

Simulating Transitions in Multinomial Probit Models

Alan Duncan ^α and Melvyn Weeks^γ

March 1998

PRELIMINARY. NOT FOR QUOTATION

Abstract

In the wake of recent developments in simulation-based inference, the multinomial probit model has received considerable attention as previously intractable barriers to estimation have been removed. However, it is still the case that outcome measures of interest to policy makers, such as the likelihood of moving between states following an exogenous shock, have been relatively neglected. In this study we examine a number of alternate transitions estimators for the multinomial probit both in the context of a Monte Carlo study and an application to labour market transitions following a policy reform.

^αDepartment of Economics and Related Studies, University of York, Heslington, York YO1 5DD, United Kingdom and Institute for Fiscal Studies. Tel: +44 (1)904 433794; Email: asd1@york.ac.uk

^γDepartment of Applied Economics, University of Cambridge, Austin Robinson Building, Sidgwick Avenue, Cambridge CB3 9DE, United Kingdom. Tel: +44 (1)223 335260; EMail: Melvyn.Weeks@econ.cam.ac.uk.

1. Introduction

In both the social and physical sciences the use of discrete choice or stimulus and response models is commonplace. This is based upon the recognition that many phenomena involve the choice between or passage through discrete, identifiable states. However, there has been relative neglect of the modelling of transitions between states in response to some exogenous shock. A number of exceptions are recent studies of labour supply behaviour where discrete choice models have been used to simulate transitions in labour force status in response to some reform to the tax-benefit system (see Duncan and Giles (1996) and Bingley and Walker (1995) among others). In addition a recent study Keane (1997) utilises panel data on consumer purchases to examine the relative importance of heterogeneity and state dependence in explaining the observed temporal persistence of consumer brand choices. A number of models are specified and one measure of fit is a comparison of observed and predicted transition frequencies across product variants. However in many cases the analyst has a single cross-section of observed choices, and conditional upon an estimated model, seeks to estimate the impact of one or more counterfactuals. Obviously in this context transitions occur without reference to an observable event, such that there is not a ready made metric for model comparison.

While such applications are undoubtedly of great relevance, the policy impact of transitions studies of this kind tend to be offset to a degree by the notorious, empirically observed tendency for outcome-based measures of fit in such models to be relatively poor in cross-sectional studies. The literature abounds with empirical studies where the authors report a systematic over- or under-prediction of certain state frequencies¹. Whilst problems of this kind may simply indicate some form of specification error, there is some suggestion that even in functionally well-specified models the predictive performance is poor, particularly where some states are relatively densely or sparsely represented in the data. The question of whether any discrepancies between observed and predicted choices really matters is, in part, a question of aggregation. For example, Hausman and Wise (1978) note that at the level of the individual accurate computation of choice behaviour is not necessary

¹see Cramer (1997) for a review of this problem in both the social and natural sciences

if the aggregate expected frequencies are accurate given the operation of the law of large numbers. Counter to this, Cramer (1991) notes that, when the statistical model is part of a wider microsimulation study, then predicted outcomes for each individual are indeed required. For example, in utilising a discrete choice model of labour supply to examine the impact of tax reform we typically compare predicted labour market states before and after the reform to determine the distribution of "movers" and "stayers" and the existence of any systematic patterns in inter-state mobility. It is therefore important in such microsimulation studies for predicted transitions to be reliable at a disaggregated level.

The statistical reliability of predicted state transitions is often ignored in the literature, perhaps because the distribution of the transition probability estimator is not obvious even if one has a reasonably good approximation for the distribution of the underlying parameters of the discrete choice model. At the very least, the distribution of a transition probability estimator has a potential for asymmetry irrespective of the characteristics of the distribution of the model itself, which rules out the so-called "delta" method as a basis for significance tests.

In this paper we examine alternative transitions estimators which control for the systematic under- or over-prediction of state frequencies inherent in many discrete choice models. In section 2 we introduce the statistical model and consider the problem of estimating transitions frequencies in the context of a trinomial probit model. In section 3 we present a thorough analysis of the derivation of transition probabilities using a graphical exposition. We also discuss how one might condition discrete choice model predictions on observed behaviour as a benchmark against which to measure the probability of transition from one state to another. In section 4 we conduct a Monte Carlo exercise to compare the various transitions estimators (either unconditional or conditioned on observed behaviour) described in the paper. Our alternative to the more traditional transitions estimator is seen to perform well within the confines of the experimental design chosen for simulation. Finally, we describe a bootstrap approach to the estimation of confidence intervals for transitions estimators as a means to comment on the statistical significance of predicted state transitions, and discuss how one might apply such measures when simulating labour market transitions following some tax-benefit

reform.

2. The Model Framework

We consider the following trinomial model of discrete choice

$$\begin{aligned} u_1^a &= x_1^0 + \epsilon_1 \\ u_2^a &= x_2^0 + \epsilon_2 \\ u_3^a &= x_3^0 + \epsilon_3 \end{aligned} \quad (2.1)$$

where u_j^a denotes the utility of the j th choice, $x = (x_1; x_2; \dots; x_k)$ is an observable set of k characteristics, $x_j^0 = (x_{j1}; x_{j2}; \dots; x_{jk})^0$, and ϵ_j is a stochastic term. These latent variables are related to observed quantities $y \in \{1; 2; \dots; J\}$ via the mapping

$$y = \arg \max_j [u_j^a; j = 1; \dots; 3] \quad (2.2)$$

If we normalise 2.1 using alternative 1 we have

$$\begin{aligned} u_{21}^a &= x_{21}^0 + \epsilon_{21} \\ u_{31}^a &= x_{31}^0 + \epsilon_{31} \end{aligned} \quad (2.3)$$

where $x_{21}^0 = x_2^0 - x_1^0$; $x_{31}^0 = x_3^0 - x_1^0$; and generally, $x_{jk}^0 = x_j^0 - x_k^0$; $\epsilon_{jk} = \epsilon_j - \epsilon_k$ ($i; j = 1; \dots; 3$):

If $\epsilon = (\epsilon_1 \ \epsilon_2 \ \epsilon_3)^0 \gg N(0; S)$; where S is the 3×3 covariance matrix, then $\epsilon_{i1} = (\epsilon_{21} \ \epsilon_{31})^0 \gg N(0; S_{i1})$; where S_{i1} is the 2×2 covariance matrix for the differenced model.

