\mathbf{k}_a, \mathbf{x}^0\right) = E_{\mathbf{w}}\left[Pr\left(\mathbf{y}_a = \mathbf{k}_a | y_3 = 1, \mathbf{x}^0, \mathbf{w}\right) - Pr\left(\mathbf{y}_a = \mathbf{k}_a | y_3 = 0, \mathbf{x}^0, \mathbf{w}\right)\right],$$

then we use consistent estimates of coefficients and covariances to estimate $Pr\left(\mathbf{y}_a = \mathbf{k}_a | y_3, \mathbf{x}^0, \mathbf{w}\right)$ *and finally we average over the sample*

$$\widehat{APE_{y_3}\left(\mathbf{k}_a, \mathbf{x}^0\right)} = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{\Phi_{\hat{R}_1}\left(s_{k_1}\hat{h}^0_{1,i}, s_{k_2}\hat{h}^0_{2,i}, \mathbf{x}^{0\prime}\hat{\boldsymbol{\beta}}_3 + \mathbf{w}_i'\hat{\boldsymbol{\delta}}_3\right)}{\Phi\left(\mathbf{x}^{0\prime}\hat{\boldsymbol{\beta}}_3 + \mathbf{w}_i'\hat{\boldsymbol{\delta}}_3\right)} - \frac{\Phi_{\hat{R}_0}\left(s_{k_1}\hat{h}^0_{1,i}, s_{k_2}\hat{h}^0_{2,i}, -\mathbf{x}^{0\prime}\hat{\boldsymbol{\beta}}_3 - \mathbf{w}_i'\hat{\boldsymbol{\delta}}_3\right)}{1 - \Phi\left(\mathbf{x}^{0\prime}\hat{\boldsymbol{\beta}}_3 + \mathbf{w}_i'\hat{\boldsymbol{\delta}}_3\right)}\right]$$

where [3]

$$\hat{R}_1 \equiv \begin{pmatrix} 1 & . & . \\ s_{k_2}s_{k_1}\hat{\rho}_{12} & 1 & . \\ s_{k_1}\hat{\rho}_{31} & s_{k_2}\hat{\rho}_{32} & 1 \end{pmatrix} \text{ and } \hat{R}_0 \equiv \begin{pmatrix} 1 & . & . \\ s_{k_2}s_{k_1}\hat{\rho}_{12} & 1 & . \\ -s_{k_1}\hat{\rho}_{31} & -s_{k_2}\hat{\rho}_{32} & 1 \end{pmatrix}.$$

Remark 11. An extreme case covered by Proposition 9 is when $\mathbf{q}_a$ exhausts the whole amount of LH in the model, that is $\mathbf{q}_a = \mathbf{q}$. It is also clear that common variables in $\mathbf{q}_a$ and $\mathbf{q}_b$ violates A.5 and so are a problem for the identification of APEs.

Remark 12. In the context of the multivariate probit model with all exogenous regressors, one can further restrict the model by maintaining also A.3($\mathbf{y}_a$): $Pr\left(\mathbf{y}_a = \mathbf{k}_a | \mathbf{x}, \mathbf{q}\right) = Pr\left(\mathbf{y}_a = \mathbf{k}_a | \mathbf{x}, \mathbf{q}_a\right)$ for any $\mathbf{k}_a$ and A.4($\mathbf{q}_a$): $D\left(\mathbf{q}_a | \mathbf{w}\right) = D\left(\mathbf{q}_a\right)$. This makes it possible to reverse the analysis, conditioning on $\mathbf{y}_a$ (notice that in this case $D\left(\mathbf{q}|\mathbf{x}\right) = D\left(\mathbf{q}\right)$, making the conditioning on $\mathbf{w}$ useless). In the recursive

[3] This example is couched in terms of a model with all strictly exogenous regressors. To be consistent with a recursive model, one has simply to adjust the notation to allow that $\left(\hat{h}^0_{1,i}, \hat{h}^0_{2,i}\right)$ be different between $\Phi_{\hat{R}_1}$ and $\Phi_{\hat{R}_0}$. See the end of Remark 12 for the reason why this is so.

model, such reverse conditioning is both useless and arbitrary. It is useless since interest centers only on the effects of $\mathbf{y}_b$ on $\mathbf{y}_a$. It is arbitrary since $Pr(\mathbf{y}_a = \mathbf{k}_a|\mathbf{x}, \mathbf{q})$ does depend on $\mathbf{q}_b$ if the latter is part of the $\mathbf{y}_b$ sub-system. This can be easily seen in the bivariate recursive model:

$$
\begin{aligned}
Pr(y_1 = 1|\mathbf{x}, \mathbf{q}) &= Pr(y_1 = 1|y_2 = 1, \mathbf{x}, \mathbf{q}) \, Pr(y_2 = 1|\mathbf{x}, q_2) + \\
&\quad Pr(y_1 = 1|y_2 = 0, \mathbf{x}, \mathbf{q}) \, Pr(y_2 = 0|\mathbf{x}, q_2) \\
&= \Phi_{C(\rho_{12})}(\alpha_1 + \mathbf{x}'\boldsymbol{\beta}_1 + \lambda + q_1, \alpha_2 + \mathbf{x}'\boldsymbol{\beta}_2 + q_2) + \\
&\quad \Phi_{C(-\rho_{12})}(\alpha_1 + \mathbf{x}'\boldsymbol{\beta}_1 + q_1, -\alpha_2 - \mathbf{x}'\boldsymbol{\beta}_2 - q_2),
\end{aligned}
$$

$\mathbf{q} = (q_1, q_2)$, where the complication lies in the fact that, because of the λ term, the two bivariate normal distributions have a different first argument and so the sum of the two distributions does not collapse to a univariate normal, as it otherwise would in the standard bivariate model.

Remark 13. In recursive models *average treatment effects* (ATEs) are typically of interest. In the bivariate model, for example, we may estimate

$$
ATE_{y_2}\left(\mathbf{k}_1, \mathbf{x}^0\right) = E_{q_1}\left\{Pr\left(y_1^1 = 1|\mathbf{x}^0, q_1\right) - Pr\left(y_1^0 = 1|\mathbf{x}^0, q_1\right)\right\},
$$

where $y_1^1 = \mathbf{1}(\varepsilon_1 > -\alpha_1 - \mathbf{x}'\boldsymbol{\beta}_1 - \lambda - q_1)$ and $y_1^0 = \mathbf{1}(\varepsilon_1 > -\alpha_1 - \mathbf{x}'\boldsymbol{\beta}_1 - q_1)$ (Wooldridge 2010, pp. 586, and Bhattacharya, Goldman, and McCaffrey (2006) report analogous formulas without the q_1 component; Entorf (2012) instead focuses on coefficient estimates and APEs based on marginal and conditional probabilities, without reporting the formulas used; Fichera and Sutton (2011) do not report the formulas for their partial effects). $ATE_{y_2}\left(\mathbf{k}_1, \mathbf{x}^0\right)$ is a counterfactual APE since both y_1^1 and y_1^0 are latent. It measures the population-averaged difference in response probability between the two hypothetical situations of treatment and non-treatment for the same random draw. In the bivariate recursive model, the ATE is based on the marginal probabilities of y_1^1 and y_1^0. If $m > 2$ the ATEs may involve marginal as well as joint probabilities. In either situation the results in Sections 2 and 3 apply assuring identification of ATEs under just A.1 and A.2. Greene (1998) tests y_2 being exogenous in a bivariate model and does not reject it, so $Pr\left(y_1^1 = 1|\mathbf{x}^0, q_1\right) = Pr\left(y_1 = 1|y_2 = 1, \mathbf{x}^0, q_1\right)$ and

$Pr\left(y_1^0 = 1|\mathbf{x}^0, q_1\right) = Pr\left(y_1 = 1|y_2 = 0, \mathbf{x}^0, q_1\right)$ and the two concepts would coincide therein.