2.1. measuring predictive performance

Consider now the evaluation of model (2.1) based on a sample of data of size N . Let the set $\{y_i; x_i\}$ for $i = 1; \dots; N$ denote observations on y and x ; and define the $N \times J$ matrix $Y = [Y_1; \dots; Y_J]$ with typical element $Y_{ij} = 1(y_i = j)$ for all $i; j$. For a given parameter vector β , we may compile a matrix P of sample probabilities with typical element $P_{ij}(x_i; \beta) = \Pr(y_i = j | x_i; \beta)$. A common empirical practice is to

model discrete outcomes on the basis of probabilities $P_{ij}(x_i; \bar{\cdot})$ using a maximum probability rule. If we let y_i represent the predicted state, then

$$y_i = \arg \max_j [P_{ij}(x_i; \bar{\cdot}); j = 1; \dots; J]; \quad (2.4)$$

One outcome-based performance measure for a qualitative model of this form would compare within sample frequencies of predicted and observed outcomes over the range of y .² Defining the $N \times J$ matrix Ψ_{MP} with typical element $\Psi_{MP} = 1(y_i = j)$ for all $i; j$, where $1(\cdot)$ represents the indicator function, we can summarise the proximity of predicted to observed states in terms of the $J \times J$ matrix $\frac{1}{N} Y^0 \Psi_{MP}$, the trace of which represents the proportion of observations for which the predicted and observed states coincide. For a model to predict perfectly requires that $\text{tr}(\frac{1}{N} Y^0 \Psi_{MP}) = 1$, an empirical feat which is rare.³

Consider now an exogenous shock x^R which impacts on state-specific utilities. Based upon the same parameter vector $\bar{\cdot}$, we may simulate the counterfactual state y_i^R for the i th observation by employing the same maximum probability rule as before, but in terms of the counterfactual x_i^R , giving

$$y_i^R = \arg \max_j [P_{ij}(x_i^R; \bar{\cdot}); j = 1; \dots; J]; \quad (2.5)$$

Standard practice builds a matrix of transitions frequencies as follows: define a matrix Ψ_{MP}^R with typical element $\Psi_{MP}^R = 1(y_i^R = j)$ for all $i; j$. Then, a summary of transitions frequencies based on the maximum probability rules (2.4) and (2.5) can be expressed in terms of the $J \times J$ matrix

$$T_{MP} = \Psi_{MP}^0 \Psi_{MP}^R; \quad (2.6)$$

the trace of which represents the number of observations for which the predicted states remain the same when one moves from the base to the counterfactual regime. Clearly, $\text{tr}(T_{MP}) = N$ for a counterfactual where no transitions are predicted.

²This is not the only measure of fit available; see Windmeijer (1995) for a useful comparison of goodness-of-fit measures in binary choice models.

³This particular measure is misleading since it has no asymptotic justification. That is to say, $\text{plim} \text{tr}(\frac{1}{N} Y^0 \Psi_{MP}) \neq 1$ even in the case of a well specified model.

However, since the transitions frequencies are constructed using the maximum probability rule, there is no guarantee that they will converge to their true values even if the behavioural parameters of the discrete choice model are themselves consistent. It is therefore difficult to place any real faith in predicted transitions frequencies of this form unless and until we are able to correct in some way the finite sample bias that exists in predictions of frequencies under the maximum probability rule.

3. An Alternative Estimator

Below we consider an alternative estimator based upon a more efficient use of information.

For utilities $u_{ij}^a = x_i^a \gamma_j + \epsilon_{ij}$; $j = 1; \dots; J$, we may derive probabilities $P_{ij} = \Pr(y_i = j | x_i; \epsilon)$ for all states j given a specific distribution for ϵ_{ij} . Our interest centres on the effects on these probabilities following an exogenous shock $x_i^R \notin x_i$ which alters utilities from u_{ij}^a to $u_{ij}^R = (x_i^R)^a \gamma_j + \epsilon_{ij}$. That is, we are in general interested in the probability $P_{i(j^a | j^R)}$ of transition from any state j^a to j^R ; where in general $P_{i(j^a | j^R)} = \Pr(u_{ij^R}^R > u_{ij^a}^R | \epsilon_{ij^a} > u_{ij^a}^a | \epsilon_{ij^a} > u_{ij^a}^a)$. When utilities are expressed in linear form, $P_{i(j^a | j^R)} = \Pr(y_i = j^a | x_i; \epsilon) \Pr(\epsilon_{ij^a} - \epsilon_{ij^R} < (x_i^R)^a \gamma_{j^R} - (x_i^a)^a \gamma_{j^a} | \epsilon_{ij^a} - \epsilon_{ij^R} < (x_i^a)^a \gamma_{j^R} - (x_i^a)^a \gamma_{j^a})$. By cumulating transitions probabilities (either unconditional or conditioned on y_i) over the sample, we may derive theoretical transitions frequencies $n_{i(j^a | j^R)}$ where

$$n_{i(j^a | j^R)} = \sum_{i=1}^N P_{i(j^a | j^R)} \text{ for all } j^a, j^R = 1; \dots; J: \quad (3.1)$$

Below we consider the estimator in more detail for the trinomial probit model

3.1. Trinomial Transition Probabilities

We are interested in the following set of state transitions probabilities

$$P_{kj} = \Pr(y^R = j; y = k) \quad k; j = 1; 2; 3 \quad (3.2)$$

where y (y^R) is an indicator for, respectively, post and pre-shock states of the world. Below we consider a particular example

Example 3.1. $P_{12} = \Pr(y = 1; y^R = 2 | x; x^R; \bar{\mu}; \bar{S}_{i-1})$

We consider the calculation of the joint probability that $y = 1$ and $y^R = 2$, namely $\Pr(y = 1; y^R = 2; x; x^R; \bar{\mu}; \bar{S}_{i-1})$, where $\bar{\mu}$ and \bar{S}_{i-1} are, respectively, the vector of mean parameters and the differenced covariance matrix.

$$P_{12} = P_R \int_{\substack{u_{21}^{aR} > 0; u_{23}^{aR} > 0 \\ u_{21}^a < 0; u_{31}^a < 0}} \frac{1}{\sqrt{|\bar{A}|}} \exp\left\{-\frac{1}{2} \begin{bmatrix} (x^R)^0(-2 \text{ } i \text{ } -1) + \mu_{21} > 0; \\ (x^R)^0(-2 \text{ } i \text{ } -1) \text{ } (x^R)^0(-3 \text{ } i \text{ } -1) + \mu_{21} \text{ } \mu_{31} > 0 \\ x^0(-2 \text{ } i \text{ } -1) + \mu_{21} < 0 \\ x^0(-3 \text{ } i \text{ } -1) + \mu_{31} < 0 \end{bmatrix} \bar{A}^{-1} \begin{bmatrix} (x^R)^0(-2 \text{ } i \text{ } -1) + \mu_{21} > 0; \\ (x^R)^0(-2 \text{ } i \text{ } -1) \text{ } (x^R)^0(-3 \text{ } i \text{ } -1) + \mu_{21} \text{ } \mu_{31} > 0 \\ x^0(-2 \text{ } i \text{ } -1) + \mu_{21} < 0 \\ x^0(-3 \text{ } i \text{ } -1) + \mu_{31} < 0 \end{bmatrix}\right\} d\mu_{21} d\mu_{31} \quad (3.3)$$

Example 3.2. From 3.3 we note that the four conditions that need to be satisfied are:

$$\begin{aligned} (x^R)^0(-2 \text{ } i \text{ } -1) + \mu_{21} > 0; & \quad \mu_{21} > (x^R)^0(-2 \text{ } i \text{ } -1) \\ (x^R)^0(-2 \text{ } i \text{ } -1) \text{ } (x^R)^0(-3 \text{ } i \text{ } -1) + \mu_{21} \text{ } \mu_{31} > 0; & \quad \mu_{31} < (x^R)^0(-3 \text{ } i \text{ } -1) \text{ } (x^R)^0(-2 \text{ } i \text{ } -1) + \mu_{21} \\ x^0(-2 \text{ } i \text{ } -1) + \mu_{21} < 0; & \quad \mu_{21} < x^0(-2 \text{ } i \text{ } -1) \\ x^0(-3 \text{ } i \text{ } -1) + \mu_{31} < 0; & \quad \mu_{31} < x^0(-3 \text{ } i \text{ } -1) \end{aligned} \quad (3.4)$$