Remark 14. It is well known that the control function approach can be used for estimation and inference in probit models with continuous endogenous regressors (Rivers and Vuong (1988), Wooldridge 2010, pp. 585-594). As already observed in Subsection 3.4, the recursive probit model provides a computationally inexpensive method to accommodate binary endogenous regressors. This section's results permit to combine the two methods within a general procedure that accommodates binary and continuous endogenous variables at the same time, allowing consistent estimation of ATEs and APEs. Consider, for example, the bivariate recursive model with

$$\begin{aligned} y_1^* &= \alpha_1 + \gamma y_2 + \beta_1 x + q_1 + \varepsilon_1, \\ y_2^* &= \alpha_2 + \mathbf{z}'\boldsymbol{\beta}_2 + q_2 + \varepsilon_2, \end{aligned}$$

$y_j = \mathbf{1}\left(y_j^* > 0\right)$, $j = 1, 2$, $x = \mathbf{z}'\boldsymbol{\beta}_3 + w$, $\boldsymbol{\varepsilon}|\mathbf{z}, x, w, \mathbf{q} \sim N\left(\mathbf{0}, R\right)$ and $\mathbf{q}|\mathbf{z}, x, w \sim N\left(\boldsymbol{\mu}, \Omega\right)$,

$$R \equiv \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

$$\boldsymbol{\mu} \equiv \begin{pmatrix} \eta_1 + \delta_1 w \\ \eta_2 \end{pmatrix}$$

and

$$\Omega \equiv \begin{pmatrix} \omega_1 & 0 \\ 0 & \omega_2 \end{pmatrix}.$$

In the structural equation of interest

$$y_1^* = \alpha_1 + \gamma y_2 + \beta_1 x + q_1 + \varepsilon_1,$$

y_2 and x are both endogenous because, respectively, $\rho \neq 0$ and the latent variable w is a common factor for q_1 and x. Since assumptions A.1-A.5 are verified, APEs based $E_{q_1}\left[Pr\left(y_1|y_2, x^0, \mathbf{z}^0, q_1\right)\right]$ can be consistently estimated into two steps: Step 1 estimates w using the residuals from the OLS regression of x on $\mathbf{z}$, $\hat{w}$. Step 2 implements a recursive bivariate model with explanatory variables $y_2, x, \mathbf{z}$ and $\hat{w}$ and computes the APEs using the conditional probability estimates averaged over the sample values of $\hat{w}$. The ATE of y_2 can be computed similarly.

5 A simple estimation procedure

Mullahy (2011) derives the formula of the PE of the joint probability in m-variate probit models, $\partial_x \Phi_{C(.)}(.)$, and uncovers that it involves evaluation of a multivariate cumulative normal over $m-1$ dimensions only. Equation (17) shows that this computational benefit is retained by APEs with conditional independent heterogeneity and arbitrary covariance matrix, provided that in the PE formula the scaled coefficients and the scaled and repositioned covariances replace, respectively, the model coefficients and the model covariances. In addition, given **Results 1**, **2** and Lemma 2, a bivariate probit model made by any two equations i and j of System (15) will provide consistent estimates of $(\alpha_h + \eta_h)/\sqrt{1+\omega_{hh}}$, $\boldsymbol{\beta}_h/\sqrt{1+\omega_{hh}}$, $\boldsymbol{\delta}_h/\sqrt{1+\omega_{hh}}$, $h = i, j$ and ξ_{ij}. This suggests a simple two-step procedure for estimating APEs, which partly dispenses with simulation algorithms.

Step 1. Carry out all possible $m' \equiv m(m-1)/2$ bivariate probit regressions to provide consistent Quasi-ML estimators of all parameters of interest. An estimate of the combined covariance matrix of the foregoing estimators, $\widehat{V}$, is obtained by applying the covariance estimator in White (1982). This yields m' covariance estimates and, for each equation $j = 1, ..., m$ of System (15), $m-1$ vectors of coefficient estimates, of which only one vector is selected at random for use in Step 2.

Step 2. Assemble the estimated APE as in Equation (17) by plugging the m' covariance estimates and the m vectors of coefficient estimates retained in *Step 1* into $\partial_x \Phi_{C\left(s_{k_1} s_{k_j} \rho_{ij}\right)}\left(s_{k_1} h^0_{1,i}, ..., s_{k_m} h^0_{m,i}\right)$ using the formulas in Mullahy (2011) for $\partial_x \Phi_{C(.)}(.)$ and then average the resulting values over the sample.

We obtain the variance of the the estimated APEs through the delta method as suggested by Greene (2012), pp. 738-739. We first work out the $(m \cdot k + m') \times 1$ vector of partial derivatives of a given APE with respect to the $k \cdot m$ coefficients and the m' covariances, $\boldsymbol{\partial} APE_i$, $i = 1, ..., n$. Then, we evaluate all $\boldsymbol{\partial} APE_i$ in the sample at the parameter estimates to get $\widehat{\boldsymbol{\partial} APE}_i$ and average all the $\widehat{\boldsymbol{\partial} APE}_i$ values over the sample to get the $(m \cdot k + m') \times 1$ column vector $\widehat{\boldsymbol{\partial} APE} = (1/n) \sum_1^n \widehat{\boldsymbol{\partial} APE}_i$. Finally, we evaluate the variance of the estimated APE as $\widehat{\boldsymbol{\partial} APE}' \widehat{V} \widehat{\boldsymbol{\partial} APE}$, where $\widehat{V}$ is the covariance matrix obtained in *Step 1.*

Remark 15. Step 1 dispenses completely with simulation algorithms and is close in spirit to the GMM approach of Bertschek and Lechner (1998), which is based on the m marginal univariate normal distributions and estimates covariances in a subsequent step. Here, we focus on marginal bivariate normal distributions and so coefficients and covariances are jointly estimated. Step 2 always permits to exploit in full the dimensional benefit evidenced in Mullahy (2011), where it is proved that the PE formulas for m-variate probit models involve only $(m-1)$-variate normal distributions. Hence, in trivariate models simulation algorithms can be avoided altogether, as the next example shows.

Example 16. *The APE of the generic regressor x in the trivariate probit model with conditionally independent heterogeneity is*

$$\begin{aligned} APE_i(\mathbf{k}, \mathbf{x}^o) &= E_{\mathbf{w}} \left[\phi\left(s_{k_1} h_1^0\right) \Phi_{C(\varrho_{23}(\mathbf{k}))} \left(\frac{s_{k_2} h_2^0 - \xi_{12}(\mathbf{k})\, s_{k_1} h_1^0}{\sqrt{1-\xi_{12}^2}}, \frac{s_{k_3} h_3^0 - \xi_{13}(\mathbf{k})\, s_{k_1} h_1^0}{\sqrt{1-\xi_{13}^2}} \right) \frac{\beta_{1i}}{\sqrt{1+\omega_{11}}} \right. \\ &+ \phi\left(s_{k_2} h_2^0\right) \Phi_{C(\varrho_{13}(\mathbf{k}))} \left(\frac{s_{k_1} h_1^0 - \xi_{12}(\mathbf{k})\, s_{k_2} h_2^0}{\sqrt{1-\xi_{12}^2}}, \frac{s_{k_3} h_3^0 - \xi_{23}(\mathbf{k})\, s_{k_2} h_2^0}{\sqrt{1-\xi_{23}^2}} \right) \frac{\beta_{2i}}{\sqrt{1+\omega_{22}}} \\ &+ \left. \phi\left(s_{k_3} h_3^0\right) \Phi_{C(\varrho_{12}(\mathbf{k}))} \left(\frac{s_{k_1} h_1^0 - \xi_{13}(\mathbf{k})\, s_{k_3} h_3^0}{\sqrt{1-\xi_{13}^2}}, \frac{s_{k_2} h_2^0 - \xi_{23}(\mathbf{k})\, s_{k_3} h_3^0}{\sqrt{1-\xi_{23}^2}} \right) \frac{\beta_{3i}}{\sqrt{1+\omega_{33}}} \right] \quad (29) \end{aligned}$$

where

$$\begin{aligned} \varrho_{23}(\mathbf{k}) &= \xi_{23}(\mathbf{k}) - \xi_{12}(\mathbf{k})\,\xi_{13}(\mathbf{k}) \\ \varrho_{13}(\mathbf{k}) &= \xi_{13}(\mathbf{k}) - \xi_{12}(\mathbf{k})\,\xi_{23}(\mathbf{k}) \\ \varrho_{12}(\mathbf{k}) &= \xi_{12}(\mathbf{k}) - \xi_{13}(\mathbf{k})\,\xi_{23}(\mathbf{k}) \\ \xi_{ij}(\mathbf{k}) &= s_{k_i} s_{k_j} \xi_{ij},\ i, j = 1, 2, 3\ i \neq j, \end{aligned}$$

and ξ_{ij} is defined as in (14). Inspection of Equation (29) shows that only the univariate Normal density *and the bivariate Normal distribution are required. Hence, consistent estimator of* $APE_i(\mathbf{k}, \mathbf{x}^o)$ *can be obtained without having to evaluate trivariate normal distributions.*