Based upon 3.4 the form of the joint probability expression is as follows:

$$\int_{(x^R)^0(-2 \text{ } i \text{ } -1)}^{x^0(-2 \text{ } i \text{ } -1)} \int_{\min((x^R)^0(-3 \text{ } i \text{ } -1) + \mu_{21}; x^0(-3 \text{ } i \text{ } -1))}^{\min((x^R)^0(-3 \text{ } i \text{ } -1) + \mu_{21}; \mu_{31})} \int_{\mu_{21}}^{\mu_{31}} \int_{\mu_{21}}^{\mu_{31}} d\mu_{21} d\mu_{31} \quad (3.5)$$

The problem of evaluating P_{12} in a binary choice problem is trivial (see Duncan and Weeks (1997)). For the 3 choice problem, and for the above example, the principal complication is that the random variable μ_{21} appears in the upper limit for μ_{31} : However, since we know that $P(A \setminus B) = P(A)P(B|A) = P(A)[P(B|A^0) + P(B|A^0)]$ where $A^0 \setminus A^0 = A$; then using a similar framework of decomposing the conditioning variable, and denoting $a_{12} = x^0(-2 \text{ } i \text{ } -1)$; $b_{12} = (x^R)^0(-2 \text{ } i \text{ } -1)$; $a_{13} = x^0(-3 \text{ } i \text{ } -1)$; $b_{13} = (x^R)^0(-3 \text{ } i \text{ } -1)$; and $c = b_{13} - b_{12}$; we may write ?? as

$$\int_{a_{12}}^{\min(c; a_{13})} \int_{b_{12}}^{\min(c; \mu_{21}; a_{13})} \int_{\mu_{21}}^{\mu_{31}} \int_{\mu_{21}}^{\mu_{31}} d\mu_{21} d\mu_{31} \quad (3.6)$$

$$= \int_{b_{12}}^{a_{12}} \int_{i=1}^{\infty} \int_{\min(c-x_{21}, a_{13})}^{\infty} \int_{j=1}^{\infty} \mathbf{1}(\dots) d x_{31} d x_{21}$$

where $\mathbf{1}(\dots) \gg N(1_{31} + B(x_{21} - 1_{21}); \frac{3}{4}x_{21}^2 (1 + \frac{1}{2}x_{21}^2))$:

From 3.6 we note that the inclusion of an endogenous component (x_{21}) in the upper integrand of the inner integral in 3.6. However, since the marginal support of x_{21} is $a_{12} < x_{21} < b_{12}$, then dependent upon the evaluation of $\mathbf{1}(x_{21} < a_{13} + c)$ the min condition on the inner integral of 3.6 evaluates to $c + x_{21}$ or a_{13} ; and in turn will determine the support of x_{21} : By comparing $a_{13} + c$ with the range of x_{21} we may identify three mutually exclusive regimes for P_{12} . These are:

1. $a_{13} + c < b_{12}$

In this instance the support of x_{31} and x_{21} is, respectively, $j = 1 < x_{31} < a_{13}$ and $b_{12} < x_{21} < a_{12}$; such that the probability for regime 2 may be written

$$\int_{b_{12}}^{a_{12}} \int_{i=1}^{\infty} \int_{a_{13}}^{\infty} \int_{j=1}^{\infty} \mathbf{1}(\dots) d x_{31} d x_{21} \quad (3.7)$$

This problem is depicted in Figure 1(a), using line r_1 :

2. $a_{13} + c > a_{12}$

The support of x_{31} and x_{21} is, respectively, $j = 1 < x_{31} < c + x_{21}$ and $b_{12} < x_{21} < a_{12}$: Utilising these limits we may write the probability for regime 1.

$$\int_{b_{12}}^{a_{12}} \int_{i=1}^{\infty} \int_{c-x_{21}}^{\infty} \int_{j=1}^{\infty} \mathbf{1}(\dots) d x_{31} d x_{21} \quad (3.8)$$

This problem is depicted in Figure 1(a), using line r_2 :

Figure 1 (a)
 Case B: $a_{31} \mid c < {}^2_{21}$
 Case C: $a_{31} \mid c > {}^2_{21}$

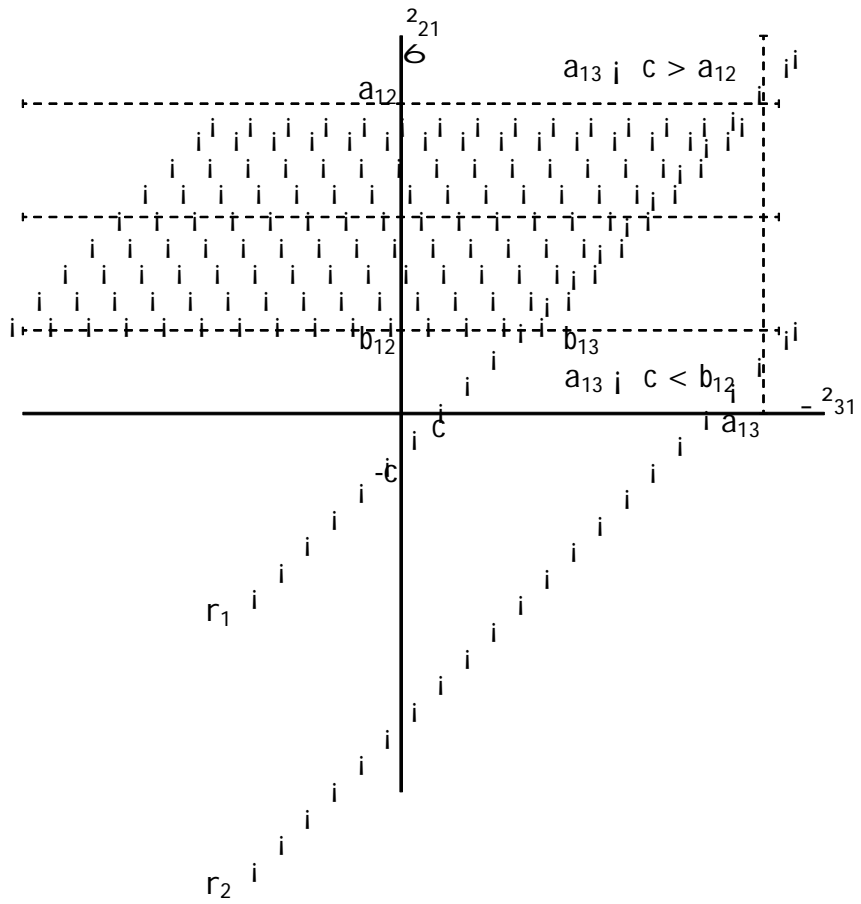
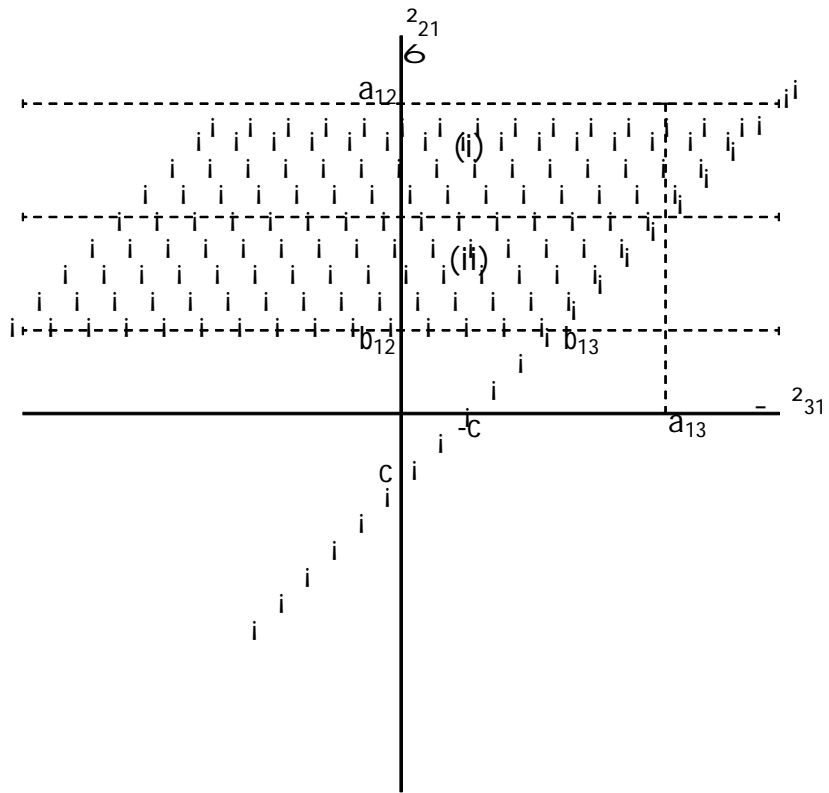