The procedure is simply implemented through official commands in Stata and is referred to as `combined biprobit`, throughout. Each of the m' bivariate probit regressions in *Step 1* is executed through `biprobit` and the m' estimation results are then combined together through `suest` to get the estimated covariance matrix $\widehat{V}$ into memory. The APE and standard error formulas of *Step 2* can then be either processed through the command `predictnl` or from within the Mata environment in Stata.

Remark 17. The `combined biprobit` procedure is not implementable in recursive models, since necessarily at least one of the bivariate subsystem that are needed for estimating covariances is not recursive, that is with a response variable in the right hand side of the first equation and a second equation that is not peculiar to that response variable. In this case simulation-based estimators are indispensable.

6 Monte Carlo experiments

We now report on two distinct batteries of Monte Carlo experiments. The first battery focuses on the estimation of APEs in a trivariate probit model and evaluates the finite-sample performances of the `combined biprobit` procedure of Subsection 5 in comparison with two popular simulation based Stata codes, `mvprobit` and `cmp`. The second battery estimates the finite sample bias of APEs of conditional probabilities under the assumptions spelled out in Section 4.

We generate multivariate normal variables using a result in Rao (1973):

Lemma 18. *given the* $m \times 1$ *random vector* $\mathbf{z}$*, then*

$$\mathbf{z} \sim N(\boldsymbol{\mu}, \Sigma)$$

if and only if

$$\mathbf{z} = \boldsymbol{\mu} + B\mathbf{u},$$

where $\mathbf{u}$ *is a* $p \times 1$ *random vector such that* $\mathbf{u} \sim N(\mathbf{0}, I)$*,* B *is an* $m \times p$ *matrix with* p *rank and* $\Sigma = BB'$*.*

6.1 APEs in trivariate probit

We consider a trivariate probit model with independent, or conditionally independent, LH.

We first generate a vector of six (pseudo) independent standard normal variables $(u_1\ u_2\ u_3\ u_4\ u_5\ u_6)' \sim N(\mathbf{0}, I_6)$. Then, we set $\boldsymbol{\varepsilon} = B(u_1\ u_2\ u_3)'$ and $\boldsymbol{\nu} = B(u_4\ u_5\ u_6)'$, where

$$B = \begin{pmatrix} 0.6 & -0.8 & 0 \\ -0.8 & 0.6 & 0 \\ 0.8 & 0 & 0.6 \end{pmatrix}.$$

From Lemma 18, $\boldsymbol{\varepsilon}$ and $\boldsymbol{\nu}$ are independent random numbers from the same normal distribution

$$N\left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & & \\ -0.96 & 1 & \\ 0.48 & -0.64 & 1 \end{pmatrix}\right].$$

We also specify the latent regression model

$$y_j^* = 2 + x + q_j + \varepsilon_j, \tag{30}$$

and define $y_j = \mathbf{1}\left(y_j^* > 0\right)$, $j = 1, 2, 3$, where $q_j = w + \nu_j$,

$$\begin{pmatrix} x \\ w \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \\ \tau & 1 \end{pmatrix}\right],$$

ν_j and ε_j are independent and both independent of $(x\ w)$ and τ alternates between 0 and 0.5. Therefore, x and q_j are either independent, if $\tau = 0$, or only independent conditional on w if $\tau = 0.5$. It also has that

$$R + \Omega = 2 \begin{pmatrix} 1 & & \\ -0.96 & 1 & \\ 0.48 & -0.64 & 1 \end{pmatrix}$$

and so the normalized coefficient on x in all equations is $1/\sqrt{2}$, while the normalized covariances, ξ_{ij}, are exactly equal to the original ones, so $\xi_{21} = \xi_{32} = -0.96$ and $\xi_{31} = 0.48$. We consider samples of 500, 1000 and 10000 observations. The number of simulation points in `mvprobit` alternates among 5 (the default), 10 and 100. True and estimated APEs of x are averaged over the sample. Monte Carlo biases and root mean squared errors for the APE and the three coefficients on x, along with the three covariances, are computed through 1000 replications. Results for independent ($\tau = 0$) and conditionally independent ($\tau = 0.5$) heterogeneity are reported in Tables 1 and 2, respectively.

Table 1: Monte Carlo biases and RMSEs for APEs, regression coefficients and covariances with independent heterogeneity ($\tau = 0$)

n ╲ parameters	Combined biprobit							Trivariate mvprobit						
	APE	β_1	β_2	β_3	cov12	cov13	cov23	APE	β_1	β_2	β_3	cov12	cov13	cov23
	BIAS[a]													
n=500	0.000	0.013	0.008	0.012	-0.013	-0.000	-0.010	0.021	0.132	0.012	0.056	0.447	-0.114	0.248
n=1000	0.001	0.003	0.006	0.004	-0.007	-0.004	-0.006	-0.023	0.012	0.115	-0.065	0.312	-0.065	0.122
n=10000	-0.000	0.001	0.000	0.000	-0.001	-0.000	-0.001	-0.007	-0.006	-0.049	-0.003	0.325	-0.060	0.170
n ╲ parameters	**RMSE**[a]													
n=500	0.016	0.097	0.093	0.098	0.038	0.096	0.107	0.021	0.132	0.012	0.056	0.447	0.114	0.248
n=1000	0.011	0.066	0.067	0.066	0.030	0.070	0.076	0.023	0.012	0.115	0.065	0.312	0.065	0.122
n=10000	0.003	0.021	0.020	0.021	0.009	0.021	0.023	0.007	0.006	0.049	0.003	0.038	0.096	0.107
n ╲ parameters	**Computation time per replication**													
n=500	0.5”							3.4”						
n=1000	0.5”							3.8”						
n=10000	1.7”							22.0”						

[a]1000 replications

Table 2: Monte Carlo biases and RMSEs for APEs, regression coefficients and covariances with conditionally independent heterogeneity ($\tau = 0.5$)

n ╲ parameters	`Combined biprobit`							`Trivariate mvprobit`						
	APE	β_1	β_2	β_3	cov12	cov13	cov23	APE	β_1	β_2	β_3	cov12	cov13	cov23
	BIAS[a]													
n=500	0.001	0.008	0.011	0.009	-0.011	0.001	-0.012	0.009	0.012	-0.034	-0.027	0.390	-0.160	0.182
n=1000	0.001	0.002	0.011	0.005	-0.008	-0.004	-0.005	-0.019	-0.074	-0.011	0.010	0.265	-0.034	0.126
n=10000	-0.000	0.001	-0.001	0.000	-0.001	-0.001	-0.001	0.007	0.007	-0.037	-0.005	0.297	-0.056	0.175
n ╲ parameters	**RMSE**[a]													
n=500	0.016	0.107	0.105	0.106	0.037	0.095	0.099	0.009	0.012	0.034	0.027	0.390	0.160	0.182
n=1000	0.012	0.074	0.074	0.075	0.029	0.068	0.071	0.019	0.074	0.011	0.010	0.265	0.034	0.126
n=10000	0.004	0.022	0.024	0.024	0.009	0.022	0.022	0.007	0.007	0.037	0.005	0.297	0.056	0.175
n ╲ parameters	**Computation time per replication**													
n=500	0.4”							2.8”						
n=1000	0.4”							3.9”						
n=10000	1.8’							25.7”						

[a] 1000 replications

Monte Carlo results decidedly support the `combined biprobit` procedure. For all sample sizes and for both $\tau = 0$ (Table 1) and $\tau = 0.5$ (Table 2), `combined biprobit` yields APE estimates that almost always are virtually unbiased and, anyhow, with smaller bias than `mvprobit`. While the `mvprobit` estimates have always negligible standard deviations, with RMSEs that virtually equal biases, the latter are far from being negligible to the extent that the RMSE of the `combined biprobit` APE estimates are almost always smaller than `mvprobit`'s. We also notice that the `mvprobit` finite-sample performance seems particularly poor when it comes to the bias of estimated coefficients and, especially, estimated covariances. Finally, `combined biprobit` is from 7 to 14 times faster than `mvprobit`, with computing times that are also less sensitive to sample sizes.