Figure 1 (b)
 Case A: $a_{21} < a_{13}$; $c < b_{12}$



3.2. Calibrated Transition Probabilities

In some circumstances we may require our transitions estimator to be calibrated to reproduce observed (pre-shock) behaviour with certainty. For example, in a microsimulation study of the impact of tax reform on labour market behaviour the researcher may wish to examine transitions relative to observed rather than predicted behaviour in order for the simulated costs of the reform to be as close as possible to those simulated in static microsimulation models. We demonstrate below how one might guarantee such a benchmark using a discrete choice behavioural model.

Consider again a set of J state-specific latent utilities $u_j^a = x_j^0 + \epsilon_j$ related to observed quantities $y \in \{1, 2, \dots, J\}$: To force this model to predict the observed outcome y through the maximum probability mapping (2.4), we must place bounds on the values of the unobservable components of utility ϵ_j . By exploiting any assumptions that are made about the stochastic distribution of the model, we may generate a realisation e_j which can then be factored into latent relationships of the form $u_j^a = x_j^0 + e_j$ to recover discrete predictions \hat{y} which coincide exactly with y .

Suppose that we observe $y = j^a \in \{1, \dots, J\}$: If our assumed model relies on the latent relationships $u_j^a = x_j^0 + \epsilon_j$ for $j = 1, \dots, J$; then for $\hat{y} = j^a$ through the mapping (2.2) requires that $u_{j^a}^a > u_j^a$ for all $j \neq j^a$; which implies that

$$(\epsilon_j - \epsilon_{j^a}) < x_j^0 - x_{j^a}^0 \text{ for all } j \neq j^a: \quad (3.10)$$

Our two proposed calibration methods derive realisations $e_j \in e_{j^a}$ which respect the bounds (3.10) on the unobserved components of utility and which will guarantee that $\hat{y} = y$. We may evaluate $e_j \in e_{j^a}$ as the conditional expectation

$$\overline{(e_j - e_{j^a})} = E[\epsilon_j - \epsilon_{j^a} | \epsilon_{j^a} > x_j^0 - x_{j^a}^0]; \quad (3.11)$$

we refer to this expression as the (conditional) prediction error, otherwise known as the generalised residual. Alternatively we may realise $e_j \in e_{j^a}$ by drawing at random from the conditional distribution of u^a given y

$$f(\epsilon_j - \epsilon_{j^a} | \epsilon_{j^a} > x_j^0 - x_{j^a}^0): \quad (3.12)$$

We may factor either realisation of the random components back into the original (differenced) model, giving $u_j^{ae} - u_j^{ce} = x^0(-j - j^a) + (e_j - e_j^a)$ or $u_j^{ae} - u_j^{ce} = x^0(-j - j^a) + (e_j - e_j^a)$ when based on (3.11) and (3.12) respectively. Hence, $u_j^{ae} - u_j^{ce} < 0 \forall j \in j^a$ for $y = j^a - 2 (1; \dots; J)$ which guarantees that $y^e = \arg \max_j [u_j^{ce}; j = 1; \dots; 3] = j^a$ when calibrated using (3.11). In the same way, we may define a calibrated predictor $y^e = \arg \max_j [u_j^{ae}; j = 1; \dots; 3]$ based on (3.12).

4. Bootstrap-Based Confidence Intervals

In a number of recent studies (see Horowitz (1994) and Godfrey (1997)), analysts have begun to utilise bootstrap techniques to approximate the empirical distribution function, thereby facilitating the construction of confidence intervals which are not dependent upon potentially unreliable asymptotic arguments. To date we have demonstrated how transition frequencies may be estimated using a number of alternate methods without recourse to the underlying distribution of the sample transition frequencies. Given that the distribution of these frequencies is unknown and is likely to exhibit considerable asymmetries, we utilise a parametric bootstrap technique to construct confidence intervals. This is described below.

Transition frequencies are generated according to the following process. Consider the aggregate frequency of movers between states j and j^0 based on the maximum probability rule (2.2);

$$T_{jj^0} = \sum_{i=1}^X 1(u_{ijk}^a > 0) \cap 1(u_{ij^0k}^r > 0) \quad \forall j, j^0 \in k \in 2 \dots J$$

where $u_{ijk}^a = u_{ij}^a - u_{ik}^a$: Since u_{ijk}^a are unobserved estimated aggregate transition frequencies are given by $T_{jj^0} = \sum_{i=1}^P 1(\psi_i = j) \cap 1(\psi_i^R = j^0) \quad \forall j, j^0 \in k \in 2 \dots J$: Parametric bootstrap confidence intervals are constructed using the following algorithm..

1. Based upon an estimation sample of size N , and data pairs $(y_i; x_i)$ we estimate the parameters of the discrete choice model

$$u_j^a = x^0 - j + \epsilon_j; \quad j = 1; \dots; J \quad (4.1)$$

as introduced in section 2.

2. Generate S samples of size N by resampling from the fitted model $f(y; \hat{\beta})$, where $\hat{\beta}$ is a vector of (estimated) mean and covariance parameters. For the s th sample construct two $N \times J$ matrices, u^{s} and Y^s with typical elements k $u_{ij}^{s} = f(x_i; \hat{\beta})$ and k $Y_{jk}^s = 1(u_{ijk}^{s} > 0)$. The s th bootstrap sample comprises the data pairs $(y_i^s; x_i)$, where $y_i^s = \arg \max_j (u_{ij}^{s}; j = 1, \dots, J)$:
3. Given $(y_i^s; x_i)$ estimate the s th bootstrap parameters $\hat{\beta}^s(\hat{\beta})$. Dependent upon the structure of the transition estimator, we then construct Ψ^s :
4. Consider a counterfactual $x_i^R \notin x_i$ and construct the counterfactual analogues of Ψ^s , namely Ψ^{sR} .
5. Form the transition estimator $\hat{T}^s = \Psi^{s0} \Psi^{sR}$
6. For each of the \hat{T}^s transition matrices there are J^2 'separate' transition cells. Let $\hat{T}_l^s(i)$ denote the i th ordered value for the l th cell. An estimate of the p th quantile of T is the $(S + 1)$ th ordered value of \hat{T}_l^s : For $S = 1000$ the 95% bootstrap confidence interval is $P(T_{l(25)} \leq T_l \leq T_{l(975)}) = 0.95$:

5. Computational Results

To compare the various transitions estimators covered in this paper, we propose initially to assess performance within the framework of a simulated sample design. By doing so we enjoy an element of control sufficient to highlight the conditions under which predicted transitions deviate from what we know is the true data generating process. However, we recognise (as do Skeels and Vella (1995)) that Monte Carlo results based on such an approach are clearly design-specific, and may not relate directly to the sorts of economic problems most regularly confronted by applied researchers.

5.1. simulated design

We simulate a trinomial version of the general statistical design described above based on latent variables u_{ij}^a for three states $j = 1, \dots, 3$: These latent variable is assumed to depend linearly on a set of characteristics $x_i = [1; x_{i2}; x_{i3}]'$ through

the parameter vectors $\bar{\gamma}_j$ to give latent relationships of the form

$$u_{ij}^a = x_i^0 \bar{\gamma}_j + \epsilon_{ij}; \quad j = 1; \dots; 3 \quad (5.1)$$

with observability rule

$$y_i = \arg \max_j [u_{ij}^a; \quad j = 1; \dots; 3];$$

and where the unobserved components $\epsilon_i = \epsilon_{i1}; \epsilon_{i2}; \epsilon_{i3}$ are assumed to be independent across i and distributed multivariate normally with mean 0 and variance matrix \mathbf{P} .

For our Monte Carlo simulations we set $x_{i2} \gg N(0; 1)$ and $x_{i3} \gg U(j; 1; 1)$. We choose to parameterise directly the variance matrix \mathbf{P}_{i-1} of error differences $\epsilon_{i-1} = \epsilon_{i2}; \epsilon_{i1}; \epsilon_{i3}; \epsilon_{i1} g^0$. Specifically, we set

$$\mathbf{P}_{i-1} = \begin{matrix} \mathbf{A} \\ \times \\ \end{matrix} \begin{matrix} 1 & 0.5 \\ 0.5 & 2 \end{matrix} :$$

To complete this parameterisation we define the following: $\bar{\gamma}_1 = \epsilon_{11}; 0; 0 g$, $\bar{\gamma}_2 = \epsilon_{11}; 0; 1 g$ and $\bar{\gamma}_3 = \epsilon_{11}; 0; 1 g$. By adjusting the value of the constant term ϵ_{11} , we are able to control the probability of observing $y = 1, 2$ or 3 in the simulations: We choose $\epsilon_{11} = 1.5$; making $y_i = 1$ the most likely state to be observed.

We examine transitions following an impact on x_{2i} . For the simulated design, we predict transitions once x_{2i} is superseded by x_{2i}^R , where

$$x_{2i}^R = x_{2i} + z : (U[j; 1; 1] + \Phi); \quad (5.2)$$

The parameters z and Φ in (5.2) control, respectively, the scale and the direction of the impact on x_{2i} : The larger is z , the larger is the scale of the impact. If $\Phi = +1$ (-1) the impact is entirely positive (negative), whereas for $\Phi = 0$ the impact is balanced.

Monte-Carlo results for this experimental design are reported in Tables 5.1 and 5.2. The sample size is set at $N=500$. We generate 1000 replications of ϵ_i (for fixed x_i) and cumulate transitions matrices following impacts of the form $z = 1, \Phi = 2 (j; 1; 0)$: We report in each table the true frequencies $n_{(j)}$ for states $j = 1; 2; 3 g$, the true transitions frequencies $n_{(j|k)}$ from state j to state k , for

true frequencies:	applies (3.2) to the true parameters \bar{p}_j
maximum probability:	averages \hat{p}_{MP}^s over repeated estimates of b_j^s
exact probability:	averages (3.2) over repeated estimates of b_j^s
resampled probability:	resamples (3.2) using (2.2) over repeated estimates of b_j^s
conditional expectation:	calibrates (2.2) using (3.11) over repeated estimates of b_j^s
single conditional draw:	calibrates (2.2) using (3.12) over repeated estimates of b_j^s
multiple conditional draw:	calibrates (2.2) using multiple (3.12) over repeated estimates of b_j^s

$j; k = f1; 2; 3g$; and a range of measures relative to these benchmarks for the following six transitions estimators:

Tables 5.1 and 5.2 summarise the results of two Monte Carlo experiments, the first relating to the balanced transition shock and the second focussing on the unbalanced design. The first panel in each table lists the true frequencies and the estimated frequencies for each alternative estimator, with empirical standard deviations in parentheses. The second panel records the absolute bias of each transitions frequency relative to the true figure, with 90% empirical confidence bounds (measured relative to each transitions estimate) in square parentheses. Thus, if the absolute bias falls outside the range covered by the empirical confidence bound, the bias can be considered significant at the 90% level.

For the balanced design, the results in Table 5.1 indicate a degree of unreliability in the maximum probability estimator (2.2) for the majority of experimental designs, although biases are not typically significant. Even though the predicted state frequencies are broadly correct, we find a systematic under-prediction of diagonal transitions for most measures based on a maximum probability rule, whilst the estimator derived from exact probabilistic expressions performs well.

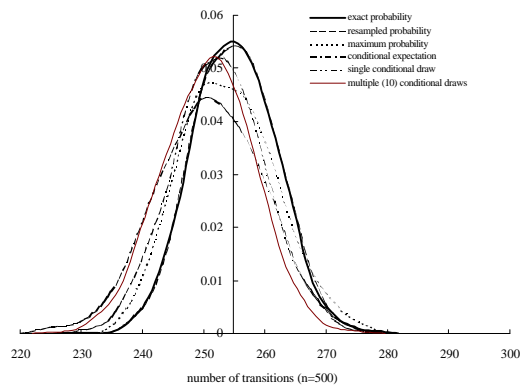
Figures ?? to ?? show the corresponding resampled distributions for each transition probability under the balanced Monte Carlo design.