The `mvprobit` estimates reported in Tables 1 and 2 have been obtained using 5 simulations points (draws in the `mvprobit` jargon), which is the default option for the procedure. At the cost of enormously higher computing times, we have also used 10 and 100 simulation points, observing a steady, although very slow, improvement in accuracy. With a sample size of 1000, the magnitude of the bias decreases from 0.023, to 0.020 and 0.012 as the number of draws increases from 5 to 10 and 100, respectively (see Table 5 in appendix). We therefore conjecture that only with a number of draws close to the sample size, which may be hardly feasible in many circumstances, the finite-sample bias of `mvprobit` would be comparable to that of `combined biprobit`.

As a further check, we have also tried the recent `cmp` code by Roodman (2011), a general procedure for non-linear seemingly unrelated equations that can estimate multivariate probit models based on simulations. Table 6 shows that, although with a largely better performance than `mvprobit`, `cmp` is still behind `combined biprobit` in all respects.

6.2 Conditioning on response subvectors: experiments on bivariate probit

We here consider a bivariate probit model with independent, or conditionally independent, LH. We base the random part of the model on a vector of six (pseudo) independent standard normal variables $(u_1\ u_2\ u_3\ u_4\ u_5\ u_6)' \sim N(\mathbf{0}, I_6)$. Then, we set $\boldsymbol{\varepsilon} \equiv (\varepsilon_1\ \varepsilon_2)' = B_{\varepsilon}(u_1\ u_2)'$ and $\boldsymbol{\nu} \equiv (\nu_1\ \nu_2)' =$

$B_\nu\,(u_3\; u_4)'$ and $(x\; w)' = B_\kappa\,(u_5\; u_6)'$ where

$$B_\varepsilon = \begin{pmatrix} 0.6 & -0.8 \\ -0.8 & 0.6 \end{pmatrix},\; B_\nu = \begin{pmatrix} 1 & 0 \\ c & 1.5 \end{pmatrix},\; B_\kappa = \begin{pmatrix} 1 & 0 \\ 0.8 & 0.6 \end{pmatrix},$$

From Lemma 18 $\boldsymbol{\varepsilon}$, $\boldsymbol{\nu}$ and $(x\; w)'$ are independent and

$$\boldsymbol{\varepsilon} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \\ -0.96 & 1 \end{pmatrix}\right],\; \boldsymbol{\nu} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \\ c & c^2 + 2.25 \end{pmatrix}\right],$$

$$\begin{pmatrix} x \\ w \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \\ 0.8 & 1 \end{pmatrix}\right].$$

We also specify the bivariate latent regression model

$$\begin{aligned} y_1^* &= 1 + \lambda y_2 + \beta x + q_1 + \varepsilon_1, \\ y_2^* &= 1 + x + q_2 + \varepsilon_2 \end{aligned}$$

with $y_j = \mathbf{1}\left(y_j^* > 0\right)$, $j = 1, 2$, and

$$\begin{aligned} q_1 &= \delta w + \nu_1, \\ q_2 &= \nu_2. \end{aligned}$$

We wish to evaluate models with and without right-hand-side endogenous binary variables, therefore we let λ alternate between 0 and 1 with $\beta = 1 - \lambda$. A.1-A.4 are always satisfied and so is A.5 when $c = 0$. Thus, letting c vary across 0, 0.5, 1, 2 and 10 permits to evaluate the impact of departures from A.4 in the second equation. We also let δ alternate between 0 and 1 to shift from independence to conditional independence of x and q_1.

The population object of interest is $E_{q_1}\left[Pr\left(y_1=1|y_2=1,x^0,q_1\right)\right]$:

$$
\begin{aligned}
E_{q_1}\left[Pr\left(y_1=1|y_2=1,x^0,q_1\right)\right] &= \\
E_{q_1}\left[\frac{Pr\left(y_1=1,y_2=1|x^0,q_1\right)}{Pr\left(y_2=1|x^0,q_1\right)}\right] &= \\
E_{q_1}\left\{\frac{\Phi_{C(\xi)}\left[1+\lambda+\beta x^0+q_1,\left[1+x^0+c\left(q_1-0.8\delta x^0\right)/\left(1+0.36\delta^2\right)\right]/\kappa\right]}{\Phi\left[\left(1+x^0+c\left(q_1-0.8\delta x^0\right)/\left(1+0.36\delta^2\right)\right)/\kappa\right]}\right\}. && (31)
\end{aligned}
$$

where $\xi=-0.96/\kappa$ and

$$\kappa=\sqrt{3.25+\frac{0.36c^2\delta^2}{1+0.36\delta^2}}.$$

When $\delta=0$ and $c=0$, then q_1 and (y_2,x) are independent and so $E_{q_1}\left[Pr\left(y_1=1|y_2=1,x^0,q_1\right)\right]=Pr\left(y_1=1|y_2=1,x^0\right)$, or

$$E_{q_1}\left[Pr\left(y_1=1|y_2=1,x^0,q_1\right)\right]=\frac{\Phi_{C\left(\frac{-0.96}{\sqrt{6.5}}\right)}\left[\left(1+\lambda+\beta x^0\right)/\sqrt{2},\left(1+x^0\right)/\sqrt{3.25}\right]}{\Phi\left[\left(1+x^0\right)/\sqrt{3.25}\right]}. \quad (32)$$

As discussed in Subsection 3.4, regardless of y_2 being endogenous in the first equation, conventional bivariate probit estimation yields consistent estimators of the scaled coefficients and the scaled-repositioned covariance for any values of δ and c and so of conditional probabilities. Let $\hat{P}_{11}\left(x^0,w_i\right)$ denote the bivariate probit estimate of $Pr\left(y_1=1|y_2=1,x^0,w_i\right)$ and let

$$\widehat{EP}_{11}^{0}\equiv(1/n)\sum_{i=1}^{n}\hat{P}_{11}\left(x^0,w_i\right).$$

Given Proposition 8, $\widehat{EP}_{11}^0$ is a consistent estimator of

$$\frac{E_{\mathbf{q}}\left[Pr\left(y_1=1,y_2=1|x^0,q_1,q_2\right)\right]}{E_{q_2}\left[Pr\left(y_2=1|x^0,q_2\right)\right]} =$$
$$\frac{E_w\left[Pr\left(y_1=1,y_2=1|x^0,w\right)\right]}{Pr\left(y_2=1|x^0\right)} =$$
$$\frac{E_w\left\{\Phi_{C\left(\frac{-0.96}{\sqrt{2(c^2+3.25)}}\right)}\left[\left(1+\lambda+\beta x^0+\delta w\right)/\sqrt{2},\left(1+x^0\right)/\sqrt{c^2+3.25}\right]\right\}}{\Phi\left[\left(1+x^0\right)/\sqrt{c^2+3.25}\right]}. \tag{33}$$

where the first equality is implied by independence of q_2 and x in our data generating process and the last by the fact that A.1 and A.2 hold. The foregoing expression coincides with the population object of interest in Equation (31) if and only if $c=0$. When $\delta=0$, Equation (33) simplifies to

$$\frac{E_{\mathbf{q}}\left[Pr\left(y_1=1,y_2=1|x^0,q_1,q_2\right)\right]}{E_{q_2}\left[Pr\left(y_2=1|x^0,q_2\right)\right]} =$$
$$\frac{\Phi_{C\left(\frac{-0.96}{\sqrt{c^2+3.25}}\right)}\left[\left(1+\lambda+\beta x^0\right)/\sqrt{2},\left(1+x^0\right)/\sqrt{c^2+3.25}\right]}{\Phi\left[\left(1+x^0\right)/\sqrt{c^2+3.25}\right]}. \tag{34}$$

We consider two bivariate models: with only exogenous regressors ($\beta=1$, $\lambda=0$, $\delta=0$) and with a binary endogenous regressor ($\beta=0$, $\lambda=1$, $\delta=(0,1)$). We fix $x_0=1$ in either model.