Table 5.1: Monte-Carlo Results: balanced transition

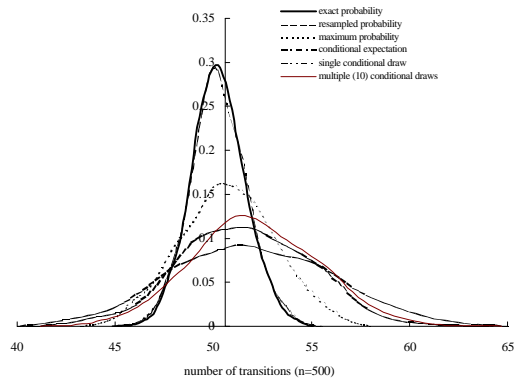
Transitions Frequency	$n_{(1! 1)}$	$n_{(1! 2)}$	$n_{(1! 3)}$	$n_{(2! 1)}$	$n_{(2! 2)}$	$n_{(2! 3)}$	$n_{(3! 1)}$	$n_{(3! 2)}$	$n_{(3! 3)}$
	Monte Carlo estimates, N=500, S=1000 (standard deviations)								
True frequency	254.80	50.60	45.80	25.00	43.00	7.00	19.10	9.90	44.80
resampled freq.	255.10 (6.80)	50.30 (1.40)	45.60 (1.70)	24.90 (1.60)	43.10 (3.70)	7.20 (1.40)	19.20 (1.90)	10.20 (1.80)	44.50 (3.80)
exact formula	255.10 (6.70)	50.30 (1.30)	45.60 (1.60)	24.90 (1.60)	43.10 (3.60)	7.20 (1.40)	19.20 (1.90)	10.20 (1.70)	44.50 (3.80)
max. probability	254.00 (7.80)	50.70 (2.40)	44.40 (3.00)	27.90 (2.60)	42.90 (5.70)	7.70 (2.10)	20.70 (2.10)	9.80 (2.30)	41.80 (5.30)
cond.expectation	252.40 (7.40)	51.60 (3.30)	46.80 (3.50)	26.60 (3.20)	42.00 (4.50)	6.50 (2.00)	21.50 (3.30)	10.50 (2.30)	42.00 (4.50)
single cond. draw	251.10 (8.60)	52.00 (4.10)	47.80 (5.00)	26.50 (3.30)	41.20 (4.70)	7.40 (2.20)	21.30 (3.40)	10.60 (2.50)	42.10 (4.60)
multiple cond. draw	250.80 (7.40)	52.10 (3.20)	48.00 (3.70)	26.50 (3.00)	41.20 (4.40)	7.40 (1.90)	21.20 (3.10)	10.60 (2.10)	42.20 (4.40)
	$n_{(1! 1)}$	$n_{(1! 2)}$	$n_{(1! 3)}$	$n_{(2! 1)}$	$n_{(2! 2)}$	$n_{(2! 3)}$	$n_{(3! 1)}$	$n_{(3! 2)}$	$n_{(3! 3)}$
	absolute bias [90% bootstrap con...dence bounds]								
resampled freq.	0.30 [-8.3,9.0]	-0.30 [-1.7,1.8]	-0.20 [-2.2,2.1]	-0.10 [-2.1,2.1]	0.10 [-4.4,4.9]	0.10 [-1.6,1.8]	0.00 [-2.5,2.5]	0.40 [-2.3,2.2]	-0.30 [-4.8,5.0]
exact formula	0.30 [-8.4,8.9]	-0.30 [-1.6,1.8]	-0.20 [-2.0,2.1]	-0.10 [-2.0,2.0]	0.10 [-4.5,5.0]	0.10 [-1.6,2.0]	0.00 [-2.5,2.3]	0.30 [-2.2,2.1]	-0.20 [-4.8,5.2]
max probability	-0.70 [-10.0,10.0]	0.20 [-3.0,3.0]	-1.40 [-3.0,5.0]	2.90 [-4.0,3.0]	-0.10 [-7.0,8.0]	0.70 [-3.0,2.0]	1.60 [-3.0,2.0]	-0.10 [-3.0,3.0]	-3.00 [-7.0,7.0]
cond. expectation	-2.30 [-9.0,10.0]	1.10 [-4.0,4.0]	1.00 [-5.0,4.0]	1.60 [-4.0,4.0]	-1.00 [-6.0,6.0]	-0.50 [-2.0,3.0]	2.40 [-4.0,5.0]	0.60 [-2.0,4.0]	-2.80 [-6.0,6.0]
single cond. draw	-3.70 [-11.0,11.0]	1.40 [-5.0,5.0]	2.00 [-6.0,6.0]	1.50 [-4.0,5.0]	-1.80 [-6.0,7.0]	0.40 [-3.0,3.0]	2.20 [-4.0,5.0]	0.70 [-3.0,3.0]	-2.70 [-6.0,6.0]
multiple cond. draw	-4.00 [-9.9,9.3]	1.50 [-3.9,4.2]	2.20 [-4.5,4.8]	1.50 [-3.8,3.7]	-1.80 [-5.3,6.0]	0.30 [-2.6,2.4]	2.10 [-4.0,4.1]	0.70 [-2.6,3.0]	-2.60 [-5.8,5.6]

Table 5.2: Monte-Carlo Results: unbalanced transition

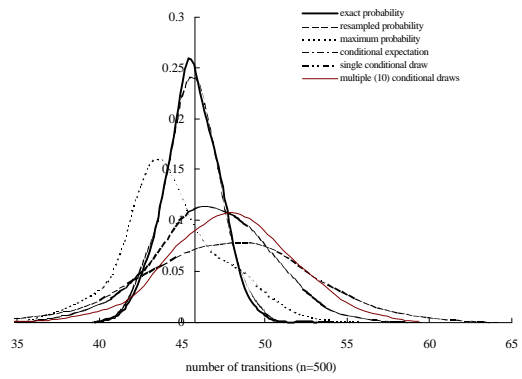
Transitions Frequencies	$\pi_{(1 1)}$	$\pi_{(1 2)}$	$\pi_{(1 3)}$	$\pi_{(2 1)}$	$\pi_{(2 2)}$	$\pi_{(2 3)}$	$\pi_{(3 1)}$	$\pi_{(3 2)}$	$\pi_{(3 3)}$
	Monte Carlo estimates, N=500, S=1000 (standard deviations)								
True frequency	351.20	0.00	0.00	62.10	12.80	0.10	52.10	0.50	21.20
Resampled freq.	350.90 (6.90)	0.00 (0.00)	0.00 (0.00)	61.90 (4.30)	12.70 (2.90)	0.50 (0.70)	52.50 (4.10)	0.60 (0.70)	20.90 (3.60)
Exact frequency	350.90 (6.90)	0.00 (0.00)	0.00 (0.00)	61.90 (4.30)	12.70 (2.90)	0.50 (0.70)	52.50 (4.10)	0.60 (0.70)	20.90 (3.50)
Max. probability	349.20 (9.80)	0.00 (0.00)	0.00 (0.00)	75.50 (7.00)	2.90 (2.20)	0.00 (0.10)	67.90 (6.50)	0.30 (0.60)	4.20 (3.10)
Cond. expectation	350.90 (6.90)	0.00 (0.00)	0.00 (0.00)	71.00 (4.60)	4.00 (2.10)	0.10 (0.30)	63.20 (5.60)	0.40 (0.70)	10.30 (5.00)
Single cond. draw	350.90 (6.90)	0.00 (0.00)	0.00 (0.00)	68.10 (4.60)	6.80 (2.80)	0.20 (0.70)	59.40 (5.90)	0.60 (0.90)	14.00 (5.40)
Multiple cond. draw	350.90 (6.90)	0.00 (0.00)	0.00 (0.00)	68.10 (4.20)	6.80 (1.90)	0.20 (0.60)	59.40 (5.10)	0.60 (0.70)	14.00 (4.70)
	$\pi_{(1 1)}$	$\pi_{(1 2)}$	$\pi_{(1 3)}$	$\pi_{(2 1)}$	$\pi_{(2 2)}$	$\pi_{(2 3)}$	$\pi_{(3 1)}$	$\pi_{(3 2)}$	$\pi_{(3 3)}$
	absolute bias [90% bootstrap con...dence bounds]								
Resampled freq.	-0.20 [-8.8,8.5]	0.00 [0.0,0.0]	0.00 [0.0,0.0]	-0.20 [-5.8,5.4]	-0.10 [-3.2,4.1]	0.40 [-0.2,1.4]	0.40 [-5.5,5.1]	0.10 [-0.3,1.3]	-0.40 [-4.3,4.8]
Exact formula	-0.20 [-8.8,8.5]	0.00 [0.0,0.0]	0.00 [0.0,0.0]	-0.20 [-5.7,5.3]	-0.10 [-3.3,4.1]	0.40 [-0.2,1.3]	0.40 [-5.3,5.3]	0.10 [-0.3,1.2]	-0.40 [-4.3,4.7]
Max. probability	-1.90 [-12.0,13.0]	0.00 [0.0,0.0]	0.00 [0.0,0.0]	13.40 [-10.0,8.0]	-9.80 [-3.0,3.0]	-0.10 [0.0,0.0]	15.70 [-11.0,7.0]	-0.20 [0.0,1.0]	-17.00 [-1.0,6.0]
Cond. expectation	-0.30 [-9.0,9.0]	0.00 [0.0,0.0]	0.00 [0.0,0.0]	8.90 [-6.0,6.0]	-8.70 [-2.0,3.0]	-0.10 [0.0,0.0]	11.10 [-7.0,6.0]	0.00 [0.0,1.0]	-10.90 [-5.0,5.0]
Single cond. draw	-0.30 [-9.0,9.0]	0.00 [0.0,0.0]	0.00 [0.0,0.0]	6.00 [-6.0,6.0]	-6.00 [-3.0,4.0]	0.10 [0.0,1.0]	7.30 [-7.0,7.0]	0.10 [0.0,2.0]	-7.20 [-5.0,8.0]
Multiple cond. draw	-0.30 [-9.0,9.0]	0.00 [0.0,0.0]	0.00 [0.0,0.0]	5.90 [-5.2,5.2]	-5.90 [-2.1,2.6]	0.10 [0.0,0.7]	7.30 [-6.4,6.0]	0.10 [-0.4,1.1]	-7.30 [-4.1,6.3]