Table 3 reports the finite-sample bias of $\widehat{EP}_{11}^0$ with respect to the population object of interest:

$$BIAS=E\left(\widehat{EP}_{11}^0\right)-E_{q_1}\left[Pr\left(y_1=1|y_2=1,x^0,q_1\right)\right]$$

as well as the finite-sample bias with respect to the probability limit of $\widehat{EP}_{11}^0$:

$$BIAS_{plim}=E\left(\widehat{EP}_{11}^0\right)-\frac{E_{\mathbf{q}}\left[Pr\left(y_1=1,y_2=1|x^0,q_1,q_2\right)\right]}{E_{q_2}\left[Pr\left(y_2=1|x^0,q_2\right)\right]}.$$

We consider estimation samples of 500, 1000 and 10000 observations and compute $BIAS$, $BIAS_{plim}$, $plim\left(\widehat{EP}_{11}^0\right)$ when $\delta\neq 0$, and $E_{q_1}\left[Pr\left(y_1=1|y_2=1,x^0,q_1\right)\right]$ when $c\neq 0$ or $\delta\neq 0$, as Monte Carlo averages over 10000 replications (when $\delta=0$, $plim\left(\widehat{EP}_{11}^0\right)$ is exactly computed through Equation

(34) and, for both models, takes on the values reported in Table **??** in appendix; when $\delta = 0$ and $c = 0$, $E_{q_1}\left[Pr\left(y_1 = 1|y_2 = 1, x^0, q_1\right)\right]$ equals $plim\left(\widehat{EP}^0_{11}\right)$ [4].)

Table 3: Biases in average conditional probabilities

Bivariate model with a unique exogenous regressor:

$$\begin{aligned} y_1^* &= 1 + x + q_1 + \varepsilon_1 \\ y_2^* &= 1 + x + q_2 + \varepsilon_2 \end{aligned}$$

	n=500		n=1000		n=10000	
(c, δ)	$BIAS$	$BIAS_{plim}$	$BIAS$	$BIAS_{plim}$	$BIAS$	$BIAS_{plim}$
$(0, 0)^a$	0.000	0.000	0.000	0.000	0.000	0.000
$(0.5, 0)$	0.005	0.000	0.004	0.000	0.004	0.000
$(1, 0)$	0.012	0.000	0.012	0.000	0.012	0.000
$(2, 0)$	0.031	0.000	0.030	0.000	0.030	0.000
$(10, 0)$	0.069	-0.000	0.069	-0.000	0.070	-0.000

Bivariate model with a binary endogenous regressor:

$$\begin{aligned} y_1^* &= 1 + y_2 + q_1 + \varepsilon_1 \\ y_2^* &= 1 + x + q_2 + \varepsilon_2 \end{aligned}$$

(c, δ)	$BIAS$	$BIAS_{plim}$	$BIAS$	$BIAS_{plim}$	$BIAS$	$BIAS_{plim}$
$(0, 0)^a$	-0.001	-0.001	-0.001	-0.001	-0.000	-0.000
$(0.5, 0)$	0.009	-0.001	0.010	-0.000	0.010	-0.000
$(1, 0)$	0.024	-0.001	0.025	-0.000	0.025	0.000
$(2, 0)$	0.063	-0.000	0.063	-0.000	0.063	0.000
$(10, 0)$	0.268	-0.000	0.268	-0.000	0.268	-0.000
$(0, 1)$	-0.004	-0.004	-0.002	-0.002	-0.000	-0.000
$(0.5, 1)$	0.014	-0.003	0.015	-0.002	0.017	-0.000
$(1, 1)$	0.036	-0.002	0.038	-0.001	0.038	-0.000
$(2, 1)$	0.081	-0.002	0.082	-0.001	0.083	-0.000
$(10, 1)$	0.134	-0.001	0.135	-0.000	0.135	0.000

[a]In this case $BIAS = BIAS_{plim}$

Results are clear-cut. We proved in Section 4 that restricting LH to be independent across equations ($c = 0$) results in consistent estimators for conditional probabilities, and so for the corresponding APEs. We find here that this translates into a negligible finite sample bias ($BIAS$), which always declines with the size of the estimation sample and becomes virtually zero when $n = 10000$. While $BIAS$ remains

[4]For the sake of validation of our Monte Carlo computations, we evaluated Monte Carlo averages also for the cases where exact results were available, with virtually zero differences. These further Monte Carlo results are available on request from the Authors.

quite small when $c = 0.5$, it increases steadily with c. $BIAS_{plim}$ is always negligible, virtually zero when $n = 10000$. The foregoing Monte Carlo findings hold regardless of the binary endogenous variable being or not included into the first equation, as well as regardless of x and q_1 being independent or conditionally independent. The Monte Carlo results also confirm a virtually zero bias in the estimate of the scaled coefficient as well as the scaled and re-positioned covariance (unreported, but available on request from the authors).

Bhattacharya, Goldman, and McCaffrey (2006) conduct similar Monte Carlo experiments, focusing on ATEs with no LH and restricting to the case of $\lambda \neq 0$. They find that bivariate probit always outperforms competitor estimators, showing virtually zero biases in estimated PEs. Thus, our results confirm that the excellent finite-sample performance of the bivariate probit model is robust to conditionally independent heterogeneity, provided that the covariance between the omitted variables in the two `biprobit` equations is small, if not zero. We stress that the latter qualification has to be met also for the seemingly harmless bivariate probit model with no endogenous regressors (see Remark 12).

7 An application to immigrants' models of trans-national ethnic identities

As an empirical illustration of our results we estimate APEs in the context of measurement and analysis of immigrants' ethnic identity. This topic is nowadays largely debated since, following a number of EU-directives, during the last decade one of the common goals of policy makers across the European countries has been to provide a good level of integration to the increasing number of immigrants from within or outside Europe. This seems to be an appropriate action for avoiding episodes of social danger. In this view, the analysis of ethnic identity in association with the individual economic performance seems particularly useful.

Most of the initial economic literature on immigrants' integration focussed on measures of wage or occupational differentials with respect to natives. Here we take inspiration from a recent strand of the literature, based on an application of the theoretical models of Akerlof (1997) and Akerlof and Kranton (2000), and consider immigrants' economic performance and ethnic identity as both endogenous variables that are possibly interrelated. While we do not attempt to extract a casual link in one direction or the other, we set up a new model of joint determination of immigrants' economic performance and ethnic identity that identifies interesting response probabilities and the related APEs.

The process of formation of individual's sense of belonging to a country is complex: first, it is subjectively determined and can be different across people coming from a same country; second, it is essentially dynamic; third, it involves many aspects beyond the individual economic performance, such as the social, political and cultural dimensions. From an empirical point of view, different approaches to the measurement of ethnic identity have been proposed in the literature.

The majority of studies focus on one or more aspects of ethnic identity that are modeled, basically, as a linear process where commitment to the home and the host country are mutually exclusive: the more an individual commits and feels for one country the less she commits and fells for the other. In other words, individuals are assumed to adopt so-called *oppositional identities.* Figure 1 illustrates this process of identity formation. Imagine to represent the relationship between commitment to the origin and commitment to the host country on a bi-dimensional graph with the two axis dedicated, and suppose that we can measure each type of commitment on a 0-1 scale. All the points for which the two commitments are mutually exclusive lie on the diagonal from (1,0) to (0,1). The two extremes of the diagonal are the (0,1) point of maximum commitment to the host country, that denotes full adaptation of immigrants, and the (1,0) point of maximum commitment to the home country, that denotes the maximum level of *ethnicity* in the sense that ethnic identity has not been affected by the host country. Along the diagonal it is possible to the define what Constant, Gataullina, and Zimmermann (2009) call the *one-dimensional measure of ethnic identity or ethnosizer,* that implicitly captures the idea of immigrant low/high assimilation in economic research, one that is easy to measure because in practice requires information on the commitment only for one country.

Constant, Gataullina, and Zimmermann (2009) develop a more general model for ethnic identity, based on the psychological studies of Berry (1997). In this context ethnic identity is considered as a personal mix between the sense of belonging to the home or the host country, i.e. as the result of a dynamic acculturation process that happens in a trans-national dimension and can generate more than two extreme outcomes because commitment to the home and the host countries can also co-exist, although with different strengths. . In terms of Figure 1, our measure of ethnic identity can be any point in the area delimited by the points (0,0), (0,1), (1,1) and (1,0). Constant, Gataullina, and Zimmermann (2009) call this measure the *two-dimensional ethnosizer* because it requires information on the commitment to *both* the home and the host cultures on as many aspects as possible (values, norms, languages, culture, etc.). According to the definitions in Berry (1997), when a strong identification with the host culture is coupled with a weak dedication to the ancestry the immigrant's type of acculturation is called *Assimilation* and the ethnosizer falls in the upper-left quadrant; any point in the upper-right quadrant (strong dedication to both the home and the host culture) describes the process o*f Integration*; the state opposite to Assimilation is called *Separation* (lower-right quadrant); and the case of weak dedication to both home and host country is referred to as *Marginalisation* (lower-left quadrant).

There are only a few empirical applications based on the the two-dimensional ethnosizer (Constant, Gataullina, and Zimmermann 2009 for Germany, Nekby and Rodin 2010 for Sweden, and Drydakis 2012 for Greece). Although with different data and number of aspects available for constructing the ethnosizer, all of them follow an approach into two separate stages. First, they analyze the determinants of the four ethnic identity outcomes separately through OLS regressions. Then, in the spirit of Akerlof (1997) and Akerlof and Kranton (2000), they work out correlations between some measures of economic performance and ethnic identity.

Here we adhere to a concept of two-dimensional ethnosizer, but from a new methodological stance. Explicitly admitting the endogenous status of, on the one hand, the two dimensions of ethnic identity and, on the other, an indicator of economic performance, we estimate joint and conditional response probabilities, with the related APEs. In doing this, we extend the evidence from the existing literature to the Italy.

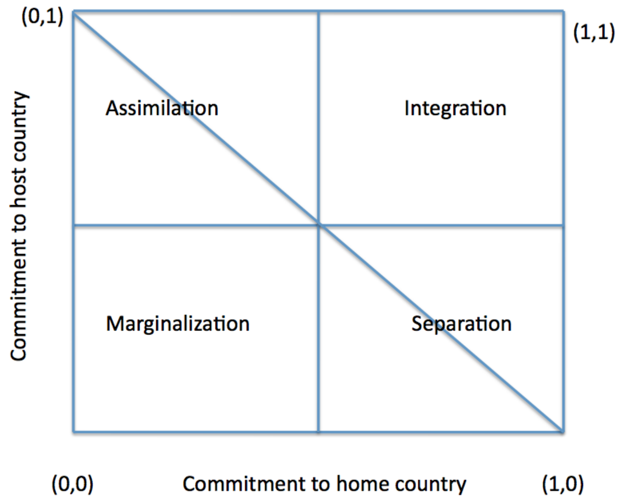

Fig. 1. The two-dimensional ethnosizer

7.1 A Model of trans-national ethnic identities

We set up a trivariate probit model with no constraints beyond normalization, i.e. System (11) specialized to $m = 3$. The first two equations describe the immigrant's latent commitment to the culture of host, y_1^*, and home country y_2^*, whereas the last equation describes the immigrant's latent propensity to find an occupation y_3^*:

$$
\begin{aligned}
y_1^* &= \alpha_1 + \mathbf{x}'\boldsymbol{\beta}_1 + q_1 + \varepsilon_1 \\
y_2^* &= \alpha_2 + \mathbf{x}'\boldsymbol{\beta}_2 + q_2 + \varepsilon_2 \\
y_3^* &= \alpha_3 + \mathbf{x}'\boldsymbol{\beta}_3 + q_3 + \varepsilon_3
\end{aligned}
$$

with the observed binary variables $y_i = \mathbf{1}(y_i^* > 0)$, $i = 1, 2, 3$. This latent regression system can be thought of as the reduced form of a simultaneous equation model, where y_1^*, y_2^* and y_3^* may be right-hand side variables in some or all of the equations. Combining the outcomes of y_1 and y_2 leads to the classification in Berry (1997): 1) *Integrated* ($y_1 = 1, y_2 = 1$); 2) *Assimilated* ($y_1 = 1, y_2 = 0$); 3) *Separated* ($y_1 = 0, y_2 = 1$); and 4) *Marginalized* ($y_1 = 0, y_2 = 0$). A finer partition of eight possible outcomes arises when the four different types of ethnic identities are combined with the economic status *Occupied* ($y_3 = 1$) or *Not occupied* ($y_3 = 0$). Accordingly, eight APEs of the generic regressor x are identified under A.1 and A.2 alone

$$APE_x(k_1, k_2, k_3) = E_{\mathbf{q}}\left[\partial_x Pr\left(y_1 = k_1, y_2 = k_2, y_3 = k_3 | \mathbf{x}^0, \mathbf{q}\right)\right],$$

$k_1, k_2, k_3 = 0, 1$ and, given the unconstrained nature of the system, with an arbitrary covariance matrix for the $\mathbf{q}$'s, as seen in Section 3.

Interesting insights on the association between economic performance and ethnic identity can be grasped by estimating the average gap between conditional response probabilities

$$APE_{y_3}(k_1, k_2) = E_{\mathbf{q}}\left[Pr\left(y_1 = k_1, y_2 = k_2 | y_3 = 1, \mathbf{x}^0, \mathbf{q}\right) - Pr\left(y_1 = k_1, y_2 = k_2 | y_3 = 0, \mathbf{x}^0, \mathbf{q}\right)\right], \quad (35)$$

$k_1, k_2 = 0, 1$. The foregoing APEs are identified under the restricted framework of Section 4, with A.3 satisfied by construction and A.4 and A.5 that read here as $D(q_3|\mathbf{w}) = D(q_3)$ and $D(q_3|\mathbf{w}, q_1, q_2) = D(q_3|\mathbf{w})$.

The reference individual described by vector $\mathbf{x}^0$ is a catholic male living in the north macro-region of Italy, of current age and age at arrival equal to the sample mean points, 38 and 32 years old respectively. The APEs are then obtained averaging over the control variables, $\mathbf{w}$: education dummies and country-of-origin dummies. As established in Subsection 3, the APEs can be estimated using either estimators based on the GHK or the computationally inexpensive and also more precise `combined biprobit` procedure.

7.2 Data and variables' definition

The data used for our analysis have been collected by Fondazione ISMU (Foundation for Initiatives and Studies on MUltietnicity) between October 2008 and February 2009 through a questionnaire asked to more than 12,000 foreign immigrants aged more than 18. The method of collection called *by centres*[5], allows to construct proper weights of observations so that, although collected for a subset of regions, data are representative of the whole Italian population of foreign immigrants. In particular,

[5]Details of the sampling method can be found in Blangiardo and Cesareo (2011)

the regions covered are 13: Piemonte, Lombardia, Trentino Alto Adige, Veneto, Emilia Romagna, Toscana, Marche, Abruzzo, Lazio, Campania, Molise, Puglia, Sicilia.