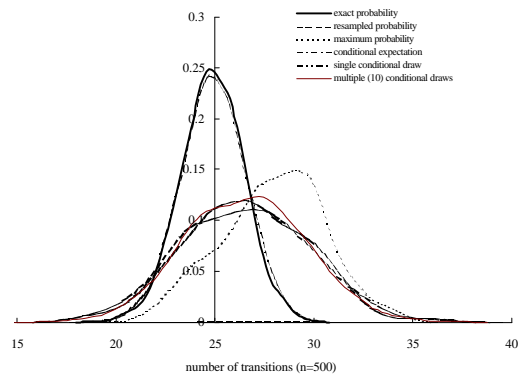
balanced transition: $\Pr(1 | 1)$



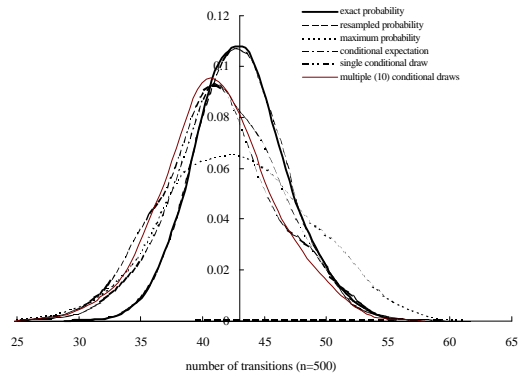
balanced transition: $\Pr(1 | 2)$



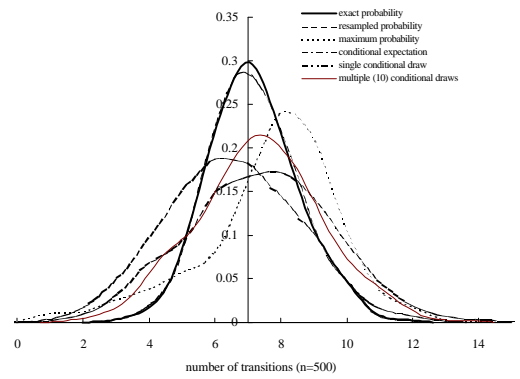
balanced transition: $\Pr(1 | 3)$



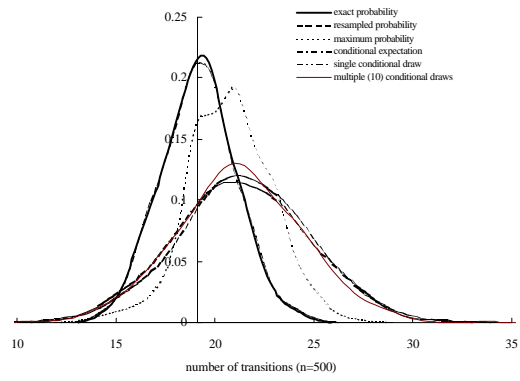
balanced transition: $\text{Pr}(2 | 1)$



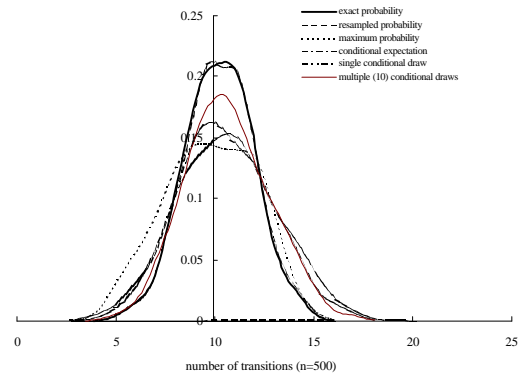
balanced transition: $\text{Pr}(2 | 2)$



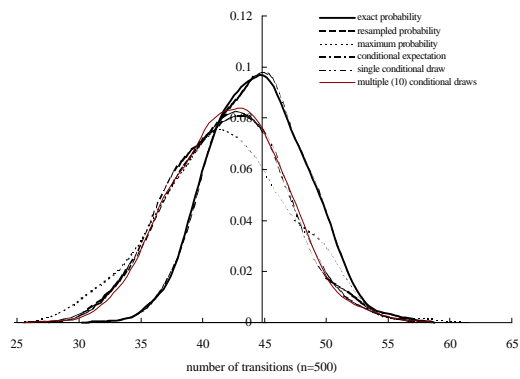
balanced transition: $\text{Pr}(2 | 3)$



balanced transition: $\text{Pr}(3 | 1)$

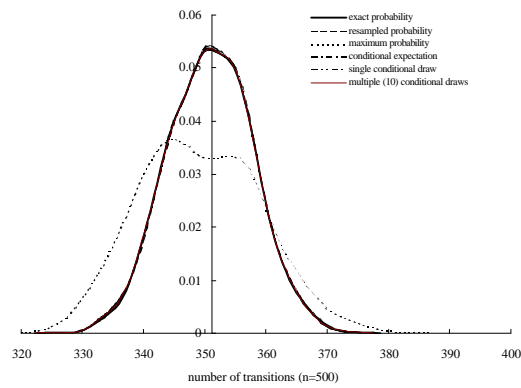


balanced transition: $\text{Pr}(3 | 2)$

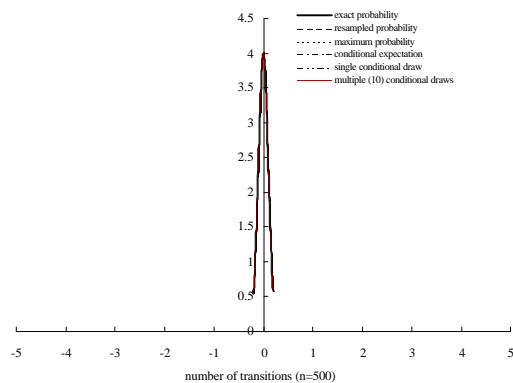


balanced transition: $\text{Pr}(3 | 3)$

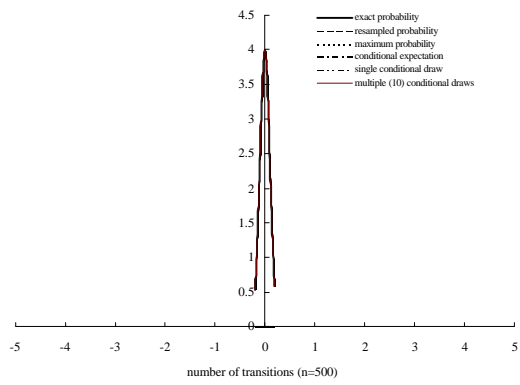
Table 5.2 reports similar results for the unbalanced design, with more severe biases relative to the true frequencies for the maximum probability rule and calibrated estimators. Notice that, when the design becomes unbalanced, the predicted state frequencies from T_{MP} deviate markedly from the truth. As noted by Windmeijer (1995), the use of the naive maximum probability rule tends to under-predict (over-predict) the sparse (dense) state due to the wasteful nature of the metric which translates the predicted probability into a discrete state prediction. The over-prediction of σ -diagonal transitions by T_{CE} is also manifestly clear, with significant errors for the unbalanced designs. Again, the alternative transitions estimators derived from exact probability measures perform well.