Various are the advantages of using ISMU data. First of all, to date they are the only set of data in Italy that oversamples foreign immigrants, therefore being highly representative of this population. Second, they are collected with the specific purpose of studying the concept of integration. Third, they provide two variables asked symmetrically in the direction of the host and the home country, therefore allowing to provide two-dimensional measures of identity à la Zimmermann. The first question is a general self-assessment of the sense of belonging to a country: "How much do you feel you belong to Italy?" and "How much do you feel you belong to your home country?" . The second question asks: "To what extent are you interested in knowing what happens in Italy?" and "To what extent are you interested in knowing what happens in your home country?". In both cases the possible answers respect the following scale: "not at all","a little", "rather/sufficiently", and "a lot". The two lowest levels of intensity have been aggregated and valued 0, and the two highest levels have been aggregated and valued 1.

As explanatory variables we use: age, ethnicity (aggregated in four groups: Eastern-Europe, North-Africa, Other-Africa, Latin America), age at arrival, education (available in four levels: none, compulsory, secondary, laurea and more/tertiary[6]), religion (muslim, catholic, orthodox, copto, evangelistic/evangelical, other christian, hindu, sikh, other, none), region (aggregated in North, Centre, South).

7.3 Results

Estimation results are reported in Table 4. We first focus on the eight APEs of each of the three regressors: *age*, *age at arrival* and *sex* , $APE_x\left(k_1,k_2,k_3\right)$, and then on the four APEs of the employed/not-employed dummy variable, $APE_{y_3}\left(k_1,k_2\right)$.

Estimation is carried out through `combined biprobit` and `mvprobit`. The `mvprobit` estimators are implemented using the default number of draws, 5. In the case of the four $APE_{y_3}\left(k_1,k_2\right)$ we have also used 100 and 1000 draws.

We start commenting results for the eight $APE_x\left(k_1,k_2,k_3\right)$. Results are generally close between

[6]Notice that those are the degrees achieved in the home country, and therefore can be heterogeneous.

Table 4: Trivariate probit model of ethnic identity and economic performance (age = 38, $age\ at\ arrival$ = 32, sex = $male$, $religion$ = $catholic$)

Parameters\Statuses	`combined biprobit`					`mvprobit`[a]			
	employed					employed			
	integrated	assimilated	separated	marginalized		integrated	assimilated	separated	marginalized
APE - age	0.0272*** (0.0025)	0.0036*** (0.0007)	-0.0036 (0.0028)	0.0007** (0.0003)		0.0264*** (0.0026)	0.0038*** (0.0007)	-0.0031 (0.0029)	0.0007** (0.0003)
APE - age at arrival	-0.0293*** (0.0026)	-0.0045*** (0.0007)	0.0079*** (0.0030)	-0.0009*** (0.0003)		-0.0276*** (0.0028)	-0.0045*** (0.0008)	0.0065** (0.0030)	-0.0009*** (0.0004)
APE Women	0.0185 (0.0171)	0.0037 (0.0030)	-0.0439** (0.0191)	-0.0001 (0.0017)		0.0206 (0.0184)	0.0069* (0.0036)	-0.0482** (0.0202)	0.0011 (0.0019)
Parameters\Statuses	not-employed					not-employed			
	integrated	assimilated	separated	marginalized		integrated	assimilated	separated	marginalized
APE age	-0.0068*** (0.0013)	-0.0001 0.0003	-0.0205*** (0.0020)	-0.0005*** (0.0002)		-0.0072*** (0.0013)	-0.0003 (0.0003)	-0.0196*** (0.0020)	-0.0006*** (0.0002)
APE age at arrival	0.0058*** (0.0014)	-0.0003 (0.0003)	0.0210*** (0.0023)	0.0004** (0.0002)		0.0063*** (0.0014)	-0.0001 (0.0003)	0.0198*** (0.0022)	0.0005** (0.0002)
APE Women	0.0157* (0.0086)	0.0007 (0.0010)	0.0024 (0.0115)	0.0007 (0.0010)		0.0152* (0.0089)	0.0013 (0.0011)	-0.0008 (0.0114)	0.0013 (0.0011)
Parameters\Statuses	integrated	assimilated	separated	marginalized	draws	integrated	assimilated	separated	marginalized
APE employed (APE_{y_3})	0.046*** (0.004)	-0.024*** (0.003)	-0.005 (0.004)	-0.016*** (0.003)	5	0.039*** (0.003)	-0.018*** (0.002)	-0.007** (0.003)	-0.013*** (0.002)
					100	0.048*** (0.004)	-0.026*** (0.003)	-0.004 (0.004)	-0.018*** (0.003)
					1000	0.048*** (0.004)	-0.027*** (0.003)	-0.004 (0.004)	-0.017*** (0.003)
cov12, cov13, cov23	-0.263***, 0.032, 0.195*** (0.022, 0.020, 0.027)					-0.269***, 0.034, 0.195*** (0.039, 0.035, 0.045)			
computing time	6'34"								

[a]Default number of draws, 5, except where indicated

`combined biprobit` and `mvprobit`, in terms of coefficient signs, magnitudes and standard errors and we will therefore base our comments on the results from the former procedure. Focussing on the APE of *age* we find that growing one year older makes it more likely being "employed" and either "integrated", "assimilated", or "marginalized", with a greater APE (+2.7%) on the association "employed-integrated". It also makes it less likely being "not-employed" in general, with a greater APE on the association "not-employed-separated" (-2.0%). Turning, then, on the APE of *age at arrival,* we notice that arriving one year later makes it less likely to be employed and either integrated, assimilated, or marginalized with a strong negative effect on the association with "integrated" (-2.9%). It also makes it more likely being separated, whether employed or not, with a greater APE on the association "not-employed-separated" (+2.1%). As regards the differences between women and men[7], we find that they are not generally significant, with two notable exceptions: in comparison with men, women have 1) a smaller probability of being "employed-separated" (-4.4% smaller); 2) a greater probability of being "not-employed-integrated" (+1.6%).

Finally, turning to the four estimators of $APE_{y_3}(k_1, k_2)$, we find that being employed has a significantly positive impact on the chances of being integrated and significantly reduces those of being either assimilated or marginalized. It has, though, a small and statistically insignificant impact on separation. From a policy perspective, these findings show that policies stimulating the labour market participation of immigrants in Italy are likely to increase integration, but only at the expenses of assimilation and marginalization, not separation. If the policy target is on a lower degree of immigrants' separation, different forms of interventions should be thought of.

On the computation side, we notice that `mvprobit` provides estimates that are largely close to `combined biprobit`. Of course greater accuracy, at the expenses of enormously higher computation times, is obtained when the number of draws departures from 5, increasing to 100 and 1000. It is interesting that in these cases the `mvprobit`'s APE estimates get even closer to `combined biprobit`'s, so confirming the dimensionality gains exploited by the computationally inexpensive combined procedures.

[7] Our APE estimates are to be thought of as first-order Taylor approximations of the differences in probability statuses between women and men.

8 Conclusion

This paper has dealt with identification and estimation of APEs in multivariate probit models with conditionally independent LH.

We have proved that APEs based on joint or marginal response probabilities are identified by a reduced-from multivariate probit model with the same constraints as the structural model, provided that the structural-error covariance matrix is unconstrained beyond normalization (UBN). To be UBN, it must have a minimal set of normalization constraints given the other model restrictions. In models with cross-equation equality restrictions or linear non-homogenous restrictions, a structural covariance matrix in correlation form does not respect the foregoing caveat, with the consequence that for the APEs to be identified by a conformably restricted reduced-form model the LH covariance matrix has to be restricted as well.

In principle, the problem could be overcome by estimating the reduced-form model with a covariance matrix that is UBN. This is rather easy for the multinomial models estimated by Stata's `asmprobit` and for models with linear non-homogenous restrictions, such as willingness-to-pay models (Wooldridge 2005). But standard multivariate models with cross-equation equality constraints, such as the panel probit model (Bertschek and Lechner 1998 and Greene 2004), meet the limitation common to many probit routines that the error covariance matrix cannot be more general than the correlation form. This is the case of Stata's `biprobit`, `mvprobit` and `cmp` or Limdep's `BIVARIATE PROBIT` and `MPROBIT`. The *multinomial probit model with i.i.d. errors*, implemented by Stata's `mprobit`, provides another example of a covariance matrix that is not UBN and as such does not support arbitrary covariance for the LH components.

We have also proved that, in the case of APEs of conditional response probabilities, conditional independence of the LH components is not sufficient for consistency of the reduced-form multivariate probit estimator, and further independence assumptions are needed. This holds regardless of the model being standard or recursive. Monte Carlo experiments prove that in the absence of the additional independence assumptions the finite-sample bias may be severe.

A computational result of the paper is that the analytical formulas available for PEs in multivariate probit models (Greene 2012; Mullahy 2011) can be used for robust estimation of APEs, taking advantage of the reduced-dimensionality benefit evidenced by Mullahy (2011). In this respect, we propose a combined bivariate probit method for estimation of robust multivariate-probit APEs that proves to be not only dramatically faster, but also more accurate and precise than existing simulation-based estimators, such as those implemented by `mvprobit` and `cmp`, as we demonstrate in a battery of Monte Carlo experiments. Further efficiency gains, at the expenses of computational ease though, could be obtained in principle by implementing a minimum distance estimator averaging all of the available vectors of coefficient estimates from the $m\,(m-1)\,/2$ bivariate probit regressions. This is part of our research agenda.

References

Akerlof, G. A. (1997): "Social distance and social decisions," *Econometrica*, 65, 1005–1027.

Akerlof, G. A., and R. E. Kranton (2000): "Economics and identity," *Quarterly Journal of Economics*, 115, 715–753.

Berry, J. W. (1997): "Immigration, acculturation and adaptation," *Applied Psychology: An International Review*, 46, 5–68.

Bertschek, I., and M. Lechner (1998): "Convenient estimators for the panel probit model," *Journal of Econometrics*, 87, 329–371.

Bhattacharya, J., D. Goldman, and D. McCaffrey (2006): "Estimating probit models with self-selected treatments," *Statistics in Medicine*, 25, 389–413.

Blangiardo, G. C., and V. Cesareo (2011): *Integration indexes*, no. 2/2011 in Quaderni Ismu. Fondazione ISMU, Milan.

Cappellari, L., and S. P. Jenkins (2003): "Multivariate probit regression using simulated maximum likelihood," *The Stata Journal*, 3, 278–294.

Christofides, L. N., T. Stengos, and R. Swidinsky (1997): "On the calculation of marginal effects in the bivariate probit model," *Economics Letters*, 54, 203–208.

Constant, A., L. Gataullina, and K. F. Zimmermann (2009): "Ethnosizing immigrants," *Journal of Economic Behavior and Organization*, 69(3), 274–287.

Dawid, A. P. (1979): "Conditional independence in statistical theory," *Journal of the Royal Statistical Society B*, 41, 1–31.

Drydakis, N. (2012): "The effect of ethnic identity on the employment of immigrants," *Review of Economics of the Household*, forthcoming.

Entorf, H. (2012): "Expected recidivism among young offenders: Comparing specific deterrence under juvenile and adult criminal law," *European Journal of Political Economy*, 28, 414–429.

Evans, W. N., and R. M. Schwab (1995): "Finishing High School and Starting College: Do Catholic Schools Make a Difference?," *The Quarterly Journal of Economics*, 110, 941–974.

Fichera, E., and M. Sutton (2011): "State and self investment in health," *Journal of Health Economics*, 30, 1164–1173.

Geweke, J. (1989): "Bayesian inference in econometric models using Monte Carlo integration," *Econometrica*, 57, 1317–1340.

Greene, W. H. (1996): "Marginal effects in the bivariate probit model," Working Paper EC-96-11, NYU, Stern School of Business.

Greene, W. H. (1998): "Gender Economics Courses in Liberal Arts Colleges: Further Results," *Journal of Economic Education*, 29, 291–300.

Greene, W. H. (2004): "Convenient estimators for the panel probit model: Further results," *Empirical Economics*, 29, 21–47.

——— (2012): *Econometric Analysis*. Upper Saddle River, NJ: Prentice Hall, seventh edn.

Hajivassiliou, V. A. (1990): "Smooth simulation estimation of panel data LDV models," Discussion paper, Department of Economics, Yale University.

Keane, A. (1994): "A computationally practical simulation estimator for panel data," *Econometrica*, 62, 95–116.

Maddala, G. S. (1983): *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge University Press, Cambridge.

Mullahy, J. (2011): "Marginal effects in multivariate probit and kindred discrete and count outcome models, with applications in health economics," NBER WP SERIES 17588, NBER.

Nekby, L., and M. Rodin (2010): "Acculturation identity and employment among second and middle generation immigrants," *Journal of Economic Psychology*, 31, 35–50.

Rao, C. R. (1973): *Linear Statistical Inference and Its Applications*. New York: Wiley.

Rivers, D., and Q. H. Vuong (1988): "Limited information estimators and exogeneity tests for simultaneous probit models," *Journal of Econometrics*, 39, 347–366.

Roodman, D. M. (2011): "Fitting fully observed recursive mixed-process models with cmp," *The Stata Journal*, 11, 159–206.

Train, K. (2002): *Discrete Choice Methods with Simulation*. Cambridge University Press, Cambridge.

White, H. (1982): "Maximum likelihood estimation of misspecified models," *Econometrica*, 50(1), 1–25.

Wooldridge, J. M. (2005): "Unobserved heterogeneity and estimation of average partial effects," in *Identification And Inference For Econometric Models: Essays In Honor Of Thomas Rothenberg*, ed. by D. W. K. Andrews, and J. H. Stock. Cambridge University Press, New York.

——— (2010): *Econometric Analysis of Cross Section and Panel Data*. The MIT Press, Cambridge, MA, 2nd edn.

Yatchew, A., and Z. Griliches (1985): “Specification error in probit models,” *Review of Economics and Statistics*, 67, 134–139.

Table 5: Monte Carlo results for `mvprobit` with 10 draws (100 where indicated) and $\tau = 0$.

n ＼ parameters	APE	β_1	β_2	β_3	cov12	cov13	cov23
	BIAS						
n=500	0.023	-0.035	0.007	-0.046	0.121	-0.039	0.073
n=1000	-0.020	-0.036	0.060	-0.024	0.130	-0.081	0.186
n=1000 (100 draws)	-0.012	-0.093	0.107	-0.035	0.056	-0.014	-0.028
n=10000	-0.004	0.009	-0.025	0.017	0.206	0.006	0.078
n ＼ parameters	**RMSE**						
n=500	0.023	0.035	0.007	0.046	0.121	0.039	0.074
n=1000	0.020	0.036	0.060	0.024	0.130	0.081	0.186
n=1000 (100 draws)	0.012	0.093	0.107	0.035	0.056	0.014	0.028
n=10000	0.004	0.009	0.025	0.017	0.206	0.006	0.078

A Further Monte Carlo results

Table 6: Monte Carlo results for `cmp` with 10 draws (cmp's default draws where indicated) and $\tau = 0$.

n ╲ parameters	APE	β_1	β_2	β_3	cov12	cov13	cov23
	BIAS						
n=500	-0.008	0.008	-0.017	0.001	0.203	-0.035	0.098
n=1000	-0.008	0.002	-0.019	0.004	0.201	-0.031	0.097
n=1000 (default)	-0.003	0.008	-0.013	0.003	0.039	0.003	0.024
n=10000	-0.009	-0.001	-0.025	-0.002	0.207	-0.036	0.101
n ╲ parameters	**RMSE**						
n=500	0.018	0.097	0.098	0.098	0.218	-0.106	0.135
n=1000	0.014	0.066	0.069	0.067	0.207	0.072	0.115
n=1000 (default)	0.011	0.068	0.067	0.066	0.052	0.065	0.071
n=10000	0.009	0.022	0.033	0.022	0.208	0.042	0.103
n ╲ parameters	**Computation time per replication**						
n=500	3.37"						
n=1000	3.97"						
n=1000 (default)	12.03"						
n=10000	12.96"						