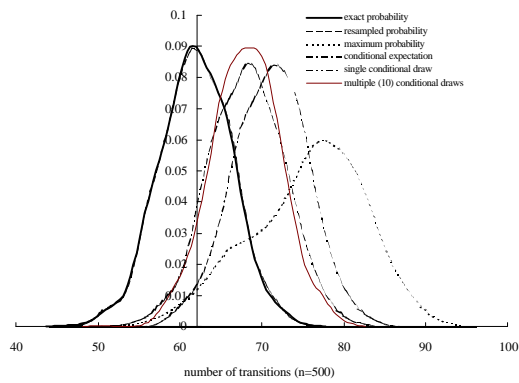
unbalanced transition: Pr(1 ! 1)



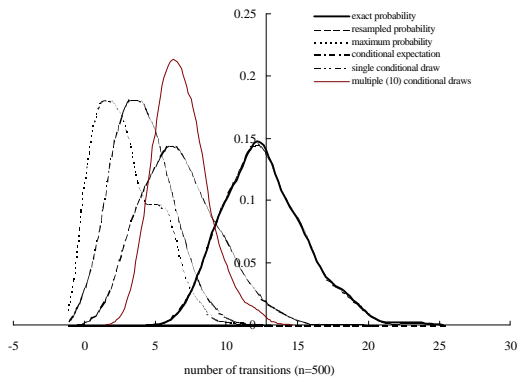
unbalanced transition: Pr(1 ! 2)



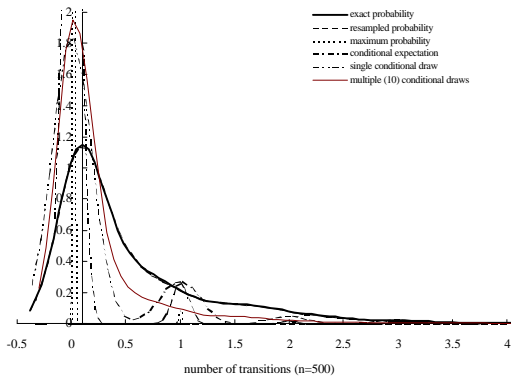
unbalanced transition: $\Pr(1 \neq 3)$



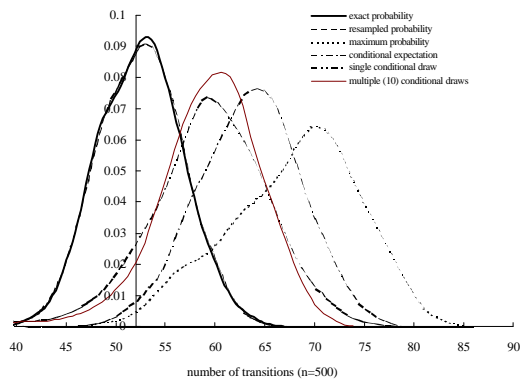
unbalanced transition: $\Pr(2 ! 1)$



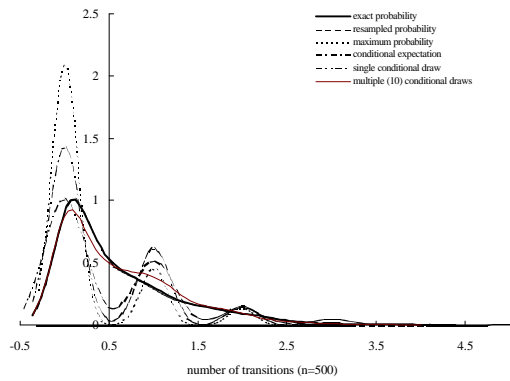
unbalanced transition: $\Pr(2 ! 2)$



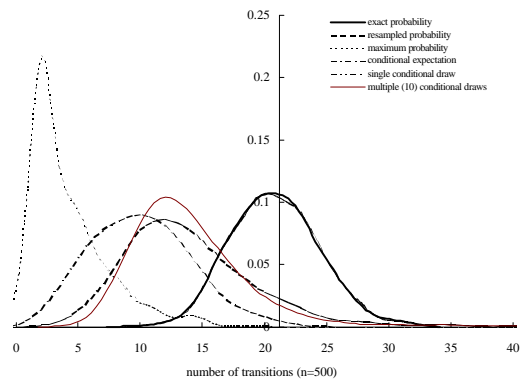
unbalanced transition: $\Pr(2 ! 3)$



unbalanced transition: $\Pr(3 \neq 1)$



unbalanced transition: $\Pr(3 \neq 2)$



unbalanced transition: $\Pr(3 \neq 3)$

Repeating these simulations over different sample sizes and for transitions impacts of different magnitudes, the same overall patterns are observed.

6. Conclusions

We focus in this paper on various methods by which transitions following an exogenous shock to the deterministic component of models of discrete choice. Concentrating on low-dimensional problems, we have been able to confirm the bias of naive discrete state predictors based on the standard maximum probability rule, and show that the problems associated with such predictors extend to the reliability of standard methods by which transitions are simulated. In a three-dimensional framework, explicit forms for transitions frequencies have been proposed, the reliability of which have been confirmed using simulated Monte-Carlo designs. We exploit various methods by which discrete choice models may be calibrated to offer alternative transitions estimators which perform well in Monte-Carlo simulations. An extension is suggested which offers a robust and unbiased alternative to transitions estimator in higher-dimensional problems.

References

- Bingley, P., and I. Walker (1995): "Labour Supply, Unemployment and Participation in in-work Transfer Programmes," Discussion paper, Institute for Fiscal Studies.
- Cramer, J. (1991): *The Logit Model*. Edward Arnold.
- Duncan, A., and C. Giles (1996): "Labour Supply Incentives and Recent Family Credit Reforms," *Economic Journal*, 106, 142–155.
- Hausman, J., and D. Wise (1978): "A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognising Interdependence and Heterogeneous Preferences," *Econometrica*, 46, 403–26.

Keane, Michael, P. (1997): "Modeling Heterogeneity and State Dependence in Consumer Choice Behaviour," *Journal of Business and Economic Statistics*, 15(3), 310–327.

Skeels, C., and F. Vella (1995): "Monte Carlo evidence on the robustness of conditional moment tests in Tobit and Probit models," Discussion paper, mimeo, Australian National University.

Windmeijer, F. (1995): "Goodness-of-Fit Measures in Binary Choice Models," *Econometric Reviews*, 14, 101–116.